2023

Assessment of Acquisition, Retention, and Evolution of Randomly Acquired Characteristics with Wear

Nathaniel Weston
West Virginia University, nrw0010@mix.wvu.edu

Assessment of Acquisition, Retention, and Evolution of Randomly Acquired Characteristics with Wear

Nathaniel R. Weston, B.S.

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

Jacqueline Speir, Chair

Casper Venter

Tina Moroose

Matthew Marvin

Department of Forensic & Investigative Science

Morgantown, West Virginia
2023

Keywords: footwear, randomly acquired characteristics, continued wear, dissimilarity, acquisition, retention, geometric evolution, impression-wide similarity

ABSTRACT

Assessment of Acquisition, Retention, and Evolution of Randomly Acquired Characteristics with Wear

Nathaniel R. Weston, B.S.

The recovery of known-source shoes for the purpose of comparison to crime scene impressions often occurs with a temporal lag. During this passage of time, the outsole can be altered due to continued wear. These changes may impact forensically relevant characteristics of use known as randomly acquired characteristics (RACs). Continued wear may cause the formation of new RACs, cause RACs to undergo some degree of geometric change, and/or lead to the loss of formerly existing RACs. Consequently, the correspondence between a test impression from a known mated shoe with continued wear and a questioned impression with less wear deposited at a scene may diminish. In order to systematically examine these changes, the acquisition, retention, and geometric change of RACs were investigated as a function of step-count. Eight participants were recruited to walk in a new pair of Nike® Quest 3 shoes on a fixed route for 10 collections at 20K step-intervals up to a maximum of 200K total steps of use. Between each step-interval, high quality exemplars were created, digitized, and registered. The corresponding outsole was examined for the presence of RACs using oblique illumination and 4X magnification, and each feature's geometry was traced using Adobe Photoshop Elements 10®. In addition, one pair of shoes was remarked in replicate for each of the 10 step-intervals to inform uncertainty of detection and measurement, forming a quality control (QC) dataset.

RACs were tracked from the time of their acquisition through a total of 200K steps. The average gain in RAC count increased at a rate of 1.6X to 2.0X to 2.4X for 40K, 60K, and 80K steps of additional wear, respectively. However, uncertainty demonstrated that acquisition equal to or less than two RACs is likely attributed to variability in analyst detection. Likewise, the average loss, when normalized by the number of RACs present, increased at a rate of 1.6X to 2.0X to 2.3X for 40K, 60K and 80K steps of additional wear, respectively. Unlike the uncertainty reported for RAC gain, the quality assurance methodology used in this study effectively limited (nearly) all missed detections. As a result, variability in RAC loss as a function of missed analyst detection cannot be quantified, but is believed to be negligible.

For the purpose of detecting the spatial distribution of RAC acquisition and loss, 20 heatmaps were populated as a function of cumulative wear. Comparison of the spatial distributions of RACs for both acquisition and loss suggests possible clustering, with a slight increase in the number of RACs gained in the medial edge of the toe, and lost in the lateral toe and heel, for later step-intervals.

With the aim of quantifying the geometric change of RACs through continued wear and cumulative use, probability density functions and receiver operating characteristic (ROC) curves were generated to compare the numerical similarity — based on average Hausdorff distance (HA) — between known mated (KM) RACs that differ with wear, versus all other known non-mated (KNM) RACs in the dataset. If uncertainty in the method and the ability of the analyst to repeatedly trace RACs known to be identical is ignored, then across all step-interval differences (up to 180K) and cumulative wear (up to 200K of total steps), there was a 0.86 probability that mated RACs which

persisted through continued wear were more numerically similar to themselves than non-mates. Moreover, for 20K steps of additional wear for a lightly-worn shoe versus a well-worn shoe, this probability was relatively constant (0.95 - 0.90), failing to detect any differences over the life of the shoe (at least up until 200K steps of use). However, for increasing additional steps, the probability of a randomly selected RAC being deemed more similar to a worn version of itself as opposed to a non-mated RAC decreased by about 4% for every 20K steps of additional wear between 20K and 100K total steps. When considering uncertainty in the analyst's ability to repeatedly mark RACs known to be identical, 95% of the QC mated RACs had a numerical similarity score less than 1.12 HA units. At 20K steps of additional wear, nearly 75% of the KMs-with-wear distribution overlapped with the KM-QC scores indicating either no substantial change based on wear, or an inability for this method to detect any difference due to wear. Another 10% of scores for KMs-with-wear were less than KNM scores, but 15% were actually greater (suggesting a numerical false exclusion). For all scores where a false exclusion between known mates is more numerically likely, the ratio of KMs-with-wear versus KNM scores was 19% for 20K additional steps. This value increased to 20%, 30% and 35% for 40K, 60K and 80K steps of additional wear, respectively.

In order to assess the effect of continued wear and cumulative age on the degree of global (or impression-wide) self-similarity, vectors of RAC area per spatial bin were transformed using principal component analysis (PCA), whereafter accuracy in group membership was evaluated using cosine similarity (CS) and linear discriminant analysis (LDA). When using CS as a classifier, source association accuracy decreased by approximately 6% for 20K - 100K steps of use (72%, 66%, 61%, 54%, 50%). This matches intuition in that a greater degree of continued wear will result in lower self-similarity. Lower classification accuracy was observed for earlier cumulative step-intervals, until at least six RACs were acquired to generate sufficient differentiation. These findings were corroborated with a hold-one-out validation (HOOV) using LDA as a classifier, such that self-association accuracy was below 75% until 160K steps were accrued. Additionally, little-to-no increase in classification accuracy was observed when a participant's left/right shoe designation was ignored. In other words, RAC maps for a left or right shoe worn by a single participant were not more similar than other close non-mates in this dataset. For the wear-induced deviation in self-similarity using CS and LDA, at least 8% and 23% (respectively) could be attributed to the combined method and/or analyst's variation in marking. Thus, greater global self-similarity than that reported may exist for moderately-worn shoes (perhaps as high as 80% for 20K steps of wear accumulated after 100K and measured using CS, and >>90% after 160K total steps when using LDA).

In summary, this research quantified RAC (i.) acquisition, (ii.) loss, (iii.) spatial distribution, and (iv.) numerical geometric similarity — as well as (v.) global impression-wide self-similarity — for a specific make/model of shoe, tracked over 200K steps of total wear, as worn by an opportunistic set of participants, walking a single/specific terrain. Results indicate a relatively equal rate of acquisition and normalized loss with increasing wear, such that loss can dominate for shoes with a greater number of initial RACs while acquisition can dominate for lightly-worn shoes with few RACs. In addition, there is evidence that approximately 15% of KMs-with-wear will start to exhibit changes in RAC geometry with 20K additional steps of wear that numerically cause them appear more similar to KNMs. As the step-interval difference increases to a total of 80K, this rises to approximately 25%. Finally, global self-similarity after 20K steps is very poor for new/lightly-worn shoes (owning to the lack of features on which to form a valid association), but this metric is generally greater than 80% for moderately-to-well-worn shoes (100K for CS and 160K for LDA).

Naturally, as the number of steps walked between the collection of two test-impressions increases, similarity decreases. If the outcomes observed in this study persist with other datasets that include other shoe makes/models, participants, and terrains, then the results frame the range of variation in RAC number and self-similarity with less-worn mates that a forensic footwear examiner might reasonably expect between test impressions that differ in total wear.

Contents

1. Introduction

1.1 Statement of the Problem

A variety of forensic evidence can be collected for the eventual use of generating a linkage between a suspect and a crime scene. Footwear impressions are one such example, with the ability to provide a source association between questioned impressions collected at a scene and known test impressions collected from the outsole of a suspect's shoe. However, knowing and recreating the exact circumstances of how questioned impressions are deposited is nearly impossible. In addition, impressions can vary greatly in quality and sufficiency. As a result, part of the comparison process requires a footwear examiner to understand and account for possible variations and explanations for dissimilarities between questioned and known impressions. One possible factor that can contribute to this variation is the accumulation, loss of, and change in 'characteristics of use,' which can evolve as a function of continued wear of an outsole between the deposition of its impression at a scene and the ultimate submission of the shoe to a laboratory for analysis.

If shoes are worn following the deposition of an impression at a crime scene, impressions collected later will presumably differ to some degree from those deposited at the scene. However, the degree of change that can be expected is currently unknown. There is evidence to indicate that a shoe may change very little and look the same, that it may acquire new wear patterns and Randomly Acquired Characteristics (RACs), and/or that pre-existing RACs may either be lost or undergo a degree of geometric change. Since these possibilities are influenced by continued wear (assuming the absence of purposeful damage on the part of the wearer), the nature and degree of use is expected to lead to variations and dissimilarities.

There are various factors that may contribute to the change of RACs, such as steps taken, terrain walked upon, activities performed, and participant attributes. Unfortunately, the time, expense, and participant-commitment required in order to investigate all factors is highly prohibitive, making it difficult to report an expected range of variation that may be possible under a specific set of conditions. It is even more challenging to extrapolate past observations to new situations. One major obstacle is that most studies are metered by 'time' rather than 'steps' taken. Time since the deposition of an impression does not translate directly to a specified degree of continued wear (*e.g.,* a pair of shoes can be in a person's possession for weeks, months, or years, and only be periodically used). It is expected that increasing continued wear positively correlates with the degree of change observed. However, this magnitude of change is not well characterized as a function of step count in the existing literature. Furthermore, none of these changes have been investigated as a function of cumulative wear. For example, RACs may change more in the first 10 - 20 miles of wear when a shoe is brand new than the last 10 - 20 miles of wear when the shoe is near the end of its useful life and near being retired by its user. Moreover, no pre-existing research investigates dissimilarities in RAC geometry attributed to continued wear. The aim of this study is to provide a baseline understanding of the degree of change in individualizing features on footwear which can be expected and attributed to additional and cumulative wear.

1.2 Literature Review

1.2.1 Forensic Footwear Examinations

Footwear examinations typically involve several tasks beginning with an examiner's determination of the quality, clarity, and rarity of class, subclass and individualizing characteristics present on a questioned impression. Broadly, class characteristics (*e.g.,* shoe size and brand) are evaluated in order to exclude or include a set of known exemplars as a possible source of the unknown impression. Thereafter, wear and RACs are identified and compared for correspondence with a test impression, including their spatial location and physical geometry. Based on the totality of information present, the degree of similarity observed, and the degree to which dissimilarities can be explained, the examiner forms an opinion [1]. The laboratory scale used to report this opinion/conclusion can vary by agency, but typically includes associations of class, as well as extremely strong source-associations (such as an identification), and disassociations (such as an exclusion). Continued wear of a pair of shoes beyond deposition of its impression at a scene has the potential to influence the observed agreements and disagreements between mated and non-mated pairs, thereby potentially influencing examiner conclusions.

1.2.2 Characteristics Of Use

General wear is the degradation of the tread pattern on an outsole. Specific wear describes the pattern, position and depth of tread loss. Although specific wear is considered more discriminating than general wear, it is not widely accepted as a means for source-association [1]. Conversely, the location and shape of a RAC, depending on its clarity and complexity, can be used as the basis for identifications. All three types of characteristics of use (general wear, specific wear, and RACs) can vary between replicate impressions collected at different points in time (assuming the passage of time is synonymous with continued wear). However, the rate at which these features vary is not believed to be uniform across an entire outsole. For example, in a study of 650 pairs of shoes, the heels of the outsoles were analyzed for wear patterns and it was reported that the front edge of the heel yielded the heaviest concentration of erosion [2]. This was believed to occur due to the heel striking the ground first with greater pressure, thus increasing the likelihood of loss. Another study, this time of 380 shoes, sought to investigate RAC independence in terms of location, shape type, orientation, and outsole pattern [3]. It was found that these features are not independent of one another, and additionally, that the RACs were not independent of the shoe sole pattern. This suggests a spatial dependency on the external factors influencing acquisition and/or loss. In a comparable study of 1,300 outsoles and more than 72,300 randomly acquired characteristics, RAC frequency, when normalized for tread in contact with the ground, was lower than expected on the edge of the heel, but greater than expected on the medial ball of the toe [4]. Thus, there is preliminary evidence to indicate that RACs may not develop and/or persist uniformly across an entire outsole. Instead, RACs vary as a function of the way the wearer walks, and therefore may not change (be acquired, lost or evolve in geometry) uniformly either. As a consequence, the degree of dissimilarity in characteristics of use may vary spatially across an outsole, requiring a need for not only an estimation of rates, but their spatial distribution as well.

1.2.3 Acquisition

RAC acquisition is believed to be influenced by a variety of factors, including pressure, gait, foot pronation/supination, and activity. This is illustrated in a study of six participants (three men and three women) who hiked Mount Bierstadt while wearing a new pair of Altitude II hiking boots. Each participant traversed approximately 7 miles (round-trip) and changed boots between the ascent and descent (resulting in 12 outsoles for analysis) [5]. After use, the shoe was divided into 38 tread elements, and on average, at least 1 RAC was detected on 44% of the tread elements on each men's boot and on 33% of the tread elements on each women's boot. Moreover, when comparing all outsoles to each other, all were sufficiently different in terms of acquired features to allow for individualization [5]. Since hiking rigid terrain like a mountain may not be indicative of everyday-use, a separate study [6] investigated acquisition across multiple terrains more representative of typical wearer travel, such as dirt and gravel paths, concrete sidewalks, and cross-country courses. However, neither the percentage of tread elements with RACs present, nor the total distance traversed with the associated RAC count was reported, which unfortunately makes it difficult to compare and contrast findings. Moreover, this "more typical-use terrain" study involved a single participant, who ran in 39 different pairs of men's size 12 Adidas® Supernova Classic shoes across an 8 year time period. Each pair of shoes was worn for 341 miles on average (with a 55-mile minimum and 449-mile maximum) across the same three aforementioned terrains. The shoe was divided into 39 areas defined by tread elements and total RAC count was recorded in each area across all shoes once they were retired by the user. RAC presence ranged from 1 to 61 over all shoes, with an average of 11 for right shoes and 20 for left shoes [6]. This study design did not allow for insight into when within the continued wear interval the RACs were acquired (or if any RACs were lost altogether, potentially influencing the count reported) and therefore highlights a need for further research into these areas. In a third continued wear study [7], again with only one participant, five pairs of ladies Land's End® shoes were worn for 30 days each. A spatial grid (array) was superimposed over each outsole, and RAC counts were collected per spatial cell [7]. This 2-dimensional dataset was rasterized into a single vector, and data reduction was performed using principal component analysis (PCA). Only quantity and location of individual features were recorded, but a distance metric utilized to determine vector similarity in PC-space revealed a high true positive classification rate [7]. Thus, with 30 days of more "typical" wear (*e.g.,* not a fixed activity like running, such as in the Wilson [6] study), RAC acquisition and location were considered sufficiently useful to differentiate outsoles. Unfortunately, these results conflict with more recent investigations. In a San Jose State University study of 10 participants with different shoes and degrees of wear, 0 new identifiable RACs were detected after 45 days of use [8]. Similarly, Sheets *et al.* (2013) added artificial/simulated RACs to 11 pairs of shoes, which were observed over 7 weeks of use. Again, only 1 pair (of the 11) showed the creation of a new RAC [9]. Thus, either the rate of RAC acquisition is extremely variable, and/or time is a poor proxy for use. In summary, these studies show variations in terrain, activity, and the nature used to assess usage (two of the five summarized studies [5, 6] used actual miles (albeit step count can vary when a mile is traveled dependent on participant stride length/activity performed), while the remaining three used time as a proxy for wear). Thus, variation in study design may very well explain the wide variation in results, but only additional investigation, with well-known activities/terrains, and wear measured through pedometers in terms of step-count, can help elucidate these differences.

1.2.4 Retention

Retention of RACs, or their persistence following acquisition, can be fairly high. In the San Jose State University study, all RACs that were originally present during initial enrollment were retained after 45 days [8]. Likewise, when an unspecified number of 2-3 mm deep cuts and 1 mm deep holes were artificially created on outsoles by Sheet *et al.* (2013), only two of the 2-3 mm deep cuts on the heel were lost altogether by the end of the study (7 weeks) [9]. In contrast, Wyatt *et al.* (2005) [10] reported that a RAC was lost, on average, within approximately seven days of continuous wear. However, considerable variation was observed, including the highest observed loss of 5 RACs in a single day of wear, and the highest retention was all RACs (or 0 losses) over 30 days of wear. In Cassidy (1980), RAC retention also varied widely with continued wear. Heel impressions were collected from one participant for a single new pair of Cat's Paw boots prior to use, once for three consecutive days, and once weekly thereafter, to a maximum of 68 days [11]. A total of 36 RACs were identified across the study and were tracked from acquisition to loss. It was reported that 9 RACs were retained for 1 to 3 days, 1 RAC was retained for 13 days, 4 RACs were retained for 18 to 20 days, 3 RACs were retained for approximately 35 days, 7 RACs were retained for approximately 48 days, 6 RACs were retained for approximately 59 days, and 6 RACs were retained for approximately 65 to 68 days [11]. With such a large degree of variation in RAC retention, and with little being known regarding the spatial distribution of loss and retention, additional study is needed to determine the existence of possible trends, and if these vary spatially across an outsole.

1.2.5 Evolution

The evolution of RACs post-acquisition has not been widely researched either. In the Wyatt *et al.* (2005) [10] study, impressions from 54 participants were evaluated two months apart in order to determine possible source associations. The visual comparison was performed by three separate analysts, and overall, 22 out of the 54 sets of prints could be mated to their source after a maximum of two months of wear. For those that could not be confidently mated, the authors' attribute the false negative association to the loss of features, features that either changed or were covered up by new features, and an uncounted number of shoes that simply did not have RACs at the start of the experiment. In contrast, Petraco *et al.* (2010) [7] reported that mated shoes remained clustered together in PC-space through continued wear, and that the shoes could be differentiated with a much higher degree of accuracy ($\geq 77\%$) than reported in Wyatt *et al.* (2005) [10]. In Sheets *et al.* (2013) [9], PCA was similarly performed; however, the feature vectors were constructed with both wear and RACs, rather than just RACs as used by Petraco *et al.* (2010) [7]. Nonetheless, it was again reported that the degree of variability was such that the intra-shoe variation was lower than that of inter-shoe variation [9]. Thus, additional study is needed; there is little understanding of how RACs vary or erode in shape with wear, the range of variation that is reasonable if continued wear has occurred, and the range that is deemed unreasonable and should be considered a basis for exclusion.

1.3 Research Objectives

So long as accidental characteristics have sufficient correspondence (*e.g.,* definition, clarity, location, etc.) between a questioned and known impression, then a positive association can be made with a single RAC [1]. However, an examiner's confidence in making such a conclusion based on a single feature can vary with the geometric complexity of the RAC in question. For example, a small

pin prick is more likely to lead to chance associations than a long and jagged line, and therefore the weight afforded to each RAC is a function of its perceived rarity in shape. However, the degree of correspondence between a mated pair of questioned and known impressions is expected to vary. Impressions collected from crime scenes vary widely in medium, substrate, and possible background interference, so precise correspondence is not necessarily expected. Instead, an examiner considers what types of changes are reasonable given the scene and case circumstances. One such factor is continued wear. Although one can never know exactly how RACs are acquired, changed, and eroded under all circumstances, it is still useful to understand what types of changes are 'reasonable' under known and controlled circumstances. The purpose of this study was to quantify this space. More specifically, using a controlled shoe make and model (and therefore outsole material), well-defined step-intervals, activities and terrain, a baseline understanding of the type and rates of change in RAC count, spatial location, and geometry were investigated. The results reported in this study were used to achieve four main objectives, which were (1) determination of the occurrence of acquisition/loss as a function of accumulated and additional wear, (2) characterization of the spatial distribution of RAC acquisition, frequency, and retention, (3) investigation of the degree of evolution in RAC geometry with continued wear, and (4) examination of accuracy in source-associations through the use of numerical classifiers.

2. Methods

2.1 Data Collection

2.1.1 Participant Solicitation and Questionnaire

This project has an approved protocol (No. 2104288546) on file with the West Virginia University (WVU) Institutional Review Board (IRB). A total of eight participants were recruited via flyers and announcements (electronic and verbal), and enrolled following standard informed consent procedures. Participants were assigned alpha-numeric identifiers that were paired to all data. Participants completed a background survey at the onset of the study, reporting physical attributes (*e.g.,* height, weight, etc.), as well as their physical shoe size.

2.1.2 Shoe Selection

The Nike® Quest 3 shoe was selected for this study (Figure 2.1). The shoe was selected for a host of reasons, including: (i.) a rubber outsole (rather than microcellular), (ii.) a large surface area of tread in contact with the ground, and (iii.) a fixed number of tread elements regardless of shoe size.

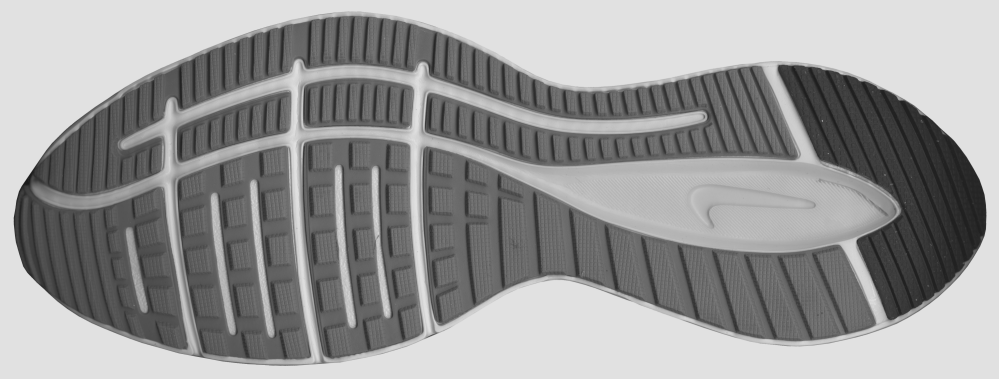

Figure 2.1: Digital scan of the Nike® Quest 3 shoe selected for this wear study.

