# Interpol review of shoe and tool marks 2016-2019

Martin Baiker-Sørensen*, Koen Herlaar, Isaac Keereweer, Petra Pauw-Vugts, Richard Visser

Section of Firearms and Tools, Division of Chemical and Physical Traces, Netherlands Forensic Institute (NFI), Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands

## ARTICLE INFO

## ABSTRACT

This review paper covers the forensic-relevant literature in shoe and tool mark examination from 2016 to 2019 as a part of the 19th Interpol International Forensic Science Managers Symposium. The review papers are also available at the Interpol website at: https://www.interpol.int/content/download/14458/file/Interpol%20Review%20Papers%202019.pdf.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

This chapter provides an overview of articles and theses relevant to shoe- and toolmark examiners, which were published between May 2016 and December 2018, and is the sequel to the review for the 18th Interpol International Forensic Science Managers Symposium in October 2016 by Martin Baiker-Sørensen (Baiker, 2016) [1] (available online (Interpol, 2019) [2]. It is based on relevant articles from all volumes of a selection of forensic journals (American Journal of Forensic Medicine and Pathology, Forensic Science International, International Journal of Legal Medicine, Journal of Forensic and Legal Medicine, Journal of Forensic Identification, Journal of Forensic Sciences, Legal Medicine, Science and Justice, The Australian Journal of Forensic Sciences, AFTE Journal and Forensic Science, Medicine and Pathology) of that period as well as peer reviewed articles from other journals and otherwise relevant sources, which were found by browsing the internet.

### Disclaimer

The authors and the Netherlands Forensic Institute do not intend to advocate any commercial software or system mentioned in this collection of reviews but rather give examples of systems that are available and that the authors are aware of.

The traditional way of shoe- and toolmark examination includes subjectivity in the process. This may lead to variation in the conclusions of different examiners. In recent years, the demand for more objective approaches that lead to less variability of the outcome of an examination, that rest upon a sound scientific base and that have a strong statistical underpinning, has increased. In the United States this was expressed in a report of the National Academy of Sciences (NAS), published in 2009 [3] and more recently in a report of the President's Council of Advisors on Science and Technology (PCAST), published in 2016 [4].

In the United States, this lead to the appointment of the National Commission on Forensic Science hosted by the Department of Justice and the Organization for Scientific Area Committees (OSAC) for Forensic Science at the National Institute of Standards and Technology (NIST). The OSAC groups are "a collaborative body of more than 500 forensic science practitioners and other experts who represent local, state, and federal agencies; academia; and industry" and their goal is to "support the development and promulgation of forensic science consensus documentary standards and guidelines, and to ensure that a sufficient scientific basis exists for each discipline." The relevant OSACs for shoe-, tire- and toolmark examiners are the OSAC Footwear and Tire Subcommittee [5] and the OSAC Firearms and Toolmarks Subcommittee [6], which are both part of the OSAC Physics/Pattern Interpretation Scientific Area Committee [7]. Based on the documents of the previous Scientific Working Groups (some documents are still online [8]), the OSAC working groups are developing best practice guidelines, of which the first were recently published and some are in an advanced phase of being approved at the time of writing.

In our last review we elaborated on the many ways to render daily casework more objective and to statistically underpin the interpretation of evidence [1] and following approved standards and guidelines as provided by the OSACs in casework is an essential step. In addition, there are many examiners and researchers that strive to render their field more objective by improving traditional methods and providing a statistical foundation, by developing new

---

* Corresponding author.
  E-mail address: m.baiker@nfi.nl (M. Baiker-Sørensen).

methods or by introducing new technologies into their labs, to supplement the examiner's findings with objective measures. In the following sections, the highlights, trends and details in the various disciplines covered in this review are presented.

## 1.1. Structure of this review article

In the following sections, first the highlights and trends in the various disciplines covered in this review are summarized. After that, more details on the publications in each discipline are given, subdivided into sections covering shoemarks, evidence comparison based on physical properties and striated and impression toolmarks. The latter was further split into publications about conventional and invasive toolmarks respectively, as the latter is a fairly new but very exciting branch of examination. Although it is still in its infancy, it has great potential. Progress in the past was mainly hindered by the fact that invasive traumas are available to the forensic pathologist/anthropologist and a certified toolmark examiner is often not on the premises. Hopefully this will change in the future, as there might be much more information hidden in invasive traumas that are perceived to date and thus a more close collaboration between examiners of different disciplines might yield very promising results.

The subsections follow the steps of toolmark examination in practice and, as the strategy of assessing evidence is similar for different disciplines, the structure of the individual sections is roughly the same. As there are many contributions describing methods for automated comparison of shoe- and toolmarks, sections focusing on either of the two are given separately. Alongside the development of automated methods, several groups integrated their approaches into software in the form of graphical user interfaces (GUIs) during the last years, which enable examiners to test the methods with their own data. These are presented in a separate section as well. Note that the articles are categorized based on their main focus, but parts may also be presented in other sections. If for instance an article describes a novel and interesting way of data acquisition, while that is not the main focus of the article, the article will be discussed in several sections.

## 2. Highlights and trends

### 2.1. Shoemarks

Shoemarks are commonly encountered at crime scenes but are often neglected, although they occur much more frequently than e.g. DNA or fingerprint evidence. One reason is that DNA evidence is typically compared automatically and thus objectively, while shoemark analysis in most labs is done in the traditional manner by human examiners. So the field would greatly benefit from an "injection of technology and associated modern analytical tools" [9], by using automated comparison based on more objective data and more objective analysis methods. Mainly two developments in recent years are very promising in this respect. The first is that the technology for 3D shoe impression mark acquisition is close to being applicable in practice. There are systems available now that have proven to satisfy the varying requirements at a crime scene like lighting conditions and substrate materials [10]. In addition, the techniques are nowadays providing a high level of detail. Using this new technology in casework would be a great improvement, as it has been shown that 3D data has superior properties compared to conventional 2D photography data and that "identification can be obtained in a higher percentage of cases" [11,12]. There are also several software solutions available nowadays, some free of charge, that can be used to visualize and compare data manually [13,14]. Another exciting development is that algorithms for 2D shoemark

retrieval, i.e. finding the make and model of a shoe in a database, are getting close to the transition from being purely academic to being useful in practice. Unlike in our last review three years ago, there are several algorithms around now that show high retrieval performance, even when confronted with potentially bad quality data from crime scenes [15–18], including partial marks and marks on strongly varying backgrounds. In addition, authors compared their algorithms frequently with those of others. Unfortunately that did not always happen using publicly available databases and therefore it is still difficult to objectively compare the best performing algorithms. Authors should therefore use publicly available and representative databases including real crime scene marks. There are also some commercial systems available now for automated shoemark retrieval [19,20]. So far all presented automated retrieval systems are using 2D images of shoemarks, but in the future it might be interesting to include impression evidence as well by also supporting 3D surface data.

Regarding the interpretation of shoemark evidence the recent literature is limited. Only one publication was found that aimed at calculating a likelihood ratio (LR) and in that work a limited database was used [21]. As the publications that studied the frequency of occurrence and the distribution of randomly acquired characteristics (RACs) used varying amounts of categories of RAC shapes and shape definitions [22–24], it is difficult to compare the results. Rather than choosing a range of user defined categories it would be useful to find out, which RAC shapes are actually the most relevant and distinctive. In addition, the RAC shapes were typically determined manually and it was shown that on the one hand, repeated prints made under controlled conditions in the lab might contain varying amounts and appearance of RACs [25] and on the other hand, annotations by different examiners are not always consistent [22]. Therefore there is a need for standardization of RAC shapes, in order to compare studies to determine frequencies, co-occurrence and distribution, as well as the way a shoe sole is to be subdivided for local RAC frequency assessment. There is also further research required that studies the degree of dependence between individual RACs as it has been shown, that the independence assumption originally proposed by Stone [24], may not be valid [23].

### 2.2. Striated and impression toolmarks

In our last review article, we discussed in detail the various possibilities to render casework more objective by using 3D surface data of toolmarks and presented a large range of imaging technologies like confocal microscopy, focus variation or photometric stereo [1]. In many pilot studies it had been shown that those technologies have the potential to accurately acquire 3D data of toolmarks. In recent years, forensic laboratories are working on integrating 3D surface metrology into actual casework and it is crucial that this process is accompanied by a quality assurance guideline to make sure that the devices are performing according to specification and are evaluated in practice. This means that acquisition hardware has to be calibrated, ideally using standardized specifications [26] and physical measurement standards, to test the technical specifications under known circumstances. In addition, the devices have to be evaluated for the specific forensic applications with samples that are representative for casework [27]. One example of a system that has been fully validated [28] is now being used in actual casework by the FBI. In combination with a Virtual Comparison Microscope [29], it is used to manually compare cartridge case impressions and aperture shear marks. The system was developed to compare firearm evidence, but a modified system could also be used for toolmarks.

Automated toolmark comparison methods in the past often

used one score to determine toolmark (dis-)similarity and methods were tested with experimental marks generated in the lab. In recent years two methods were proposed that use a multi feature similarity score [30] or a convolutional neural network [31]. Using multi feature vectors instead of choosing a single similarity metric is attractive, as several features of variable 'strength' might yield better performance if combined. However, this strategy also requires a reliable feature detector rather than using the data as it is. Neural networks have the advantage that it is not necessary to explicitly define a measure of similarity at all, but rather let an algorithm decide which features are valuable to distinguish between matching and non-matching marks. The disadvantage is that typically a lot of data is needed to properly train a neural network. Nevertheless it will be interesting to see further developments in this direction and the results of testing the algorithms with large datasets. Particularly test datasets should include crime scene quality data as well. In addition, the vast majority of algorithms focused on striated toolmarks only and not on impression marks. To our knowledge there is only one algorithm that was developed for, and tested on impression toolmarks. The method uses the mark contour and local appearance [32]. The real crime scene data used for this study is available online. Clearly there is a need for algorithms that are designed for comparing tool impression marks and that are tested using realistic data. Such algorithms do exist for firearms impressions [33,34], but they have not been tested on toolmark impressions.

Reliably weighing evidence within an LR framework requires an available population database of representative marks. If different laboratories however use a different population database for the analysis, the statistical results of comparing the same pieces of evidence could differ. Therefore the NIST, the FBI and the NFI started a collaboration project in 2018 that aims at developing the infrastructure for a centralized and permanently monitored database of striated and impression marks. To this end, the existing infrastructure of the software package 'Scratch', which was developed at the NFI to automatically compare striated toolmarks, build local databases and calculate LRs, was extended with additional functionality to automatically compare impression marks. In addition, the database functionality was greatly extended, to allow setting up a large and diverse database of striated and impression marks (Reference Population Database of Firearm Toolmarks or RPDFT [35]). In the future, a great variety of similarity metrics, determined with a consistent background population database with all relevant meta information, will be available to third parties. For now, the main focus is on firearm marks, but the framework will be versatile enough to host toolmarks as well in the future.

### 2.3. Invasive striated and impression toolmarks

A large body of literature was published in the field of invasive toolmark examination. As it is still a relatively new field, in most of the articles the variability and individuality of marks created with saws, knives and other sharp objects were studied. Furthermore all articles were focusing on class characteristics (shape parameters) rather than individual characteristics, because the latter are rarely encountered in bone marks and are difficult to capture. Data acquisition was typically done using conventional 2D microscopy, however by most authors only for qualitative purposes. For quantitative analysis, authors more frequently used 3D volume imaging techniques, especially Micro-CT. This technique was recently shown to provide more observable detail than 2D microscopy [36]. However, only one publication compared Micro-CT to measurement standards [37] to see whether the CT measurements were accurate. More focus should be put on this issue in the future. An interesting development is the usage of morphometrics to quantify

shape features and a subsequent statistical analysis to determine differences between e.g. different tools. Most authors determined shape features manually however using dedicated software packages. It would be desirable to determine shape features automatically as different examiners might interpret the data differently and shape descriptors may be ambiguous. Furthermore, shape descriptors vary greatly between publications. Rather than choosing shape descriptors manually it should be studied which descriptors are the most powerful to distinguish between different types of tools. Subsequently, authors should use similar descriptors. Another reason why study results are not easy to compare is the fact that the experimental setups vary, e.g. with respect to bone species (human vs. a variety of animals), bone state (fleshed, defleshed, macerated, frozen) and the way of applying the marks (controlled with a machine or by hand). As most of these factors have shown to have an influence, the most realistic scenario should be chosen to create experimental marks, thus using human, fleshed bones if possible and applying marks by hand and with realistic force.

Regarding the specificity of marks with respect to different types of tools, some studies conclude that serrated and non-serrated knives can be distinguished [38–40], while others don't [41]. Studies focusing on saws typically conclude that different saw types, thus hand saws, hacksaws and reciprocating saws, can be distinguished [42–47]. However, only one author also compared the shape features of the marks with the actual tool that was used to create it [44]. This is an essential step though and should get more attention in research projects in the future. On a whole the combination of 3D imaging (Micro-CT) in combination with morphometrics and statistical analysis of the measurement results is a great step, but the conditions under which experiments are conducted should be closer to realistic conditions and there should be a consensus regarding useful shape parameters for saw and cut marks in bone.

## 3. Shoemarks

Many steps have to be taken to compare shoemarks in the lab, from documenting the marks at the crime scene to creating and acquiring experimental shoeprints in the lab with suspect shoes to the comparison of the shoe sole characteristics and finally the interpretation of the examination results. A good book [9] to give an overview of the whole process of shoemark examination can be found in Bodziak [11], a book "to share the authors foundation of knowledge and experience to provide novice and experienced examiners, crime scene technicians, investigators, prosecution and defense counsel with a comprehensive source of information on forensic footwear evidence". Included are chapters on detection, acquisition, enhancement, casting and examination of footwear evidence. In addition several production steps from design to molding and cutting are discussed. Furthermore, the book contains case examples. In addition, an overview of standard operating procedures for shoe and toolmarks can be found in another book by Petraco et al. [48]. It includes the whole pipeline from the crime scene to the examination. These two books mainly focus on the more traditional approaches to forensic shoemark analysis. A book that contains chapters on the acquisition and comparison of marks as well as the interpretation of the results which also includes recent technological advances was published by Bennett and Budka [9]. Their focus lies mainly in the field of ichnology, which is the study of geological records like dinosaur foot impression marks, but techniques for e.g. data acquisition are similar and thus might be useful for the field.

The following chapters provide an overview of recent publications regarding one or several steps of forensic shoemark

examination. As there seems to be some discrepancy between the terminologies used by different authors, we in this review article stick to the following: For 2D representations of a shoe created in the *lab*, we use *shoeprint* and for those found at a *crime scene*, we use *shoemark*. For 3D representations of a shoe created in the *lab*, we use *shoe imprint* and for those found at a *crime scene*, we use *shoe impression mark*. In general we use *shoemarks* to address shoe evidence, if the context is not relevant.

