



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: A Quantitative Assessment of Shoeprint Accidental Patterns with Implications Regarding Similarity, Frequency and Chance Association of Features

Author(s): Jacqueline A. Speir

Document Number: 251522

Date Received: February 2018

Award Number: 2013-DN-BX-K043

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.



West Virginia University®

**EBERLY COLLEGE OF ARTS AND SCIENCES
FORENSIC & INVESTIGATIVE SCIENCE**

Final Technical Report: 2013-DN-BX-K043

Project Title: A Quantitative Assessment of Shoeprint Accidental Patterns with Implications Regarding Similarity, Frequency and Chance Association of Features

Principal Investigator: Jacqueline A. Speir, West Virginia University
208 Oglebay Hall, PO Box 6121, Morgantown, WV 26506
Jacqueline.Speir@mail.wvu.edu, 304.293.9233

Submitting Official: John T. Childress; JTChildress@mail.wvu.edu
Recipient Organization: West Virginia University Research Corporation, 866 Chestnut Ridge Road, PO Box 6845, Morgantown, WV 26506, **DUNS:** 191510239, **EIN:** 550665758

Award Period: 01/01/2014 - 05/31/2017
Reporting Period End Date: 05/31/2017 - *Final Report*

Keywords: shoeprints, footwear, randomly acquired characteristics, similarity, frequency, random/chance co-occurrence

Table of Contents

1.	Executive Summary	1
2.	Research Overview	4
2.1	Review of Selected Works	4
2.2	Research Methodology & Results	7
2.2.A	Data Acquisition	7
2.2.B	Pre-processing of High Quality Prints	10
2.2.C	Registration	13
2.2.D	Segmentation	15
2.2.E	Processing	18
2.2.F	Outsole Size & Shape Normalization	24
2.2.G	Shape Descriptor	26
2.3	Similarity Assessment	28
2.3.A	RAC Loss	29
2.3.B	Comparison	31
2.3.C	Individual RAC Similarity	34
2.3.D	Similarity as a Function of RAC Shape	34
2.3.E	Similarity as a Function of RAC Size	35
2.4	Differentiating between KM and KNM Crime Scene RACs using Similarity Metrics	41
2.4.A	RAC Map Correlation	45
2.5	Chance Co-occurrence	47
2.5.A	Use of Similarity in the Web Application	50
2.6	Impact, Outcomes, Evaluation & Dissemination	52
2.7	Added Value	52
2.8	Publications & Abstracts	53
2.9	Meetings, Presentations & Invited Talks	55
2.9.A	Limitations	56
A.	Appendices	57
A.1	Bibliography	57
A.2	Author Response to Reviewer Comments	59

1. Executive Summary

The National Academy of Sciences (NAS) 2009 report on *Strengthening Forensic Science in the United States* revealed several research recommendations related to forensic footwear examinations, including the need for greater clarity concerning the variability of outsole class and individual (randomly acquired) characteristics (RACs), the validity and reliability of current methods and practices, the relative frequency of features, and the appropriate use of statistical standards (NAS, 2009). In response to this request, this project performed foundational research to clarify the empirical frequency and shape distribution of randomly acquired characteristics on outsoles collected from a general population.

To achieve this goal, an outsole database was generated, resulting in summary statistics and frequency estimates on 72,306 randomly acquired characteristics extracted from 1,300 outsoles. The subsequent results are based on a combination of automated and analyst-derived image extraction and processing tools, with the human-dependent step of RAC detection and marking. Given some unavoidable subjective steps in the image processing chain, inter- and intra-analyst variability in RAC marking was assessed using a quality control/assurance program that included the duplicate marking of 5,477 randomly acquired characteristics across 160 shoes (320 RAC maps). The results indicate that RAC *detection* is the largest variable not easily controlled (even with training), but when RACs are equally detected in repeat analyses, they are marked *relatively consistently*, with mean polar coordinate localization differences of less than $r \pm 0.2\text{mm}$, and $\theta \pm 0.1^\circ$, and shape attribution (*e.g.*, isometric, elongated or irregular) agreement nearly 75% of the time.

Post-detection and extraction, each RAC was broadly characterized in terms of its degree of linearity, circularity and triangularity. Using geometric shape classification rules, automated shape attribution was compared to human-perceptual assignments and found to be in agreement between 68% to 95% of the time, across 1,352 comparisons, and depending on the complexity of the dataset presented for analysis. Overall, the results indicate limited utility in classifying complex features into prescribed shape classes (such as circles, lines, curves, rectangles, triangles, etc.), and that future work should consider alternative mechanisms (such as shape clustering), as opposed to strict categorization, as a means of grouping randomly acquired characteristics in terms of shape similarity.

Next, outsole size and shape normalization was performed. This step, although not ideal, was deemed unavoidable in order to create sufficient power in the inter-comparison of all 1,300 shoes in the database, regardless of outsole style/shape and size. Post normalization, each RAC was localized to one of 990 possible spatial bins, each 5mm x 5mm in size. Post-localization and binning, estimates of co-occurrence and similarity were possible. This was accomplished by computing the Fourier descriptor of each RAC, and for RACs with positional co-occurrence, pairwise comparisons were performed using five similarity metrics (Euclidean distance (ED), Hausdorff distance (HD), modified cosine similarity (MCS), matched filter (MF), and modified phase only correlation (MPOC)). Variation in similarity score as a function of RAC shape, perimeter and area were computed and are reported, along with receiver operator and cumulative match characteristic curves that provide insight on the use of numerical metrics

to rank-order RACs from different sources. Results indicate superior performance with distance metrics (HD and ED), making Hausdorff distance the best candidate (of those metrics compared) for computing score-based likelihood ratios. More specifically, it was noted that both HD and ED had statistically indistinguishable AUCs (area under the curve) of 0.82, and that both were significantly better than MCS, MF and MPOC. However, alternative metrics, including deep learning, might prove equally or more useful, and additional work is needed to fully appreciate the strengths and weaknesses associated with the use of numerical shape comparisons within the field of forensic footwear examinations.

Equipped with RACs with known positional co-occurrence and shape similarity, three questions related to chance co-occurrence were asked. First, what was the empirical frequency of finding a pair of RACs with positional similarity anywhere on an outsole within this dataset? Second, what was the empirical frequency of selecting two shoes at random and finding shape similarity at a specific location? Lastly, what kind of numerical/quantitative similarity is expressed by RACs with positional co-occurrence?

With regard to positional co-occurrence *anywhere* on an outsole, the empirical frequency is extremely high (1 in 4 for elongated features, to 1 in 10 for isometric features). However, when positional co-occurrence in a specific location is queried, median results range from 1 in 2,080 for elongated features, to 1 in 9,279 for irregularly-shaped features. In addition, the worst case scenario (greatest chance association) was found to be 1 in 281 for elongated features, while the best case scenario (lowest chance association) was found to be 1 in 844,350 or better (this value is limited by the size of the database).

However, RACs with positional co-occurrence (and even identical shape categorizations), are not necessarily geometrically similar (*e.g.*, two linear elements could vary in orientation, length, thickness, curvature, etc). Thus, the mathematical similarity of RACs with coincidental positional and shape similarity were computed based on 6,993 known match comparisons, and 3,239,114 known non-match comparisons. The results indicate that 13% of known non-matches have likelihood ratios (LR) greater than 1.0, but even for known non-match RACs with some degree of *numerical similarity*, very few were found to be *visually indistinguishable*. In fact, to assess the possibility of numerical versus visual confusion, a subset of 19,800 of the most similar RACs from different sources (1,000 shoes) were visually compared. More specifically, all pairwise comparisons for RACs from a subset of 1,000 shoes were compared and ranked (for a total of 2,022,595 known non-matches), and the five most similar RACs per spatial bin were examined to determine visual differentiability. Of the almost 20,000 visual comparisons that were performed, all but 25 pairs were deemed distinguishable based on RAC geometry (orientation, size, shape, complexity, etc.), with an associated probability of confusion on the order of 1.2E-05 (or 0.001% of the time, assuming an effective ranking). Moreover, when the 25 indistinguishable *RAC pairs* were further characterized, all were found to be differentiable based on shoe *class characteristics* (differences in make/model, size and/or degree of wear).

In conclusion, there is evidence to assert that RACs possess a high degree of forensic discrimination potential. However, the widespread and general applicability of any associated probability and chance association computed based on this study must be considered within the confines that bound the research dataset and methodology. More specifically, all results are

a function of the nature of the footwear population studied, which was predominately athletic gear (86%), men's wear (72%), and of sizes 9 through 11 (53%). Moreover, all shoes have been inter-compared without regard for class characteristics, which required the *less-than-ideal* step of normalization as a function of outsole size and shape. As such, the probability of confusion, reported to be on the order of 1.2E-05, must be interpreted within the confines of this footwear population, and with full understanding of the nature of the data analyses that lead to these conclusions.

2. Research Overview

This work proposed five deliverables:

- To analyze and characterize more than 400 exemplar and questioned prints (referred to as the ‘footwear database’);
- To subsequently pairwise compare exemplar RACs and RAC maps using a quantitative metric (correlation);
- To provide frequency estimates relating RAC ‘type’ and location;
- To report on the random co-occurrence of accidental features;
- To create a user-friendly graphical interface that allows the analyst to rapidly extract similarity and frequency information from the database, including quantitative measures of similarity and estimates of chance co-occurrence of features in terms of geometry and positional information.

To date, the following products have been realized:

- The collection (purchase and donation) of 2,028 outsoles;
- The analysis and characterization of 1,300 outsoles (900 more than originally contracted to deliver);
- The quantitative comparison of RAC similarity using 5 metrics (4 more than originally contracted to deliver, including the strengths and weaknesses associated with each);
- A web-based application that provides frequency estimates describing RAC type and location (<http://www.4n6chemometrics.com/database/>);
- A web-based application reporting on the random co-occurrence of accidental features (<http://www.4n6chemometrics.com/database/>);
- Long-term and on-going soft-benefits associated with an increased understanding of fundamental phenomenology such as the spatial prevalence and coincidental association of accidental features and feature patterns in random and specific footwear populations.

2.1 Review of Selected Works

The power associated with demonstrating a linkage between footwear and an impression left at the scene of a crime is directly related to the perceived rarity of the shoeprint itself, which is a function of observed class and accidental characteristics (including clarity and quality). When individualizing characteristics are present, their relative position, orientation, size and

shape are examined and compared with known exemplars (SWGTHREAD, 2006) in an effort to formalize the strength of the suspected linkage. However, the degree to which a feature, or a collection of features, might repeat is less well understood. This latter issue is referred to as chance co-occurrence (or the random match probability (RMP) in DNA/population genetics), which inevitably impacts the discrimination potential associated with any form of forensic evidence, including forensic footwear analysis.

Within the pattern sciences, the likelihood of a close non-match is often formalized by an examiner’s accumulated expertise, wherein an analyst, with years of experience, develops internalized knowledge as to the likelihood that a feature (or set of features) would reproduce by random chance alone between two known non-matches. Despite the reasonable validity associated with using accumulated expertise to inform evidence interpretation, the need for external theoretical models and empirical investigations to support these inferences still exists. Unfortunately, providing objective data in support of these endeavors is not an easy feat, and most attempts are constrained as a function of model assumptions and sample size (wherein the latter obstacle should not be underestimated since it is extremely difficult and time-consuming to obtain a sufficiently large sample size (or footwear database) for reliable estimation).

Past efforts in support of increased knowledge concerning close non-matches and chance association are based on theoretical models and small/modest-scale empirical research efforts, as well as studies with alternative goals (not necessarily related to answering the question of chance association, but by virtue of research design, have shed light on this question) (Cassidy, 1995; Champod et al., 2000; Davis and DeHaan, 1977; Davis and Keeley, 2000; Fawcett, 1970; Hannigan et al., 2006; Petraco et al., 2010; Sheets et al., 2013; Shor and Wiesner, 2015; Skerrett et al., 2011; Stone, 2006; Wilson, 2012). Although not an exhaustive review, the remainder of this section will highlight three notable works ((Fawcett, 1970), Stone (2006), and Cassidy (1995)) that provided motivation for this project.

To begin, Fawcett (1970) modeled the chance agreement of accidental characteristics between test and crime scene impressions using the binomial coefficient. The model determines the number of ways of obtaining (x) unordered accidental marks in a scene impression from a total of (s) possible defects using Eq. (1).

$${}_s C_x = \frac{s!}{(x-s)!(s)!} \quad (1)$$

To account for variation in resolution, as well as aberrant pseudo-accidentals that are a function of contamination of the questioned impression, differences in scene versus exemplar prints were modeled according to Eq. (2) where (z) represents the number of features in the questioned print and (p) defines the number of scene and exemplar accidentals that correspond.

$$\text{Probability of Chance Coincidence} = \frac{{}_s C_x}{\sum_p ({}_z C_p) ({}_{s-z} C_{x-p})} \quad (2)$$

Although Eq. (2) can be simplified by considering a single term in the denominator (where

$p = p$ *only*, instead of summing from $p = p$ to $p = x$ or z), Fawcett (1970) states that the computation will always underestimate the true probability since the metric is binary in nature (*e.g.*, a defect is present or absent in the questioned and exemplar impressions, without regard for shape or variation in size).

In a comparable approach, Stone (2006) purported to compute the theoretical probability of the random duplication of accidental features with increasing degrees of complexity (points, lines, curves, etc.) but assuming questioned impressions of full resolution and free of pseudo-accidentals. The model employed by Stone (2006) assumes a hypothetical flat-soled men's size 8.5 shoe with a working area of $16,000\text{mm}^2$ and a 1mm^2 resolvable limit. Using these model parameters, Stone (2006) computed a 1 in 16,000 probability of random duplication of a single point feature, defined as a defect with 'no discernible length or width'. This computation was repeated for larger defects (such as lines and curves), that possess a greater number of attributes including binned estimates for length, orientation and curvature. Although the results are extremely useful and intuitive, Stone (2006) acknowledges the many ways in which a theoretical model can diverge from reality when the complexity and variation of empirical data are considered. For example, the computed probability is a function of the model assumptions, impacted by model violations such as (1.) shoes with surface areas less than $16,000\text{mm}^2$, (2.) resolution capabilities less than 1mm^2 , (3.) outsoles with a raised instep or unequal contact with terrain, and (4.) observations suggesting that accidentals are not equally likely to occur on the entire outsole (as a function of a raised instep or otherwise). Despite this, a model should not be invalidated simply because it is bounded. In fact, bounding is typically a model requirement; partially driven by the need to formulate simplifying assumptions, and partially based on our understanding of phenomenology. However, the physics implicit in the model should be continually expanded using empirical observations. For example, crime scene prints are often partial impressions deposited in variable media and therefore of variable resolution. As such, the resolution capability of 1mm^2 may not be reasonable under all circumstances. In addition, there is considerable known variation in shoe size, characteristics, and the likely capture of an accidental. Consider the evidence provided by Bodziak (2000) from a survey of 450 athletic shoes showing that outsoles that differ in size between 6 and 12 vary in length between approximately 3.8 and 5.5cm. Similarly, Davis and DeHaan (1977) reported evidence based on a random analysis of 650 pairs of men's shoes showing that shoe sizes that differ between 6 and 11 can lead to variation in heel and sole widths of approximately 3cm. Moreover, Davis and DeHaan (1977) report that this sample highlighted many shoes with deep 'cleats', thereby reducing the area of the outsole susceptible to damage. This is compounded by raised insteps that can further limit the surface area of a shoe in contact with the terrain. In addition, the study by Davis and DeHaan (1977) reported that defects were rarely encountered in the rear and central portions of the heel, which the authors attribute to areas of high wear and continual erosion. As a corollary, Davis and DeHaan (1977) attribute a higher significance to any accidental that is found on the mid- to rear-section of the heel.

In contrast with the theoretical models presented by Fawcett (1970) and Stone (2006), Cassidy (1995) attempted to answer questions concerning chance reproduction of individual characteristics based on empirical studies. Using groups of police recruits engaged in activities that promote the chance reproduction of accidental characteristics, Cassidy (1995) concluded that there was a 1 in 6 chance of finding 10 'minute' characteristics and a 1 in 20 chance of finding

3 moderate-sized characteristics possessing *coincidental similarity in position* for the heel of compared shoes. However, Cassidy (1995) acknowledged that similarities based on mold may very well be included in these figures, and an extensive discussion of the similarity in feature *quality* was not explicitly provided. Similar empirical observations collected from two different sets of test shoes (presumably less likely to exhibit mold subclass characteristics) suggest that a single moderate-sized characteristic has a 1 in 38 to a 1 in 60 chance of possessing coincidental similarity with known non-match heels (Cassidy, 1995). Of course, the heel of a shoe is only a fraction of the size of the shoe modeled by Stone (2006), but even assuming that the heel is a fourth of the total area of the outsole, the empirical results presented by Cassidy (1995) suggest a higher probability of chance similarity in position. Although one could argue that this difference is a function of the shared activities performed by the police recruits (*e.g.*, activities believed to favor the duplication of shared features), this is only speculative since an empirical baseline does not exist that can describe the chance duplication of random characteristics for individuals and activities that are unrelated or random (which this research project addresses).

2.2 Research Methodology & Results

Characterizing the presence, geometric shape and utility of a randomly acquired characteristic is not, in fact, a straightforward process. Instead, the physical evidence record available for analysis is typically corrupted by a host of factors that can include variations in material properties, deposition conditions, temporal and spatial factors, as well as attempts at collection and enhancement (not to mention the size, quality and clarity of the feature under consideration). With this in mind, the major focus of this project was to report on similarity scores and population frequency estimates that result when comparing *high quality (HQ) (exemplar)* known matches (KM) and known non-matches (KNM), with a limited focus on the increased variability that results when considering mixed-media and crime scene-like (CS) conditions. In order to generate sufficient data upon which to argue some pretext in terms of statistical power, a large footwear database was needed, therefore demanding a rigorous approach to data (1.) *acquisition*, (2.) *pre-processing*, (3.) *registration*, (4.) *segmentation*, (5.) *processing*, and (6.) *comparison*. The remainder of this section will expand upon each of these steps.

2.2.A Data Acquisition

Footwear was obtained by donation, request, and purchases from Goodwill[®] and similar thrift stores. Table 1 reports the total number of shoes procured by the research group (2,028), including source. Note that nearly half (45%) resulted from purchases, while the remaining 55% were acquired based on generous corporate donations.

Of the 2,028 footwear collected, 1,300 have been fully characterized. Tables 2 - 6 summarize the attributes of the resulting footwear dataset as a function of type, degree of wear, the presence/absence of features (such as microcellular material and/or Schallamach patterns), manufacturer, and finally, gender/size.

Table 1: Break down of footwear by source.

Source	Total Shoes	~ Percent %
Goodwill®	914	45%
Decker	72	4%
Nike	538	27%
Reebok	274	13%
Under Armour	230	11%
Total	2,028	100%

Table 2: Footwear database by shoe type.

Type	Number	~ Percentage %
Athletic	1,122	86%
Dress	89	7%
Boot	69	5%
Sandal	20	2%
Total	1,300	100%

Table 3: Footwear database by degree-of-wear . Note that ‘light’ describes an outsole with texture still present throughout most of outsole, ‘moderate’ describes an outsole with texture that remains apparent but may be accompanied by small bald spots, and ‘heavy’ describes an outsole nearly devoid of any remaining texture, many or large bald spots, and possible holes where the sole has been worn away.

Degree of Wear	Number	~ Percentage %
Light	324	25%
Moderate	673	52%
Heavy	303	23%
Total	1,300	100%

Table 4: Footwear database as a function of microcellular material and Schallamach patterns.

Microcellular Material	Number	~ Percentage %
Present	155	12%
Absent	1,145	88%
Total	1,300	100%
Schallamach Pattern	Number	~ Percentage %
Present	921	71%
Absent	379	29%
Total	1,300	100%

Table 5: Footwear database as a function of manufacturer. Note that ‘other’ reports on shoes from manufacturers that individually contributed a very small number of shoes (less than 1% of the total), but in totality, define 23% of the database.

Manufacturer	Number	~ Percentage %
Adidas	32	2%
Asics	30	2%
Brooks	10	1%
Converse	36	3%
Hoka	36	3%
New Balance	22	2%
Nike	513	39%
Puma	15	1%
Reebok	152	12%
Skechers	16	1%
Under Armour	114	9%
Unknown	24	2%
Other	300	23%
Total	1,300	100%

Table 6: Footwear database as a function of intended gender and manufacturer’s reported size. Note that 8% of the outsoles in this database have unknown gender and/or size, and that each row reports data for both the whole and half-size (e.g., 10 and 10.5).