2.1.3 Participant Shoe Wear/Transfer

A total of 10 collections took place at 20K step-intervals (∼10 miles) for a total of 200K steps (∼100 miles). Participants walked on the same asphalt-paved trail following the same route over the course of this study. After acquiring 20K steps, participants submitted shoes and the outsoles were scanned at a resolution of 600 PPI using an Epson Expression 1100XL Graphic Arts flatbed scanner. A pre-existing methodology was followed for Handiprint exemplar creation [12]. A high quality (HQ) Handiprint exemplar (2-3160 CSI Forensic Supply) of each outsole was created using a Zephyr brush (A-10200), Lightning Black Powder (1-4005 CSI Forensic Supply), and preserved using a clear polyester cover (2-3160 CSI Forensic Supply). These HQ exemplars were scanned at a resolution of 600 PPI using the aforementioned scanner. The outsole and exemplar image scans

were co-registered into a single coordinate system for data analysis (hereafter referred to as Eigen-registered). A graphical user interface (GUI) displayed both outsole and mated exemplar images collected at the zero-week step-interval for registration. For all step-intervals thereafter, exemplar images were registered using the post-registered (or Eigen) exemplar at the zero-week step-interval. To conduct Eigen-registration, sharp/reproducible geometric points on the image of the outsole and the test impression were selected pairwise between the original and most recent impression for a total of eight point selections. Chosen points varied between step-intervals due to accumulated wear. However, the selected points were dispersed as evenly as possible across the length and width of the shoe/Handiprint. Using a previously defined methodology [12], mathematical computations were applied to the most recent step-interval Handiprint image to align the toe-to-heel in a North-South orientation, and co-register all images using a single coordinate system. Using oblique illumination and a 4X magnifier, the outsoles were examined for the presence of RACs. A RAC needed to be visually apparent on both the outsole and the registered Handiprint exemplar to be marked. For any RAC that met this criteria, the RAC geometry was digitally traced on the Handiprint image using Adobe Photoshop Elements 10®, effectively creating a RAC map. RACs which were present on the outsole but absent on the Handiprint due to lack of tread element detail transfer were marked with a nine-pixel grayscale box at the approximated position of the missing RAC's centroid. The purpose of this was to isolate RACs that could not be used to assess changes in geometry (but were still valid to assess changes in acquisition and loss). Right shoe RAC maps were flipped to correspond with left shoe RACs maps, and both left and right RAC maps were normalized to a men's size 10.5 for spatial analysis of extracted features. A total of 1,435 RACs were extracted and mapped between Cartesian and polar coordinates and normalized with θ and r_{norm} using a previously described procedure [12]. Of these, 206 lacked geometric information (RACs present on the outsole but failing to reproduce on the associated Handiprint, but localized using a nine-pixel grayscale box) and an additional 21 were part of the zero-week step-interval. As a result, the dataset was reduced to 1,414 for acquisition/loss evaluation, and 1,208 for geometric analysis.

2.2 RAC Tracking

2.2.1 Patch Maps

Figure 2.2 demonstrates a patch map (bottom) generated from its mated marked Eigen-registered Handiprint (top). A patch map is an image that highlights the locations of marked RACs via square "patches." In other words, if a RAC was detected, then the patch image reproduced the details of the Handiprint tread features and the RAC (or RACs) at this location (in a 4.65 mm x 4.65 mm square). The Handiprint for a given step-interval was marked using the method described above (oblique illumination and 4X magnification). After marking, a patch-map was created for the Handiprint. For the subsequent step-interval, the process was repeated, but before concluding the marking process, the patch map from the previous step-interval was overlaid on the newly marked Handiprint from the next/following step-interval. If a RAC on the previous step-interval (referenced by a patch) was present on the next step-interval but was missed during the initial marking procedure, the "missed" feature was marked in order to avoid loss of RACs due to a missed-detection by the researcher (versus loss of RACs based on wear). If a patch did not contain a RAC (lost due to wear), no additional action was taken. After completing this step, the most recent step-interval marked Handiprint reflected all new RACs acquired by wear, and all pre-existing RACs not lost by wear, after controlling for (limiting) the possibility of inflated loss due to the analyst's missed detection of a feature.

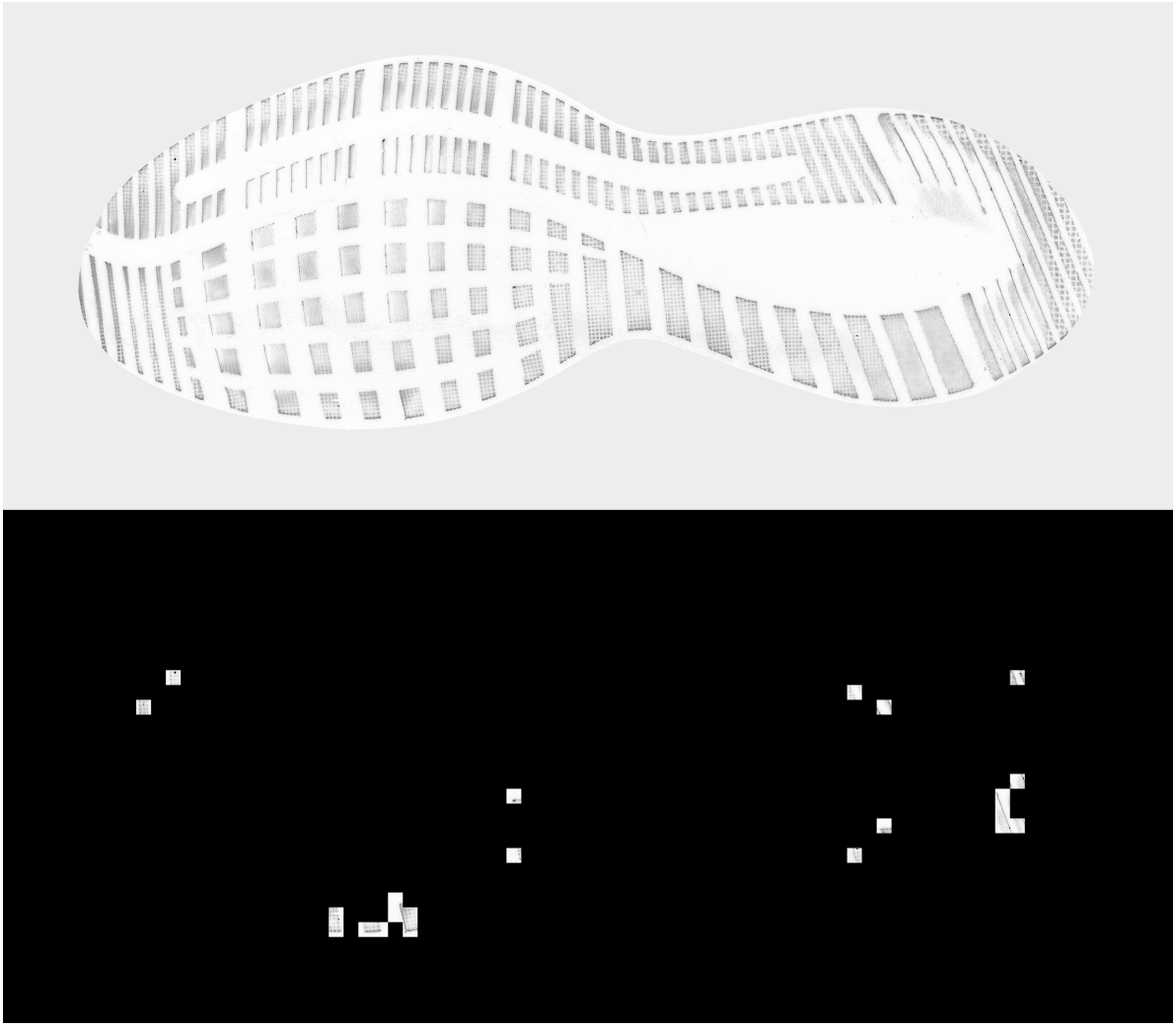

Figure 2.2: Marked Eigen-registered Handiprint (top) and its mated patch map (bottom). RACs present on the marked Eigen-registered Handiprint are displayed (with Handiprint tread detail) in patches where they were identified, thus creating a patch map.

2.2.2 RAC Renaming

After marking, RACs were automatically extracted and assigned any arbitrary numeral label based on scanning the image from left-to-right and top-to-bottom. As such, the numerical label assigned to a RAC that persisted between step-intervals varied as a function of the RACs acquired/persisting/lost in any given step-interval. This can be observed in Figure 2.3 image (A) and (B), which illustrate RAC acquisition and loss between consecutive step-intervals. For example, image A (earlier step-interval) has five marked RACs. Similarly, image B (subsequent step-interval) also has five marked RACs, but one lost and one gained when compared to the previous step-interval. In order to track a RAC that persisted across more than one step-interval, it was most convenient to resort and relabel so that a RAC's numerical label remained constant across all step-intervals, and when lost, the label was retired. This is illustrated by image B' which is the

renamed version of B in Figure 2.3. The RAC labeled 2 in the earlier step-interval (A) was lost, and is retired, and the RAC labeled 2 in the subsequent step-interval (B) is newly acquired, and is therefore labeled 6 after processing (B'), such that image A and image B' are used moving forward.

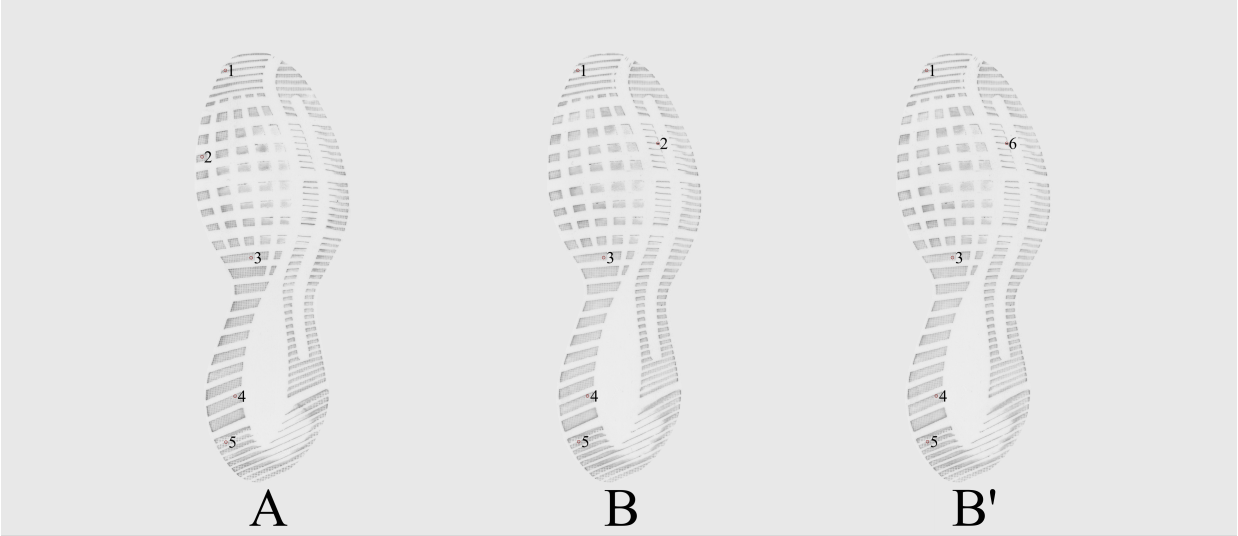

Figure 2.3: RAC renaming between step-intervals. Image A and B represent marked Eigen-registered Handiprints at step-intervals eight and nine, respectively, where one RAC was acquired and one RAC was lost in the latter. In both images, five RACs are emphasized by red circles and numbered as they would be during the first step of RAC extraction. Image B' represents how RACs in image B were renamed to simplify the tracking of RAC persistence, loss and acquisition.

In order to automate the RAC renaming process, each RAC's centroid was extracted and labeled by a triplet of θ, r, and r_{norm} values. Using this triplet, if a RAC on the outsole in a subsequent step-interval was in close proximity to a RAC on a previous step-interval, it became a candidate for renaming. If more than one RAC in the subsequent step-interval was in close proximity to a RAC on the previous step-interval, the nearest RAC was selected as the best candidate. However, this tie-breaking process, although numerically valid based on proximity, was supervised to ensure that (a.) the correct candidate was selected for renaming, and (b.) a candidate should have been renamed at all (a RAC could have been newly acquired in a location where a previous RAC was lost, and in this case, should receive a new and unique identifier). The decision criteria for the latter was supported by either a RAC's centroid shifting position to an unreasonable degree (defined as a difference in θ, r and/or r_{norm} greater than 25 units) and/or if a RAC's geometry changed significantly (*e.g.*, a linear RAC changed from an orientation of 0° - 180° to 45° - 225°, or a linear RAC was reduced to a pinprick). Finally, the supervised verification process was completed using an `R-Shiny` application (illustrated in Figure 2.4). RAC pairs nominated as candidates to be mates were displayed, along with the triplet used to denote the RAC's centroid, and a binary decision was made based on a visual comparison of the RAC-pair candidate to confirm or refute if the pair was truly mated. RACs with only one candidate were visually confirmed or denied and subsequently renamed, if necessary. If a RAC had multiple candidates, all were observed, the tie was manually broken, and the other RACs that were nominated were assigned a new label if they were not candidates for another RAC. If a RAC did not have any candidates, it was a new RAC and was therefore assigned a new label, continuing from where the last RAC in the previous step-interval left off.

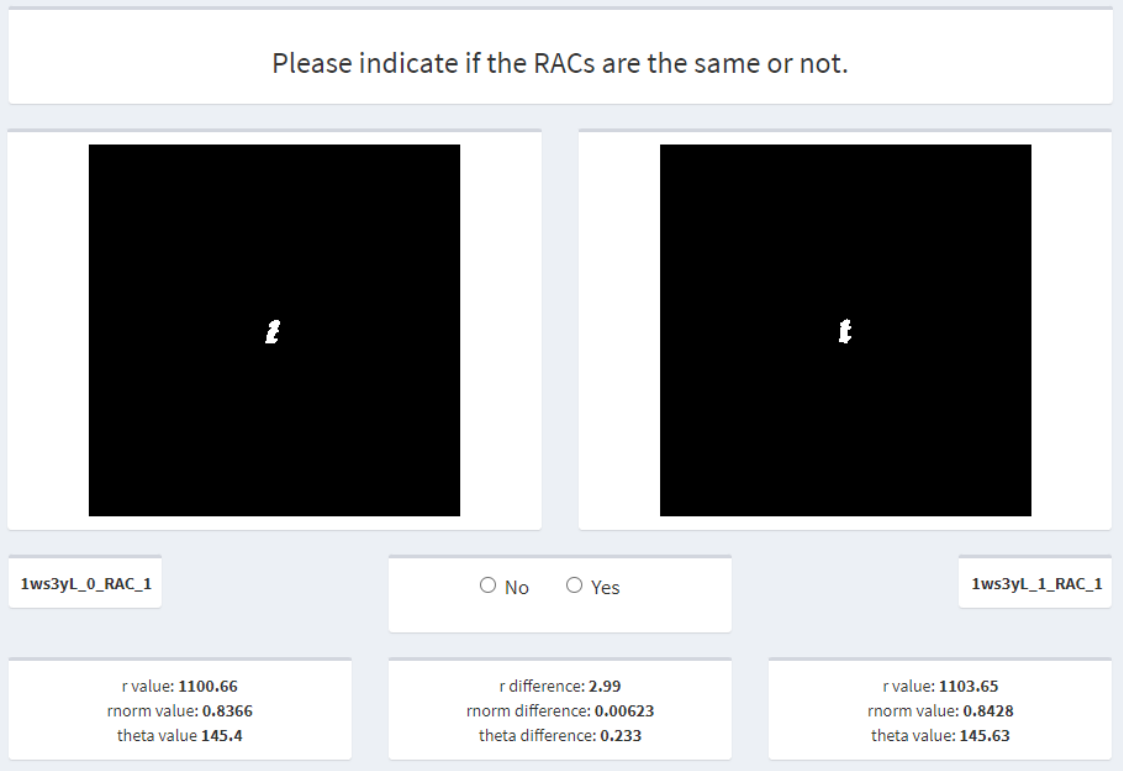

Figure 2.4: `R-Shiny` application created to visually confirm RAC renumbering. Candidates for RAC pairs were reactively displayed with their associated θ, r, and r_{norm} values and the differences between them. A binary "Yes" or "No" decision was made to reflect a mate or non-mate. For this example, "Yes" was selected due to general similarity in geometry and only minor differences in triplet values.

2.3 Quantifying RAC Changes with Wear

2.3.1 Acquisition/Retention of RACs

Bar plots were constructed reflecting average per-participant RAC acquisition and loss for step-interval differences (P = 20K - 80K) and all cumulative step-intervals (F). In addition, loss was normalized based upon the RAC count in the previous step-interval to eliminate RAC presence in a preceding step-interval as a confounding factor for RAC loss in a subsequent step-interval.

2.3.2 Spatial Distribution of RACs

The acquisition and retention of RACs were represented spatially as heatmaps. These heatmaps were constructed for each of the cumulative step-intervals, resulting in 10 heatmaps describing spatial acquisition, and 10 heatmaps describing spatial loss.

2.3.3 RAC Geometric Evolution

RAC Preprocessing

All non-grayscale/binary RAC images were convolved with a Prewitt filter (PF) to detect and extract RAC boundary information as shown in Equations 2.1 - 2.3 [13]. The horizontal (G_X) and vertical (G_Y) Prewitt operators were applied to each pixel ($F_{s,t}$) of the binary RAC images, resulting in two convolved images. Next, the gradient magnitude (G_{st}) was computed, resulting in a composite image.

$$G_X = F_{s+1,t+1} + F_{s+1,t} + F_{s+1,t-1} - (F_{s-1,t+1} + F_{s-1,t} + F_{s-1,t-1}) \tag{2.1}$$

$$G_Y = F_{s-1,t-1} + F_{s,t-1} + F_{s+1,t-1} - (F_{s-1,t+1} + F_{s,t+1} + F_{s+1,t+1}) \tag{2.2}$$

$$G_{st} = \sqrt{G_X^2 + G_Y^2} \tag{2.3}$$

Fourier descriptors (FDs) were generated from the Prewitt RAC perimeter images using a previously defined methodology [14] to represent RAC geometries. Two example image sets are illustrated in Figure 2.5, where the binary RAC image is on the left and the associated PF-FD is on the right for each pair.

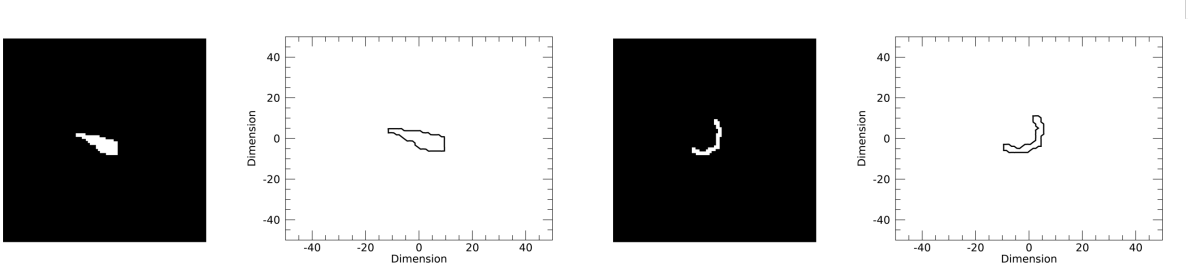

Figure 2.5: Binary RAC images (left-most in each pair) and associated Prewitt Fourier descriptor (right-most in each pair).

Similarity Metrics

A total of seven similarity metrics were computed. Percent area overlap (%A) was computed on binary RAC images, while the remaining six were computed using the Prewitt Fourier descriptor as input; these included absolute value Euclidean distance (AE) [15], total Euclidean distance (TE) [15], maximum Hausdorff distance (HM) [16], average Hausdorff distance (HA) [16], and matched filter (MF) [17], and Hermitian angle (HRA) [15]. Of those computed on Fourier descriptors, only average Hausdorff distance is explained in greater detail below; the reader is referred to primary resources [15 - 17] for details concerning all others.

Average Hausdorff distance is a computation of Euclidean distances (EDs) between each point of a boundary for one RAC compared to all points on another RAC, the minimum distance is selected, whereafter the average of these distances is retained as shown in Equations 2.4 - 2.6 for boundary

points a and b from RACs A and B [16], $\overrightarrow{y}$ and $\overrightarrow{z}$ correspond to the real and imaginary portions of the PF-FDs of RACs, and R corresponds to the number of elements in the vector ($R = 350$). This is a non-scaled similarity metric, and a lower HA indicates a more similar RAC pair.

$$\mathrm{HA}(A, B) = \max(\mathrm{ha}(A, B), \mathrm{ha}(B, A)) \tag{2.4}$$

$$\text{where ha}(A, B) = \frac{\sum_{a \epsilon A}^{R} (\min_{b \epsilon B}(\mathrm{ED}\,\{a, b\}))}{R} \tag{2.5}$$

$$\text{where ED} = \sqrt{(\overrightarrow{y}_1 - \overrightarrow{y}_2)^2 + (\overrightarrow{z}_1 - \overrightarrow{z}_2)^2} \tag{2.6}$$

Each of the seven similarity metrics were evaluated by examining receiver operating characteristic (ROC) curves, which were generated using the `pROC` package in `R` [18]. Using this package, the functions `auc`, `power.roc.test`, `ci`, and `roc.test` were used to compute the power of the ROC curve, the area under the curve (AUC), the confidence interval of the AUC (C-I), and to test for statistically significant differences between classifiers using the DeLong method [19], respectively. The optimal similarity metric among the seven compared was selected, ROC curves were generated using this metric for all step-interval differences and all cumulative step-intervals. Thereafter, power, AUCs, C-Is, and significance testing between groups was computed as previously described.