### 3.1. Detection/creation

Creating detailed and complete experimental shoeprints in the lab is an essential step in the daily work of the forensic shoemark examiner, and care should be taken to create prints as consistent as possible to reduce variability of characteristics among prints. In addition, shoeprints should be an accurate copy of the shoe outsole, including class characteristics and RACs. Whitlow studied different ways to create experimental shoeprints in the lab [49] and compared them based on the amount and visibility of RACs. To this end, prints of two types of shoes, worn work boots and sneakers, were acquired with two different methods, the Identicator inkless shoeprint system and using Handiprint lifting material with black fingerprint powder. For both methods, three ways of creating a shoeprint were considered, dynamic step, static step and rolled. From images of the prints, RACs were located manually and validated using the shoe outsoles. Finally, the fractions of RACs present, considering the previously mentioned conditions, were determine and compared quantitatively. The results show that both methods of capturing shoeprints perform equally well. There were differences however between the way the prints were made, with the dynamic and rolled conditions showing significantly more RACs, basically as a result of the larger shoe sole area that was printed compared to the static print. The authors therefore conclude that the way the shoeprints are made has a larger effect on the amount of RACs than the way the prints are captured.

### 3.2. Acquisition

Typically the acquisition of shoemarks is done using 2D photography [11] and several techniques exist that use oblique light from different angles to make the image look more 'plastic'. However, 2D photographs do not provide depth information and are dependent on lighting conditions and it has been shown that identification can be obtained in a higher percentage of cases, when additional casts of impressions are made [11,12]. As the properties of substrate materials vary greatly however (sand, snow, mud etc.) and evidence might be destroyed, casting can be complicated. Casts also have the disadvantage that they have to dry sometimes for longer periods of time. Alternatively, 3D data can be determined instead of casting, or the cast itself can be acquired by 3D technology [9]. Using 3D models enables easy storing and quantitatively analyzing impressions as well as sharing with others.

In recent years, novel developments were mainly in the field of 3D image acquisition and we therefore will focus on new technologies in this field in the following sections. Possible techniques for 3D acquisition are laser scanning devices, both for long distance, low resolution, and short distance, high resolution acquisition of impressions, which have been around for quite some time already [9]. Authors used for example the NextEngine laser scanner [50,51]. In addition, more techniques were tested in recent years. Thompson et al. [52] presented using a 3D structured light scanning device (4D Dynamics PicoScan [52,53]) for shoemark impression acquisition. The authors state that the technique provides morphological information, next to color info and provides higher resolution at lower equipment costs compared to laser scanning devices. To

demonstrate the technique, shoe impressions are acquired using seven shoes, new and used, in sand and soil and subsequently compared manually using 3D software. Based on visual assessment of the results, the authors state that in principle the technique is suitable for the examined substrates, but that a more in-depth quantitative analysis of the acquisition results is required. Other systems for structured light scanning include the GOM Atos [54], the Fraunhofer Kolibri [10,55], the Artec Eva [56] and the David SLS-2 [57]. Photogrammetry was employed by Faulkner [58], who studied the usability of applying the commercially available software package PhotoScan [59] to reconstruct 3D shoe impressions. Photogrammetry is used to construct a 3D model based on typically two or more 2D images, taken from different angles (typically the more images are used, the better the resulting 3D model is [60]). To this end, 3D data was reconstructed from impressions in sand from a used running shoe. Qualitative analysis results indicate that the software can produce useful results in practice, given that the illumination conditions in the 2D photographs are similar. Yet another acquisition technique is photometric stereo (e.g. Evident EverLSS 360 [61]).

The main criteria for the applicability of devices in practice are, among others, price, ease of use at a crime scene, acquisition speed, resolution and performance under realistic crime scene conditions (e.g. varying substrate materials, temperatures and lighting conditions), with the last two items being of high importance. Of all the mentioned techniques so far, 3D structured light based systems are the most promising for acquiring not only class characteristics but also RACs, as these systems typically yield the highest resolution and thus provide most details. In general systems are easy to use at a crime scene and some systems were specifically designed for and tested under a variety of crime scene conditions [61,62]. Summarizing the recent advances, 3D shoemark acquisition technology is expected to be introduced into casework in the near future [62].

### 3.3. Enhancement

Photographs of crime scene shoemarks are often noisy and lack contrast. Therefore Reddy [63] proposed an automated shoemark enhancement algorithm that produces well illuminated images with balanced contrast and little noise. The algorithms is tested qualitatively and quantitatively using ten images of three databases and shows to be superior to previous approaches. In future work, the impact on the application of the technique in the forensic context will have to be demonstrated.

### 3.4. Casting/preservation

Shoemarks encountered at a crime scene are found in many different substrates, e.g. in soil, sand, gravel or snow at an outdoor scene or on wooden floors, carpets, tiles etc. at an indoor scene. In addition, marks can be made in dust or can be made with wet shoes (e.g. blood or mud). As a result, specific casting/lifting techniques are required to optimally recover and preserve shoemarks from substrates with greatly varying properties. Authors presented techniques for casting in snow [64], soluble food products [65] and sand [12] as well as for lifting wet marks [66].

Petraco et al. [64] describe a method to cast footwear impressions in snow, using commercially available bio-foam blocks. For casting, the authors suggest modifying the blocks by attaching a cardboard to one side of the block, putting the other side on the impression and apply pressure manually. For testing they cast several impressions and qualitatively assess the casts. Overview images of the results are provided in the article. They conclude that class and random acquired characteristics can be discerned in the casts and that the method could also be used for casting

impressions in sand, dried soil or mud.

Sabolich [65] studied ways to best preserve footwear impressions in a water soluble food product (Swiss Miss Chocolate Mix). To this end a used hiking shoe, in which additional small cuts were made, was used to create impressions, two for each condition. Subsequently eight different products like waterproofing spray, hairspray and antiperspirant were used to cover the impressions. Finally the marks were cast using dental stone. The class and individual characteristics of the casts were compared qualitatively with a control cast and the results show that only class characteristics can be discerned. The best results are obtained using a sequential treatment of several products.

A comparison between photography and casting of footwear impressions in sandy soils frequently encountered in the United States was presented by Snyder [12]. Two worn athletic shoes with author-applied additional RACs in the outsoles were used to create impressions in a variety of sands, fill dirt, crushed coquina and top soil. Afterwards photographs were taken and casts made with dental stone. Subsequently the fractions of in total nineteen RACs that could still be visually discerned on the photographs and the casts were determined and compared. The results show that in all cases, RACs were significantly better retained in casts. The soil type did have a large impact on the amount of retained RACs, ranging from below 30% (three types) to more than 60% (three types).

Hong [66] studied qualitatively, if footwear marks made by wet soles can successfully be lifted after drying using an electrostatic dust print lifter device (EDPL). The authors consider several types of underground, namely overhead projector film, a painted road, a wooden floor, a stone floor and asphalt. Lifting was done after waiting different periods of time after deposition, up to 28 h, to allow dust to settle on the marks. Furthermore, the effect of the shoe sole drying up by walking was considered. The authors conclude that dried up marks can successfully be lifted from all surfaces, even after waiting up to 28 h, and that wet shoe soles produce useable marks for several steps, but that there is more research needed to assess the influence of the surface properties on the marks.

### 3.5. Variability/individuality of shoemark characteristics

To be able to assess the similarity and dissimilarity between shoemark characteristics made under uncontrolled circumstances, e.g. at a crime scene, it is necessary to know the (dis-) similarity between shoeprints made under controlled circumstances in the laboratory. Several articles in the last years focused on studying a variety of factors that may influence the appearance of shoemarks and the occurrence and absence of shoe characteristics. Typically, authors studied the variability of randomly acquired characteristics (RACs) (or accidental marks) caused by abrasion and damages, rather than class characteristics, depending on substrate type [67], shoemark medium (e.g. blood or dust) [67,68], shoe abrasion [69] or repeated shoeprint creation under the same condition (repeatability of RACs) [25].

McElhone et al. [67] simulated several conditions typically encountered at crime scenes to study the quality of shoemarks dependent on these conditions. Specifically they considered two types of blood (human and equine blood), two types of flooring surface (wood and tiles), three categories of footwear tread depth (non-existent, shallow and deep) and varying periods of time that the blood was allowed to dry (0–24 min). Shoeprints were made with an apparatus specifically designed for this purpose and the resulting blood patterns acquired with a digital camera. Three repetitions were made for each condition. Subsequently, the quality of the image, measured as the amount of retained detail in the print, was assigned to one of five categories. Finally, the category

ranking was compared among the conditions using statistical significance testing. In general, human blood marks were of better quality compared to equine blood marks. The substrate did not seem to have an effect on the quality, except when blood was allowed to dry for a couple of minutes. The tread depth did have an effect, with the shallow depth yielding better results than non-existent and deep treads. The dryness of the blood had an impact as well, with the mark quality decreasing but only under some conditions. The authors conclude that the results highlight the importance of taking external factors into account during the interpretation step of an examination.

Two articles from the last years focus on the variability of RACs. One with respect to repeated shoeprint creation in the lab [25] and the other with respect to different circumstances in which the marks were created (lab vs. crime scene) [68]. In Shor et al. [25] the authors state that often only one shoeprint was created in the lab for an examination and considered as a genuine representation of the shoe outsole. This approach however was based on the assumption that RACs are present consistently in repeated prints. To test this assumption, the authors studied the repeatability of RACs. They created test prints with seven worn shoes that contained one to eighteen RACs in the outsoles. Orange fingerprint powder was applied to the soles, prints were created on clear adhesive lifters and subsequently photographed. To locate RACs a semi-automated method was used, in which a qualified examiner first roughly defines the surrounding area of a RAC, after which an in-house developed software determines its contour, which can again be adjusted by the examiner if desired. This was done for twenty-five repetitions of each outsole. The repeatability was then assessed qualitatively, by providing color-coded contour overlay images, and quantitatively, by defining a contour dissimilarity measure. The results show that for some RACs the contours were determined very consistently among the prints. For others however, particularly resulting from shallow scratches, partially torn material as well as height variations of sole elements, the contours varied substantially. The authors suggest creating several experimental shoeprints under varying conditions in the lab, to assess the repeatability of the prints for a particular shoe.

Yee et al. [69] used the publicly available software package CloudCompare [13] to measure the abrasion of shoe soles. To this end, new running shoes were worn by fifteen participants over a ten month period (350 km approximately) and subsequently four left shoes with identical shoe sizes were selected. Data acquisition before and after wear was done with a David structured light scanner (SLS-2) to determine 3D surfaces of the soles, which subsequently were loaded into the software and aligned semi-automatically. Finally, wear was determined by calculating the distance between the surface before and after usage. For analysis, the distance was shown locally color-coded on the surface of the new shoe such that different colors indicate different distances and hence local differences in abrasion.

The conditions in the lab can be controlled to reduce the variability of experimental shoeprints. This is not the case at a crime scene however, where many factors like the substrate or the deposition process play a role in shoemark creation and influence mark characteristics. Therefore, Richetelli et al. [68] studied the difference between RACs on high quality shoeprints created in the lab and simulated crime scene marks made with the same shoes. More specifically, the loss and similarity of RACs was analyzed quantitatively based on shape, perimeter, area and common source of RACs. To create shoeprints, fifty shoes were scanned and used to generate Handiprints. Crime scene-like prints were created by covering the shoe outsoles with shoe polish and then walking normally over acetate sheets. Two such created prints were subsequently lifted. Based on the global quantitative analysis of RAC

features the authors came to three conclusions. First that assessing RACs globally might not be specific enough, second that the absence of RACs in a crime scene mark should not be the reason for exclusion and third that there is no basis for the demand for a minimum number of corresponding features. In fact, the results show that even when simulated crime scene marks are used, which are supposed to be of better quality than real crime scene marks, the amount of lost RACs varied between 33% and 100%, with an average of 85%.

### 3.6. Automated mark retrieval/comparison

A large body of literature the last three years was dedicated to automated comparison of shoemarks, particularly for retrieval of shoeprints from reference databases that show high similarity with query shoemarks from crime scenes. In the following, a short overview of shoemark retrieval systems is given. More thorough overviews can be found elsewhere [9,15–17,70].

Mark retrieval typically is a two-step approach consisting of 1.) a feature-description step and 2.) a similarity measurement step. Features are shoemark properties that have the potential to distinguish different types of shoes from one another. In early systems, features are user annotated (so-called tags) geometrical shapes like lines or circles [71]. In fully automated systems, algorithms determine the useful features, which can e.g. be derived from periodic patterns of the sole or specific interest points like shoe profile corners. More advanced algorithms take the geometrical relation between features or regions of specific pattern of similar features into account. In addition, some approaches consider several scale levels to include more local and more global information within the same framework.

The features and the relationship between features finally have to be encoded into a representation, with which different data can be compared in the similarity measurement step. If for all shoeprints in a database the same abstract representation exists, an automated database query then calculates a measure of similarity between the abstract representations of the query data and the database entries and subsequently ranks the similarity values. Depending on how the similarity measure is implemented, e.g. the highest similarity would then be on rank 1, the second highest on rank 2 etc. If the algorithm performs well and the shoemark in question was made with a shoe in the database, the ranked list contains the print of that shoe at the beginning of the list.

The determination of suitable features for automated shoemark comparison is a challenging task, as shoemarks can be rotated, translated, scaled, deformed w.r.t. the shoeprint of the same shoe and can be incomplete. In addition, the sole properties could vary due to wear. Furthermore, shoemark images from crime scenes might include background patterns, multiple marks or are influenced by the way the marks were made (e.g. blood, dust, dirt etc.). In the following, the contributions of articles of the last three year are presented.

Gwo et al. [72] present a region based method for shoeprint from database retrieval, based on earlier work (Wei et al. [73]). The method uses binary (black/white) images to first determine the outer contours and based on that a core point using the entire shoeprint. Subsequently the print is subdivided in circular regions and Zernike moments based features are calculated as search criterion for automated retrieval. The approach is tested using 5 laboratory quality prints made with each of 246 shoes, hence in total 1230 shoeprints. Database retrieval performance is not reported.

Richetelli et al. [70] highlight the importance of testing algorithms under realistic circumstances and therefore compared an algorithm developed by the authors [74] with those that showed promising performance in previous studies by Luostarinen et al.

[75] and Almaadeed et al. [76] (both discussed in detail in our last review [1]). The tested algorithms were Fourier transform based methods employing either the power spectrum (Fourier-Mellin transform, FMT) or the phase spectrum (phase only correlation, POC) and methods using local interest points and applying a scale invariant feature transform (SIFT), combined with random sample consensus (RANSAC) comparison. The test database consisted of full (100 shoes) and partial (full prints divided into six areas) high quality (HQ) shoeprints as well as crime scene-like full shoeprints (36 shoes) with variations in media type (blood, dust), substrate (ceramic tiles, vinyl tiles, acetate sheets, paper) and chemical/optical print enhancement procedures. Subsequently HQ full and partial as well as crime scene-like prints were compared to the HQ full database and the methods compared quantitatively. The results show, that for the HQ full vs. HQ full comparison, all methods performed very well. In all other circumstances, the POC method performed best with a probability of approx. 58% that the correct match is within 1% of the database size (1 image) and a probability of approximately 85% that the correct match is within 10% of the database size (10 images). In addition, the local feature based approach (SIFT + RANSAC) performed better than FMT for partial HQ marks but the results were more or less the same for crime scene-like marks. In the conclusions, the authors stress that not one algorithm necessarily performs best in all circumstances though. For example, as Fourier transform based methods (FMT and POC) rely on repetitive structures in the outsoles (which typically occur in more than 60% of the cases according to Ref. [77]), for structures with few repetitive patterns local feature based methods might work better. Further research is required to assess this further.