Men’s Size	Number	~ Percentage %	Women’s Size	Number	~ Percentage %
5	4	<1%	4	4	<1%
6	10	<1%	5	2	<1%
7	34	3%	6	18	1%
8	93	7%	7	69	5%
9	215	17%	8	82	6%
10	271	21%	9	51	4%
11	193	15%	10	23	2%
12	87	7%	11	10	<1%
13	30	2%	12	2	<1%
Total	937	72%	Total	261	20%

2.2.B Pre-processing of High Quality Prints

As necessary, each shoe was gently washed (using warm water) to remove debris (*i.e.*, this research does not account for the possible presence of transient RACs, such as rocks, gum, etc.). When dry, each outsole was scanned at 600PPI with an Epson Expression 11000XL Graphic Arts Scanner. Post-outsole scanning, Handiprint exemplars were created (Bodziak, 2000) using a Zephyr[®] brush (A-1-0200 Arrowhead Forensics, trimmed to a total length of approximately 1 inch), Lightning[®] Black Powder (1-4005 CSI Forensic Supply) and Handiprint sheets with clear polyester covers (2-3150 CSI Forensic Supply). To create each exemplar, the Handiprint sheet was prepared by removing the clear polyester sheet and allowing the flexible Handiprint material to rest (reform shape, adhesive side-up) while lightly dusting the outsole with the powder and Zephyr[®] brush.

During powder application, the outsole was brushed in at least three directions; North-South (toe/heel), East-West (medial/lateral) and diagonally to ensure full coverage. After dust application, the shoe was tapped three-four times to dislodge excess dust, before placing the outsole on top of the prepared Handiprint sheet sitting on the laboratory benchtop. The Handiprint+shoe combination was slowly pulled off of the benchtop toward the analyst, while the researcher used his or her hands to gently add pressure on the non-adhesive side of the Handiprint (pressing the outsole against the tacky side of the Handiprint to maximize tight contact). When the Handiprint+shoe was fully removed from the laboratory benchtop, the analyst then used a paper towel or fingerprint roller to gently reapply pressure between the Handiprint and outsole to again maximize contact. When complete, the Handiprint was pulled from the outsole and laid flat on the benchtop. The clear polyester cover was then slowly re-applied from bottom to top in a type of rastering process to minimize the introduction of air pockets between the Handiprint and protective cover. After development, the Handiprint was likewise scanned at 600PPI; both are illustrated in Fig. 1 for a size 9 men’s Converse Chuck Taylor[®] All Star[®] with moderate wear and Schallamach patterns.

Crime Scene-Like Images

Using a random number generator, 50 pairs of shoes were selected from the aforementioned exemplar database. Five analysts of differing height, weight and shoe size were selected and randomly assigned 10 pairs of shoes to aid in print creation. In order to best replicate crime scene conditions, each analyst wore his or her randomly assigned shoes when creating impressions (*note that this methodology differed from that used in exemplar creation which entailed pressing a dusted outsole against an adhesive sheet*). Each outsole was lightly covered with shoe polish and analysts walked four steps over clear acetate sheets, thereby creating two replicate impressions per shoe for a total of 200 crime scene-like quality prints. Each impression was then developed using black magnetic powder (Lightning Powder Co. Black Magnetic 1-0160) and lifted using white gelatin lifters (13cm x 36cm BVDA Gellifters, Batch no. 2015033). Fig. 2 illustrates one ‘best case’ and one ‘worst case’ reproduction scenario. *Note: the authors acknowledge that the crime scene-like impressions created in this dataset are still far superior than prints collected at the majority of real scenes by actual examiners.*



Figure 1: Example of outsole and Handiprint exemplar scans (size 9 men's Converse Chuck Taylor[®] All Star[®]).

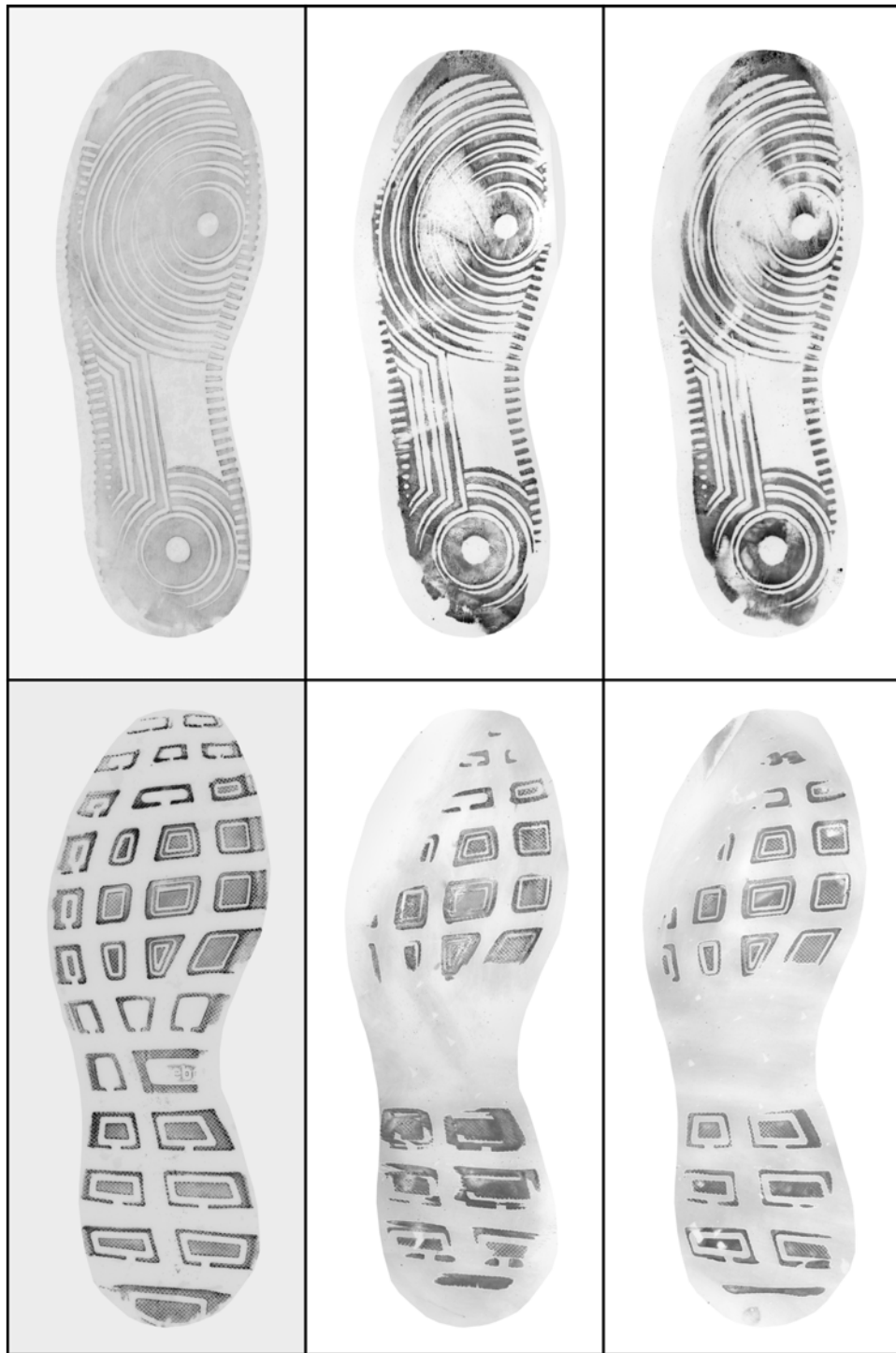


Figure 2: A 'best case' scenario (top row) and a 'worst case' scenario (bottom row) for crime scene-like impression production. Handprint exemplar (left) and two crime scene-like replicates (center, right). Note: the authors acknowledge that the crime scene-like impressions created in this dataset are still far superior than prints collected at the majority of real scenes by actual examiners.

Pre-Processing of Crime Scene-Like Prints

After lifting, all crime scene impressions were likewise scanned at 600PPI with the Epson Expression 11000XL Graphic Arts Scanner. The lifters were affixed to a scanning board designed to raise the gel surface off the scanner bed by less than 1mm, thus allowing for clear, focused prints, without direct interaction between the lifter and the scanner's glass surface. After scanning, lifts were covered and stored for future reference.

2.2.C Registration

In order to facilitate the automated downstream extraction of RAC shape and position, the outsole and exemplar (and crime scene-like images) were background subtracted and registered using identified control points. This process required the analyst to identify eight common geometric shapes that were patent on both the outsole and the exemplar. The features selected for registration varied per shoe, but needed to be distributed as evenly as possible around the perimeter of the outsole (a minimum of two on the toe, two on the heel, and the remaining four on the lateral and medial sides of the shoe) and generally consisted of class characteristics with sharp boundaries, such as corners in polygonal-geometric shapes (and lettering in logos, if applicable).

To expedite this process, a simple graphical user interface was constructed that opened two paired images (the scanned version of the outsole, and the mirrored version of the Handiprint exemplar). With both images in a common orientation, the analyst used the cross-hair of the cursor on his or her mouse to designate mated-points between the images (open windows). Using this process, any number of mated points could have been selected, but as a compromise in terms of efficiency and accuracy, eight total ground control points were selected. Of the two possible images to use as a base, the outsole was selected, which meant during transformation, the Handiprint exemplar was translated, rotated and scaled (as necessary) to bring it into registration with the outsole. This transformation was performed using a first order polynomial with least-squares fitting (note that a first order polynomial was selected over an affine transformation in order to handle slight shearing in the toe and heel that is not uncommon when creating Handiprint exemplars). Note that registration of crime scene-like images was accomplished using the same methodology.

In addition to this co-registration, the background (non-tread areas) of both the outsole and exemplar (and crime scene-like images) were removed. This was accomplished in a rather rudimentary or primitive way, using the aforementioned graphical user interface, wherein the analyst simply traced the perimeter of the outsole using the cross-hair of the cursor, thus automatically generating a binary image that labeled every pixel as either belonging to the outsole or belonging to the background. Once generated, this map was saved and mathematically multiplied with other images downstream (*e.g.*, the outsole and Handiprint exemplar) to effectively increase image signal to noise ratios. As such, the background (or non-tread areas) of both the outsole and exemplar were removed (Fig. 3) to ensure the highest quality imagery moving forward (*e.g.*, removal of remnants of the analyst's hands that may have been captured during scanning when pressure was applied to the outsole to promote a nearly planar surface, and/or removal of extraneous dust and fingerprints on Handiprint exemplars).

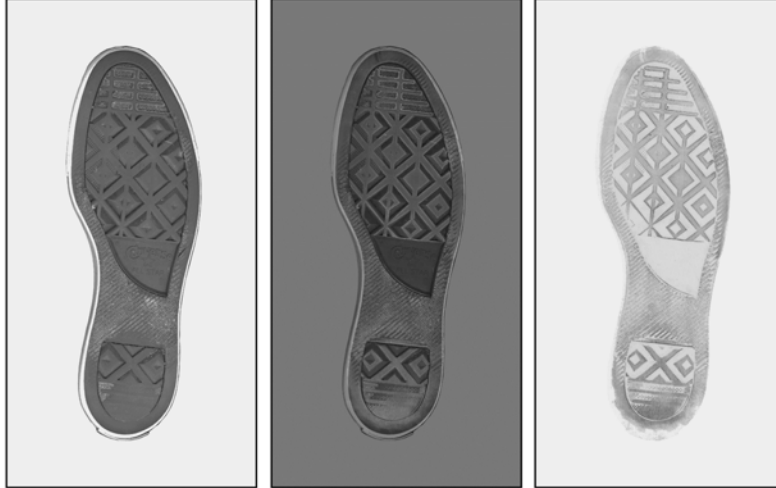


Figure 3: Registered and background subtracted outsole scan (left) and Handiprint scan (right). The middle image is an overlay of the outsole and Handiprint illustrating co-registration (size 9 men's Converse Chuck Taylor[®] All Star[®]).

Finally, the outsole and exemplar (and crime scene-like images) were collectively translated and rotated to ensure that all were centered within the image frame (8,961 x 8,961 pixels) and oriented such that the long-axis of the shoe (toe-to-heel) was North-South. This was most easily accomplished using the binary image that was created in the previous step, wherein each pixel was defined as either outsole or background. From this image, the midpoint of the outsole was mathematically computed (x_o, y_o) , defined as the x-pixel halfway between the maximum width of the shoe and the y-pixel halfway between the maximum length of the shoe. Since the image frame was 8,961 x 8,961 pixels, the image frame center was located at pixel coordinate (4,481,4,481), so the outsole and Handiprint exemplar images were centered by translating the imagery such that (x_o, y_o) was coincident with (4,481, 4,481).

To ensure that the shoe's long-axis was North-South, the binary map defining outsole versus background was treated as a bivariate normal distribution, amenable to eigen-decomposition. After decomposition, the resulting eigen-vectors defined the major and minor axes of the best-fit ellipse conforming to the (x,y) coordinates of the pixels that defined the outsole. Ergo, the deviation of the major axes from vertical defined the degree of rotation necessary to ensure that the final imagery was oriented as close to North-South as possible within the image frame.

2.2.D Segmentation

Following registration and background subtraction, randomly acquired characteristics present on both the exemplar and crime scene-like images were marked. This process required the analyst to physically examine each outsole with oblique illumination and 4X magnification. Upon identifying a RAC that appeared on both the outsole and the exemplar, the analyst blacked out the RAC pixels on the Handiprint image using the pencil tool in Adobe[®] Photoshop[®] Elements 10. This was completed by tracing the edge of the RAC with the pencil tool (set at 2-pixels wide) and then filling in the RAC (if necessary), with the paint bucket tool while viewing the exemplar at a minimum magnification of 200X. When complete, each feature was examined to ensure that every pixel included within the traced perimeter of the RAC was fully labeled (converted to black). For features found on the edge of the shoe, a lug, or a tread element, the boundary of the RAC was interpolated by hand if the distance for interpolation was short and relatively linear (Fig. 4 (a)). In instances when the edge could not be dependably interpolated (*e.g.*, along an irregular segment, a curved surface, or near a large void area), the RAC was traced, but not closed, in order to avoid the introduction of interpolation variability (Fig. 4 (b)).

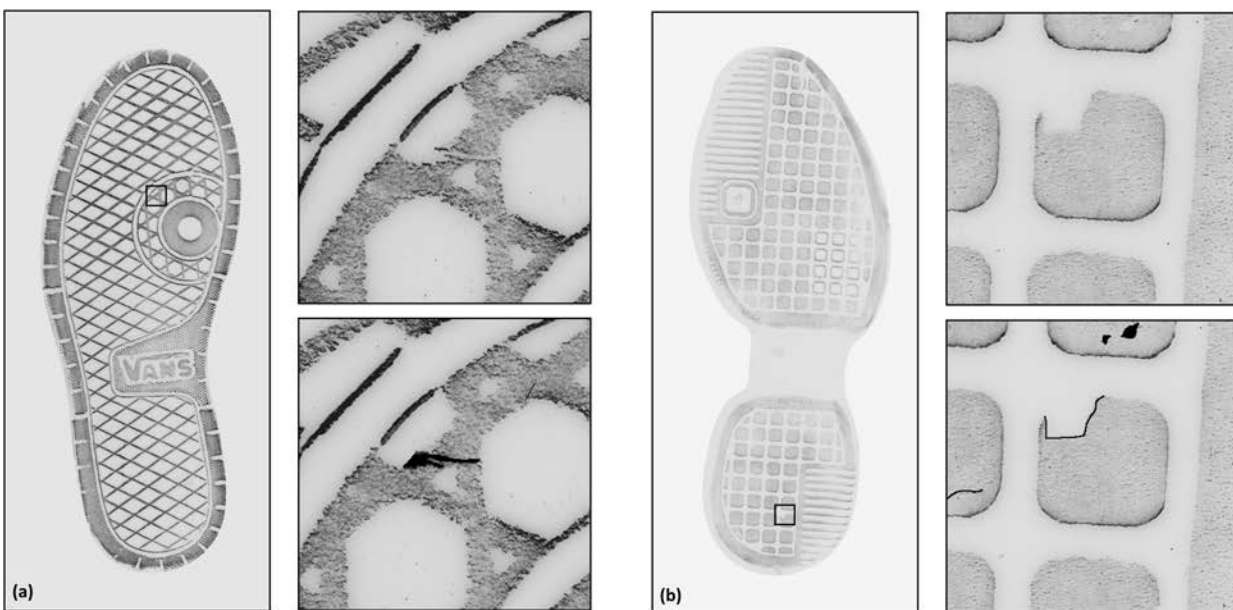


Figure 4: (a) Illustration of RAC on edge of linear tread element. Note that the edge of the RAC (terminating on the edge of a short and linear tread element), has been interpolated and the entire RAC has been filled in (size 9.5 men's Vans[®] sneaker, Skink Mid model shoe). (b) Illustration of RAC on edge of curved tread element. Note that the edge of the RAC (terminating on the edge of a curved tread element), has not been interpolated nor filled in (size 9 men's Adidas[®] sneaker, Pro Feather model).

When this registered and marked image was subtracted from its registered (but unmarked) counterpart, the result was a RAC map that highlighted the location and geometry associated with each randomly acquired feature (Figs. 5 and 6). Using the standard image processing technique of connected components, the location of each RAC was sequentially characterized using three parameters that were readily available based on x,y pixel coordinates; the radius (r) or distance (in pixels) between the shoe's midpoint and the RAC's centroid (geometric average of the RAC's x,y pixel coordinates), the angular (θ) position (in degrees) between the RAC's centroid and zero degrees (defined as a horizontal line drawn directly East of the shoe's midpoint), and the normalized distance (r_{norm}) equal to r divided by the distance (in pixels) between the shoe's midpoint and the perimeter of the shoe at angular position θ (obtained by casting out a vector from the shoe's midpoint to the shoe's perimeter at angle θ).



Figure 5: Registered and marked Handprint image (left) and resulting RAC map (right) (size 9 men's Converse Chuck Taylor® All Star®).

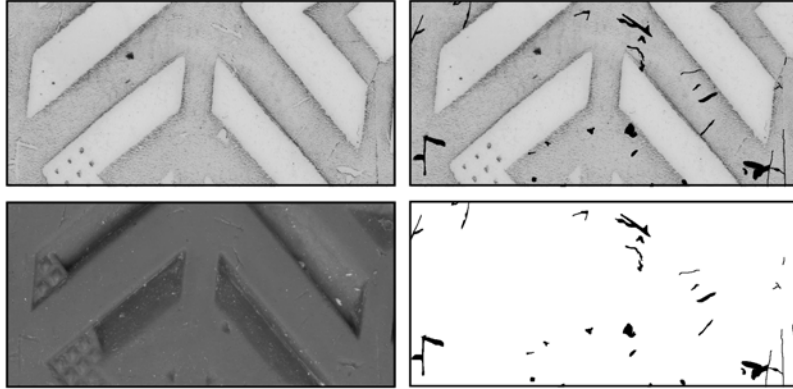


Figure 6: Example of a selected portion of the Converse Chuck Taylor[®] All Star[®]. Handprint (top left), outsole (bottom left), marked Handprint (top right), RAC map (bottom right). Note that the outsole image shown in this figure has been scanned on a flat bed scanner, but that all RACs were detected using 4X magnification and oblique illumination.

Following localization, each feature was automatically numbered (via its connected component value) and extracted from the total RAC map. The resulting subimages (Fig. 7) were then evaluated to define RAC shape and geometry, based on a five-dimensional RAC feature vector, before transformation into individual RAC Fourier descriptors (FD).

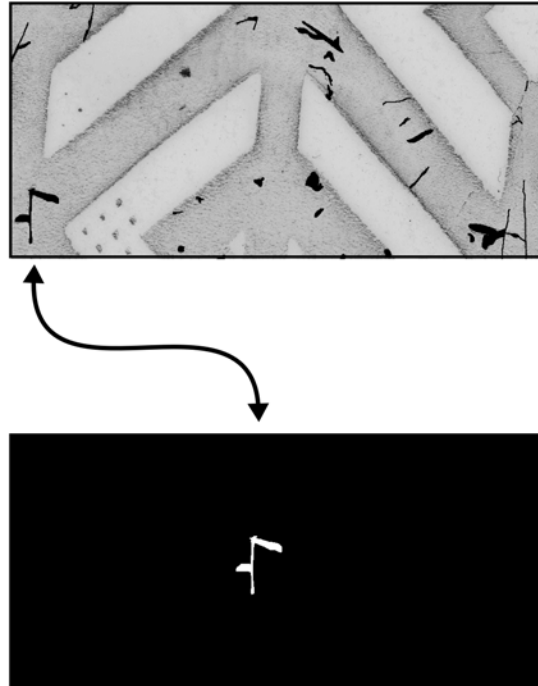


Figure 7: Subsection of RAC map and example of connected component subimages. This particular RAC was numbered #101 and located at a normalized radius of 0.55 and an angle of 104°.