2.4 Quantifying Outsole Evolution with Wear

2.4.1 Data Preprocessing

Global evolution of outsoles with continued wear was evaluated using cosine similarity (CS) and linear discriminant analysis (LDA) after data reduction using principal component analysis (PCA). Feature vectors of RAC area per normalized spatial bin were created for each participant (c) and for each step-interval, and vectors with at least one observation in any cell were concatenated together to form an l (step-interval) x q (spatial bins) matrix (or a 10 x 1145 matrix) describing each participant's outsole (left and right treated independently). An example of this matrix can be seen in Figure 2.6.

$$\begin{bmatrix} o_{1_1} & o_{2_1} & \cdots & o_{l_1} \\ o_{1_2} & o_{2_2} & \cdots & o_{l_2} \\ \vdots & \vdots & \ddots & \vdots \\ o_{1_q} & o_{2_q} & \cdots & o_{l_q} \end{bmatrix}$$

Figure 2.6: Example participant matrix. Total RAC area (o) by position q (of 1145 bins) in the feature vector at step-interval l.

2.4.2 Principle Component Analysis (PCA)

All participant matrices were concatenated together and decomposed into associated eigenvectors and eigenvalues. Using the eigenvalues, estimates of retained variance after data reduction were computed, and the eigenvectors were used to describe participant information post-data reduction for further data analysis.

2.4.3 Cosine Similarity (CS)

Cosine similarity was computed between all step-intervals and all participants using the PCA transformed data as seen in Equation 2.7, where $\overrightarrow{x}_{c_l}$ is a vector of total RAC area observations in PC-space for a fixed-participant (c) and a fixed step-interval (l), $\overrightarrow{a}_{c_l}$ is a non-mated (differing participant or same participant differing step-interval) vector of total RAC area observations in PC-space for a fixed step-interval, and $|\overrightarrow{x}_{c_l}|$ and $|\overrightarrow{a}_{c_l}|$ are the respective magnitudes. The resulting angle is inversely related to vector similarity. Confusion matrices were generated for each participant and each step-interval, whereafter the classification accuracy was computed.

$$\Theta = cos^{(-1)}\left(\left[\frac{\overrightarrow{x}_{c_l}}{|\overrightarrow{x}_{c_l}|}\right]^T\left[\frac{\overrightarrow{a}_{c_l}}{|\overrightarrow{a}_{c_l}|}\right]\right) \tag{2.7}$$

2.4.4 Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) was performed on the PCA transformed data using Equations 2.8 and 2.9, where K is the total number of shoes, c is a fixed participant, n is the number of step-intervals per participant that contained RACs (*i.e.*, non-zero vectors), $\mathbf{S_c}$ is the covariance matrix for participant c, $\overrightarrow{x}_{c_l}$ is a vector of total RAC area observations in PCA space for a fixed-participant (c) and a fixed step-interval (l), $\overrightarrow{\bar{c}}$ is a vector of means for participant c across the 10 step-intervals (or all step-intervals with non-zero vectors), $\mathbf{S^{-1}_{pooled}}$ is the inverse of the pooled covariance matrix, and LD_{lc} is the LDA distance. Analogous to cosine similarity, the resulting distance is inversely related to a vector's similarity to class centroid.

$$\mathbf{S_{pooled}} = \frac{\sum_{c=1}^{K}(n_c - 1)\mathbf{S_c}}{N - K} \tag{2.8}$$

$$LD_{lc} = (\overrightarrow{x}_{c_l} - \overrightarrow{\bar{c}})^T\mathbf{S^{-1}_{pooled}}(\overrightarrow{x}_{c_l} - \overrightarrow{\bar{c}}) \tag{2.9}$$

A hold-one-out validation (HOOV) was conducted, and confusion matrices were generated for each participant and each step-interval, whereafter the classification accuracy was computed.

2.5 Uncertainty of Measurement

2.5.1 Data Collection

One pair of shoes was randomly selected without replacement for remarking in each step-interval throughout the study, generating 20 repeated markings (2 shoes x 10 step-intervals = 20 repeated markings). Since two repeats (and two originals) had no identified RACs, the quality control (QC) dataset was reduced from 20 to 18 total comparisons between 36 impressions. In hindsight, the two shoes without RACs that were randomly selected for inclusion in the quality control study should have been replaced with a valid sample. Each pair of shoes was remarked at least once. Due to the shoes being identifiably different in size and color, the examinations were not fully blinded; however, repeated markings were performed after all shoes within the step-interval were marked and at a minimum of one day apart from when the shoe was originally marked.

2.5.2 Global Correlation and Detection

A 2-dimensional correlation coefficient was used to assess global similarity between repeated markings. For an M x N image (8691 x 8691), the global correlation between mated RAC images was computed according to Equations 2.10 - 2.12 [20], where $\mathbf{f_1}$ defines the original RAC map (denoted as $\hat{\mathbf{f}}_\mathbf{1}$ after normalization), $\mathbf{f_2}$ defines the mated repeat RAC map (denoted as $\hat{\mathbf{f}}_\mathbf{2}$ after normalization), and s and t define pixel coordinates in respective rows and columns. Missed analyst's detections were tabulated to report average change in detection between mated QC markings.

$$\hat{\mathbf{f}_\mathbf{1}}(s,t) = [\mathbf{f_1}(s,t) - mean(\mathbf{f_1})]/std(\mathbf{f_1}) \tag{2.10}$$

$$\hat{\mathbf{f}_\mathbf{2}}(s,t) = [\mathbf{f_2}(s,t) - mean(\mathbf{f_2})]/std(\mathbf{f_2}) \tag{2.11}$$

$$r = \frac{1}{MN}\sum_{s=1}^{M}\sum_{t=1}^{N}\hat{\mathbf{f}_\mathbf{1}}(s,t)\,\hat{\mathbf{f}_\mathbf{2}}(s,t) \tag{2.12}$$

2.5.3 Geometric Similarity Score Distributions

Variability in RAC tracing was investigated for repeated markings using probability density functions (PDFs) of scores computed using the selected optimal similarity metric (previously explained) to describe the probability of a score if sampled from known mated (KM), known non-mated (KNM), and quality control (QC) pairs. The `density` package was used in `R` to compute kernel density estimates (KDEs) of similarity scores using a Gaussian kernel with a width (standard deviation) equal to Silverman's rule [21] for step-interval differences P = 20K - 80K. The dissimilarity between RAC pairs which was not attributed to the analyst's variation in marking was determined by the percentage of KM and KNM densities above the similarity score which captured 95% of the QC density. The KM and KNM distributions were integrated between this QC threshold and where density heights were equal (or a likelihood ratio of 1) to quantify the change in distributions which can be attributed to continued wear.

2.5.4 Quantifying Outsole Evolution with Wear

Each of the 36 (18 repeated) quality control RAC area vectors were projected into the same PC-space as the full dataset. These impressions comprise the quality control dataset, which are based on 18 individuals sets, where a set is defined as two mated feature vectors based on repeated markings of the same outsole in a fixed step-interval. Of the 18 sets, three had another mate with additional wear. For example, left outsole E (or EL) was marked once in step-interval 60K, comprising one set, and again in step-interval 160K, again comprising one set; therefore, both sets have a known mated set with more and less wear in the 160K and 60K step-intervals, respectively.

Cosine Similarity

Cosine similarity was computed pairwise for all 36 quality control vectors (18 original markings with 18 repeats). Vectors were classified based on the smallest computed angle. Theoretical perfect accuracy and discrimination potential based on RACs would result in angles of 0^o *within* each set, and

larger angles *between* sets. Deviations from this ideal performance are based on inconclusive/non-discriminating vectors, and/or analyst variation in marking, forming a basis to estimate uncertainty in the use of cosine similarity to classify vectors.

As previously mentioned, vectors were mated based on the minimum computed angle. However, this meant that mated and non-mated vectors with differing wear could be considered most similar to each other. In other words, all pairwise comparisons regardless of step-interval were computed or (36 x 35)/2 generating 630 angles. For the three repeated QC datasets previously mentioned, this meant that it was possible for a mated outsole in a QC dataset to match more closely to itself with additional wear, and still be considered an accurate classification. Similarly, any known non-mated vectors with differing accumulated wear could also be deemed most similar to each other.

Linear Discriminant Analysis

Each QC set (comprised of two PC-transformed vectors) was used to generate a QC-step-interval centroid. Using the inverse of the pooled covariance matrix generated by the wear data, each of the 36 vectors were held-out and classified using LDA. In other words, scaling was accomplished by a fixed/constant inverse of a pooled covariance matrix, with only distances between centroids or mated QC vectors allowing for variation.

Ideal performance (no variation in marking and high discrimination potential of input vectors) should result in zero misclassifications. Correct classification was defined as any vector (within or between) QC sets that was classified to itself, regardless of additional wear. Again, deviations from ideal performance were used to form a basis to estimate uncertainty in the use of LDA to classify vectors in this dataset.

3. Results and Discussion

3.1 Data Collection

3.1.1 Study Step Acquisition

A total of eight participants were recruited and walked a total of 200,000 steps during this study. Each participant wore a fixed pedometer and traversed the same route 10 times. See appendix section 5.1 for pedometer reliability testing and section 5.2 for self-reported step count per-cumulative step-interval of collection.

3.1.2 Participant Physical Attributes

Each participant completed a questionnaire at the onset of the study describing physical attributes. Participant responses were mated to assigned alpha-numeric identifiers and stored in a single-copy encrypted file on a secure network drive. De-identified summary information is provided in Table 5.3 in Appendix section 5.3.

3.1.3 Data Acquisition

Over the course of this study, a total of 1,435 RACs were identified and marked. Of these, 206 lacked geometric information (RACs present an outsole but failing to reproduce on the associated Handiprint, but localized using a nine-pixel grayscale box) and an additional 21 were part of the zero-week step-interval. As a result, the dataset was reduced to 1,414 for acquisition/loss evaluation, and 1,208 for geometric analysis.

3.2 Acquisition/Retention of RACs

In order to illustrate the total number of differences tracked over the life of the shoes with continued wear, bar plots were constructed reflecting RAC acquisition and lack of retention (or loss). The goal of constructing this visual was two-pronged. First, does the degree of continued wear (or step-interval difference) affect the number of acquisitions or losses? For example, does a shoe with 20K steps of additional wear (or a 20K step-interval difference) have a different number of acquisitions and/or losses than a shoe with 40K steps of additional wear (or a 40K step-interval difference)? Second, does the shoe's age (or cumulative wear) affect the number of acquisitions or losses? For example, does a heavily-worn shoe have a higher expected number of acquisitions and/or losses than a lightly-worn shoe or vice versa?

Figure 3.1 is a plot of the average number (and standard error) of RAC acquisitions (for n = 16 shoes) for a fixed step-interval difference (P = 20K, 40K, 60K and 80K) as a function of shoe total usage. See Appendix section 5.4 for data per participant per step-interval. Although statistical testing was not conducted, there appears to be a positive correlation or general upward trend in acquisitions with continued wear. Additionally, acquisition appears to be bimodal, with the two

distributions separating at 100K steps of cumulative wear. A visual inspection was performed on outsole scans for the 100K and 120K step-intervals to investigate whether or not there was a noticeable change that would explain this occurrence (*e.g.*, an extreme increase of general wear). However, no remarkable differences between these step-intervals were observed during this comparison, so a presumptive suggestion for the bimodal data is not yet known.

If the data is considered bimodal, then the standard error bars for the earliest distribution do not overlap with those in the later distribution. This could indicate a difference in physical phenomenology, or a change in the marking criteria of the researcher as a function of increasing experience. Assuming the former, then this could mean that rates of acquisition are dependent upon cumulative step-interval. In other words, a shoe in its early stages of use (*e.g.,* 0K - 20K steps) would be expected to have a lower rate of acquisition than in its later stages of use (*e.g.,* 180K - 200K steps). This difference could be attributed to the rigidity of the shoe's outsole in its early stages of life which likely decreases the probability of gaining new RACs. However, the results should be repeated and validated since researcher criteria in marking cannot be eliminated as a confounding variable in this study.

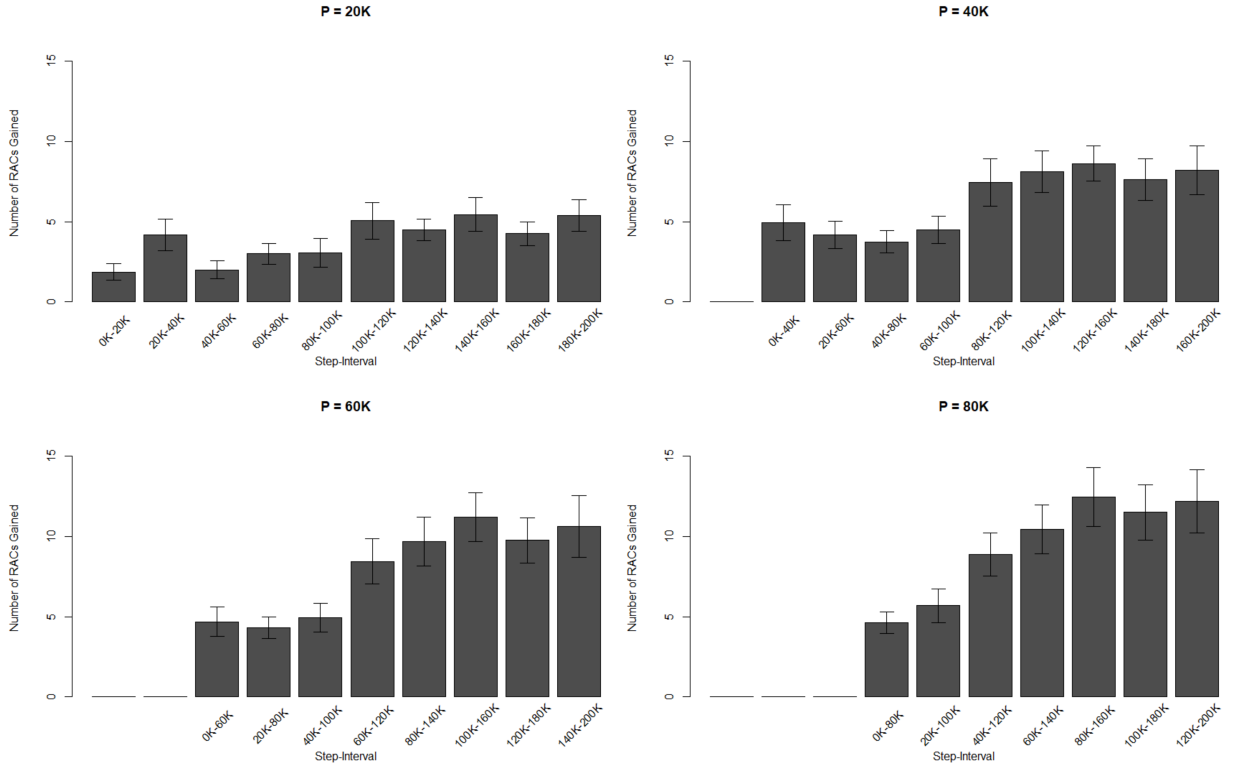

Figure 3.1: Bar plots (and standard error) of RAC gain (for n = 16 shoes) for a fixed step-interval difference (P = 20K, 40K, 60K and 80K) as a function of shoe total usage. The x-axis reports the specific step-interval under investigation, while the y-axis reports the average number of RACs acquired for all participants. For example, the 20K - 40K observation in the P = 20K step-interval difference plot corresponds to the number of RACs gained on average when comparing the 20K step-interval to the 40K step-interval.

Figure 3.2 is a plot of the average number (and standard error) of RAC losses (for n = 16 shoes) for a fixed step-interval difference (P = 20K, 40K, 60K and 80K) as a function of shoe total usage. Since these plots depict loss prior to normalization by the number of RACs present in the proceeding step-interval, they are provided in the interest of full transparency, but only allow for limited interpretation. In general, shorter step-intervals of continued use (P = 20K) (regardless of total usage or age) resulted in an average of approximately two lost RACs. As the step-interval before the deposition of a questioned impression and collection of a test impression of the same shoe increases up to a step-interval difference of P = 80K, the average loss increases to approximately four RACs. This is of interest because an additional 60K steps of use, without considering RAC count, resulted in only two more RACs being lost on average, and therefore warranted further investigation into RAC loss post-normalization with RAC count.

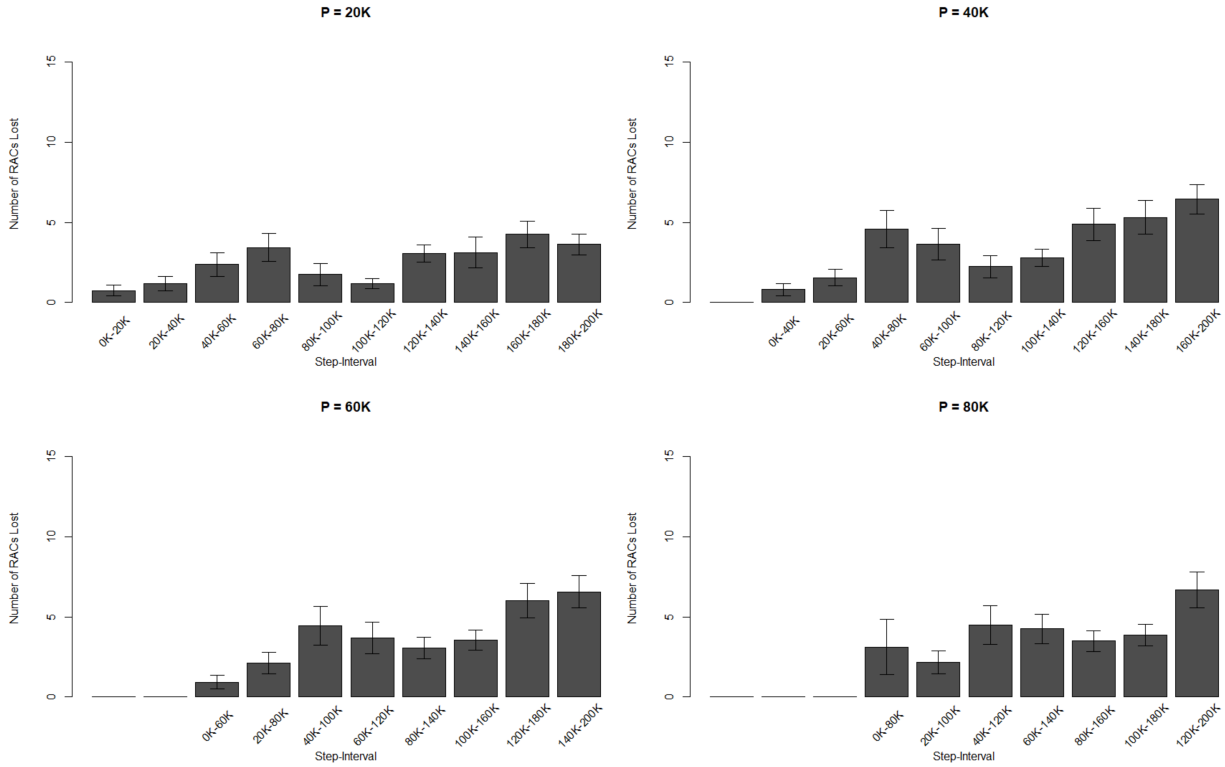

Figure 3.2: Bar plots (and standard error) of RAC loss (for n = 16 shoes) for a fixed step-interval difference (P = 20K, 40K, 60K and 80K) as a function of shoe total usage. The x-axis reports the specific step-interval under investigation, while the y-axis reports the average number of RACs lost for all participants. For example, the 20K - 40K observation in the P = 20K step-interval difference plot corresponds to the number of RACs lost on average when comparing the 20K step-interval to the 40K step-interval.

