JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2010, Vol. 55, No. 2
doi: 10.1111/j.1556-4029.2009.01221.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Benjamin Bachrach,[1] Ph.D.; Anurag Jain,[1] M.S.; Sung Jung,[1] M.S.; and Robert D. Koons,[2] Ph.D.

A Statistical Validation of the Individuality and Repeatability of Striated Tool Marks: Screwdrivers and Tongue and Groove Pliers*

ABSTRACT: Tool mark identification relies on the premise that microscopic imperfections on a tool's working surface are sufficiently unique and faithfully transferred to enable a one-to-one association between a tool and the tool marks it creates. This paper presents a study undertaken to assess the validity of this premise. As part of this study sets of striated tool marks were created under different conditions and on different media. The topography of these tool marks was acquired and the degree of similarity between them was quantified using well-defined metrics. An analysis of the resulting matching and nonmatching similarity distributions shows nearly error-free identification under most conditions. These results provide substantial support for the validity of the premise of tool mark identification. Because the approach taken in this study relies on a quantifiable similarity metric, the results have greater repeatability and objectivity than those obtained using less precise measures of similarity.

KEYWORDS: forensic science, tool mark identification, 3D imaging, automated comparison of microscopic tool mark evidence, striations, screwdrivers, tongue and groove pliers, statistical methodology

The ability to perform tool mark to tool mark comparisons based on microscopic features observed on the tool mark's surface is at the core of tool mark identification. Supreme Court decisions such as *Daubert versus Merrill Dow* (1) and *Kumho Tire versus Carmichael* (2) are making it increasingly necessary to further formalize scientific evidence presented in court. Furthermore, the development of DNA identification techniques and the level of accuracy achievable in the estimation of the associated error rates have raised the expectations for the quantitative precision that may be achieved in forensic analysis. Quantitative evidence regarding the validity of the basic premise of tool mark comparison would provide additional support for the admissibility of tool mark evidence. The Federal Bureau of Investigation (FBI) and Intelligent Automation, Inc. (IAI) have undertaken an extensive study to verify the premise that the microscopic features transferred from a tool's working surface to the marks created by it are sufficiently unique and repeatable to enable the association of a tool with its marks. This paper reports the results of this study for the case of striated tool marks (a paper reporting the results for impressed tool marks is in preparation). In particular, we consider two types of tools: screwdrivers and tongue and groove pliers. In addition to considering the comparison of striated tool marks created under the same conditions, we also evaluated the effect of the media onto which the tool marks are created. In the case of screwdrivers, the effect of the variation of angle of attack in the creation of striated tool marks was also evaluated.

An important element of this study was the use of topographical (3D) data for the characterization of tool marks. The concept of using a 3D characterization of a surface for identification purposes goes as far back as 1958, when Davis (3) proposed the idea of the "Striagraph" for ballistic identification. The application of 3D methodologies to obtain characteristic information about striated marks on bullets has also been reported by DeKinder (4,5). Geradts (6) has presented a system capable of performing, in an automated way, comparisons between 3D topographical measurements of striated tool marks. Bachrach (7) has described an automated comparison system that uses 3D information of a bullet's surface to perform automated comparisons. More recently, Banno (8) has reported on the 3D visualization and comparison of features on fired bullets by using 3D surface topography data. The principles of tool mark identification can be found in Miller (9). An often cited study on the criteria for identification for firearm and tool mark identification was published by Biasotti and Murdoch (10). Another significant effort that examines the theory of identification as it pertains to tool marks and the criteria for their identification has been reported by Miller (11). An exhaustive review of the literature pertaining to the identification criteria for firearms and tool mark identification has been more recently carried out by Nichols in 1997 (12) and 2003 (13). This study builds upon and extends the results of the previous studies by providing consistent quantitative measures in 3D of tool mark similarity.

As part of the study reported in this paper, a confocal microscope was used to acquire topographical data of tool mark samples. A significant number of striated tool mark samples were created under controlled conditions on a variety of media. Algorithms were

[1]Intelligent Automation, Inc., 15400 Calhoun Drive, Suite 400, Rockville, MD 20855.

[2]Counterterrorism and Forensic Science Research Unit, FBI Laboratory, FBI Academy, Quantico, VA 22135.

