2015

Quantitative assessment of the discrimination potential of class and randomly acquired characteristics for crime scene quality shoeprints

Nicole Richetelli

Quantitative assessment of the discrimination potential of class and randomly acquired characteristics for crime scene quality shoeprints

Nicole Richetelli, B.A.

Thesis submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic & Investigative Science

Dr. Jacqueline Speir, Ph.D., Chair

Dr. Nicholas Petraco, Ph.D.

Dr. Gerald Lang, Ph.D.

Department of Forensic & Investigative Science

Morgantown, West Virginia
2015

Keywords: Forensic, footwear, quantitative, crime-scene-like, automated classification, randomly acquired characteristics, similarity

ABSTRACT

Quantitative assessment of the discrimination potential of class and randomly acquired characteristics for crime scene quality shoeprints

Nicole Richetelli, B.A.

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 the 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 was three-fold.

The first (and most time consuming obstacle) of this study was data acquisition. In order to efficiently process footwear exemplar inputs and extract meaningful data, including information about randomly acquired characteristics, a semi-automated image processing chain was developed. To date, 1,000 shoes have been fully processed, yielding a total of 57,426 RACs characterized in terms of position (θ, r, r_{norm}), shape (circle, line/curve, triangle, irregular) and complex perimeter (*e.g.,* Fourier descriptor). A plot of each feature versus position allowed for the creation of a heat map detailing coincidental RAC co-occurrence in position and shape. Results indicate that random chance association is as high as 1:756 for lines/curves and as low as 1:9,571 for triangular-shaped features. However, when a detailed analysis of the RAC's geometry is evaluated, each feature is distinguishable.

The second goal of this project was to ascertain the baseline performance of an automated footwear classification algorithm. A brief literature review reveals more than a dozen different approaches to automated shoeprint classification over the last decade. Unfortunately, despite the multitude of options and reports on algorithm inter-comparisons, few studies have assessed accuracy for crime-scene-like prints. To remedy this deficit, this research quantitatively assessed the baseline performance of a single metric, known as Phase Only Correlation (POC), on both high quality and crime-scene-like prints. The objective was to determine the baseline performance for high quality exemplars with high signal-to-noise ratios, and then determine the degree to which this performance declined as a function of variations in mixed media (blood and dust), transfer mechanisms (gel lifters), enhancement techniques (digital and chemical) and substrates (ceramic tiles, vinyl tiles,

and paper). The results indicate probabilities greater than 0.850 (and as high as 0.989) that known matches will exhibit stochastic dominance, and probabilities of 0.99 with high quality exemplars (Handiprints or outsole edge images).

The third and final aim of this research was to mathematically evaluate the frequency and similarity of RACs in high quality exemplars versus crime-scene-like impressions as a function of RAC shape, perimeter, and area. This was accomplished using wet-residue impressions (created in the laboratory, but generated in a manner intended to replicate crime-scene-like prints). These impressions were processed in the same manner as their high quality exemplar mates, allowing for the determination of RAC loss and correlation of the entire RAC map between crime scene and high quality images. Results show that the unpredictable nature of crime scene print deposition causes RAC loss that varies from 33-100% with an average loss of 85%, and that up to 10% of the crime scene impressions fully lacked any identifiable RACs. Despite the loss of features present in the crime-scene-like impressions, there was a 0.74 probability that the actual shoe's high quality RAC map would rank higher in an ordered list than a known non-match map when queried with the crime-scene-like print. Moreover, this was true despite the fact that 64% of the crime-scene-like impressions exhibit 10 or fewer RACs.

Acknowledgements

I would first like to acknowledge my chair, Dr. Jacqueline Speir. Thank you for all of your help and inspiration throughout the last two years. I have achieved more than I ever thought possible.

I would also like to thank the other members of my committee who took time to read my work and offer suggestions: Dr. Nicholas Petraco, Dr. Gerald Lang, and Dr. Patrick Buzzini.

Lastly, I could not have completed all of this work without the help of everyone else working in the lab! Thank you all: Mike Fagert, Adam Epler, Madonna Nobel, Renu Watalingam, Mike Hite, Carly Lasky, Anika Redden, Shreya Kamath, Josh Davis, Jonathan Thurston.

Portions of this project were supported by Award No. 2013-DN-BX-K043, awarded by the National Institute of Justice, Office of Justice Program, U.S. Department of Justice. The opinions, findings, conclusions and recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice.

This work is dedicated to my family, who have always supported me in my academic endeavors.

Contents

1. Introduction

Footwear impression evidence can provide invaluable information to forensic scientists in order to link a suspect to a crime scene or reconstruct the series of events leading up to a crime. Despite this utility, shoeprint evidence is often "undervalued by investigators, attorneys, and the courts due to their limited knowledge of it" (1). As evidence of this, the *Census of Publicly Funded Forensic Crime Laboratories* reported that only 11,000 footwear examinations were conducted out of a total of 4 million requests in all of 2009 (2). This amounts to less than 0.3% of all forensic work carried out by 397 forensic laboratories in the United States over the course of a single year. Although it is unreasonable to anticipate that shoeprints can be detected and recovered at each and every crime scene, it is still valid to assume that some kind of footwear evidence may be present, and that the current statistics indicate underutilization of shoeprint impression evidence within the forensic community. With this in mind, the remainder of this document is divided into five chapters.

The first chapter (this chapter) reviews the current state of the footwear field, including (i.) a discussion of the relevant features in shoeprint comparisons that are typically evaluated during a forensic examination, (ii.) the types of conclusions that are reached after such an exam, (iii.) the education and training requirements associated with examiner expertise, as well as (iv.) a summary of major research contributions to the field, before closing with (v.) a brief list of topical areas that could benefit from additional research (including the three objectives of this body of work).

This introduction is followed by three chapters that are intended as stand-alone, draft, publication-quality journal submissions designed to address specific objectives from the aforementioned research list. The first describes a semi-automatic image processing chain that was implemented to streamline data acquisition for over 1,000 pairs of shoes, thereby allowing the efficient prediction of random co-occurrence in frequency of outsole features (in terms of location, shape, and similarity). This is followed by an assessment of the baseline performance of Phase Only Correlation (POC) as an automated classification technique to classify both high quality and crime-scene-like impressions. Finally, the last draft-publication reports the reproducibility of accidentals in crime-scene-like quality prints, including the assessment of similarity metrics such as Modified Phase Only Correlation, Matched Filter, Modified Cosine Similarity, Euclidean Distance and Hausdorff Distance, as well as the ability of a RAC map to mate back to its known match within a database. These three publication drafts are followed by one final (fifth) chapter that briefly reviews ways in which this research can be further prioritized and expanded.

1.1 Analysis and Interpretation of Footwear Impression Evidence

The methods of analysis of footwear impression evidence are often compared to that of fingerprint impression evidence. Fingerprint examination includes an evaluation of first, second, and third level detail. First level detail describes the overall pattern design, such as arch, loop, or whorl. The first level detail of fingerprints is least individualizing, but allows for impressions to be grouped. The same three basic friction ridge patterns are exhibited throughout the population and although this detail cannot lead to an identification, it helps to narrow down the comparison group (3). The second level detail includes specific ridge path and the presence of minutiae (ridge endings, bifurcations, dots, etc.). In fingerprints, comparison of this second level detail can lead to a positive identification. The locations of these details and the frequencies of the occurence of such features in a given location help to define the "uniqueness" of a fingerprint; that is, no two people are believed to exhibit the same fingerprints (3). Lastly, the third level detail is encompassed in the pore structures and small details contained within the ridges. Throughout the comparison process, an examiner records any disagreement between the latent and the known prints (3).

Similar to fingerprints, shoeprints exhibit some characteristics that are present throughout the population and some characteristics that are believed to be random and individualizing. A shoe's class or manufacturing characteristics include the size, shape, style, and pattern design. By definition, a manufacturing characteristic will be shared by many other shoes, as compared with the first level characteristics of fingerprints (1). Individual or acquired characteristics of a shoe include cuts, scratches, tears, gum, shoe patches, and holes; these are comparable to the bifurcations, ridge endings, and dots of a fingerprint. Footwear examiners conduct a methodical analysis of these aforementioned features in order to reach a conclusion about the origin of a given shoe impression. First and foremost, the known shoes must be used to make exemplar prints for comparison. In addition, the original lift or cast of an unknown impression must be obtained, if possible; if this evidence is not available, high quality photographs of this evidence must be used for comparison. Typically, the first of these analyses is based on class characteristics. This generally starts with a comparison of the outsole design, given that this is the most obvious feature of a shoe. The design of the questioned impression and the exemplar impression must be in agreement in order to move forward with the comparison. Next, the physical shape and size of the shoe and its design elements must be compared. This analysis includes both measuring the physical dimensions/perimeter of the outsole, if possible, as well as examining the size of the individual design features (1). The size of the design should be examined because as a sole changes size, something must change on the outsole, and this can happen in one of two ways (1):

1. The design element size will not change, but instead the number of design elements will change

2. The design element size will vary, but the number of design elements will remain constant

Should the known and the questioned impressions still correspond after the comparison of physical and design size, an analysis of randomly acquired characteristics (RACs) may ensue. The term randomly acquired refers to features that are "not planned or intentionally manufactured, and that their combined position, orientation, size, and features are unlikely to re-occur" (1). Following all of the aforementioned comparisons, the examiner can then come to a conclusion about the origin of the questioned shoeprint impression. SWGTREAD has published a standard for all forensic shoeprint examiners to reference when making conclusions about comparisons between shoeprints. These conclusions and the reasoning for each conclusion are as follows (4):

1. **Lacks sufficient detail**
    - No comparison was conducted.
    - A comparison was conducted but the impression did not have enough detail for a meaningful comparison.
2. **Exclusion**
    - Enough detail and differences were present to conclude that the known exemplar was not the source of the impression.
3. **Limited association of class characteristics**
    - Some similar class characteristics were present, but there were factors which limited the ability to make a stronger association.
4. **Association of class characteristics**
    - Class characteristics of physical design and size are consistent between the known and the questioned impression. Some correspondence of wear may also be present.
5. **High degree of association**
    - Correspondence of general physical design, size, and wear, as well as one or more acquired characteristics between the known and the questioned impression.

6. **Identification**

    - The known and the questioned impression correspond in class and acquired characteristics. It is the opinion of the examiner that the known footwear was the source of the questioned impression.

After conducting all relevant analyses, a footwear examiner can reach one of these conclusions, which are based on the guidelines in the SWGTREAD standards. Currently in the field, footwear evidence comparisons and the resulting conclusions are largely based on an examiner's experience, leading to some criticism. However, an analyst's experience is acquired from extensive education, training, casework and a wealth of acquired knowledge as illustrated in section 1.2.

1.2 Acquisition of Experience for the Forensic Footwear Examiner

Becoming a forensic footwear examiner requires a combination of education and training. Requirements include (5):

1. Bachelor's Degree in a physical or natural science, or
2. Associate's Degree plus 2 years experience, or
3. High School Diploma plus 4 years experience.

In addition to the above, comprehensive training is required to help candidates learn standards in terminology, evidence handling, as well as legal considerations. Further, a section of supervised casework is completed and an examination may be conducted to ensure that the body of knowledge has been mastered. As outlined by SWGTREAD, the following topics are included as part of the training program for forensic footwear examiners (5):

1. **Introduction to forensic footwear examination**

    - History; value of footwear evidence.

2. **Terminology**
3. **Evidence handling procedures**

    - Procedures and protocol; relationship of forensic footwear evidence to other forensic disciplines; collection and preservation; marking and documentation; chain of custody.

4. **Examination of impressions**

    - Protocols; theory of individualization; case organization; note taking; evidence evaluation and comparison; conclusions and findings; report writing.

5. **Laboratory instrumentation and equipment**

    - Procedures and protocol; photographic equipment; measuring devices; light sources; computers and peripherals; other relevant laboratory equipment.

6. **Photography**

    - Theory of photography; basic camera operation; general crime scene photography; examination quality photography; two- and three- dimensional impressions; various lighting techniques; filters.

7. **Recovery by lifting**

    - Electrostatic lifting; gelatin, adhesive, and other lifting methods.

8. **Recovery by casting**

    - Dental stone; fixatives and release agents; snow casting.

9. **Detection of impressions**

    - Visible impressions; specialized lighting; electrostatic lifting; physical and chemical methods.

10. **Enhancement**

    - Photographic, chemical, physical, imaging software.

11. **Manufacturing**

12. **Preparation of test impressions**

13. **Court testimony and legal issues**

    - Expert witness qualifications; legal decisions; preparations of exhibits; moot court.

1.3 Previous Research

While many forensic footwear examiners agree that the characteristics on a shoe are "unique" and can be used for identification, the field is often challenged due to lack of statistical evidence (6). Some footwear evidence has even faced admissibility challenges in the courtroom given the lack of quantitative support for the qualitative evidence, because the Daubert decision rejects the "general acceptance" rule (7). In an attempt to offer quantitative support to footwear examinations, a limited number of studies have assessed the discriminating potential of footwear, with implementation of mathematical methods, offering support to the assertion that footwear evidence can produce an identification of source.

Detailed below is a review of the current literature which attempts to incorporate statistics into footwear impression evidence and further support the claim that the features present on outsoles can lead to an identification. This claim is largely based on analysis of three major features associated with footwear, including:

1. Class characteristics: design, manufacturing process, etc.
2. Subclass characteristics: air bubbles, incomplete mixing, etc.
3. Randomly acquired characteristics: nicks, holes, tears, etc.

As aforementioned, class characteristics include manufacturing features, design, and size; however, shoes rarely exhibit only one class characteristic. In fact, the combination of class characteristics can greatly reduce the sample set for comparison, as illustrated in Fig. [1.1]. Though class characteristics cannot be used for an identification, the combination of these features can greatly aid in narrowing down the possible sources of a given impression by excluding shoes which do not correspond in manufacturing features (such as model, size, etc.).

Further, air bubbles, a subclass characteristic, also help to narrow down the possibilities of source of a given impression. For example, Champod et al. (2000) examined the presence of air bubbles in polyurethane shoe outsoles; the goal of the study was to "gather statistical data on the occurrence and measured features of air bubbles on shoesoles, in order to extend our knowledge of the stochastic behaviour of air bubbles" (8). The analysts examined seventy-one pairs of the same shoe and analyzed bubbles found only in the ball portion of the shoe. The results indicate that the configuration of air bubbles is highly variable, even between two shoes of the same pair.

Though class and subclass characteristics are often useful for discriminating purposes, these features cannot actually be used to reach an identification (1). Rather, an analysis of acquired characteristics is necessary in order to determine whether a given shoe was the source of a crime scene print. Wilson (2012) compared the outsoles of thirty-nine pairs of the same shoe, worn by the same person, to determine if the acquired characteristics varied enough from shoe to shoe to separate the pairs of shoes and make an identification

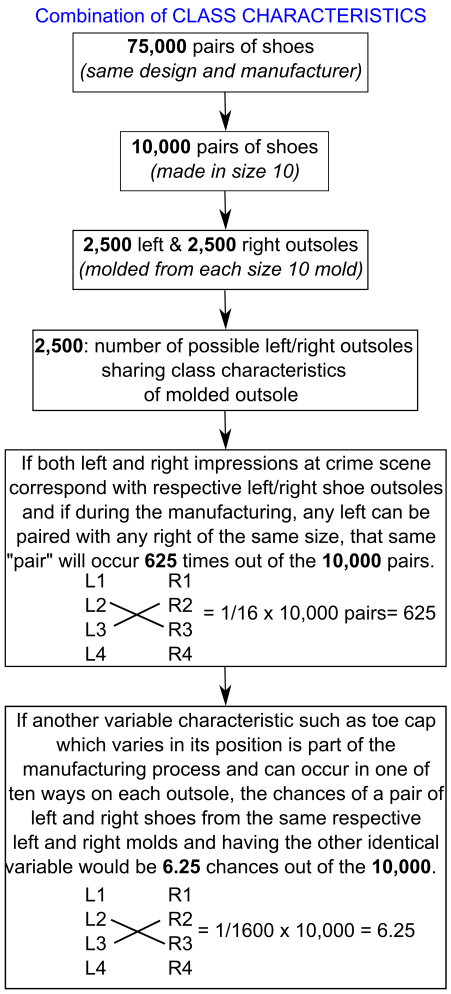


Figure 1.1: An illustration of how the presence of combined class characteristics can significantly decrease the number of suspect shoes (adapted from Bodziak (2000)).

(9). Each outsole was examined and acquired characteristics (such as cuts, gouges, and tears) were marked, counted, and recorded. Wilson (2012) found that even in shoes with a comparable number of acquired characteristics, visual examination could quickly differentiate between the location, size, and shape characteristics. In conclusion, the results indicate that "the likelihood of the characteristic(s) repeating in the same area of the same shoe with the same size and tread design is so small that it is the opinion of the experienced examiner that one would never observe the same amount of similarity between two different shoes" (9). While the results from this research indicate that the number, shape, and size of randomly acquired characteristics are not repeated, a large scale study

should be conducted in order to determine whether these results are repeatable.

While Wilson (2012) offered evidence that accidental characteristics are truly identifying, Cassidy (1995) provides some numerical estimates for the *probability* of repeated accidentals based on a dataset that consisted of boots worn by police recruits. For this study, because all of the shoes were worn for the same time span and over the same terrain, conditions favored the chance reproduction of accidental characteristics and wear (10). Two impressions from each of 97 shoes were recorded, for a total of 194 impressions. The shoes were broken into group A (59 shoes) and group B (38 shoes). From each impression, three accidentals were chosen and then compared against all other impressions to determine whether any of the accidentals were duplicated in the same position on another shoe. Results indicated that minute accidental characteristics were more prevelant on lightly worn shoes, while moderate or significant accidentals were much more likely to occur on more heavily worn shoes (10). For minute characteristics, the results indicated a 1 in 6 chance of encountering a duplicate accidental (10). However, the results for moderate characteristics indicated that these are less likely to be duplicated, likely due to increased size and/or complexity. For group A shoes, the chance of a coincidental similar accidental position was about 1 in 20, while this chance for group B shoes was even lower at approximately 1 in 38 (10). According to Cassidy (1995), the quality of accidentals greatly impacted the number required for an identification. More specifically, accidentals that are small or of poor quality require a larger number of features to reach an identification than larger or more rare characteristics (10). This study offers some numerical estimates about the probability of encountering accidentals in the same position on two different shoes. However, given the small sample size of this study, there remains a need for a large scale study concerning mathematical comparison of accidental characteristics.

While all of the above studies focus on empirical data, Stone (2006) utilized theoretical probabilities to describe the individuality of RACs. Stone (2006) identified five standardized individual characteristics. In addition, the researcher analyzed theoretical acquired characteristics based on their position, configuration, and orientation. To arrive at the computed probabilities, a hypothetical 16,000 square millimeter grid was superimposed on the theoretical shoeprint. The author then determined the hypothetical probability of encountering a given accidental on another shoe provided that the hypothetical shoe did not produce the print. For a point characteristic, the probability of a random duplication was modeled as 1 in 16,000. For a line characteristic, the length, orientation, and position were combined to obtain the probability of encountering a duplication of the given line characteristic valued at 1 in 384,000. For a curve characteristic the position, length, orientation, direction of curvature, degree of curvature, and apex location were all combined in the probability calculation to yield a 1 in 19,200,000 probability of finding a given curve characteristic in another shoe. As the characteristics became more complex, the probability of a random duplication greatly decreased. In addition, assuming that acquired characteristics occur at random and are independent from other characteristics, the probabilities of encountering several different characteristics could be multiplied together to obtain the probability of the random duplication of the entire collection. The results

reveal, "information about the sometimes incomprehensible magnitude of the 'uniqueness' of these types of characteristics when they occur in multiples or combinations", though the author explains that these probabilities are theoretical and that several validation studies are required (11). While this study indicates very robust results, the empirical data obtained by Cassidy (1995) indicates a lower discrimination potential, suggesting the need for further work to determine the true value of accidental characteristics.

Petraco et al. (2010) examined footwear impressions using facial recognition techniques, namely principal component analysis (PCA) (12). PCA assumes that variance provides information about a given dataset. Ideally, PCA serves as a data reduction technique which still captures most of the variability of the original dataset (13). For this study, the authors examined five pairs of the same type of shoe, which were each worn by the same individual. The Abbott Grid Locator was utilized in order to record the position of any accidentals on the shoe outsole, ignoring size and shape. After recording accidentals and completing PCA on the dataset, the authors used Maximum likelihood Gaussian-linear classification analysis (MLG-LCA) in order to determine the similarity between patterns based on distance in principal component space (13). This metric is essentially a Mahalanobis distance using the Z (principal component derived matrix) and the pooled covariance matrix. Mahalanobis distance determines how many standard deviations away a point is from the mean of a distribution. The results indicated that the average correct identification rate of the five pairs of shoes was approximately 92%. These results indicate that shoe prints, even when the shoes are worn by the same person and exhibit the same manufacturing characteristics, are statistically separable and the acquired characteristics provide enough information to potentially be used for identification. This study provides one method of statistically analyzing footwear impression evidence and even calculates an error rate as is required by the Daubert decision.

Furthermore, Sheets et al. (2013) aimed to determine the persistence of acquired characteristics over time, similar to Petraco et al. (2010). The goal of the study was to determine the rate at which wear affects the persistence of randomly acquired characteristics. For this study, eleven pairs of the same shoe were analyzed; a set of "accidentals" was cut into the outsole of the shoe in the same location on each pair. Participants in the study wore the pair of shoes for a period of seven weeks and the outsoles were examined at four times throughout the period of wear. A square grid was used to record the size of the accidentals via percentage of grid occupied by the accidental. Therefore, only the size of the acquired characteristics was recorded and the location and shape were ignored during the analyses. PCA was utilized to determine the variation within repeated measures of the same shoes and between different shoes at each time interval. Throughout the study, intra-shoe variation was much lower than inter-shoe variation. Thus, even with additional wear, each shoe better matched itself than any other shoe to which it was compared (14).

Therefore, several studies support the theory of individuality of shoeprints and provide evidence, both empirical and theoretical, for the identification of source based upon an analysis of randomly acquired characteristics on shoe outsoles (10; 9; 12; 14).

1.4 Advancing the Field with Further Research Support

While research exists to support the merit of footwear evidence in narrowing down a suspect pool or even identifying the source of an impression, the current body of work focuses largely on (i.) high quality evidence samples, and generally those of (ii.) very limited sample size. Therefore, additional work is needed. Moreover, very little focus has been directed at the interpretation and discriminating power of degraded and variable *crime scene samples*. With these goals in mind, the forensic footwear community could greatly benefit from several additional research thrusts, not limited to but including:

1. Development and implementation of an automated image processing chain, allowing for several different inputs (*i.e.,* high quality exemplars, lifts, photographs, and casts of crime scene impressions), which could facilitate efficient extraction and characterization of features to be used for comparisons and assessment of similarity between two prints or specific elements on several impressions;

2. Large scale study of RACs to include an assessment of postion and shapes of accidentals as well as the potential for co-occurrence;

3. A detailed evaluation of a single automated classification method. This work should include an assessment of the discrimination potential for crime scene quality impressions as well as an analysis of method limitations;

4. Analysis of the reproducibility of RACs in crime scene quality impressions as well as the potential for the features observed in evidence samples to be linked to source footwear;

5. Evaluation of the variability in examiner conclusions, including estimations of "error" rates for footwear comparisons (thus satistfying the Daubert standard);

6. Development of a national footwear database, comparable to AFIS for fingerprints;

7. Assessment of the frequency of different outsole styles and sizes;

8. Quantification of intra- and inter-analyst variability in RAC marking;

9. Investigation of the degree of uncertainty in footwear comparisons and conclusions;

10. Development of software which can automatically extract RACs for comparison from a variety of inputs, thus eliminating the need for manual identification of accidentals.

In an attempt to address some of the aforementioned research needs for footwear impression evidence, the current study executed a three-pronged research design. First, an image processing chain, with both automated and user-fed algorithms, was implemented allowing for digitization of varying impressions and extraction of feature information for

comparison between different inputs. Using this methodology, a collection of 1,000 shoes has been analyzed, including an assessment of RAC frequencies and potential for co-occurrence for 57,426 identified accidentals (thus addressing items one and two from the recommended research list). Subsequently, a random sample of 36 shoes was selected from this set to be used for creation of 108 blood and 72 dust impressions. These prints were utilized to conduct a quantitative assessment of the performance of Phase Only Correlation (POC), an automated classification method (recommended research item three) on high quality versus crime-scene-like images. This evaluation was based on a total of 1,525 high quality and 3,096 crime-scene-like print comparisons. Lastly, 200 crime-scene-like prints were examined to determine the fidelity of the impression transfer process; namely, the ability of the 6,762 randomly acquired characteristics observed in the exemplar impressions from 100 shoes to appear in the crime scene prints (recommended research item 4). In addition, an evaluation of similarity between 1,766 known match RAC mates (RACs identified in the exemplar impressions which were visible in the crime-scene-like prints) was conducted. The following three chapters are copies of draft, publication-quality journal submissions addressing said items from the recommended research list.

2. Technical Note

Technical note: Quantifying the frequency of shape and position of randomly acquired characteristics on outsoles

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 the 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 the type of results available for subsequent statistical analysis after processing a collection of more than 1,000 shoes.

Keywords: Footwear, Shoeprints, Randomly Acquired Characteristics, Accidentals, Fourier Descriptors, Feature Vectors, Semi-Automated

Introduction

Though footwear impression evidence can provide a wealth of information about a crime, including potential suspects, the total number of possible offenders, and the most probable series of events associated with a reconstruction, this evidence is often undervalued (or even overlooked) due to limited knowledge about how to collect, analyze, and interpret footwear impressions [1]. Part of the reason for this disconnect may be the difficulty associated with collecting sufficient-sized and community-shared databases for extensive research and study, which would allow the legal and forensic community to fully appreciate the value of this type of evidence. The fact is, footwear research is extremely time-consuming and labor intensive, regardless of whether the analyst is focused on class, randomly acquired characteristics (RACs), or both. Although class features hold incredible value, this project deliberately disregards class characteristics and instead focuses on RACs or accidental features such as nicks, tears, holes, and cuts that typically develop on outsoles as a function of wear. The reason for this narrow focus in scope is primarily four-fold. First, class features have received some research attention in the past [2–11] and this trend is likely to continue in the future. As a result, this investigative effort intentionally sought out the less-traveled parallel track concerning characterization of accidental features, while simultaneously collecting sufficient data to allow for subsequent class analysis downstream. Second, the National Academy of Sciences (NAS) 2009 report on *Strengthening Forensic Science in the United States* encouraged studies to shed light on the variability of randomly acquired characteristics, including

relative frequency of features, and the appropriate use of statistical standards [12]. Third, the Scientific Working Group for Shoeprint and Tire Tread Evidence (SWGTREAD) requested focused research in the area of "*Random Placement Shape and/or Placement of Randomly Acquired Characteristics*" [13], and finally, SWGTREAD also requested focused research in the area of "*Mathematical Probabilities of Randomly Acquired Characteristics*" [14]. Given these challenges, the first goal (and bottleneck) of this project was data acquisition. The remainder of this technical note describes the manner in which more than 1,000 worn shoes (obtained from a variety of sources including personal donations, corporate donations, and purchases from local thrift stores) were sequentially processed via a combination of automated and user-fed algorithms allowing for identified RACs to be extracted and characterized in terms of shape, geometry, and physical location.

Material and Methods

Available defining characteristics associated with more than 1,000 shoes have been recorded, including make, model, size, manufacturer product code, degree of wear, and the presence of either microcellular material or Schallamach patterns as detailed in Tables [1] - [6]. 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 [1] and 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.