Since a RAC must be acquired before it can be lost, it is necessary to account for RAC quantity present at the beginning of a step-interval (or the number of RACs that have the potential to be lost). This is shown in Figure 3.3, which plots average RAC loss (and standard error) per step-interval difference (P = 20K, 40K, 60K and 80K) as a percentage of the number of RACs present. The data is the same as that shown in Figure 3.2, but transformed such that the reported loss

is a function of the number of RACs present at the start of a step-interval (first value in x-axis label). Within a plot for a specific P = XK, the relatively uniform height of the bars (versus the sinusoidal-like distribution seen in Figure 3.2) demonstrates that differences in average loss for the beginning step-intervals (*e.g.,* 0K to 60K steps) are attributed to the lower count of RACs that were present at the onset of the study. Furthermore, differences in average loss for later step-intervals (*e.g.,* 160K to 200K steps) are attributed to the higher count of RACs that were present near the end of the study. In addition, the standard error bars generally overlap across the cumulative step-intervals, which reinforces that average loss is unlikely to be statistically different between cumulative step-intervals. In contrast, the step-intervals with 0K as the lower boundary, as well as the 60K - 80K step-interval, fall outside of this trend. Nonetheless, these observations are believed to be explainable. Features marked during the zero step-interval were those present prior to being worn by the participants (and therefore present based on random pre-use activity such as shoe fitting). However, in hindsight, perhaps these were mold defects or sub-class characteristics, and therefore fundamentally different from wear-based characteristics, which could explain the lower rate of loss. Likewise, features marked in the 60K step-interval and prior had the potential to be outsole texture rather than true RACs (see Appendix section 5.5) and may have resulted in over-marking, which could explain the sudden increase of loss seen in the 60K - 80K step-interval (*e.g.,* the researcher stopped marking these features). Assuming that the above explains these outliers, then the trend is that the rates do not differ between earlier and later step-intervals. As a result, for these shoes, the terrain walked in this study, and the assumption of no change in general terrain for any degree of usage, special attention does not need to be afforded to the degree of general wear when considering RAC loss, as a similar rate of loss was observed for shoes with both lower and higher amounts of cumulative wear when considering the number of RACs present in the previous step-interval of comparison.

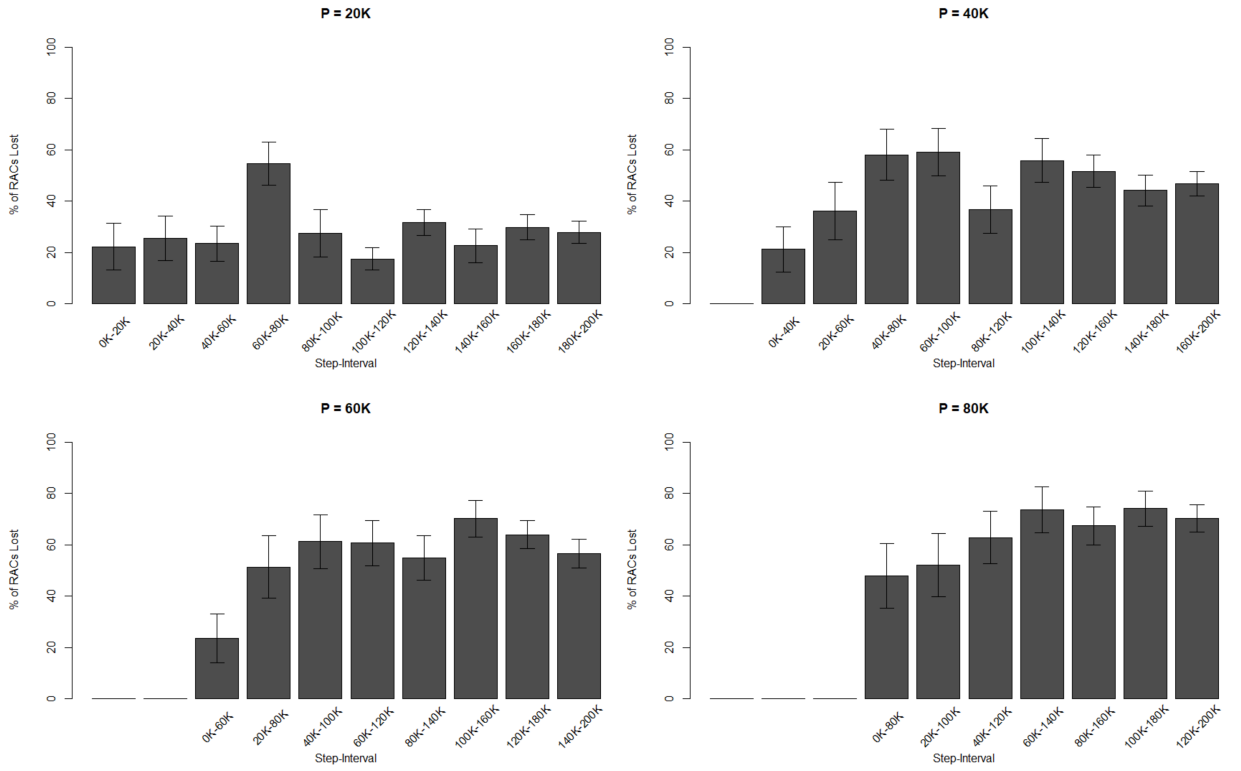

Figure 3.3: Bar plots (and standard error) of RAC loss (for n = 16 shoes), shown as a percentage of the number of RACs present, for a fixed step-interval difference (P = 20K, 40K, 60K and 80K) as a function of shoe total usage. The x-axis reports the specific step-interval under investigation, while the y-axis reports the average percent of RACs lost out of the number of RACs present (first value in x-axis label) for all participants. For example, the 20K - 40K observation in the P = 20K step-interval difference plot corresponds to the average percent of RACs lost out of the number of RACs present in the 20K step-interval when comparing the 20K step-interval to the 40K step-interval.

Table 3.1 reports the average gain, average loss, percent average loss and one standard deviation for each step-interval difference P, which allows for a comparison between RAC loss and gain through additional use. As the degree of wear increases between the deposition of a shoeprint at a scene and the collection of an exemplar, the average number of RACs gained increases from 1.6X (for a 20K difference or 6.4/3.9) to 2.0X (for a 40K difference or 8.0/3.9) to 2.4X (for a 60K difference or 9.4/3.9). Likewise, the average percent of RACs lost increases similarly from 1.6X (for a 20K difference or 45.5/28.3) to 2.0X (for a 40K difference or 55.3/28.3) to 2.3X (for a 60K difference or 64.0/28.3). For these shoes and the terrain walked in this study, an increasing number of RAC gain was observed in comparison to RAC loss. However, this observation is specific to this dataset, as shoes were studied from "new" (or no use) and therefore did not have any/many RACs present at the onset of the study (meaning few RACs had the potential to be lost for the step-intervals of 20K - 80K reported in Table 3.1). In practice, this indicates — that for this dataset — if many (>> 15) RACs are present to begin with, as the number of additional steps of use increases (from 20K to 40K to 60K to 80K) RAC gain will not match or outrun loss, and a decreasing number of total RACs will be observed on the exemplar than the less-worn questioned impression. Conversely, if

few (<< 15) RACs are present to begin with, gain will outrun loss and more total RACs will be observed on the exemplar than the less-worn questioned impression. If this observation is reconfirmed in additional studies with different shoes, then the results form a baseline for what a footwear examiner could reasonably expect when impressions with differing degrees of wear are compared in criminal investigations. For 20K steps of additional wear in this dataset, an examiner could expect approximately four RACs to be gained, and 71.7% (100% - 28.3%) of pre-existing RACs to be retained. Moreover, for 80K steps of additional wear, an examiner could expect approximately nine RACs to be gained and 36% (100% - 64%) of pre-existing RACs to be retained.

Table 3.1: Columns correspond to step-interval difference (P), average gain, average loss, percent average loss and one standard deviation for each observation.

P	Average Gain ± SD	Average Loss ± SD	Percent Average Loss ± SD
20K	3.9 ± 3.5	2.5 ± 2.9	28.3 ± 28.7
40K	6.4 ± 4.9	3.6 ± 3.7	45.5 ± 34.8
60K	8.0 ± 5.9	3.8 ± 3.8	55.3 ± 36.5
80K	9.4 ± 6.5	4.0 ± 4.3	64.0 ± 37.8

3.3 Spatial Distribution of RACs

Spatial heatmaps were constructed for gain and loss across step-intervals to visually represent locations in which RACs were acquired and lost. The forensic purpose of this investigation was to identify any recurring patterns of loss or gain across the outsole. For example, do RACs which are acquired in the heel get lost more than other areas, and do more RACs get acquired in the medial ball of the toe than other regions?

Figure 3.4 illustrates the total count gained (left-most heatmap) and lost (left-of-center heatmap) (shaded in a blue-red color gradient) for all participants across all step-intervals. Its interpretation is supplemented by the inclusion of a Handiprint collected at 0K steps of cumulative wear (right-of-center Handiprint), and a magnified view of a Handiprint collected at 200K steps of cumulative wear (right-most Handiprint). There is an observable difference in maximum RAC count between the loss and gain heatmaps on the medial edge of the ball of the toe, which may suggest that acquisition contributes more to the expected differences with continued wear in this area. The absence of observations in the center instep and center heel region are a function of non-contact areas, which can be seen in the left-of-center Handiprint shown in Figure 3.4. An entire lack or low number of RACs are acquired in the center toe region (and therefore a lack or low number of RACs lost). This is hypothesized to be the case due to the presence of Schallamach in this area, which can be seen in the magnified image of the ball of the toe in a Handiprint collected after 200K steps of wear. At the onset of the study, shoes gained this Schallamach and therefore inhibited the acquisition (and therefore eventual loss) of RACs *of other types* in these locations. In other other words, Schallamach patterns were not marked in this study, even though they are defined as a type of RAC [22].

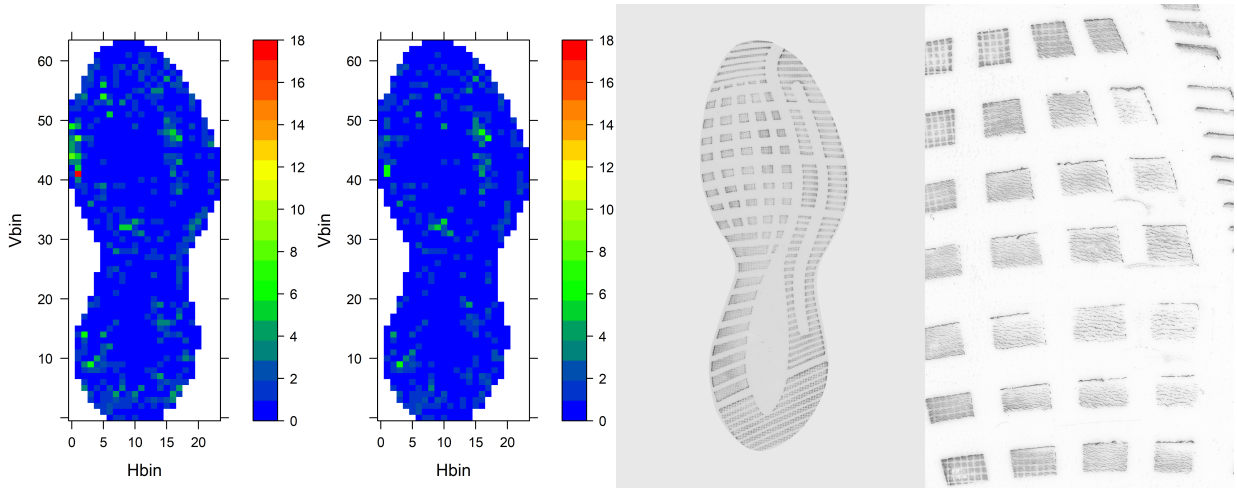

Figure 3.4: Spatial distribution of gain (left-most heatmap) and loss (left-of-center heatmap) across all step-intervals. Heatmaps are accompanied by Handiprints collected at 0K steps (right-of-center Handiprint), and a magnified view of the center toe region of a Handiprint collected at 200K steps (right-most Handiprint). Total number of acquisitions and losses per spatial cell across all step-intervals are represented in a blue-red color gradient.

Spatial distribution of RAC gain parsed into individual step-intervals is represented in Figures 3.5 - 3.6. Each individual heatmap corresponds to a comparison between a fixed step-interval F to step-interval $F + 1$. For example, the first heatmap reflects the spatial distribution of gain when comparing 0K - 20K steps. Total count of gain is shaded in a blue-red color gradient. Although tests for autocorrelation or complete spatial randomness were not performed, there is only very limited evidence of increased acquisition in the medial edge of the toe in later step-intervals.

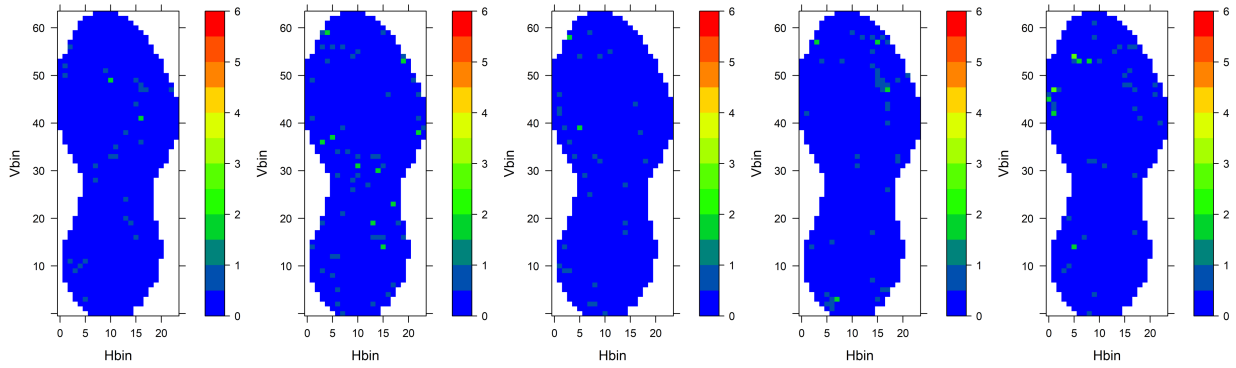

Figure 3.5: Spatial heatmaps of gain for earlier step-intervals. Total number of acquisitions for comparisons per spatial cell between step-intervals are represented in a blue-red color gradient. Each heatmap reflects a comparison of step-interval F to $F + 1$, ranging from $0KF/0KF + 1$ to $80KF/80KF + 1$ (*e.g.*, the leftmost heatmap corresponds to all RACs gained from the 0K step-interval to the 20K step-interval).

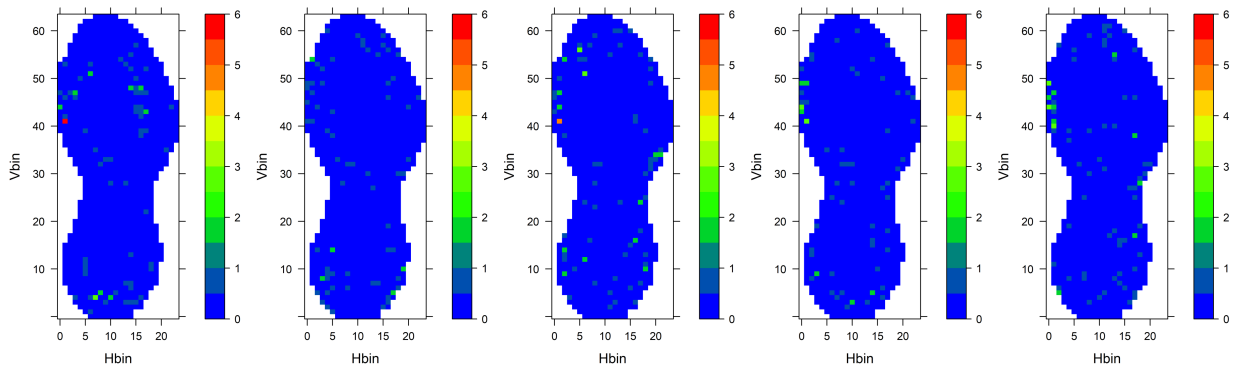

Figure 3.6: Spatial heatmaps of gain for later step-intervals. Total number of acquisitions for comparisons per spatial cell between step-intervals are represented in a blue-red color gradient. Each heatmap reflects a comparison of step-interval F to $F+1$, ranging from $100KF/100KF+1$ to $180KF/180KF+1$ (*e.g.*, the leftmost heatmap corresponds to all RACs gained from the 100K step-interval to the 120K step-interval).

Figures 3.7 - 3.8 illustrate the spatial distribution of RAC loss parsed into individual step-intervals. The heatmaps represent a comparison between a fixed step-interval F to step-interval $F+1$, and count is depicted in a blue-red color gradient. Again, although tests for autocorrelation or complete spatial randomness were not performed, there is only limited evidence of increased loss in the lateral toe and heel in some step-intervals.

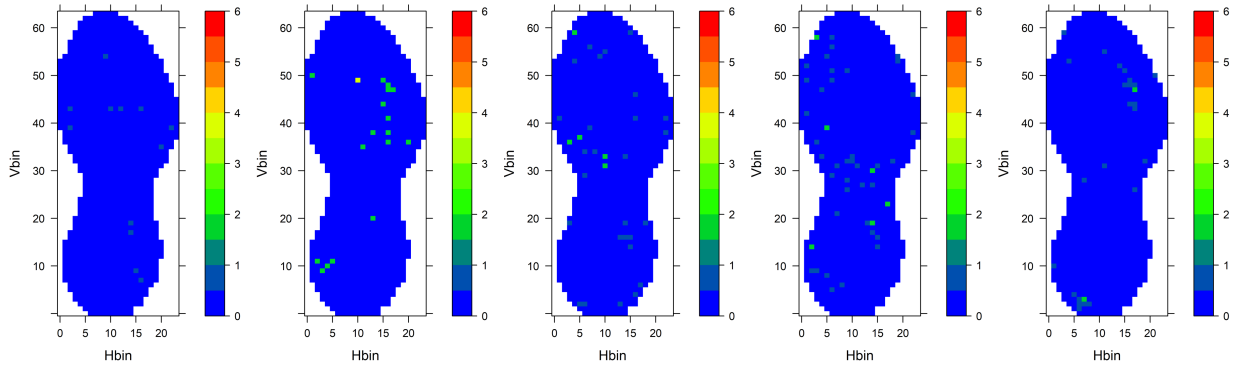

Figure 3.7: Spatial heatmaps of loss for earlier step-intervals. Total number of losses for comparisons per spatial cell between step-intervals are represented in a blue-red color gradient. Each heatmap reflects a comparison of step-interval F to $F+1$, ranging from $0KF/0KF+1$ to $80KF/80KF+1$ (*e.g.*, the leftmost heatmap corresponds to all RACs lost from the 0K step-interval to the 20K step-interval).

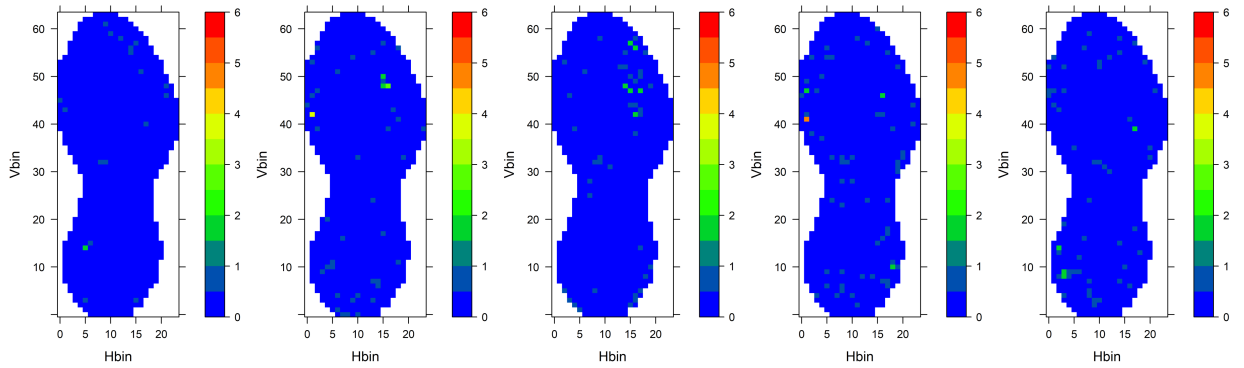

Figure 3.8: Spatial heatmaps of loss for later step-intervals. Total number of losses for comparisons per spatial cell between step-intervals are represented in a blue-red color gradient. Each heatmap reflects a comparison of step-interval F to $F+1$, ranging from $100KF/100KF+1$ to $180KF/180KF+1$ (*e.g.*, the leftmost heatmap corresponds to all RACs lost from the 100K step-interval to the 120K step-interval).

3.4 RAC Geometric Evolution

For the purpose of quantifying the degree of geometric change expected with continued and cumulative wear, the pairwise similarity of RAC contours was computed for seven different RAC-based similarity metrics (see Appendix section 5.6 for information on previous analytical approaches). Similarity scores were used to construct receiver operating characteristic (ROC) curves for each step-interval difference (P) and cumulative step-interval (F). From each ROC curve, the area under the curve (AUC) was computed, and each AUC was compared pairwise to test for statistically significant differences. ROC curves generated for all metrics were used to determine the optimal similarity metric among the seven selected for comparison. This similarity metric was then used to answer two forensically relevant questions. First, does the shape/geometry of RACs that are mated between step-intervals express more or less numerical similarity than non-mated RACs? For example, does a RAC which persists through 20K steps of additional wear (or a 20K step-interval difference) have a greater numerical self-similarity than a non-mate? Second, do any trends exist for RAC geometry based on accumulated wear (or cumulative step-interval)? For example, does a heavily-worn shoe have a diminished degree of numerical self-similarity in comparison to a lightly-worn shoe or vice versa?