Kortylewski et al. [15,78] propose a method based on the Active Basis Model (ABM). Their approach captures local information about geometry and appearance (texture) of patterns at multiple scale levels and includes a deformation model to deal with possibly deformed shoemarks. In addition, the ABM is extended with an occlusion model to separate the relevant parts of the mark from possible background information. To test the approach on a wide range of realistic data, the authors, together with several state criminal police offices, created the publicly available FID-300 dataset [79], which consists of real crime scene shoemarks, including partial and deformed marks. In addition, the database contains 1175 HQ gallery images. The authors compared their previous approach based on local Fourier spectra comparison [77] with two more Fourier transform based methods, a local appearance based method and an approach based on matching local geometric primitives. The results show that for the FID-300 dataset, their proposed method significantly outperforms all other methods with a probability of 58% that the correct match is within 1% of the database size (12 images) and a probability of 80% that the correct match is within 10% of the database size (120 images).

The same FID-300 database was also used by Kong et al. [18,80] to test a novel algorithm using multi-channel deep feature matching based on multi-channel normalized cross correlation, embedded in a convolutional neural network (Siamese network). While most methods seek to be invariant to geometrical distortions, their methods uses a dense template search over translations and rotations as initialization. Using the FID-300 dataset, the authors compared their algorithm to the one presented by Kortylewski et al. in Ref. [15] and show that they obtain a better performance, namely 79% that the correct match is within 1% of the database size (12 images). The authors also added experiments for testing the robustness of their algorithm with respect to different relative sizes of the partial marks. The results are 84%, 86%, 79% and 64% for full, ¾, half and ¼ the size of a full mark, that the correct match is within 1% of the database size (12 images). The higher performance for ¾ prints is explained by the fact that often the full

marks had more background noise and/or had overlapping marks. Finally, the effect of background noise was studied by specifically selecting marks with a lot and little background noise. The difference is very severe, with a drop in probability from 72% to 15% that the correct match is within 1% of the database size (12 images). Based on this last result the authors state that in a real examination, the examiner could coarsely draw the boundaries of the mark manually and therefore reduce the influence of background noise.

Yet another study made use of the FID-300 database for testing. Alizadeh et al. [81] present a shoemark retrieval method based on blocked sparse representations of the images. In order for the method to work, several pre-processing steps to remove rotation, scaling and noise are required. Testing was done by manually selecting 83 less noisy images from the 300 and processing these manually to remove noise. They report a probability of 35% that the correct match is within 1% of the database size (1 image) and a probability of 60% that the correct match is within 10% of the database size (8 images). In addition, they created an own database of crime scene-like prints of 190 times 5 prints, thus 950 prints, which is meant to be publicly available. At the time of writing however, the link given in the article did not work.

In Wang et al. [16] the authors present an improvement to a previously published method based on manifold ranking and using hybrid features of region and appearance [82]. As often multiple shoemarks of the same shoe are present at a crime scene, they adjusted their method such that multiple images of the same mark, which might contain complementary information, can be used to improve retrieval performance. In addition, examiner provided scores of the relevance of those multiple mark to the query are included. For testing, 72 query images from real crime scenes were compared to a subset of a database including 10,096 shoemark images (the original database size is much larger [83]) from real crime scenes, which consists of 9592 original images, the 72 query images and 432 synthetically manipulated versions (rotation, translation, scaling) of the query images. In total eight state of the art methods (including [74,77]) including their previously presented method [82] were tested with this dataset and with a probability of 90% that the correct match is within 1% of the database size (101 images), the method performs significantly better than all other methods, including their own previous method (probability of 85% that the correct match is within 1% of the database size, thus 101 images [82]). This is the first method so far that demonstrates how multiple marks on a scene can be exploited to improve the retrieval performance. Unfortunately, the database is not publicly available.

Using the same database Cui et al. [17] test their robust shoeprint retrieval method based on local-to-global features. They use a local feature point extraction step in combination with a deep belief network (DBN) to render the step robust with respect to image noise and then employ spatial pyramid matching (SPM) to incorporate local and global information to be able to handle partial images. The method was tested with 536 query images against a database of 34,768 crime scene images [83] and compared to the performance of four other methods, including [82], the method proposed by Wang et al. in 2014. The authors report slightly better results. Their probability is 81% that the correct match is within 0.3% of the database size (104 images) against 78% [82], but the results are very similar to Wang et al., which are 82% [16]. This method does not take multiple marks into account though and that might further improve the performance. In addition, their method is several times faster.

As the performance of local interest point based methods is strongly dependent on how accurately local interest points can be detected, Vagac et al. [84] presented a strategy for robust detection of shoe sole features using a deep neural network (DNN). For testing, 13 publicly available image found online were used and feature detection performance compared to several other feature detectors (Sobel, Canny, Haralick, Marr-Hildreth-log edge detectors and Line Segment Detector). The authors conclude that their approach qualitatively outperforms the other approaches.

### 3.6.1. Summary

Early systems for automated shoemark to shoeprint comparison were relatively simple and fast and reported retrieval performance was high, but they were typically tested with high quality marks and prints made in the laboratory. As several recent studies show, the performance of these systems decreases drastically in experiments including real crime scene quality data [9,15–18,70,81]. More recent methods therefore are more advanced and the most successful methods take local as well as global image information into account [15–18,81]. As retrieval performance is degrading with image noise [18], some authors explicitly include a noise model in their approach [15] or propose to coarsely draw mark boundaries manually, prior to automated analysis [16]. As often multiple marks are present at a crime scene, one method allows to use several marks of the same shoe, determined by the forensic examiner, to further improve the retrieval performance [16].

Most methods with the highest retrieval performance were also tested with databases including large numbers of real crime scene marks and some authors even used the same databases. Unfortunately some of these databases are not publicly available [16,18] and therefore it is difficult to judge what kind of variation is present in the crime scene marks. In addition, only one author [18] focused specifically on the performance of the algorithm dependent on partiality and noise levels using *real crime scene data* (this distinction was presented frequently by others, but with high quality prints). The others did not make this distinction. This might help though to further improve algorithms by focusing on the aspects that influence the retrieval performance the most. For example by allowing a minimum user interaction as mentioned above, if that improves the performance significantly. Besides retrieval performance, also retrieval speed is an important aspect and there are great differences between algorithms, ranging from a few milliseconds [17] to 18 ms [16], 54 ms [74] and 255 ms [85]. Again, allowing some user interaction, might allow the algorithm to be more time efficient. The time investment at the beginning might be easily compensated during retrieval, especially for very large databases.

On a whole it is not possible at this point to objectively compare the best performing methods based on the literature and this should be considered for future research. Methods should be tested using *publicly available* databases, with special focus on the performance, given varying noise levels, partiality and deformations in *real crime scene data*. Nevertheless, the results presented in the literature (best method yields a probability of 90% that the correct match is within 1% of the database size and other methods are close) are very promising and it is expected that robust automated methods with high retrieval performance are close to implementation in practice.

Finally we want to point out that all methods so far are based on 2D images of shoemarks. As 3D data typically is more accurate and contains more details compared to 2D images, it would be interesting to see if algorithms can be improved using 3D data in the future.

### 3.7. Digital reference databases

For automated shoeprint retrieval and weighing the evidence after a comparison alike, large and representative shoemark databases are required. In the following an overview of *publicly available*

databases including a description of their composition is given.

To the best of our knowledge, three databases became publicly available in recent years. The FID-300 datasets includes three hundred crime scene marks [79], that were acquired by scanning gelatin lifters or by photography. In addition, 1175 reference prints are provided. For each crime scene mark a corresponding reference print is available. Two databases were set up with crime scene-*like* marks [81,86]. The first contains 190 times five repeated prints, thus 950 prints in total. Data acquisition was done with a scanner. To simulate crime scene conditions, participants were wearing used shoes and walked freely. Although the database is presented as publicly available, the link to the database did unfortunately not work at the time of writing. Finally, a crime scene-*like* shoeprint database that includes different substrate materials (ceramic, vinyl, acetate and paper) as well as different print media (blood and dust) was presented in Ref. [86]. Eighteen pairs of shoes were used to create these marks, with participants walking freely. Marks were acquired by lifting and scanning (dust) or by scanning after drying (blood). High quality reference prints of one hundred shoes, including the ones used for the crime scene-like prints, are also available.

## 3.8. Weighing the evidence/interpretation

In the recent literature there are basically two approaches aiming at rendering shoemark evidence comparison more objective, either by employing a likelihood ratio (LR) based system to determine the evidential strength [21] or by studying properties of RACs like frequency of occurrence as well as shape features [22,23] that can subsequently be used to derive statistical measures for interpretation of the evidence.

The articles presented in section 'Automated mark retrieval' all define some measure of similarity between shoemarks. However, most of the contributions use the measure of similarity to produce a ranked list, as their goal is image retrieval from a database and not determination of the weight of the evidence. An exception is the PhD thesis of Park [21], in which a semi-automated method to calculate a score (signature) between shoeprints is presented, which subsequently can be used to determine the strength of the evidence, the likelihood ratio (LR). The score takes class characteristics as well as unique wear and tear patterns into account and is calculated by combining six features using 1.) a random forest or 2.) a Bayesian Additive Regression Trees (BART). In addition, they propose a score based likelihood ratio determination system to calculate an LR. The algorithm is tested using full prints of 150 pairs of shoes, of which five prints each were acquired in their lab. The results show that the known match (KM) and known non-match (KNM) score distributions are well separated.

In the report of the National Academy of Sciences (NAS report) 2009 [3], the authors state that research on the "Random shape and/or placement of RACs" and the "Mathematical probabilities of RACs" is required. Work towards fulfilling this goal was presented by Speir et al. [22] and Kaplan Damary et al. [23]. The first article describes studying the discrepancies between examiners and automated labelling of RACs, the consistency with which human examiners determine and label RACs and location dependent co-occurrence of RACs on used shoe soles. The used data were scans and Handiprints, made from 1000 worn soles of mostly athletic shoes, which were pre-processed to align and enhance the images. In total, 57,426 RACs were located on the images and the areas drawn manually by in total seven in-house trained analysts. Subsequently, the location as well as a shape descriptor was determined for each RAC and categorized into four categories (lines/curves, circles, triangles and irregular) by an automated algorithm and the analysts. A comparison of the results shows that the chosen categories were not always consistent among examiners and compared to the automated method. The authors therefore state that it might be more robust to use only three categories (irregular, elongated and approximate isometry for circles and triangles). The results further show that RACs were labelled consistently by the analysts as long as they were detected, but that the detection step was less consistent. Finally, a grid of 5 times 5 mm sized boxes was employed to determine frequency and location specific information for RACs, as well as the chance of co-occurrence of RACs in a grid cell. The results are provided by means of a heatmap, a color-coded representation of the frequencies, and are available online [87]. The authors stress that location and shape does not account for clarity, quality and complexity of geometric features, while these aspects might also be very important to forensic shoemark examiners in practice.

In Kaplan Damary et al. [23] the authors investigate the relationship among several RAC features, namely location, shape type and orientation. The goal of their work was to test whether assuming independence between individual RACs is valid, as previous studies did assume independence and thus multiplied individual probabilities of RAC occurrence (e.g. Stone [24]). In contrast to Speir et al. [22], the authors proposed using the seven categories for shape used by the Israeli Police Division of Identification and Forensic Science: Scratch, Hole, Cut-off corner, Rift, Foreign object, Schallamach and Missing part. During analysis however "Foreign object" and "Missing part" were omitted as a result of low occurrence. The hypotheses were tested with 13,500 examiner annotated RACs found on 380 shoeprints. The results show that all individual features were dependent on one another, with dependencies between shape type and location as well as orientation and shape type being strongest. In additional experiments, the authors reduce the set of shapes to "Scratches" and "Holes" and show that these are independent on location, but not orientation. In addition, they show that RAC size is dependent on location, shape and orientation. The authors state that the used shape types have to be redefined and that RAC annotation by human examiners might be prone to error.

There is a need for standardization of RAC shapes, in order to compare studies to determine frequencies, co-occurrence and distribution, as well as the way a shoe sole is to be sub-divided for local RAC frequency assessment. In fact, different authors used different categories. Speir et al. [22] used four categories, but suggest three, Kaplan Damary et al. [23] used five categories (two omitted) and Stone [24] also used five (but different) categories.

It seems like the assumption of independence is not valid, at least given the categories that were used in Kaplan Damary et al. [23]. It might be possible to rely on a small set of shapes that occur very often (like holes and scratches) only instead of defining a large variety of shapes that in turn might be difficult or ambiguous to annotate. If independence RAC features can be demonstrated for such a subset, the individual probabilities could be multiplied [24]. Otherwise, dependencies between RAC features have to be determined and taken in to account when calculating the evidential strength.

Class characteristics of shoe outsoles are relatively constant over time. The RACs may change over time, but that is typically with respect to the distribution and the total amount, at least for a limited time period, as damages to the outsoles are permanent. Bily et al. [88] presented a study that demonstrates, that certain shoe outsole materials can temporarily contain imprint patterns of the substrate. These temporary patterns can then be seen in shoeprints that were made right after walking on a substrate with a distinctive pattern. Specifically, ethylene vinyl acetate (EVA) outsoles, which are used in TOMS Men's and Women's Classics, seem to have this property, in case the tread is worn away as a result of heavy usage.

The authors therefore took TOMS shoes and stepped on twenty different substrate materials, including tile, indoor and outdoor carpeting, welcome mats and bath mats. Right afterwards, shoeprints were made and studied for patterns. For eleven of the twenty substrates, the authors could find the substrate pattern in the shoeprint. The most likely substrates to cause these patterns were the non-yielding ones. The imprinted pattern disappeared after a short period of time, but the authors stress that in shoes containing EVA, these temporary patterns can occur and be misleading during an examination.

Typically the distinctive value of class characteristics is limited, as they are present in all shoes of the same brand, make and size. Sanuk Vagabond and TOMS classic shoes however consist of a mixed-rubber outsole and an additional textile layer, which both contain class characteristics. In Gokool et al. [89] the authors therefore studied whether the two overlapping patterns of class characteristics yield a distinctive pattern. For both brands, four pairs of new shoes were taken and the outsoles scanned. Several features were defined and annotated manually on the scans using Adobe Photoshop. Repeated annotation was simulated by randomly displacing the features. Subsequently, the similarities between configurations of features of known matching and known non-matching patterns were calculated using an in-house developed software. The results showed that the KM and KNM similarity scores were clearly separated, indicating that the combination of patterns with class characteristics can indeed yield a highly distinctive pattern.