In order to facilitate the automated downstream extraction of RAC shape and position, the outsole and exemplar were background subtracted and registered using identified control points. This process required the analyst to identify common geometric shapes (usually class characteristics) that were patent on both the outsole and the exemplar. The mating of these common points allowed for the automatic computation of variations in translation, rotation, and scale between the outsole and the exemplar. Once detected, the outsole and exemplar were co-registered or adjusted to ensure that they occupied the same location in image space (centered and oriented such that the long-axis of the shoe (toe-to-heel) was North-South within the image frame). In addition to this co-registration, the background (non-tread areas) of both the outsole and exemplar were removed (Fig. [2]) 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).

Following registration and background subtraction, randomly acquired characteristics present on both the outsole and exemplar 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. When this registered and marked image was subtracted from its registered (but unmarked) counterpart, the result was a RAC map that high-

lighted the location and geometry associated with each randomly acquired feature (Figs. [3] & [4]). Using the standard image processing technique of connected components, the location of each RAC was sequentially characterized using three parameters; the radius (r) or distance (in pixels) between the shoe's center and the RAC's centroid, the angular (θ) position (in degrees) of the RAC's centroid using the shoe's center as the origin, and the normalized distance (r_{norm}) equal to r divided by the distance (in pixels) between the shoe's center and the perimeter of the shoe at angular position θ.

Following localization, each feature was automatically numbered (via its connected component value) and extracted from the total RAC map. The resulting subimages (Fig. [5]) were then evaluated to define RAC shape and geometry, based on a 5-dimensional RAC feature vector, before transformation into individual RAC Fourier descriptors (FD).

RAC Feature Vector

Each randomly acquired characteristic was attributed to one of four categories: lines/curves, circles, triangles, and irregular-shaped features. To determine this categorization, 5 attributes per RAC were required, including area, perimeter, linearity, circularity, and triangularity. The first 2 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 coordinates [15]. 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. [1] [16], where A is the area of the object, and P is the length of its perimeter:

$$\begin{aligned} R_c &= \frac{4\,\pi\,A}{P^2} \\ R_c &= \text{maximum of 1.0 for a perfect circle} \end{aligned} \tag{1}$$

The fifth and final metric was a triangularity value computed using central moments (Eq. [2]) that are invariant to translation, scale, and rotation.

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q \tag{2}$$

As per Rosin (2003) [17], the variable I_1 in Eq. [3] equals $\frac{1}{108}$ for any triangle that has been affine transformed into a perfect right-angled triangle:

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4} \tag{3}$$

As such, the triangularity measure can be normalized to vary between $0.0 - 1.0$ according to Eq. [4] [17]:

$$T = \begin{cases} 108\,I_1 & \text{if } \; I_1 \leq \frac{1}{108} \\ \frac{1}{108\,I_1} & \text{otherwise} \end{cases} \tag{4}$$

Categorization Parameters

The 5-dimensional feature vector (Fig. [6]) 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 4 groups; line/curve, circle, triangle, or irregular.

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. [7]).

Shape Descriptor

In addition to shape categorization, each RAC was treated as a closed planar figure yielding a Fourier description [18–20]. This description was generated by tracing the contour of the shape $(x(t), y(t))$ where $t = 0, \ldots N-1$ with $N = 350$ for this dataset) and assuming a complex plane $z(t) = x(t) + i\,y(t)$ (where $i = \sqrt{-1}$). The resulting one-dimensional complex sequence of numbers was then mapped to the frequency domain via the discrete Fourier transform [19] where R_m and θ_m are the magnitude and phase of the m^{th} coefficient, respectively [19]:

$$
\begin{aligned}
Z(m) &= \sum_{t=0}^{N-1} z(t)\, e^{(-i2\pi mt/N)} = R_m\, e^{(i\theta_m)} \\
m &= -N/2, \ldots, -1, 0, 1, \ldots, N/2-1
\end{aligned}
\tag{5}
$$

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 [19]:

$$
\begin{aligned}
&Z(0) = 0 && \Rightarrow \text{translation invariance} \\
&R_m = \tfrac{R_m}{R_1} && \Rightarrow \text{scale invariance} \\
&\theta_m = \theta_m - \tfrac{\theta_{-1}+\theta_1}{2} && \Rightarrow \text{rotation invariance} \\
&\theta_m = \theta_m + m\tfrac{\theta_{-1}-\theta_1}{2} && \Rightarrow \text{start point invariance}
\end{aligned}
\tag{6}
$$

To illustrate, consider Figs. [8] & [9]. Fig. [8] depicts a single RAC (A), along with four synthetic modifications (B-E showing changes in scale, rotation, and translation). The resulting normalized Fourier descriptors are plotted in Fig. [9]. The x- and y-axes are arbitrary dimensions since the images have been normalized, but note that all contours are normalized to the same configuration, save a single π radian ambiguity [21]. Unless otherwise noted, all subsequent uses of RAC Fourier descriptors make use of both translation and start point invariance modifications.

Results

Database Statistics

To date, more than 1,000 shoes have been pre-processed. The definining characteristics of the first 1,000 are detailed in Tables [1] - [6]. The majority

of shoes in this collection are athletic in nature (Table [1]), due to generous corporate donations and the availability of shoes for purchase from local thrift stores. Table [2] reports the degree of wear of each shoe, which is not quite balanced between light, moderate, and heavy. For this study, shoes with "light wear" are those that exhibit discernible texture throughout. Conversely, the label "moderate wear" describes shoes with a reasonable degree of wear, resulting in both lost texture and possible bald spots. Finally, the term "heavy wear" is reserved for shoes with a near complete loss of texture, many or large bald spots, and possible holes or areas where the outsole has completely worn through.

Table [3] shows that nearly 90% of the collection lacks microcellular material in outsole composition. This is fortuitous since the presence of microcellular material is likely to increase intra- and inter-analyst variability in identifying randomly acquired characteristics. Conversely, approximately three-quarters of the database show Schallamach patterns (Table [4]); this is likewise fortuitous. Although current RAC data does not include the quantification of these features, the discrimination potential of Schallamach patterns can be explored in future studies.

Table [5] reports shoe frequency as a function of manufacturer and/or brand. Results indicate that almost 30% of the shoes processed thus far are from Nike® while another 28% are comprised of a small number of shoes, but from numerous manufacturers. Finally, Table [6] breaks down the database according to size and intended market (men or women). The results here are not random, but selective in the sense that our group did not capture data for shoes with a physical outsole size greater than the maximum length of

a sheet of Handiprint currently available for purchase (or approximately 13 inches in total length).

The shoes in Table [1] generated a total of 57,426 RACs (average of 57, minimum of 1, and maximum of 410). The majority (45%) were categorized as lines/curves, with another 38% falling into the irregular category. Circles filled a distant third group, comprising only 11% of the database, with triangles completing the remaining 6% (Table [7]). This data has been transformed into an interactive web-based heat map that currently reports frequency data for a "normalized" shoe, based on 57,426 RACs extracted from 1,000 shoes in the database. The normalization step, although not ideal, is unfortunately a necessity since it is near impossible to collect a sufficient number of shoes of a given make, model, and size to allow for statistical data analysis. Instead, the semi-random selection of shoes was normalized to create a *single idealized shoe* so that all RACs could be compared as if they occupied the same image space (as per θ and r_{norm}). In short, 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 (*Note: we also have the capacity to report frequency values as absolute, physical or non-normalized values using θ and r upon request. 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.*)

A static version of the web-based heat map is illustrated in Fig. [10]. In the associated frequency table, the numerical values in the top row remain constant regardless of the user's interaction with the web-page, displaying data associated with total RAC count for the entire database (regardless of cell location). Conversely, the two remaining rows automatically update to display RAC count and frequency for individual cells (5mm x 5mm) when queried by the user (in this static version, data is provided for a single cell outlined in black). The heat map allows the analyst to visually and quantitatively evaluate the spatial density of randomly acquired characteristics according to location and category, in response to the National Academy of Sciences (NAS) 2009 request for relative frequency of features, as well as SWGTREAD's request for research on *Random Placement Shape and/or Placement of Randomly Acquired Characteristics*" [13], and the "*Mathematical Probabilities of Randomly Acquired Characteristics*" [14].

Despite this positive step (pun intended), the authors acknowledge that this database must be used with caution. The utility of the density information is its ability to shed light on the random and variable nature of RAC frequency and *possible co-occurrence*. However, the heat map data is not intended to be a quantitative collection of independent wear-related events that can be multiplied to provide a cumulative probability of occurrence for a constellation of RACs on a randomly selected outsole. Moreover, density and categorization does little to account for the clarity, quality, and complexity of a geometric feature, which is as much (if not more important) to the forensic footwear comparison than the simple assessment of presence or absence. As such, the examiner's responsibilities cannot be deduced to a simple table of

frequencies, and a great deal more is required to both interpret and understand how best to utilize the database this project is generating. Despite this caveat, now that the data exists and is accessible to the community, our new focus is how best to present it to maximize value, along with estimates of uncertainty in frequency, analyst-variability, and quantitative metrics of shape similarity.

Table 1: Frequency of shoe type.

Type	Number
Athletic	838
Dress Shoe	88
Boot	56
Sandal	18
Total	**1,000**

Table 2: Degree of wear. Shoes with light wear have discernible texture. Shoes with moderate wear may show some bald spots and lost texture. Shoes with heavy wear have a near complete loss of texture, many or large bald spots, and possible holes or areas where the outsole has worn away.

Wear	Number
Light	281
Moderate	456
Heavy	263
Total	**1,000**

Table 3: Presence of microcellular material on the outsole.

Microcellular Material	Number
Present	108
Absent	892
Total	**1,000**

Table 4: Presence of Schallamach pattern on the outsole.

Schallamach Pattern	Number
Present	743
Absent	257
Total	**1,000**

Figure 1: Example of outsole and Handiprint exemplar scans.

Table 5: Frequency of manufacturer/brand.

Manufacturer/Brand	Number
Adidas	28
Asics	30
Brooks	10
Converse	30
Hoka	36
New Balance	20
Nike	294
Puma	14
Reebok	160
Skechers	12
Under Armour	60
Unknown	26
Other (fewer than 10 shoes)	280
Total	**1,000**

Table 6: Frequency of men's and women's shoe sizes. Note: shoes of unknown size account for the remaining 106 shoes (approximately 10%) of the database. Please note that size includes the full and half size; for example, a size 6 includes size 6 and size 6.5.

Men's Size	Number	Women's Size	Number
Size 5	2	Size 4	4
Size 6	4	Size 5	2
Size 7	28	Size 6	10
Size 8	54	Size 7	56
Size 9	148	Size 8	70
Size 10	200	Size 9	46
Size 11	162	Size 10	22
Size 12	62	Size 11	8
Size 13	14	Size 12	2
Total	**674**	**Total**	**220**

Table 7: Frequency of RAC shape categories in 1,000 shoes.

Category	Number	Percentage
Lines/Curves	25,826	45%
Irregulars	22,092	38%
Circles	6,288	11%
Triangles	3,242	6%
Total	**57,426**	**100%**

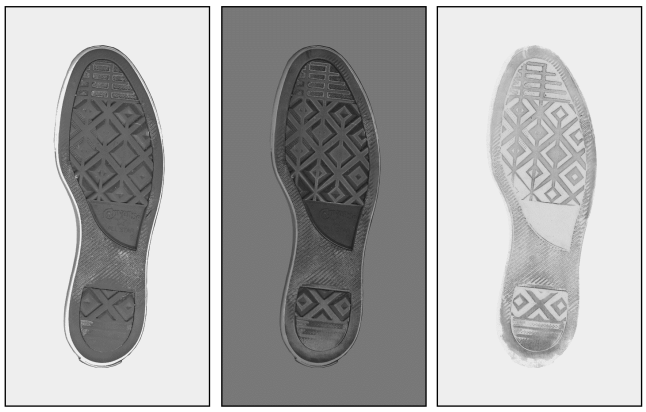


Figure 2: Registered and background subtracted outsole scan (left) and Handiprint scan (right). The middle image is an overlay of the outsole and Handiprint illustrating co-registration.

Figure 3: Registered and marked Handiprint image (left) and resulting RAC map (right).

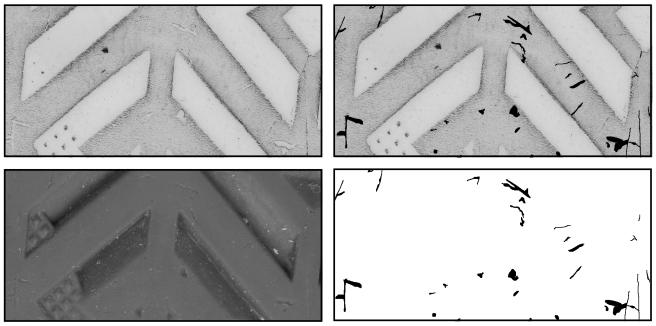


Figure 4: Example of a selected portion of the Converse Chuck Taylor® All Star®. Handiprint (top left), outsole (bottom left), marked Handiprint (top right), RAC map (bottom right). Note that the outsole image shown in this figure has been scanned on a flat bed scanner, but that all RACs were detected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

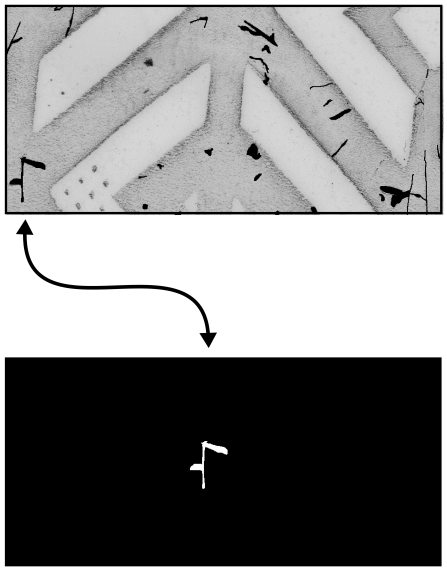


Figure 5: Subsection of RAC map and example of connected component subimages. This particular RAC was numbered #101, located at a normalized radius of 0.55 and an angle of 104^o

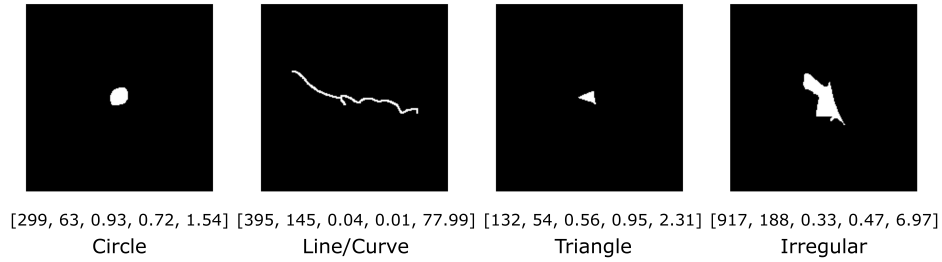


Figure 6: Four RAC images with their corresponding feature vectors [area, perimeter, circularity, triangularity, linearity].

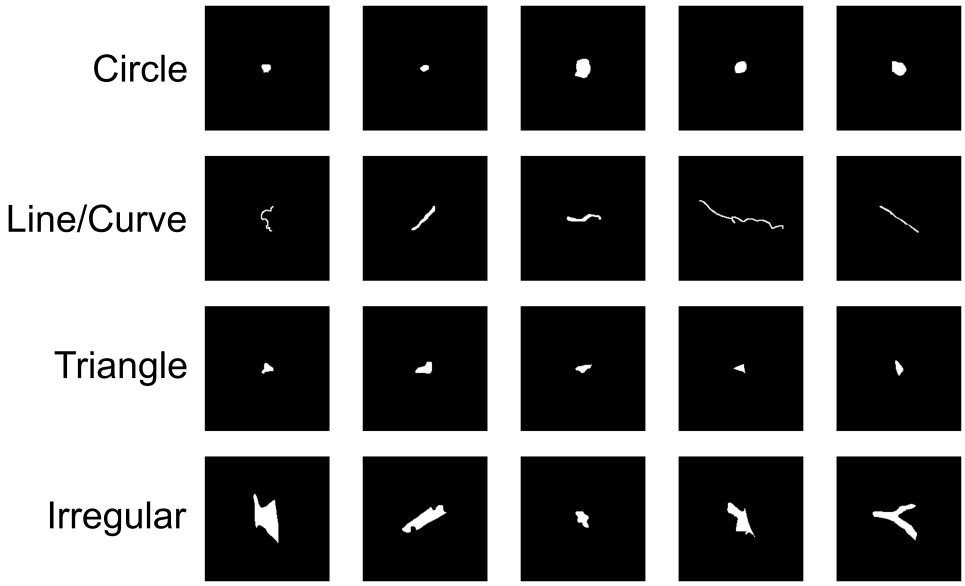


Figure 7: Examples of RACs classified as circles, lines/curves, triangles, and irregulars.

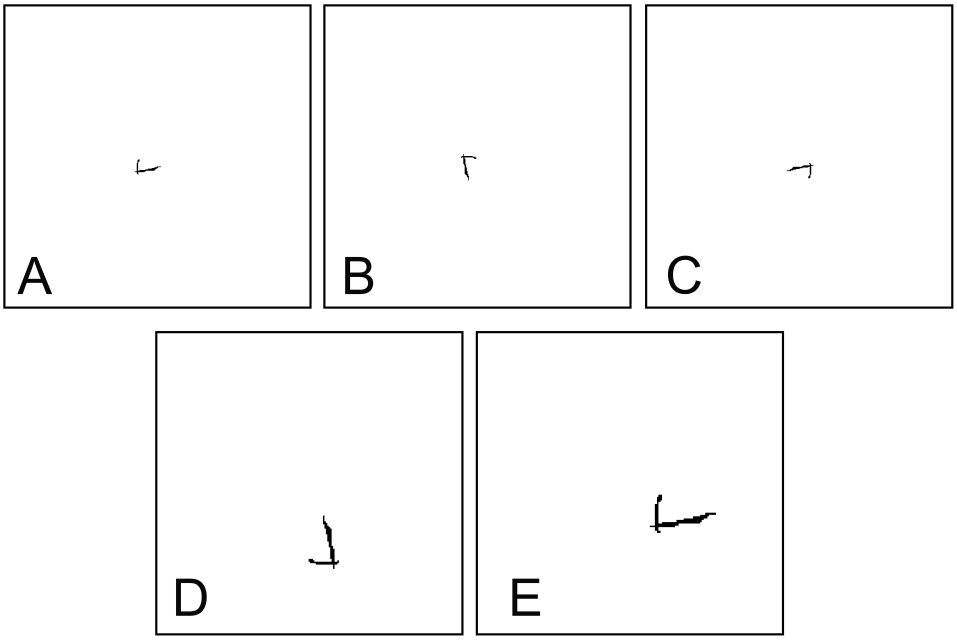


Figure 8: (A) Original RAC, (B) Rotated, (C) Rotated, (D) Rotated, Translated, and Scaled (E) Scaled and Translated.

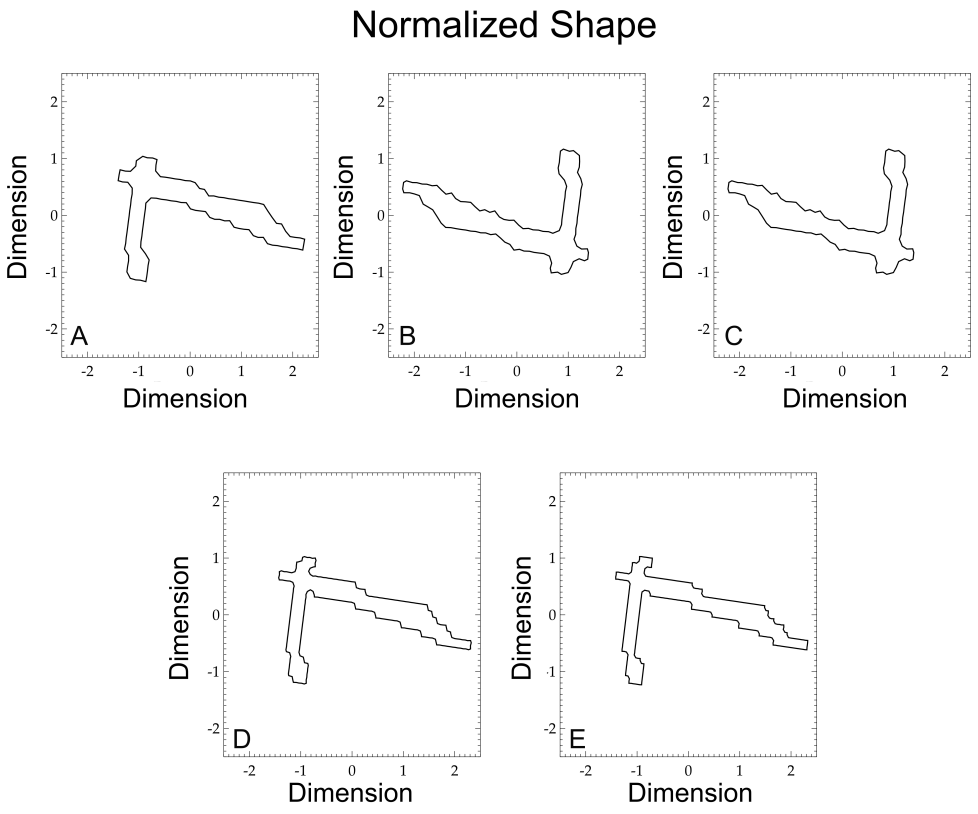


Figure 9: Plot of normalized Fourier shapes derived from the RACs shown in Fig. [8].

Description	Any Shape	Irregular	Circle	Triangle	Line/Curve
Total: In Database	57,426	22,075	6,287	3,242	25,822
Total: In Cell	86	35	12	0	39
Chance of Finding RAC in Cell	1:667	1:1,640	1:4,785	1:-,142	1:1,472

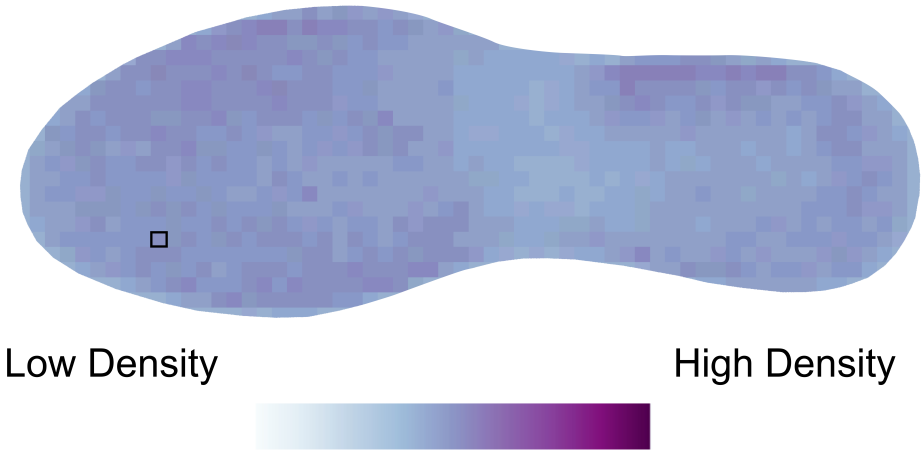


Figure 10: Static illustration of web-based heat map for a normalized shoe. Numerical values in the top row of the associated frequency table remain constant regardless of the user's interaction with the heat map, displaying data associated with total RAC count for the entire database (regardless of cell location). Conversely, the middle and bottom rows automatically update to display RAC count and frequency for individual cells (5mm x 5mm) when queried by the user. In this static example, the results are shown for a single cell outlined in black near the toe. Note that the normalized shoe was a size 10 men's Reebok® walking shoe with an area of 21,235mm^2.

References

[1] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery and Examination, Second Edition, CRC Press, 2000.

[2] A. Alexander, A. Bouridane, D. Crookes, Automated classification and recognition of shoeprints, Image Processing and its Applications 2(465) (1999) 638–641.

[3] Z. Geradts, J. Keijzer, The image-database REBEZO for shoeprints with developments on automatic classification of shoe outsole designs, Forensic Science International 82 (1996) 21–31.

[4] P. de Chazal, J. Flynn, R. B. Reilly, Automated processing of shoeprint images based on the Fourier transform for use in forensic science, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 341–50.

[5] T. Kiely, Forensic Evidence: Science and the Criminal Law, Second Edition, CRC Press, 2006.

[6] T. Hannigan, L. Fleury, R. Reilly, B. O'Mullane, P. de Chazal, Survey of 1276 shoeprint impressions and development of an automatic shoeprint pattern matching facility, Science & Justice 46 (2006) 79–89.

[7] M. Gueham, A. Bouridane, D. Crookes, Automatic recognition of partial shoeprints based on phase-only correlation, IEEE 4 (2007) 441–444.

[8] M. Gueham, A. Bouridane, D. Crookes, O. Nibouche, Automatic recog-

nition of shoeprints using Fourier-Mellin transform, NASA/ESA Conference on Adaptive Hardware and Systems (2008) 487–491.

[9] R. Xiao, P. Shi, Computational Forensics: Lecture Notes in Computer Science, Springer, 2008.

[10] A. Bouridane, Imaging for Forensics and Security: From Theory to Practice, Springer, 2009.

[11] M. Jing, W. Ho, L. Chen, A novel method for shoeprints (sic) recognition and classification, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics (2009) 2846–2851.

[12] NAS, 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, 2009.

[13] SWGTREAD, Scientific Working Group for Shoeprint and Tire Tread Evidence - Reccomendations for Research: Random placement shape and/or placement of randomly acquired characteristics (http://www.swgtread.org/research/recommendations-for-research/8-footwear/50-random-placement-shape-and-or-placement-of-randomly-acquired-characteristics) (2015 (Accessed November)).

[14] SWGTREAD, Scientific Working Group for Shoeprint and Tire Tread Evidence - Reccomendations for Research: Mathematical probabilities of randomly acquired characteristics

(http://www.swgtread.org/research/recommendations-for-research/10-footwear-and-tires/42-mathematical-probabilities-of-randomly-acquired-characteristics) (2015 (Accessed November)).

[15] U. Park, A. K. Jain, Face matching and retrieval using soft biometrics, IEEE Transactions on Information Forensics and Security 5(3) (2010) 406–415.

[16] R. Gonzalez, R. E. Woods, Digital Image Processing, 3rd Edition, Pearson Prentice Hall, New Jersey, 2008.

[17] P. Rosin, Measuring shape: Ellipticity, rectangularity, and triangularity, Machine Vision and Applications 14 (2003) 172–184.

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

[19] I. Bartolini, P. Ciaccia, M. Patella, WARP: Accurate retrieval of shape using phase of Fourier descriptors and time warping distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(1) (2005) 142–147.

[20] C. Dalitz, C. Brandt, S. Goebbels, D. Kolanus, Fourier descriptors for broken shapes, EURASIP Journal on Advances in Signal Processing 161 (2013) 1–11.

[21] A. Folkers, H. Samet, Content-based image retrieval using Fourier descriptors on a logo database, Proceedings of the 16th International Conference on Pattern Recognition 3 (2002) 521–524.