Figure 3.9 is a plot of classification performance for each of the seven similarity metrics used to compare RACs through continued wear. Absolute value Euclidean distance (AE), total Euclidean distance (TE), maximum Hausdorff distance (HM), average Hausdorff distance (HA), matched filter (MF), Hermitian angle (HRA), and percent area overlap (%A) were the similarity metrics tested; the x-axis corresponds to the false positive rate and the y-axis corresponds to the true positive rate. For these ROC curves, a KM pair is defined as RAC pairs which persist through any degree of continued wear at any cumulative age, and a KNM pair is defined as any non-mated pair at any degree of continued wear or any cumulative age. Data associated with these ROC curves is shown in Table 3.2, where columns correspond to the similarity metric classifier, the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number

of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I). All 1,208 RACs were compared pairwise resulting in 729,028 comparisons — 1,448 of which were KM pairs (created from 943 unique RACs) and 727,580 were KNM pairs. Average Hausdorff distance was selected as the similarity metric for investigating change in geometry with cumulative and continued wear due to it having the highest AUC. When grouped regardless of additional wear (up to 180K additional steps) and regardless of total shoe usage (20K, 40K, etc. up to 200K of total usage), there is a 0.86 probability that a RAC is more similar to itself than a KNM when assessed by HA for this dataset. Figure 3.10 reflects the results of pairwise testing for statistically significant differences ($\alpha = 0.05$) between the classification performance of the similarity metrics. The x-axis and the y-axis are the classifiers, and each data point corresponds to a statistically significant (blue shading) or non-significant (red shading) difference between classifier AUCs, computed using DeLong's approach [19]. All classifiers were found to be statistically different than HA (the best-performing metric), and therefore HA was the only metric used moving forward.

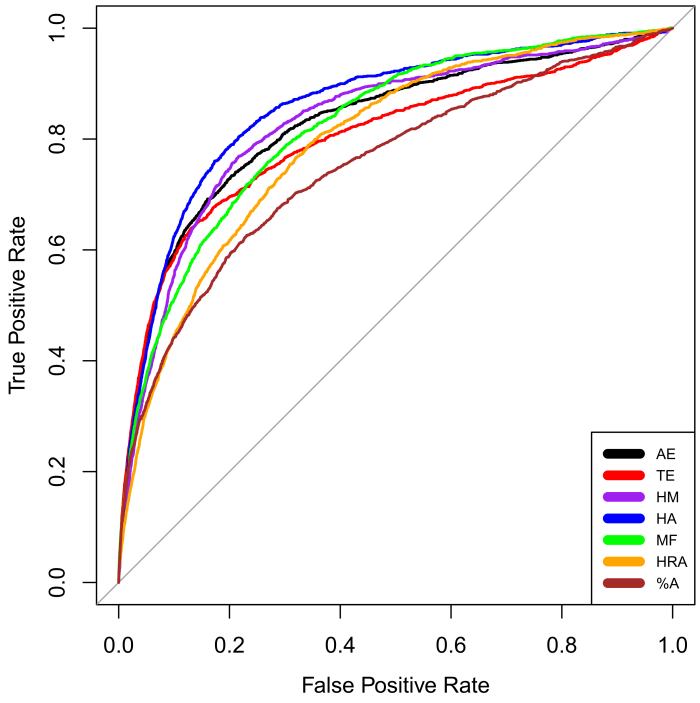

Figure 3.9: Classifier receiver operating characteristic (ROC) curves. The x-axis corresponds to the false positive rate of classification, and the y-axis corresponds to the true positive rate of classification. Each ROC curve corresponds to the classification performance for the classifiers of: absolute value Euclidean distance (AE), total Euclidean distance (TE), maximum Hausdorff distance (HM), average Hausdorff distance (HA), matched filter (MF), Hermitian angle (HRA), and percent area overlap (%A).

Table 3.2: Classifier receiver operating characteristic (ROC) curve data. Columns correspond to classifiers (acronyms are previously explained in text and Figure 3.9 caption), the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I).

Classifier	**#KM**	n **for #KM**	**#KNM**	n **for #KNM**	**Power**	**AUC**	**C-I**
AE	1,448	943	727,580	1,208	1.00	0.828	0.816-0.840
TE	1,448	943	727,580	1,208	1.00	0.800	0.786-0.814
HM	1,448	943	727,580	1,208	1.00	0.828	0.817-0.840
HA	1,448	943	727,580	1,208	1.00	0.856	0.845-0.866
MF	1,448	943	727,580	1,208	1.00	0.822	0.812-0.832
HRA	1,448	943	727,580	1,208	1.00	0.793	0.782-0.805
%A	1,448	943	727,580	1,208	1.00	0.750	0.734-0.764

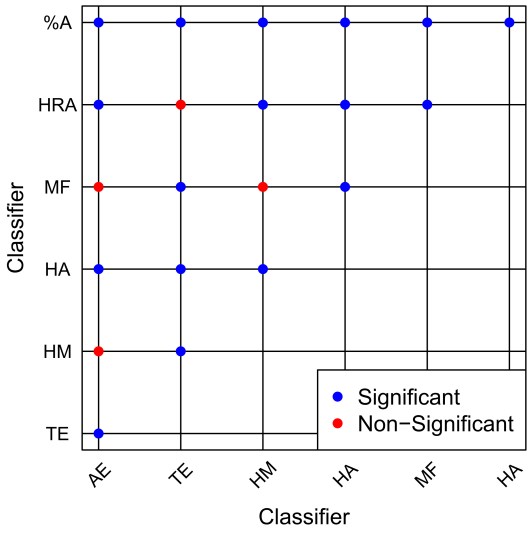

Figure 3.10: Classifier area under the curve (AUC) pairwise significance testing. The x-axis and y-axis denote classifier acronyms. Each observation indicates a statistically significant (at $\alpha = 0.05$) (blue shading) or non-significant (red shading) difference between AUCs describing classifiers.

Figure 3.11 is a plot of classification performance of average Hausdorff distance of RAC similarity measured using HA per step-interval difference (P). The x-axis corresponds to the false positive rate and the y-axis corresponds to the true positive rate. For these ROC curves, a KM pair is defined as

RAC pairs which persist through the specified degree of continued wear at any cumulative age, and a KNM pair is defined as any non-mated pair at the same specified degree of continued wear at any cumulative age (meaning each KNM set parallels the KM set). Data associated with these ROC curves is shown in Table 3.3, where columns correspond to step-interval difference (P), the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I). All step-interval differences are included in the interest of completeness; however, step-interval differences of P = 140K - 180K have power below the verbal equivalent of "good" (or > 0.80) [23], and were therefore not further considered when forming any conclusions. Generally, the probability of a KM being correctly classified as itself decreases with increasing continued wear. Of interest is the approximate 4% decrease in the probability of a RAC being classified as itself over a non-mate for an increasing step-interval difference of 20K, 40K, 60K, etc. for every step-interval difference with "good" power compared to the subsequent one (except for P = 100K vs. 120K [which is an approximate 8% decrease]). This suggests that the probability of a mated pair with wear appearing more similar than a non-mated pair decreases linearly with increasing step-interval differences. When comparing the two extremes of the step-interval differences with acceptable power (P = 20K vs. 120K), the AUC decreases by a probability of 0.241 (AUC = 0.902 vs. 0.661) which represents a 36.5% decrease (100% x ((0.902-0.661)/0.661) = 36.5%). However, the fact that the difference in probability with an additional 100K steps of continued wear (P= 120K - 20K = 100K additional steps of continued wear) is less than 25% indicates that numerical geometric self-similarity is reasonably retained through continued wear. At lower step-interval differences (*i.e.,* P = 20K, 40K, 60K), there is a higher probability of a known mate being correctly classified as itself. For example, for 20K steps of continued wear, which can be accrued in as few as 1 - 3 days, there is a 0.90 probability of RACs retaining greater self-similarity in geometric size and shape as measured by average Hausdorff distance than non-mated RACs. For 80K steps of continued wear, which can be accrued in as few as 4 - 6 days, this drops to a probability of 0.79, which is still quite promising. Pairwise significance testing was conducted for each of the step-interval differences with "good" power (see Appendix section 5.7).

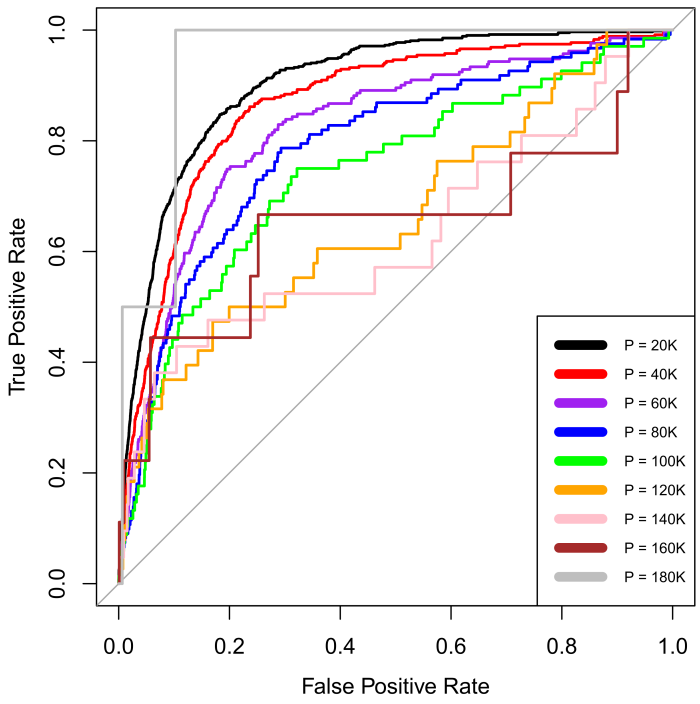

Figure 3.11: Step-interval difference receiver operating characteristic (ROC) curves. The x-axis corresponds to the false positive rate of classification, and the y-axis corresponds to the true positive rate of classification. Each ROC curve corresponds to the classification performance of average Hausdorff (HA) computed for a step-interval difference, ranging from P = 20K - 180K.

Table 3.3: Step-interval difference receiver operating characteristic (ROC) curve data. Columns correspond to step-interval difference (P), the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I).

P	**#KM**	**n for #KM**	**#KNM**	**n for #KNM**	**Power**	**AUC**	**C-I**
20K	623	920	152,127	1,208	1.00	0.902	0.892-0.913
40K	354	598	129,075	1,208	1.00	0.866	0.847-0.885
60K	211	390	103,594	1,208	1.00	0.822	0.792-0.852
80K	122	236	83,082	1,208	1.00	0.788	0.745-0.831
100K	68	136	61,071	1,208	0.999	0.742	0.676-0.808
120K	38	76	46,465	977	0.940	0.661	0.563-0.760
140K	21	42	34,724	756	0.546	0.628	0.479-0.776
160K	9	18	21,883	502	0.365	0.651	0.399-0.903
180K	2	4	6,750	243	0.687	0.945	0.851-1.00

Figure 3.12 is a plot of classification performances of average Hausdorff distance of RACs for each cumulative step-interval versus the subsequent one (F vs. F+1). The x-axis corresponds to the false positive rate and the y-axis corresponds to the true positive rate. For these ROC curves, a KM pair is defined as RAC pairs which persist through 20K steps of continued wear for each cumulative step-interval versus the subsequent one, and a KNM pair is defined as any non-mated pair at the same 20K steps of continued wear at the same cumulative age comparison (meaning each KNM set parallels the KM set). Data associated with these ROC curves is shown in Table 3.4, where columns correspond to cumulative step-interval compared against the following one (F vs. F+1), the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I). All AUCs are generally similar across each of the cumulative step-interval comparisons. This indicates that increasing general wear does not alter the probability that a RAC is correctly classified as itself for this dataset. Pairwise significance testing was conducted for each of the cumulative step-intervals (see Appendix section 5.7) and generally supports the reported trend.

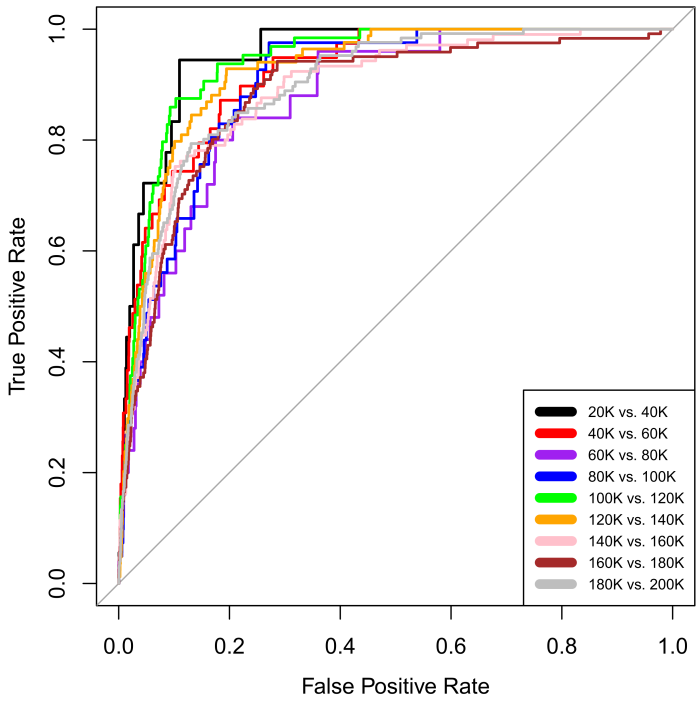

Figure 3.12: Cumulative step-interval receiver operating characteristic (ROC) curves. The x-axis corresponds to the false positive rate of classification, and the y-axis corresponds to the true positive rate of classification. Each ROC curve corresponds to the classification performance of average Hausdorff (HA) computed between a cumulative step-interval and the subsequent one (F vs. F+1), ranging from F = 20K - 180K.

Table 3.4: Cumulative step-interval receiver operating characteristic (ROC) curve data. Columns correspond to the comparison between a cumulative step-interval and the subsequent one (F vs. F+1), the number of known mate comparisons (#KM), the number of unique RACs contributing to the number of known mate comparisons (n for #KM), the number of known non-mate comparisons (#KNM), the number of unique RACs contributing to the number of known non-mate comparisons (n for #KNM), the power of the ROC curve, the area under the curve (AUC), and the confidence interval of the AUC (C-I).

F vs. F+1	**#KM**	**n for #KM**	**#KNM**	**n for #KNM**	**Power**	**AUC**	**C-I**
20K vs. 40K	18	36	2,414	107	1.00	0.951	0.921-0.981
40K vs. 60K	39	78	5,357	147	1.00	0.920	0.885-0.955
60K vs. 80K	25	50	4,945	140	1.00	0.876	0.819-0.932
80K vs. 100K	41	82	6,399	162	1.00	0.901	0.867-0.935
100K vs. 120K	64	128	12,724	231	1.00	0.939	0.919-0.959
120K vs. 140K	84	168	20,905	290	1.00	0.923	0.902-0.944
140K vs. 160K	105	210	27,528	334	1.00	0.889	0.859-0.918
160K vs. 180K	121	242	33,368	366	1.00	0.882	0.851-0.912
180K vs. 200K	126	252	38,487	358	1.00	0.899	0.875-0.922

3.5 Quantifying Outsole Evolution with Wear

For the purpose of understanding the influence of wear (both additional and cumulative) on the global self-similarity of shoes, confusion matrices were generated with cosine similarity and LDA as classifiers, and the accuracy of assignments were tabulated. The objective of these analyses were to answer three questions. First, does the degree of continued wear (or step-interval difference) affect classification accuracy? For example, is a shoe with 20K steps of additional wear (or a 20K step-interval difference) classified more accurately than a shoe with 40K steps of additional wear (or a 40K step-interval difference)? Second, does the shoe's age (or cumulative wear) affect classification accuracy? For example, is a more heavily-worn shoe classified more accurately than a lightly-worn shoe or vice versa? Third, is a shoe's closest non-mate its mated right/left shoe? For example, if a right shoe is not classified as itself, is it classified as the left shoe worn by the same participant?

3.5.1 Principal Component Analysis

Principal component analysis (PCA) was performed on vectors of RAC area per spatial bin (1145 features) that had at least one RAC (150 test impressions out of a possible 160 or 10 step-intervals x 8 participants x 2 shoes (left and right)). The 10 samples which had zero identified RACs, and were therefore removed from the data set, are tabulated in Appendix Tables 5.8 and 5.9. The PCA transformed data set resulted in data reduction from 1145 spatial bins to 57 new features (based on linear combinations of the original data). These 57 new features captured 95% of the original variability in the data. The first three principal component (PC) axes captured 16.9%, 11.4%, and 8.1% of the total variance, respectively. In contrast, Petraco *et al.* (2010) reported that 59.7% of the total variance was retained in the first three PCs as opposed to the ∼36.6% retained in this study [7].

3.5.2 Cosine Similarity

Cosine similarity (CS) was computed using Equation 2.7 on the PC-reduced data between all step-intervals and all participants, and each sample (PC-reduced vector) was assigned to the shoe with the smallest angle. For example, the 20K step-interval for participant AL was compared to the 40K - 200K step-intervals for participant AL and to the 20K - 200K step-intervals for all other participants. Table 3.5 reflects the cumulative accuracy of these classifications for all participants at the first, second, and third rank positions (or smallest, second smallest, and third smallest angles), where P corresponds to step-interval difference (*e.g.*, P = 20K encompasses both the 20K - 40K step-interval and the 80K - 100K step-interval), SS corresponds to shoe-specific accuracy or the number of times a shoe was classified as itself, and PS corresponds to participant-specific accuracy or the number of times a shoe was classified as itself or its mated left or right shoe. The cells of Table 3.5 highlighted in yellow represent comparisons with accuracy $\geq 75\%$. As the step-interval difference increases (to 40K, 60K, etc.), a shoe's self-similarity to its less worn self generally decreases. Table 3.6 reports the shoe-specific accuracy parsed into an arbitrary label mated to a participant shoe for each step-interval difference (P).

Table 3.5: First, second, and third rank classification accuracy using cosine similarity for each step-interval difference (*e.g.*, P = 20K, 40K, etc.) as a function of shoe total usage. In other words, P = 20K includes 20K vs. 40K, 40K vs. 60K, 60K vs. 80K, etc. while P = 160K includes 20K vs. 180K and 40K vs. 200K. Shoe-specific (SS) corresponds to accuracy of assignment when a shoe is classified as itself, and participant-specific (PS) corresponds to accuracy of assignment when a shoe is classified as itself or its mated left or right shoe. Cells highlighted in yellow reflect accuracy $\geq 75\%$. Zero step-interval vectors (meaning the appearance of the outsole prior to use for this study) were not analyzed. Outsoles at any step-interval with zero identified RACs are tabulated in Appendix section 5.9.

P	Rank 1		Rank 1 or 2		Rank 1, 2, or 3	
	Accuracy (SS)	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**
20K	104/144 (72.2%) [77.6%]†	104/144 (72.2%) [77.6%]†	117/144 (81.2%)	119/144 (82.6%)	119/144 (82.6%)	123/144 (85.4%)
40K	85/128 (66.4%) [72.0%]†	86/128 (67.1%) [72.8%]†	97/128 (75.7%)	99/128 (77.3%)	100/128 (78.1%)	109/128 (85.1%)
60K	68/112 (60.7%) [66.6%]†	70/112 (62.5%) [68.6%]†	76/112 (67.8%)	78/112 (69.6%)	81/112 (72.3%)	89/112 (79.4%)
80K	52/96 (54.1%) [60.4%]†	54/96 (56.2%) [62.7%]†	61/96 (63.5%)	64/96 (66.6%)	63/96 (65.6%)	72/96 (75.0%)
100K	40/80 (50.0%) [57.1%]†	43/80 (53.7%) [61.4%]†	46/80 (57.5%)	50/80 (62.5%)	48/80 (60.0%)	55/80 (68.7%)
120K	27/64 (42.1%) [50.0%]†	28/64 (43.7%) [51.8%]†	35/64 (54.6%)	36/64 (56.2%)	38/64 (59.3%)	40/64 (62.5%)
140K	25/48 (52.0%) [64.1%]†	25/48 (52.0%) [64.1%]†	28/48 (58.3%)	28/48 (58.3%)	32/48 (66.6%)	33/48 (68.7%)
160K	18/32 (56.2%) [75.0%]†	18/32 (56.2%) [75.0%]†	20/32 (62.5%)	20/32 (62.5%)	21/32 (65.6%)	21/32 (65.6%)
180K	3/16 (18.7%) [27.2%]†	3/16 (18.7%) [27.2%]†	4/16 (25.0%)	5/16 (31.2%)	5/16 (31.2%)	6/16 (37.5%)
Average	422/720 (58.6%) [66.1%]†	431/720 (59.9%) [67.6%]†	484/720 (75.9%)	499/720 (69.3%)	507/720 (70.4%)	548/720 (76.1%)

† Classification accuracy when inconclusives (step-intervals with zero RACs identified) are removed from the dataset

Regardless of the cumulative wear associated with a shoe (as in previous past usage of 10 miles (20K steps or light wear) versus 100 miles (200K steps or heavy wear)), if exemplars are collected at 20K step-intervals, shoes are more similar to themselves (versus any other non-mated shoe in this dataset) at least 72.2% of the time across all shoes and participants (Table 3.5). From Table 3.6, which reports the first rank classification accuracy per participant, shoe "HR" and "CL" seem to contribute the greatest number of erroneous classifications. When "HR" is removed, overall classification accuracy increases to 78.9% (101/128). Likewise, when both shoes "HR" and "CL" are removed, overall accuracy increases to 85.7% (96/112). The considerable increase in overall accuracy following the removal of these shoes is attributed to these samples having larger quanti-

ties of RAC acquisition and loss in later step-intervals (therefore resulting in a diminished degree of self-similarity and a subsequent decrease in classification accuracy — see Appendix Tables 5.4 - 5.5). Overall, results indicate that even with modest wear (10 miles), a shoe's overall RAC map may vary enough numerically to result in false first, second and third rank associations and disassociations if compared using RAC total area per spatial bin and using the numerical metric of cosine similarity. Of course, this is not to indicate that a visual examination would lead to a false inclusion or exclusion by a footwear examiner, but highlights that numerical changes in total RAC area per spatial bin are large enough to confuse the numerical metric of cosine similarity.