### 3.9. Software for shoemark comparison and retrieval

Several open source software solutions exist that allow the user to visualize and compare 3D shoemark datasets as well as perform 3D measurements. One author used for example the CloudCompare software [13] and in Bennett et al. [9], an in-house software package DigTrace is presented [14]. The latter was specifically designed for footwear analysis and ichnology and contains many options specific to shoemark examination. Yet another option is Meshlab [14], although the main focus of this package is the processing and editing of 3D datasets, rather than comparing them.

There are some commercial systems available that can be used for image retrieval from a database. All of these are working with 2D images. Two examples of which the authors are aware of are a system called PRIDE Shoeprint Matcher by Hobbit Imaging Solutions [19] and EverASM (Automated Shoeprint Matcher) by Everspry [20]. The first includes a Fast Fourier Transform based algorithm for retrieval and the second is based on a deep neural network approach. The first system is at the time of writing evaluated by law enforcement agencies in several countries and it is expected that it will be implemented in casework in the near future [90].

## 4. Comparison of physical properties

In the relevant period, only one article was found including a comparison of physical properties of evidence. Nienaber et al. [91] assessed the potential value of analyzing the chemical and physical properties of plastic cable ties. Twenty packets of black plastic cable ties (nominally 200 mm times 4.8 mm) were purchased in packet sizes ranging from twenty-five to one hundred and representative samples were subsequently compared within and between packets, based on visual inspection, chemical composition, measured physical dimensions such as width, thickness and tooth-count of the grip section and stable isotopic composition ($\delta^2H$, $\delta^{13}C$ and $\delta^{15}N$). The results show that cable ties of the same packet were indistinguishable with respect to all characteristics. Cable ties from ten of the twenty packets could be distinguished by visual inspection, in some cases also for ties from the same manufacturer. Measuring the physical properties did not provide additional discrimination. Nineteen of the twenty packets were uniquely characterized by their isotopic composition, based on $\delta2 H$ and $\delta15 N$ measurements. The authors conclude that isotopic composition comparison is the most effective approach but that visual examination can provide a rapid and inexpensive first step of an examination.

## 5. Striated and impression toolmarks

Forensic toolmark comparison requires many steps, from documenting the marks at the crime scene to creating and acquiring experimental toolmarks in the lab as well as comparing toolmarks and evaluating the results. An overview of standard operating procedures for shoe and toolmarks can be found in a book by Petraco et al. [48]. It includes the whole pipeline from the crime scene to the examination.

The following chapters provide an overview of recent publications regarding one or several steps of forensic toolmark examination. Besides advances to the traditional approach of manual comparison and evaluation of toolmark evidence, special attention is paid to new technological developments that aim at rendering a mark comparison more objective by using 3D toolmark data instead of 2D images or by employing automated toolmark similarity determination and subsequent calculation of statistically meaningful measures of the evidential strength.

### 5.1. Detection/creation

Vehicles and firearms are typically labelled with unique serial numbers. These are frequently removed by criminals to make it difficult to determine the rightful owner of the stolen goods or to claim ownership by creating new numbers. Therefore techniques are required to restore obliterated numbers and in the last years, several methods were proposed to restore numbers in iron, steel and copper alloys [92–94]. As using titanium, aluminum and possibly polycarbonate are getting more frequently used for car and firearm parts in the future, number restoration possibilities were studied with these materials as well (titanium [95], aluminum [94], polycarbonate [96]). Finally, an article studying the optimum temperature for creating experimental toolmarks in wax was presented [97].

The usual method of alteration in Israel is polishing with an abrasive disk, and to restore serial numbers, Fry's solution (90 g copper II chloride, 120 ml HCl and 100 ml distilled water) can be used [93]. However a new technique to remove serial numbers seems to be heating with a localized melting system or a flame torch. In Tsach et al. [93], the authors studied whether heated serial numbers could be recovered with Fry's solution. Eight samples of chassis, of which it was suspected that the original number was altered by local heating and re-stamping, were taken and treated with Fry's solution. The experiments revealed that the original number could not be recovered, but that the solution did cause a circular area around some of the digits while this did not happen for others. The circular areas turned out to be present around the altered digits and thus the authors conclude that although number retrieval is not possible, using Fry's solution on digits that were altered by localized heating can at least be used to detect that digits were altered.

Fortini et al. [92] studied the usability of five etching reagents, including Fry's solution, to restore numbers on steel, that were obliterated by various depths of erasure up to 60 μm. To this end, fifty stamped steel disks were provided by Beretta and represent

the material that is typically used for manufacturing firearms. Half of the disks were normalized and tempered and the other half austempered. The characters, seven on each disk, were removed to varying depths by honing. Of each group, five plates were taken for each of the five reagents. In total, three hundred fifty images (fifty disks with seven characters) were taken before and after restoration and compared visually to each other by in total thirty observers, of which each was assigned to study twenty five random images. The results show that Fry's solution had the highest sensitivity, with the most characters restored up to 60 μm. A solution of nitric acid (25% concentrated $HNO_3$ and 75% water) resulted in the major number of characters being restored. The authors also studied whether sex and age of the observers did influence the results, but statistical significant effects were not encountered.

Sharma et al. [94] provide several examples from cases in which obliterated vehicle serial numbers in aluminum, iron and copper alloys were restored. The authors also present a flow chart indicating the preferred reagent for the different metal types. They suggest Fry's solution or nitric acid for iron, a mix of glycerol, hydrofluoric acid and nitric acid (30 ml, 20 ml and 10 ml respectively) for aluminum and a mix of iron(II) chloride, hydrochloric acid and water (20 g, 10 ml and 250 ml respectively) for copper alloys. The number restoration performance was assessed qualitatively.

Increased usage of titanium in modern firearms raises the question whether traditional restoration methods can still be applied. To test this, Schultheis [95] took eight room temperature and heated reagents (including Fry's solution) and applied these to titanium samples. As only concentrated hydrochloric acid seemed to cause a reaction, this reagent was studied more thoroughly on eleven heated titanium samples with four different methods of marking application. Visual inspection of the samples showed that in ten out of the eleven, the serial numbers could be fully or partially restored. The reagent application time was dependent on the way the number was applied originally. Metal deformation techniques like stamping required relatively more time compared to metal removal techniques like laser engraving.

Polymers have attractive properties and may be used for replacing parts in automobiles and firearms that traditionally are made from metal [96]. As traditional etching techniques cannot be applied to polycarbonate, Parisien et al. [96] propose an approach based on Raman spectroscopy for this purpose. With this technique, residual mechanical strain and local structural changes can be detected. In addition, the method is non-destructive. In a pilot study, the authors successfully recovered stamped letters (120 μm deep) that were obliterated by milling and state that the estimated maximum depth of recovery is approximately 750–800 μm.

Typically, experimental toolmarks are created in lead, as very fine details can be observed in this material and it is soft enough to not alter the state of a tool. Wax could be a cheap and non-toxic alternative to lead and it has been shown, that toolmarks in wax, created at room temperature, are of similar quality than those in lead for the most relevant range of details [98]. Some types of wax however, might be too soft to create toolmarks reliably at room temperature. Therefore Finkelstein et al. [97] studied the influence of the substrate temperature on the details in toolmarks. To this end, a flat screwdrivers was used to create marks with 45° angle of attack in four different types of wax (LectroStik Stikkiwax, Elgad Multiwax, Sonneborn Multiwax and Chemtrec beeswax), at five different wax temperatures, ranging from −30 to 25° Celsius. Marks for each condition, i.e. brand and temperature, were repeated ten times. In addition, marks at −18° were made, as this is the typical temperature of domestic freezers. Based on visual comparison of the marks, the authors state that the marks made in wax at −18° and −30° were significantly better than those made at higher

temperatures for all tested wax types. After toolmark creation, the marks can be stored at room temperature without compromising the marks.

### 5.2. Acquisition

Standard reference scales are frequently used by forensic examiners, to document physical dimensions of objects. As many labs use the same scale for many years, Ferruci et al. [99] studied the accuracy of ABFO (American Board of Forensic Odontology) No. 2 standard reference plastic scales from four different vendors. Five scales from each vendor were purchased and tested with respect to length and circle diameter measurements, the error in placement of the circle centers as well as leg perpendicularity. These criteria were assessed twice, right after purchase and after four years of usage. The results show that after purchase, all length scales satisfied the ABFO No. 2 specifications, while the internal and external circle diameters and center-to-center distances lacked adherence to specifications. The within variation for the vendors was low. Regarding the leg angle did more than half of scales not satisfy ABFO No. 2 specifications and within variability was high. The measurements after four years showed minimal changes, except for the leg angle, that had changed significantly, up to several degrees for some vendors. The authors therefore suggest conducting scale quality checks frequently.

### 5.3. Classification

Typically a toolmark examination starts with a taxonomical study of toolmark and tool characteristics. Klees [100] therefore proposes a classification system including toolmark terms and descriptions as well as tool actions to augment the common classification systems found in literature and to reach a more standardized format. The article proposes the following tool action type classification: abrading, chopping, compression, crimping, engraving, firing, gripping, leveraging, pinching, piercing, sawing, shearing, slicing and torquing. For each of the classes, a short description together with one or several example images is provided.

### 5.4. Variability/individuality of toolmark characteristics

The variability of marks of tools like screwdrivers or chisel may be high, as it is dependent on many parameters like the angle of attack, the substrate, the axial rotation angle as well as the depth of the mark. In the past, publications mainly focused on the influence of the angle of attack, thus the angle between a tool and a plane orthogonal to the substrate, on the similarity between toolmarks and subsequently the effect on the separation between KM and KNM similarity score distributions [1]. Recently, two publications focused on the effect of the *axial* rotation of a tool during tool mark creation, thus a rotation with respect to the longitudinal axis of a tool [101,102]. The rationale is that in real crime scene marks it might not always be clear under which axial rotation angle they were made and that there might be several marks, created at different axial rotation angles, originating from the same tool present at a crime scene.

In Macziewski et al. [101] the authors studied the influence of varying the angle of attack and the axial rotation angle on a toolmark similarity score (T1), which was described in earlier work [103]. They used ten sequentially made flat head screwdriver tips and created five toolmarks each in lead with angles of attack 40, 55 and 70° and axial rotation angles 0, 10, 20 and 30°. The marks were acquired using an Alicona G4 Infinite Focus Microscope [104] and automatically compared using the Mantis software [105]. The

results show that when known matching toolmarks taken at identical angles are compared, the score is significantly higher than scores of comparing known non-matching marks, however as the axial rotation angle increases, the scores are slightly decreasing and the deviation is increasing. With the axial angle fixed, varying the angle of attack causes the score to drop significantly and at an angle difference of larger than 10°, the scores are in the range of known non-matching scores. The same holds for varying the axial rotation with the angle of attack fixed. Varying both angles causes the score to drop significantly instantly. The explanation for the decrease of the score for changes in angle of attack is given as the result of the change in amplitude of the striation profile and for changes in axial rotation angle as a result of obstruction and compression of striations.

Garcia et al. [102] used five electricians chisels and created toolmarks in lead at a constant angle of attack of 0°, varying axial rotation angles from 0, 15, 30, 45 to 60 and 75°. Marks were acquired using an Alicona G4 IFM and automatically compared using the in-house developed software Scratch [106,107]. Comparison was done between two marks at identical angles as well as marks at different angles. The results show that the same angle scores were much larger than known non-matching scores and the variance was low. The absolute angle did influence the score though, with higher angles leading to a slightly decreasing score. Comparing different angles showed that while an angle difference of 15° still yielded scores similar to same angle scores, the scores dropped at 30° and higher to the known non-matching range. The reason for this is the fact that the registration algorithm implemented in Scratch does not only correct for translation, but also to a small degree for scaling. For small angle difference, the scaling seems to compensate for the compression. In a second set of experiments it was tested, whether a toolmark of an axially rotated chisel, thus a compressed toolmark, could be re-sized (stretched) and compared to a toolmark made at 0°. Results show that the obtained scores are lower than same angle comparisons but still much higher than known non-matching scores. The decrease in score was explained by the fact that details are disappearing, that geometric relations between striations are distorted and that striations are obstructed. Finally, 3D surfaces of the chisel tips were acquired and used to create virtual toolmarks [108] for an in-depth assessment of what happens when the tool is rotated axially and to predict the axial rotation angle from a real toolmark. This seems to be possible up to a rotation angle of 45° with an accuracy of about three degrees.

One author examined the individuality of lathe chuck jaw impressions. In Finkelstein et al. [109], the authors studied whether the chuck jaws of lathes, which are frequently used by criminals to fix and modify firearm parts, leave distinctive impression marks in improvised rifle barrels. They took forty five metal rods and tubes, wrapped them in 1.5 mm thick lead sheets and fixed them in ten different lathe chucks, all of the same type. The resulting impression marks were then studied qualitatively with comparison microscopes. The authors conclude that class as well as individual characteristics are present in the mark and that these are distinctive for a particular lathe chuck.

## 5.5. Automated mark comparison

To date, most algorithms to automatically compare striated toolmarks are based on explicitly choosing the similarity metric, with which two marks are compared, e.g. global [103,107] and local [103,110] cross correlation or relative distance [111]. In two recent articles however, mark similarity is based on multi feature vector comparison [30] and employing a convolutional neural network [31].

Hare et al. [30] developed a multi feature score based algorithm

to compare bullet land impressions. The first part of the article is focusing on pre-processing striated marks on bullets specifically, but the similarity measurement step in the second part could also be applied to toolmarks. In contrast to using one measure of similarity, they suggest to compare mark feature vectors, which are constructed by first identifying peaks and valleys in striated mark profiles and then measure a series of five features. A decision tree is employed to finally predict if two marks are a match or not, based on the feature vector. The method is tested on the Hamby dataset [112] and 88 lands from unknown bullets are compared to 118 lands of known bullets. The results show that all actual matches resulted in a significantly higher predicted match probability than the non-matches.

To date, most algorithms aim at comparing toolmarks for subsequent determination of error rates or likelihood ratios. These systems are often rather slow and comparison of a query with hundreds or thousands of marks in a database may take a long time. Therefore, Keglevic et al. [31] present a search engine that can be used for fast toolmark retrieval from a database. The approach is based on using the convolutional neural network TripNet for fast calculation of similarities between toolmarks. The aim of the method is invariance to lighting conditions, as the method is supposed to work with 2D images, substrate and angle of attack. The authors test the method on a publicly available dataset of the Netherlands Forensic Institute consisting of 250 screwdriver marks (fifty tools at five angles). For each, 3D surface data and a 2D RGB image (both acquired with an Alicona G4 IFM microscope) as well as a 1D profile are available. For each angle of attack, the network was trained with data from all *other* angles. They compared the performance to a baseline based on elastic shape matching and showed that their TripNet outperforms the baseline, particularly for toolmarks that differ largely in angle of attack. With increasing difference though, also the TripNet performance decreased. Retrieval of toolmark images of unseen tools however performed similar to the baseline. The authors conclude that the retrieval result is robust with respect to large differences in the angle of attack, but that the algorithm to date has difficulties to generalize to unseen tools.