2.2.E Processing

RAC Feature Vector

Initially, each randomly acquired characteristic was attributed to one of four categories (lines/curves, circles, triangles, and irregular-shaped features), as inspired by work conducted by Stone (2006) (Table 7).

Table 7: Description of accidental characteristics (adapted from Stone (2006)).

Feature	Discrete Position	Discernible Length	Discernible Width	Orientation	Curvature	2D Shape	Elevation
Point	Yes	No	No	N/A	N/A	N/A	N/A
Line	Yes	Yes	No	Yes	No	N/A	N/A
Curve	Yes	Yes	No	Yes	Yes	N/A	N/A
Enclosure	Yes	Yes	Yes	Yes	N/A	Yes	No
3D	Yes	Yes	Yes	Yes	N/A	Yes	Yes

However, geometric characterization was determined using an automated technique, and based on five RAC attributes, including: area, perimeter, linearity, circularity and triangularity. The first two descriptions (area and perimeter) were readily available; area describes the total number of pixels comprising the RAC and perimeter evaluates the distance in pixels along a line/curve, or around a two-dimensional shape.

The linearity metric was also readily available and was obtained by computing the ratio of the first and second eigenvalues (λ_1 and λ_2) generated from eigen decomposition of the RAC itself (Park and Jain, 2010). Using this approach, when λ_1 is much greater than λ_2 , the RAC in question has a greater length than width and can be classified into the line/curve category.

The fourth measurement was a circularity metric, computed according to Eq. 3 (Gonzalez and Woods, 2008), where A is the area of the object, and P is the length of its perimeter:

$$R_c = \frac{4\pi A}{P^2} \tag{3}$$

$$R_c = \text{maximum of 1.0 for a perfect circle}$$

The fifth and final metric was a triangularity value computed using central moments (Eq. 4) that are invariant to translation, scale and rotation. As per Rosin (2003) (Rosin, 2003), the variable I_1 in Eq. 5 equals $\frac{1}{108}$ for any triangle that has been affine transformed into a perfect right-angled triangle:

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q \tag{4}$$

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4} \tag{5}$$

As such, the triangularity measure can be normalized to vary between 0.0 – 1.0 according to Eq. 6 (Rosin, 2003):

$$T = \begin{cases} 108 I_1 & \text{if } I_1 \leq \frac{1}{108} \\ \frac{1}{108 I_1} & \text{otherwise} \end{cases} \quad (6)$$

The five-dimensional feature vector (Fig. 8) describing area, perimeter, linearity, circularity and triangularity served as a primary descriptor and comparison parameter for each randomly acquired characteristic. In addition, it was used to categorize the randomly acquired characteristics into one of the four groups previously mentioned; line/curve, circle, triangle or irregular.

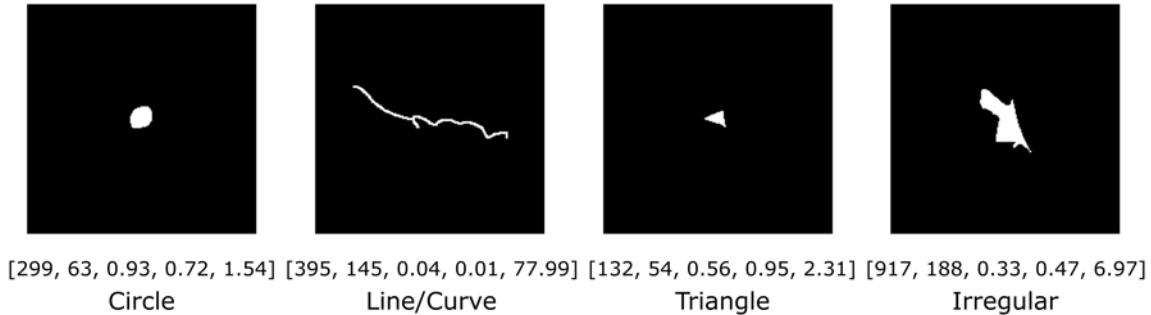


Figure 8: Four RAC images with their corresponding feature vectors [area, perimeter, circularity, triangularity, linearity].

Based on a survey of known geometric shapes, absolute categorization rules were developed. More specifically (and for this dataset), circles have a circularity measure greater than or equal to 0.8, triangles have a circularity measure less than 0.8 *and* a triangularity greater than or equal to 0.9, while lines/curves have a linearity ratio greater than 5 *and* a triangularity measure less than or equal to 0.3; any shape not satisfying one of the above rules defaults into the irregular category (Fig. 9).

Although this categorization was relatively straightforward when it came to idealized shapes, it was clear that RACs rarely took on an idealized form/habit, and therefore, categorization into narrowly defined bins proved to be somewhat unproductive, and even sometimes at odds with what a human analyst might decide if asked to categorize a real randomly acquired characteristic. However, this disagreement cannot be defined as a failing of the algorithm, since shape categorization by a human observer is, by necessity, both subjective and personal. Nonetheless, when quantified, the agreement between ‘automated’ and ‘human’ categorization of RACs ranged between 95% and 68%, depending on the complexity and imperfections of the shape under review. For example, using a test set of 74 ‘stylized’ shapes (manually created in ImageJ (Rasband, 2016) with an intended geometry), plus 110 randomly selected RACs, the overall agreement or accuracy in categorization was computed to be 95%. This was determined by taking the total test set of 184 images and presenting them via a graphical user

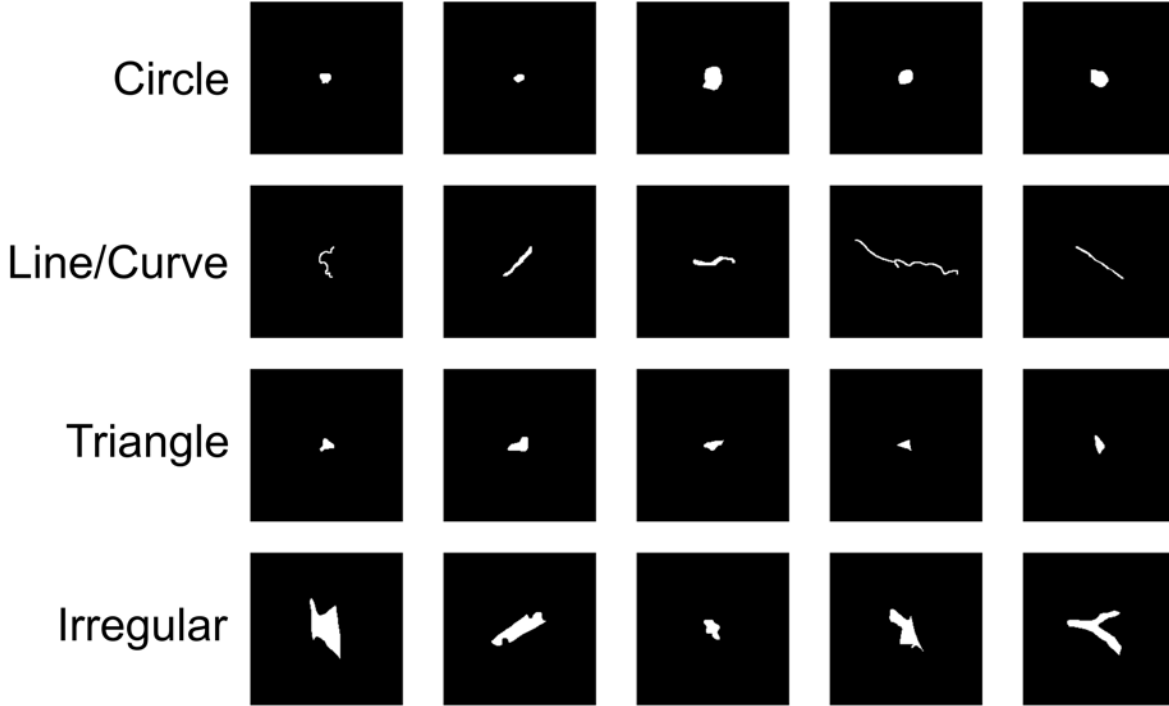


Figure 9: Examples of RACs classified as circles, lines/curves, triangles, and irregulars by the algorithm.

interface to analysts seated at a computer. When presented with each image, in a randomized order, the analyst was asked to categorize the shape as either a circle, triangle, line/curve or irregular-shaped feature by clicking on a corresponding toggle button. The same shape was automatically categorized using the decision rules determined during our training phase, and the results for three separate analysts (for a total of 552 human-perceptual estimates of shape categorization) were combined into the confusion matrix shown in Table 8 with an overall agreement of 95%.

Table 8: Confusion matrix for automated categorization of 184 shapes (74 stylized and 110 real RACs) as assessed by three analysts for a total of 552 human-perceptual assessments of shape. The column headers represent the algorithm report while the rows designate human-perception. Total agreement equals 95%.

Label	Circle	Triangle	Line/Curve	Irregular
Circle	99	0	0	0
Triangle	0	90	2	4
Line/Curve	0	0	214	8
Irregular	1	5	5	124

Conversely, for a total of 800 randomly selected RACs (zero stylized shapes), assessed by four analysts (200 each, with a total of 746 human-perceptual estimates of shape categorization of which 27 RACs happen to repeat during the randomized selection), the equivalent confusion matrix (shown in Table 9) was found to have an overall agreement of 68%. Despite the clear

decrease in agreement, the authors assert that this should not be defined as an ‘error rate’ since it is based on human-perception of shapes, which cannot be expected to agree among or between individuals. The problem is that there is no appropriate reference by which to define ‘ground truth’ as soon as shapes become complex and imperfect. To illustrate this, consider Figs. 10 and 11 which show a sampling of RAC images that lead to disagreement in the ‘human-perception’ versus ‘automated-algorithm’ study, contributing to the results shown in Table 9. In both figures, the top row denotes the automated categorization label, while the cell label indicates the human analyst choice. Depending on the viewer and the image, there are some instances where the human’s reasoning seems more ‘accurate’, and some instances where the algorithm’s choice seems more ‘accurate’. Overall, the results suggest that it would not be robust to keep a large number of RAC shape groupings, due to both human-perceptual differences and RAC complexity/imperfections. Given this observation, the authors suggest a maximum of three groups that may be useful moving forward; ‘irregular’ for complex structures, ‘elongated’ to describe lines and curves, and a new grouping defined as ‘approximate isometric’ to include circular and triangular structures.

Table 9: Confusion matrix for automated categorization of 746 unique RACs and 27 repeated RACs as assessed by four analysts for a total of 800 human-perceptual assessments of shape. The column headers represent the algorithm report while the rows designate human-perception. Total agreement equals 68%. Note that early categorizations also included a ‘rectangle’ group, defined as a linear element with width/thickness greater than 30 pixels (or 1.3 mm).

Label	Circle	Triangle	Rectangle	Line/Curve	Irregular
Circle	46	10	0	0	19
Triangle	12	25	0	1	19
Rectangle	3	3	0	2	3
Line/Curve	0	7	0	340	58
Irregular	9	80	0	29	134

It is also important to note that several analysts were involved in marking randomly acquired characteristics. Therefore, to assess inter- and intra-analyst variation in RAC marking, a random set of 100 pairs of shoes (approximately 8% of the database) were selected for periodic reassessment. On an approximate once-weekly basis, each analyst selected the next available shoe from the randomized list (which may or may not be a shoe he or she has already marked), and repeated the marking process on the post-registered and background subtracted image. Subtraction of the newly marked RAC image from its registered and unmarked mate created a secondary RAC map. Differences between replicate maps then served as a basis for assessing inter- and intra-analyst variation in marking.

To date, the quality assessment program has obtained 161 paired RAC maps, prepared by 5 analysts, over a 15-month time period (although at the time of writing, only 160 have been used in data analysis). The information contained in each of the 320 RAC maps (2 markings x 160 shoes) has been assessed in two ways. First, the data has been converted into a one-dimensional (1D) vector by rastering across image rows and down image columns, collecting total RAC size per cell using a fixed bin width of 150 x 150 pixels (approximately 6mm x 6mm). The resulting 1D feature vectors of RAC size (per cell) for paired RAC maps were

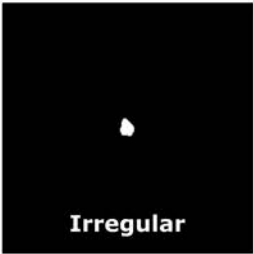
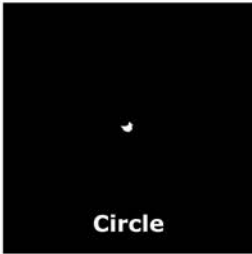
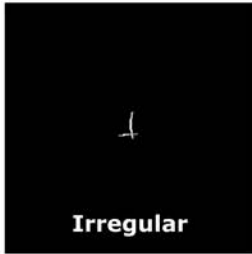
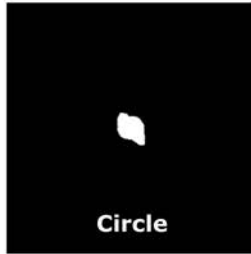
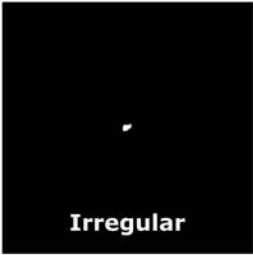
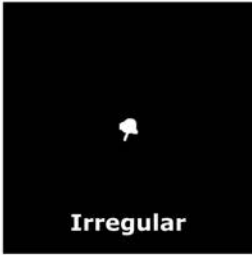

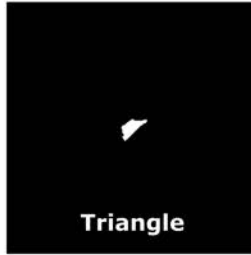
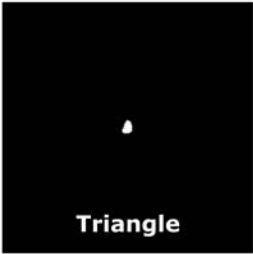
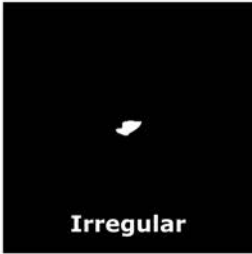

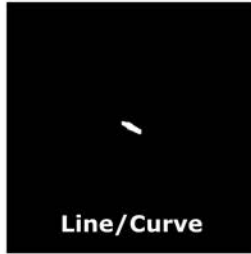
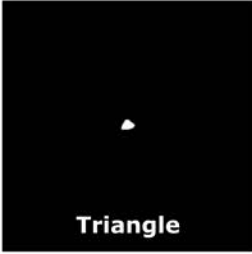
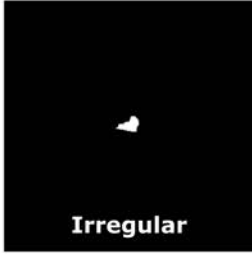
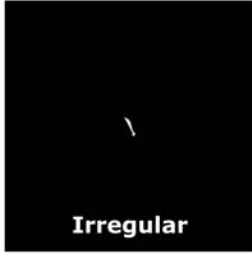
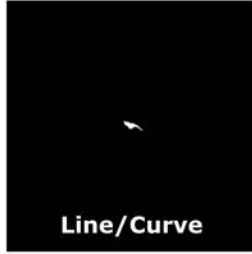
Automated Classification			
Circle	Triangle	Line/Curve	Irregular
 Irregular	 Circle	 Irregular	 Circle
 Irregular	 Irregular	 Irregular	 Triangle
 Triangle	 Irregular	 Irregular	 Line/Curve
 Triangle	 Irregular	 Irregular	 Line/Curve

Figure 10: Illustration of disagreement in human-perception of shape categorization (cell labels) versus automated categorization based on training rules (column header). To account for this disagreement a reduction in grouping complexity (from four to three) is suggested: irregular, elongated (lines and curves) and approximate isometry (a combination of circular and triangular structures).


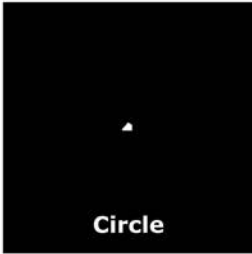
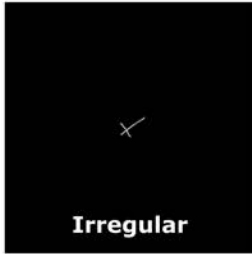
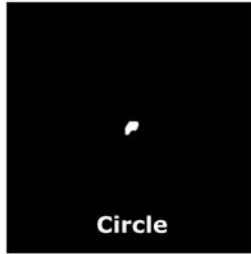
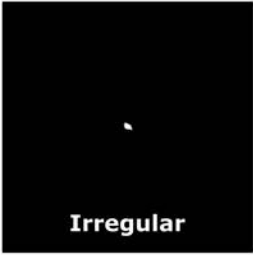
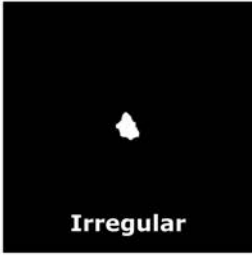
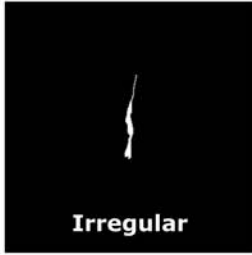
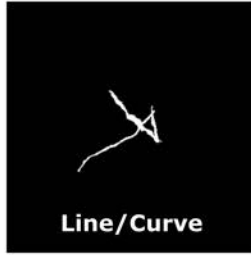
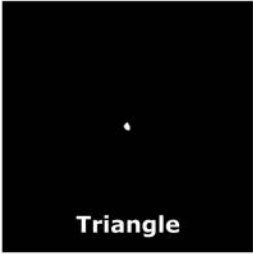
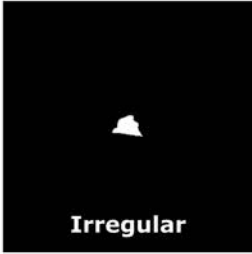
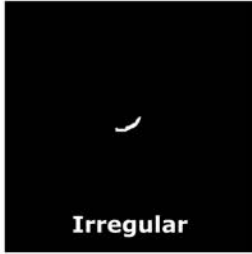
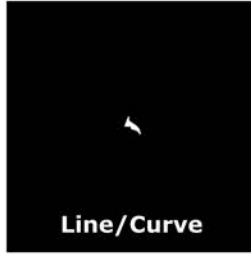
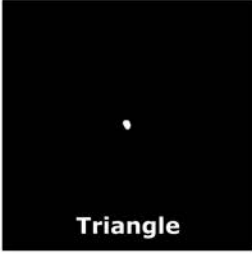
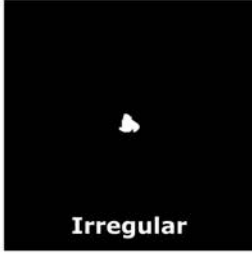
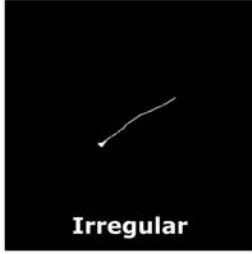
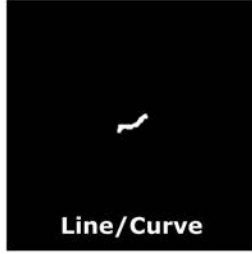
Automated Classification			
Circle	Triangle	Line/Curve	Irregular
 Irregular	 Circle	 Irregular	 Circle
 Irregular	 Irregular	 Irregular	 Line/Curve
 Triangle	 Irregular	 Irregular	 Line/Curve
 Triangle	 Irregular	 Irregular	 Line/Curve

Figure 11: Illustration of disagreement in human-perception of shape categorization (cell labels) versus automated categorization based on training rules (column header). To account for this disagreement a reduction in grouping complexity (from four to three) is suggested: irregular, elongated (lines and curves) and approximate isometry (a combination of circular and triangular structures).

then evaluated to determine the average correlation coefficient of similarity. Inter-analyst variation produced an average correlation coefficient of 0.66 with a variance of 0.057, based on 137 paired RAC maps. To date, the dataset has allowed for the computation of intra-analyst correlation, but thus far, based on only 23 paired RAC maps for 2 analysts in the research group; the combined average correlation coefficient is 0.80 with a variance of 0.016. In addition to the image-wide correlation scores, individual uncertainty of measure for θ , r , and r_{norm} has been computed based on this same dataset. Table 10 reports the mean, variance and range of measurements associated with each value based on 160 shoes and 5,477 duplicate marked randomly acquired characteristics (combined inter- and intra-analyst markings). In addition, illustrations of duplicate markings of known match RACs are shown in Fig. 12, along with individual measurement differences. The results indicate that angular differences are very small (less than a 1°) and that radial distances differ by 0.16 ± 1.9 mm. The interpretation of each quality metric (correlation versus measurement uncertainty) indicates that the greatest variation is within the RAC *detection* process; but when a RAC is detected, it is marked, on average, in such a manner to limit variation in actual localization.