Equally interesting is that the participant-specific accuracy only increases slightly compared to the shoe-specific accuracy at any given step-interval. This suggests (at least for this dataset) that in the absence of general wear, specific wear, and/or Schallamach patterns (none of which were evaluated in this study), the closest non-mate in terms of RAC maps — as assessed via cosine similarity — is not necessarily the left/right shoe from a pair.

Table 3.6: First rank classification accuracy using cosine similarity for each participant for each step-interval difference (*e.g.*, P = 20K, 40K, etc.) as a function of shoe total usage. In other words, P = 20K includes 20K vs. 40K, 40K vs. 60K, 60K vs. 80K, etc., while P = 160K includes 20K vs. 180K and 40K vs. 200K. Shoe-specific (SS) corresponds to accuracy of classification when a shoe is classified as itself. Participant IDs were randomly assigned to an arbitrary identifier (A-H) and to its left (L) or right (R) identity. Zero step-interval vectors (meaning the appearance of the outsole prior to use for this study) were not analyzed. Outsoles at any step-interval with zero identified RACs are tabulated in Appendix section 5.9.

P	**AL**	**AR**	**BL**	**BR**	**CL**	**CR**	**DL**	**DR**	**EL**	**ER**	**FL**	**FR**	**GL**	**GR**	**HL**	**HR**
	Accuracy (SS)															
20K	7	7	6	6	5	7	6	6	6	9	7	7	9	7	6	3
40K	7	7	5	3	3	2	4	7	5	7	6	6	7	5	8	3
60K	5	6	4	3	2	1	2	7	4	7	5	5	7	3	6	1
80K	4	5	4	2	1	0	5	4	3	5	4	4	6	2	3	0
100K	2	2	2	1	1	2	4	4	2	4	3	3	5	1	3	1
120K	2	1	4	1	1	1	2	2	1	3	2	2	3	1	0	1
140K	2	2	2	1	2	1	2	2	0	3	1	1	3	1	1	1
160K	1	2	2	1	2	1	1	1	0	2	0	0	2	1	1	1
180K	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0

Table 3.7 reflects the cumulative accuracy of the aforementioned classifications for the first, second, and third rank, for a fixed step-interval difference of P = 20K, where each observation corresponds to comparisons of a cumulative step-interval F vs. F+1 for all participants (*e.g.*, 20K vs. 40K reflects the comparison of F = 20K and F+1 = 40K), SS refers to shoe-specific or the accuracy of assignment when a shoe is classified as itself, and PS refers to participant-specific or the accuracy of assignment when a shoe is classified as itself or its mated left or right shoe. The cells highlighted in yellow represent those with accuracy $\geq$ 75%. Early cumulative step-intervals (F $\leq$ 40K) reflect a lower accuracy of classification, whereas cumulative step-intervals thereafter reflect an increasing accuracy of classification. The comparison of F = 80K is an outlier in this general trend since all other comparisons at cumulative step-intervals $\geq$ 60K result in accuracy greater than $\geq$ 75%. Overall, the cumulative age of the shoe does impact classification accuracy, presumably because increased wear permits differentiation. From a forensic perspective, this means that special consideration needs to be afforded to shoes which appear to have been very lightly worn.

Table 3.7: First, second, and third rank classification accuracy using cosine similarity for each cumulative step-interval (*e.g.*, F = 20K, 40K, etc.) at a fixed step-interval difference of P = 20K. In other words, P = 20K includes 20K vs. 40K, 40K vs. 60K, 60K vs. 80K, etc. while P = 160K includes 20K vs. 180K and 40K vs. 200K. Shoe-specific (SS) corresponds to accuracy of classification when a shoe is classified as itself, and participant-specific (PS) corresponds to accuracy of assignment when a shoe is classified as itself or its mated left or right shoe. Cells highlighted in yellow reflect accuracy $\geq 75\%$. Outsoles at any step-interval with zero identified RACs are tabulated in Appendix section 5.9.

F vs. F+1	**Rank 1**		**Rank 1 or 2**		**Rank 1, 2, or 3**	
	Accuracy (SS)	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**
20K vs. 40K	4/16 (25.0%) [36.3%]†	4/16 (25.0%) [36.3%]†	4/16 (25.0%)	5/16 (31.2%)	5/16 (31.2%)	6/16 (37.5%)
40K vs. 60K	5/16 (31.2%) [38.4%]†	5/16 (31.2%) [38.4%]†	7/16 (43.7%)	7/16 (43.7%)	8/16 (50.0%)	9/16 (56.2%)
60K vs. 80K	13/16 (81.2%) [86.6%]†	13/16 (81.2%) [86.6%]†	14/16 (87.5%)	14/16 (87.5%)	14/16 (87.5%)	14/16 (87.5%)
80K vs. 100K	10/16 (62.5%) [66.6%]†	10/16 (62.5%) [66.6%]†	15/16 (93.7%)	15/16 (93.7%)	15/16 (93.7%)	15/16 (93.7%)
100K vs. 120K	13/16 (81.2%)††	13/16 (81.2%)††	14/16 (87.5%)	14/16 (87.5%)	14/16 (87.5%)	15/16 (93.7%)
120K vs. 140K	14/16 (87.5%)††	14/16 (87.5%) ††	15/16 (93.7%)	16/16 (100%)	15/16 (93.7%)	16/16 (100%)
140K vs. 160K	14/16 (87.5%)††	14/16 (87.5%)††	16/16 (100%)	16/16 (100%)	16/16 (100%)	16/16 (100%)
160K vs. 180K	16/16 (100%)††	16/16 (100%) ††	16/16 (100%)	16/16 (100%)	16/16 (100%)	16/16 (100%)
180K vs. 200K	15/16 (93.7%)††	15/16 (93.7%)††	16/16 (100%)	16/16 (100%)	16/16 (100%)	16/16 (100%)
Average	104/144 (72.2%) [77.6%]†	104/144 (72.2%) [77.6%]†	117/144 (82.3%)	119/144 (82.6%)	119/144 (82.6%)	123/144 (85.4%)

† Classification accuracy when inconclusives (step-intervals with zero RACs identified) are removed from the dataset
†† No change in accuracy with or without inconclusives

When considering the differences in SS accuracy and PS accuracy in Tables 3.5 and 3.7, there is little to no increase in classification success. In other words, each shoe's closest non-mate was infrequently its mated left or right shoe. Again this suggests (at least for this dataset) that with regard to RACs (as a reminder, general wear, specific wear and Schallamach patterns were not evaluated) there is little evidence that left/rights are closer non-mates than other shoes of the same make and model.

Similar to Figures 3.1 and 3.2, bar plots with standard error bars were created to illustrate the average total number of RACs present on a shoe in this dataset (Figure 3.13). Table 3.8 was generated as a supplement to Figure 3.13, where F corresponds to cumulative step-interval (*e.g.*, 20K steps, 40K steps, etc.), Z corresponds to the total count of RACs present, $\bar{Z} \pm SD$ corresponds to the average across all participants and one standard deviation, and "Min" and "Max" correspond to the minimum and maximum Z. Based on the total number of RACs present for each 20K step-interval of cumulative use, this plot suggests the presence of at least six RACs (Table 3.8 — obtained after approximately 100K steps of wear) is required in order to have a sufficient number of features to compare such that 80% or more of the time a shoe will possess maximum self-similarity (based on the numerical metric of cosine similarity). In other words, sufficient use is required to create a sufficient number of differences between shoes in order to differentiate them from close non-mates using the metric of cosine similarity (when ignoring general/specific wear and Schallamach patterns).

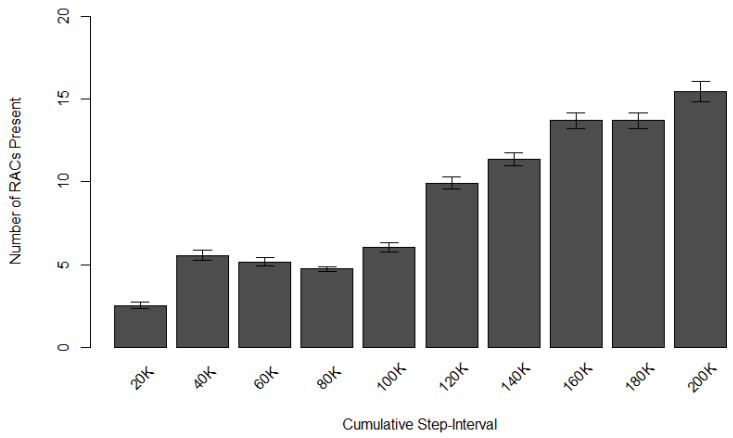

Figure 3.13: Bar plots (and standard error) of RAC count (for n = 16 shoes). The x-axis reports the specific cumulative step-interval under investigation, while the y-axis reports the average number of RACs present for all participants. For example, the 20K observation corresponds to the number of RACs present on average in the 20K cumulative step-interval.

Table 3.8: Columns correspond to cumulative step-interval (F) (*e.g.*, 20K steps, 40K steps, etc.), the total count of RACs present (Z), the average and one standard deviation across all participants ($\bar{Z} \pm SD$), and the minimum and maximum Z.

F	Z	$\bar{Z} \pm SD$	Min	Max
20K	41	2.6 ± 2.8	0	9
40K	89	5.6 ± 4.9	0	14
60K	83	5.2 ± 3.9	0	14
80K	76	4.8 ± 2.6	0	10
100K	97	6.1 ± 4.3	1	16
120K	159	9.9 ± 5.9	1	20
140K	182	11.4 ± 6.4	2	20
160K	219	13.7 ± 7.4	5	30
180K	219	13.7 ± 8.0	5	33
200K	247	15.4 ± 9.9	6	45

Table 3.9 reflects the accuracy for the first rank when considering any step-interval difference, where each observation corresponds to comparisons of a cumulative step-interval F vs. any F for all participants (*e.g.*, "20K vs. any" reflects the comparison of F = 20K versus 40K, 60K, 80K etc. up to 200K), SS refers to shoe-specific or the accuracy of assignment when a shoe is classified as itself, and PS refers to participant-specific or the accuracy of assignment when a shoe is classified as itself or its mated left or right shoe. In comparison to Table 3.7, a considerable increase in accuracy in classification is observed for earlier cumulative step-intervals (less than about 60K); thereafter, the accuracy in both tables is comparable. This increase in reported accuracy for the 20K and 40K step-intervals suggests that outsoles cluster more with themselves through continued

wear than they do with any non-mated outsole. Inspection of the feature vectors (four accurate at 20K in Table 3.7 and seven accurate at 20K in Table 3.9) did not reveal a trend or pattern which could explain this occurrence.

Table 3.9: First rank classification accuracy using cosine similarity for each cumulative step-interval (*e.g.*, F = 20K, 40K, etc.) compared to any step-interval difference. Shoe-specific (SS) corresponds to accuracy of classification when a shoe is classified as itself, and participant-specific (PS) corresponds to accuracy of assignment when a shoe is classified as itself or its mated left or right shoe. Cells highlighted in yellow reflect accuracy $\geq 75\%$. Outsoles at any step-interval with zero identified RACs are tabulated in Appendix section 5.9.

F vs. any F	**Rank 1**	
	Accuracy (SS)	**Accuracy (PS)**
20K vs. any	7/16 (43.7%) [63.6%]†	8/16 (50.0%) [72.7%]†
40K vs. any	11/16 (68.7%) [84.6%]†	11/16 (68.7%) [84.6%]†
60K vs. any	12/16 (75.0%) [80.0%]†	12/16 (75.0%) [80.0%]†
80K vs. any	13/16 (81.2%) [86.6%]†	13/16 (81.2%) [86.6%]†
100K vs. any	14/16 (87.5%)††	14/16 (87.5%)††
120K vs. any	13/16 (81.2%)††	13/16 (81.2%)††
140K vs. any	14/16 (87.5%)††	14/16 (87.5%)††
160K vs. any	15/16 (93.7%)††	15/16 (93.7%)††
180K vs. any	16/16 (100%)††	16/16 (100%)††
Average	115/144 (79.9%) [85.8%]†	116/144 (80.6%) [86.6]†

† Classification accuracy when inconclusives (step-intervals with zero RACs identified) are removed from the dataset
†† No change in accuracy with or without inconclusives

3.5.3 Linear Discriminant Analysis

A hold-one-out validation (HOOV) was performed using the PC-reduced data for all step-intervals and all participants with linear discriminant analysis (LDA) as a classifier, using Equations 2.8 and 2.9. Each shoe at each step-interval was classified to the participant-shoe combination with the lowest LDA distance. For example, the 20K step-interval left shoe for participant AL was held-out, each group centroid (mean vector) and individual covariance matrix was computed, followed by the dataset's pooled covariance matrix, and distances were calculated between the held-out sample and each group (where the number of groups = 16 or one per participant per left/right shoe). Table 3.10 reflects the cumulative accuracy of HOOV classifications for all held-out RAC vectors at the first, second and third rank positions (or smallest, second smallest and third smallest LDA distance), where F corresponds to the cumulative step-interval of the held-out sample, SS corresponds to shoe-specific accuracy or the number of times a shoe was classified to its own group, and PS corresponds to participant-specific accuracy or the number of times a shoe was classified to its own group or its mated left or right group. The cells highlighted in yellow represent comparisons with accuracy $\geq 75\%$. When using LDA as a classifier, the classification of a held-out sample with greater cumulative use reflects an equal or higher accuracy than a held-out sample with lesser cumulative use. Conversely, the classification of a held-out sample with lesser cumulative use reflects an equal or lower accuracy than a held-out sample with greater cumulative use. The held-out cumulative step-intervals F = 80K and F = 140K fall slightly outside of this general trend; however, this deviation in the trend for accuracy was only a difference of two correct classifications or less. The increased accuracy in classification for higher cumulative step-intervals is likely attributed to both the removal of the aforementioned empty set step-intervals and the higher RAC count of later step-intervals. In other words, the 10 empty set vectors all occurred prior to the 80K step-interval, and as a consequence, group means/covariances would be more strongly dominated by later step-

intervals, therefore resulting in a greater degree of similarity (or a higher accuracy) between the mean and later step-intervals. The overall lower accuracy reported for this dataset contrasts with the findings reported by Petraco *et al.* (2010). [7]. However, this former research was based on RAC spatial information that was collected on specific days during the study, rather than at specific step-intervals. As a result, the total duration of use may have been much less (30 days) than the total number of steps taken in this study. In addition, Petraco *et al.* (2010) [7] partitioned the study such that shorter step-intervals were used when forming the mean and covariance matrices. That said, the trends reported in both studies are consistent in that later step-interval accuracy increases over earlier step-intervals.

When comparing between the accuracy of the two classifiers — cosine similarity (Table 3.7) and linear discriminant analysis (Table 3.10) — there is a notable decrease in performance for the latter. This fits intuition, suggesting that a centroid-based classifier (as opposed to the use of an individual end-member-based classifier) is a less useful measure of the similarity of RAC maps (when ignoring general/specific wear and Schallamach), especially when the class mean is formed over a large degree of cumulative wear (*e.g.,* 200K steps).

Table 3.10: First, second, and third rank hold-one-out validation (HOOV) classification accuracy using LDA for a held-out sample with cumulative step-interval F (*e.g.*, F = 20K, 40K, etc.). Shoe-specific (SS) corresponds to accuracy of assignment when a shoe is classified to its own group (participant and left or right shoe), and participant-specific (PS) corresponds to accuracy of assignment when a shoe is classified to its group or its mated left or right group. Cells highlighted in yellow reflect accuracy $\geq 75\%$. Outsoles at any step-interval with zero identified RACs are tabulated in Appendix section 5.9.

Held-Out F	**Rank 1**		**Rank 1 or 2**		**Rank 1, 2, or 3**	
	Accuracy (SS)	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**	**Accuracy (SS)**	**Accuracy (PS)**
20K	5/16 (31.2%) [45.5%]†	5/16 (31.2%) [45.5%]†	5/16 (31.2%)	5/16 (31.2%)	6/16 (37.5%)	7/16 (43.7%)
40K	8/16 (50.0%) [61.5%]†	8/16 (50.0%) [61.5%]†	10/16 (62.5%)	10/16 (62.5%)	12/16 (75.0%)	12/16 (75.0%)
60K	10/16 (62.5%) [66.7%]†	11/16 (68.7%) [73.3%]†	10/16 (62.5%)	11/16 (68.7%)	11/16 (68.7%)	12/16 (75.0%)
80K	8/16 (50.0%) [53.3%]†	9/16 (56.2%) [60.0%]†	10/16 (62.5%)	11/16 (68.7%)	10/16 (62.5%)	11/16 (68.7%)
100K	10/16 (62.5%)††	11/16 (68.7%)††	12/16 (75.0%)	12/16 (75.0%)	12/16 (75.0%)	12/16 (75.0%)
120K	11/16 (68.7%)††	11/16 (68.7%)††	11/16 (68.7%)	11/16 (68.7%)	12/16 (75.0%)	12/16 (75.0%)
140K	10/16 (62.5%)††	10/16 (62.5%)††	11/16 (68.7%)	11/16 (68.7%)	12/16 (75.0%)	12/16 (75.0%)
160K	13/16 (81.2%)††	13/16 (81.2%)††	15/16 (93.7%)	15/16 (93.7%)	16/16 (100%)	16/16 (100%)
180K	13/16 (81.2%)††	13/16 (81.2%)††	14/16 (87.5%)	14/16 (87.5%)	15/16 (93.7%)	15/16 (93.7%)
200K	12/16 (75.0%)††	12/16 (75.0%)††	13/16 (81.2%)	13/16 (81.2%)	14/16 (87.5%)	14/16 (87.5%)
Average	100/160 (62.5%) [66.7%]†	103/160 (64.3%) [68.7%]	111/160 (69.4%)	113/160 (70.6%)	120/160 (75.0%)	123/160 (76.9%)

† Classification accuracy when inconclusives (step-intervals with zero RACs identified) are removed from the dataset
†† No change in accuracy with or without inconclusives

3.6 Uncertainty of Measurement

With the intention of quantifying the variation which can be attributed to analyst marking (for both RAC detection and geometry), four analytical approaches were evaluated. The first metric used to meter changes in detection and geometry was a global 2-dimensional correlation coefficient, which was computed for each mated QC image pair (36 total impression mark-ups). The second metric investigated was the average change in RAC detection between repeated markings. The third metric was aimed at reporting changes in RAC geometry due to wear (P = 20K - 80K) deemed greater than intra-analyst variation in marking, as measured by the overlap of probability density functions describing HA scores for KNMs, KMs-with-wear, and KMs-with-analyst-variation-in-marking. Fi-

nally, the QC vectors were classified. For both cosine similarity and linear discriminant analysis, each quality control PC-transformed vector was assigned to the closest QC PC-transformed vector using both classifiers. Overall, this investigation aimed to examine the degree of change in the detection of RACs, their geometry, and the overall deviation from group membership which can be attributed to the examiner (as opposed to the degree of change attributed to experimental factors). For example, what is the analyst's level of variation in identifying and tracing RACs, and is the degree of change observed due to wear for this dataset greater than the level of "noise" that can be attributed to analyst-variation?

3.6.1 Global Correlation and Detection

For all mated QC pairs, variability in correlation and RAC detection was investigated. Table 3.11 reflects summary data from this investigation, where columns correspond to the variable of interest, the average and one standard deviation of the associated variable, and the minimum and maximum value for the associated variable. More specifically, r is a 2-dimensional correlation coefficient computed between the mated QC RAC maps, "#New" is the number of additional RACs marked in the repeat, and "#Dropped" is the number of RACs not detected in the repeat. Values for "#New" and "#Dropped" per QC data set (where a set corresponds to a single outsole that was repeatedly marked, with a total of n = 18 useful sets) were averaged, since both are changes in RAC detection, and there is nothing special to differentiate the temporal order of marking. Data used to generate this summary table is located in Appendix section 5.8. On average, there was a 0.73 global correlation between mated QC markings, with a minimum of 0.41 and a maximum of 0.94, respectively. With the patch map mechanism limiting loss between step-intervals, differences in global correlation can be attributed primarily to differences in the detection of new RACs and in tracing their geometry. On average, two additional new RACs were marked during a repeated marking. However, the worst-case scenario corresponded to a maximum of nine differences in detection. As a result, reported rates of acquisition equal to or less than two are most likely due to the analyst's variation in RAC detection. Reported rates of loss may be indirectly affected as well. Since a RAC can only be lost after it is acquired, an increased/decreased level of detection may have resulted in increased/decreased rates of loss. However, since the patch map mechanism was implemented in order to limit the possibility of missed analyst's detections contributing to loss, the contribution/uncertainty in the frequency of loss cannot be quantified.

Table 3.11: Variation in the 18 repeated QC image sets (see Appendix Table 5.7 for individual data per QC set). Columns correspond to the variable of interest, the average and one standard deviation of the associated variable, and the minimum and maximum value for the associated variable. Regarding the variables of interest, r is a 2-dimensional correlation coefficient computed between the mated QC RAC maps, "#New" is the number of additional RACs marked in the repeat, and "#Dropped" is the number of RACs not detected in the repeat. Values for "#New" and "#Dropped" were averaged, since both contribute to changes in detection.