*Presented in part at the Association of Firearms and Tool Marks Examiners (AFTE) conferences in 2005 and 2006.

Received 24 April 2008; and in revised form 5 Jan. 2009; accepted 11 Jan. 2009.

developed and implemented to generate the necessary tool mark signatures and well-defined metrics were used to objectively evaluate the degree of similarity between known matching and nonmatching tool mark pairs. The distributions of the degree of similarity values obtained from the comparison of known matching and nonmatching pairs of tool marks were analyzed using established statistical techniques. While it is not possible to prove uniqueness statistically (14), the results of this study provide support for the concept that tool marks contain measurable features that exhibit a high degree of individuality.

Methods

The main goal of the study under consideration was to assess the degree of individuality and repeatability of the features transferred from the working surface of a tool to the tool marks created by it in an objective and repeatable manner. The approach selected to achieve this goal was by development of an automated tool mark comparison system. An automated comparison system provides both objective and repeatable results, since it applies the same algorithms and similarity metric to each tool mark pair under comparison. Moreover, such a system is capable of comparing large numbers of tool marks in a short period of time.

In addition to the development of an automated comparison system, a rigid methodology was formulated and followed for the creation of sample tool marks for the following three scenarios of interest:

Scenario (a) Comparison of tool marks when both the medium and the conditions under which different tool marks are created are the same.

Scenario (b) Comparison of tool marks when the conditions under which tool marks are created are the same, but the media are different.

Scenario (c) Comparison of tool marks when the medium onto which different tool marks are created is the same, but the conditions are different (this scenario was considered for the variations in the screwdriver's angle of attack only).

By analyzing the statistical distributions of similarity values resulting from the comparison of known matching and nonmatching pairs of tool marks, it is possible to assess the degree to which tool marks created by the same tool are repeatable and distinguishable from tool marks created by other tools. In this section, we provide an overview of the automated tool mark comparison system, the associated similarity metric, and the methodology followed for the creation of the tool mark samples used in this study.

3D-Based Automated Tool Mark Comparison System

The implementation of an automated comparison system requires two main components: (i) data acquisition hardware and (ii) data analysis software. The data acquisition hardware is responsible for capturing the physical characteristics of the specimen being analyzed. The data analysis software is responsible for the storage, management, processing, and comparison of the data acquired by the data acquisition hardware. In the following sub-sections, we describe these two components.

Data Acquisition Hardware—From the inception of this study, it was decided that topographical images (often referred to as 3D data) as opposed to photographical images (referred to as 2D data) would be used to characterize the tool marks under comparison. Both topographical imaging and photographical imaging are processes which translate physical properties of the specimen into an array of numerical values. In the case of photographical images, these values correspond to the intensity of the light reflected by the specimen; in the case of topographical images, they correspond to the depth of the specimen's surface with respect to a reference plane. The use of topographical data has a number of important advantages over photographical data. Figure 1 shows an example of a topographical image on the left and a photographical image on the right corresponding to a striated tool mark created by a pair of tongue and groove pliers. Figure 1 demonstrates the vulnerability of photographical images to variations in the reflectivity of the medium onto which the tool mark is created. Other parameters which can influence photographical images are illumination conditions (intensity, angle, type of illumination, etc.), and camera angle. Topographical imaging is virtually immune to these variables, and therefore, provides a significantly more robust process to capture the relevant features of a specimen. In terms of flexibility, topographical data has the significant advantage of allowing for dimensionally-preserving geometric transformations of the data. For example, topographical data can be mathematically "rotated" without distortion. This property plays an important role in the processing of the data. Figure 2 provides a visual representation of this characteristic. The images seen in Fig. 2 correspond to the same data, but from a different point of view. This is not always possible for photographical data (at least not accurately, unless multiple images are taken). Also, as seen in Fig. 2, topographical data allows for the identification and isolation of "waviness" (usually due to class characteristics) and "roughness" (often associated with individual characteristics). A more extensive discussion of the advantages of topographical data as opposed to photographical data can be found in (9).