3. POC

Classification of Footwear Outsole Patterns using Phase Only Correlation. Part I: Baseline Performance

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. To date, several different automated, semi-automated and user-driven classification models have been developed and discussed in the primary literature. Although each approach has demonstrated some level of success, most are multi-phased and susceptible to failure owing to one or more weaknesses in the image processing chain. Furthermore, there have been limited attempts to compare and quantify success when confronted with crime scene quality prints. The research presented here examines the performance of a single semi-automated shoeprint classification algorithm (based on Phase Only Correlation (POC)) for the classification of both high quality and crime-scene-like quality impressions. More specifically, the work is divided into two parts. Part I characterizes the baseline performance of POC and the loss in discrimination potential associated with this algorithm when presented with crime-scene-like prints that vary in terms of media (blood and dust), transfer mechanisms (gel lifters), enhancement techniques (digital and chemical) and variations in print substrate (ceramic tiles, vinyl tiles and paper). The results indicate probabilities greater than 0.850 (and as high as 0.989) that positive

samples (known matches) will order higher in a ranked list than negative samples (known non-match) when confronted with mixed media (blood and dust), transfer mechanisms (gel lifters), enhancement techniques (digital and LCV) and variations in print substrate (ceramic tiles, vinyl tiles, and paper). Based on this success, Part II has been initiated to further identify weaknesses. These results are forthcoming, wherein the authors intend to further characterize weaknesses in the image processing chain as a function of scale, translation, rotation, and partial print reproduction to help the footwear examiner better identify *a priori* how best to employ POC for use with crime scene marks.

Keywords: Footwear, Database, Classification, Phase Only Correlation, Crime Scene

---

Introduction

With the increased popularity of crime solving dramas on television, the public is much more aware of what crime scene investigators are looking for while processing a scene. This knowledge, whether accurate or not, has altered the jury's expectations during a criminal trial [1]. If this knowledge has affected the jury, it is equally likely to have altered how criminals attempt to conceal their crimes, putting greater importance on evidence types currently outside of the limelight of the media. These alternate forms of evidence, footwear included, provide information that is critical in linking suspects to victims, crime scenes, and even multiple crime scenes to each other. In fact, when the crime scene is void of all other forms of impression

evidence, footwear may be the only probative information at the scene. If present, footwear class and accidental characteristics may afford the analyst the ability to focus a criminal investigation, link high volume crimes together, or otherwise provide information vital to the successful resolution of a case. Ideally, the strength of this linkage will be highly discriminating, which is often a function of several factors, including the presence of randomly acquired characteristics (RACs). Despite this desired result, linkage based on accidental characteristics is not always possible, and a common misconception is that impression evidence must lead to identification for it to be useful; on the contrary, class features, if present in sufficient quantity and quality, can be extremely valuable [2].

Presuming little debate over the expressed utility of footwear classification in forensic investigations, analysts are left with deciding how best to go about creating and searching a database of possible exemplars. Ideally, this classification process should be simple, efficient, and automated, thereby freeing specialized investigators to concentrate on more demanding tasks. Regardless of the mechanism employed to elicit possible matches, footwear classification faces many challenges, including low signal to noise ratios (SNR), manufacturing variations, limited or partial data, and variability in user-input.

Footwear Classification Challenges:

Analyst versus Automated Methods

Since shoeprints began being sorted and collected, there has been a use for reference sets. In 1937 the FBI started a small rubber heel file which grew into the current reference collection (comprised of thousands of photographs, catalogs, and digital impressions) [3]. A similar effort was undertaken by the

National Police Agency in Tokyo, Japan, also evolving into a computerized reference collection [3]. However, these are only a collection of images of shoes, not a way to compare an unknown shoe to a set of exemplars. Since these inaugural efforts, scientists have progressed from completely manual comparisons to more automated processes [4–12], but not without encountering challenges along the way, including the trade-off between efficiency and accuracy.

In order to best evaluate crime scene quality impressions, a classification method must preserve structural outsole information while simultaneously reducing or removing noise. For this pattern recognition problem, the human observer has proven to be exceptionally skillful in every regard. Conversely, the human programmer must work diligently to overcome recognition loss when faced with what might seem like trivial differences between two images with content that varies only in size, image registration, image type (*e.g.,* photograph, cast, etc.), or signal to noise ratio.

For example, the human analyst has little difficulty classifying two shoes that differ only in physical size, when they match in tread design. Conversely, size variation is a particularly difficult issue to tackle in a mathematical comparison, especially because the size difference can be introduced by the manufacturer through more than one modification. For instance, shoe size can be altered by either increasing or decreasing the size of individual tread elements present on the outsole (*e.g.,* molded soles). Alternatively, the design of individual tread elements can remain constant, but the totality of the overall pattern is successively truncated (*e.g.,* die-cut soles) during the production of smaller sized soles. These size variation changes all depend

on the sole production method and what the company's final vision is for the shoe line. Shoe soles that are stamped out of a large piece of sole material will have the same pattern across the entire surface but as the shoe size increases, the number of repeating elements (*e.g.,* rows, columns, triangles) will also increase [13], thus altering the final appearance of the sole. Hence, the severity of how shoe size alters the pattern depends heavily on the above aforementioned manufacturing method. While an examiner can easily identify two outsoles of the same class, even considering differences in size or location of tread elements, successful classification with an automated system under the same conditions must include a mathematical representation or normalization step for the aforementioned variations in order to avoid classification failure even under high signal to noise ratio conditions.

Another challenge arises when a reference shoe and an unknown are not in the same image space, specifically in the case of partial impressions. Except under extreme conditions (*e.g.,* a partial impression of less than a few tread elements) a footwear analyst can relatively easily account for missing information and visually assess region correspondence between two impressions. However, the same spatial realignment is not as easily or seamlessly achieved with a computer algorithm in an automated fashion, as evidenced by the multitude of image registration models currently accessible in the literature (see review papers by Zitova and Flusser [2003] [14] and Wyawahare et al. [2009] [15] for details).

Additionally, large variations in the appearance of footwear evidence can complicate outsole classification. As varied as the ways in which shoeprints can be deposited, are the methods by which they are collected and enhanced.

Shoeprint information at a scene can be preserved and documented using a number of methods, including photography, lifting, and casting of impressions. Examiners (regardless of their level of experience) are innately able to account for these variations in appearance in order to determine the type of shoe which made an impression (human observers co-mingle and synthesize the content of numerous media types on a daily basis (*e.g.,* print, video, cell phone, Internet)). However, this large variability in evidence collection and preservation method becomes an obstacle for an automated system, somewhat analogous to the traditional multi-sensor fusion problem (integration of multiple sensors that vary in terms of signal to noise ratio, temporal resolution, spatial resolution, spectral resolution, distortion, perspective, etc.). In terms of footwear classification, the fusion dilemma is the comparison of imagery that likewise differs in terms of user-input (exemplar method (*e.g.,* Handiprint, ink, Magna-brush method), collection preferences (*e.g.,* photograph, digital scan), resolution settings (PPI), dimensionality (2D or 3D), media (*e.g.,* blood, dust, mud), substrate (*e.g.,* tile, vinyl, carpet, wood-flooring), enhancement mechanisms (*e.g.,* physical, chemical, digital), etc.).

In short, examiners have exceptional pattern recognition skills, even in the presence of overwhelmingly low signal to noise ratios that are often encountered when presented with low quality crime scene evidence. Conversely, automated systems require intentional and robust mathematical solutions. So why all the effort to accomplish something already elegantly solved by the evolution of the human observer? Efficiency; for all but the most commonly encountered shoes and questioned impressions with the lowest SNR, manual classification methods are inefficient and impractical in today's rapid

forensic discipline. The footwear examiner has extremely specialized pattern recognition intellect which is better suited toward the skillful comparison tasks required post-classification. As such, the pursuit and accomplishment of successful automated classification frees the analyst, allowing her to devote time to other higher-level tasks that cannot be accomplished from the benefit of today's ever-expanding computing efficiency.

Models Used for Automated Classification

To date, a variety of different classification algorithms have been evaluated for use on footwear impression evidence, including identification of local interest points in tread elements and correlation of the entire outsole design [4–12, 16, 17]. While most of these attempts are highly successful when provided with high signal to noise input, the real test is how well the algorithm can tolerate degraded and variable imagery.

Several of the existing classification methods utilize the Fourier transform in one form or another. Geradts and Keijzer [1996] generated 'Fourier-features' to identify and compare the shape of outsole design elements [16]. Later attempts focused on a fully automated classification process that used the two-dimensional discrete Fourier transform (DFT), the power spectral density, and the two-dimensional correlation coefficient as the similarity metric, thus considering the entire outsole design rather than focusing on specific design features [5, 7]. Gueham et al. [2008] computed a fast-Fourier transform (FFT) on shoeprint images and, after filtering and log-polar mapping, computed the 2D correlation of the new Fourier magnitudes [18]. Conversely, some Fourier methods compute a correlation of the Fourier phase information to automatically classify outsole patterns [6, 17]. While the magnitude

of an image is important, the phase information obtained from the Fourier transform holds the contextual information necessary for image reconstruction, thus providing the ability to accurately analyze images of low quality, which is the most probable form of footwear evidence [19].

Moment invariants, a shape description method, are commonly used in the object recognition field because they can be invariant to rotation, translation, and scale differences between shapes [20]. For example, AlGarni and Hamiane [2008] created a feature vector for each shoe that contained seven Hu moment invariants [8]. Similarly, in order to account for the shape irregularity and complexity of shoe tread patterns, Xiao and Shi [2008] incorporated an orthogonal polynomials-based (Zernike moments) shape descriptor method [7].

Another commonly used classification metric utilizes distance metrics to assess the similarity of feature vectors compiled using information contained within a shoeprint. Patil and Kulkarni [2009] used a Gabor transform and Euclidean distance to compare shoeprint images. For this method, images were convolved with a Gabor filter and a feature vector was constructed for each shoe. These feature vectors were compared using Euclidean distance, thus obtaining a similarity score for comparison between images [9]. Beyond simple distance metrics, statistical values such as the Mahalanobis distance can be utilized to asses the similarity of outsole textured regions as was done by Dardi et al. [2009] [10].

Another area of study regarding automated classification is local interest points. For these methods, interest points or features are extracted using a detector and these points are then mathematically compared using different

similarity metrics (*e.g.,* k-nearest neighbors, cosine similarity, etc.) [11, 12].

Regardless of the method employed, efficient and automated shoeprint classification has been extensively explored and success has been achieved, *though largely on high quality or synthetically degraded impressions.* Unfortunately, few studies have addressed the performance of these algorithms on crime scene impressions. Of course, algorithm validation is a two-step process; if an algorithm fails when presented with high quality imagery, there is no reason to move on and try to obtain a more realistic indicator of performance in the presence of complicated inputs. However, once an algorithm shows some level of success when presented with laboratory synthetic samples, it becomes appropriate to identify its strengths and weaknesses and determine its utility in actual case usage.

To date there have been two attempts to evaluate the performance of different automated methods when presented with crime scene footwear evidence. In 2009, Cervelli et al. [21] sought to compare the performance of three metrics: power spectral density (PSD) [5], Modified Phase Only Correlation (MPOC) [6], and texture based Mahalanobis distance (MD) [10]. In order to account for a range in print quality, two different sets of shoe marks were compiled. The first set was comprised of high quality exemplars with synthetic additions of noise and blur, while the second set consisted of real crime scene marks. For the synthetic shoe marks, all algorithms performed well, with MPOC exhibiting the best matching capacity [21]. In almost all cases, the highest MPOC score corresponded to the correct known match for each query print. However, when real crime scene impressions were tested, these results quickly diminished, indicating (as expected) that

synthetic crime scene marks are not an accurate indicator of an algorithm's performance on case quality evidence. Though the correct match percentage dropped greatly for real crime scene marks, MPOC still out-performed the other models [21].

More recently, Luostarinen and Lehmussola [2014], evaluated the accuracy of seven different automatic classification algorithms including PSD, Fourier transform, Hu's moment invariants, Mahalanobis distance, Gabor transform, local interest points with RANSAC, and spectral correspondence of local interest points [22]. More specifically, these methods were tested using three different image datasets of differing quality impressions, including real crime scene marks. Furthermore, partial and rotated prints were examined using all algorithms [22]. Overall, the method employing local interest points and RANSAC performed the best. However, many of the algorithms proved inconsistent and inaccurate when confronted with "non-ideal" input (*e.g.,* crime scene quality impressions, rotations, etc.) [22]. Again, these results indicate that laboratory quality prints, although useful as a first-pass when comparing an algorithm's potential performance, do little to really allow the research analyst to truly understand the strengths and weaknesses of a given classification metric.

In the end, after comparison of a multitude of algorithms, the community is still left with much uncertainty as to how best to move forward. In truth, no single classification algorithm is likely to out-perform all others in every single scenario. Instead, each metric is multi-phased and susceptible to failure owing to one or more weaknesses in the image processing chain. Therefore, the goal of the current work is *not to prove that one algorithm out-*

competes all others, but to (i.) characterize the loss in discrimination potential associated with a successful classification algorithm when presented with crime-scene-like prints, and (ii.) to help identify weaknesses in the image processing chain. More specifically, the goal of this work is two-phased. First, identify an algorithm that shows some level of success and characterize its baseline performance. Phase Only Correlation was selected as the algorithm of choice following a detailed literature survey and the work conducted by Cervelli et al. [2009] (based on the results from Luostarinen and Lehmussola [2014] local interest points with RANSAC could have been another likely candidate for study). Second, assess the weakest link(s) in the image processing chain associated with the selected algorithm to offer solutions that may help strengthen and reinforce weaknesses, and when not possible, offer comment so that the examiner is aware *a priori* of the expected failing.

Material and Methods

Experimental Design

A total of sixty-five shoes were selected as high quality controls. When possible, available defining characteristics associated with each shoe were recorded, including make, model, size, degree of wear, and the presence of Schallamach patterns. As necessary, each shoe was gently washed to remove debris (*i.e.,* this research does not account for the possible presence of transient RACs such as rocks, gum, etc.). A subset of fifty outsoles were selected; these were scanned at 600PPI (3 replicates) with a Canon CanoScan8800F flatbed scanner, downsampled by 10, converted to binary, and transformed using a Canny Edge Detector [23]. The premise was to test the utility of us-

ing a memory-lean, low-resolution, binary edge image as the database proxy for all image types moving forward. For the remaining fifteen shoes, traditional Handiprint exemplars (3 replicates) [2] were created and scanned at 600PPI using an Epson Expression 11000XL Graphics Arts Scanner.

Using a random number generator, a total of thirty-six shoes were selected for crime scene print creation. Six analysts of differing heights, weights, and shoe sizes were selected to aid with crime scene print creation. Each analyst was randomly assigned 6 shoes (3 for dust and 3 for blood) as illustrated in Fig. [1]. Each shoe was used to create a crime scene print on different substrates, which included ceramic tiles, vinyl tiles, clear acetate sheets, and paper (dust only). Therefore, each analyst created a total of twenty-one crime scene prints (12 in dust and 9 in blood). The total number of prints created is detailed in Table [1]. All prints were co-registered using control points to minimize misclassification as a function of image registration (*Note: POC is a traditional image registration technique and can theoretically be modified to achieve both image registration and classification downstream*).

Crime-Scene-Like Print Creation

In order to best replicate crime scene conditions, analysts wore the shoes and walked over each substrate for the creation of all crime scene prints.

Dust Impressions

For creation of dust prints, analysts stepped in a tray of collected vacuum dust and walked over each substrate. Latent prints on tiles and acetates were then lifted using black gelatin lifters (13cm x 36cm BVDA Gellifters, Batch no. 2014198) and covered with the provided clear sheet; impressions on paper

were not lifted (Fig. [2]). Subsequently, the covered lifts and the latent impressions on paper were scanned at 600PPI using the Epson Expression Graphic Arts 11000XL and enhanced in Adobe®Photoshop®Elements 10 to increase contrast and minimize noise.

Blood Impressions

Certified pathogen-free human blood was utilized for creation of crime-scene-like blood impressions. A paper towel was saturated with blood and analysts stepped onto the paper towel (then over two newspapers in order to minimize blood pooling), before stepping onto the substrate. These impressions were allowed to dry and then scanned at 600PPI using the Epson Expression Graphic Arts 11000XL. After initial scanning, the impressions were enhanced using leuco-crystal violet (LCV), prepared as detailed in [2]. The enhanced impressions were again digitized after drying (Fig. [3]).

Post-Processing

Following crime scene print creation, enhancement, and digitization, all images were registered and background subtracted [**REF: Technical Note**]. In total, 66 blood and 106 dust prints were available for POC comparisons (8 dust prints and 2 blood prints were eliminated due to lack of a discernible tread pattern).

POC Comparison Metric

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. [1] & [2] where the subscripts refer to the two images under comparison [17].

$$G_1(u,v) = A(u,v)e^{j\sigma(u,v)} \tag{1}$$

$$G_2(u,v) = B(u,v)e^{j\theta(u,v)} \tag{2}$$

Once the Fourier transform of each input image has been calculated, the Phase Only Correlation can be computed according to Eq. [3] [5–7] where F^{-1} is the inverse Fourier transform and G_2^* is the complex conjugate of G_2 [17].

$$\begin{aligned} POC_{g_1 g_2} &= F^{-1}\left[\frac{G_1(u,v)G_2^*(u,v)}{\mid G_1(u,v)G_2^*(u,v) \mid}\right] \\ &= F^{-1}\left[e^{j[\sigma(u,v)-\theta(u,v)]}\right] \end{aligned} \tag{3}$$

Results & Discussion

In order to obtain a baseline for POC performance on outsole classification, replicate high quality exemplars for each method were compared (*i.e.,* outsole scans were compared to outsole scans and Handiprints were compared to Handiprints). Of the two, the Handiprint exemplar is the more traditional impression for comparison. However, the lower-resolution edge image was included with the hope that it would contain sufficient information to support a high degree of classification success for two major reasons. First, reduced computer storage needs, and second, almost every type of crime-scene print can be converted to an edge image (of some variety), potentially reducing variations in user-input.

For the known match (KM) comparisons, 3 replicates from each shoe were compared, yielding N=195 scores (65(3) = 195). For the known non-match (KNM) comparisons, a single replicate was used, yielding N=1,330 scores ($n(n-1)/2 = 50(49)/2 + 15(14)/2 = 1,330$). Fig. [4] illustrates the results from POC comparisons for high quality KM (solid line) versus KNM (dotted line) images. Of note is the distinct bimodal shape for the known non-match scores. The smaller, leftmost peak represents the POC scores of Handiprint KNMs while the larger, broader, and rightmost peak corresponds to the POC scores for the outsole edge images. Clearly the probability density indicates that the POC metric prefers the Handiprint exemplar as input imagery over the binary, downsampled edge image exemplars. Based on these results, the Handiprint exemplars were selected as the database image against which all 172 crime-scene-like query images (66 blood and 106 dust) were compared.

Fig. [5] illustrates the probability density of the log of match scores for KM and KNM comparisons of bloody crime-scene-like prints versus high quality Handiprint exemplars. Overall, the KM (N=106) scores for blood are lower than those for high quality impressions. As a result, there is more overlap between the KM and KNM densities (N=106(17)=1,802 scores) wherein ambiguous classifications can occur. Fig. [6] shows that the results from dust exhibit an even larger region of overlap (KM of N=66 scores and KNM=66(17)=1,122 scores), likely due to the decreased signal to noise ratio expected from dust impressions (and as depicted in Fig. [2]).

In order to compare the POC performance for the three datasets, a receiver operator characteristic (ROC) curve was constructed. Fig. [7] plots the false positive versus true positive rate for each comparison scenario. In

order to evaluate the performance of the classifier, the area under the curve (AUC) was computed. For a perfect classifier, the AUC=1, indicating perfect stochastic dominance. Based on the POC results, high quality impressions exhibit a 0.989 probability that a randomly sampled positive and negative pair will be correctly ordered in a ranked list. For bloody crime-scene-like prints, the AUC is slightly lower, with a correct rank probability of 0.974. However, this probability drops to approximately 0.895 for dusty impressions.

Conclusions

The results from this study remain consistent with previous research findings in that POC performs exceptionally well when classifying *high quality* footwear imagery [21]. In addition, the current work has attempted to generate a baseline level of performance for POC when presented with relatively good quality crime-scene-like prints. Results to date demonstrate reasonable levels of success (AUCs above 0.85) suggesting that the algorithm can handle a degree of variation in media (blood and dust), transfer mechanisms (gel lifters), enhancement techniques (digital and LCV), and substrate (ceramic tiles, vinyl tiles and paper).

Although these results are quite promising, the authors remain cautious, acknowledging that many vulnerabilities have yet to be tested. For example, as can be seen from Figs. [2] & [3], the "crime-scene-like" imagery generated for this comparisons is still of very respectable quality and, to adopt a term from analytical chemistry, not nearly able to stress the "the limit of detection" (LOD) for this algorithm (nor near the limit in quality observed by a footwear examiner on a routine basis). Moreover, the majority of the

test images were more than 80% complete, so future work must assess how well partial imagery can be classified. To this end, previous research suggests that image overlap (or partials) as low as 30% [24] can still be registered. This is extremely promising and suggests that a similar level of success may be achieved here, although it is important to note that shoeprint imagery is very different from the types of imagery traditionally registered using POC (medical and remotely sensed aerial imagery) so it is difficult to discern if this level of robustness will translate readily. Also of note is the need to move to a fully automated image registration mechanism (capable of handling variations in scale, rotation, and translation). Since POC is amenable to this implementation by conversion to log-polar space, further testing will move away from the use of ground control points for registration, testing image comparison with a significantly reduced need for user pre-processing [25]. These attributes (fully-automated registration, partial prints, and more extensive image degradation) are the topic of phase-two moving forward.

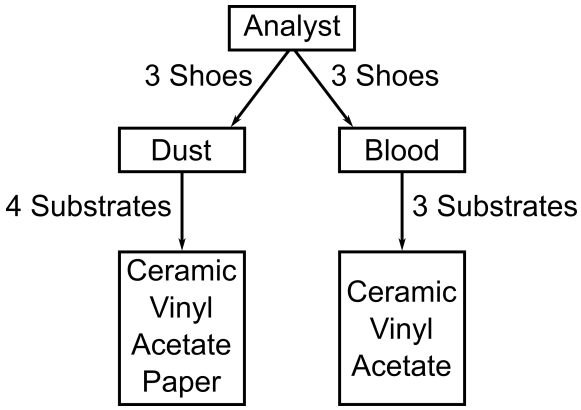


Figure 1: Work flow for crime scene print creation.

Table 1: Total number of crime scene prints created for dust, blood, and blood enhanced with leuco-crystal violet (LCV).

	Ceramic	Vinyl	Acetate	Paper	Total
Dust	n=18	n=18	n=18	n=18	72
Blood	n=18	n=18	n=18	n=0	54
Blood+LCV	n=18	n=18	n=18	n=0	54

364

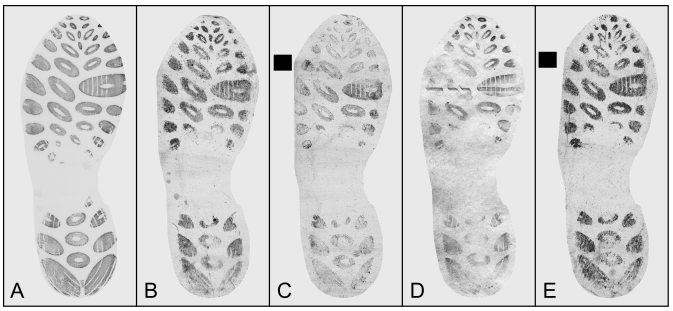


Figure 2: A) High quality exemplar; B) Digitally enhanced dust impression lifted from clear acetate; C) Digitally enhanced dust impression lifted from ceramic tile; D) Digitally enhanced dust impression from paper; E) Digitally enhanced dust impression lifted from vinyl tile. *Note: The sharp edge (denoted by a square) on the toe portion of shoe C) and E) is the demarcation of the gel lifter.*

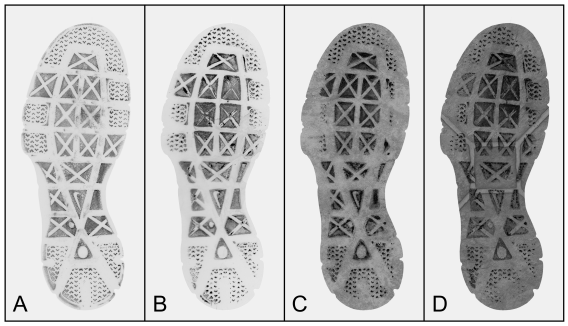


Figure 3: A) High quality exemplar; B) LCV enhanced blood impression on clear acetate; C) LCV enhanced blood impression on ceramic tile; D) LCV enhanced blood impression on vinyl tile.

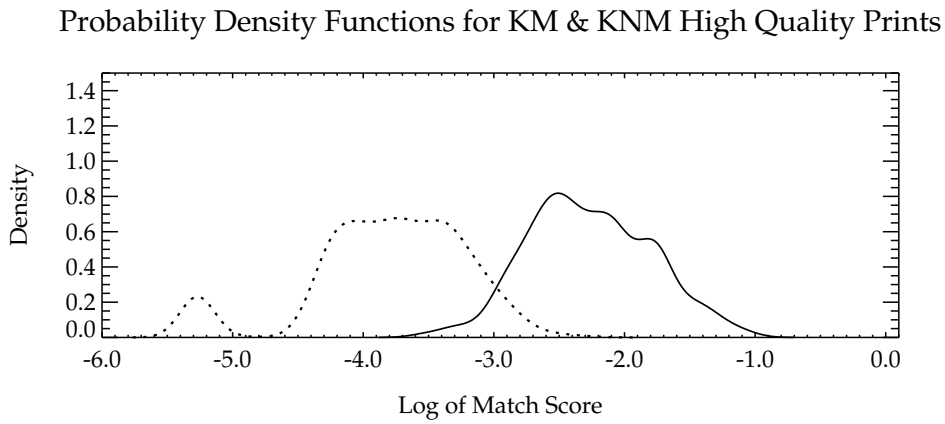


Figure 4: Probability density functions (PDFs) for KM (solid line) and KNM (dotted line) high quality prints. There were N=195 KM comparisons and N=1,330 KNM comparisons for the high quality exemplar dataset. PDFs were constructed using a Gaussian Kernel Density Estimator (KDE). The bin width for each density estimate was set equal to one quarter of the standard deviation of the log of the respective POC scores (or 0.118 for KMs and 0.120 for KNMs).

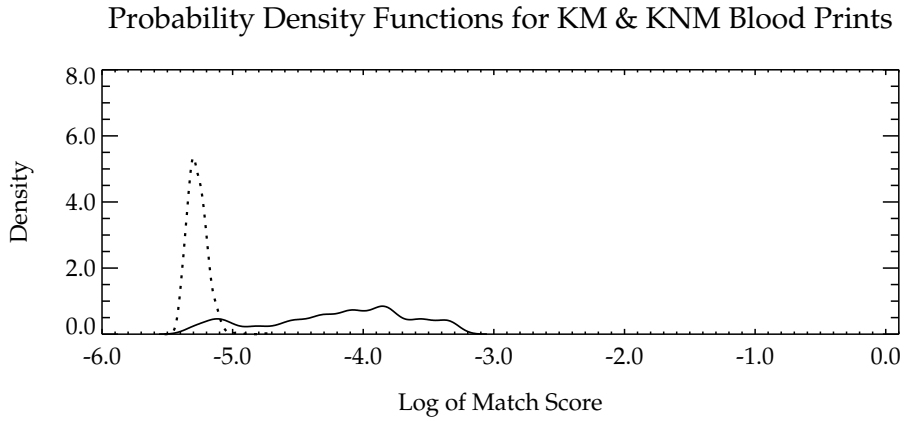


Figure 5: Probability density functions (PDFs) for KM (solid line) and KNM (dotted line) blood prints. There were N=106 KM comparisons and N=1,802 KNM comparisons for the blood print dataset. PDFs were constructed using a Gaussian Kernel Density Estimator (KDE). The bin width for each density estimate was set equal to one quarter of the standard deviation of the log of the respective POC scores (or 0.080 for KMs and 0.011 for KNMs).