Variable	**Average ± SD**	**Min**	**Max**
r	0.73 ± 0.15	0.41	0.94
#New/#Dropped	2.06 ± 2.16	0	9

3.6.2 Geometric Similarity Score Distributions

Probability density functions of the average Hausdorff distance (HA) were generated using a Gaussian kernel density estimator for all quality control pairs (QCs), all known non-mated pairs (KNMs), and all known mated pairs (KMs) for step-interval differences P = 20K - 80K, and are illustrated in Figure 3.14. Each PDF was estimated using a Gaussian kernel with bandwidth equal to Silverman's rule [21]. For P = 20K - 80K, these values were 0.09, 0.11, 0.15, and 0.21 for KM pairs, respectively. For KNM pairs, these values were 0.11, 0.12, 0.12, and 0.12, respectively. Since the QC dataset was a fixed group, its bandwidth remained constant at 0.08. The vertical gray lines in Figure 3.14 at HA = 1.12 represent the upper limit for an integral (0 to HA = 1.12) that captures 95% of the QC density. Likewise, the vertical purple lines at HA = 1.38, 1.48, 1.58, and 1.72 for P = 20K, 40K, 60K, and 80K, respectively, represent the HA score where the height of the density curves is equal for the KMs and the KNMs (and therefore a likelihood ratio (LR) of 1.0). Table 3.12 can be used to supplement the understanding of these PDFs, where columns correspond to step-interval difference (P) and the integrals for the KM and KNM distributions at the aforementioned colored thresholds. The QC dataset describes the variation in known mated RAC HA scores due solely to variation in marking by the research analyst. Since 95% of QC similarity scores are less than 1.12, an HA of 1.12 or greater is considered the cut-off for detecting "true" variation in RAC geometry due to wear and/or between non-mated RAC pairs. Thus, for P = 20K, 27% of KM HA values (1 - KM "0 - gray" integral in Table 3.12), and 88% of KNM HA values (1 - KNM "0 - gray" integral in Table 3.12) are larger than 95% of the HAs explainable by analyst marking. When considering P = 80K (or 60K steps of additional wear over the P = 20K plot), the percentage greater than the QC threshold increases to 51% for KMs, but stays relatively constant at 90% for KNMs.

Focusing on the portions of the distributions that are not explainable by intra-analyst variation in marking, it is found after 20K steps of continued wear, the probability of an HA score between 1.12 and 1.38 ("gray - purple" integral in Table 3.12) is at most 2.1X more likely if sampled from KMs than from KNMs. Conversely, beyond 1.38, the score is more likely to occur in the KNM group. For scores greater than 1.38, it is observed that 15% of the KM scores ("purple - ∞" integral in Table 3.12) after 20K steps of additional wear are large enough to be more likely sampled from the KNMs (LRs less than 1.0). After 80K steps of continued wear, the probability of an HA score between 1.12 and 1.72 ("gray - purple " integral in Table 3.12) is at most 2.7X more likely if sampled from KMs than KNMs, which is a 28.6% increase (100% x ((2.7-2.1)/2.1) = 28.5%) in likelihood in comparison to P = 20K. For P = 80K, 25% of the KM scores are large enough to be more likely sampled from the KNMs (LRs less than 1.0), which is a 65.8% increase (100% x ((0.252-0.152)/0.152) = 65.8%) from the percentage observed for P = 20K.

If known mated RACs are compared that differ in additional steps of wear by 20K, 40K, 60K or 80K, a modest but increasing *percentage* of HA similarity scores (15%, 16%, 23%, and 25%, respectively) or *ratio* of KMs-with-wear versus KNM scores (19%, 20%, 30%, and 35%, respectively), are more likely to be attributed as originating from a KNM distribution. In other words, the additional wear is creating a numerical dissimilarity that is causing the conditional probability of an HA score if sampling from KM to dip below that if sampling from KNMs. This matches intuition, but the fact that the percentage only increases by 10% between 20K and 80K additional steps, and that it is never greater than 25%, is promising.

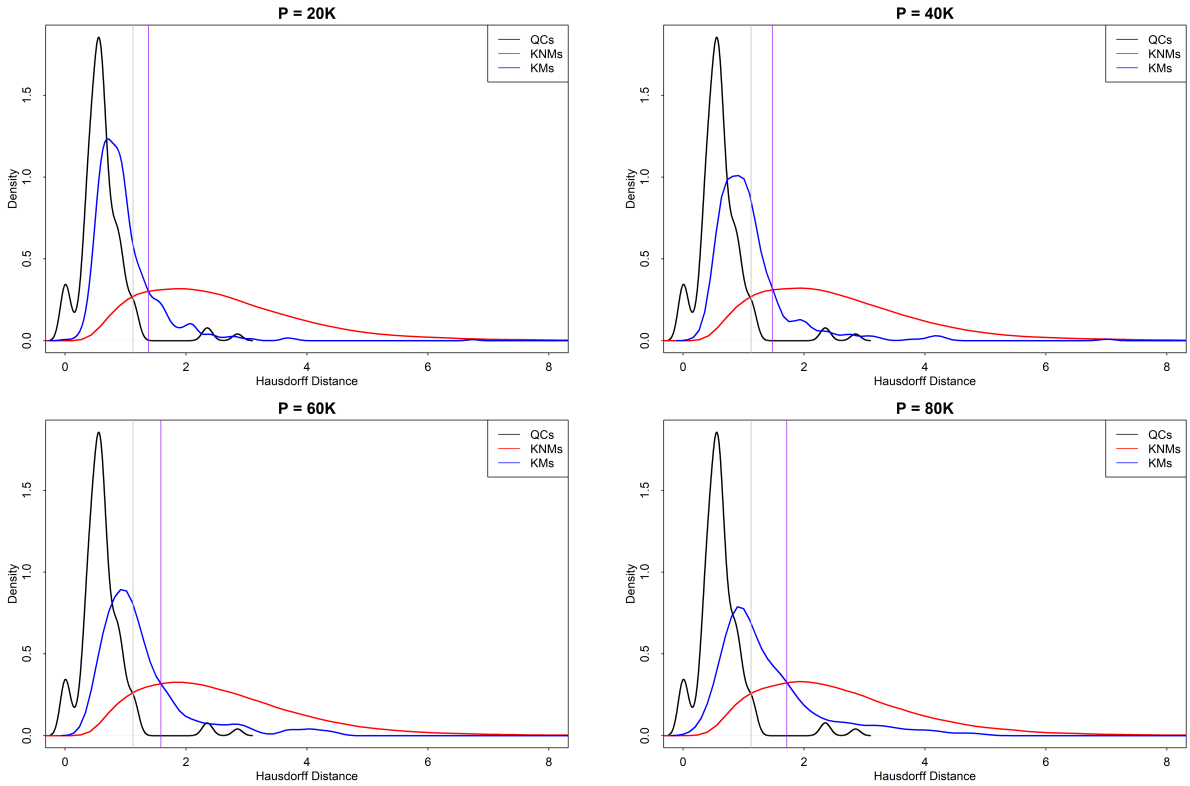

Figure 3.14: Average Hausdorff distance (HA) probability density functions. Plots illustrate the density of similarity scores generated with a kernel density estimator for quality control pairs (QCs), known non-mated pairs (KNMs), and known mated pairs (KMs) for step-interval differences P = 20K - 80K. See text for an explanation of HA score thresholds denoted by the "purple" and "gray" vertical lines.

Table 3.12: Probability density function integrals. Columns correspond to step-interval difference (P), and the integrals for the known mate (KM) and known non-mate (KNM) distributions for various average Hausdorff score thresholds. See text for an explanation of HA score thresholds denoted by the "purple" and "gray."

P	**Integrals**					
	0 - gray		**gray - purple**		**purple - ∞**	
	KM	**KNM**	**KM**	**KNM**	**KM**	**KNM**
20K	0.743	0.115	0.105	0.070	0.152	0.810
40K	0.648	0.112	0.193	0.106	0.158	0.782
60K	0.558	0.109	0.217	0.138	0.225	0.753
80K	0.488	0.105	0.260	0.176	0.252	0.719

3.6.3 Quantifying Outsole Evolution with Wear

Cosine Similarity

Pairwise computation of cosine similarity between all 36 QC PC vectors resulted in an accuracy of 91.7% (33/36), which means at least 8.3% of the previously reported classification inaccuracy of cosine similarity could be attributed to the analyst's variation in marking and/or the methodology used.

Linear Discriminant Analysis

A HOOV was performed using LDA as the classifier, where each group centroid was formed by a same-interval mated original and repeated QC PC vector. Table 3.13 reflects the accuracy of these HOOV classifications for all held-out PC-RAC vectors at the first rank position (or smallest LDA distance), where F corresponds to the cumulative step-interval of the held-out sample, SS corresponds to shoe-specific accuracy or the number of times a shoe was classified to its own group, and PS corresponds to participant-specific accuracy or the number of times a shoe was classified to its own group or its mated left or right group. The HOOV resulted in an overall 77.7% classification accuracy, meaning 23.3% of the HOOV inaccuracy observed in the full data set could be attributed to the analyst's variability in marking test impressions and/or the methodology used.

Table 3.13: First rank hold-one-out validation (HOOV) classification accuracy using LDA for a held-out quality control sample with cumulative step-interval F (*e.g.*, F = 20K, 40K, etc.). Shoe-specific (SS) corresponds to accuracy of assignment when a shoe is classified to its own group (participant and left or right shoe), and participant-specific (PS) corresponds to accuracy of assignment when a shoe is classified to its group or its mated left or right group.

	Rank 1	
Held-Out F	**Accuracy (SS)**	**Accuracy (PS)**
20K	1/1 (100%)	1/1 (100%)
40K	1/2 (50.0%)	1/2 (50.0%)
60K	1/1 (100%)	1/1 (100%)
80K	1/2 (50.0%)	1/2 (50.0%)
100K	2/2 (100%)	2/2 (100%)
120K	1/2 (50.0%)	2/2 (100%)
140K	2/2 (100%)	2/2 (100%)
160K	2/2 (100%)	2/2 (100%)
180K	2/2 (100%)	2/2 (100%)
200K	1/2 (50.0%)	1/2 (50.0%)
Average	14/18 (77.7%)	14/18 (77.7%)

4. Conclusions

The purpose of this research was to quantify the degree of change in randomly acquired characteristics that can be expected with continued wear for a specific make/model of outsole, as worn by an opportunistic set of participants traversing a fixed terrain. More specifically, the following variables were evaluated: (i) the rates of acquisition and loss, (ii) the spatial distribution of these occurrences, (iii) the degree of geometric change in RAC shape/morphology and, (iv) the deviation of an impression from accurate group classification. To investigate these phenomena, eight participants walked a total of 200K steps, and RACs were tracked as a function of steps accrued. Although the results of this study are specific to the selected make and model of shoe, the terrain walked, and the activities undertaken by participants, the observations can assist footwear examiners in providing a point of reference for dissimilarities observed between impressions with continued wear.

4.1 Acquisition/Retention of RACs

Naturally, the rate of acquisition is positively correlated with additional wear; for step-interval differences between 20K to 80K, the average gain in RAC count increases from 1.6X to 2.0X to 2.4X. However, these results are highly impacted by uncertainty. The QC dataset revealed that acquisition equal to or less than two RACs is most likely due to intra-analyst variation in detection.

Similarly, when loss is normalized by the number of RACs present, the rate of loss between 20K and 80K is also positively correlated with wear, changing from 1.6X to 2.0X to 2.3X. Uncertainty cannot be assigned to reported rates of loss due to the patch map mechanism limiting missed analyst's detection. As a consequence, count regarding RAC loss is considered more robust than count describing RAC acquisition.

Overall, results indicate that dissimilarities observed during a comparison of mated questioned and known impressions are dependent upon RAC count present in the original questioned impression; for outsoles with a higher RACs count, the combination of gain and loss (over 20K to 80K additional steps of use) is likely to lead to an overall decrease in the number of total RACs retained. Conversely, for outsoles with few RACs initially, gain outruns loss leading to an overall increase in the number of RACs observed in the test impression created from the more well-worn outsole.

4.2 Spatial Distribution of RACs

Visual assessment of the cumulative spatial heatmaps indicated that acquisition contributes more to the expected differences with continued wear in the medial edge of the ball of the toe. The absence of observations in the center instep and center heel region are a function of non-contact areas. Moreover, the presence of Schallamach patterns in the center toe region was observed to stagnate rates of change in acquisition and loss for other types of RACs, and therefore warrants further investigation. Visual assessment of the step-interval spatial heatmaps indicated a minor increase in the acquisition of RACs in the medial edge of the toe in later cumulative step-intervals (or more moderately worn shoes). Likewise, a minor increase in the loss of RACs was observed

in the lateral toe and heel in some step-intervals (most of which occurred for more moderately worn shoes). Although complete spatial randomness and autocorrelation tests were not performed, the general (but weak) trends observed, consisting of increased acquisition in the medial toe, and increased loss in the lateral toe and heel, can only be observed after moderate wear permits an accumulation of sufficient RACs.

4.3 RAC Geometric Evolution

The geometric evolution of RAC shape proved to be the most challenging attribute to quantify. No single numerical metric is well-suited for every type and shape of RAC, and visual perception of changes by human observers is equally problematic due to large intra- and inter-inconsistency in opinions.

However, if average Hausdorff distance is considered the most informative metric, and if results are *averaged* across all step-interval differences (up to 180K) and cumulative wear (up to 200K of total steps), then results indicate that there is a 0.86 probability that two randomly selected mated RACs that differ in wear will be deemed more similar to each other than any non-mated RACs pair.

Considering step-interval specific results, as the step-interval increases (from 20K, to 40K, to 60K, etc. up until approximately 100K of additional steps) this probability consistently drops by about 0.04 (0.902, to 0.866, to 0.822, to 0.788, to 0.742).

Finally, it was found that increasing general wear did not alter the probability that a RAC was correctly classified as itself. More specifically, the probability that a randomly selected mated RAC pair with additional wear would be more similar to each other than two non-mated RACs was relatively constant and high (between 0.95 and 0.90) for 20K steps of additional wear for a new shoe (20K to 40K) as compared to 20K steps of additional wear for a well-worn shoe (180K to 200K).

However, these results are impacted by uncertainty in geometry as assessed at the score level. After determining the HA score that captured 95% of the analyst's variation in marking QC RAC pairs, it was determined that 74% of KM HA values with 20K steps of continued wear, and 11% of KNM HA values — also with 20K steps of additional wear — could not be differentiated from analyst-variation in tracing RAC geometry. This does not mean that only 26% of the KMs exhibited geometry changes large enough to be of interest. Instead, it may be that the sensitivity of the methodology employed is insufficient to capture more than 27% of the wear-induced variation. Thus, it is most conservative to consider the ratio of HA scores for KMs-with-wear versus KNMs. This amounts to a normalized comparison. As a result, it can be reported that additional steps of wear by 20K, 40K, 60K to 80K results in a modest but increasing *ratio* of KM-with-wear HA similarity scores versus KNM scores (19%, 20%, 30%, and 35%, respectively), that are more likely to be attributed as originating from a KNM distribution.

4.4 Quantifying Outsole Evolution with Wear

Across all shoes and participants, 72.2% of test impressions were more similar to themselves after 20K additional steps using cosine similarity, regardless if lightly-worn or well-worn. As expected,

decreased accuracy was observed for greater degrees of continued wear when using cosine similarity as a classifier. More specifically, accuracy dropped by about 6% for every additional 20K steps of use (72%, 66%, 61%, 54%, 50%) up until about 120K total steps. However, results (and intuition) indicate that if evaluating impressions with little-to-no indications of general wear, the limited number of RACs present will make the outsole more similar to every other shoe with the same make, model and size, making it impossible to limit the population of possible sources for the impression. For this dataset, this was true up until about 80K to 100K steps of total use, after which impressions developed enough RACs (approximately six) to limit numerical similarity with non-mated sources, such that self-similarity started to climb above 85%. However, at least 8.3% of the inaccuracy observed in classification can be attributed to the methodology and/or analyst variation in marking.

Interestingly, when a sample did not reflect maximum self-similarity, the closest non-mate was infrequently its mated left or right shoe. This suggests for this dataset that the spatial relationship of individualizing features for the mated left and right shoes do not share greater similarity than unrelated shoes with comparable class characteristics.

Lastly, linear discriminant analysis reflected an overall accuracy of classification of 62.5%. However, uncertainty showed that 22.3% of the lack of successful classification could be attributed to the method and/or analyst's variation in marking. LDA had a lower classification accuracy in comparison to cosine similarity, which indicates that a centroid-based classifier may perform more poorly when investigating the changes in group membership with continued wear (when general/specific wear and Schallamach patterns are ignored). Accuracy in classification at the first rank position was below 75% until 160K steps of cumulative wear. Petraco *et al.* (2010) [7] shared similar findings, where it was reported that lightly worn outsoles with with fewer RACs were difficult to differentiate in PC-space, while more heavily worn shoes demonstrated increased separability.

4.5 Further Investigation

To strengthen the experimental approach outlined in this work, and to corroborate or amend conclusions, four future investigative efforts are proposed. First, it was difficult to identify a metric which could capture minute changes of RACs with continued wear. Therefore, a similarity metric which is able to capture small changes in geometry should be further investigated. Second, the results are highly impacted by the researcher's ability to mark features with high repeatability. Global correlation between mated QC markings was 0.73 on average, with a minimum of 0.41 and a maximum of 0.94. Future studies might consider comparing image "patches" that reproduce identified RACs along side neighboring tread elements, and then asking trained forensic footwear examiners to compare the similarity between patches that differ in wear. A dataset that comprised a mixed set of these pairwise comparisons, although not able to numerically quantify observed changes, could instead define the range of change deemed acceptable (by wear) within the community, and the step-interval(s) beyond which changes are deemed too different to allow for an association. Third, findings presented in this study are limited to this dataset, and therefore future work is needed to determine which results are specific and which are generalizable. In other words, to form a meaningful baseline regarding the expected change in RACs with continued wear in a controlled environment (*e.g.,* same shoes worn, terrain traversed, activity performed), this type of study (barring suggested improvements in methodology) should be repeated to determine if the

findings are robust. As an extension of this work, different terrains should also be included. Lastly, the impact of wearer-attributes on observed results should be explored.

Despite the noted limitations and suggestions for improvement in future studies, the results of this research characterize the changes in global and local (RAC-level) similarity that can be encountered from test-impressions that differ in general and accumulated wear. Since exemplars created from seized shoes can exhibit varying degrees of wear and continued use, this variation must be characterized and accounted for when comparing known test-impressions to impressions deposited at crime scenes. This research provides a point-of-reference for the range of changes that can be observed in an opportunistic dataset, and future studies are expected to clarify which observed results are generalizable to other populations.

5. Appendices

5.1 Pedometer Reliability

Pedometers were worn for the purpose of testing reliability following heavy and continuous use. A total of 10 trials were conducted for each pedometer, where each trial was comprised of 40 manually counted steps traversed in a straight line. After this desired step count was reached for each trial, the pedometer was read and recorded. Results across the 10 trials for 7 of the pedometers used over the course of the continued wear study are displayed in Table 5.1 (one pedometer was unacquirable from a participant). The pedometer column corresponds to assigned identifiers mated to pedometers for tracking purposes, the $X \pm SD$ column corresponds to the average step count and the standard deviation, the minimum column corresponds to the lowest step count across trials, and the maximum column corresponds to the highest step count across trials.

Table 5.1: Pedometer reliability summary data. The columns represent an arbitrary pedometer ID, the average and one standard deviation of steps, and finally, the minimum and maximum count across all 10 trials.

Pedometer	$X \pm SD$	Minimum	Maximum
1	39.1 ± 1.51	36	41
2	40.6 ± 0.809	40	42
3	40.3 ± 0.467	40	41
4	40.5 ± 0.688	40	42
5	40.3 ± 0.467	40	41
6	40 ± 0	40	40
7	40.9 ± 0.831	40	42

5.2 Participant Step Count

Although precautions were put into place to safeguard against traveling over the desired step count in a step-interval (*e.g.,* swapping for different shoes when reaching 10K steps, walking in groups for accountability, etc.) a few instances still occurred where the desired step count was exceeded. Table 5.2 depicts self-reported summary data regarding step count for participants across all step-intervals. The table columns report participants' alpha-numeric identifier, pedometer usage (some participants opted to use a personal step count tracker instead of the supplied pedometer), the average step count and associated standard deviation, and finally, the lowest and highest step count self-reported across all step-intervals.

Table 5.2: Self-reported summary data for participant step count. Columns represent participant ID, the anecdotally self-reported percentage time the participant used the supplied pedometer (as opposed to a personal tracker), the average step count and one standard deviation, and finally, the minimum and maximum counts reported.

Participant	% Pedometer	$X \pm SD$	Minimum	Maximum
q7vhg	100	20013.4 ± 39	20000	20123
41evv	50	20001.4 ± 3.4	20000	20011
is4dv	100	20000.2 ± 0.63	20000	20002
bty4d	100	20034.5 ± 94	20000	20300
9asbt	100	20000.3 ± 0.95	20000	20003
u5ynj	100	20000.3 ± 0.95	19999	20002
o55y9	95	20002.8 ± 5.2	20000	20013
1ws3y	50	20001.0 ± 1.3	20000	20003

5.3 Participant Attributes

Participant attributes were recorded at the onset of the study. Participant information was de-identified, and associated physical attributes are reported in Table 5.3, where columns correspond to a random participant identifier, gender, age, height, and weight.

Table 5.3: Columns correspond to a de-identified participant label, gender, age, height, and weight.