Table 10: Variation in analyst duplicate marking of 5,477 randomly acquired characteristics across 160 shoes (320 RAC maps).

Metric	θ (degrees)	r (pixels)	r (mm)	r_{norm}
Mean	0.0922	4.27	0.167	0.00177
Variance	0.0178	91.7	3.61	0.0000121
Maximum	0.6990	112	4.40	0.03000

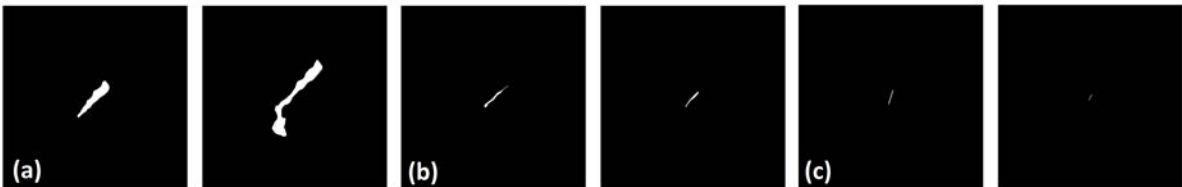


Figure 12: Duplicate markings of known match RACs with the following marking variations: (a) $\Delta\theta = 0.235$, $\Delta r = 67.2px/2.64mm$, $\Delta r_{norm} = 0.0170$; (b) $\Delta\theta = 0.567$, $\Delta r = 88.5px/3.49mm$, $\Delta r_{norm} = 0.0270$; (c) $\Delta\theta = 0.551$, $\Delta r = 112px/4.40mm$, $\Delta r_{norm} = 0.0290$.

However, when the shape categorization of duplicate marked RACs are compared, repeated concerns about shape classification are noted. For example, Tables 11 and 12 report confusion matrices for RAC shape classification when high quality images marked by different analysts (or the same analyst repeatedly) are pairwise compared (Table 11; 5,477 features), and when high quality images are pairwise compared with lower quality images created to mimic crime scene-like conditions (Table 12; 1,766 RACs). As is evident from the results, the overall consistency in shape label varies between 74% and 68%, respectively.

2.2.F Outsole Size & Shape Normalization

At this point in data acquisition, each RAC has a geometric description, location and well-defined origin (from a left or right shoe with a known pattern, a known manufacturer (usually),

Table 11: Confusion matrix for RAC shape categories for duplicate RAC markings among high quality RACs that were part of the quality control/assurance study. Note that approximately 74% of known match RACs maintained the same shape class.

Label	Elongated	Isometric	Irregular
Elongated	2047	35	270
Isometric	37	476	353
Irregular	287	381	1341

Table 12: Confusion matrix for RAC shape categories for duplicate RAC markings between high quality RACs and those created to mimic crime scene-like conditions. Note that approximately 68% of known match RACs maintained the same shape class.

Label	Elongated	Isometric	Irregular
Elongated	487	19	128
Isometric	5	184	176
Irregular	89	141	537

a known size, etc.) and can be assessed as such. However, for any given shoe size (or pattern, or brand, etc.) the database itself is limited in sample size. With this in mind, an interim solution (at least until the database grows to such a size that sampling is considered robust) is to transform the frequency information into a normalized space that allows for numerical assessment regardless of shoe size, shape, pattern, etc. Naturally, this simplification bounds the utility of the frequency information, and the authors urge the user to be cognizant of this moving forward, but the transformation in no way invalidates provisional usefulness.

Normalization was achieved using a *single idealized shoe* corresponding to a men’s size 10 Reebok[®] walking shoe with an outsole surface area of approximately 21,235mm². Beginning from the top medial portion of the shoe, the outsole was divided into 5mm x 5mm cells through a rastering process, creating 990 total cells of which 860 were complete, and 130 were partial (or straddling the perimeter/edge of the outsole as illustrated in Fig. 13). By mapping between Cartesian and polar coordinates, each RAC could be localized via θ and r_{norm} . Essentially, this meant that a RAC near the edge of the medial part of the heel on a women’s size 6.5 could have the same θ and r_{norm} as a RAC on the edge of the medial part of the heel of a men’s size 10.0, and therefore map to the same 5mm x 5mm cell in the normalized outsole. (*Note: we also have the capacity to report frequency values as absolute, physical or non-normalized values using θ and r . This would be equivalent to taking a stack of Handiprints, centering all shoes in the middle of each sheet with the toe-heel oriented North-South, and drilling down through all sheets at a fixed location, regardless of shoe size. To further elaborate, in the aforementioned example, the RAC on the medial heel portion of the women’s size 6.5 shoe would likely fall somewhere in the lower-instep area of the men’s size 10.0.*)

The shoes described in Tables 2 - 6 generated a total of 72,306 RACs. The minimum number on a single outsole was 1, and the maximum was 465 (with an average of 56). The mean, standard deviation and maximum number of RACs per heatmap bin is summarized in Table 13. The total across all shoes/bins reveals 25,420 irregularly shaped features (35%), 32,549

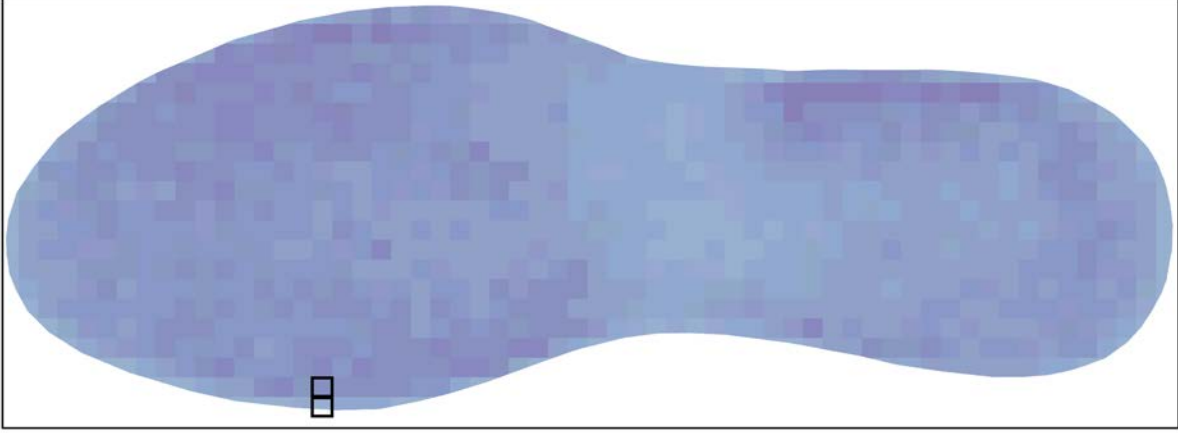


Figure 13: Illustration of a full and partial 5mm x 5mm cell on the normalized outsole.

linear features (45% lines/curves), and 14,337 isometric features (20% circles/triangles).

Table 13: Descriptive statistics of RAC frequency per heatmap bin as a function of shape.

Statistic	Any Shape	Isometric	Elongated	Irregular
Mean	73	14	33	26
SD	36	8	19	13
Maximum	182	45	109	63

Table 14 reports the frequency of bins with a given range of randomly acquired characteristics, revealing that the highest percentage of bins (35%) have between 61-90 RACs. Graphically, this same data is provided as a histogram in Fig. 14, but with a binwidth/resolution of 12 (*e.g.*, 0-12 RACs, 13-24 RACs, etc.) showing that the largest frequency of bins have between 72-84 RACs.

Table 14: Frequency of bins with a given range of RACs.

Number of RACs	Frequency of Bins	~ Percentage %
0	14	> 1
1-30	138	14
31-60	191	19
61-90	342	35
91-120	202	20
121-150	89	9
151-180	14	>1
Total	990	100

2.2.G Shape Descriptor

Based on the limitations noted using general shape categorization, additional methods to discern RAC similarity/dissimilarity were sought. To this end, each RAC was treated as a

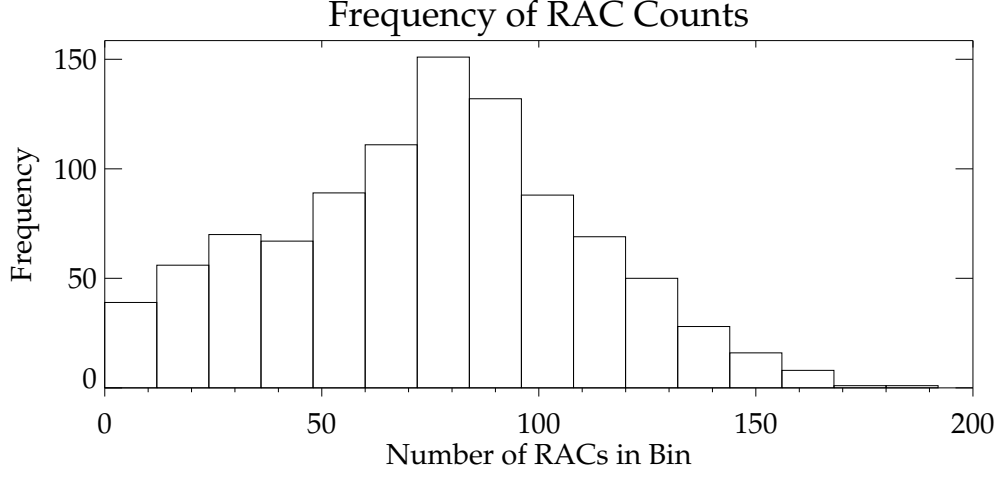


Figure 14: Frequency of RACs per bin (cell/bin resolution of 5mm x 5mm) and histogram bins of size 12 (i.e., 0-12, 13-25, etc.).

closed planar figure yielding a Fourier description (Bartolini et al., 2005; Dalitz et al., 2013; Wallace and Mitchell, 1980). This description was generated by tracing the contour of the shape $(x(t), y(t))$ (where $t = 0, \dots, N - 1$ with $N = 350$ for this dataset) and assuming a complex plane $z(t) = x(t) + iy(t)$ (where $i = \sqrt{-1}$). The resulting one-dimensional complex sequence of numbers was then mapped to the frequency domain via the discrete Fourier transform (Bartolini et al., 2005) where R_m and θ_m are the magnitude and phase of the m^{th} coefficient, respectively (Bartolini et al., 2005):

$$Z(m) = \sum_{t=0}^{N-1} z(t) e^{(-i2\pi mt/N)} = R_m e^{(i\theta_m)} \quad (7)$$

$$m = -N/2, \dots, -1, 0, 1, \dots, N/2 - 1$$

As necessary, the coefficients can be normalized and forced to be invariant to translation, scale, rotation and contour/sequence start point according to the following modifications (Bartolini et al., 2005):

$$\begin{aligned} Z(0) = 0 & \Rightarrow \text{translation invariance} \\ R_m = \frac{R_m}{R_1} & \Rightarrow \text{scale invariance} \\ \theta_m = \theta_m - \frac{\theta_{-1} + \theta_1}{2} & \Rightarrow \text{rotation invariance} \\ \theta_m = \theta_m + m \frac{\theta_{-1} - \theta_1}{2} & \Rightarrow \text{start point invariance} \end{aligned} \quad (8)$$

To illustrate, consider Fig. 15; the first row depicts a single RAC (A), along with four synthetic

modifications (B-E showing changes in scale, rotation and translation), and the second row illustrates the associated normalized Fourier descriptors. Note that the x- and y-axes are arbitrary since the images have been normalized, but all shape contours are normalized to the same configuration, save a single π radian ambiguity (Folkers and Samet, 2002). Unless otherwise noted, all subsequent uses of RAC Fourier descriptors make use of both translation and start point invariance modifications.

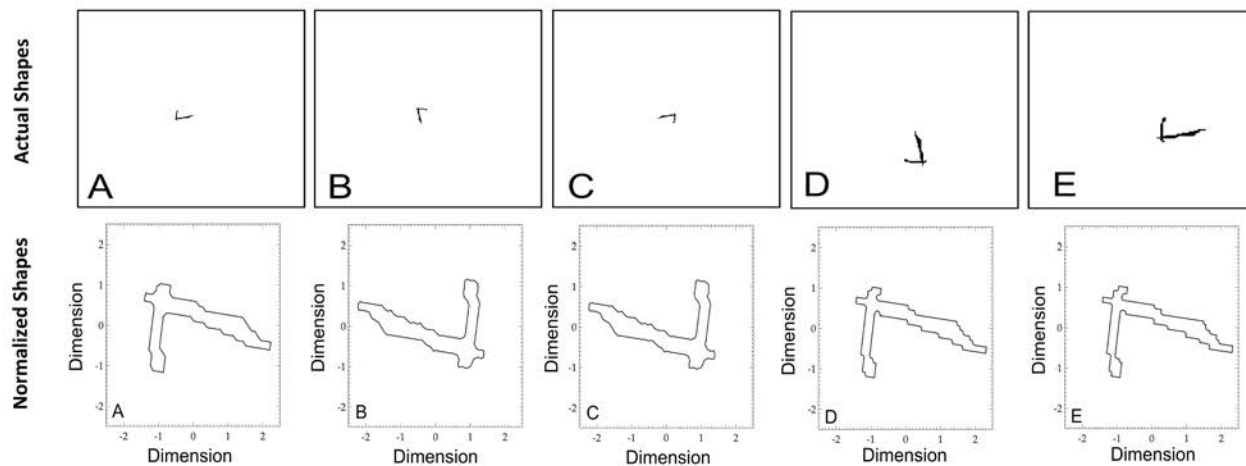


Figure 15: First row: (A) Original RAC, (B) Rotated, (C) Rotated, (D) Rotated, Translated, and Scaled, (E) Scaled and Translated. Second row: Plot of normalized Fourier shapes derived from the RACs shown in first row.

2.3 Similarity Assessment

The aforementioned normalization step yields RAC frequency information and the potential for chance co-occurrence of RACs within a 5mm x 5mm cell on an outsole (in other words, the dataset can empirically estimate the random chance of discovering two or more accidentals in the same position on shoes previously known to be unrelated). This can be further divided by geometry in terms of the chance co-occurrence of elongated, isometric and irregular shaped RACs within a 5mm x 5mm cell. However, chance co-occurrence in position and general category does not mean coincidental association in actual geometry since general categorization does not sufficiently describe RAC complexity. Thus, chance association in position and general shape should be extended to include an estimate of the actual similarity exhibited between the co-occurring features. This can be accomplished using numerical metrics, visual assessment, and/or both. In support of this, one of the goals of this work was to examine the strengths and weaknesses of a host of similarity metrics. More specifically, the project was contracted to deliver results for correlation using full RAC maps, however, the research team actually investigated and provide results for five different metrics, as assessed individually when comparing high quality RACs to their crime scene-like reproductions, as well as one metric that compared total RAC maps (where a map is a collection or constellation of RACs in totality). The goal was to:

- Assess the strengths and weaknesses of each metric;

- Report the actual separability of known matches and known non-matches as a function of RAC maps;
- To reduce the number of pairwise comparisons that would be required if an analyst wished to ‘visually’ assess similarity. In other words:
 - To perform all pairwise similarity comparisons for RACs with positional overlap (which required a total of 3,239,114 pairwise comparisons);
 - Sort the ordered list in every bin/cell to reveal the top (most similar) 8 paired RAC candidates;
 - Allow an analyst to visually examine the short-list (pairs deemed *most similar* mathematically) to determine the degree to which the coincidental association in position and categorical-shape is associated with *visually indistinguishable* RACs, and therefore of forensic significance.

With the above in mind, the remainder of this section will describe the similarity metrics of interest, the strengths and weaknesses associated with each metric, the separation of known matches and known non-matches, and the manner in which a numerical metric can be used within a web-based application.

Identification of Known Match (KM) RAC Pairs

A database of 2,159 randomly acquired characteristics was compiled by marking features on the 200 crime scene-like (CS) impressions (originating from 100 shoes) as detailed in Richetelli et al. (2017). In order to compare KM RACs, it was necessary to identify correspondences between accidentals on high quality exemplars and crime scene-like prints. For exemplars that were repeatedly marked within the quality control/assurance program, correspondences were readily available. For crime scene-like images, there was much greater chance that the print deposition and replication process would create positional variations, so RAC pair candidates were nominated if the angular (θ) and normalized radial values (r_{norm}) between a RAC on an exemplar fell within $1 - 2^\circ$ and 0.1, respectively, of a corresponding θ and r_{norm} on its associated crime scene-like image. These (large) thresholds were selected in order to minimize loss of candidate RAC mates, but all resulting candidates were manually verified (and adjusted as necessary) before moving forward. Fig. 16 illustrates a set of these known match pairs, as well as the corresponding location information for each accidental, of which a total of 1,766 paired candidates were detected (Richetelli et al., 2017).

2.3.A RAC Loss

Given the inherent inconsistency present in shoeprint creation (such as variation in pressure, torque, substrate, etc.), it is expected that reproduction of RACs in crime scene-like quality prints will be variable in comparison to high quality exemplars collected by pressing a dusted outsole against an adhesive sheet, thus ensuring full and even contact. Based on the results from this study, an average of 85% of RACs were not reproduced in crime scene-like impressions (Table 15). In addition, zero RACs were reproduced in 10% of the images (20 out of 200 impressions). Loss was further broken down by shape, perimeter and area to determine if RAC

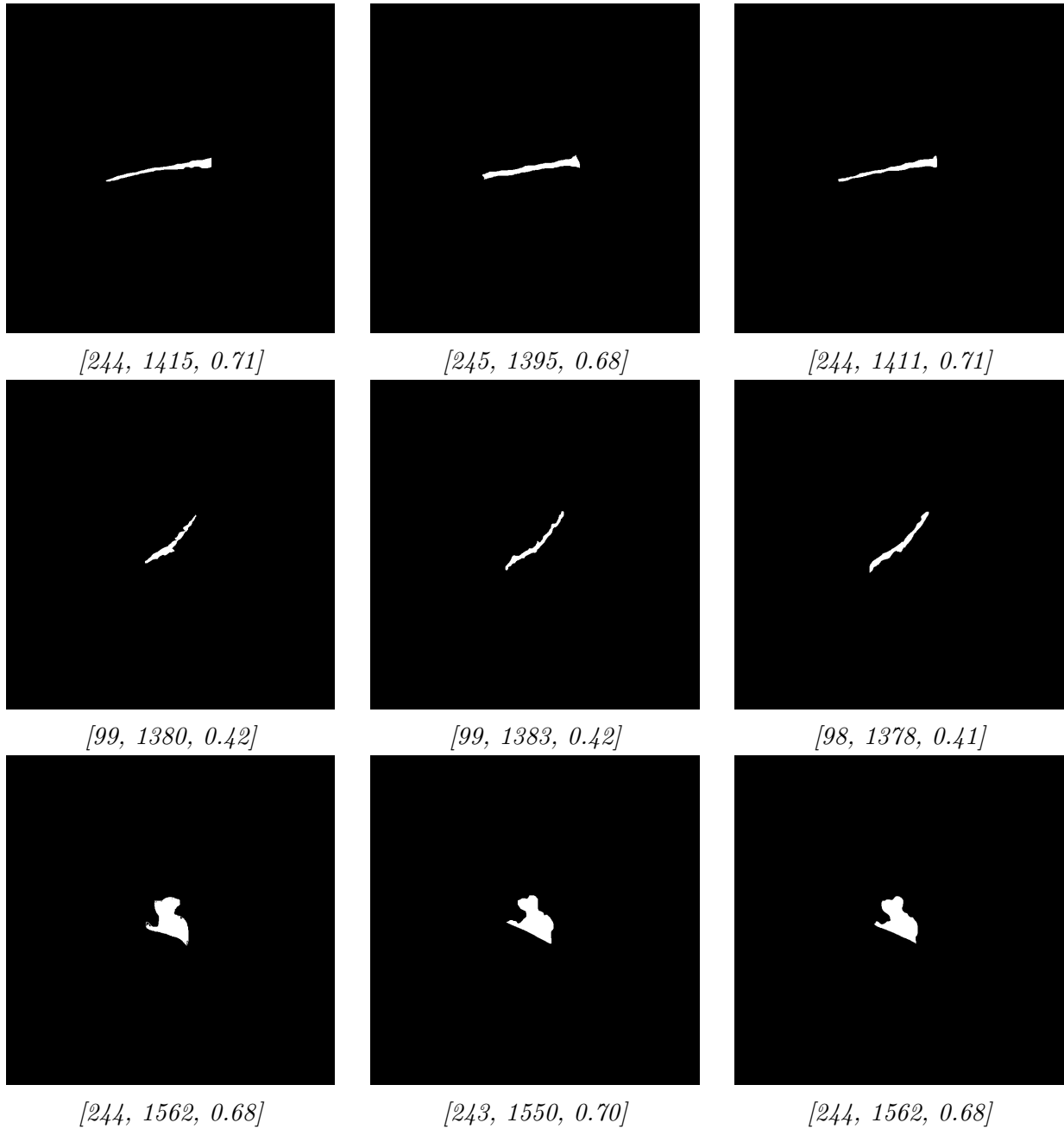


Figure 16: RAC image mates with their corresponding location information $[\theta$ (degree), r (pixel), r_{norm}]. High quality RAC image (left) with its detected crime scene RAC mates from each replicate (center, right).

reproduction varied as a function of any of these factors. As detailed in Table 16, RAC loss (77% - 84%) exhibited very little variation across shape categories. However, greater variation can be observed as a function of RAC size; in fact, significant differences in loss as a function of RAC size were detected as per the Chi-square test (McHugh, 2013) with $\alpha = 0.05$. To summarize the observed trend, as a feature’s size increased (in either total area or perimeter), the percent loss decreased (Tables 17 and 18). Of course, this matched intuition in that ‘larger’ defects were more likely to persist and withstand the variation introduced during reproduction in a crime scene setting as compared to smaller features that were more easily occluded by erratic conditions (such as differences in media, substrate, motion, pressure, etc.).