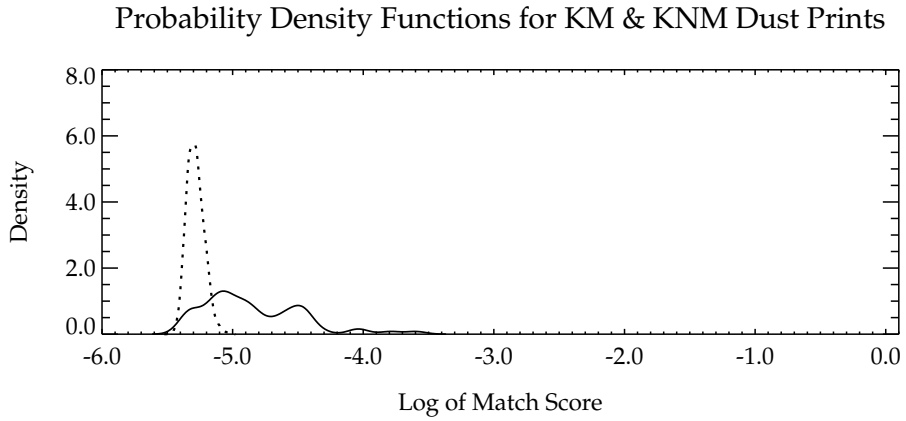


Figure 6: Probability density functions (PDFs) for KM (solid line) and KNM (dotted line) dust prints. There were N=66 KM comparisons and N=1,122 KNM comparisons for the dust print dataset. PDFs were constructed using a Gaussian Kernel Density Estimator (KDE). The bin width for each density estimate was set equal to one quarter of the standard deviation of the log of the respective POC scores (or 0.077 for KMs and 0.009 for KNMs).

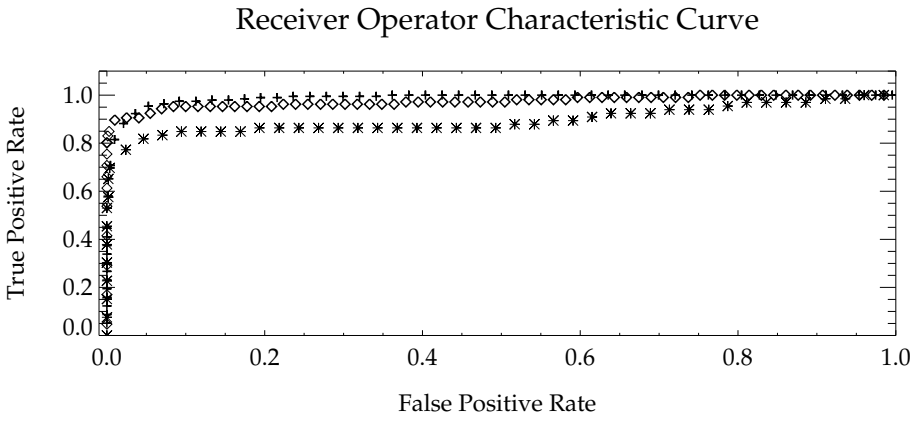


Figure 7: ROC curve for POC results. Crosses represent high quality impressions, diamonds represent the blood prints, and stars represent dust impressions. The area under the curve for high quality impressions is 0.989, for blood impressions is 0.974, and for dust impressions is 0.895.

[1] D. Shelton, The "CSI Effect": Does it really exist?, National Institute of Justice Journal 259 (2008) 1–7.

[2] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery and Examination, Second Edition, CRC Press, 2000.

[3] T. Kiely, Forensic Evidence: Science and the Criminal Law, Second Edition, CRC Press, 2006.

[4] A. Alexander, A. Bouridane, D. Crookes, Automatic classification and recognition of shoeprints, Image Processing and its Applications 2 (1999) 638–641.

[5] P. de Chazal, J. Flynn, R. B. Reilly, Automated processing of shoeprint images based on the Fourier transform for use in forensic science, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 341–50.

[6] M. Gueham, A. Bouridane, D. Crookes, Automatic recognition of partial shoeprints based on phase-only correlation, IEEE 4 (2007) 441–444.

[7] R. Xiao, P. Shi, Computational Forensics: Lecture Notes in Computer Science, Springer, 2008.

[8] G. AlGarni, M. Hamiane, A novel technique for automatic shoeprint image retrieval, Forensic Science International 181(1) (2008) 10–14.

[9] P. Patil, J. Kulkarni, Rotation and intensity invariant shoeprint matching using gabor transform with applicaiton to forensic science, Pattern Recognition 42 (2009) 1308–1317.

[10] F. Dardi, F. Cervelli, S. Carrato, An automatic footwear retrieval system for shoe marks from real crime scenes, Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (2009) 668–672.

[11] O. Nibouche, A. Bouridane, D. Crookes, M. Gueham, M. Laadjel, Rotation invariant matching of partial shoeprints, Proceedings of the 13th International Machine Vision and Image Processing Conference (2009) 94–98.

[12] M. Pavlou, N. Allinson, Automated encoding of footwear patterns for fast indexing, Image and Vision Computing 27(4) (2009) 402–409.

[13] W. Bodziak, Manufacturing processes for athletic shoe outsoles and their significance in the examination of footwear impression evidence, Journal of Forensic Sciences 31 (1986) 153–176.

[14] B. Zitova, J. Flusser, Image registration methods: A survey, Image and Vision Com 21 (2003) 977–1000.

[15] M. Wyawahare, P. Patil, H. Abhyankar, Image registration techniques: An overview, International Journal of Signal Processing, Image Processing and Pattern Recognition 2(3) (2009) 11–28.

[16] Z. Geradts, J. Keijzer, The image-database rebezo for shoeprints with developments on automatic classification of shoe outsole designs, Forensic Science International 82 (1996) 21–31.

[17] A. Bouridane, Imaging for Forensics and Security: From Theory to Practice., Springer, 2009.

[18] M. Gueham, A. Bouridane, D. Crookes, O. Nibouche, Automatic recognition of shoeprints using Fourier-Mellin transform, NASA/ESA Conference on Adaptive Hardware and Systems (2008) 487–491.

[19] A. Oppenheim, J. Lim, The importance of phase in signals, IEEE 69 (1981) 529–541.

[20] J. Flusser, Moment invariants in image analysis, International Journal of Computer, Electrical, Automation, Control and Information Engineering 1(11) (2007) 3708–3713.

[21] F. Cervelli, F. Dardi, S. Carrato, Comparison of footwear retrieval systems for synthetic and real shoe marks, Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis (2009) 684–689.

[22] T. Luostarinen, A. Lehmussola, Measuring the accuracy of automatic shoeprint recognition methods, Journal of Forensic Sciences 59(6) (2014) 1627–1634.

[23] R. Maini, H. Aggarwal, Study and comparison of various image edge detection techniques, International Journal of Image Processing 3(1) (2002) 1–11.

[24] A. Alba, R. Aguilar-Ponce, J. Vigueras-Gómez, E. Arce-Santana, Phase correlation based image alignment with subpixel accuracy, Advances in Artificial Intelligence - Lecture Notes in Computer Science 7629 (2013) 171–182.

[25] J. Sarvaiya, S. Patnaik, K. Kothari, Image registration using log polar transform and phase correlation to recover higher scale, Journal of Pattern Recognition Research 7 (2012) 90–105.

4. Wet Residue

Evaluation of shoeprint similarity via analysis of randomly acquired characteristics: A comparison of high quality exemplars and crime scene prints

Abstract

Forensic footwear evidence can prove invaluable to a criminal investigation by providing information about the nature of a crime or who may have committed it. However, limited knowledge about the discrimination potential of this evidence can lead to challenges in court. Though experienced forensic footwear examiners agree that these impressions can be just as discriminating as a fingerprint, general acceptance of this assertion can benefit from quantitative research. While there are several studies detailing classification of outsole patterns, these manufacturing characteristics cannot be used for an identification. Instead, randomly acquired characteristics (RACs) must be utilized for this purpose. Although empirical studies exist describing the discriminating power and frequency of these features, there have been limited attempts to characterize the utility of accidentals in crime scene quality prints. 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. This study mathematically compares the presence of RACs in high quality exemplars versus crime-scene-like quality impressions as a function of RAC shape, perimeter,

and area. Furthermore, the total RAC map, a binary representation of all RACs present in an impression, has been used to illustrate the bounds by which crime-scene-like laboratory samples can be linked back to their high quality exemplars. Results indicate that the unpredictable conditions associated with crime-scene print production promotes RAC loss that varies between 33%-100% with an average of 85%, and that this loss increases proportionally as a function of RAC perimeter and area. Furthermore, when the entire outsole is taken as a constellation of features, 64% of the crime-scene-like impressions exhibit 10 or fewer RACs. Despite this, there was a 0.74 probability that the match score for a randomly selected pair of positive (known match) and negative (known non-match) samples would be correctly ranked in an ordered list. Overall, the results indicate that footwear comparisons cannot be reduced to a "simple point counting" procedure; instead, more abstract qualities less amenable to quantitation (such as RAC shape, clarity, and complexity) are extremely important and unmistakably relevant in the comparison process.

Keywords: Footwear, Randomly Acquired Characteristics, Accidentals, Frequency, Shape Descriptors, Feature Vectors, Crime Scene

---

Introduction

Footwear impression evidence, which is left at almost every crime scene, can be invaluable for forensic scientists in order to link a suspect to a crime scene or reconstruct the series of events leading up to a crime. In order to maximize the utility of this evidence, it is therefore necessary to understand

how to properly evaluate and interpret footwear impression evidence. To this end, several studies have examined the individuality and utility of three major aspects of footwear impressions: class, subclass, and randomly acquired characteristics [1–9], the latter of which are the specific focus of this research. Of these three attributes, class characteristics (outsole design, size of outsole elements, and shoe dimensions) are the least discriminating. In combination, these features can greatly aid in narrowing down the possible sources of a given impression by excluding shoes of a given size, brand, model, etc., however, they cannot be used for identification of source [10]. When present, subclass characteristics (*e.g.,* air bubbles, stippling, remnants of inconsistent mixing of outsole material), which are a result of the manufacturing process that may vary for different shoes or molds, can provide additional means to reduce the set of possible sources for footwear impression evidence [11]. Through the use of class and subclass characteristics, an examiner may be able to eliminate an extremely large number of possible source shoes, greatly narrowing the number of reasonable leads and possible contributors in a criminal investigation. However, to determine a more precise likeness and actual source attribution between a questioned and known footwear exemplar, the examiner must proceed to compare the quantity, quality, clarity and complexity of what are termed accidental or randomly acquired characteristics (RACs) (*e.g.,* tears, nicks, stones, holes, etc.). If these features have reproduced in the crime scene print, and are in "sufficient agreement", the examiner is permitted to reach an actual identification of source. To reiterate and borrow a statement from the appellate court in the case of the State of Illinois vs. Charles A. Campbell [1991], "*shoe print evidence may*

be as reliable and as trustworthy as any other evidence...even one individual characteristic, depending on the nature and uniqueness, could be enough for a valued comparison" [10].

Since source attribution is a function of RACs, an active area of research is how best to demonstrate the degree to which information contained within shoeprints (specifically accidental features) are random and variable [6–9]. Much of this work has been affirmative of previous assertions (high discrimination) when conducted on high quality impressions or with theoretical data, however, there has not been a focused effort to determine how this discrimination might vary as a function of RAC size or shape under dynamic print production. As such, this study focused on quantifying RAC loss and variation during the production of crime-scene-like prints in order to better characterize the bounds by which case prints can be linked back to their high quality exemplars.

Sources of Variability in Footwear Impression Evidence

Numerous factors can affect the appearance of footwear impressions collected in criminal investigations. Consequently, examination and interpretation of this evidence is innately challenging and requires extensive training and accumulated expertise. More specifically, the entire process tends to be influenced by variations in print creation, collection, and enhancement, and in order for analysts to reasonably compare crime scene impressions to high quality exemplars obtained from suspect shoes, it is imperative that the sources of variability be understood and accounted.

Creation of Crime Scene Impressions

Despite the numerous methods of crime scene print creation, there are two major classes: two- and three-dimensional. Within each of these classes, however, exist a number of different factors which can contribute greatly to the variability present in the appearance of crime scene shoeprints.

Two-dimensional impressions include those which sit on top of a surface and have no discernible depth [12]. Positive impressions result from a transfer of material from the outsole to a substrate; examples include prints in blood, grease, and dust [10]. Conversely, a negative impression is left when an outsole lifts a residue from a surface. These often occur when a clean shoe comes into contact with a dirty surface and removes accumulated dirt or dust from the substrate. For negative impressions, the outsole elements are depicted in the void pattern. Clarity and quality of the impression often depends on the surface of deposition (*i.e.,* a waxed floor tile will likely capture a more detailed impression than carpet) as well as the media in which the print is made (*e.g.,* blood, grease, dust, etc.) [12].

Conversely, three-dimensional impressions result in deformation of the surface, resulting in an impression with depth. These prints can be found in soil, sand, and snow and the detection, preservation, and forensic utility of these impressions vary depending on a multitude of environmental conditions, including substrate composition [13, 14].

Collection and Enhancement of Impressions

Given the variability in the initial appearance of footwear impressions, the methods for collecting and enhancing this evidence can differ greatly depending on the conditions of deposition. For example, two-dimensional prints

are lifted to improve visibility and allow further examination, while three-dimensional impressions are often cast in order to preserve the entire depth of the impression [15]. Furthermore, the lifting method employed depends on the material deposited as well as the substrate containing the impression (*e.g.,* electrostatic lifters for dry impressions on non-porous surfaces, gelatin lifters for wet origin prints on non-porous surfaces [16–18]).

In addition to collection of crime scene prints for examination, enhancement methods may be employed to maximize visual detail. In general, impressions can be enhanced in four major ways: chemically, physically, digitally, or via electromagnetic radiation. In order to increase contrast between the impression and the background, chemical methods are carefully selected depending on the material in which the print is deposited, as well as the substrate properties. Extensive research exists detailing which methods are appropriate in a variety of scenarios [19–23]. Likewise, physical enhancement can be utilized to maximize contrast. This technique involves increasing contrast via the use of powders [10]. For example, by applying a fingerprint or fluorescent powder to an impression on a waxed surface, the evidence will retain the powder and can be easily distinguished from the background. Further, digital enhancement techniques can be used alone or in conjunction with another technique. These methods aim to use computer programs to increase image quality by maximizing the signal to noise ratio, thus increasing the amount of information available to the analyst for comparison purposes [24, 25]. Lastly, enhancement via electromagnetic radiation includes the use of specialized light sources (*e.g.,* ultraviolet, infrared, etc.) to maximize contrast of the impression against the background and therefore increase the

clarity and detail of evidence [26–29].

Given the inherent variability and complexity of footwear impression deposition, as well as the number of physical factors which can influence the appearance of prints (*e.g.,* media, substrate, enhancement methods, etc.), it is reasonable to expect variability in the appearance (clarity, quality, detail, etc.) of crime scene evidence. In short, a crime scene impression will rarely be an exact replicate of the source shoe or a corresponding high quality exemplar print. More specifically, the RACs which are visible in a high quality image are unlikely to consistently reproduce in crime scene evidence impressions. This is especially true given that RACs show variability in reproduction among high quality replicates even when prepared under ideal conditions in the laboratory! In fact, to account for this inherent variation, several replicate exemplars are typically created in the laboratory for both case and research purposes [5, 8], which further exemplifies the need to better understand RAC variation as a function of shape, perimeter, and area.

Methodology

Using a random number generator, 50 pairs of shoes were selected from a total of 400 to be used for crime scene print creation.

Pre-Processing

Available defining characteristics associated with each shoe were recorded, including make, model, size, manufacturer product code, degree of wear and the presence of Schallamach patterns. As necessary, each shoe was gently washed to remove easily dislodged 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 Graphics Arts Scanner. Post-outsole scanning, Handiprint exemplars were created [10], and each exemplar was likewise scanned at 600PPI.

Exemplar Processing

In order to facilitate the automated downstream extraction of RAC shape and position, the outsole and exemplar were background subtracted and registered using identified control points as detailed in [**REF: Technical Note**]. Post-registration and background subtraction, randomly acquired characteristics present on both the outsole and exemplar were marked and subsequently localized [**REF: Technical Note**]. Each feature was then automatically numbered and extracted from the total RAC map using connected components. The resulting subimages were then evaluated to define RAC shape and geometry based on a 5-dimensional classification feature vector, and then finally transformed into individual RAC Fourier descriptors (FD) [**REF: Technical Note**].

Crime-Scene-Like Print Creation

Five analysts of differing heights, weights and shoe sizes 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 the shoes when creating impressions (*note that this methodology differed from that used in exemplar creation which entailed pressing an outsole onto 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. [1] illustrates one "best case" and one "worst case" reproduction scenario.

Processing of Crime-Scene-Like Prints

After lifting, all impressions were scanned at 600PPI with an Epson Expression 11000XL Graphics Arts Scanner. The lifters were afixed to a scanning board designed to raise the gel surface off the scanner bed by approximately 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.

The digitized crime-scene-like print images were then background subtracted and registered to the corresponding high quality exemplars using identified and corresponding control points [**REF: Technical Note**]. This process ensured that all images (outsole scan, exemplar and both crime-scene-like impressions) were co-registered in the same image space.

Following registration and background subtraction, RACs on the crime-scene-like prints were marked and localized [**REF: Technical Note**]. Each individual RAC was again automatically numbered and saved into an individual file via connected components. Finally, feature vectors were created detailing RAC shape parameters and location information [**REF: Technical Note**].

Identification of Known Match RAC Pairs

In order to compare RACs, it was necessary to identify correspondences between accidentals on high quality exemplars and crime-scene-like prints. This was accomplished using RAC subimage location information. Features from the exemplars were automatically nominated as candidate matches if the angular (θ) and normalized radial value (r_{norm}) fell within $1 - 2^o$ and 0.1, respectively, of the corresponding θ and r_{norm} for the crime-scene-like RAC [**REF: Technical Note**]. These thresholds were selected in order to minimize loss of candidate RAC mates, which were subsequently manually verified (and adjusted as necessary) before moving forward. Fig. [2] illustrates a set of these known match pairs, as well as the corresponding location information for each accidental.

Similarity Metrics

Five metrics were used to analyze the similarity of known match crime scene to high quality RAC pairs. These metrics were Modified Phase Only Correlation (MPOC), matched filter (MF), a modified cosine similarity (MCS), Hausdorff distance (HD), and Euclidean distance (ED).

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. [1] & [2] (where the subscripts reference the images under comparison and $i = \sqrt{-1}$) [30].

$$G_1(u, v) = A(u, v)e^{i\sigma(u,v)} \tag{1}$$

196

$$G_2(u,v) = B(u,v)e^{i\theta(u,v)} \tag{2}$$

197 Once the Fourier transform of each input image has been calculated, the
198 Phase Only Correlation can be computed according to Eq. [3] [2, 3, 31]
199 where F^{-1} is the inverse Fourier transform and G_2^* is the complex conjugate
200 of G_2 [30].

$$\begin{aligned} POC_{g_1 g_2} &= F^{-1}\left[\frac{G_1(u,v)G_2^*(u,v)}{\mid G_1(u,v)G_2^*(u,v) \mid}\right] \\ &= F^{-1}\left[e^{i[\sigma(u,v)-\theta(u,v)]}\right] \end{aligned} \tag{3}$$

201 Eq. [3] can be modified by application of a frequency filter that selectively
202 limits frequencies used in the computation such that $F[g(x,y) \cdot h(k,l)] =$
203 $G(u,v)$. In this instance, each image $g(x,y)$ was modified by the windowing
204 function shown in Eq. [4] with $\alpha = 0.2$ and where $k = l = N$ which is the
205 size of the RAC image in pixels (1600 x 1600):

$$\begin{aligned} h(k) &= \alpha - (1-\alpha)\, cos\left[\frac{2\pi k}{N}\right] \\ k &= 0, 1, \ldots, N-1 \end{aligned} \tag{4}$$

206 *Fourier Descriptors (FD)*

207 With the exception of MPOC which was computed using 1600 x 1600
208 pixel imagery, all remaining similarity metrics were based on perimeter in-
209 formation. More specifically, the RAC was treated as a closed planar figure

yielding a Fourier description (FD) [32–34]. This description was generated by tracing the contour of the shape $(x(t), y(t))$ where $t = 0, \ldots N-1$ with $N = 350$) and assuming a complex plane $z(t) = x(t) + i\,y(t)$ (where $i = \sqrt{-1}$). The resulting one-dimensional complex sequence of numbers was then mapped to the frequency domain via the discrete Fourier transform [33] where R_m and θ_m are the magnitude and phase of the m^{th} coefficient, respectively [33]:

$$\begin{aligned} Z(m) &= \sum_{t=0}^{N-1} z(t)\, e^{(-i2\pi mt/N)} = R_m\, e^{(i\theta_m)} \\ m &= -N/2, \ldots, -1, 0, 1, \ldots, N/2-1 \end{aligned} \tag{5}$$

The coefficients were then transformed to ensure invariance to translation, and contour/sequence start point according to the following modifications [33]:

$$\begin{aligned} &Z(0) = 0 && \Rightarrow \text{translation invariance} \\ &\theta_m = \theta_m + m\tfrac{\theta_{-1}-\theta_1}{2} && \Rightarrow \text{start point invariance} \end{aligned} \tag{6}$$

Matched Filter (MF)

The matched filter similarity metric between two shapes $Z_1(m)$ and $Z_2(m)$ was computed as illustrated in Eq. [7] [35] where $Z(m)$ is normalized according to $\frac{Z(m)}{\sqrt{\sum_t |z(t)|^2}}$ such that 0.0 is the minimum (least similar) and 1.0 is the maximum (most similar):

$$m = \text{argmax} \left| \frac{1}{N} \sum_{t=0}^{N-1} Z_1(m)\ Z_2(m) e^{(i2\pi mt/N)} \right| \tag{7}$$

Modified Cosine Similarity (MCS)

Cosine similarity is a commonly used metric that can assess the similarity between two data vectors [36]. 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 are defined as FDs (or complex numbers $z = x + i\,y$), each complex vector was reduced to its magnitude $|z| = \sqrt{x^2 + y^2}$ before employing the traditional cosine computation shown in Eq. [8], where (T) represents the transpose of a vector.

$$\theta = cos^{-1}\left[\frac{a^T b}{\sqrt{a^T a}\sqrt{b^T b}}\right] \tag{8}$$

Euclidean Distance (ED)

Euclidean distance was the fourth metric employed for comparison. The distance (D) between elements in the complex vectors is obtained as detailed in Eq. [9], where x_1 and y_1 denote the real and imaginary parts of the first vector, respectively [36]. Likewise, x_2 and y_2 denote the real and imaginary parts of the second vector for comparison, respectively. The total distance is normalized by dividing the summation by the maximum number of elements in the vectors ($N = 350$ for this datset), yielding an average distance. Naturally, as elements become more dissimilar, the distance between them increases.

$$D = \frac{1}{N}\sqrt{\sum (x_1 - x_2)^2 + \sum (y_1 - y_2)^2} \tag{9}$$

Hausdorff Distance (HD)

Using the Euclidean distance, Hausdorff distance was likewise computed. This is more of 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 Manhattan distance, but this is something that can be remedied moving forward). In this computation, the distance ($d(a, b)$) is computed between a point (*e.g.,* a_1) on the perimeter of RAC (A) and all points on the perimeter of RAC (B) (Fig. [3]) using any desired distance metric (such as ED). Following all computations, the smallest distance from a_1 to B is retained. This process is then repeated for all points on A (*i.e.,* $a_2...a_n$), wherein $h(A, B)$, or the maximum of these minimums, is retained [37]. This same process is 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 is then the maximum of these two values ($h(A, B)$ and $h(B, A)$) as illustrated in Eq. [10].

$$H(A, B) = max\{h(A, B), h(B, A)\} \tag{10}$$
$$\text{where } h(A, B) = max_{a \in A}\{min_{b \in B}\{d(a, b)\}\}$$

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 similar metric*" or the degree to which the wet-residue images could be linked back to their source. This was accomplished using image-wide Phase Only Correlation according to Eq. [3] (without windowing), and on full RAC maps 8691 x 8691 pixels in dimension.

Results & Discussion

RAC Loss

Given the inherent inconsistency present in shoeprint creation, such as 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 [1]). 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 reproduction varied as a function of any of these factors. As detailed in Table [2], RAC loss (77% - 84%) exhibited very little variation across shape classes. However, greater variation can be observed as a function of RAC size. Generally, as a feature's size increased (in either total area or perimeter), the percent loss decreased (Tables [3] & [4]). This matched intuition in that "larger" defects are likely to persist and withstand the variation introduced during reproduction in a crime scene setting as compared to smaller features that may be more easily occluded by erratic conditions (such as differences in media, substrate, motion, etc.). Overall, Tables [1] - [4] suggest one major outcome. As with fingerprint comparisons, there is no scientific basis on which to demand a minimum number of features in order to judge source attribution in footwear comparisons. Moreover, the utility of an accidental feature should not be reduced to a simple counting exercise; its presence (and therefore its "*uniqueness*") should not be reduced to an independent

wear-related event that is multiplied to provide a cumulative probability of occurrence among a constellation of other RACs on a randomly selected outsole. By the same token *a RAC's absence is not a valid reason for an exclusion*, ergo, *absence of evidence is not evidence of absence.*

Individual RAC Similarity

Five metrics were utilized to determine similarity between crime-scene-like RACs and their high quality mates (MPOC, MF, MCS, HD, and ED). This study assessed the differences in scores as a function of RAC shape, perimeter, and area. The results were illustrated in two ways: binned bar plots and probability density functions (PDFs). The bar plots display trends in the data and allowed for comparison via Chi-square tests. However, the binning process inherently fragmented the data, so full probability density functions (PDFs) were also constructed using Gaussian kernel density estimators (KDEs) to allow for an unabridged view of the spread in scores for a given set of conditions.

Similarity as a Function of RAC Shape

Differences in similarity scores based on RAC shape were detected for MPOC and MCS as per the Chi-square test [38] with $\alpha = 0.05$. In other words, the similarity scores for different shapes were significantly different from those expected if the variables were independent.

For MPOC, circles exhibited higher similarity scores, while lines and curves exhibited lower similarity scores (Figs. [4] & [5]). This trend was significant for all binned scores and 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. [6]). Conversely, an elongated feature, when rotated, is likely to exhibit a drastic decrease in correlation between its known match. As such, a metric such as MPOC (with a theoretical maximum of 1.0) is likely to show higher dissimilarity for linear RAC features unless forced to be rotationally insensitive.

Dependence in similarity scores for RAC shape were also detected for MCS θ values. As illustrated in Figs. [7] & [8], circles have the greatest density in smaller angular bins, while lines and curves dominant in frequency for larger angular bins. This difference is again likely due to rotational variations. As a circular object rotates, its overall appearance and orientation (and therefore similarity with another circular object) is likely to remain relatively unchanged (Fig. 6). However, if a linear RAC is skewed, the orientation of the feature is likewise changed, again resulting in feature vectors with detectable differences.

The remaining similarity metrics (MF, ED and HD) were not found to depend on RAC shape. The dependence of MPOC and MCS (as well as the lack of dependence of MF, ED and HD on RAC shape) are equally significant results. For example, one might argue that circular features are less discriminating than linear features. The premise for this argument is that circular features have more degrees of freedom compared to linear features, and that this assertion is especially true if you begin to consider features in combination (*e.g.,* comparing two circles versus two lines with some fixed spatial relationship). However, it is also likely that not all metrics are equally sensitive to these differences, proving that *numerical* metrics of similarity,

although *objective* and *impartial*, are still *biased* estimators that require exploration, testing and understanding before deployment.