The same authors applied a different type of convolutional neural network for database retrieval of impression marks made by adjustable wrenches on lock cylinders. In Keglevic et al. [32] the FORMS-Locks database is presented, consisting of comparison microscopy images of marks of forty-eight distinct tools (ninety-six tool jaws of wrenches), acquired with eleven different angles of illumination. Besides these images, manually annotated impression mark contours and local image patches along these contours are provided for all illumination conditions. To measure image similarity, a convolutional neural network was implemented based on three selections of local image patches and applied for mark retrieval. The first was including all patches at fixed orientation, the second was including all patches at random orientations and the last was including only patches at the same location along the contour. The results show that with 31.68% false positive rate at 95% recall the best performance was achieved in the last condition. The authors state that there is room for improvement and that these results should be considered as baseline. The Cumulative Match Probability was provided in Ref. [113] and was 70% within the first 20% of the size of the database (twenty-five image).

Hadler et al. [110] describe a possibility to improve a previously published method by the same authors [103], that aims at automatically comparing striated toolmark profiles. In the original algorithm, a set of correlations is determined in local data windows that are shifted, but as the selection of the locations of these windows is random, the resulting distribution of similarity scores can be slightly different each time the algorithm is executed with the

same two marks. In addition, the windows cannot be considered independent. As a remedy, the authors proposed to normalize the profiles by subtracting the baseline and use deterministic window selection to remove randomness. As the windows are not allowed to overlap, they are assumed to be independent. The modified algorithm was tested using fifty sequentially manufactured screwdriver tips and marks made at 30, 40 and 50° angle of attack (with respect to the substrate). In total, fifty pairs of matching and non-matching marks were compared. The results show that the algorithm provides comparable separation between KM and KNM U statistic values as reported in the original article and that the U statistic distribution is now normally distributed. The latter is interpreted as a proof of the independence of the individual U statistic values.

Using multi feature vectors instead of choosing a single similarity metric is attractive, as several features of variable 'strength' might yield better performance if combined. However, this strategy also requires a reliable feature detector rather than using the data as is. Neural networks have the advantage that it is not necessary to explicitly define a measure of similarity at all, but rather let an algorithm decide which features are valuable to distinguish between matching and non-matching marks. The disadvantage is that typically a lot of data is needed to properly train a neural network. In fact, the method presented in this section relying on a neural network does not generalize well [31].

All the presented methods are either mainly applied and tested in research environments or the performance is not good enough yet for application in practice.

### 5.6. Digital reference databases

Digital reference databases for toolmarks are very rare. One publicly available database is provided by the Netherlands Forensic Institute [114], and contains three hundred datasets of fifty different flat-head screwdrivers (five angles of attack per screwdriver plus repeated measurements at one angle of attack). The marks were made in wax sheets, casted and 3D surface datasets acquired with a focus variation acquisition device that also provides 2D RGB images of the toolmarks. This database does not include real crime scene marks. Another database is FORMS-Locks [115], which consists of comparison microscopy images of real crime scene marks of forty-eight distinct wrenches (ninety-six tool jaws), acquired with eleven different angles of illumination. Besides these images, manually annotated impression mark contours and local image patches along these contours are provided for all illumination conditions.

### 5.7. Weighing the evidence/interpretation

In Dutton [116], the author studies the feasibility of using the likelihood ratio (LR) or Bayesian approach in Australian laboratories instead of the currently used AFTE range of conclusions [117] and discusses practical benefits and future challenges of the framework. In short, the LR approach includes testing the probability of an outcome of an examination given two competing hypotheses (same source and different source hypothesis) and yields a continuous numerical outcome. In contrast, the AFTE range of conclusions approach yields one of a range of categorical conclusions like "identification" or "elimination" [117]. In the article, the biggest advantages are pinpointed as the possibility to more accurately provide weight to the evidence where an identification framework would yield an "inconclusive" and that the framework is logically defensible. The main disadvantages are that implementing an evaluative framework is a long and cumbersome process that requires extra staff and sufficient means and that understanding the

framework may be a challenge for examiners and jurors alike. In addition, large mark databases are required to calculate an LR reliably and that to date in many disciplines these databases are not available yet. However it is also noted that a verbal scale could be used if the amount of available data is insufficient. In addition, the author states that different jurors may give incorrect weight to the verbal scales. Finally, the article describes a path towards implementation of an LR framework in practice.

As firearm and toolmark examiners are frequently confronted in US courts with the claim that the results of examinations lack a scientific basis, Murdock et al. [118] present a paper with the purpose to provide counter arguments to lawyers and academics claiming that there are no random match probabilities and error rates available for forensic firearm and toolmark examination. In addition, a review of literature dealing with random match probabilities and statistical applications is provided.

More articles were published on this topic in the last years (e.g. Ref. [119–121]), but are typically targeted specifically at firearm mark examination. However as the evidence evaluation step is typically also applicable to toolmarks, we refer the interested reader to the chapter 'Examination of Firearms' by E. J. A. T. Mattijssen in this collection of reviews.

### 5.8. Software for toolmark analysis

In daily practice toolmark evidence is typically compared using comparison microscopes. To this end an experimental mark created in the lab is manually moved relative to a suspect mark, to determine (dis-)similarities of toolmark characteristics, e.g. striations. The examiner moves the marks in real time and studies (dis-) similarities directly on what can be seen through the microscope. Comparing 3D surface data of marks is more complicated, as the data is not available in real time and has to be acquired first. To provide examiners the means to compare 3D data in a familiar environment, so-called Virtual Comparison Microscopes (VCM) were developed in recent years. Duez et al. [29] demonstrate software, that contains such a VCM. The software was developed specifically for firearm mark comparisons, but can also be used for comparing toolmarks. After loading two 3D surface datasets of toolmarks, both are presented side by side and can be translated and rotated independently, just like in a conventional microscope. In addition, it is possible to zoom in and zoom out to mimic different microscope magnifications (note that the original resolution of the data does *not* change when zooming). After aligning the marks, the viewers can be locked, to simultaneously translate, rotate or scale the marks. The system was validated by fifty-six participants at fifteen laboratories using cartridge case impressions and aperture shear marks and the results show that trained examiners can successfully use virtual microscopy in casework. In fact, the firearms/toolmarks unit of the Federal Bureau of Investigation is already using VCM in daily casework. In the future, VCM software should be validated for comparing toolmark evidence as well. Software for viewing marks is available free of cost [122].

A mobile system that combines an optical 3D topography scanner with software to acquire and compare toolmarks was proposed by Chumbley et al. [123]. The hardware consists of an Alicona SL IFM, a compact and portable 3D surface acquisition system and a laptop with the Alicona data acquisition software and the in-house developed Mark and Tool Inspection Suite (MANTIS) software. After acquisition, the tool or mark data can be imported into MANTIS and studied visually. For manually comparing marks a Virtual Comparison Microscope, and for automated comparison, an objective mark similarity determination [103] is available. The software also allows deriving virtual profiles from measured tool surfaces, depending on the angle of attack, and search for the most

likely angle of attack with which a mark of the tool was made. The authors note that although the system can be used for automated comparison of marks with higher complexity than striated tool-marks, e.g. impression marks, the performance of the algorithm will decrease. They point out that the comparison algorithm was initially developed for striated marks and encourage other parties to contribute to their software with more advanced algorithms. The system is built using open-source libraries and software and therefore enables integration of third party algorithms.

So far, available software packages mainly focus on visualization and manual and/or automated alignment of marks and determination of mark similarity, which could subsequently be used for database retrieval. The software package 'Scratch' [106] provides a graphical user interface to visualize and automatically compare striated marks of tools and firearms (e.g. land engraved areas or LEAs and primer shear marks), also with multiple LEAs simultaneously, using an algorithm presented earlier [107]. In addition, the software can determine virtual toolmarks from tool surface data and compare them to experimental toolmarks for angle of attack retrieval. Furthermore, the software provides a simple interface to set up toolmark and firearm mark databases and determine reference known match and known non-match distributions, which can subsequently be used to determine likelihood ratios for a comparison result. A virtual comparison microscope is also included.

The current structure of Scratch only supports building local databases and provides a limited amount of metadata. Based on the existing infrastructure, the functionality was extended [35] to be able to automatically compare not only striated but also impression marks, by incorporating algorithms developed by NIST [33,34]. In addition, the database functionality was greatly extended, to allow setting up large and diverse databases. A variety of objective measures of similarity and statistical statements of uncertainty will also be available in the future. The setup of the new system is mainly focusing on firearm marks like striated bullet marks, aperture shear marks and breech face impression marks, but the setup is such that it could also be used for toolmarks.

## 6. Invasive striated and impression toolmarks

### 6.1. Detection/creation

Forensic examiners need to create experimental mark in the lab and the circumstances should ideally be identical to the situation at a crime scene. As this is not possible, alternative methods have to be used but it has to be shown, that those yield similar results. One of the variables that play a role is the substrate material, which should be similar to the material in which the suspect marks were made. As human bones are not readily available, animal bones could be an alternative.

Croker et al. [124] compared the major limb bones (humerus, radius, femur and tibia) of fifty adults as well as the corresponding bones of sheep, pigs, cattle, large dogs and kangaroos. Specifically, the authors determined bone shaft diameter, cortical bone thickness and a cortical thickness index, the sum of the thickness of both cortices divided by the diameter, at various points along the shaft. They show that although the absolute thickness varies, the cortical thickness index does only slightly vary between the species. Properties like bone density however have not been studied.

As bone material might have to be frozen prior to creating toolmarks the question arises whether that has an influence on the bone properties. In Hale et al. [125], the authors studies the impact of freezing over time on bone mineral density (BMD). For eight fetal pigs, the BMD was determined using an X-Ray acquisition device, first on fresh samples and then repeatedly over a period of twenty weeks, after which they were thawed again. Based on the

measurement results the authors conclude that freezing seems to not influence the BMD but that samples should be thawed entirely to avoid erroneous measurements in the X-Ray images.

Realistic application of stabs to bones is also important. Benson et al. [126,127] present a prototype of a stabbing machine with an interchangeable knife holder. Using a motorized arm and a pneumatic system, sixty unique stabbing positions can be set up and the stabbing force is variable with a maximum of 221 N. The machine was evaluated with textile cuts, but might also be useful to create stabbing marks in bone.

### 6.2. Acquisition

A noticeable trend in the acquisition of invasive marks is a large variation in used techniques and methods. Still conventional 2D microscopy is used, but the majority of articles describe different methods such as computed tomography (CT), 3D microscopy, reflectance transformation imaging (RTI) and scanning electron microscopy (SEM). Conventional 2D microscopy is mostly used for qualitative assessment of marks. The other methods have been demonstrated to be superior for quantitative assessment of mark properties and are therefore applied more frequently nowadays, particularly Micro-CT. Qualitative assessment is also still used on the other 3D methods, such as Dittmar [128] describing qualitative toolmark assessment using SEM on archaeological material.

A comparison of stereomicroscopy with Micro-CT is provided in Pelletti et al. [36], where thirty-two false starts were created with four different types of hand saws in human bone and subsequently analyzed using stereomicroscopy and Micro-CT. The authors were particularly focusing on the potential of the imaging techniques to determine the morphology of the marks. The qualitative analysis results showed that false starts and their shape can be more accurately determined with Micro-CT. In a sequel study [37], the same authors studied the accuracy, precision and inter-rater reliability with respect to manual saw mark analysis on Micro-CT images. Three forensic pathologists and/or radiologists were asked to measure a set of four features, including kerf width and depth, on twenty-four false start lesions in bone, created with three different saw types. The measurement results were subsequently compared statistically and the authors conclude that they were reproducible and robust.

The previous studies were conducted on a limited set of samples. Norman et al. [44] studied whether using Micro-CT is a suitable technique for saw mark analysis using 270 samples. Based on measurements of a set of seven features by two independent raters they conclude that the technique is powerful and reliable to determine toolmark class characteristics, as the reproducibility was high.

A variant on stereo-microscopy was described in Cerutti et al. [129]. They describe a method of making thin cross-sections from the inflicted lesions which can be analyzed by light microscopy as it is being done in (medical) histology analysis. Although the method is destructive on the lesion, very detailed analysis of the morphology of the cross-section is possible. In this study the lesions were inflicted on old bone material. Therefore, the conclusions might not be valid for lesions inflicted into fresh bone material.

In order to perform robust quantitative assessments, it is necessary to study the robustness of the acquisition technique as well as comparing different acquisition techniques with each other to choose the right technique for an application. Most articles that follow are describing this.

Shamata et al. [130] describe the key considerations and best practice of using 3D scanning with structured light. Only focusing on the area of interest, combining three scans and elimination

background noise by using a black background gives the best results. The application is injuries on living individuals. Reynolds et al. [131] describe the robustness of using CT imaging combined with CAD software for measuring anthropological features of postcranial bones. Both the intra-observer and inter-observer variation is small, resulting in a highly repeatable approach. The measured features are relatively large compared to features used in toolmark analysis.

LeGarff et al. [132] describes the importance of knowing the precision of a Micro-CT imaging device and the effect of using a registration method. Using registration in Micro-CT imaging increases the precision of measurements.

Clarke et al. [133] describes the pro and cons of using reflectance transformation imaging (RTI) to preserve and analyze saw marks in bone. RTI was found to be excellent for visualizing toolmarks on bone, though more successful in shallow details than deeper marks. Large file sizes and time consumption are limitations for RTI, while a low direct cost of equipment is a pro.

Besides accurate acquisition it might sometimes be required to demonstrate 3D models of evidence including toolmarks in course to support reports. This requires accurate 3D data acquisition on the one hand, but accurate data reproduction on the other hand. In Baier et al. [134,135] the authors present a system that first scans an object accurately in 3D with a Micro-CT scanner, then uses dedicated software to process the data and segment relevant bone structures and subsequently prints a copy of bone models. They successfully applied this technique in two cases with a fractured humerus [135] and an injured skull [134]. Although the reported scan resolutions 36–80 μm might not be sufficient yet to accurately reproduce details in toolmarks, the resolution of 14 μm reported in Pelletti et al. [36,37] might be. In addition there are already Micro-CT scanners available that provide a much higher resolution, but that typically comes with a decrease in possible object size (e.g. Refs. [136]).