Table 15: Quantifying RAC loss between high quality exemplars and replicate crime scene-like impressions (HQ = high quality, CS = crime scene).

RACs	HQ	CS Rep 1	CS Rep 2
Total Number	6,896	1,049	1,110
Number Lost	-	5,847	5,786
Percent Lost	-	85%	84%
Mean Number per Shoe \pm 1 SD	69 ± 72	10 ± 12	11 ± 12
Maximum Number	307	66	61
Minimum Number	2	0	0

Table 16: RAC loss between high quality exemplars and replicate crime scene-like impressions as a function of RAC shape (HQ = High Quality).

Shape	Total HQ RACs	Lost HQ RACs	\sim % Loss
Circle	1,024	863	84%
Line/Curve	2,685	2,239	83%
Irregular	2,732	2,173	80%
Triangle	455	348	77%

Table 17: RAC loss between high quality exemplars and replicate crime scene-like impressions as a function of RAC perimeter (HQ = High Quality).

Perimeter (P) (mm)	P (pixels)	Total HQ RACs	Lost HQ RACs	\sim % Loss
$0 < P < 2$	0-45	2,936	2,623	89%
$2 \leq P < 4$	46-91	2,413	1,939	80%
$4 \leq P < 6$	92-137	828	599	72%
$6 \leq P < 8$	138-183	337	217	64%
$P \geq 8$	≥ 184	382	245	64%

2.3.B Comparison

To assess the degree of similarity that can be expected when comparing high quality exemplars with crime scene-like impressions, five metrics of similarity were considered, including:

Table 18: RAC loss between high quality exemplars and replicate crime scene-like impressions as a function of RAC area (HQ = High Quality).

Area (A) (mm ²)	A (pixels ²)	Total HQ RACs	Lost HQ RACs	% Loss
$0.00 \leq A < 0.25$	0-131	3,994	3,548	89%
$0.25 \leq A < 0.50$	132-264	1,408	1,080	78%
$0.50 \leq A < 0.75$	265-396	589	419	71%
$0.75 \leq A < 1.00$	397-528	294	201	68%
$1.00 \leq A < 2.00$	529-1058	391	253	65%
$A \geq 2.00$	≥ 1059	220	122	55%

modified phase only correlation (MPOC), matched filter (MF), a modified cosine similarity (MCS), Hausdorff distance (HD), and Euclidean distance (ED). *Note: the results comparing high quality versus crime scene-like imagery were assessed prior to characterization of all 1,300 outsoles in the current database, and are therefore based on the first 1,000 outsoles that were analyzed.*

Modified Phase Only Correlation (MPOC)

The Fourier transform $F[g(x, y)] = G(u, v)$ of a spatial domain image $g(x, y)$ gives the analyst access to frequency information associated with image amplitude $A(u, v)$ and phase $\sigma(u, v)$ as illustrated in Eqs. 9 and 10 (where the subscripts reference the images under comparison and $i = \sqrt{-1}$) (Bouridane, 2009).

$$G_1(u, v) = A_1(u, v)e^{i\sigma_1(u, v)} \quad (9)$$

$$G_2(u, v) = A_2(u, v)e^{i\sigma_2(u, v)} \quad (10)$$

Once the Fourier transform of each input image has been calculated, the phase only correlation can be computed according to Eq. 11 (de Chazal et al., 2005; Gueham et al., 2007; Xiao and Shi, 2008) where F^{-1} is the inverse Fourier transform and G_2^* is the complex conjugate of G_2 (Bouridane, 2009).

$$\begin{aligned} POC_{g_1g_2} &= F^{-1} \left[\frac{G_1(u, v)G_2^*(u, v)}{|G_1(u, v)G_2^*(u, v)|} \right] \\ &= F^{-1} [e^{i[\sigma_1(u, v) - \sigma_2(u, v)]}] \end{aligned} \quad (11)$$

Eq. 11 can be modified by application of a frequency filter that selectively limits frequencies used in the computation such that $F[g(x, y) \cdot h(k, l)] = G(u, v)$. In this work, each image $g(x, y)$ was modified by the windowing function shown in Eq. 12 with $\alpha = 0.2$ and where $k = l = N$ which is the size of the RAC image in pixels (1,600 x 1,600):

$$\begin{aligned}
h(k) &= \alpha - (1 - \alpha) \cos \left[\frac{2\pi k}{N} \right] \\
k &= 0, 1, \dots, N - 1
\end{aligned} \tag{12}$$

Fourier Descriptors (FD)

With the exception of MPOC which was computed using 1,600 x 1,600 pixel imagery, all remaining similarity metrics were based on perimeter information. More specifically, the RAC was treated as a closed planar figure yielding a Fourier description (FD) (Bartolini et al., 2005; Dalitz et al., 2013; Wallace and Mitchell, 1980), and as previously described (Eqs. 7 and 8 and Fig. 15), after normalization to ensure invariance to translation, rotation and contour/sequence start point.

Matched Filter (MF)

The matched filter similarity metric between two shapes $\hat{Z}_1(m)$ and $\hat{Z}_2(m)$ was computed as illustrated in Eq. 13 (Gregga et al., 2002) where $\hat{Z}(m)$ is normalized according to $\frac{\hat{Z}(m)}{\sqrt{\sum_t |z(t)|^2}}$ such that 0.0 is the minimum (least similar) and 1.0 is the maximum (most similar):

$$\text{MF} = \operatorname{argmax} \left| \frac{1}{N} \sum_{t=0}^{N-1} \hat{Z}_1(m) \hat{Z}_2(m) e^{(i2\pi mt/N)} \right| \tag{13}$$

Modified Cosine Similarity (MCS)

Cosine similarity is a commonly used metric that can assess the similarity between two data vectors (Schott, 2007). For two *similar* inputs a and b , the resulting angle (θ) between them will be small; conversely, θ is large for two dissimilar inputs. Since the RAC perimeters were defined as FDs (or complex numbers $z(t) = x(t) + iy(t)$), each complex vector was converted to a real-valued vector (\hat{z}) by adding x and y in quadrature before employing the traditional cosine computation shown in Eq. 14, where (T) represents the transpose of a vector.

$$\theta = \cos^{-1} \left[\frac{a^T b}{\sqrt{a^T a} \sqrt{b^T b}} \right] \tag{14}$$

Euclidean Distance (ED)

Euclidean distance was the fourth metric employed for comparison. The distance (D) between elements in complex vectors was obtained as detailed in Eq. 15, where x_1 and y_1 denote the real and imaginary parts of the first vector, respectively (Schott, 2007). Likewise, x_2 and y_2 denote the real and imaginary parts of the second vector for comparison, respectively. The total distance was normalized by dividing the summation by the maximum number of elements in the vectors ($N = 350$ for this dataset), yielding an average distance. Naturally, as elements become more dissimilar, the distance between them increases.

$$D = \frac{1}{N} \sqrt{\sum (x_1 - x_2)^2 + \sum (y_1 - y_2)^2} \quad (15)$$

Hausdorff Distance (HD)

Using the Euclidean distance, Hausdorff distance was likewise computed. This is more a variant of ED than a truly unique computation since ED was used ‘under-the-hood’ in the HD computation (instead of a new metric - such as Manhattan distance - but this is something that can be remedied moving forward). In this computation, the distance ($d(a, b)$) was computed between a point (*e.g.*, a_1) on the perimeter of RAC (A) and all points on the perimeter of RAC (B) using the desired distance metric (in this case, ED). Following all computations, the smallest distance from a_1 to B was retained. This process was then repeated for all points on A (*i.e.*, $a_2 \dots a_n$), wherein $h(A, B)$, or the maximum of these minimums, was retained (Huttenlocher et al., 1993). This same process was repeated to compare all points on RAC perimeter vector B to those on RAC perimeter vector A, thus obtaining $h(B, A)$. The actual distance HD was then the maximum of these two values ($h(A, B)$ and $h(B, A)$) as illustrated in Eq. 16.

$$H(A, B) = \max\{h(A, B), h(B, A)\} \quad (16)$$

where $h(A, B) = \max_{a \in A} \{\min_{b \in B} \{d(a, b)\}\}$

2.3.C Individual RAC Similarity

Of the five metrics utilized to determine similarity between crime scene-like RACs and their high quality mates (MPOC, MF, MCS, HD and ED), each was assessed as a function of RAC shape, perimeter and area. The results are illustrated using continuous probability density functions (PDFs) constructed using Gaussian kernel density estimators, with Chi-square significance testing ($\alpha = 0.05$) as a function of 10 evenly divided discrete score bins (where bin shading in individual plots indicates significance). Note that only selected plots of interest are provided in order to illustrate relevant results.

2.3.D Similarity as a Function of RAC Shape

Differences in similarity scores based on RAC shape (linear/elongated, isometric and irregular) were detected for 99.5% of the data for all metrics, except matched filter, as per the Chi-square test (McHugh, 2013) with $\alpha = 0.05$. For MPOC, HD and ED, isometric features (circles and triangles) exhibited higher similarity scores, while linear/elongated features (lines and curves) exhibited lower similarity scores. This is illustrated in Figs. 17 and 18 for the MPOC and HD metrics, respectively. This trend is believed to be a function of rotational variation. For example, a circular RAC can tolerate orientation differences reasonably well (*i.e.*, no matter how you rotate a circle, the distance between features remains relatively consistent, as illustrated in Fig. 19). Conversely, an elongated feature, when rotated, is likely to exhibit a drastic decrease in correlation between its known match RAC. Likewise, if confronted with similarly sized but non-matching circles from two different shoes, and similarly sized but

non-matching lines from two different shoes, a rotational offset between the circles may go undetected, but the rotational offset between the linear features may very well be the only way to differentiate the non-matching lines.

Modified Phase Only Correlation for Linear, Isometric, and Irregular Shaped Features

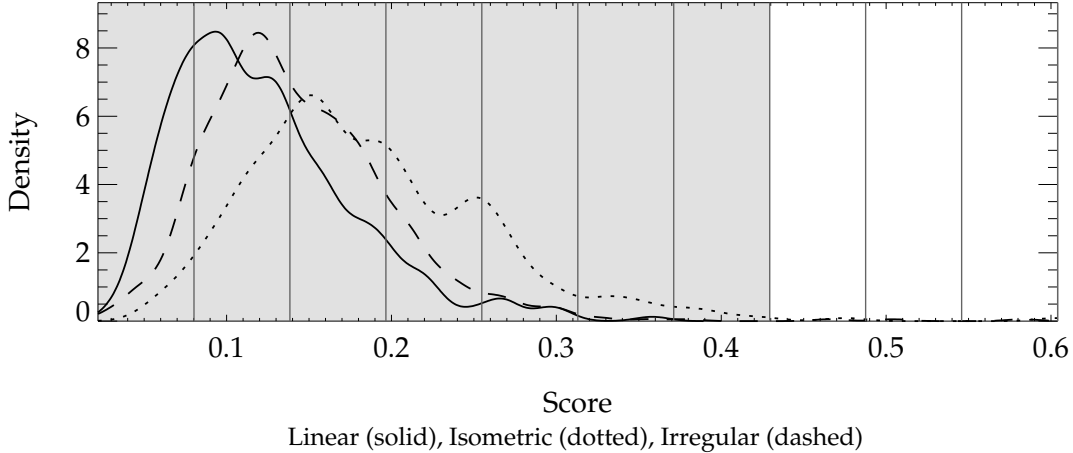


Figure 17: Modified phase only correlation scores as a function of RAC shape. Note that isometric features (circles and triangles) exhibit higher similarity scores (closer to 1.0), while linear/elongated features (lines and curves) exhibit lower similarity scores (closer to 0.0). Shaded score bins represent significance differences in score as a function of shape based on the Chi-square test. Note that differences in scores as a function of RAC shape were observed for 99.5% of the data.

Interestingly, the opposite trend (Fig. 20) was observed for MCS wherein linear/elongated features exhibited the highest similarity scores. Although it is difficult to conceptualize why this may be true, the results do match the mathematics (*e.g.*, if you compare two slightly misaligned/noisy lines and two slightly misaligned/noisy isometric features, computationally, the linear/elongated features report smaller angular differences in n-dimensional space). Also of equal importance is the fact that the matched filter expressed the *least dependence* (95.3% of the data) between score and shape (Fig. 21).

2.3.E Similarity as a Function of RAC Size

Mirroring the results for MF as a function of shape, the majority of the matched filter scores (94.2% for perimeter and 92.6% for area) *did not* exhibit dependence on RAC size, as illustrated in Figs. 22 and 23. However, differences in similarity scores based on RAC size (perimeter and area) were detected for 99.5% of the data with MPOC, HD and ED, as per the Chi-square test (McHugh, 2013) with $\alpha = 0.05$. In other words, the similarity scores for different sized RACs were significantly different from those expected if the variables were independent. Differences in similarity score based on RAC perimeter were also detected for 99.8% of the data with the MCS metric (but only minimally dependent on area for 48.7% of the data). For MPOC, HD and ED, RAC size and score varied inversely, matching intuition. This is illustrated in Figs. 24 and 25 for Euclidean distance as a function of perimeter and area,

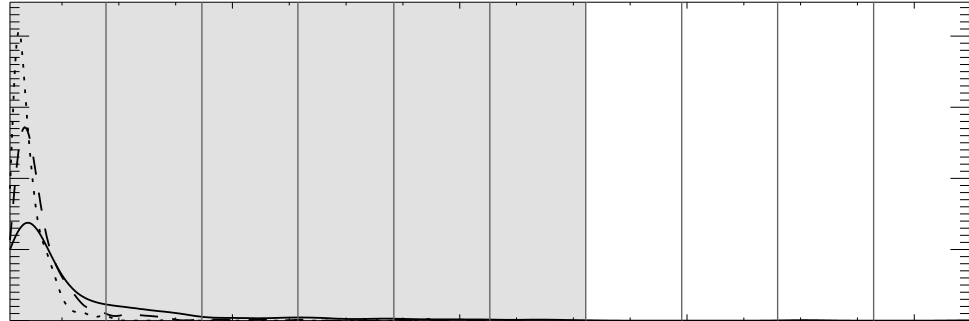


Figure 18: Hausdorff scores as a function of RAC shape. Note that isometric features (circles and triangles) exhibit higher similarity scores (closer to 0.0), while linear/elongated features (lines and curves) exhibit lower similarity scores (further from 0.0). Shaded score bins represent significance differences in score as a function of shape based on the Chi-square test. Note that differences in scores as a function of RAC shape were observed for 99.5% of the data.

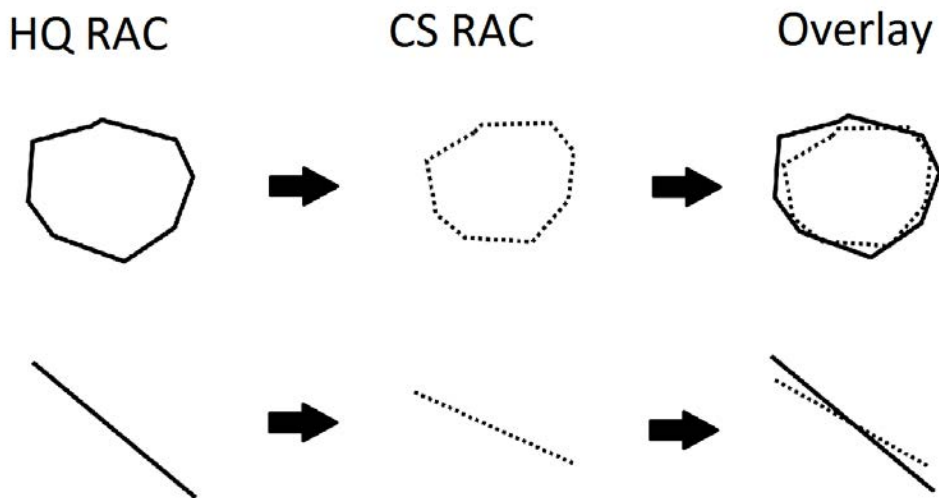


Figure 19: Example of stylized high quality (HQ) and crime scene (CS) RACs. Note that lines exhibit greater discordance (overlap very little) as compared to circular shapes when orientation differences exists (scale and rotational differences are shown for maximum emphasis).

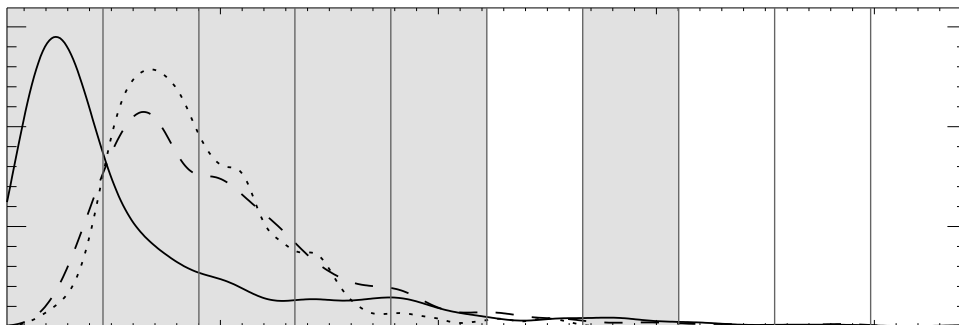


Figure 20: Modified cosine similarity scores as a function of RAC shape. Note that linear/elongated features (lines and curves) exhibit higher similarity scores (closer to 0.0). Shaded score bins represent significance differences in score as a function of shape based on the Chi-square test. Note that differences in scores as a function of RAC shape were observed for 99.5% of the data.

respectively. In other words, as RAC size increased, similarity scores decreased. This likely occurred because large features can reproduce as several smaller and segmented versions of their original, more-complex self when created under variable crime scene-like conditions (Fig. 26). Due to this phenomena, each individual smaller segment from the crime scene-like RAC may compare back to a single larger feature in the high quality impression, yielding a lower numerical score unless manual intervention or unsupervised probability models are introduced to link disconnected features back together. In other words, an automated metric can provide a baseline numerical assessment of similarity that can be very beneficial moving forward, but this illustration shows that the score still requires expert interpretation in determining the conditions that warrant sub-RAC linkages prior (or post) similarity computation.