Similarity as a Function of RAC Size

Differences in similarity scores based on RAC size (perimeter and area) were detected for MPOC as per the Chi-square test [38] 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. For MPOC, small features exhibited high similarity scores, while large RACs exhibited lower similarity scores when compared to their known match mates (Figs. [9] & [10] and [11] & [12]). 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. [13]). 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. This serves to reinforce the ultimate need for an examiner's subjective interpretation during the comparison process. Although an automated metric can provide a baseline numerical assessment of a known match that can be very beneficial moving forward, this illustration shows that a "low" *objective* similarity score still requires expert interpretation. In this instance, a visual comparison and explanation by the examiner is likely to be much more illuminating to the jury and layperson, and arguably more defensible, than a mathematical clarification of the reason for a low score.

Another interpretation of this result is that the larger the RAC, the more discriminating its potential. In other words, a single RAC, of sufficient qual-

ity, complexity, and similarity to a source is so unusual that its significance warrants source attribution. Can you *quantify* the degree of belief in this assertion with any degree of certainty? Only if you were comfortable answering this question based on the probability of encountering this score in the tail of a density estimated using a Gaussian KDE based on 200 samples. In other words, this research can be of tremendous benefit to the footwear examiner to help support courtroom testimony and conclusion protocols, but not as an automated and numerical substitute.

In addition to the differences detected by MPOC as a function of size, ED also exhibited differences in similarity scores as a function of perimeter (Figs. [14] & [15]), but the opposite trend was noted. Namely, as RAC size (area and perimeter) decreased MPOC scores were more likely to increase (exhibiting greater similarity). Conversely, as RAC size (perimeter) decreased ED scores were more likely to increase (exhibiting greater dissimilarity) (*note that this dependence was only found to be significant for ED scores between 0.01-0.02 and greater than 0.05 for* $\alpha = 0.05$). A hypothesis for this observation is possible interpolation effects. This research was conducted using pixel images, rather than vector graphics. Consequently, each RAC exists on a grid with the smallest subunit being a square picture element $42\mu m$ in size. When this feature is digitally captured for analysis, the resulting imagery includes both *inherent* fluctuations in reproduction due to variations in deposition conditions, as well as *unavoidable* inter- and intra-analyst variation from the marking phase. The end result is "jitter" (or perhaps more aptly termed, *deviation from nominal*).

If this RAC is to then be compared, it is important to note that many

similarity metrics require pre-standardization such as feature vectors of a constant size. If so, interpolation may be needed (*e.g.,* N=350). When required, minute differences from previous steps (real or analyst-based) are likely to be accentuated as a function of RAC size as shown in Fig. [16]. In this illustration, although both marked RACs (small versus large) have only a single pixel difference between them, the variation in the interpolated perimeter images are markedly different. The large RAC, with a greater number of points across its perimeter is only minimally affected by interpolation. However, the smaller the feature, the greater the artifact observed on the perimeter, at least in theory. Since this assertion is easily tested, a small study is currently underway to test this hypothesis and better characterize the divergence in results for MPOC and ED as a function of perimeter.

RAC Map Correlation

Table [5] reports the total frequency of RACs in the binary maps (as well as the corresponding high quality impressions). These maps were obtained during the subtraction process and are a binary 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 as a receiver operator characteristic (ROC) curve displaying the true positive and false positive rate (Fig. [17]). The area under the curve (AUC) indicates the probability of a randomly selected known match RAC map exhibiting a higher similarity score than a known non-match map (stochastic dominance). Based on the POC metric, there was a 0.74 probability that the match score for a randomly selected pair of positive (known match) and negative (known non-match) samples would be correctly ranked in an or-

dered list. Given that 64% of the query crime-scene-like maps contained 10 or fewer RACs for comparison (Table [5]), these results indicate that even with a minimal amount of accidental transfer, matching RAC maps will rank higher than non-matching maps approximately 74% of the time. This result is staggering and lends very strong support to the claim that footwear evidence is extremely discriminating, especially given that an average of 85% of the identified randomly acquired characteristics failed to transfer to the questioned impressions. Again, although there is no scientific basis for a minimum number of required characteristics for source attribution, these results suggest a multitude of future studies combining POC, RAC number, and RAC description (complexity, size, category (line, circle), etc.) in order to better characterize this very interesting trade space.

Conclusions

The results from this study suggest that reproducibility of RACs, in number and appearance, can vary greatly when comparing high quality and crime-scene impressions. Given that approximately 85% of these accidentals became obscured in the deposition process, it is clear that a "simple point counting" procedure cannot be used when assessing source attribution.

Furthermore, the correlation of RAC maps, from crime-scene-like impressions to high quality exemplars, offers additional support that the information contained within accidentals (*i.e.,* shape, size, and complexity) supplies greater evidence for source attribution than a simple presence, or count, of features. This is especially true considering that 64% of the crime scene prints exhibited 10 or fewer features, but that 74% of the time they ex-

hibit stochastic dominance. Since POC mimics some of the low-level spatial processing conducted by an experienced footwear examiner during the comparison process, the results indicate that source attribution is possible even when presented with *very few accidentals* provided the existing RACs exhibit sufficient discrimination potential in terms of shape, size, and complexity.

Mathematically, the results of MPOC and MCS reported lower similarity scores for linear features as compared to circular features. This is believed to be a function of both metric's sensitivity to orientation differences. However, this sensitivity was not observed for all metrics (MF, HD, nor ED). If the concept that circular features have more degrees of freedom tends to mirror an examiner's intuition when attributing significance to RACs, then it is important to evaluate the sensitivity of numerical metrics of similarity to best understand their strengths and weaknesses. Interestingly, a strength of a similarity metric when comparing known non-matches may very well be a weaknesses for the metric when comparing known matches (pushing the threshold for exclusions too high and generating too many false negatives). In contrast, an examiner can dynamically adjust this threshold as necessary, whereas doing so in an objective sense is much more difficult.

Moreover, quantitative results can actually disagree. For example, the ED results presented here were opposite to those of MPOC in that small features appeared more dissimilar than large features as a function of RAC perimeter. Whether or not this is genuine, or a function of interpolation, must be further tested. However, the major point is that while an examiner can easily associate two features with differences in appearance due to deposition or marking variability, an algorithm is forced to make a decision based on

pixel differences, and this can become challenging depending on imaging constraints. The end result proves that numerical and objective metrics are in no way a panacea for subjective assessment, and much more research is required. Instead, the authors propose a complementary relationship between examiner expertise and quantitative metrics that in combination can best define the degree of belief in source attribution for footwear evidence. What this might look like and how it might best serve the forensic community has yet to be defined. Clearly it is a long-term goal, and instead, the focus of short-term goals will be to better understand the quantitative trade space.

Table 1: Quantifying RAC loss between high quality exemplars and replicate crime-scene-like impressions.

RACs	High Quality	Crime Scene Rep 1	Crime Scene Rep 2
Total Number	6,896	1,049	1,110
Number Lost	-	5,847	5,786
Percent Lost	-	85%	84%
Mean Number per Shoe ± 1 standard deviation	69 ± 72	10 ± 12	11 ± 12
Maximum Number	307	66	61
Minimum Number	2	0	0

Table 2: RAC loss between high quality exemplars and replicate crime-scene-like impressions as a function of RAC shape.

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

Table 3: RAC loss between high quality exemplars and replicate crime-scene-like impressions as a function of RAC perimeter.

Perimeter	Total HQ RACs	Lost HQ RACs	% Loss
1-2mm	2,936	2,623	89%
2-4mm	2,413	1,939	80%
4-6mm	828	599	72%
6-8mm	337	217	64%
>8mm	382	245	64%

Table 4: RAC loss between high quality exemplars and replicate crime-scene-like impressions as a function of RAC area.

Area	Total HQ RACs	Lost HQ RACs	% Loss
0.0-0.25mm^2	3,994	3,548	89%
0.25-0.5mm^2	1,408	1,080	78%
0.5-0.75mm^2	589	419	71%
0.75-1.0mm^2	294	201	68%
1.0-2.0mm^2	391	253	65%
>2.0mm^2	220	122	55%

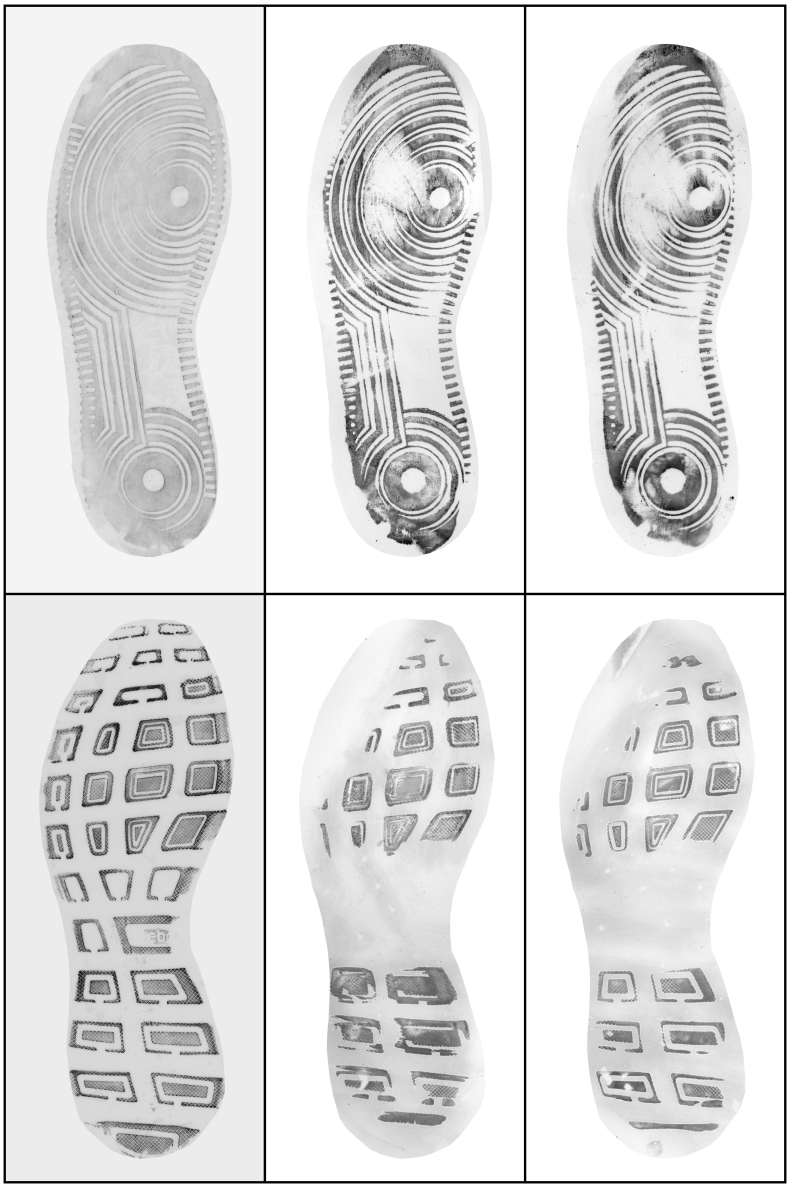
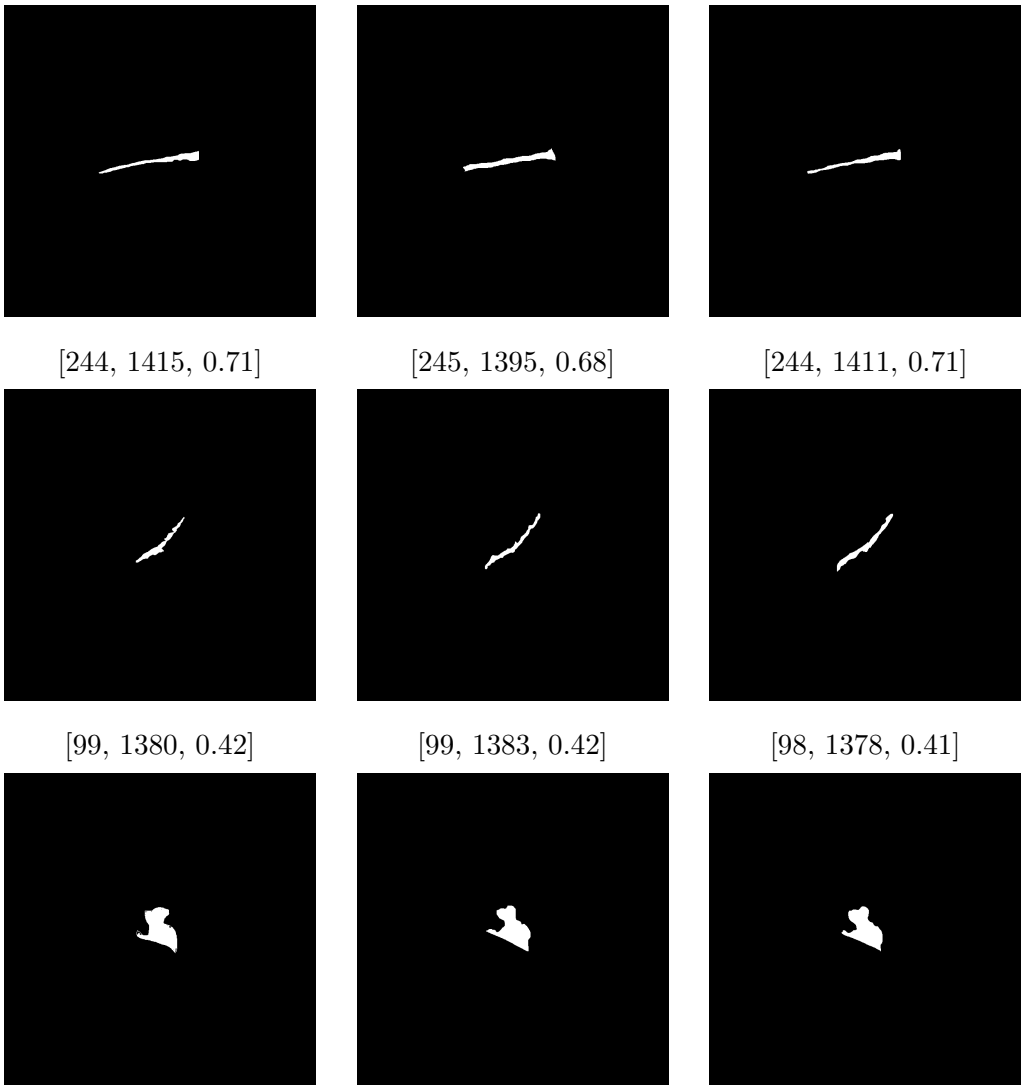


Figure 1: Top row illustrates one "best case" scenario and bottom row displays one "worst case" scenario for crime-scene-like impression production. Handiprint exemplar (left) and two crime-scene-like replicates (center, right).



[244, 1415, 0.71] [245, 1395, 0.68] [244, 1411, 0.71]

[99, 1380, 0.42] [99, 1383, 0.42] [98, 1378, 0.41]

[244, 1562, 0.68] [243, 1550, 0.70] [244, 1562, 0.68]

Figure 2: RAC image mates with their corresponding location information [θ (degree), r (pixel), r_{norm}]. High quality RAC image (right) with its detected crime scene RAC mates, one from each replicate (center, right).

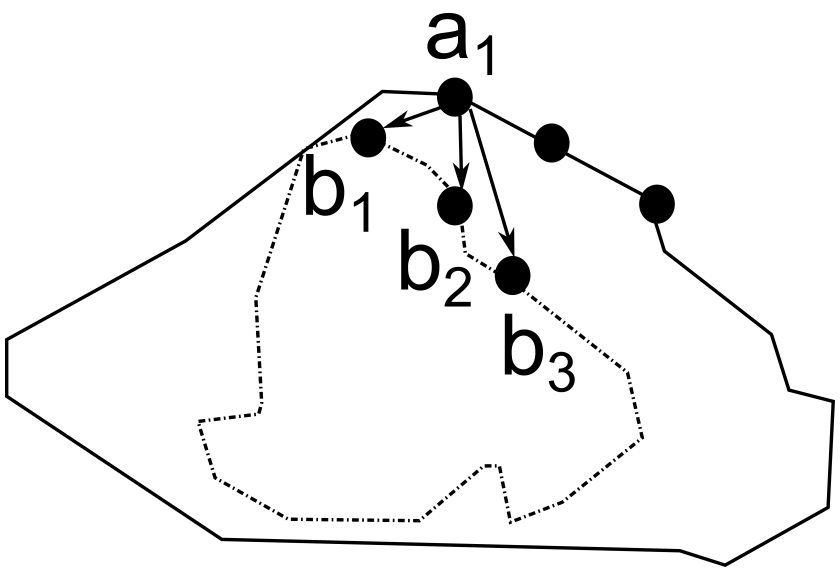


Figure 3: Two stylized RACs illustrating computation of Hausdorff distance for point a_1.

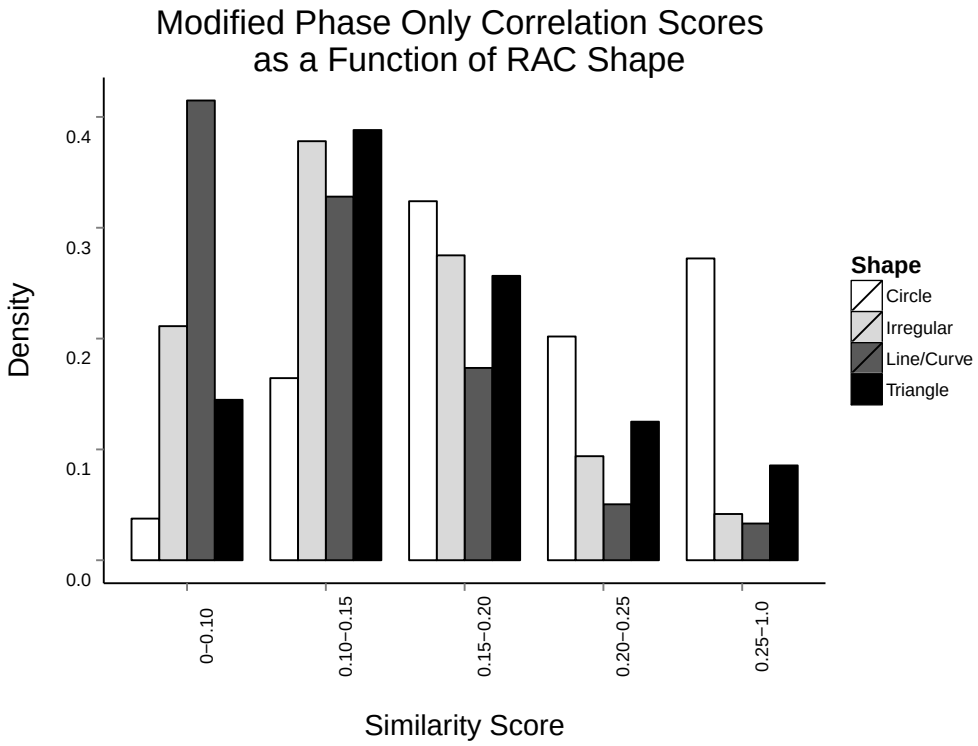


Figure 4: MPOC scores as a function of RAC shape. Circles generally have high similarity scores, while lines and curves exhibit lower similarity scores. Based on the Chi-square results, significant differences in similarity scores as a function of shape existed within all bins.

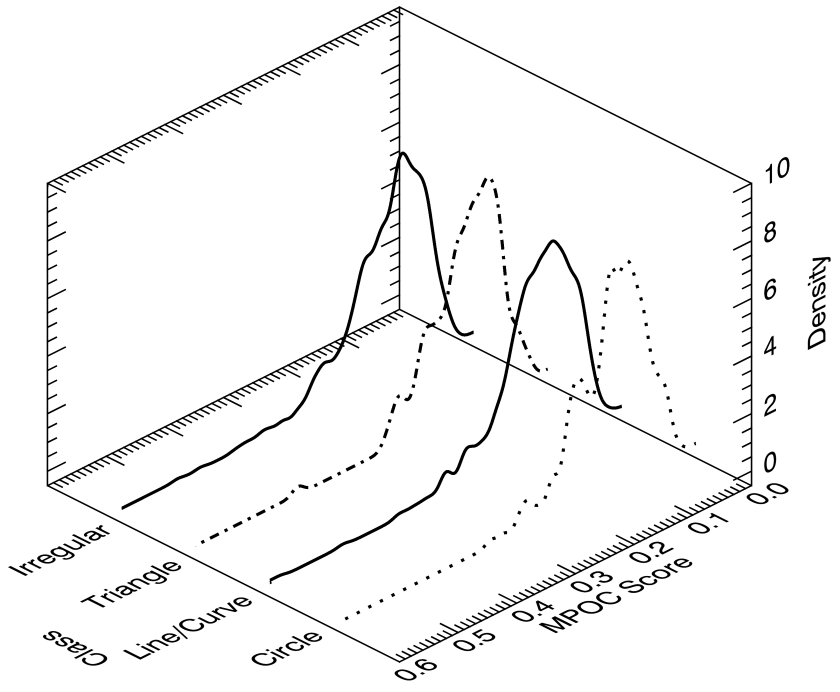


Figure 5: Probability density functions for MPOC scores as a function of RAC shape, obtained using a Gaussian kernel density estimator. Bin width information can be found in the Appendix.

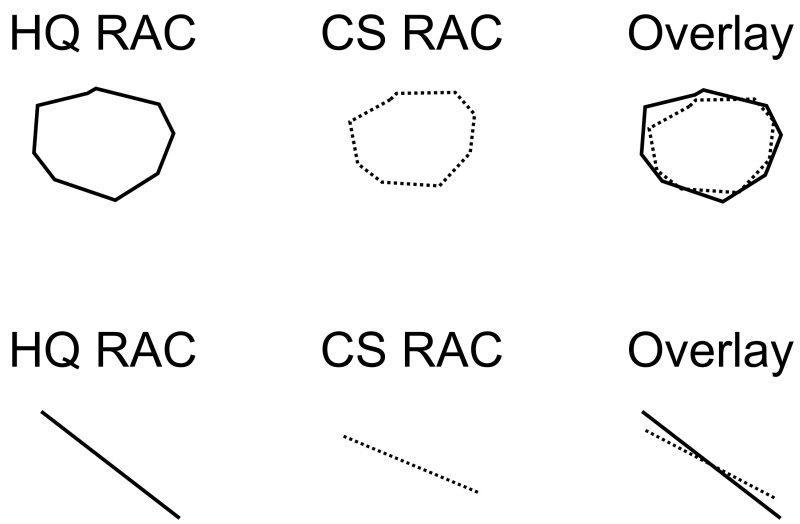


Figure 6: Example of stylized high quality (HQ) and crime scene (CS) RACs. Note that lines exhibit greater discordance (overlap very little) than circular shapes when orientation differences exists (scale and rotational differences are shown for maximum emphasis).

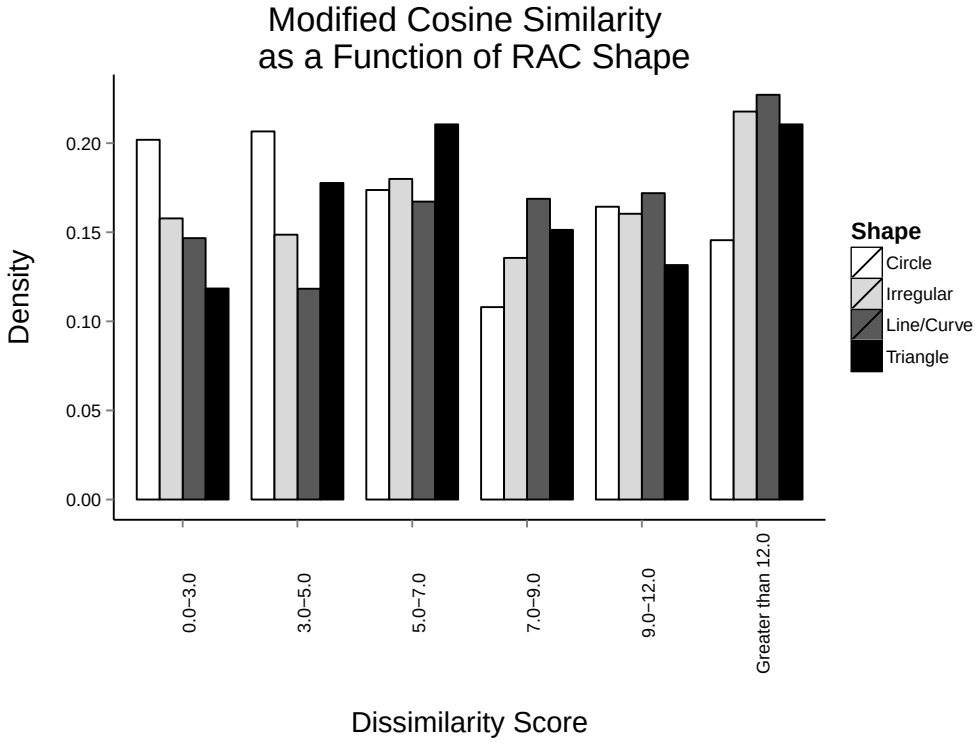


Figure 7: Modified Cosine Similarity scores as a function of RAC shape. The largest differences in RAC similarity scores, as a function of shape, exist in the bins which describe more similar features (*i.e.,* cosine angles of less than 5°). Based on the Chi-square results significant differences in MCS scores as a function of shape existed for scores ranging from 3-5°.

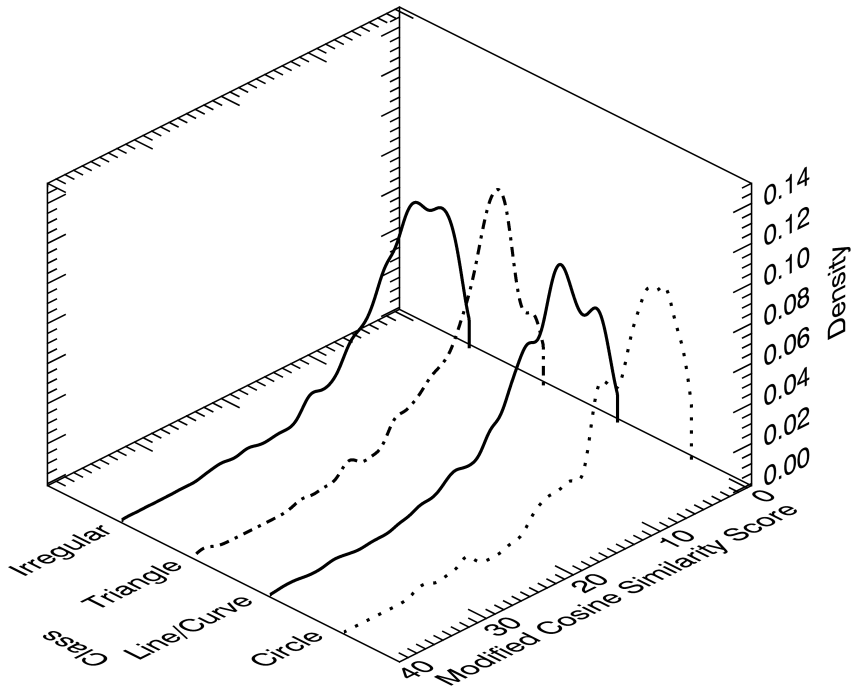


Figure 8: Probability density functions for MCS scores as a function of RAC shape, obtained using a Gaussian kernel density estimator. Bin width information can be found in the Appendix.