## 6.3. Occurrence of marks

Wood chippers are sometimes used by criminals to dispose bodies with the aim to destroy evidence that may lead to identification of the victim. In Domenick et al. [137] the authors studied the occurrence of potentially useful toolmarks on bone, after being processed in a wood chipper. They used five domestic pig limbs, put these in a home model wood chipper and subsequently assessed the size of the resulting bone fragments. In addition, they looked for potentially useful toolmarks. The most common size of the bone fragments was between 5.85 and 11.6 mm and typically the fragments were relatively flat chips. Striated toolmarks were present on some fragments, but those were rare. In addition, incomplete cuts were observed. The authors conclude that wood chippers produce useful marks for comparison.

## 6.4. Variability/individuality of invasive toolmark characteristics

Many articles in the last years were published studying the variability and individuality of toolmark characteristics in bone. As a large body of literature was focusing on saw marks, these articles were bundled in a subsection.

### 6.4.1. Saw marks

Nogueira et al. [42] studied 170 experimental false start lesions made with five different hand saw types (four with an alternating set of teeth and one with a wavy set) on pig and human femora. Three features, minimum kerf width, shape of the kerf profile and the shape of the kerf walls were measured manually with a stereomicroscope and analysis software, and subsequently compared statistically. The chosen features proved to be useful to distinguish

between the tested saw types, although some variability between lesions of the same type was encountered. Another outcome of the study is that significant differences in lesions between pig bones and human bones were encountered and the authors conclude that pig femurs might not always be a good alternative to human femurs for creating experimental saw marks. The human donors were all of high age however, which might also have an effect on the marks, as bone properties change with age. In a sequel article [43], the authors studied "secondary features" of false start lesions, in addition to the three main features mentioned above, particularly in cases where the main features lead to some ambiguity. For this study, they used the same data and analysis methods. Of these secondary features, striae on the kerf floor seemed to be useful to distinguish between an alternating vs. a wavy set of the teeth, while blade drift and bone islands may be an indication of a large saw tooth size.

Greer et al. [45] presented a study that aimed at quantifying the variation in kerf wall striations in bone lesions caused by hacksaws and reciprocating saws. In total, eighty-seven lesions were applied on juvenile pig femora with eight different hacksaw blades and six different reciprocating saw blades. Surface data of the striated walls was determined from a stack of 2D images, acquired with a stereomicroscope. Quantitative analysis of a set of surface metrology measures revealed, that while the distributions of the measured amplitudes of striations caused by hacksaws and reciprocating saws partially overlap, the amplitude of the striations produced by hacksaws is much more variable and generally higher than the amplitude of striation of reciprocating saws. Large amplitudes therefore might indicate the usage of a hacksaw.

Another article focused on the differences among different samples of the same class of saws, reciprocating saws, based on seventeen lesion characteristics, including kerf floor shape and minimum kerf width. Berger et al. [46,47] analyzed class characteristics of lesions on white-tailed deer limbs that were created with six different saw blades on bones. They used a stereomicroscope, determined all features manually and statistically analyzed differences between saw blades. They found a set of features including minimum kerf width, kerf false start shape, presence of cut surface drift and harmonics, exit chipping size and striation regularity, that have the potential to distinguish between some of the tested types. The authors note that the differences found between different blades reflect the differences that were found for hand-powered blades in earlier studies.

Finally Norman et al. [44] employed Micro-CT data of saw marks on human long bones to measure seven toolmark characteristics. The goal of the study was to determine the specificity of these measurements, whether they were similar to measurements on the tool blades and whether toolmarks differ under varying methodological conditions (controlled vs. free saw movement and fleshed vs. defleshed bones). To unravel differences, the measured features were compared statistically. Four hand saws, two reciprocating saws and two knives were used to create in total 270 saw marks. Two independent raters where then asked to determine a set of seven features like edge shape, toolmark shape and minimum kerf width. The results show that the set of features was sufficient to distinguish between the different blade types. The comparison of toolmarks and tools showed that only when marks were made under controlled conditions in defleshed bone could the tool be predicted with high accuracy. For the marks in fleshed bone made with free saw movement, the performance dropped significantly. Only one feature, the kerf width, was used for this but it is clear that the methodological condition has a large impact on the resulting mark properties.

In summary, a large body of literature was found regarding the variability and individuality of saw marks on bone. Several types of saws were studied, including hand saws [42–44], hacksaws [45]

and reciprocating saws [44−47] to inflict trauma on human [42−44], deer [46,47], and pig bones [42,43]. Analysis was done using stereomicroscopy [42,43,46,47] and Micro-CT [44], studying a varying set of class features, mainly to distinguish between different tool types based on class characteristics. The amount of features varied greatly between three [42] and seventeen [46,47] and also included surface metrology measures [45]. Lesions were inflicted on fresh bones [37,42−45], after freezing [46,47] and with maceration [138].

Based on this summary it is clear that there is a large variety of approaches and the conditions described in the articles are hardly the same, which makes results difficult to compare. Particularly so, as it has been shown in several publications in this collection of reviews and before, that e.g. macerating the bone, whether the bone is from a human or an animal or whether a mark is created by hand or under controlled circumstances might influence the results. For the future, researchers are thus encouraged to reduce potential variability as much as possible by using fresh and fleshed bones, ideally of humans and apply marks under realistic conditions. Furthermore, marks should ideally be acquired using a 3D method, preferably Micro-CT, as this has been shown repeatedly to yield accurate results. However, so far this has only been demonstrated for class characteristics and whether also individual characteristics can be assessed with Micro-CT still has to be shown. SEM has not yet proven to provide real additional value. Furthermore there is no standardized way of measuring class characteristics in saw marks and as each author chooses a different set of features it is difficult to judge which features actually best describe a mark and are the most distinctive. Further research will be required to address this issue. Finally, most articles demonstrate that marks from different saw blades can be distinguished from each other, but only one article also studies whether a mark can be related to the actual tool that created it. This is an essential step though and should get more attention in research projects in the future.

### 6.4.2. Knife and other cut marks

Many articles in the last years are studying the variability of invasive marks of knives and other tools. All are focusing on class characteristics like morphological features, i.e. characteristics that discriminate between different classes of tools, however these characteristics do not discriminate between tools from the same class. Conclusions from different studies vary. For example, some conclude that it is possible to discriminate between serrated and non-serrated knifes [38−40] while other studies are more cautious. Tennick [41] tested many morphological features and concludes that they are not useful for mark classification. Komo et al. [139] report on the complexity of the allocation of a knife to a particular bone lesion. Caution is advised regarding classifying kerf marks. Not all kerfs resulting from serrated blades show characteristic striations. Furthermore, marks made with the same knife can show variation in morphology.

Interpreting the conclusions from the different publications is very difficult since there is a lot of variation in methodologies used. Marks are made by hand [41], more representative for real casework, while other studies produce marks under more controlled circumstances with different machines [138−140]. Marks are made in fresh bone material, more representative for real casework, while other studies inflict marks into macerated or old bone material. Bone material is used from animals, mostly pigs, while other studies have human material available. However, the human material is rare and frequently related to elderly people, less representative for average casework. Furthermore, the technology used for analysis varies from 2D photography and rulers to 3D microscopy, Micro-CT and scanning electron microscopy (SEM). The analysis of the marks varies from qualitative analysis and

qualitative comparison (e.g. Ref. [38,129]) to more quantitative analysis and computer assisted comparison [139,141−143]. A comment should be made that both the qualitative and quantitative analysis in all studies is based on subjective interpretation from the examiner. In the qualitative analysis the examiner makes subjective interpretations e.g. on the shape of a kerf. In the quantitative analysis the examiner makes a subjective interpretation on the location of measurement points, since the boundaries are mostly not sharp. Some of these variations are studied by Norman et al. [40]. They analyzed the difference in the cutting mark properties mark width, wall angle and shape (Y-, T- or V-shape) between two different types of knives, plain and serrated. They used two sets of experiments, one using macerated and one using fresh porcine ribs and acquired the data using conventional microscopy and Micro-CT. They conclude that the shape properties, except the wall angle, are significantly different between the two types of knife that they are able to predict which type of knife was used and that knife edge thickness correlates with cut mark width. They compare Micro-CT with conventional microscopy for their potential to assess the cut mark shape and conclude that Micro-CT is superior. The authors conclude that the wall angle is not a reliable measure to derive the knife cutting angle.

An interesting trend from the recent literature, especially in the field of archeology and anthropology is using morphometrics for the analysis and comparison of marks [138,139,141−143]. Morphometrics helps in the statistical evaluation of the morphological features of marks and objects and in the comparison of these features. Courtenay et al. [143] use morphometrics to successfully distinguish morphological differences in cut marks produced by different lithic tool types and raw materials. Komo et al. [139] show that morphometrics could serve as a tool in a forensic examination of kerf marks in ribs, for example on the distance between walls of the kerf related to the blade thickness. Furthermore, they show the effect of maceration (after inflicting the kerf) on morphometrics. An average shrinking factor up to 8.6% was observed. Mate-Gonzalez et al. [141,142] analyzed a large number of cut marks (572) using micro-photogrammetry and morphometrics. The design of the study is related to an archaeological context, differentiating flints from different raw materials, however a similar design could be used in a more forensic context e.g. differentiating knifes. The study could not differentiate between flints from the same material. The variability of features in hatchet hacking traumas was studied in Nogueira et al. [138]. A hatchet was used to inflict thirty lesions in total in two macerated human tibiae with a specifically designed device. All lesions were then analyzed with the naked eye and stereomicroscopy and a subset of thirteen with scanning electron microscopy (SEM). Based on a morphometric assessment of features observed in the lesions, the authors conclude that it should be possible to determine that a trauma was caused by a hatchet. In this study, SEM did not seem to have an added value, as the relevant features could be observed with the naked eye and/or stereomicroscopy.

Several articles describe exposure effects on invasive marks, such as exposure to heat [144−146] or taphonomic alterations [147]. Macoveciuc et al. [144] conclude that mark signatures associated to sharp and blunt force trauma are not masked by heat exposure. Waltenberger et al. [145] conclude that width, depth, floor radius, slope and opening angle of cut marks remain stable with heat exposure. Alunni et al. [146] conclude that the features associated to hacking trauma of bone are not significantly altered by carbonization (burned). Stanley et al. [147] describe the effect of taphonomic alterations on striations in skin (porcine). They see a big effect and recommend to document skin striations as soon as possible by stereo-optical microscopy.

## 6.5. Manual mark comparison

Digitizing marks and objects does not necessary have to result in computer assisted (semi)automated comparison. A trend that is seen is that the digitized marks (2D or 3D) are manually compared with digitized objects (2D/3D). However the manual comparison is done virtually instead of physically.

Bornik et al. [148] describe software which can be used to visualize 3D data from different modalities such as CT and 3D laser scanning. By combining the different modalities into one visualization it becomes possible to virtually compare the injuries/marks with the physical shapes of an objects like a knife or hammer. Care must be taken that primarily class-characteristics can be visualized and compared.

Urbanova et al. [149] describe this virtual approach on reconstructing human skeletal remains, such a part of a skull or foot. The importance of the virtual approach increases with the complexity and state of preservation of the forensic material. The unlimited and unrestricted handling of the virtual remains enables limitless repairs and adjustments to find the "best-case reconstruction" of the remains, resulting in smaller inter-operator variation in comparison to the traditional approach.

## 6.6. Weighing the evidence/interpretation

An interesting approach is found in Park et al. [150]. They describe the use of known data on offender and victim characteristics of homicides in the past to assist in the investigation of current homicides. They found differences in offender and victim characteristics between blunt force and sharp force injuries. Blunt force is more likely to be committed by offenders who lived with the victims, using a blitz attack and weapon of opportunity. Compared to sharp force injuries, more likely to be committed by offenders who are strangers with a preselected weapon carried with them. According to the authors, the results of this study on south Korean homicides are in correspondence with results in other countries as UK, Germany, India and Sweden.

## 6.7. Software for invasive toolmark analysis

Palomeque-Gonzalez et al. [151] describe a new open source software tool for the morphometric and statistical analysis of cut marks on bone, called *Pandora*. The software is created for archaeological science. However the software seems very valuable for usage in the forensic science domain as well. The set-up of the software is designed to be able to work with input images of marks saved in 'jpg' format. Images need to be set to scale in the software and the morphology of the mark is analyzed by manually placing semi-landmarks. Then a wide range of morphometric and statistical analysis tools can be selected for further analysis of the marks. The database-like design of the software makes it easy to keep data registered to the correct input.

Mahfouz et al. [152] describe another new software tool called *Fragmento*, freely available for research. The software helps in the process of classifying bone fragments to the correct original bone. The bone fragment needs to be digitized in 3D using CT. The software subsequently tries to match and register the bone fragment to bone templates from a bone atlas. Although this work is more in the field of forensic anthropology, it related to toolmarks as well if toolmarks are present in the bone fragments. This software could help in classifying the bone fragment.

A trend visible in invasive toolmark analysis is that more frequently 3D imaging techniques from different modalities such as CT, 3D microscopy, 3D macroscopy and laser scanning are combined. Bornik et al. [148] describe a new software tool to document and present analysis results based on multi-model 3D data. Benefits are that the 3D case illustrations represent an efficient tool to present insights from case analysis to non-experts involved in court proceedings like jurists and laymen. However, there is also a risk. The persuasive power of images and illustrations can easily lead to misunderstanding and influence, especially if they present fragmentary information rather than the 'big picture'.

## 6.8. Case studies including invasive toolmarks

Many articles on invasive toolmark analysis in this review originate from the field of archeology. Valoriani et al. [153] is such an example. The questions in the field of archeology are primarily on the level of class characteristics. For forensic science this can be interesting as well, especially in the investigative phase when no murder weapon is present on the crime scene.

In Quatrehomme et al. [154] a case of a victim with a blunt trauma in the skull mimicking a gunshot wound is described. The trauma was a round hole, with typical internal beveling. As it turned out however, the hole was caused by a rib of a beach umbrella.

In Baier et al. [134,135], two cases are described including a fractured humerus [135] and an injured skull [134]. In both cases, the authors used high resolution Micro-CT scanners, to acquire 3D data of the traumas and used the resulting volume datasets and dedicate volume rendering software, to aid their investigation and to be able to more clearly demonstrate the results. In addition, 3D prints of the injured bone parts were created and used for demonstration purposes.

## Disclaimer

This is a republication in journal form of a conference proceeding that was produced for the 19th Interpol Forensic Science Managers Symposium in 2019 and was originally published online at the Interpol website: https://www.interpol.int/content/download/14458/file/Interpol Review Papers 2019.pdf. The publication process was coordinated for the Symposium by the Interpol Organizing Committee and the proceeding was not individually commissioned or externally reviewed by the journal. The article provides a summation of published literature from the previous 3 years (2016-2019) in the field of shoe and tool mark examination and does not contain any experimental data. Any opinions expressed are solely those of the authors and do not necessarily represent those of their agencies, institutions, governments, Interpol, or the journal.

## Declaration of Competing Interests

The authors have no competing interests to declare.

## References

[1] M. Baiker, The forensic examination of marks, in: Proc of the 18th Interpol International Forensic Managers Symposium, Published 2016. Available: https://www.interpol.int/how-we-work/forensics/forensic-symposium. (Accessed 16 August 2019).