Matched Filter for Linear, Isometric, and Irregular Shaped Features

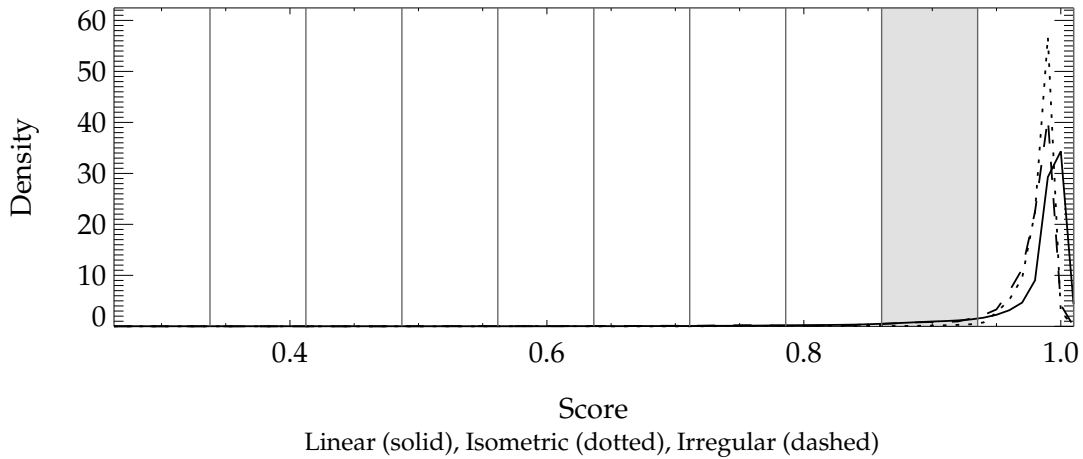


Figure 21: Matched filter scores as a function of RAC shape. Note the lack of dependence on score and shape for nearly 95.3% of the data.

Matched Filter as a Function of Increasing Perimeter

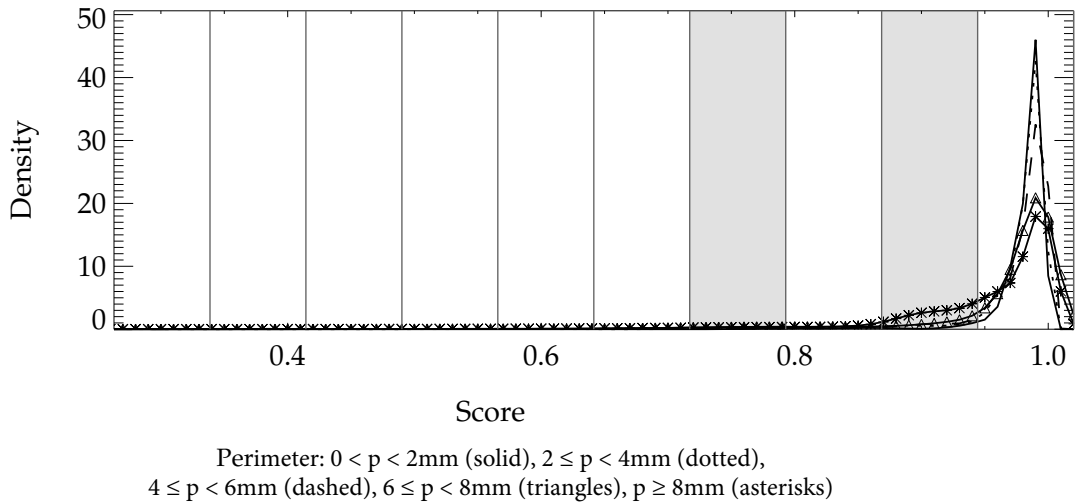


Figure 22: Matched filter similarity score as a function of RAC perimeter. Note the lack of dependence on score and perimeter for nearly 94.2% of the data.

Matched Filter as a Function of Increasing Area

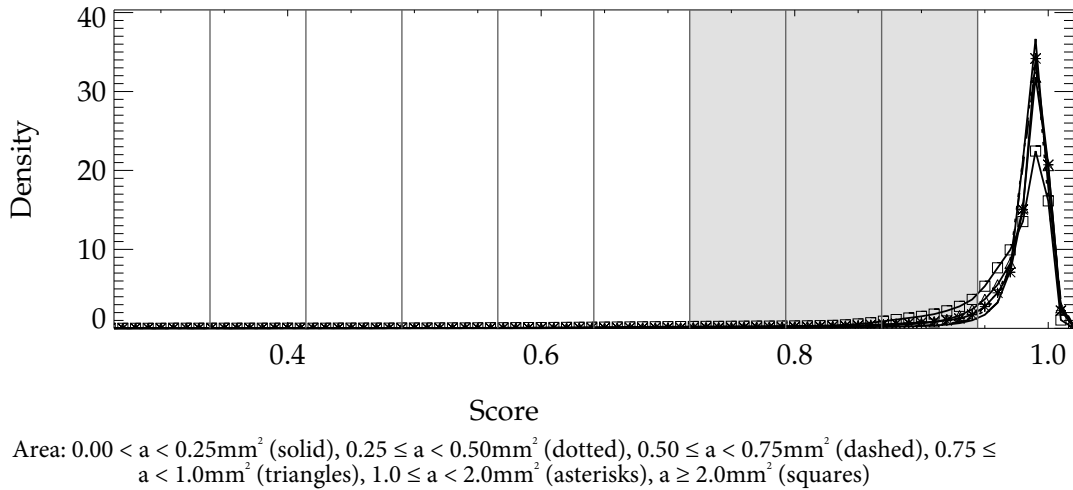


Figure 23: Matched filter similarity score as a function of RAC area. Note the lack of dependence on score and area for nearly 92.6% of the data.

Euclidean Distance as a Function of Increasing Perimeter

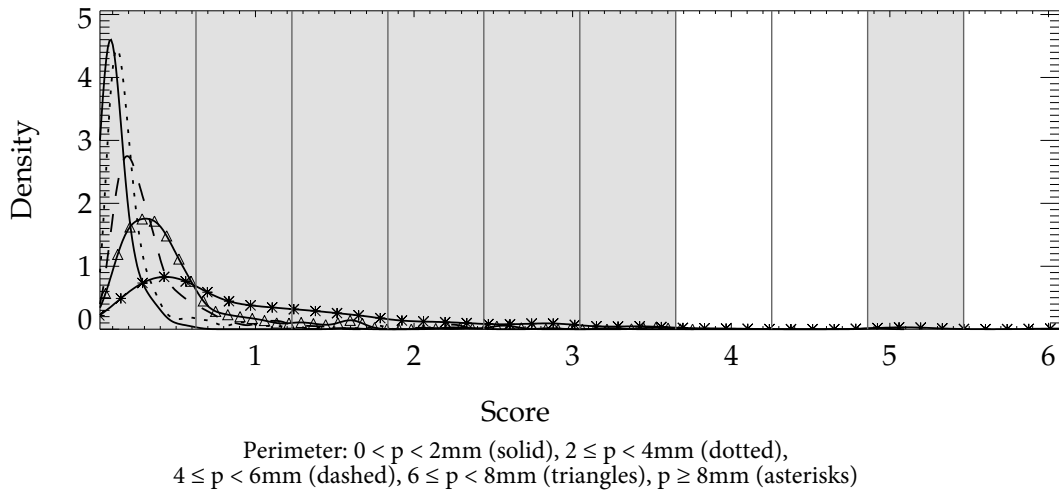


Figure 24: Euclidean distance as a function of RAC perimeter. Note that differences in score as a function of perimeter were observed for 99.5% of the data.

Euclidean Distance as a Function of Increasing Area

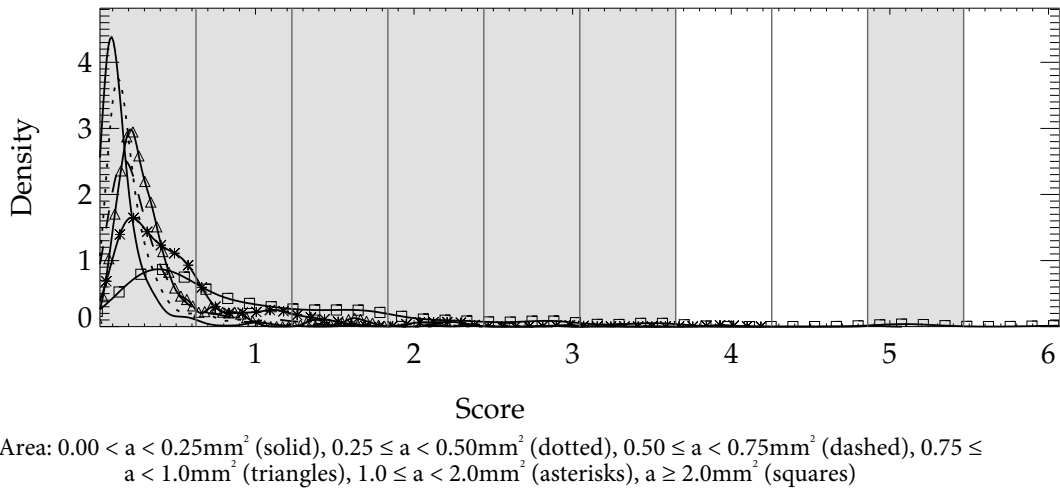


Figure 25: Euclidean distance as a function of RAC area. Note that differences in score as a function of area were observed for 99.5% of the data.

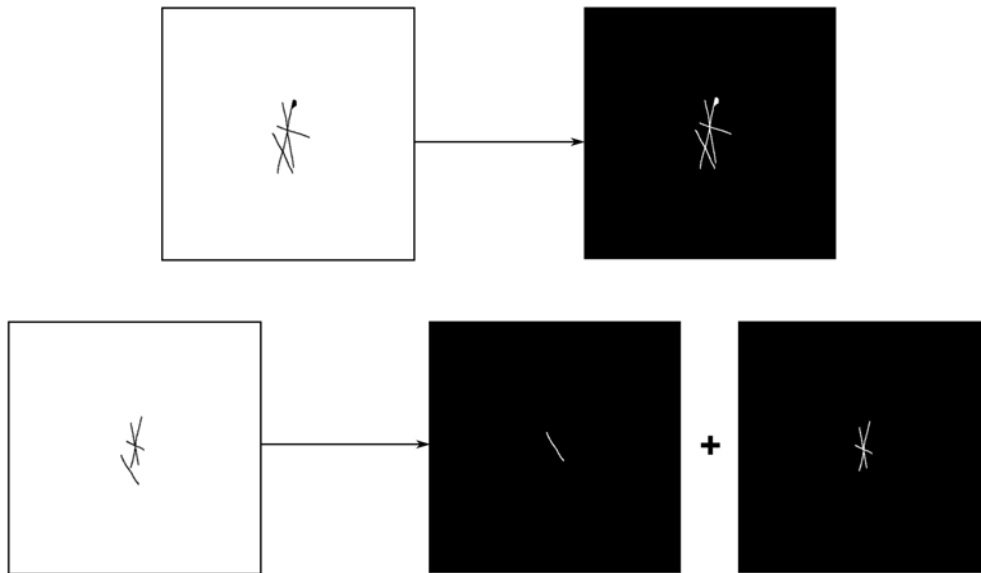


Figure 26: Original marked RAC on high quality exemplar (top left) with corresponding RAC image obtained through connected components (top right). Corresponding RAC on crime scene-like print (bottom left) and RAC images obtained through connected components (bottom center and right). The crime scene-like RACs exhibit more voids and are incomplete in comparison with their high quality counterparts.

2.4 Differentiating between KM and KNM Crime Scene RACs using Similarity Metrics

Each crime scene-like RAC was compared to all high quality KMs and KNMs with coincidental association in position within a 5mm x 5mm cell around the questioned RAC's centroid position within a database consisting of 1,000 shoes. Of the more than 57,000 high quality RACs within this database (at the time of analysis), a total of 44,230 exhibited chance association in position (within the previously defined 5mm x 5mm cells). Based on RAC density, each crime scene-like RAC was compared to an average of 72 ± 21 (one standard deviation) other RACs, with a maximum of 126 comparisons, and a minimum of 15 comparisons (*e.g.*, the most populated 5mm x 5mm cell (again, at the time of this analysis) contained 126 RACs with coincidental association in position, while the least populated had only 15 RACs with coincidental association in position). In total, 8,830 KM pairwise comparisons and 755,380 KNM pairwise comparisons were evaluated (KMs = $1,766 \times 5$ metrics = 8,830 and KNMs = $151,076 \times 5$ metrics = 755,380). The performance of each similarity metric was then evaluated using modified Cumulative Match Characteristic (CMC) and Receiver Operator Characteristic (ROC) curves — including associated Area Under the Curve (AUC) integrals.

Fig. 27 depicts the CMC curve for each similarity metric using the crime scene-like RAC as the query image, and the HQ database as possible mates. Similarly, Fig. 28 depicts the corresponding ROC curves (line designations are as follows: Hausdorff distance = solid, Euclidean distance = dotted, modified cosine similarity = dashed, matched filter = dash-dot, modified phase only correlation = dash-dot-dot-dot).

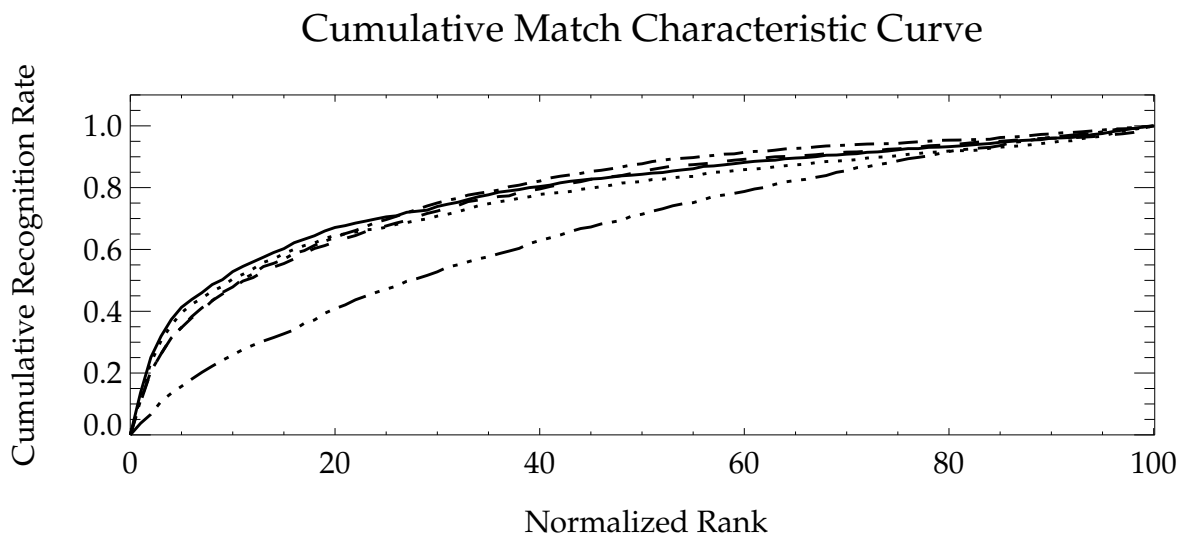


Figure 27: Cumulative match characteristic curves for similarity metrics. Line designations are as follows: Hausdorff distance = solid, Euclidean distance = dotted, modified cosine similarity = dashed, matched filter = dash-dot, modified phase only correlation = dash-dot-dot-dot.

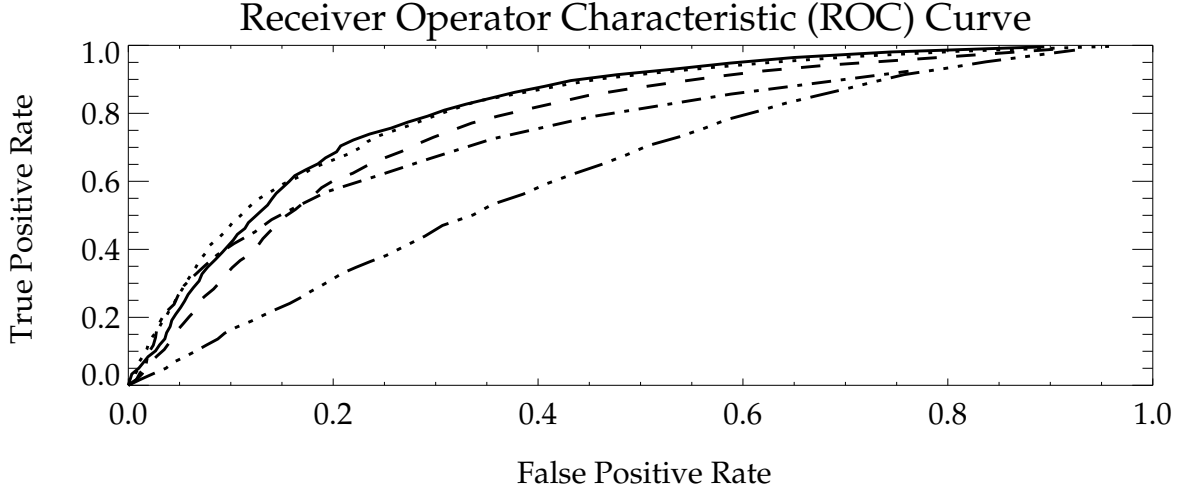


Figure 28: Receiver operator characteristic curve of RAC similarity results. Line designations are as follows: Hausdorff distance = solid, Euclidean distance = dotted, modified cosine similarity = dashed, matched filter = dash-dot, modified phase only correlation = dash-dot-dot-dot.

In addition to the CMC and ROC curves, Table 19 reports the ROC AUCs (or the probability of a randomly selected known match RAC pair exhibiting a higher similarity score than a known non-match pair (also known as stochastic dominance)). Based on all observations, Hausdorff and Euclidean distance metrics yield the best results, and this performance is significantly better than that of the remaining three techniques (at $p < 0.05$).

Table 19: AUC of ROC curves for five metrics used to assess RAC similarity. All scores are significantly different from each other at $p < 0.05$, with the exception of Hausdorff and Euclidean distances ($p = 0.986$).

HD	ED	MCS	MF	MPOC
0.8152	0.8151	0.7706	0.7462	0.6289

Note that the results regarding RAC similarity obtained using five quantitative methods suggest two additional, but equally important, conclusions. First, not all similarity metrics are created equal. This is evident based on the dependence of MPOC, HD, ED, and MCS, (as well as the lack of dependence of MF) on RAC shape and size. For example, if one were to argue that all RACs are *not* equally valuable, then it appears that *not* all numerical objective metrics of similarity are equally able to discern this inherent value. Conversely, if one were to argue that all RACs are of *equal* value, then again, not all numerical objective metrics of similarity are equally able to convey this message. However, the authors must acknowledge that shape and size (perimeter and area) may very well be interrelated factors, and this interdependence has not been tested here (*e.g.*, the dependence of score on size for a given metric could be a function of shape, say, for example, if all linear/elongated features cluster into a particular size category). Although this interdependence has not been examined, the more important point

is that numerical metrics exhibit particular behaviors, and that these behaviors can impact conclusions. In other words, the results of an objective numerical comparison, just like the results of two competing subjective experts, will not necessarily remove disagreement unless the community fully agrees on a single approach and the reason for employing a specific type of metric for all comparisons.

Second, the similarity metrics used in this study are more adept at measuring ‘exactness’, which is really not a reasonable expectation given the physical variation in print quality that is typically encountered during the commission of a crime. Although the experienced footwear examiner innately recognizes mitigating factors that can impact the quality and degree of correspondence between randomly acquired characteristics on reported matches and those anticipated by random chance alone, it is much more difficult to train an automated algorithm to handle such a wide range of possible deformations. Moreover, several exact correspondences may not be of equal or greater value than a single inexact match. For example, a single large, complex feature may be considered ‘rare’ enough to justify an identification, whereas several smaller features with basic geometries may not provide enough information, even in combination, to warrant this same conclusion. However, without expanding the quantitative metrics tested here (to incorporate complexity or additional examiner input), the smaller ‘insignificant’ correspondences may very well provide a higher numerical similarity. For a visual illustration, consider Fig. 29, which shows that the dynamic nature of print creation can result in a single RAC reproducing as several disjointed and smaller features, apparent only upon inspection of the high quality imagery (e.g., when the analyst uses *a priori* information to try to understand the scores). Thus, the performance of each of these metrics can be improved if combined with human intervention such as supervised partitioning/pairing or the use of additional probability models to re-group disjointed RACs.