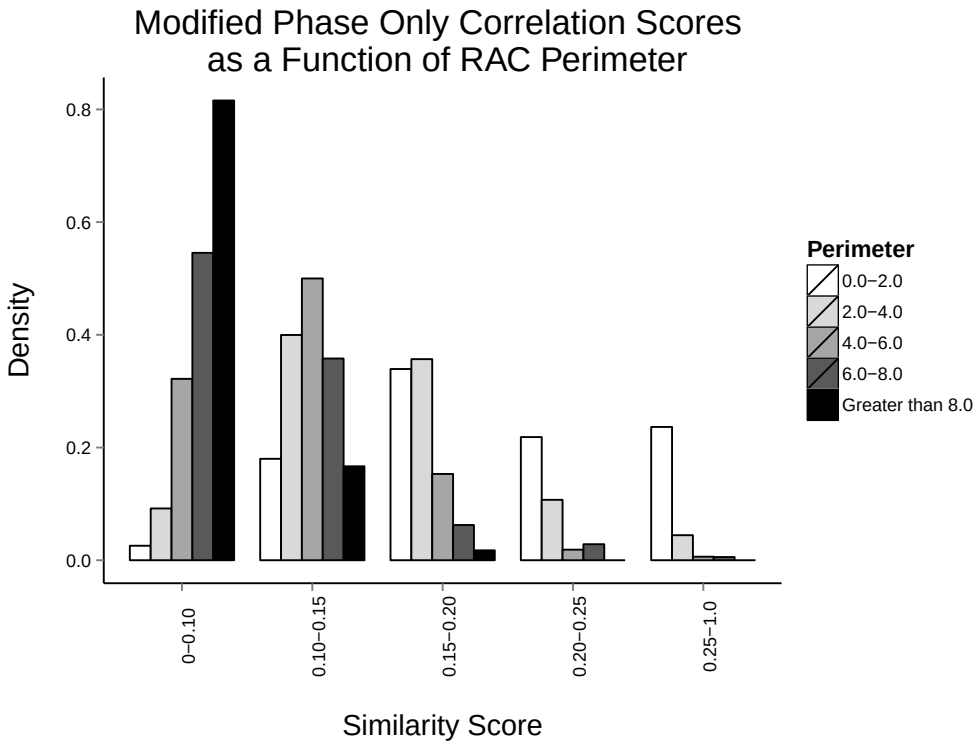


Figure 9: MPOC scores as a function of RAC perimeter. Based on MPOC, very large RACs are very dissimilar. Based on the Chi-square results, significant differences in similarity scores as a function of perimeter existed within all bins.

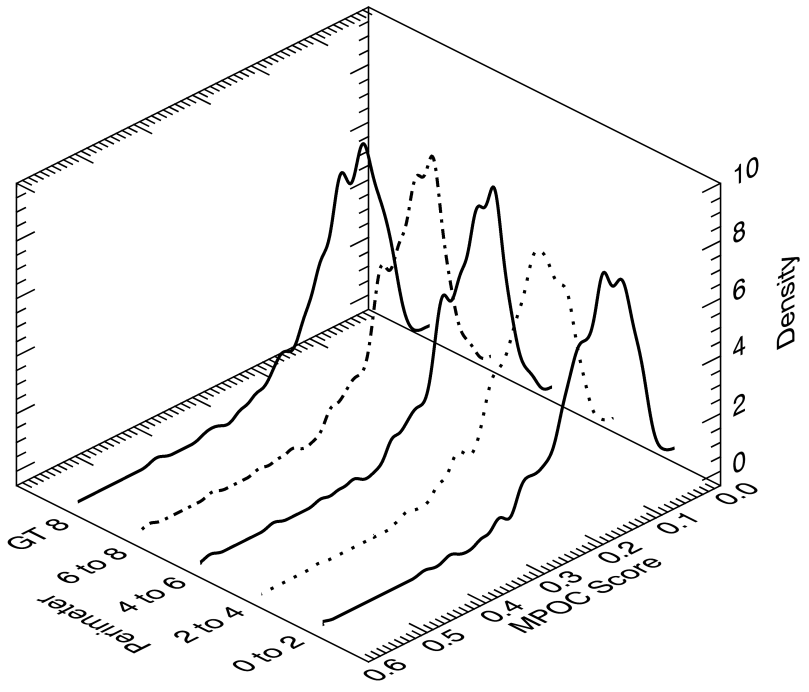


Figure 10: Probability density functions for MPOC scores as a function of RAC perimeter, obtained using a Gaussian kernel density estimator. Bin width information can be found in the Appendix.

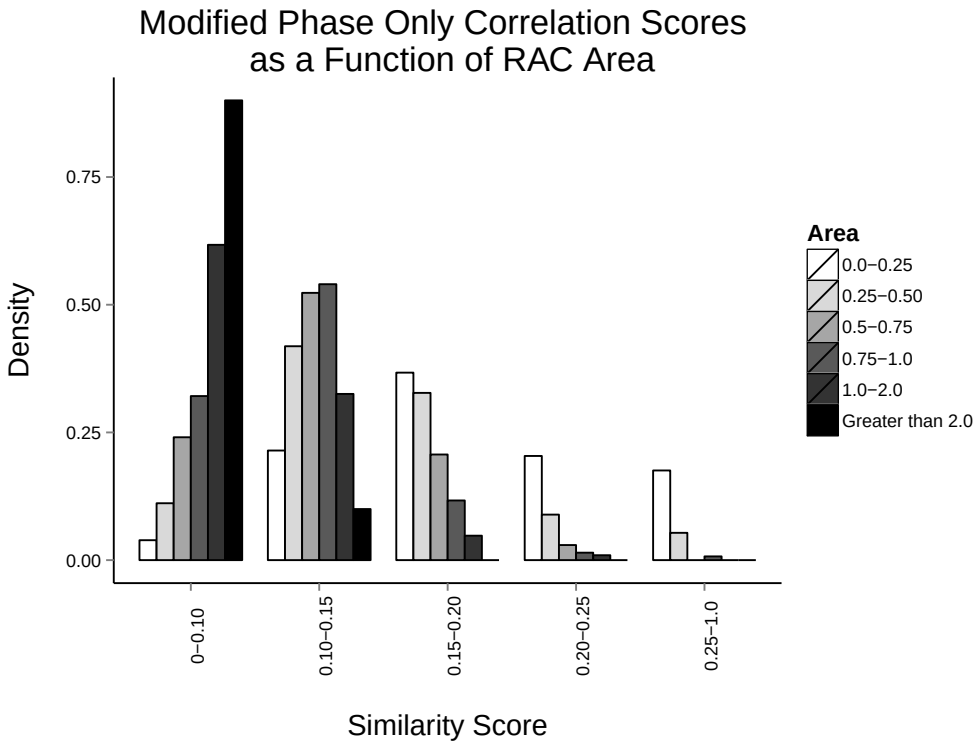


Figure 11: MPOC scores as a function of RAC area. Based on MPOC, very large RACs are very dissimilar. Based on the Chi-square results, significant differences in similarity scores as a function of area existed within all bins.

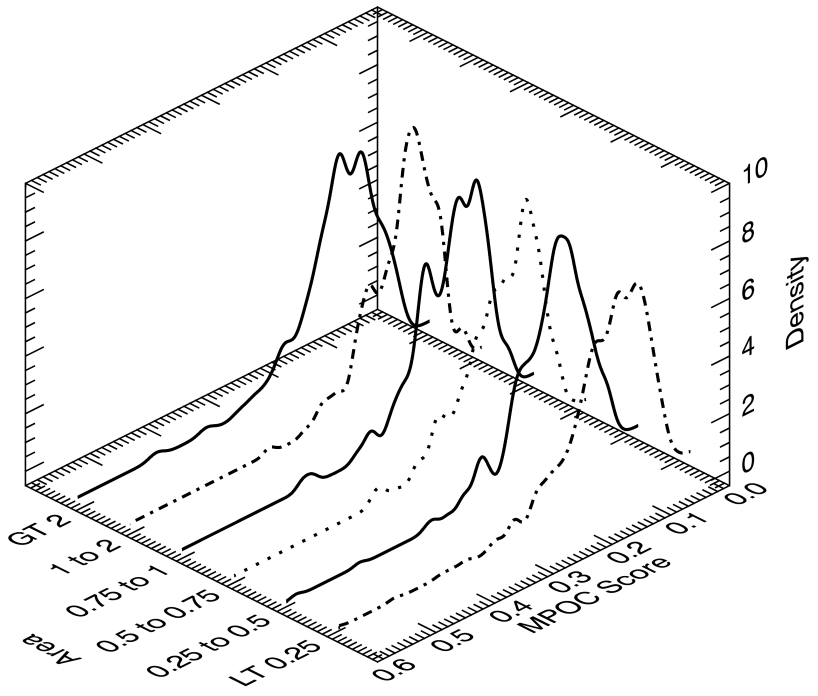


Figure 12: Probability density functions for MPOC as a function of RAC area, obtained using a Gaussian kernel density estimator. Bin width information can be found in the Appendix.

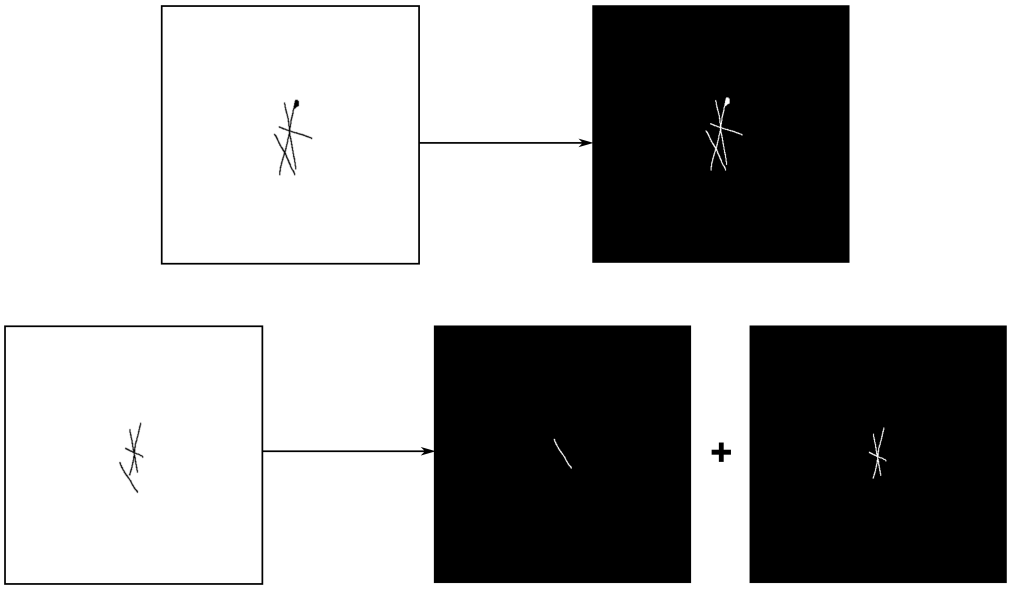


Figure 13: Original marked RAC on high quality exemplar (top left) with corresponding RAC image obtained through connected components (top right). Corresponding RAC on crime-scene-like print (bottom left) and RAC images obtained through connected components (bottom center and right). The crime-scene-like RACs exhibit more voids and are incomplete in comparison with their high quality counterparts.

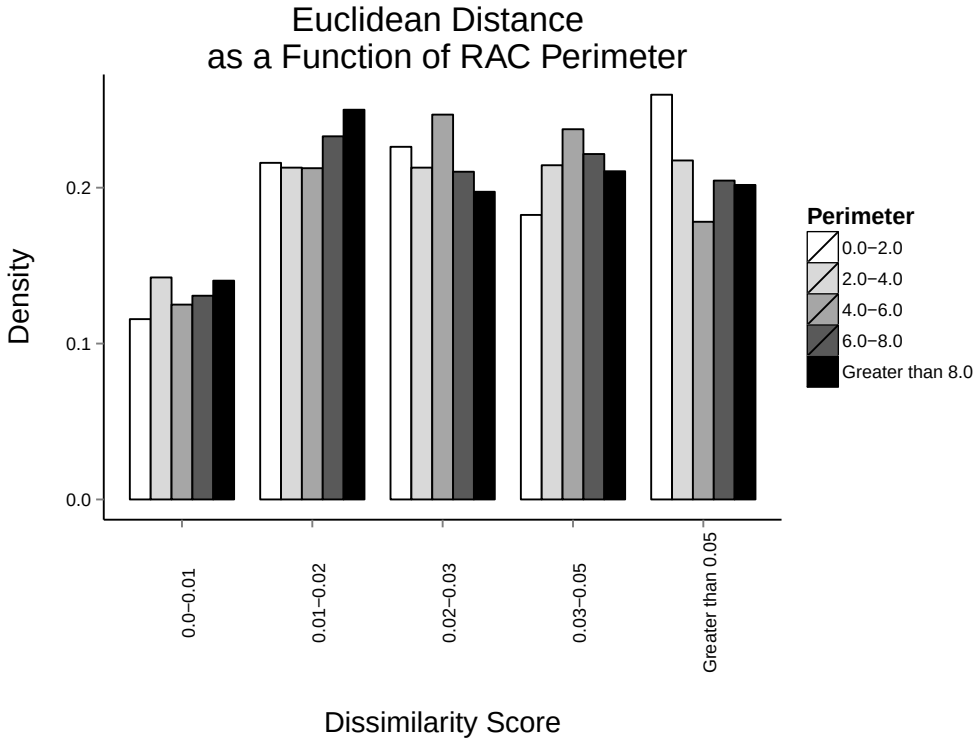


Figure 14: Euclidean distance as a function of RAC perimeter. Small RACs are more similar to their known match mates than large RACs. Based on the Chi-square results significant differences in ED scores as a function of perimeter existed for scores ranging from 0.01-0.02 and those greater than 0.05.

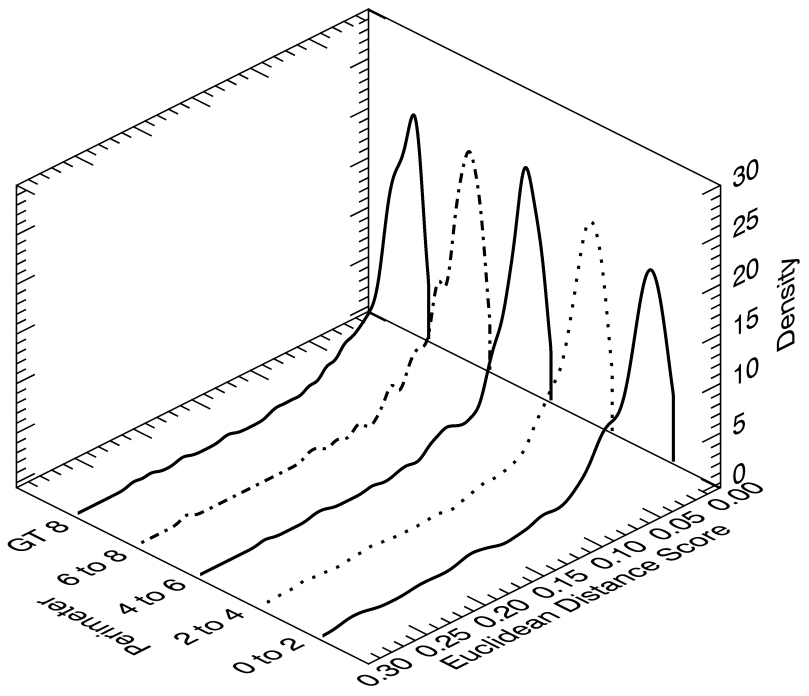


Figure 15: Probability density functions for ED as a function of RAC perimeter, obtained using a Gaussian kernel density estimator. Bin width information can be found in the Appendix.

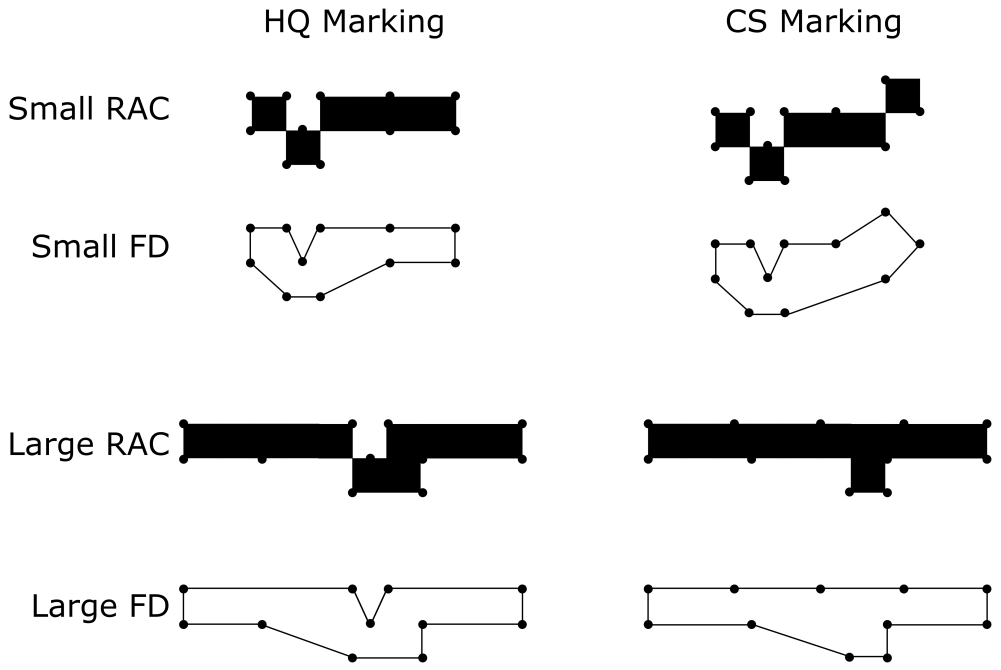


Figure 16: Example of stylized small and large high quality (HQ) and corresponding crime scene (CS) RACs. There is a 1 pixel difference between the HQ and CS marking for both the small and large features. In addition, there are 11 interpolation points which are used to extract the perimeter image of each RAC, similar to a Fourier descriptor (FD) perimeter representation. Note that the shape of the small RAC is more severely skewed by interpolation than the shape of the large feature.

Table 5: RAC map density.

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%)**

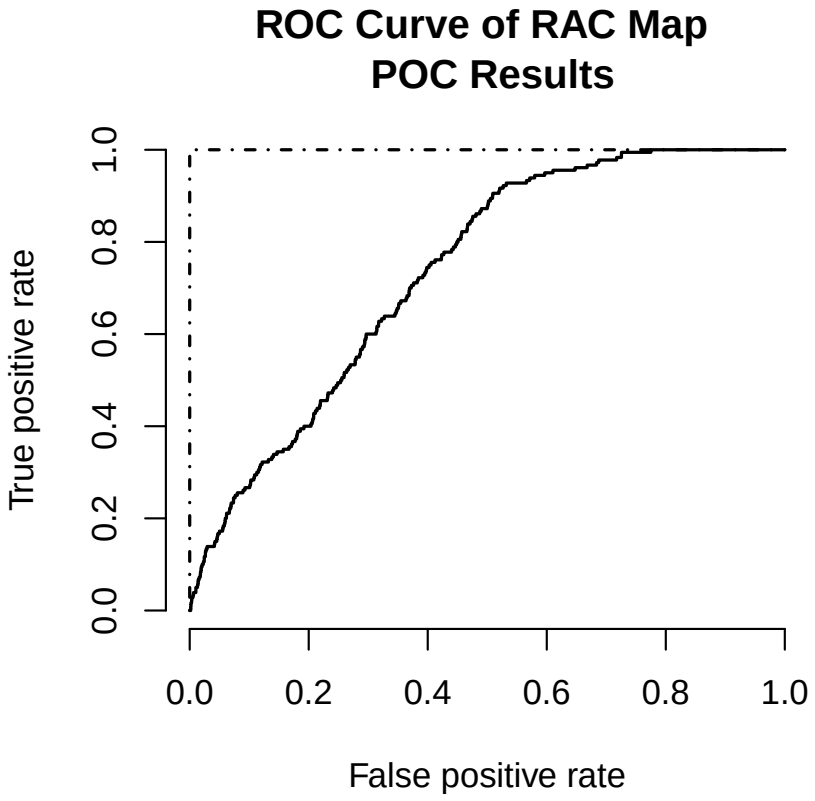


Figure 17: Receiver operator characteristic curve of RAC map POC results. High quality comparisons are represented by the dash-dotted line and exhibit an area under the curve (AUC) of 1.0. The solid line illustrates the results of crime-scene-like impressions with an AUC of 0.74.

Appendix

As previously stated, similarity score were analyzed as a function of RAC shape, perimeter, and area. Detailed in Tables 6, 7, 8 are the bin widths used to best approximate the true data when constructing the PDFs. Therefore, a total of 75 individual PDFs were constructed. Subsequently, the plots for related conditions (*i.e.,* MPOC as a function of shape) were compiled into 3-dimensional plots, thus allowing for easier visual comparison.

Table 6: Bin widths for PDFs (obtained using a Gaussian kernel density estimator) for similarity scores as a function of RAC shape. The bin width for each density estimate was set equal to one seventh of the standard deviation of the similarity scores.

	Circle	Line/Curve	Triangle	Irregular
MPOC	0.008	0.010	0.010	0.009
MF	0.003	0.002	0.004	0.003
MCS	0.804	0.827	0.942	0.864
HD	1.033	0.969	0.646	0.995
ED	0.006	0.006	0.006	0.005

Table 7: Bin widths for PDFs (obtained using a Gaussian kernel density estimator) for similarity scores as a function of RAC perimeter. The bin width for each density estimate was set equal to one seventh of the standard deviation of the similarity scores.

	0-2	2-4	4-6	6-8	Greater than 8
MPOC	0.009	0.009	0.009	0.009	0.011
MF	0.003	0.003	0.002	0.002	0.002
MCS	0.815	0.903	0.763	0.867	0.872
HD	1.109	0.949	0.843	0.981	0.886
ED	0.007	0.005	0.005	0.005	0.005

Table 8: Bin widths for PDFs (obtained using a Gaussian kernel density estimator) for similarity scores as a function of RAC area. The bin width for each density estimate was set equal to one seventh of the standard deviation of the similarity scores.

	0-0.25	0.25-0.5	0.5-0.75	0.75-1.0	1-2	Greater than 2
MPOC	0.010	0.009	0.009	0.010	0.009	0.013
MF	0.003	0.003	0.003	0.002	0.003	0.002
MCS	0.855	0.860	0.833	0.823	0.896	0.813
HD	1.043	1.059	0.799	0.914	0.884	0.754
ED	0.006	0.005	0.005	0.005	0.005	0.004

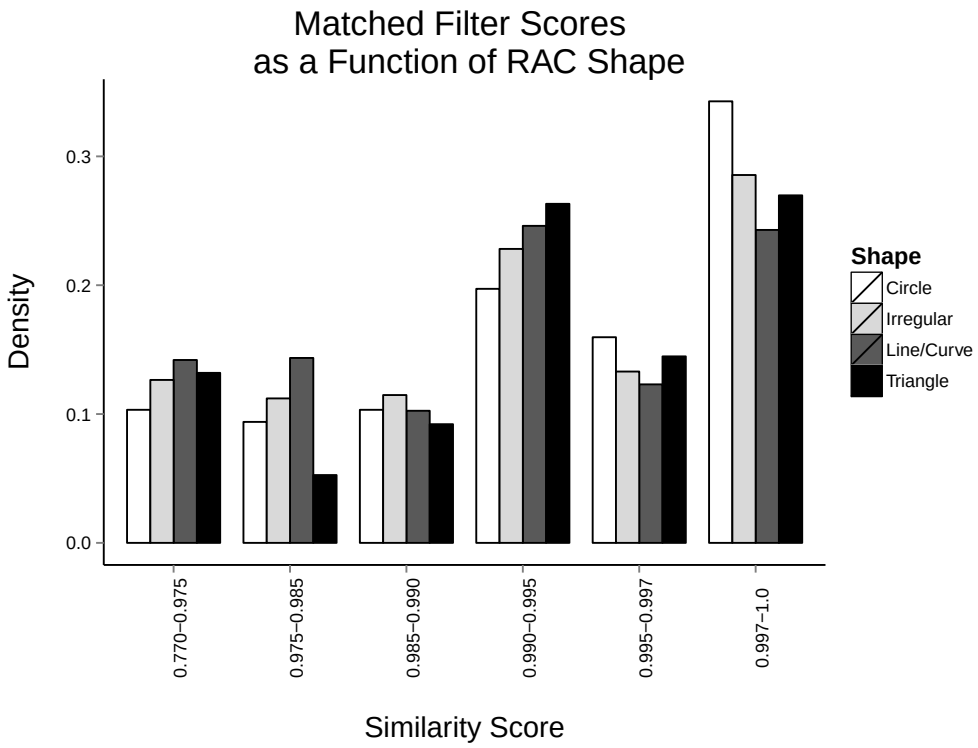


Figure 18: MF scores as a function of RAC shape.

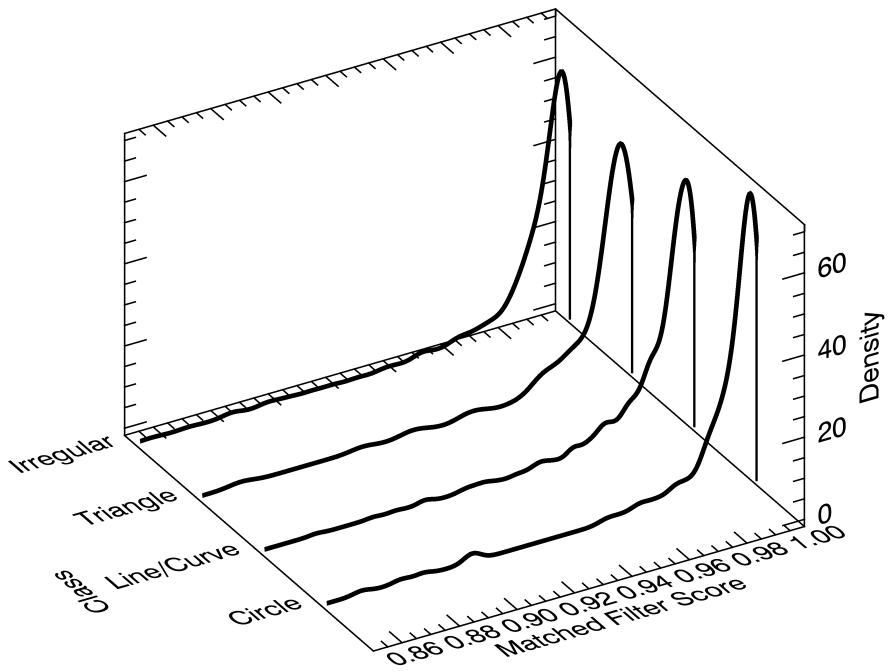


Figure 19: Probability density functions for MF scores as a function of RAC shape, obtained using a Gaussian kernel density estimator.

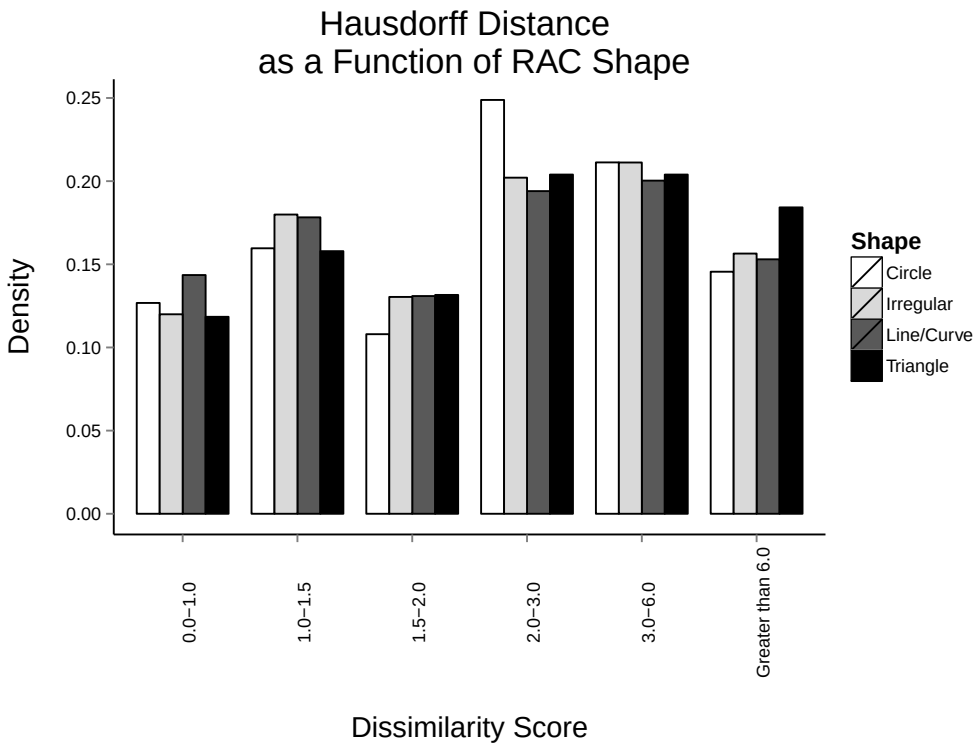


Figure 20: Hausdorff distance as a function of RAC shape.