[2] Interpol, International forensic science Managers Symposium review paper collection, Available: https://www.interpol.int/How-we-work/Forensics/Forensic-Symposium. Last accessed: 16-8-2019.

[3] Committee on Identifying the Needs of the Forensic Sciences Community: National Research Council, *Strengthening Forensic Science In the United States: A Path Forward*, The National Academies Press, Washington, DC, USA, 2009.

[4] Pcast, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, 2016.

[5] Osac, Footwear and tire subcommittee, Available: https://www.nist.gov/topics/forensic-science/footwear-tire-subcommittee, 16-8-2019.

[6] Osac, Firearms and toolmarks subcommittee, Available: https://www.nist.

gov/topics/forensic-science/firearms-toolmarks-subcommittee, 16-8-2019.

[7] Osac, Physics/pattern interpretation, Available: https://www.nist.gov/topics/forensic-science/physicspattern-interpretation-scientific-area-committee, 16-8-2019.

[8] Swgtread, Tread forensics, Available: https://treadforensics.com/index.php/about, 29-7-2019.

[9] M.R. Bennett, M. Budka, Digital Technology for Forensic Footwear Analysis and Vertebrate Ichnology, Springer, 2018.

[10] Fraunhofer, Kolibri cordless, Available: https://www.iof.fraunhofer.de/en/business-fields/photonic-sensors-and-measuring-systems/3d-measuring-systems/handheld-optical-3d-scanner-kolibri-cordless.html, 1-8-2019.

[11] W.J. Bodziak, Forensic Footwear Evidence, Practical Aspects of Criminal and Forensic Investigations, CRC Press, 2017.

[12] C. Snyder, A comparison of photography and casting methods of footwear impressions in different sandy soil substrates, J. Forensic Ident. 66 (2016) 37–58.

[13] CloudCompare, Available: https://www.danielgm.net/cc/, 6-8-2019.

[14] DigTrace, Available: http://www.digtrace.co.uk/, 15-10-2018.

[15] A. Kortylewski, Model-based Image Analysis for Forensic Shoe Print Recognition, PhD thesis, University of Basel, 2017.

[16] X. Wang, C. Zhang, Y. Wu, Y. Shu, A manifold ranking based method using hybrid features for crime scene shoeprint retrieval, Multimed. Tool. Appl. 76 (2017) 21629–21649.

[17] J. Cui, X. Zhao, N. Liu, S. Morgachev, D. Li, Robust shoeprint retrieval method based on local-to-global feature matching for real crime scenes, J. Forensic Sci. 64 (2) (2018) 422–430.

[18] B. Kong, J. Supancic, D. Ramanan, C.C. Fowlkes, Cross-domain image matching with deep feature maps, Int. J. Comput. Vis. 127 (11-12) (2019) 1738–1750.

[19] Hobbit_Imaging_Solutions, Pride - shoeprint matcher, Available: https://hobbit-is.nl/projects/forensic-intelligence/pride/?lang=en, 10-10-2018.

[20] Everspry, EverASM, Available: http://www.footprintmatcher.com/products/EverASM.pdf, 15-8-2019.

[21] S. Park, Learning Algorithms for Forensic Science Applications, PhD thesis, Iowa State University, 2018.

[22] J.A. Speir, N. Richetelli, M. Fagert, M. Hite, W.J. Bodziak, Quantifying randomly acquired characteristics on outsoles in terms of shape and position, Forensic Sci. Int. 266 (2016) 399–411.

[23] N. Kaplan Damary, M. Mandel, S. Wiesner, Y. Yekutieli, Y. Shor, C. Spiegelman, Dependence among randomly acquired characteristics on shoeprints and their features, Forensic Sci. Int. 283 (2018) 173–179.

[24] R.S. Stone, Footwear examinations mathematical probabilities of theoretical individual characteristics, J. Forensic Ident. 56 (2015) 671–692.

[25] Y. Shor, S. Wiesner, T. Tsach, R. Gurel, Y. Yekutieli, Inherent variation in multiple shoe-sole test impressions, Forensic Sci. Int. 285 (2018) 189–203.

[26] P.J. De Groot, Progress in the specification of optical instruments for the measurement of surface form and texture, Proc. SPIE 9110 (2014). Dimensional Optical Metrology and Inspection for Practical Applications III.

[27] M.T. Stocker, R.M. Thompson, J.A. Soons, T.B. Renegar, X.A. Zheng, Addressing challenges in quality assurance of 3D topography measurements for firearm and toolmark identification, AFTE J. 50 (2018) 104–111.

[28] Cadre, TopMatch-3D, Available: https://www.cadreforensics.com/, 16-8-2019.

[29] P. Duez, T. Weller, M. Brubaker, R.E. Hockensmith 2nd, R. Lilien, Development and validation of a virtual examination tool for firearm forensics, J. Forensic Sci. 63 (Jul 2018) 1069–1084.

[30] E. Hare, H. Hofmann, A. Carriquiry, Algorithmic approaches to match degraded land impressions, Law Probab. Risk 16 (4) (2017) 203–221, 0.

[31] M. Keglevic, R. Sablatnig, Retrieval of striated toolmarks using convolutional neural networks, IET Comput. Vis. 11 (2017) 613–619.

[32] M. Keglevic, R. Sablatnig, FORMS-locks: a dataset for the evaluation of similarity measures for forensic toolmark images, in: Presented at the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW), 2017.

[33] J. Song, Proposed "congruent matching cells (CMC)" method for ballistic identification and error rate estimation, AFTE J. 47 (2015) 177–185.

[34] J. Song, Proposed "NIST ballistics identification system (NBIS)" based on 3D topography measurements on correlation cells, AFTE J. 45 (2013) 184–194.

[35] X.A. Zheng, J. Soons, E. Smith, M. Baiker-Sørensen, Reference population database of firearm toolmarks (RPDFT), in: Presented at the 50th Annual Training Seminar of the Association of Firearm and Tool Mark Examiners, Nashville, TN, USA, 2019.

[36] G. Pelletti, G. Viel, P. Fais, A. Viero, S. Visentin, D. Miotto, et al., Micro-computed tomography of false starts produced on bone by different hand-saws, Leg. Med. 26 (May 2017) 1–5.

[37] G. Pelletti, G. Cecchetto, A. Viero, P. Fais, M. Weber, D. Miotto, et al., Accuracy, precision and inter-rater reliability of micro-CT analysis of false starts on bones. A preliminary validation study, Leg. Med. 29 (Nov 2017) 38–43.

[38] C. Humphrey, J. Kumaratilake, M. Henneberg, Characteristics of bone injuries resulting from knife wounds incised with different forces, J. Forensic Sci. 62 (6) (Nov 2017) 1445–1451.

[39] M. Ghui, C. Eliopoulos, M. Borrini, A Methodology in Differentiating between Knives from Cut Marks on Bone, 2016. https://www.researchgate.net/profile/Melodi_Ghui/publication/296484440_A_Methodology_in_Differentiating_Between_Knives_from_Cut_Marks_on_Bone/links/

56d5b61e08aebe4638ac63a5.pdf. (Accessed 20 March 2020).

[40] D.G. Norman, D.G. Watson, B. Burnett, P.M. Fenne, M.A. Williams, The cutting edge - micro-CT for quantitative toolmark analysis of sharp force trauma to bone, Forensic Sci. Int. 283 (Feb 2018) 156–172.

[41] C. Tennick, Cutting-edge Research: Identification and Classification of Knife Trauma on Bone, 2017.

[42] L. Nogueira, G. Quatrehomme, C. Rallon, P. Adalian, V. Alunni, Saw marks in bones: a study of 170 experimental false start lesions, Forensic Sci. Int. 268 (Nov 2016) 123–130.

[43] L. Nogueira, V. Alunni, C. Bernardi, G. Quatrehomme, Saw marks in bones: a study of "secondary features" of false start lesions, Forensic Sci. Int. 290 (Sep 2018) 157–161.

[44] D.G. Norman, W. Baier, D.G. Watson, B. Burnett, M. Painter, M.A. Williams, Micro-CT for saw mark analysis on human bone, Forensic Sci. Int. 293 (Nov 3 2018) 91–100.

[45] S.Y. Greer, Quantifying variation in kerf wall striations created by hacksaws and reciprocating saws, Forens. Anthropol. 1 (2018) 160–169.

[46] J.M. Berger, J.T. Pokines, T.L. Moore, Analysis of class characteristics of reciprocating saws, J. Forensic Sci. 63 (6) (2018) 1661–1672.

[47] J.M. Berger, J.T. Pokines, T.L. Moore, Reciprocating Saws as Tools of Dismemberment, 2017. Available online: https://www.researchgate.net/profile/James_Pokines/publication/314672877_Reciprocating_saws_as_tools_of_dismemberment/links/58e42552a6fdccc85bdf35e7/Reciprocating-saws-as-tools-of-dismemberment.pdf. (Accessed 20 March 2020).

[48] N. Petraco, N.D.K. Petraco, Impression evidence: footwear and tool marks cases from the crime scene to the courtroom, in: Forensic Science Handbook, vol. 2, CRC Press, 2016.

[49] L.R. Whitlow, An Investigation of Test Impression Methods to Accurately Reproduce Randomly Acquired Characteristics in Footwear Outsoles, *MSc thesis*, Emporia State University, 2017.

[50] S. Wiesner, Y. Shor, A. Grafit, A. Cohen, 3D scanning of shoeprints, in: Presented at the 11th Meeting of the SPTM, Prague, 2014.

[51] NextEngine, NextEngine 3D laser scanner, Available: http://www.nextengine.com/, 15-8-2019.

[52] T.J.U. Thompson, P. Norris, A new method for the recovery and evidential comparison of footwear impressions using 3D structured light scanning, Sci. Justice. 58 (2018) 237–243.

[53] 4DDynamics. PicoScan structured light scanning system, Available: http://www.4ddynamics.com/.

[54] Gom, Atos series scanners, Available: https://www.gom.com/metrology-systems/atos.html, 1-8-2019.

[55] Fraunhofer, 3D forensics, Available: https://www.3d-forensics.de/, 2-8-2019.

[56] Artec, Artec Eva scanners, Available: https://www.artec3d.com/portable-3d-scanners, 1-8-2019.

[57] David, SLS-2, Available: https://www.david-3d.com/en/support/david4/introduction, 15-8-2019.

[58] S. Faulkner, Photogrammetry of 3D Footwear Impressions Forensic Applications, MSc, Murdoch University, Western Australia, 2017.

[59] Agisoft, Available: https://www.agisoft.com/, 29-7-2019.

[60] A.F.J. Marchal, P. Lejeune, P.J.N. de Bruyn, Virtual plaster cast: digital 3D modelling of lion paws and tracks using close-range photogrammetry, J. Zool. 300 (2016) 111–119.

[61] Evident, Everspry Everlss 360 mark impression scanner, Available: https://www.shopevident.com/category/everspry/everspry-everlss-360-mark-ii-impression-scanner, 26-7-2019.

[62] Fraunhofer, Fraunhofer Institute for Applied Optics and Precision Engineering IOF, E-mail communication, 2019.

[63] K.H. Reddy, S. Sidda, A Novel Shoeprint Enhancement Method for Forensic Evidence Using Sparse Representation Method, MSc thesis, Blekinge Institute of Technology, 2017.

[64] N. Petraco, H. Sherman, A. Dumitra, M. Roberts, Casting of 3-dimensional footwear prints in snow with foam blocks, Forensic Sci. Int. 263 (2016) 147–151.

[65] A.R. Sabolich, A comparison of hydrophobic barriers for casting footwear impressions in water-soluble food products, J. Forensic Ident. 68 (2018) 207–221.

[66] S. Hong, M. Park, Collection of wet-origin footwear impressions on various surfaces using an electrostatic dust print lifter, J. Forensic Sci. 63 (2018) 1516–1520.

[67] R.L. McElhone, G.E. Meakin, J.C. French, T. Alexander, R.M. Morgan, Simulating forensic casework scenarios in experimental studies: the generation of footwear marks in blood, Forensic Sci. Int. 264 (2016) 34–40.

[68] N. Richetelli, M. Nobel, W.J. Bodziak, J.A. Speir, Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape, Forensic Sci. Int. 270 (2017) 211–222.

[69] G.L.C. Yee, Y.J. Hock, T. Nishiwaki, K. Moriyasu, K. Harano, A method to evaluate abrasion of shoe-sole using 3D scanning technique, J. Sports Sci. 4 (2016) 279–286.

[70] N. Richetelli, M.C. Lee, C.A. Lasky, M.E. Gump, J.A. Speir, Classification of footwear outsole patterns using Fourier transform and local interest points, Forensic Sci. Int. 275 (2017) 102–109.

[71] FosterFreeman, SoleMate, Available: http://www.fosterfreeman.co.uk/trace-evidence/357-sicar6-solemate-2.html, 04-01-2019.

[72] C.Y. Gwo, C.H. Wei, Shoeprint retrieval: core point alignment for pattern

comparison, Sci. Justice 56 (2016) 341–350.

[73] C.H. Wei, C.Y. Gwo, Alignment of core point for shoeprint analysis and retrieval, Inform. Sci Electron. Electric Eng. (ISEEE), 2014 Int. Conf. on 2 (2014) 4.

[74] F. Cervelli, F. Dardi, S. Carrato, Comparison of footwear retrieval systems for synthetic and real shoe marks, in: 6th International Symposium on Image and Signal Processing and Analysis, 2009, pp. 684–689.

[75] T. Luostarinen, A. Lehmussola, Measuring the accuracy of automatic shoe-print recognition methods, J. Forensic Sci. 59 (2014) 1627–1634.

[76] S. Almaadeed, A. Bouridane, D. Crookes, O. Nibouche, Partial shoeprint retrieval using multiple point-of-interest detectors and SIFT descriptors, Integrated Comput. Aided Eng. 22 (2015) 41–58.

[77] A. Kortylewski, T. Albrecht, T. Vetter, Unsupervised footwear impression analysis and retrieval from crime scene data, in: Asian Conference on Computer Vision, 2014.

[78] A. Kortylewski, T. Vetter, in: S. University of Basel (Ed.), Probabilistic Compositional Active Basis Models for Robust Pattern Recognition, 2016.

[79] A. Kortylewski, FID-300, 31-37-2019, Available: http://fid.cs.unibas.ch, 2017.

[80] B. Kong, D. Ramanan, C. Fowlkes, Cross-domain forensic shoeprint matching, in: Proceedings: British Machine Vision Conference, BMVC), 2017.

[81] S. Alizadeh, C. Kose, Automatic retrieval of shoeprint images using blocked sparse representation, Forensic Sci. Int. 277 (2017) 103–114.

[82] X. Wang, H. Sun, Q. Yu, C. Zhang, Automatic shoeprint retrieval algorithm for real crime scenes, in: Presented at the Asian Conference on Computer Vision, 2014.

[83] Everspry, Available: http://www.footprintmatcher.com/, 1-8-2019.