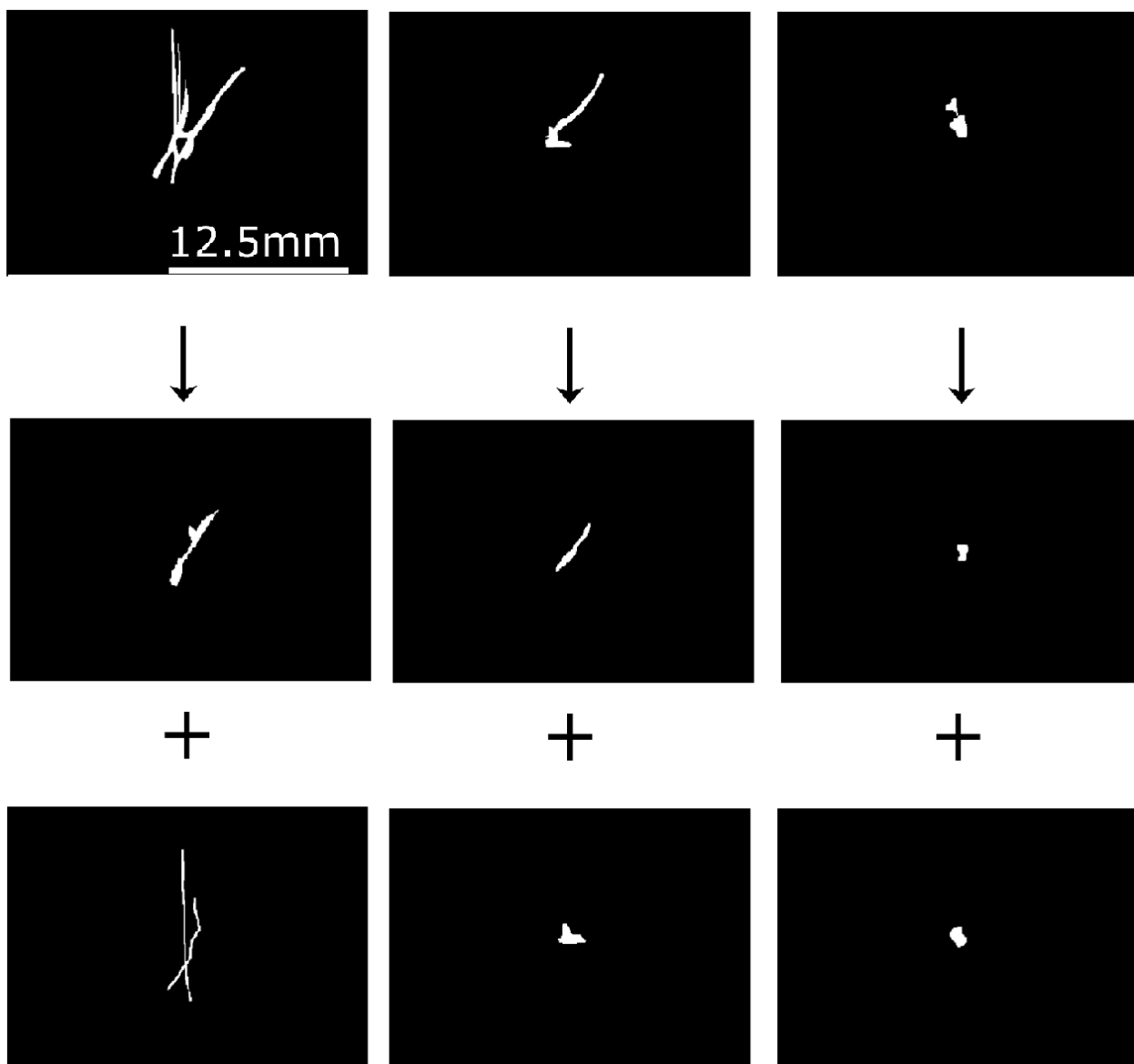


Figure 29: Illustration of segmentation of RACs in crime-scene-like impressions (center and bottom row) as compared to their high quality mate (top row). The RAC in the left column originated from the heel region of a men's size 11 Merrell® Goretex XCR hiking shoe with heavy wear. The feature in the center column was identified in the heel region of a men's size 8 Reebok® DMX Foam athletic shoe with moderate wear. Lastly, the accidental in the right column was localized in the heel region of a men's size 11.5 Nike® athletic shoe with heavy wear.

2.4.A RAC Map Correlation

In addition to individual RAC characterization and comparison, the entire RAC map for each crime scene-like print was compared back to its high quality exemplar to determine a ‘*global similarity metric*’ or the degree to which the impressions could be linked back to their source. This was accomplished using image-wide phase only correlation according to Eq. 11 (without windowing), and on full RAC maps (8,691 x 8,691 pixels in dimension).

Table 20 reports the total frequency of RACs in the binary maps, which are a comprehensive representation of all accidentals observed on an impression. The POC was computed on all possible RAC map pairs to estimate a *global* similarity score. Results are provided in Fig. 30 (a) as ROC curves displaying the true positive and false positive rate. Based on the POC metric, there was a 0.72 probability that a randomly selected pair of positive maps (known match mates) would result in a higher similarity score than a randomly selected pair of negative maps (known non-match mates). In other words, positive known matches would be correctly ranked in an ordered list 72% of the time. Given that 64% of the query crime scene-like maps contained 10 or fewer RACs, and that an average of 85% of the identified randomly acquired characteristics failed to transfer to the questioned impressions, this result is still considered promising. The same data comparison and interpretation was re-conducted on high quality RAC maps, but at a point in time when 1,261 outsoles had been fully characterized, with 161 duplicates (the duplicates are a product of the mini-study used to determine intra- and inter-analyst variation in marking). This led to 161 known-match comparisons and 794,430 known non-match comparisons, a ROC curve as illustrated in Fig. 30 (b) with an AUC of 0.995, and probability density functions as illustrated in Fig. 31.

Table 20: RAC map density (CS = Crime Scene, HQ = High Quality).

Number of RACs in Map	CS Frequency	HQ Frequency
0	20 (10%)	0 (0%)
1-5	74 (37%)	7 (7%)
6-10	33 (17%)	12 (12%)
11-15	32 (16%)	5 (5%)
16-20	14 (7%)	11 (11%)
21-25	4 (2%)	2 (2%)
Greater than 25	23 (11%)	63 (63%)
Total	200 (100%)	100 (100%)

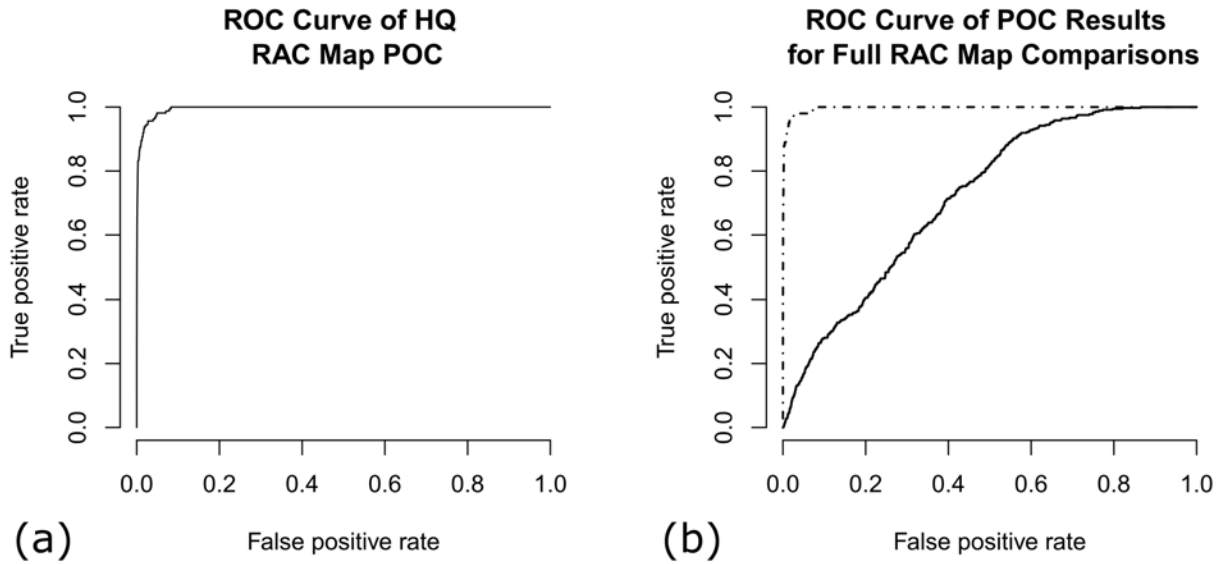


Figure 30: (a) Receiver operator characteristic curve of RAC map POC results. High quality comparisons are represented by the dash-dotted line and exhibit an area under the curve (AUC) of 0.996 (100 KMs and 19,800 KMN; 100 KMs \times 2 duplicates, each cross-compared, creating $n(n-1)/2 = 200(199)/2 = 19,900$ total comparisons, of which 100 are KMs and 19,800 are KNMs). The solid line illustrates the results of crime scene-like impressions with an AUC of 0.719. (b) Receiver operator characteristic curve of RAC map POC results based on 161 KM and 794,430 KNM comparisons, with resulting AUC of 0.995.

Probability Density Functions for KM & KNM High Quality Prints

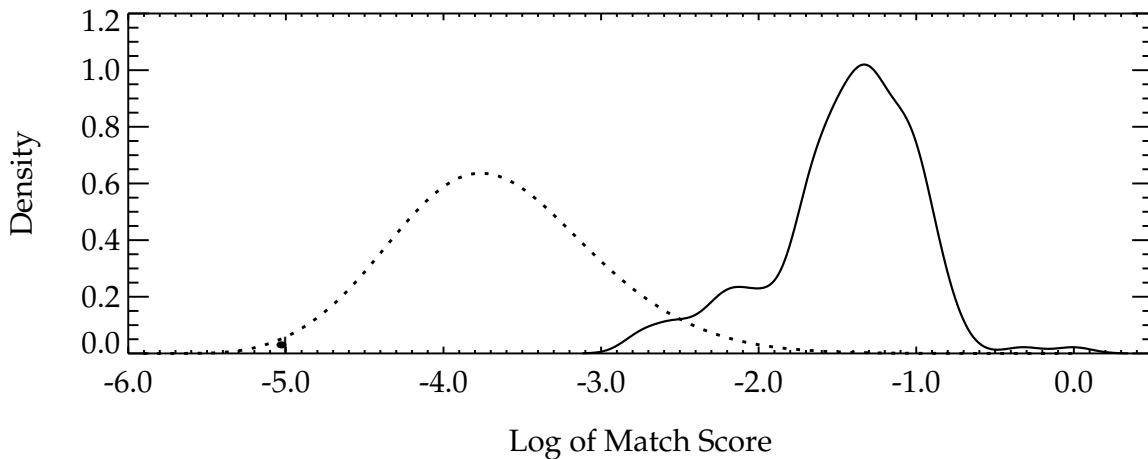


Figure 31: Probability density function describing the separability (in terms of POC match score) for 161 KM and 794,430 KNM comparisons ($n(n-1)/2 = 1,261(1,260)/2 = 794,430$).

2.5 Chance Co-occurrence

Finally, the database was queried for chance co-occurrence and random match probabilities (based on positional and shape category ‘matches’), and associated likelihood ratios as a function similarity metrics for RACs with positional co-occurrence. More specifically, the following three forensically-relevant questions were asked and answered.

1. **Question:** *What is the empirical frequency of selecting two shoes at random, and finding a pair of RACs with positional similarity anywhere on the outsole?* Using the formula for a simple combination (when order does not matter, and repetition is not allowed), the total number of ways $k = 2$ ‘outsoles’ can be selected from $n = 1,300$ (‘total shoes’ in the database) was computed according to Eq. 17.

$$\text{All (Possible Pairs)} = \frac{n!}{(n - k)!k!} = 844,350 \quad (17)$$

Next, the dataset was queried to determine (empirically) the number of times two shoes exhibited a pair of RACs with positional similarity. When this empirical result is divided by the total number of ways this could happen (or ‘All’ in Eq. 17) a normalized frequency of occurrence (or probability) is obtained. Results are summarized in Table 21 for ‘any shaped’ pairs, and for pairs where both RACs are identified as ‘elongated’, both are identified as ‘irregular’ and both are identified as ‘isometric’.

Table 21: Normalized empirical frequency (probability) and chance co-occurrence for RAC pairs of any shape, and specific shapes (elongated, irregular and isometric).

Metric	Any Shape	Elongated	Irregular	Isometric
Probability	0.58	0.27	0.21	0.095
Chance Co-occurrence	1 in 2	1 in 4	1 in 5	1 in 10

Note that the results provided in Table 21 are in alignment with reports from past empirical studies (Cassidy, 1995). More specifically, Cassidy (1995) found between a 1 in 6 chance of finding 10 ‘minute’ characteristics, a 1 in 20 chance of finding 3 moderate-sized characteristics, and 1 in 38 (to 1 in 60) chance of finding a single moderate-sized characteristic with coincidental similarity across known non-match heels (Cassidy, 1995).

2. **Question:** *What is the empirical frequency of selecting two shoes at random and finding a positional match at a specific location?* This is like asking for the empirical frequency of selecting two (unrelated) shoes at random, and finding RACs of shape type ‘A’ with positional association at location ‘B’, which is the forensic activity of identifying a crime scene print with RAC shape type ‘A’ at location ‘B’, and then asking for the random match probability (or the empirical frequency) of finding RAC shape type ‘A’ at location ‘B’ in an unrelated dataset (assuming that the empirical database described in this document can be used to model the ‘relevant population’).

To answer this question, each spatial cell in the dataset was queried to determine (empirically) the number of times two shoes exhibited a pair of RACs with positional similarity, resulting in 990 bin-specific frequencies, each divided by ‘All’ in Eq. 17, and converted to chance co-occurrence (or 1 in X), as illustrated in Table 22.

Table 22: Summary statistics (per spatial bin of the normalized outsole) describing chance co-occurrence for RAC pairs of any shape, and specific shapes (elongated, irregular and isometric). Note that the best case scenario lacks locations where no repeats were detected, and the row labeled ‘unaccounted’ reports the number of bins with zero or a single RAC.

Metric	Any Shape	Elongated	Isometric	Irregular
Mean Probability	2.251E-04	6.920E-05	1.919E-05	5.539E-05
Chance Co-occurrence	1 in 4,441	1 in 14,450	1 in 52,120	1 in 18,055
Median Probability	2.237E-03	4.808E-04	1.078E-04	3.554E-04
Chance Co-occurrence	1 in 447	1 in 2,080	1 in 9,279	1 in 2,814
Worst Probability	9.615E-03	3.559E-03	1.112E-03	1.757E-03
Chance Co-occurrence	1 in 104	1 in 281	1 in 893	1 in 569
Best Probability	1.184E-06	1.184E-06	1.184E-06	1.184E-06
Chance Co-occurrence	1 in 844,350	1 in 844,350	1 in 844,350	1 in 844,350
Unaccounted	9	15	31	22

Note that the worst case scenario (greatest chance association) was found to be a chance of 1 in 104 (or a probability of 9.61E-03) and the best case scenario (lowest chance association, excluding bins where no repeats occurred) was a chance of 1 in 844,350 (or a probability of 1.18E-06). However, it must be noted that the best case scenario is bounded by the size of the database (*i.e.*, bins with a single positional co-occurrence between all 1,300 shoes). Thus, if the database size increases, the denominator will naturally increase, but the numerator is likely to remain constant for one or more bins, thus causing a more rare ‘best case’ chance association.

3. **Question:** *How similar are RACs with positional co-occurrence?* RACs with positional co-occurrence (and even identical shape categorizations), are not necessarily geometrically similar. For example, two linear elements could vary in orientation, length, thickness, curvature, etc. Thus, the mathematical similarity of RACs with coincidental positional similarity was determined. First, the HD dissimilarity of 6,993 KM comparisons was computed (of which 5,227 were high quality versus high quality RAC comparisons (either marked by different analysts, or the same analyst repeatedly), and 1,766 where high quality versus crime scene-like RAC comparisons). Next, the HD dissimilarity of 3,239,114 KNM comparisons was computed (based on all RACs in this database with positional association). Using the maximum HD dissimilarity value, all scores were normalized such that a normalized HD score of 1.0 would signify indistinguishable RACs (within this database and the resolution possible using Hausdorff distance), and a score of 0.0 would be associated with the most dissimilar pairwise comparison between two known non-match RACs in this dataset.

All resulting normalized scores were used to create probability density functions (PDFs), and then the likelihood ratio (LR) was computed as the ratio of the KM and KNM PDFs at all normalized HD values. The cumulative results are illustrated in Fig. 32, where the

solid curve represents the probability density of the KNMs, the dotted line represents the probability density of the KMs, and the dashed line represents the associated LR value. Note that a normalized HD score between 0.9930 to 0.9931 represents the boundary LR (or an LR = 1), the LR equals 27 at the maximum KM density, and that the LR climbs to 6,945 between HD values of 0.9994-0.9995. Given the boundary LR, the number of normalized HD KNM scores greater than 0.9930 was queried, and found to occur 414,417 times (13% of all comparisons). Again, this does not mean that the compared RACs actually appear visually indistinguishable, but only that they are mathematically similar based on the similarity metric employed to compare them. In order to cross-compare the *visual similarity* of the *mathematically similar* RACs, the top 8 most-similar RAC KNM pairs with positional similarity (per spatial bin) will be physically examined — 8 pairs x 990 bins x 4 shape classes = 31,680 visual comparisons (modified by the number of bins with fewer than 5 RACs). These results are forthcoming, but based on inspection of 1,000 pairs of outsoles, with 57,426 RACs, creating 2,022,595 pairwise comparisons, and examination of the top 5 matches per bin using the matched filter (MF) similarity metric previously described, only 25 pairs (out of 19,800 examined) were found to be visually indistinguishable. Moreover, of these 25 indistinguishable pairs, all but two were on shoes that could be distinguished based on manufacturer make/model. For the remaining two pairs, one pair was on the left and the right of a matched set, and the other pair originated from shoes of the same make/model, but with a different manufacturer size, and exhibiting different degrees of wear. Thus, all 2,022,595 comparisons were ultimately deemed distinguishable.

Probability Density of HD Scores for KM and KNM RACs

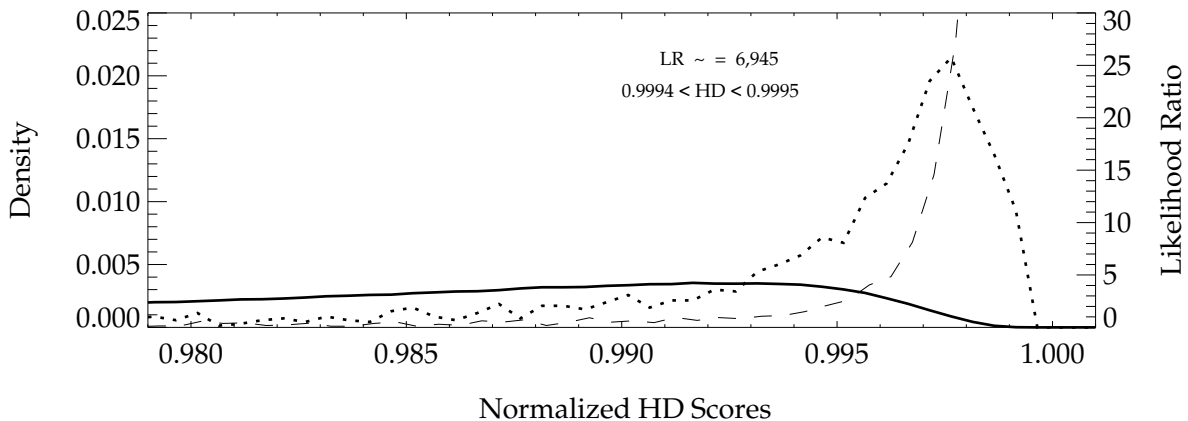


Figure 32: Probability density of normalized HD scores for KNMs (solid line), KMs (dotted line) and associated LR values (dashed line).

Note that the above can be expanded to account for additional associations in RACs at more than one location and more than one type within the dataset, but this is reserved for dynamic applications in a web-based application (described in subsection 2.5.A and illustrated in Fig. 33).

2.5.A Use of Similarity in the Web Application

Based on the results revealed in Fig. 28 and Table 19, HD was deemed the most appropriate metric to use moving forward with the web application (Fig. 33). With this in mind, every HQ RAC with positional co-occurrence (localized to a specific bin) was pairwise compared using Hausdorff distance. The results (per bin) were sorted and the top 8 ‘hits’ (or the 8 most mathematically similar RACs) are made available to the user for visual comparison. These results are available using a web-based application (Fig. 34) that allows the viewer to visually inspect RACs with positional association (leading to visual pairwise comparisons). Part of the motivation for this is the fact that the similarity metrics themselves are limited in terms of applicability, and partially to enhance efficiency (*i.e.*, inspection of all pairwise comparisons by an examiner is a tedious task, even when a bin only contains a few RACs — for example, just 10 RACs with positional similarity would give rise to $n(n - 1)/2 = 45$ pairwise comparisons). The cumulative result is refinement of the chance co-occurrence estimate based on actual RAC shape, size and complexity in order to determine *if* the random association is forensically relevant or just the numerically interesting.