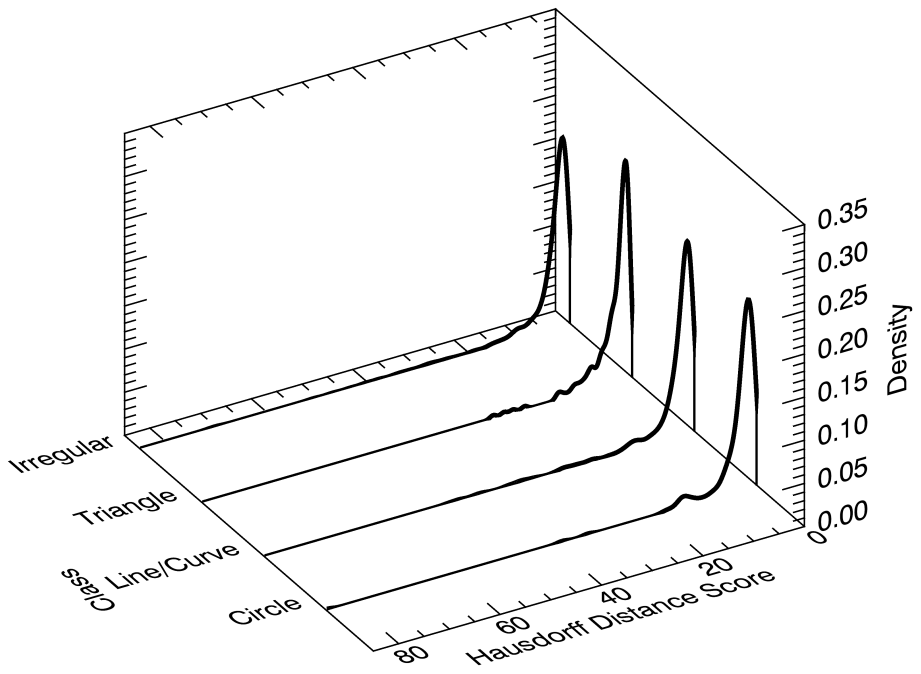


Figure 21: Probability density functions for Hausdorff distance as a function of RAC shape, obtained using a Gaussian kernel density estimator.

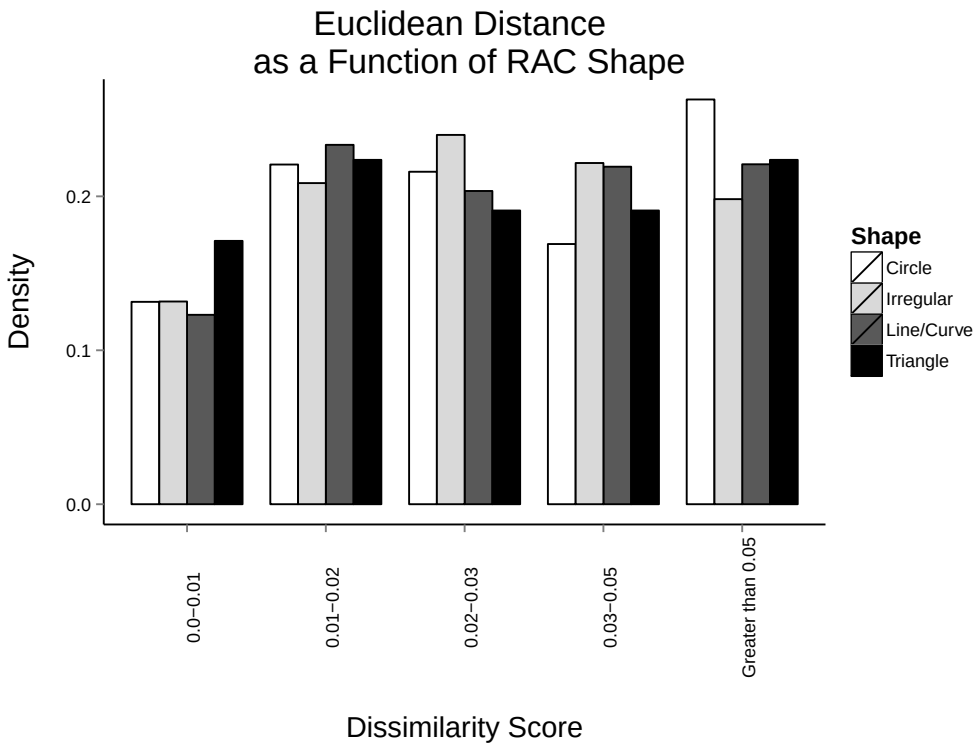


Figure 22: Euclidean distance as a function of RAC shape.

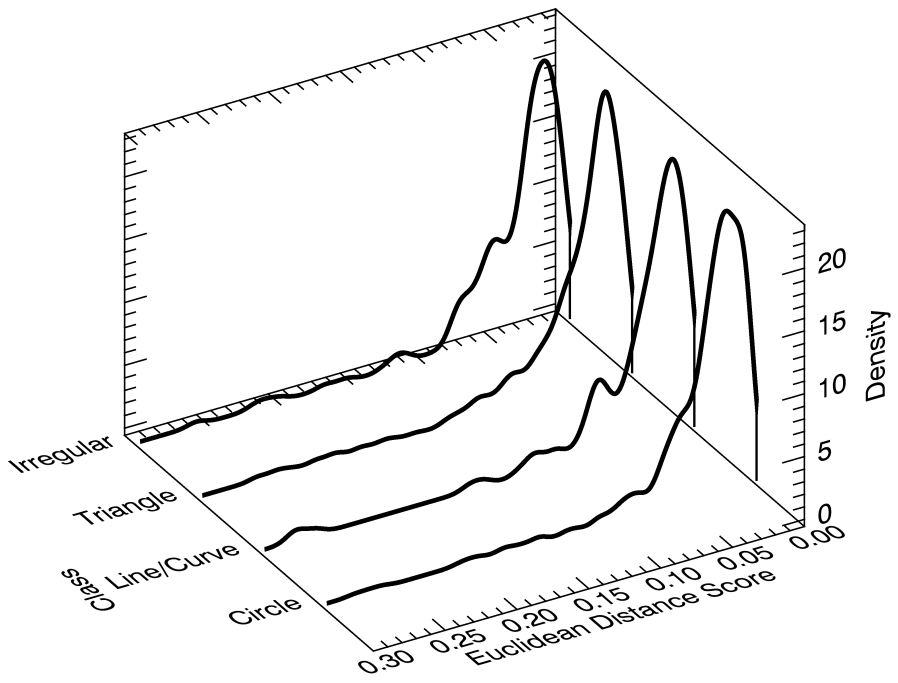


Figure 23: Probability density functions for Euclidean distance as a function of RAC shape, obtained using a Gaussian kernel density estimator.

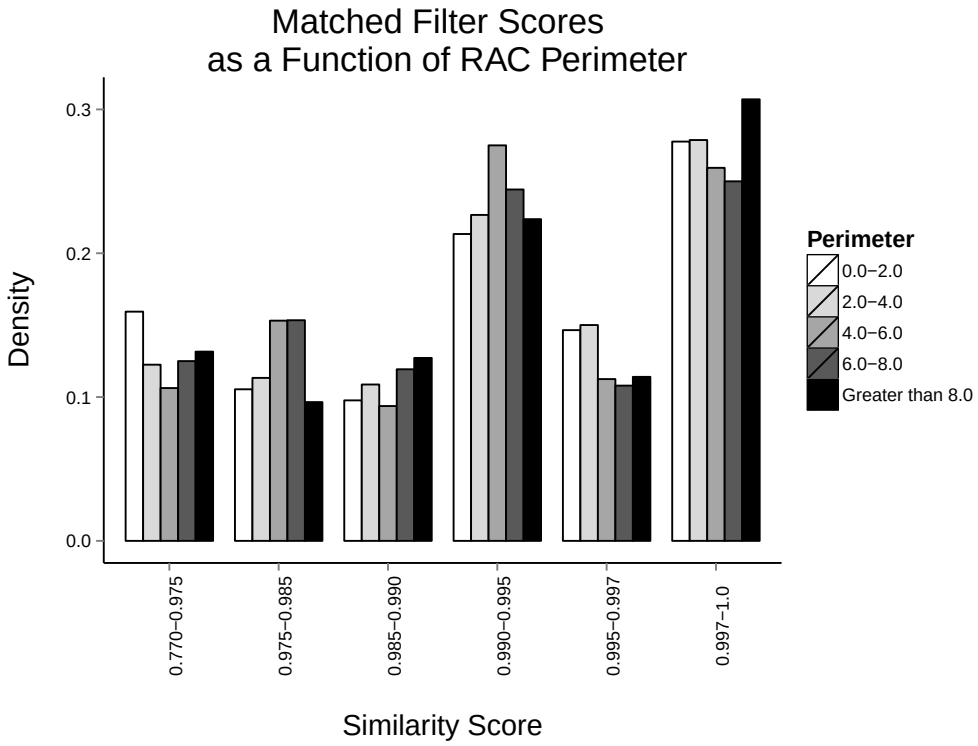


Figure 24: MF scores as a function of RAC perimeter.

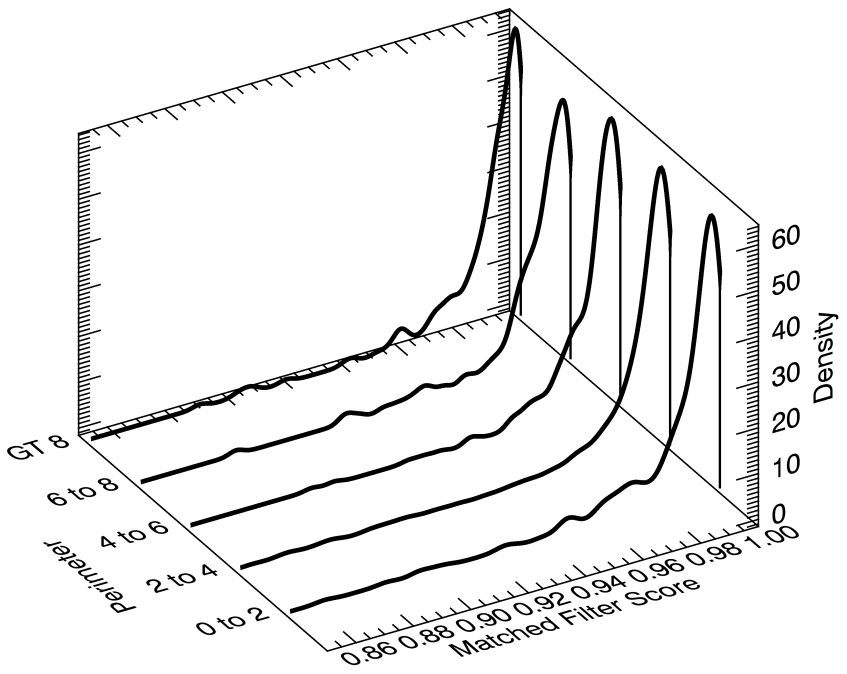


Figure 25: Probability density functions for MF scores as a function of RAC perimeter, obtained using a Gaussian kernel density estimator.

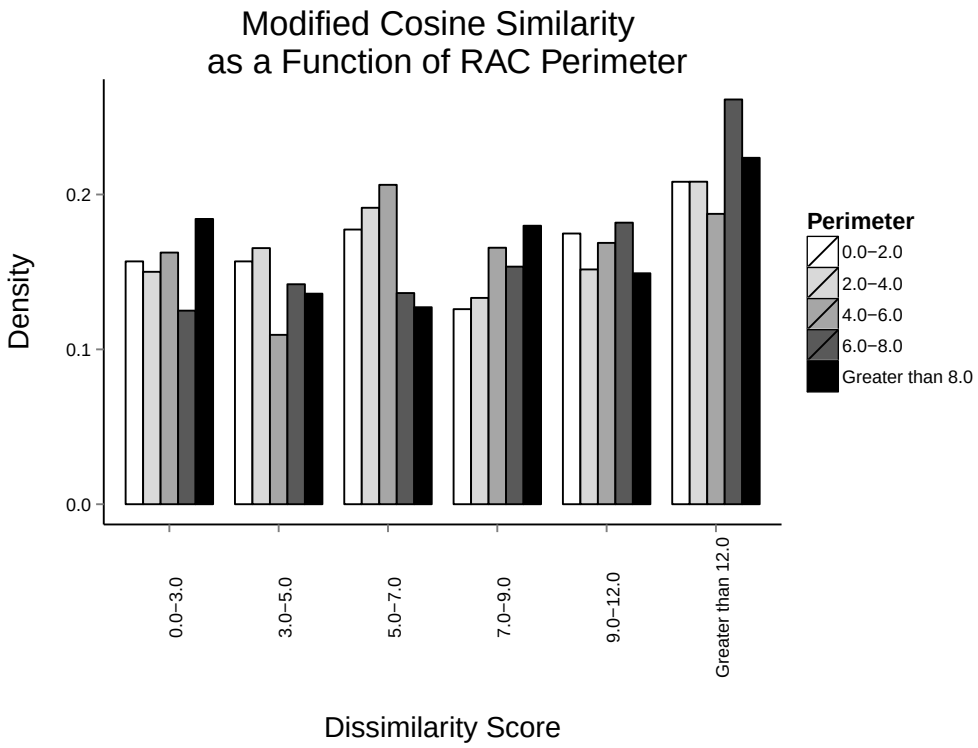


Figure 26: MCS scores as a function of RAC perimeter.

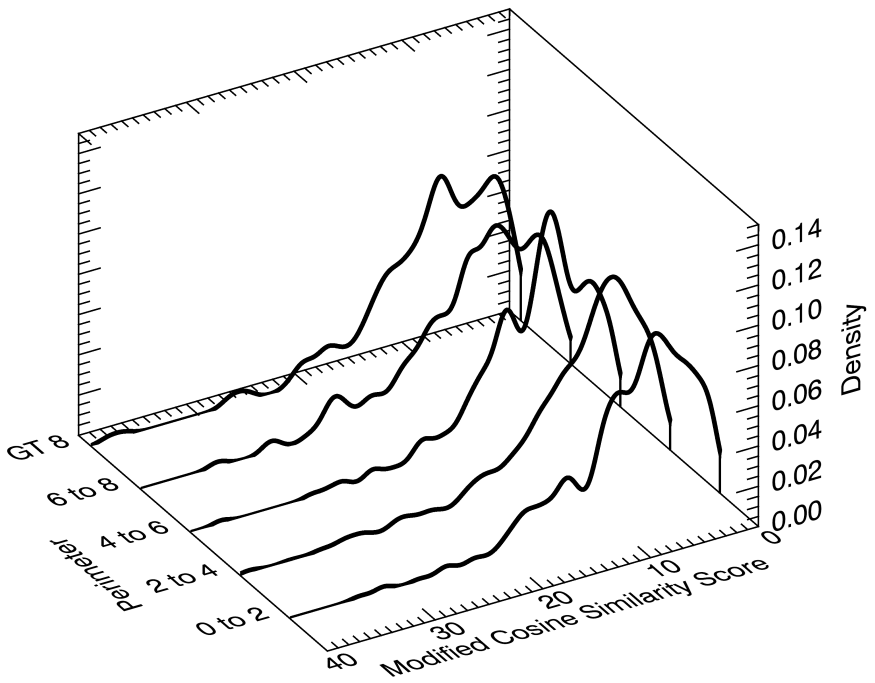


Figure 27: Probability density functions for MCS scores as a function of RAC perimeter, obtained using a Gaussian kernel density estimator.

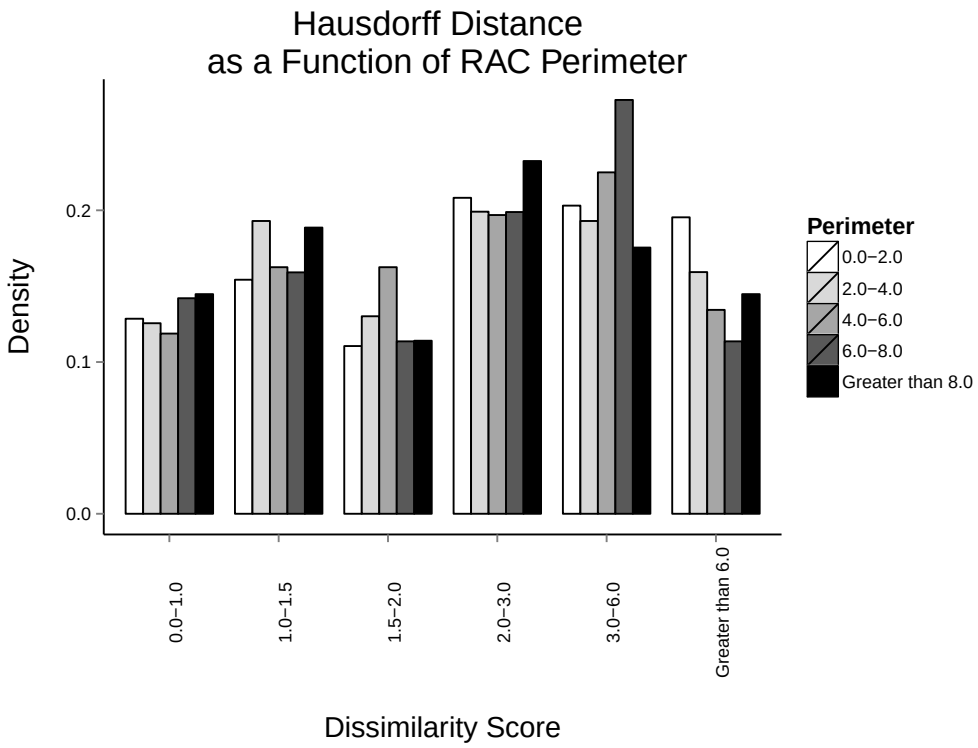


Figure 28: Hausdorff distance as a function of RAC perimeter.

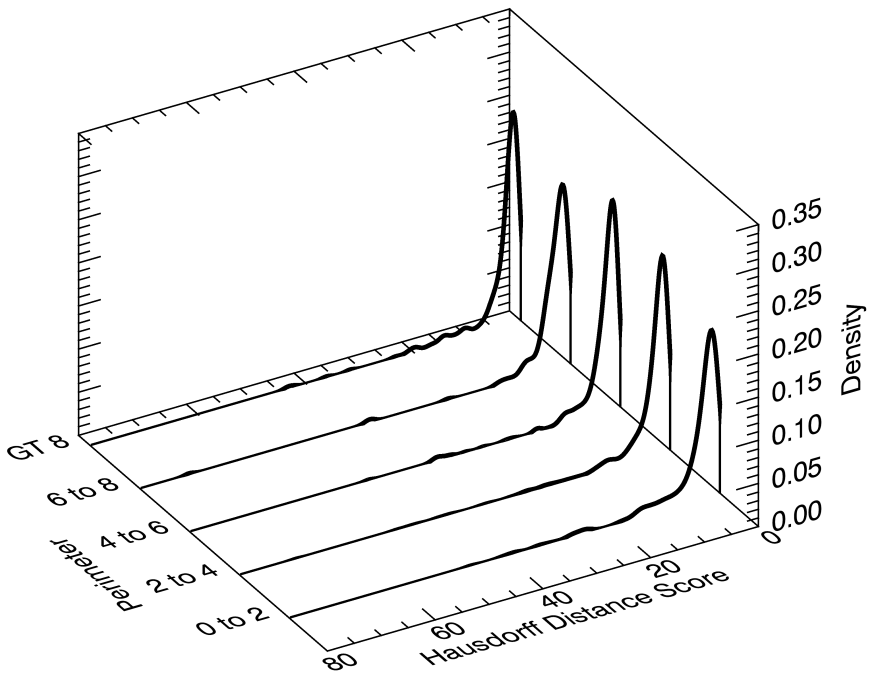


Figure 29: Probability density functions for Hausdorff distance as a function of RAC perimeter, obained using a Gaussian kernel density estimator.

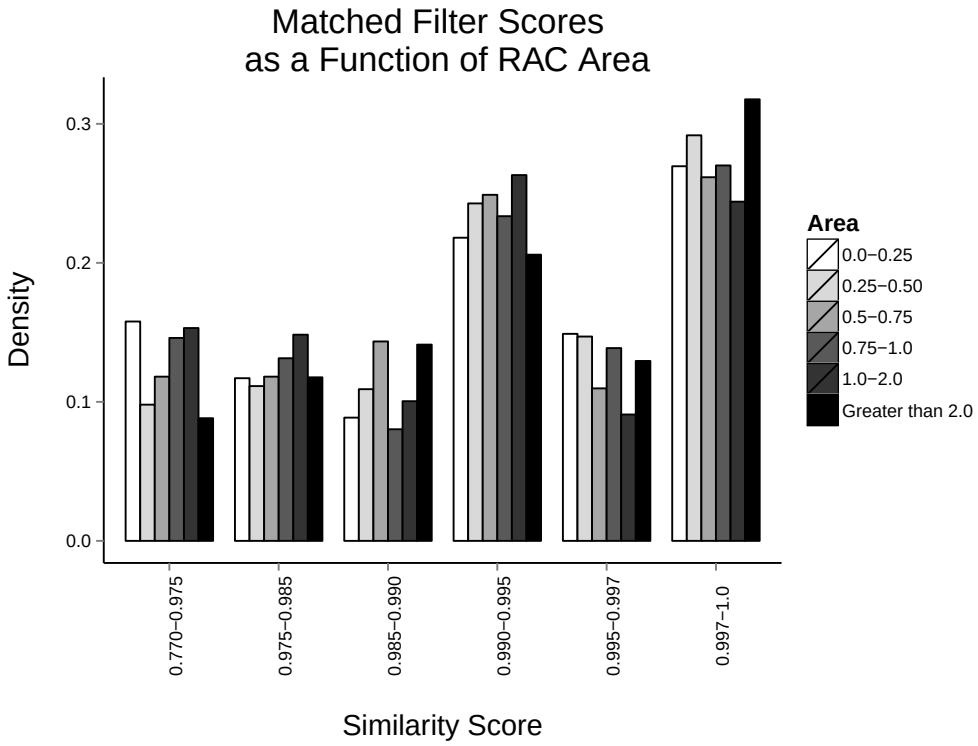


Figure 30: MF scores as a function of RAC area.

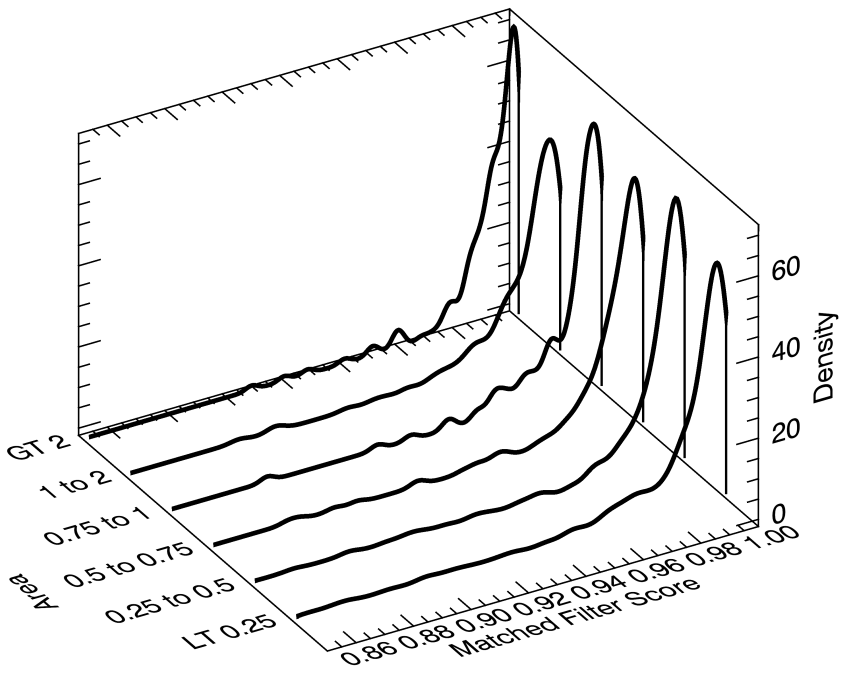


Figure 31: Probability density functions for MF scores as a function of RAC area, obtained using a Gaussian kernel density estimator.

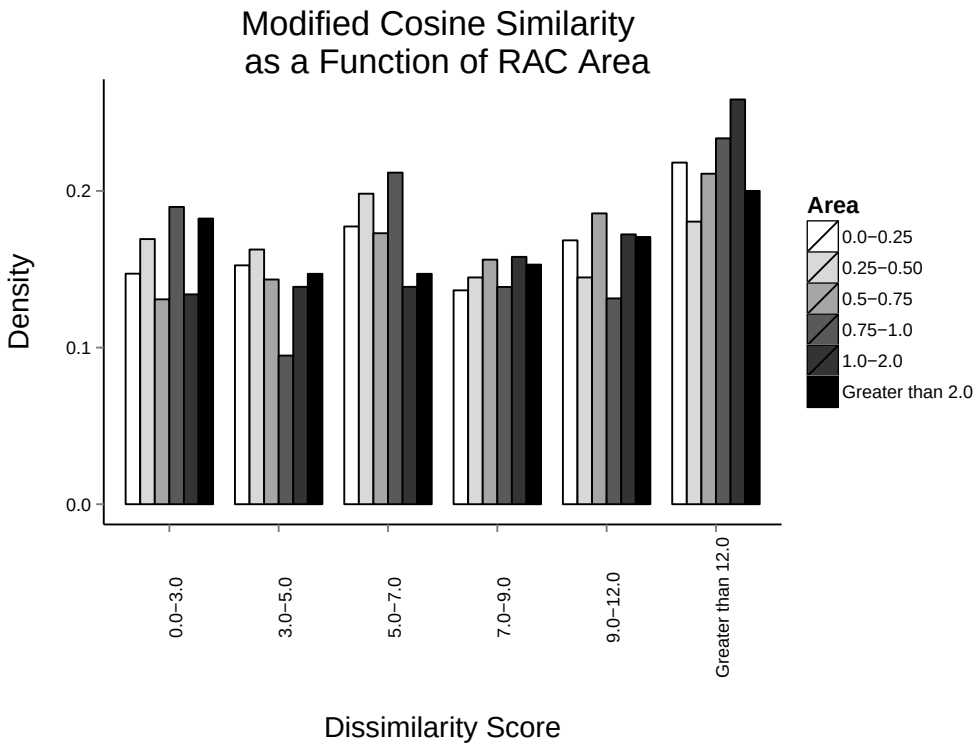


Figure 32: MCS scores as a function of RAC area.