Participant	**Gender**	**Age**	**Height**	**Weight (lbs)**
A	F	25	5'8"	159
B	F	21	5'6"	181
C	F	21	5'6"	150
D	M	22	5'10"	171
E	M	20	5'11"	146
F	F	22	5'4"	140
G	F	29	5'4"	139
H	M	22	5'11"	162

5.4 RAC Gain/Loss Data

The count of acquisition and loss was parsed into respective step-intervals (1-10) and participant's data (unique IDs) as shown in Table 5.4 and 5.5. Each count reflects the difference(s) observed from one step-interval in comparison with the subsequent step-interval (*i.e.*, "step-interval 2" represents gain and loss when comparing RAC count from "step-interval 1" to "step-interval 2").

Table 5.4: Count of gain and loss per-participant for each step-interval. This table reports the count data for gain and loss of RACs across step-interval one (20K steps) through five (100K steps) for each individual shoe assigned a random identifier, where "L" denotes left shoe and "R" denotes right shoe.

Unique ID/Step-Interval	1		2		3		4		5	
	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss
AL	2	0	4	1	0	2	6	2	3	5
AR	7	0	10	5	7	7	1	12	11	2
BL	0	0	7	0	2	4	10	6	1	10
BR	1	2	8	1	5	5	3	7	2	5
CL	2	0	1	0	0	0	6	2	3	1
CR	0	0	1	0	2	0	3	1	4	2
DL	4	5	10	4	3	8	2	2	2	0
DR	5	0	10	4	4	5	1	9	2	0
EL	0	0	0	0	0	0	1	0	1	0
ER	2	2	1	0	0	0	3	1	1	2
FR	0	0	0	0	5	0	3	1	4	1
FR	0	0	0	0	1	0	5	0	1	0
GL	3	2	6	1	0	6	1	4	0	0
GR	0	1	2	0	1	0	1	1	1	0
HL	1	0	6	0	2	1	2	6	12	0
HR	3	0	1	3	0	0	0	1	1	0

Table 5.5: Count of gain and loss per-participant for each step-interval. This table reports the count data for gain and loss of RACs across step-interval six (120K steps) through ten (200K steps) for each individual shoe, where "LS" denotes left shoe and "RS" denotes right shoe.

Unique ID/Step-Interval	6		7		8		9		10	
	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss	Gain	Loss
AL	6	1	4	5	2	1	3	4	1	2
AR	5	0	3	3	9	1	4	5	3	2
BL	2	0	1	1	2	0	3	2	1	1
BR	6	0	2	5	2	0	3	1	3	2
CL	3	1	8	7	14	10	5	5	4	7
CR	12	3	5	4	5	3	4	6	6	8
DL	2	2	5	4	3	2	1	2	2	4
DR	0	0	6	1	1	2	7	3	7	7
EL	0	1	6	0	4	2	3	0	8	7
ER	16	1	4	4	3	12	3	3	7	2
FL	3	4	6	2	12	9	5	13	7	2
FR	5	2	6	1	4	4	2	6	4	4
GL	1	0	0	1	6	0	4	4	13	2
GR	2	1	0	3	3	0	0	0	1	0
HL	7	3	7	7	4	1	9	5	5	6
HR	11	0	9	1	13	3	12	9	14	2

5.5 RAC Data Set Adjustment

Part-way through the course of the study it was realized that some instances of outsole texture had been labeled as RACs due to the examiner being a novice at the onset of data collection. Associated step-intervals and the corresponding alpha-numeric identifiers where these occurrences potentially took place were recorded for later corrective action, which would entail specific shoes and step-intervals to be re-analyzed. The first four step-intervals (zero to three) for alpha-numeric identifiers q7vhg, 9asbt, bty4d, 1ws3y, and u5ynj and the first two step-intervals (zero to one) for alpha-numeric identifiers o55y9, is4dv, and 41evv were investigated for erroneously marked RACs.

To conduct the corrective action, the patch map with RACs of the step-interval of interest was overlaid with the mated Eigen-registered outsole of the same step-interval using Adobe Photoshop Elements 10®. Patch map layer visibility was toggled on and off and each originally marked RAC was visually inspected on the Eigen-registered outsole. If the detail observed on the outsole was identified as a RAC, the corresponding RAC originally marked in this position was left unchanged. If the detail observed on the outsole was identified as texture (rather than a RAC), the corresponding RAC originally marked in this position was flagged for removal and documented. If the detail observed on the outsole was too low of a resolution to form a conclusion (such as the extreme toe and heel where the outsole curved away from the scanner glass plate), the corresponding RAC originally marked in this position was labeled as inconclusive. Inconclusive RACs were further examined in an effort to arrive at a definitive conclusion through inspection of preceding step-intervals. If the resolution was sufficient in a previous step-interval, the aforementioned criteria was utilized to arrive at a conclusion. If the resolution remained insufficient, mated RAC presence was searched

for in the earliest step-interval where marking of texture was eliminated. If the mated RAC was present, then it was in fact a real RAC and was left unchanged. If the mated RAC was not present, a conclusion could not be reached and remained as inconclusive. This entire decision process was performed for all shoes and the results from the corrective action can be seen in Table 5.6. The participant column reports alpha-numeric identifiers, the total column reports the total RAC count that was verified, the removed column reports the count of RACs removed from the data set, and the unresolved column reports the count of RACs that remain inconclusive.

Table 5.6: Corrective action results. Columns represent participant ID, the total number of RACs visually inspected, the count of RACs removed, and the number of inspected RACs that remained inconclusive or unresolved.

Participant	Total	Removed	Unresolved
1ws3yL	18	9	0
1ws3yR	60	17	4
9asbtL	26	6	4
9asbtR	28	4	7
bty4dL	35	6	1
bty4dR	48	14	0
q7vhgL	37	8	2
q7vhgR	7	0	4
u5ynjL	21	4	1
u5ynjR	10	5	0
is4dvL	0	0	0
is4dvR	6	0	0
41evvL	2	0	0
41evvR	0	0	0
o55y9L	0	0	0
o55y9R	0	0	0

Through use of the aforementioned criteria, a total of 73 RACs were removed after being identified as texture (out of a possible 285) which reflects a 25.3% reduction in the data set. Inconclusive RACs are documented, but since they cannot be conclusively identified for removal, remain in the dataset. This may result in over-reporting of changes. Conversely, if inconclusive RACs were removed from the data set and were real, expected differences in acquisition would be under-reported.

5.6 Geometric Evolution Analytical Approaches

A multitude of different analytical approaches were used to evaluate change in RAC geometry through wear. Originally, Equation 5.1 was evaluated for mated RACs that persisted across step-intervals. Examples of %A values of 25% and 75% can be seen in Figure 5.1, where RACs A and C correspond to initial RACs, and RACs B and D correspond to the features mated to RACs A and C after 20K steps of additional wear for newly worn shoes and more heavily worn shoes, respectively. The geometric change between RACs with the aforementioned thresholds is illustrated,

as well as the generally poor performance of %A as a similarity metric for linear RACs through continued wear. Consequently, %A was discarded as a viable metric to capture similarity of RACs as a function of wear.

$$\%A(1,2) = \frac{[\text{area of overlap}] \times 2}{[\text{area of 1}] + [\text{area of 2}]} \times 100 \tag{5.1}$$

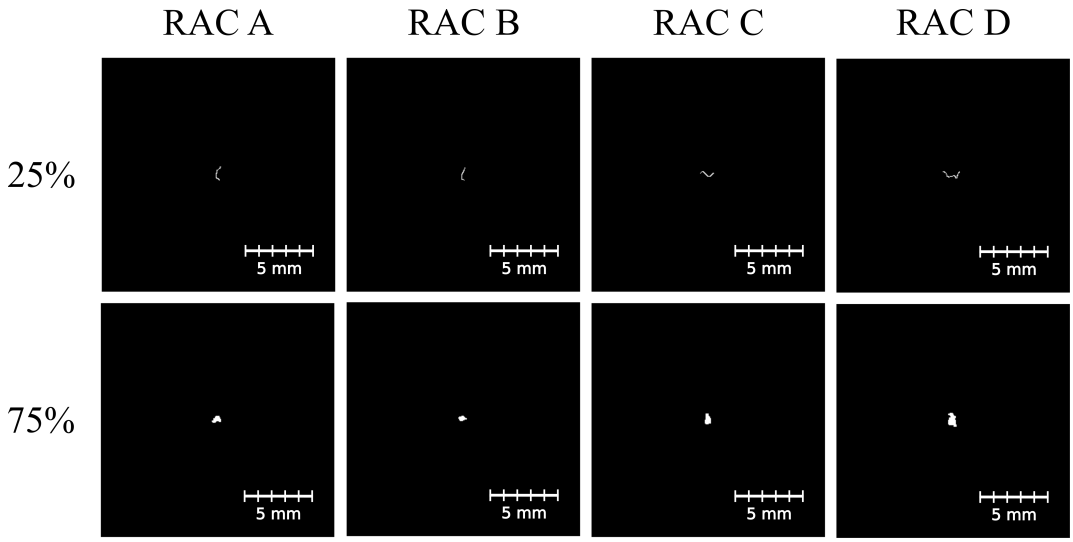

Figure 5.1: RACs with %A similarities of 25% and 75%. Images correspond to RACs with %A similarities of 25% and 75% for comparisons between RACs on newly worn shoes (A and B) and more heavily worn shoes (C and D).

Next, a visual comparison of RAC similarity was attempted by creating a graphical user interface (GUI) to display Prewitt Fourier descriptors (PF-FDs) of RACs which were mated through continued wear. The GUI displayed each RAC's PF-FD, and an overlaid version of the presented PF-FDs. A total of five raters were asked to choose the degree of wear for which the observed dissimilarities between the RAC pair could be attributed to. Four choices were provided to the raters: "Explainable by Negligible Wear," "Explainable by Minor Wear," "Explainable by Considerable Wear," and "Non-Explainable." Every 10th pair was randomly sampled from previous pairs, and redisplayed to serve as quality control pairs. For example, the 20th pair was randomly sampled from the first 19 image pairs. A subset of 400 out of 1,600 image pairs was used to assess the suitability of the proposed method, which was composed of 326 non-repeated image pairs and 36 repeated QC pairs. Intra-rater inconsistency was computed using QC pair responses as described in [24] using Equations 5.2 - 5.3, where p and q correspond to image set ratings, L corresponds to the number of images (L = 36 for this dataset), and c corresponds to a participant A-E. The IRI_c (which ranges from 0 to 1) was found to be 0.33, 0.44, 0.60, 0.62, and 0.80 (F = 36) for participants 1-5, respectively. Inter-rater reliability was computed with Krippendorff's Alpha coefficient using the `irrCAC` package in `R` with all non-repeated image pair ratings, and was found to be k-α = 0.536 (not benchmarked). It has been previously reported that a k-α 0.41 - 0.60 is considered moderate agreement [25], which was considered inadequate.

$$\text{cost(p,q)} = \begin{cases} 0: & \text{no swap} \\ 1: & \text{swaps } (1 \leftrightarrow 2),\ (3 \leftrightarrow 4) \\ 2: & \text{swaps } (2 \leftrightarrow 3) \\ 3: & \text{swap } (1 \leftrightarrow 3),\ (2 \leftrightarrow 4) \\ 4: & \text{swaps } (1 \leftrightarrow 4) \end{cases} \tag{5.2}$$

$$IRI_c = \sqrt{\frac{1}{L}\sum_{d=1}^{L} cost(p_{cd}, q_{cd})} \tag{5.3}$$

Due to a generally poor intra-rater inconsistency and a lack of consensus between responses, the visual comparison approach was deemed unsuitable. Since %A was not considered the best metric to describe linear features, an additional six numerical metrics were evaluated and compared with %A. Results indicated that average Hausdorff distance was superior, and was used for all remaining comparisons of RAC geometry.

5.7 AUC Pairwise Significance Testing

Figure 5.2 reflects the results of pairwise testing for statistically significant differences ($\alpha = 0.05$) between the classification performance of HA for each step-interval difference. The x-axis and the y-axis are step-interval differences, and each data point corresponds to a statistically significant (blue shading) or non-significant (red shading) difference between AUCs describing step-interval differences.

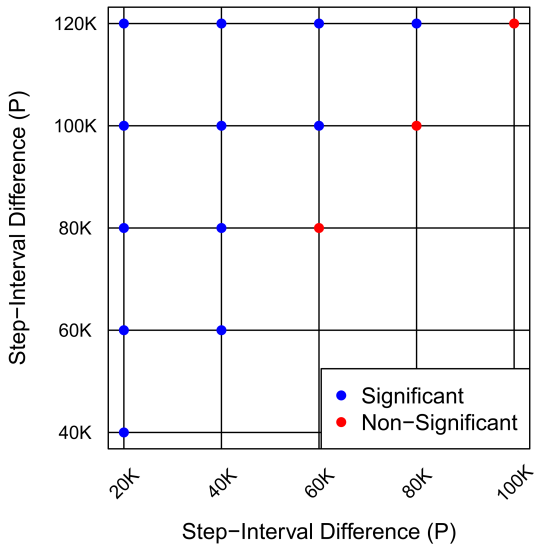

Figure 5.2: Step-interval difference area under the curve (AUC) pairwise significance testing. The x-axis and y-axis correspond to step-interval differences. Each observation indicates a statistically significant (at $\alpha = 0.05$) (blue shading) or non-significant (red shading) difference between AUCs describing step-interval differences. Pairwise testing which includes step-interval differences of P=140K - 180K were removed due to low power.

Figure 5.3 reflects the results of pairwise testing for statistically significant differences ($\alpha = 0.05$) between the classification performance of HA for two consecutive cumulative step-intervals against a differing two consecutive cumulative step-intervals. The x-axis and the y-axis are consecutive cumulative step-intervals, and each data point corresponds to a statistically significant (blue shading) or non-significant (red shading) difference between AUCs describing cumulative step-intervals.

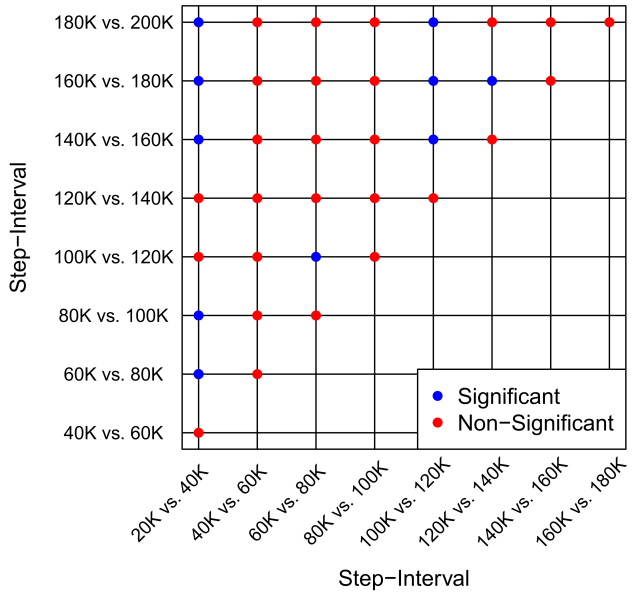

Figure 5.3: Cumulative step-interval area under the curve (AUC) pairwise significance testing. The x-axis and y-axis correspond to comparisons between cumulative step-intervals and the subsequent ones (F vs. F+1). Each observation indicates a statistically significant (at $\alpha = 0.05$) (blue shading) or non-significant (red shading) difference between AUCs describing cumulative step-intervals.

5.8 Quality Control Data

Table 5.7: Columns correspond to the identifier of a repeated marking of an image pair, the 2-dimensional correlation coefficient (r) computed between the mated RAC maps, the number of RACs in the original marking (Count A), the number of RACs in the repeated marking (Count B), the number of additional RACs marked in the repeat ("#New"), the number of RACs not detected in the repeat ("#Dropped"), and the number of RACs constant between the pair ("#Constant").

Repeated Marking	r	**Count A**	**Count B**	**#New**	**#Dropped**	**#Constant**
1	0.940	6	6	0	0	6
2	-	0	0	-	-	-
3	0.674	8	11	4	1	7
4	0.802	9	12	4	1	8
5	-	0	0	-	-	-
6	0.839	4	4	0	0	4
7	0.777	7	5	0	2	5
8	0.839	3	3	0	0	3
9	0.682	8	8	2	2	6
10	0.907	6	6	0	0	6
11	0.572	11	10	2	3	8
12	0.829	16	16	2	2	14
13	0.778	13	12	3	2	11
14	0.775	14	15	2	1	13
15	0.877	9	9	0	0	9
16	0.632	11	10	2	3	8
17	0.537	16	21	9	4	12
18	0.563	17	18	7	6	11
19	0.407	7	7	1	1	6
20	0.646	14	12	3	5	9

5.9 Empty Set Vectors

Table 5.8: Columns correspond to the cumulative step-interval (F), the number of step-intervals (I) within F, the number of participants (L), the number of empty set vectors ($\emptyset Q$), and the number of non-empty set vectors (Q).

F	**I**	**L**	**$\emptyset$Q**	**Q**
20K	1	16	5	(1 x 16) - 5 = 11
40K	1	16	3	(1 x 16) - 3 = 13
60K	1	16	1	(1 x 16) - 1 = 15
80K	1	16	1	(1 x 16) - 1 = 15
100K	1	16	0	(1 x 16) - 0 = 16
120K	1	16	0	(1 x 16) - 0 = 16
140K	1	16	0	(1 x 16) - 0 = 16
160K	1	16	0	(1 x 16) - 0 = 16
180K	1	16	0	(1 x 16) - 0 = 16
200K	1	16	0	(1 x 16) - 0 = 16

Table 5.9: Columns correspond to the step-interval difference (P), the number of cumulative step-intervals (I) within P, the number of participants (L), the number of empty set comparisons ($\emptyset R$), and the number of non-empty set comparisons (R).

P	**I**	**L**	**$\emptyset$R**	**R**
20K	9	16	10	(9 x 16) - 10 = 134
40K	8	16	10	(8 x 16) - 10 = 118
60K	7	16	10	(7 x 16) - 10 = 102
80K	6	16	10	(6 x 16) - 10 = 86
100K	5	16	10	(5 x 16) - 10 = 70
120K	4	16	10	(4 x 16) - 10 = 54
140K	3	16	9	(3 x 16) - 9 = 39
160K	2	16	8	(2 x 16) - 8 = 24
180K	1	16	5	(1 x 16) - 5 = 11

Bibliography

[1] Bodziak WJ. Footwear Impression Evidence: Detection, Recovery and Examination. CRC Press; 1999.

[2] Davis RJ, DeHaan JD. A survey of men's footwear. Journal of the Forensic Science Society. 1977;17.4:271-85.

[3] Damary NK, Mandel M, Wiesner S, Yekutieli Y, Shor Y, Spiegelman C. Dependence among randomly acquired characteristics on shoeprints and their features. Forensic Science International. 2018;283:173-9.

[4] Richetelli N, Speir J, Jackson G, Jelsema C. Statistical evaluation of randomly acquired characteristics on outsoles with implications regarding chance co-occurrence and spatial randomness [Ph.D. Thesis]. West Virginia University; 2020.

[5] Adair TW, Lemay J, McDonald A, Shaw R, Tewes R. The Mount Bierstadt study: an experiment in unique damage formation in footwear. Journal of Forensic Identification. 2007;57.2:199-205.

[6] 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.

[7] Petraco ND, 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.

[8] Pawloski S. The accumulation of wear on footwear pattern analysis. Themis: Research Journal of Justice Studies and Forensic Science. 2019;7:1-20.

[9] 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(1-3):84-91.

[10] Wyatt JM, Duncan K, Trimpe MA. Aging of shoes and its effect on shoeprint impressions. Journal of Forensic Identification. 2005;55.2:181-8.

[11] Cassidy MJ. Footwear Identification. Ottawa: Public Relations Branch of the Royal Canadian Mounted Police; 1980.

[12] Speir JA, Richetelli N, Fagert M, Hite M, Bodziak WJ. Quantifying randomly acquired characteristics on outsoles in terms of shape and position. Forensic Science International. 2016;266:399-411.

[13] Amer GMH, Abushaala AM. Edge detection methods. IEEE. 2015;2:1-7.

[14] Richetelli N, Nobel M, Bodziak J, Speir J. Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape. Forensic Science International. 2017;270:211-22.

[15] Schott J. Remote Sensing: The Image Chain Approach. Oxford Univerisity Press; 2007.

[16] Huttenlocher D, Klanderman G, Rucklidge W. Comparing images using the Hausdorff distance. IEEE Transactions on pattern analysis and machine intelligence. 1993;15(9):850-63.

[17] Gregga J, Power G, Iftekharuddin K. Correlators for rank order shape similarity measurement. Optical Pattern Recognition XII. 2002;4734:122-31.

[18] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77-85.

[19] DeLong E, DeLong D, Clarke-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

[20] Russ J, Matey J, Mallinckrodt J, McKay S. The Image Processing Handbook. vol. 8. Computers in Physics; 1994.

[21] Silverman B. Density Estimation for Statistics and Data Analysis. vol. 26. CRC press; 1986.

[22] SWGTREAD. Range of Conclusions Standard for Footwear and Tire Impression Examinations; 2013. Available from: `http://www.swgtread.org/images/documents/standards/published/swgtread_10_conclusions_range_201303.pdf`.

[23] Ryan T. Sample Size Determination and Power. John Wiley and Sons; 2013.

[24] Chugh T, Cao K, Zhou J, Tabassi E, Jain AK. Latent fingerprint value prediction: crowd-based learning. IEEE Transactions on Information Forensics and Security. 2017;13(1):20-34.

[25] Altman D. Practical statistics for medical research. CRC press. 1990.