[84] M. Vagac, M. Povinsky, M. Melichercík, Detection of shoe sole features using DNN, in: Presented at the IEEE 14th International Scientific Conference on Informatics, Poprad, Slovakia, 2017.

[85] A. Kortylewski, E-mail Communication, 2019.

[86] N. Richetelli, M.C. Lee, C.A. Lasky, M.E. Gump, J.A. Speir, Crime scene like shoeprint database, Available: http://4n6chemometrics.com/Downloads/, 2-8-2019.

[87] N. Richetelli, M. Nobel, W.J. Bodziak, J.A. Speir, RAC co-occurrence heatmap, Available: http://www.4n6chemometrics.com/database/, 6-8-2019.

[88] C. Bily, C. Mathias, Ethylene vinyl acetate outsoles and acquired character-istics, J. Forensic Ident. 67 (2017) 549–564.

[89] V.A. Gokool, J.A. John, A. Koertner, Variability of class characteristics observed in die cut outsoles composed of both rubber and textile materials, J. Forensic Ident. 68 (2018) 187–206.

[90] Hobbit, Hobbit Imaging Solutions, E-mail communication, 2019.

[91] L.M. Nienaber, S.L. Cresswell, J.F. Carter, T. Peter, A comparison of plastic cable ties based on physical, chemical and stable isotopic measurements, Sci. Justice 58 (2018) 67–75.

[92] A. Fortini, M. Merlin, C. Soffritti, G.L. Garagnani, Restoration of obliterated numbers on 40NiCrMo4 steel by etching method: metallurgical and statis-tical approaches, J. Forensic Sci. 61 (2016) 160–169.

[93] T. Tsach, N. Finkelstein, N. Volkov, Analysis of a vehicle serial number alteration technique, J. Forensic Ident. 66 (2016) 536–546.

[94] M. Sharma, S. Jha, Analysis of Obliterated Vehicle Identification Numbers: Five Case Studies, Evidence Technology Magazine, 2015, pp. 6–11.

[95] C.L. Schultheis, Serial number restoration on titanium firearms, AFTE J. 48 (2016) 11.

[96] C. Parisien, G. Kolhatkar, F. Crispino, A. Lajeunesse, A. Ruediger, Recon-struction of obliterated characters in polycarbonate through spectral imag-ing, Anal. Chem. 89 (2017) 11648–11652.

[97] N. Finkelstein, N. Volkov, T. Tsach, Improve toolmarks' impressions in soft wax, J. Forensic Sci. 62 (2017) 773–775.

[98] M. Baiker, R. Pieterman, P. Zoon, Toolmark variability and quality depending on the fundamental parameters: angle of attack, toolmark depth and sub-strate material, Forensic Sci. Int. 251 (Jun 2015) 40–49.

[99] M. Ferrucci, T.D. Doiron, R.M. Thompson, J.P. Jones 2nd, A.J. Freeman, J.A. Neiman, Dimensional review of scales for forensic photography, J. Forensic Sci. 61 (2016) 509–519.

[100] G.S. Klees, The categorization of toolmarks and tool types, AFTE J. 49 (2017) 14.

[101] C. Maczlewski, R. Spotts, S. Chumbley, Validation of toolmark comparisons made at different vertical and horizontal angles, J. Forensic Sci. 62 (2017) 612–618.

[102] D.L. Garcia, R. Pieterman, M. Baiker, Influence of the axial rotation angle on tool mark striations, Forensic Sci. Int. 279 (Oct 2017) 203–218.

[103] L.S. Chumbley, M.D. Morris, M.J. Kreiser, C. Fisher, J. Craft, L.J. Genalo, et al., Validation of tool mark comparisons obtained using a quantitative, comparative, statistical algorithm, J. Forensic Sci. 55 (Jul 2010) 953–961.

[104] Alicona, Available: https://www.alicona.com/, 6-8-2019.

[105] S. Chumbley, S. Zhang, MANTIS: portable prototype system for objective characterization of toolmarks, in: Presented at the 47th Annual Training Seminar of the AFTE, New Orleans, LA, USA, 2016.

[106] M. Baiker, R. Pieterman, Towards more objective acquisition & comparison of tool and bullet marks in casework, in: 47th Annual Training Seminar of the AFTE, New Orleans, LA, USA, 2016.

[107] M. Baiker, I. Keereweer, R. Pieterman, E. Vermeij, J. van der Weerd, P. Zoon, Quantitative comparison of striated toolmarks, Forensic Sci. Int. 242 (Sep 2014) 186–199.

[108] M. Baiker, N.D. Petraco, C. Gambino, R. Pieterman, P. Shenkin, P. Zoon, Virtual

and simulated striated toolmarks for forensic applications, Forensic Sci. Int. 261 (Apr 2016) 43–52.

[109] N. Finkelstein, A. Aronson, T. Tsach, Toolmarks made by lathe chuck jaws, Forensic Sci. Int. 275 (2017) 124–127.

[110] J.R. Hadler, M.D. Morris, An improved version of a tool mark comparison algorithm, J. Forensic Sci. 63 (May 2018) 849–855.

[111] B. Bachrach, A. Jain, S. Jung, R.D. Koons, A statistical validation of the in-dividuality and repeatability of striated tool marks: screwdrivers and tongue and groove pliers, J. Forensic Sci. 55 (Mar 1 2010) 348–357.

[112] J.E. Hamby, D.J. Brundage, J.W. Thorpe, The identification of bullets fired from 10 consecutively rifled 9mm ruger pistol barrels: a research project involving 507 participants from 20 countries, AFTE J. 41 (13) (2009).

[113] M. Keglevic, R. Sablatnig, Semi-Automatic retrieval of toolmark images, in: presented at the Oesterreichische Arbeitsgemeinschaft für Mustererkennung Workshop, Hall, Austria, 2018.

[114] Nist, NIST ballistics toolmark database, Available: https://www.nist.gov/programs-projects/nist-ballistics-toolmark-database, 15-8-2019.

[115] M. Keglevic, FORMS-Locks, Available: https://cvl.tuwien.ac.at/forms-locks/, 7-8-2019.

[116] G. Dutton, Considerations for adoption of an evaluative reporting framework for the interpretation of firearms and toolmarks evidence, AFTE J. 49 (13) (2017).

[117] Afte, Theory of identification, range of striae comparison reports and modified glossary definitions - an AFTE criteria for identification committee report, AFTE J. 22 (1990) 275–279.

[118] J.E. Murdock, N.D.K. Petraco, J.I. Thornton, M.T. Neel, T.J. Weller, R.M. Thompson, et al., The development and application of random match probabilities to firearm and toolmark identification, J. Forensic Sci. 62 (May 2017) 619–625.

[119] J. Song, T.V. Vorburger, W. Chu, J. Yen, J.A. Soons, D.B. Ott, et al., Estimating error rates for firearm evidence identifications in forensic science, Forensic Sci. Int. 284 (Mar 2018) 15–32.

[120] E.F. Law, K.B. Morris, C.M. Jelsema, Determining the number of test fires needed to represent the variability present within 9mm Luger firearms, Forensic Sci. Int. 276 (Jul 2017) 126–133.

[121] E.F. Law, K.B. Morris, C.M. Jelsema, Determining the number of test fires needed to represent the variability present within firearms of various cali-bers, Forensic Sci. Int. 290 (Sep 2018) 56–61.

[122] Cadre, Available: https://www.cadreforensics.com/, 08-08-2019.

[123] S. Chumbley, S. Zhang, M. Morris, R. Spotts, C. Maczlewski, Development of a mobile toolmark characterization/comparison system, J. Forensic Sci. 62 (Jan 2017) 83–91.

[124] S.L. Croker, W. Reed, D. Donlon, Comparative cortical bone thickness be-tween the long bones of humans and five common non-human mammal taxa, Forensic Sci. Int. 260 (Mar 2016) 104 e1–104 e17.

[125] A.R. Hale, A.H. Ross, The impact of freezing on bone mineral density: im-plications for forensic research, J. Forensic Sci. 62 (Mar 2017) 399–404.

[126] N. Benson, R. Oliveira dos Santos, K. Griffiths, N. Cole, P. Doble, C. Roux, et al., The development of a stabbing machine for forensic textile damage analysis, Forensic Sci. Int. 273 (2017) 132–139.

[127] N. Benson, R. Oliveria Dos Santos, K. Griffiths, N. Cole, P. Doble, C. Roux, et al., Erratum to "The development of a stabbing machine for forensic textile damage analysis, FSI (2017) 132-139, Forensic Sci. Int. 285 (161) (2018).

[128] J.M. Dittmar, Cut to the bone: the enhancement and analysis of skeletal trauma using scanning electron microscopy, in: D. Errickson, T. Thompson (Eds.), Human Remains: Another Dimension: The Application of Imaging to the Study of Human Remains, 2017, pp. 45–56.

[129] E. Cerutti, L. Spagnoli, N. Araujo, D. Gibelli, D. Mazzarelli, C. Cattaneo, Anal-ysis of cutmarks on bone: can light microscopy be of any help? Am. J. Forensic Med. Pathol. 37 (2016) 248–254.

[130] A. Shamata, T. Thompson, Using structured light three-dimensional surface scanning on living individuals: key considerations and best practice for forensic medicine, J. Foren. Leg. Med. 55 (Apr 2018) 58–64.

[131] M.S. Reynolds, D.M. MacGregor, M.D. Barry, N. Lottering, B. Schmutz, L.J. Wilson, et al., Standardized anthropological measurement of postcranial bones using three-dimensional models in CAD software, Forensic Sci. Int. 278 (Sep 2017) 381–387.

[132] E. Le Garff, V. Mesli, Y. Delannoy, T. Colard, J. De Jonckheere, X. Demondion, et al., The precision of micro-tomography in bone taphonomic experiments and the importance of registration, Forensic Sci. Int. 273 (Apr 2017) 161–167.

[133] S. Clarke, A.M. Christensen, Reflectance Transformation Imaging (RTI) of saw marks on bones, J. Forensic Radiol. Imag. 7 (2016) 33–37.

[134] W. Baier, J.M. Warnett, M. Payne, M.A. Williams, Introducing 3D printed models as demonstrative evidence at criminal trials, J. Forensic Sci. 63 (2018) 1298–1302.

[135] W. Baier, D.G. Norman, J.M. Warnett, M. Payne, N.P. Harrison, N.C.A. Hunt, et al., Novel application of three-dimensional technologies in a case of dis-memberment, Forensic Sci. Int. 270 (2017) 139–145.

[136] Bruker, Skyscan mircotomography, Available: https://www.bruker.com/products/microtomography.html, 16-8-2019.

[137] K. Domenick, A.M. Christensen, Experimental wood chipper trauma on bone, J. Forensic Ident. 66 (2016) 547–559.

[138] L. Nogueira, G. Quatrehomme, M.F. Bertrand, C. Rallon, P. du Jardin, et al., Comparison of macroscopic and microscopic (stereomicroscopy and

scanning electron microscopy) features of bone lesions due to hatchet hacking trauma, Int. J. Leg. Med. 131 (Mar 2017) 465–472.

[139] L. Komo, M. Grassberger, Experimental sharp force injuries to ribs: multi-modal morphological and geometric morphometric analyses using micro-CT, macro photography and SEM, Forensic Sci. Int. 288 (Jul 2018) 189–200.

[140] C. Humphrey, J. Kumaratilake, M. Henneberg, A stab in the dark: design and construction of a novel device for conducting incised knife trauma investigations and its initial test, Forensic Sci. Int. 262 (May 2016) 276–281.

[141] M.A. Maté-González, J. Yravedra, D.M. Martín-Perea, J. Palomeque-González, M. San-Juan-Blazquez, V. Estaca-Gómez, et al., Flint and quartzite: distinguishing raw material through bone cut marks, Archaeometry 60 (2018) 437–452.

[142] M.A. Maté-González, J.F. Palomeque-González, J. Yravedra, D. González-Aguilera, M. Domínguez-Rodrigo, Micro-photogrammetric and morphometric differentiation of cut marks on bones using metal knives, quartzite, and flint flakes, Archaeol. Anthropol. Sci. 10 (2018) 805–816.

[143] L.A. Courtenay, J. Yravedra, M.A. Mate-González, J. Aramendi, D. González-Aguilera, 3D analysis of cut marks using a new geometric morphometric methodological approach, Archaeolo. Anthropol. Sci. (2017).

[144] I. Macoveciuc, N. Marquez-Grant, I. Horsfall, P. Zioupos, Sharp and blunt force trauma concealment by thermal alteration in homicides: an in-vitro experiment for methodology and protocol development in forensic anthropological analysis of burnt bones, Forensic Sci. Int. 275 (Jun 2017) 260–271.

[145] L. Waltenberger, H. Schutkowski, Effects of heat on cut mark characteristics, Forensic Sci. Int. 271 (2017) 49–58.

[146] V. Alunni, L. Nogueira, G. Quatrehomme, Macroscopic and stereomicroscopic comparison of hacking trauma of bones before and after carbonization, Int. J. Leg. Med. 132 (2018) 643–648.

[147] S.A. Stanley, S.V. Hainsworth, G.N. Rutty, How taphonomic alteration affects the detection and imaging of striations in stab wounds, Int. J. Leg. Med. 132 (Mar 2018) 463–475.

[148] A. Bornik, M. Urschler, D. Schmalstieg, H. Bischof, A. Krauskopf, T. Schwark, et al., Integrated computer-aided forensic case analysis, presentation, and documentation based on multimodal 3D data, Forensic Sci. Int. 287 (2018) 12–24.

[149] P. Urbanova, A.H. Ross, M. Jurda, I. Splichalova, The virtual approach to the assessment of skeletal injuries in human skeletal remains of forensic importance, J. Forensic Leg. Med. 49 (Jul 2017) 59–75.

[150] J. Park, H. Son, Weapon use in Korean homicide: differences between homicides involving sharp and blunt instruments, J. Forensic Sci. 63 (Jul 2018) 1134–1137.

[151] J.F. Palomeque-González, M.A. Maté-González, J. Yravedra, M. San Juan-Blazquez, E.G. Vargas, D.M. Martín-Perea, et al., Pandora: a new morphometric and statistical software for analysing and distinguishing cutmarks on bones, J. Archaeol. Sci. Rep. 13 (2017).

[152] M.R. Mahfouz, A. Mustafa, E.E. Abdel Fatah, N.P. Herrmann, N.R. Langley, Computerized reconstruction of fragmentary skeletal remains, Forensic Sci. Int. 275 (Jun 2017) 212–223.

[153] S. Valoriani, C. Eliopoulos, M. Borrini, Sharp force trauma death in a young individual from medieval gloucester, Am. J. Forensic Med. Pathol. 38 (Jun 2017) 111–114.

[154] G. Quatrehomme, M.D. Piercecchi-Marti, L. Buchet, V. Alunni, Bone beveling caused by blunt trauma: a case report, Int. J. Leg. Med. 130 (May 2016) 771–775.