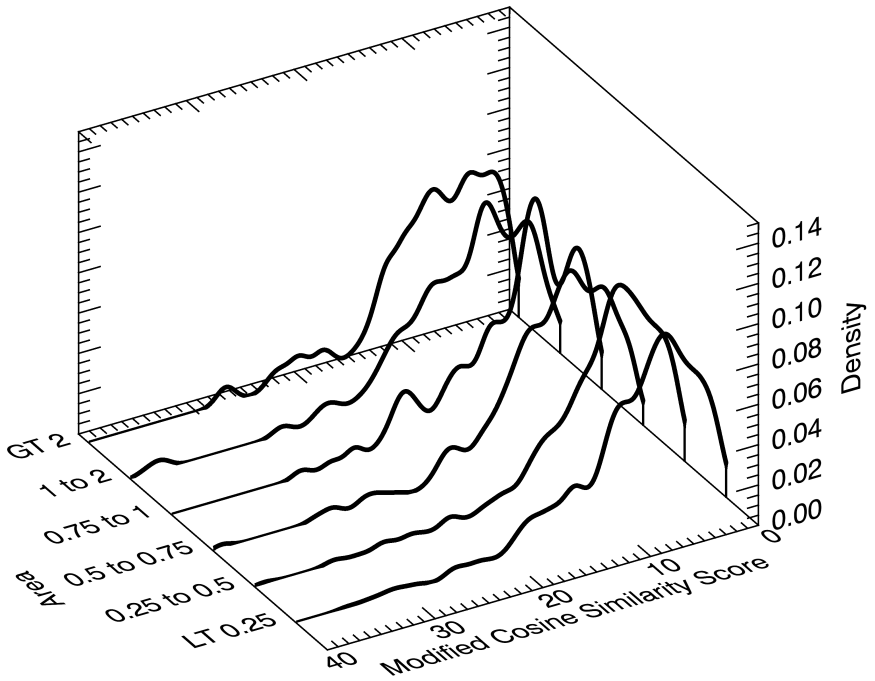


Figure 33: Probability density functions for MCS scores as a function of RAC area, obtained using a Gaussian kernel density estimator.

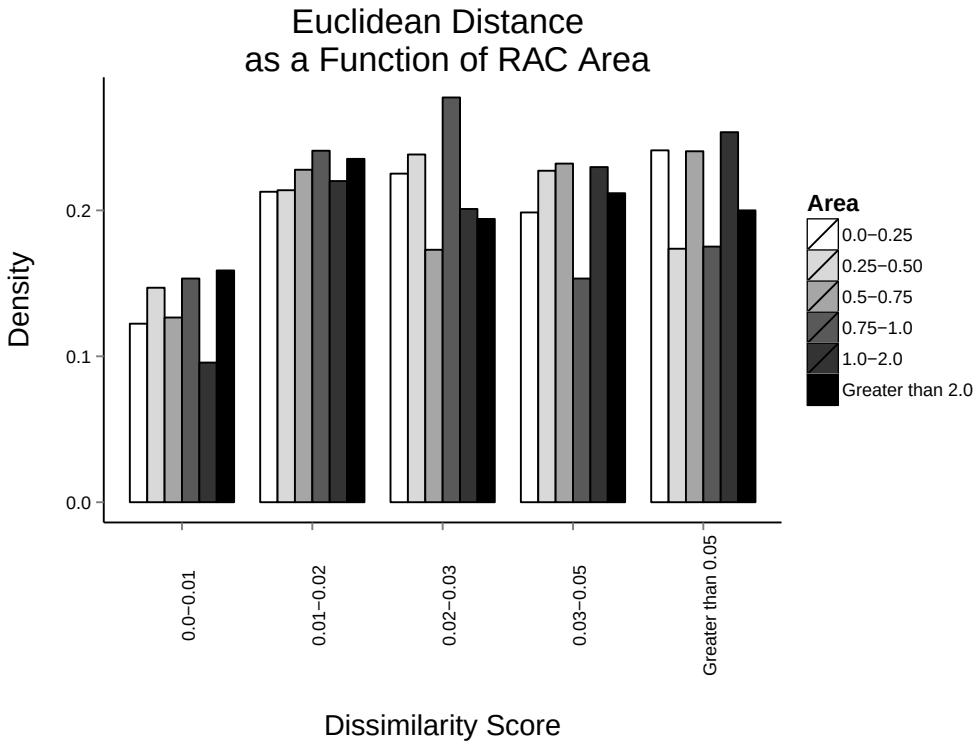


Figure 34: Euclidean distance as a function of RAC area.

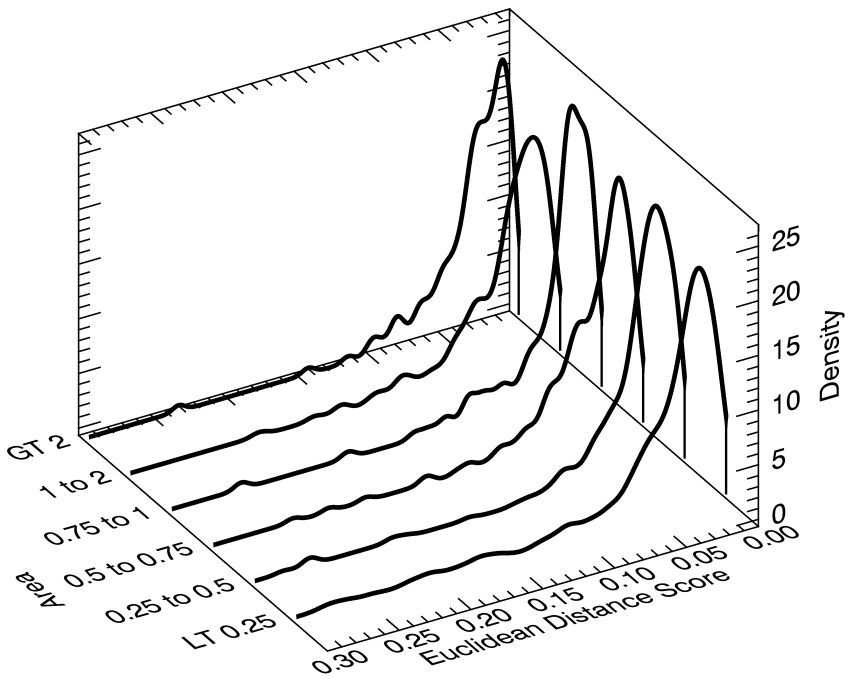


Figure 35: Probability density functions for Euclidean distance as a function of RAC area, obtained using a Gaussian kernel density estimator.

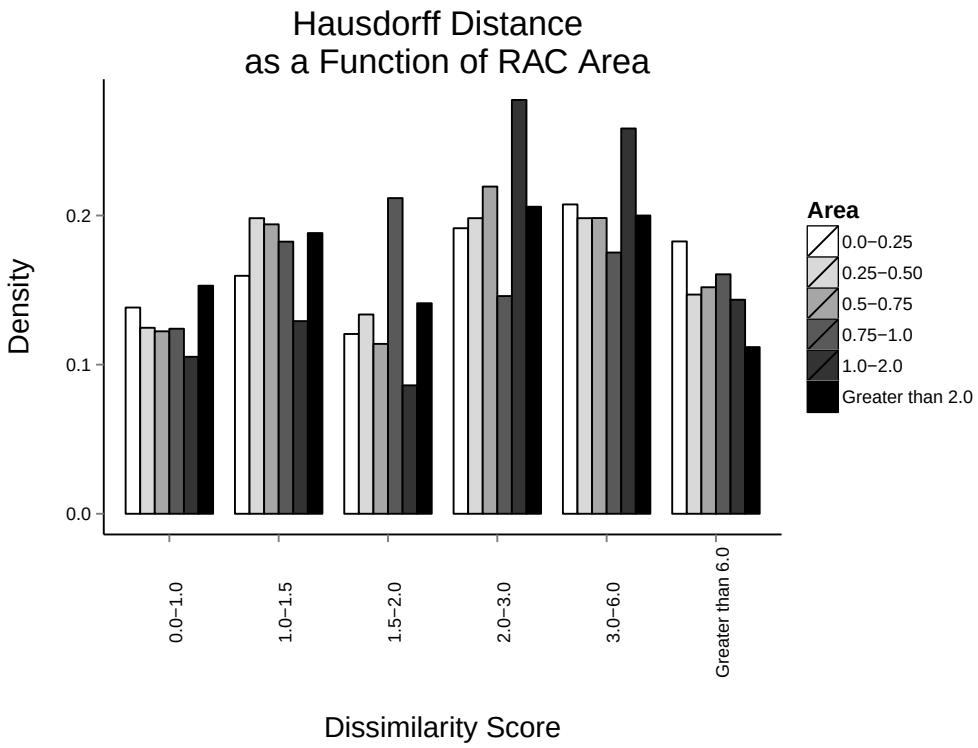


Figure 36: Hausdorff distance as a function of RAC area.

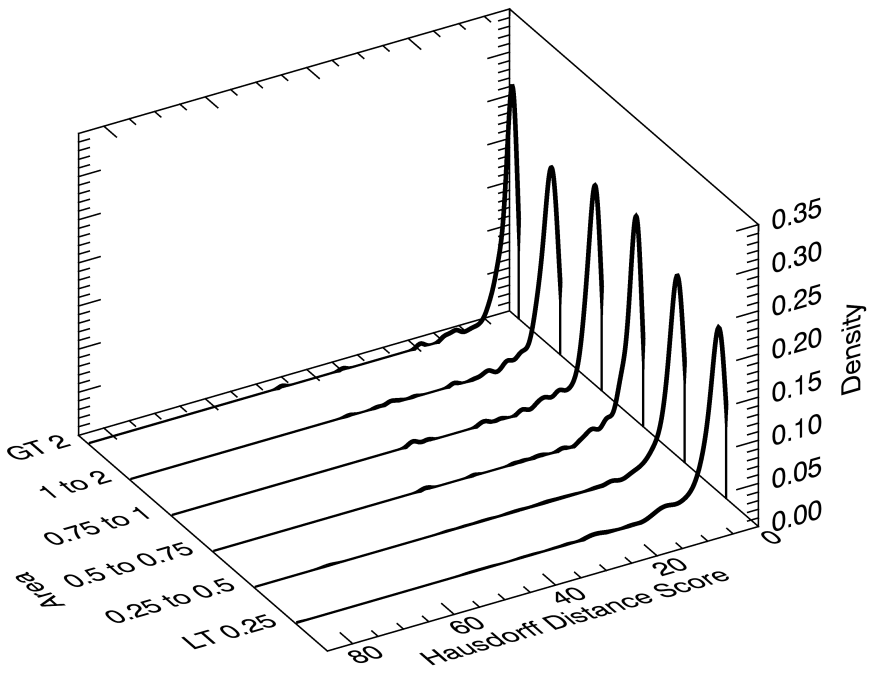


Figure 37: Probability density functions for Hausdorff distance as a function of RAC area, obtained using a Gaussian kernel density estimator.

[1] A. Alexander, A. Bouridane, D. Crookes, Automated classification and recognition of shoeprints, Image Processing and its Applications 465 (1999) 638–641.

[2] P. de Chazal, J. Flynn, R. Reilly, Automated processing of shoeprint images based on the Fourier transform for use in forensic science, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 341–50.

[3] M. Gueham, A. Bouridane, D. Crookes, Automatic recognition of partial shoeprints based on phase-only correlation, IEEE 4 (2007) 441–444.

[4] C. Champod, R. Voisard, A. Girod, A statistical study of air bubbles on athletic shoe soles, Forensic Science International 109 (2000) 105–123.

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

[6] M. Cassidy, Footwear Identification (Reprint), Lightning Powder Company, Inc., 1995.

[7] R. Stone, Footwear examination: Mathematical probabilities of theoretical individual characteristics, Journal of Forensic Identification 56 (2006) 577–599.

[8] N. Petraco, C. Gambino, T. Kubic, D. Olivio, N. Petraco, Statistical discrimination of footwear: A method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques, Journal of Forensic Sciences 55 (2010) 34–41.

[9] H. D. Sheets, S. Gross, G. Langenburg, P. J. Bush, M. A. Bush, Shape measurement tools in footwear analysis: A statistical investigation of accidental characteristics over time, Forensic Science International 232 (2013) 84–91.

[10] W. J. Bodziak, Footwear Impression Evidence: Detection, Recovery and Examination, Second Edition, CRC Press, 2000.

[11] D. S. Hilderbrand, Four basic components of a succcessful footwear examination, Journal of Forensic Identification 49 (1999) 37–59.

[12] L. Tocan, Footwear trace, concept, interpretation: Training in forensics, Romanian Journal of Forensic Science 6(90) (2013) 1517–1520.

[13] L. Hammer, J. Wolfe, Shoe and tire impressions in snow: Photography and casting, Journal of Forensic Identification 53(6) (2003) 647–655.

[14] C. Snyder, The ability of footwear to produce impressions of good detail in sandy soil substrates, Journal of Forensic Identification 65(3) (2015) 273–288.

[15] J. Vandiver, Identification of suitable plaster for crime scene casting, Journal of Forensic Sciences 23(3) (1978) 607–614.

[16] P. Iten, Recovery of shoe and fingerprint impressions by means of electrostatic transfer, Kriminalistik 10 (1986) 468–470.

[17] J. Keijzer, I. Keereweer, H. van der Heuvel, Z. Geradts, The use of gelatin lifters for the recovery of footwear impressions in combination

with latent fingerprint powders, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[18] Y. Shor, A. Vinokurov, B. Glattstein, The use of an adhesive lifter and pH indicator for the removal and enhancement of shoeprints in dust, Journal of Forensic Science 43(1) (1998) 182–184.

[19] A. Mankevich, Chemical enhancement of clay residue footwear impressions, 78th Annual Educational Conference, International Association for Identification (1993).

[20] D. Brundage, Physical developer: A chemical enhancement technique for forensic footwear impressions, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[21] L. Dinkins, Development and enhancement of footwear impressions on non-porous surfaces, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[22] A. Theeuwen, S. van Barneveld, J. Drok, I. Keereweer, J. Limborgh, W. Naber, T. Velders, Enhancement of footwear impressions in blood, Forensic Science International 95(2) (1998) 133–151.

[23] K. Farrugia, H. Bandey, L. Dawson, N. Daéid, Chemical enhancement of soil based footwear impressions on fabric, Forensic Science International 219 (2012) 12–28.

[24] C. Karazalus, T. Palmbach, H. Lee, Digital enhancement of sub-quality bitemark photographs, Journal of Forensic Sciences 46(4) (2001) 954–958.

[25] M. Sharma, J. S., Uses of softwear in digital image analysis: A forensic report, Proceedings of the Second International Conference on Digital Image Processing 7546 (2010).

[26] T. Krauss, S. Warlen, The forensic science use of reflective ultraviolet photography, Journal of Forensic Sciences 30(1) (1985) 262–268.

[27] A. Ytii, Enhancement of shoeprints with Polilight- a case report, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[28] V. Inlow, Use of forensic light sources in the detection of imrpession evidence on deceased persons, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[29] J. Hebrard, E. Dupasquier, P. Romanello, G. Poully, Experimental and comparative study of new casting materials, International Symposium of the Forensic Aspects of Footwear and Tire Impression Evidence (1994).

[30] A. Bouridane, Imaging for Forensics and Security: From Theory to Practice., Springer, 2009.

[31] R. Xiao, P. Shi, Computational Forensics: Lecture Notes in Computer Science, volume 5158, Springer, 2008.

[32] T. Wallace, O. Mitchell, Analysis of three-dimensional movement using fourier descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2, NO. 6 (1980) 583–588.

[33] I. Bartolini, P. Ciaccia, M. Patella, WARP: Accurate retrieval of sshape using phase of fourier descriptors and time warping distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 27, No. 1 (2005) 142–147.

[34] C. Dalitz, C. Brandt, S. Goebbels, D. Kolanus, Fourier descriptors for broken shapes, EURASIP Journal on Advances in Signal Processing 161 (2013) 1–11.

[35] J. Gregga, G. Power, K. Iftekharrudin, Correlators for rank order shape similarity measurement, Optical Pattern Recogntion XIII, Proceedings of the SPIE 4734 (2002) 122–131.

[36] J. R. Schott, Remote Sensing: The Image Chain Approach, 2nd Edition, Oxford University Press, 2007.

[37] D. Huttenlocher, G. Klanderman, W. Rucklidge, Comparing images using the Hausdorff distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 15(9) (1993) 850–863.

[38] M. McHugh, The chi-square test of independence, Biochemica Medica 23(2) (2013) 143–149.

5. Future Directions

5.1 Chance Co-occurrence in Position & Shape

The normalized shoe used in this study was a men's size 10 Reebok® walking shoe with an outsole surface area of 21,235mm^2. Using θ and r_{norm} each RAC (total of 57,426 RACs from 1,000 shoes) was localized into a cell measuring 5mm x 5mm in area, generating what is referred to as a heat-map or a plot of frequency versus position. Based on this plot, one bin was found to contain the greatest potential for RAC co-occurrence in position, as illustrated in Table [5.1]. When this result was further broken down as a function of shape category, the probability in co-occurrence ranged from 1:756 to 1:9,571 within a single 5mm x 5mm bin.

Table 5.1: Frequency of RACs and potential for co-occurrence as a function of position and shape for bin located approximately 5mm from the lateral edge and 70mm from the heel of the shoe.

Description	**Any Shape**	**Irregular**	**Circle**	**Triangle**	**Line/Curve**
Total: In Database	57,426	22,075	6,287	3,242	25,822
Total: In Cell	132	39	11	6	76
Chance of Finding RAC in Cell	1:435	1:1,472	1:5,220	1:9,571	1:756

Following localization, all pairwise comparisons in similarity were computed based on shape categorization and using Modified Phase Only Correlation (MPOC). The results are shown in Fig. [5.1], which report MPOC scores, RAC images, and Fourier images for the two most similar RACs detected within the bin. In contrast with all former published results, *some level of visual similarity* can be discerned. This is not to suggest that the accidentals are indistinguishable, since clearly each pair can be differentiated based on size, shape and/or orientation. However, it is relevant to note that there is *some level of expressed similarity* that should not be ignored. Moreover, most accidentals with possible co-occurrence in position and some expressed similarity in shape are *extremely minute in size*. This clearly indicates that more work is needed to better understand the limit of discrimination as a function of RAC size and complexity. To address this need, the database of 1,000 shoes will be doubled. If past RAC frequency is a good indicator

of future counts, then the end-goal of a complete database of 2,000 shoes is likely to contain more than 100,000 RACs, allowing for a detailed statistical analysis of RAC co-occurrence in terms of shape category (lines/curves, circles, triangles, and irregular-shaped features) and position (θ, r, r_{norm}). In addition, the line/curve category is currently being subdivided into lines, simples curves, and compound curves to increase discriminating power. When complete, a detailed analysis of co-occurrence in position and shape will be provided, similar to that show in Fig. [5.1], along with recommendations regarding limits in discrimination as a function of RAC size, area, geometry, and complexity.

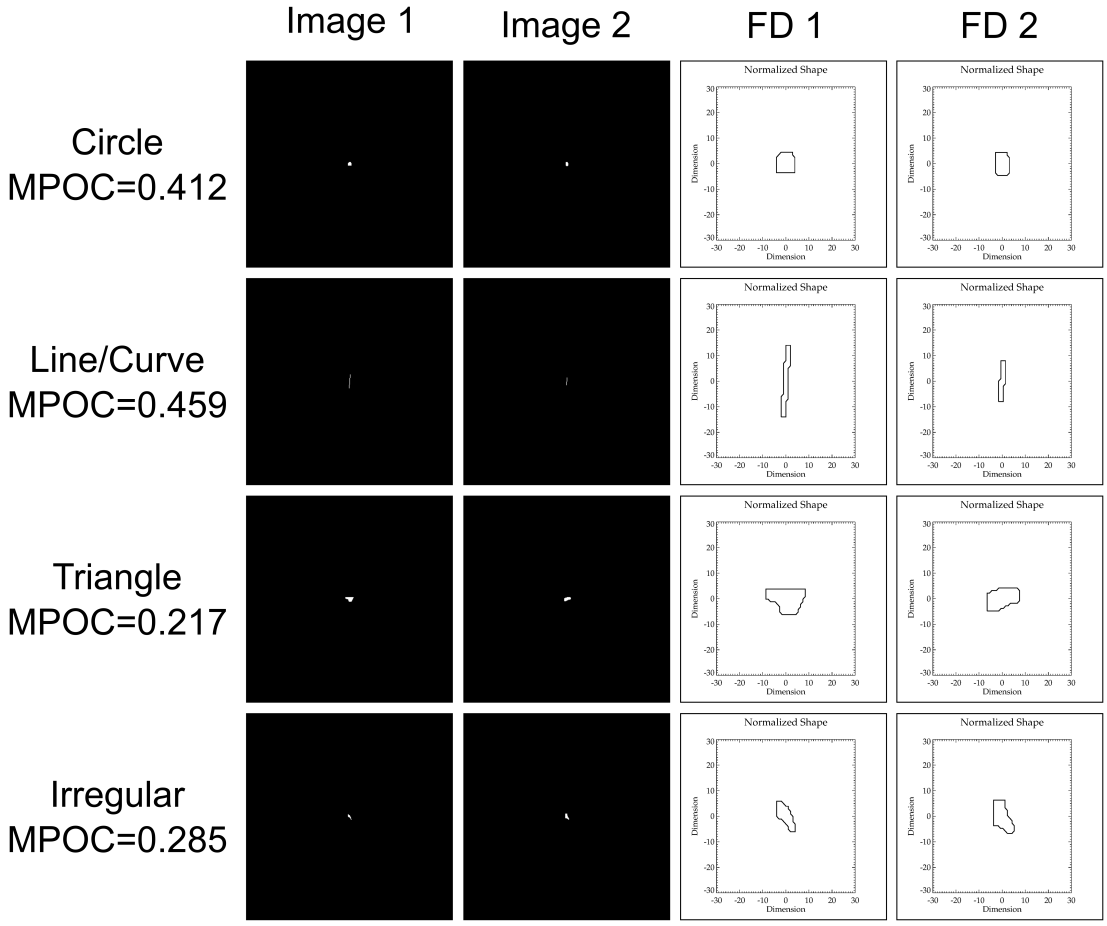


Figure 5.1: An illustration of the most similar RACs (*i.e.,* highest MPOC score) in each shape category within the bin located approximately 5mm from the lateral edge and 70mm from the heel of the shoe. The two RAC images, obtained through connected components, are displayed in the first two columns. In addition, the Fourier descriptors (FD) for both images are included for easier visualization (last two columns). Note that the most similar RACs are distinguishable based upon visual inspection and a correspondingly low MPOC score.

5.2 Phase Only Classification

5.2.1 Automated Classification

Based on the preliminary results for automated classification of outsole patterns for crime scene quality prints, Phase Only Correlation is a promising method. However, the crime-scene-like prints used for this phase of the research still contained more information than is often found at crime scenes (*i.e.,* they contain 80% or more of the outsole pattern based upon visual inspection). Thus, future imagery will be generated to increasingly stress the algorithm via fractional losses more comparable to that encountered in casework. In addition, 3-dimensional casts will be included to increase the variation in substrate/media input.

5.2.2 RAC Maps

Further research will be conducted regarding the reproducibility of RACs in crime scene impressions. Namely, the results for RAC map correlation will be broken down and analyzed in order to determine matching accuracy as a function of RAC number, size, shape and complexity. This data will offer some insight into the amount of information that is required for accurate identification of source (*i.e.,* are several small RACs equal to a few, large, complex features).

5.3 Addressing Remaining Research Needs

Although the current research attempted to address several forensic footwear research needs in order to increase efficiency and offer support to examiner conclusions (specifically regarding the interpretation of crime scene quality evidence), it was not possible to concentrate on all of the aforementioned recommended research areas. A suggested approach for addressing the remaining research topics is as follows:

- First, an assessment of the variability in examiner conclusions should be conducted. A similar study was previously conducted for fingerprint examiners (15). From the data collected, quantification of error rates for examiner conclusions was possible. Therefore, it may be possible to identify the error rates for footwear examinations, thus offering additional support to examiners in courtrooms when defending their conclusions.

- In addition, measures of uncertainty for examiner conclusions could be developed. While the current scale of conclusions allows an examiner some flexibility (*i.e.,* different levels of association between exclusion and identification), there is currently no method to report a quantitative level of uncertainty to accompany evidence comparisons. Although it is not clear how this might be achieved, research and discussion

should consider if and how uncertainty may benefit the footwear examiner moving forward.

- Further, an evaluation on intra- and inter-examiner variability in detecting and marking RACs should be pursued. Given that the automated comparison of RACs is contingent on the marking of features, it is important to understand the expected variability in identification and tracing of these accidentals. Based on an analysis of this variation, estimates of uncertainty may be developed for the results obtained through automated measures (*i.e.,* a confidence interval which accounts for inherent, and unavoidable, analyst variation in the absence of an automated extraction mechanism).

- Even better, the development of automated RAC extraction software should be explored. In theory, this algorithm could be used to search through an image for the presence of acquired features and segment the image to allow for evaluation of individual RACs or comparison of all features. This would offer tremendous increases in research efficiency, although it is clearly an extremely difficult image processing problem in its own right (*e.g.,* like finding the needle in the haystack).

- Lastly, estimates on the frequency of outsole information (make, model, size, etc.) should be gathered. This task is immensely difficult to tackle given the number of unknowns, such as number of shoes of a given type, how long shoes remain in the population before they are discarded, and the production of counterfeits for popular types of shoes. All of these factors complicate the estimation of frequency for outsole designs in the population, and therefore great attention will be required in order to account for or somehow overcome these unknown factors.

Bibliography

[1] Bodziak W. Footwear Impression Evidence: Detection, Recovery, and Examination Second Edition. CRC Press; 2000.

[2] Durose MR, Walsh KA, Burch AM. Census of Publicly Funded Forensic Crime Laboratories, 2009. U.S. Department of Justice, Bureau of Justice Statistics; 2012.

[3] Ashbaugh D. Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology. CRC; 1999.

[4] SWGTREAD. Range of Conclusions Standard for Footwear and Tire Impression Examination; 2013.

[5] SWGTREAD. Guide for Minimum Qualifications and Training for a Forensic Footwear and/or Tire Tread Examiner; 2006.

[6] Committee on Identifying the Needs of the Forensic Sciences Community NRC. Strengthening Forensic Science in the United States: A Path Forward. The National Academies; 2009.

[7] Defendant's Motion in Limine Regarding Expert Testimony on Footwear Comparisons; 2004.

[8] Champod C, Voisard R, Girod A. A statistical study of air bubbles on athletic shoe soles. Forensic Science International. 2000;109:105–123.

[9] Wilson HD. Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes. Journal of Forensic Identification. 2012;62(3):194–203.

[10] Cassidy MJ. Footwear Identification. Lightning Powder Company, Inc.; 1995.

[11] Stone RS. Footwear examinations: Mathematical probabilities of theoretical individual characteristics. Journal of Forensic Identification. 2006;56:577–599.

[12] Petraco NDK, Gambino C, Kubic TA, Olivio D, Petraco N. Statistical discrimination of footwear: a method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques. Journal of Forensic Sciences. 2010;55(1):34–41.

[13] Rencher AC. Methods of Multivariate Analysis. John Wiley & Sons, Inc.; 1995.

[14] Sheets HD, Gross S, Langenburg G, Bush PJ, Bush MA. Shape measurement tools in footwear analysis: A statistical investigation of accidental characteristics over time. Forensic Science International. 2013;232:84–91.

[15] Ulery BT, Hicklin RA, Buscaglia J, Roberts MA, Fienberg SE. Accuracy and reliability of forensic latent fingerprint decisions. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(19):7733–7738.