	Any Shape	Irregular	Elongated	Approximately Isometric
Total RACs in Database	72,306	25,420	32,549	14,337
Total RACs in Cell	182	43	109	30
Chance of Finding RAC in Cell	1 in 397	1 in 1,681	1 in 633	1 in 2,410
Chance of RAC Co-Occurrence	1 in 116	1 in 1,268	1 in 281	1 in 2,234
Similarities & Likelihood Ratios	All Shapes	Irregular	Elongated	Approximately Isometric

Few RACs  Many RACs

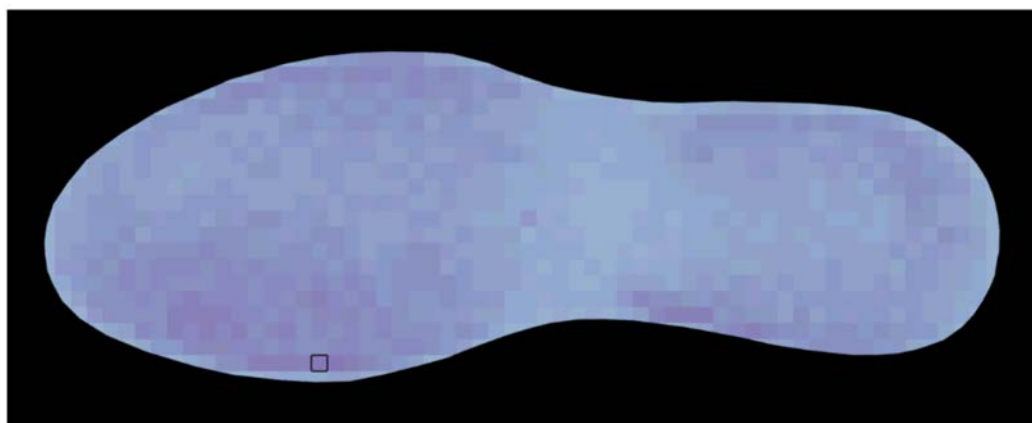


Figure 33: Static illustration of web application: <http://www.4n6chemometrics.com/database/>. The user can select a cell of interest, obtain frequency information associated with the cell, and then examine the most similar pairwise-compared RACs by selecting the buttons in the row labeled ‘Similarity & Likelihood Ratio’. Once selected, the user is directed to a second web page that allows for a side-by-side comparison of the 8 most similar RACs of interest.

However, it is also important to remind the reader that RAC-specific likelihood ratios (LR) are not yet available for this dataset. In the interim, this webpage reports generalized LRs only. More specifically, the LR is the probability of a NHD score, given that the score has been sampled from a probability density function of scores generated when evidence is known to be from the same source, versus the probability of the same NHD score, given that the score has been sampled from a probability density function of scores generated when the evidence is known to be from different sources. However, the actual RACs being compared have not been reproduced repeatedly to generate individual (e.g., RAC-specific) known match probability density functions, and although the RAC in question has been repeatedly compared when generating the known non-match probability density function, this function also contains scores based on thousands of other unrelated known non-match RAC comparisons.



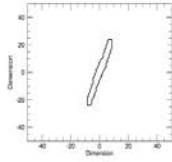
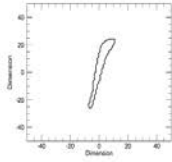

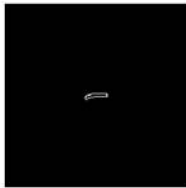
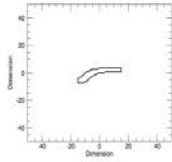
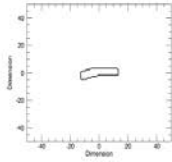
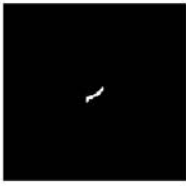

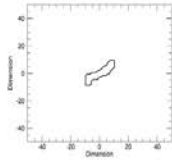
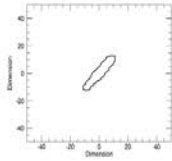
Rank	RAC 1 Image	RAC 2 Image	RAC 1 Fourier Descriptor Image	RAC 2 Fourier Descriptor Image	Score	Likelihood Ratio
1					0.99686	10
2					0.99655	7
3					0.99634	5

Figure 34: Static illustration of top 3 pairwise-compared RACs for a selected bin (note that the web page actually provides the top 8, including images of the Fourier descriptors, the normalized HD similarity score, and the associated likelihood ratio).

2.6 Impact, Outcomes, Evaluation & Dissemination

The contracted research was required to shed light on 3 research objectives of interest to NAS, NIJ and SWGTREAD (SWGTREAD, 2013):

1. Random shape and/or placement of accidental marks;
2. Mathematical probabilities of individual characteristics;
3. Frequency of style, size, design and uniqueness in footwear.

Deliverables were as follows:

1. An evaluation of the discrimination potential of a numerical metric of similarity when comparing footwear accidental patterns:
 - (a) Density functions describing the probability of similarity scores for known match and known non-match comparisons;
 - (b) False positive and false negative error rates, including receiver operator characteristics (ROC) curves;
 - (c) A Bayesian interpretation of similarity scores.
2. A tool and dataset that can be integrated into existing training procedures and policies:
 - (a) An electronic collection of exemplar and test impressions;
 - (b) A web application that reports empirical frequency estimates related to close non-matches.
3. A framework for databasing studies (frequency of make, model and accidental characteristics):
 - (a) Positional frequency and estimates of chance occurrence of accidental features;
 - (b) Frequency of accidental types.

2.7 Added Value

1. Began a collaboration with the FBI to share methodology/code that will allow the FBI to process their footwear database using the approach created by our research group.
 - 11-12-June-2014: A joint meeting between the WVU research group, the project's practitioner partner (Mr. William Bodziak) and two analysts from the FBI (Dr. Brian Eckenrode and Mr. Eric Gilkerson) allowing for both training and professional development for the research students on this project (at the time, one undergraduate and two graduate students). Note that the most significant outcome was that each student was given general and specific instructions regarding

the identification of randomly acquired characteristics on outsoles, and how to differentiate microcellular bubbles, model features, texture and Schallamach patterns. The end result was a mini-proficiency testing session where each student interacted with the expert practitioners to receive specific feedback on his or her ability to detect and mark the acquired features.

- 28-May-2015: Teleconference between WVU, NIST and the FBI group.
- 30-June-2015: Teleconference between WVU, NIST and the FBI group.
- 15-16-July-2015: Joint meeting between the WVU research group, the project's practitioner partner (William Bodziak), and FBI analysts/visiting scientists. During this visit, Mr. Bodziak provided a training session to WVU and FBI analysts on the generation and collection of crime scene prints. This meeting was extended to 17-July-2015 between the WVU research group and the FBI group. During the full three-day time period, the WVU group trained three FBI visiting scientists (Brent Allred, Nicholas Vercruyssen and Andrew Plotner) in data acquisition and extraction. In addition, we collaborated and had extended discussions concerning the project with one FBI program manager (Brian Eckenrode) and two FBI footwear examiners (Brian McVicker and Eric Gilkerson). At present, we have an unofficial agreement that the FBI will analyze their database of 776 pairs of boots using the algorithms WVU has created. On 07-July-2015, WVU gave the FBI group code responsible for image pre-process.
- 15-16-September-2016: Joint meeting between the WVU research group and three FBI employees (Dr. Brian Eckenrode, Nicholas Vercruyssen and Katherine Ky). Note that data analysis sharing is ongoing between the two teams.

2.8 Publications & Abstracts

- Richetelli, N., Lee, M., Lasky, C., Gump, M., and Speir, J. *Classification of Footwear Outsole Patterns using Fourier Transform and Local Interest Points*. **Forensic Science International**. Vol. 275, 2017, pp. 102-109.

Abstract: Successful classification of questioned footwear has tremendous evidentiary value; the result can minimize the potential suspect pool and link a suspect to a victim, a crime scene, or even multiple crime scenes to each other. With this in mind, several different automated and semi-automated classification models have been applied to the forensic footwear recognition problem, with superior performance commonly associated with two different approaches: correlation of image power (magnitude) or phase, and the use of local interest points transformed using the Scale Invariant Feature Transform (SIFT) and compared using Random Sample Consensus (RANSAC). Despite the distinction associated with each of these methods, all three have not been cross-compared using a single dataset, of limited quality (*i.e.*, characteristic of crime scene-like imagery), and created using a wide combination of image inputs. To address this question, the research presented here examines the classification performance of the Fourier Mellin transform (FMT), phase-only correlation (POC), and local interest points (transformed using SIFT and compared using RANSAC), as a function of inputs that include mixed

media (blood and dust), transfer mechanisms (gel lifters), enhancement techniques (digital and chemical) and variations in print substrate (ceramic tiles, vinyl tiles and paper). Results indicate that POC outperforms both FMT and SIFT + RANSAC, regardless of image input (type, quality and totality), and that the difference in stochastic dominance detected for POC is significant across all image comparison scenarios evaluated in this study.

- Richetelli, N., Nobel, M., Bodziak, W., and Speir, J. *Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape*. **Forensic Science International**. Vol. 270, 2017, pp. 211-222.

Abstract: Forensic footwear evidence can prove invaluable to the resolution of a criminal investigation. Naturally, the value of a comparison varies with the rarity of the evidence, which is a function of both manufactured as well as randomly acquired characteristics (RACs). When focused specifically on the latter of these two types of features, empirical evidence demonstrates high discriminating power for the differentiation of known match and known non-match samples when presented with exemplars of high quality and exhibiting a sufficient number of clear and complex RACs. However, given the dynamic and unpredictable nature of the media, substrate, and deposition process encountered during the commission of a crime, RACs on crime scene prints are expected to exhibit a large range of variability in terms of reproducibility, clarity, and quality. Although the pattern recognition skill of the expert examiner is adept at recognizing and evaluating this type of natural variation, there is little research to suggest that objective and numerical metrics can globally process this variation when presented with RACs from degraded crime scene quality prints. As such, the goal of this study was to mathematically compare the loss and similarity of RACs in high quality exemplars versus crime scene-like quality impressions as a function of RAC shape, perimeter, area, and common source. Results indicate that the unpredictable conditions associated with crime scene print production promotes RAC loss that varies between 33% and 100% with an average of 85%, and that when the entire outsole is taken as a constellation of features (or a RAC map), 64% of the crime scene-like impressions exhibited 10 or fewer RACs, resulting in a 0.72 probability of stochastic dominance. Given this, individual RAC description and correspondence were further explored using five simple, but objective, numerical metrics of similarity. Statistically significant differences in similarity scores for RAC shape and size were consistently detected for three of the five metrics (modified phase only correlation, Euclidean distance, and Hausdorff distance). Conversely, a single metric (the matched filter) expressed the least dependence between score and both shape and size. Moreover, for all crime scene-like RACs with coincidental association in position, the matched filter produced the greatest discrimination potential in sorting known matches and known non-matches. Despite this demonstrated success, numerical metrics of similarity are not without limitations, and the remainder of this work provides commentary on the difficulties associated with using objective metrics when faced with segmentation, incomplete information, and low signal-to-noise ratios.

- Speir, J., Richetelli, N., Fagert, M., Hite, M., and Bodziak, W. *Technical Note: Quantifying randomly acquired characteristics on outsoles in terms of shape and position*.

Forensic Science International. Vol. 266, 2016, pp. 399-411.

Abstract: Footwear evidence has tremendous forensic value; it can focus a criminal investigation, link suspects to scenes, help reconstruct a series of events, or otherwise provide information vital to the successful resolution of a case. When considering the specific utility of a linkage, the strength of the connection between source footwear and an impression left at the scene of a crime varies with the known rarity of the shoeprint itself, which is a function of the class characteristics, as well as the complexity, clarity, and quality of randomly acquired characteristics (RACs) available for analysis. To help elucidate the discrimination potential of footwear as a source of forensic evidence, the aim of this research is to further characterize the chance association in position, shape, and geometry of RACs on a semi-random selection of footwear. To accomplish this goal in an efficient manner, a partially automated image processing chain was required, including steps for automated feature characterization. This technical note details the methods, procedures, and type of results available for subsequent statistical analysis after processing a collection of more than 1000 shoes and 57,426 randomly acquired characteristics.

- Richetelli, N. *Master's of Science in Forensic Science Thesis: Quantitative Assessment of the Discrimination Potential of Class and Randomly Acquired Characteristics for Crime Scene Quality Shoeprints.* West Virginia University, Department of Forensic & Investigative Science, December, 2015.

2.9 Meetings, Presentations & Invited Talks

- Richetelli, N., Speir, J. *Quantifying the Chance Similarity of Randomly Acquired Characteristics in terms of Shape and Position on High Quality Footwear Exemplars.* Presentation: **International Association for Identification (IAI) Annual Educational Conference**, Cincinnati, OH, 2016.
- Speir, J. *Imaging, Classification, and Quantification of Forensic Pattern Evidence.* **Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science Seminar Series**, Rochester, NY, 11-March-2015.
- Speir, J. *A Quantitative Assessment of Shoeprint Accidental Patterns with Implications Regarding Similarity, Frequency and Chance Association of Features.* Invited Talk and Collaborative Meeting: **National Institute of Standards and Technology**, (Martin Herman, Hariharan Iyer, Simone Gittelsohn, Steve Lund, Yooyoung Lee), Gaithersburg, MD, 23-March-2015.
- Richetelli, N., Fagert, M., Hite, M., Speir, J. *Estimates of Randomly Acquired Characteristic Frequency in High Quality Footwear Exemplars.* Poster: **National Institute of Justice Impression Pattern and Trace Evidence Symposium**, San Antonio, TX, 2015.
- Richetelli, N., Fagert, M., Epler, A., Bodziak, W., Speir, J. *Preliminary Efforts to Quantify the Chance Similarity in Shape and Position of Randomly Acquired Characteristics*

in Footwear. Presentation: **International Association for Identification (IAI) Annual Educational Conference**, Sacramento, CA, 2015.

- Speir, J. *Preliminary Efforts to Quantify the Chance Similarity in Shape and Position of Randomly Acquired Characteristics in Footwear*. Invited Talk: **NSF Statistical and Applied Mathematical Sciences Institute (SAMSI) Forensics Opening Workshop**, Research Triangle Park, NC, 31-August-2015 through 04-September-2015.

2.9.A Limitations

Use and interpretation of the results presented in this summary document are bounded by the following limitations:

1. The database samples are such that a single shoe type, manufacturer, brand and size does not repeat with appreciable frequency, providing only limited power to allow for statistical estimates based on *a priori* class association. Instead, the results ignore class features (including outsole size, perimeter and geometry) and apply only to the *normalized scenario* describe in subsection 2.2.F.
2. The results/metrics (*i.e.*, likelihood ratios, random chance probability, etc.) are specific to the samples collected in this dataset, and should be extrapolated to other scenarios with caution, and only when doing so reasonably approximates the relevant population. In addition, the spatial limit of detection for RACs extracted from high quality imagery is most certainly much lower than that expected in forensic impressions (*e.g.*, only a subset of the largest RACs in this dataset are likely to reproduce in crime scene imagery, and the entire dataset and associated chance co-occurrence estimates should be modified to account for this if any extrapolation to casework is made).
3. Randomly acquired characteristics were manually extracted by analysts (no known automated extraction algorithm has proven to be effective for this purpose), and as such, positional and shape/geometry frequency estimates are limited by intra- and inter-analyst variability in marking.
4. Score-based likelihood ratios are limited by the discrimination potential associated with the similarity metric utilized. In addition, the numerator in all LRs is a generic probability density function (created using scores obtained when sampling known matches, and not the repeated sampling and comparison any specific RAC).
5. Chance co-occurrence in position and position/categorical-shape does not equal a *forensically significance* random match probability.

A. Appendices

A.1 References

- Bartolini, I., P. Ciaccia, and M. Patella (2005). WARP: Accurate retrieval of shape using phase of Fourier descriptors and time warping distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(1), 142–147.
- Bodziak, W. J. (2000). *Footwear Impression Evidence: Detection, Recovery and Examination, Second Edition*. CRC Press.
- Bouridane, A. (2009). *Imaging for Forensics and Security: From Theory to Practice*. Springer.
- Cassidy, M. (1995). *Footwear Identification (Reprint)*. Lightning Powder Company, Inc.
- Champod, C., R. Voisard, and A. Girod (2000). A statistical study of air bubbles on athletic shoe soles. *Forensic Science International* 109, 105–123.
- Dalitz, C., C. Brandt, S. Goebbels, and D. Kolanus (2013). Fourier descriptors for broken shapes. *EURASIP Journal on Advances in Signal Processing* 161, 1–11.
- Davis, R. and J. DeHaan (1977). A survey of men’s footwear. *Journal of the Forensic Science Society* 17, 271–285.
- Davis, R. and A. Keeley (2000). Feathering of footwear. *Science & Justice* 40, 273–276.
- de Chazal, P., J. Flynn, and R. Reilly (2005). Automated processing of shoeprint images based on the Fourier transform for use in forensic science. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 341–350.
- Fawcett, A. (1970). The role of the footmark examiner. *Journal of the Forensic Science Society* 10, 227–244.
- Folkers, A. and H. Samet (2002). Content-based image retrieval using Fourier descriptors on a logo database. *Proceedings of the 16th International Conference on Pattern Recognition* 3, 521–524.
- Gonzalez, R. and R. E. Woods (2008). *Digital Image Processing, 3rd Edition*. Pearson Prentice Hall, New Jersey.
- Gregga, J., G. Power, and K. Iftekharrudin (2002). Correlations for rank order shape similarity measurement. *Optical Pattern Recognition XIII, Proceedings of the SPIE* 4734, 122–131.
- Gueham, M., A. Bouridane, and D. Crookes (2007). Automatic recognition of partial shoeprints based on phase-only correlation. *IEEE*, IV (441–444).
- Hannigan, T., L. Fleury, R. Reilly, B. O’Mullane, and P. deChazal (2006). Survey of 1276 shoeprint impressions and development of an automatic shoeprint pattern matching facility. *Science & Justice* 46, 79–89.

- Huttenlocher, D., G. Klanderman, and W. Rucklidge (1993). Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(9), 850–863.
- McHugh, M. (2013). The Chi-square test of independence. *Biochemica Medica* 23(2), 143–149.
- NAS (2009). Strengthening forensic science in the united states: A path forward; committee on identifying the needs of the forensic sciences community, National Research Council. Technical report, <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>.
- Park, U. and A. K. Jain (2010). Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security* 5(3), 406–415.
- Petraco, N., C. Gambino, T. Kubic, D. Olivio, and N. Petraco (2010). Statistical discrimination of footwear: a method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques. *Journal of Forensic Sciences* 55, 34–41.
- Rasband, W. (1997-2016). ImageJ, U.S. National Institutes of Health. *Bethesda, Maryland, USA*, <http://imagej.nih.gov/ij/>.
- Richetelli, N., M. Nobel, W. J. Bodziak, and J. A. Speir (2017, jan). Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape. *Forensic Science International* 270, 211–222.
- Rosin, P. (2003). Measuring shape: Ellipticity, rectangularity, and triangularity. *Machine Vision and Applications* 14, 172–184.
- Schott, J. R. (2007). *Remote Sensing: The Image Chain Approach, 2nd Edition*. Oxford University Press.
- Sheets, H. D., S. Gross, G. Langenburg, P. J. Bush, and M. A. Bush (2013). Shape measurement tools in footwear analysis: A statistical investigation of accidental characteristics over time. *Forensic Science International* 232, 84–91.
- Shor, Y. and S. Wiesner (2015). Methodological shift in scientifically evaluating shoeprint: A computerized system to statistical aid the expert reach the degree of certainty. *Presented at International Association for Identification (IAI) Centennial Conference Sacramento, CA*.
- Skerrett, J., C. Neumann, and I. Mateos-Garcia (2011). A bayesian approach for interpreting shoemark evidence in forensic casework: Accounting for wear features. *Forensic Science International* 210, 26–30.
- Stone, R. (2006). Footwear examination: Mathematical probabilities of theoretical individual characteristics. *Journal of Forensic Identification* 56, 577–599.
- SWGTTREAD (2006). Guide for the examination of footwear and tire impression evidence; <http://www.swgtread.org/images/documents/standards/published/08-examination-final-03-2006.pdf> (accessed February 2013).

SWGTHREAD (2013). Scientific working group for shoeprint and tire tread evidence recommendations for research; <http://www.swgtread.org/research/recommendations-for-research> (accessed March 2013).

Wallace, T. and O. Mitchell (1980). Analysis of three-dimensional movement using Fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2(6), 583–588.

Wilson, H. (2012). Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes. *Journal of Forensic Identification* 62(3), 194–203.

Xiao, R. and P. Shi (2008). *Computational Forensics: Lecture Notes in Computer Science, Volume 5158*. Springer.