There are a variety of technologies for the acquisition of topographical data that have been utilized in commercially available systems. For the purposes of this study, the candidate choices were constrained by the requirement that only noncontacting acquisition techniques be considered. The rationale for this requirement was that a contact-based system would pose the risk of damaging the tool mark under consideration or altering the data if the same tool mark had to be acquired multiple times. At the start of this study, we considered several commercially available 3D imaging systems. These instruments utilize different technologies as indicated in Table 1. Among these, only the MicroSurf white light confocal microscope manufactured by NanoFocus AG (NanoFocus, Inc., Glen Allen, VA) and the NT series of white light interferometers manufactured by Veeco Instruments, Inc. (Chadds Ford, PA) provided the performance required for this project. Both these

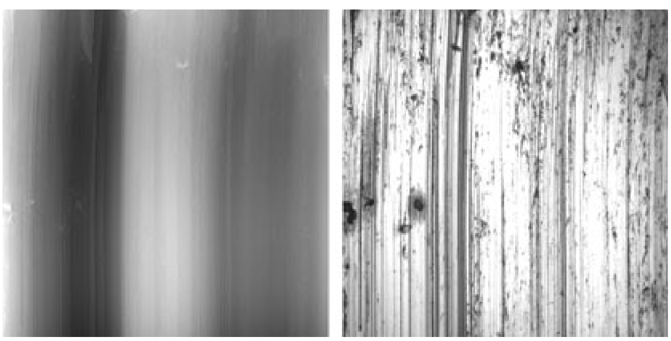

FIG. 1—*Example of topographical (left) and photographical (right) data for a striated tool mark.*

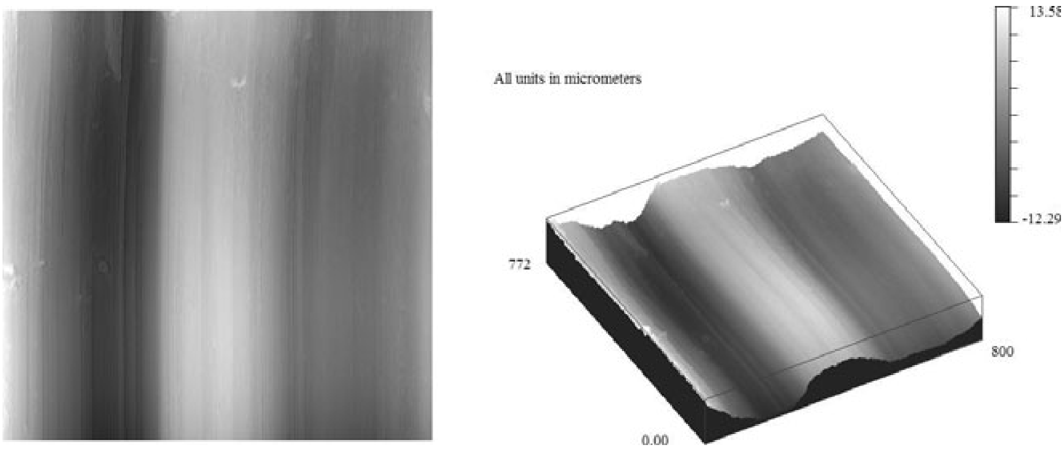

FIG. 2—*Geometric transformation of striated tool mark topographical data.*

TABLE 1—*Data acquisition systems evaluated.*

Manufacturer	Model	Technology	Data	Evaluation
LMI Technologies	LTS series	Triangulation	Single point	Inadequate lateral resolution
STIL	CHR	Chromatic Aberration	Single point	Inadequate parameters
NanoFocus	MicroScan	Dynamic Focusing	Single point	Limited range
Optimet	ConProbe/ConoLine	Conoscopic Holoraphy	Point/line	Inadequate lateral resolution
Veeco	NT series	WL Interferometry	Patch	Excellent performance
NanoFocus	MicroSurf	WL confocal microscope	Patch	Excellent performance

systems have exceptional lateral and depth resolution, and have the capability to acquire rectangular "patches" of points as opposed to single points or lines of data. The relative performance of the white light confocal microscope against the white light interferometer sensor is still a subject of debate within the metrology community. Nonetheless, there is evidence to suggest that the white light confocal microscope can handle steeper slopes than its white light interferometer counterpart. On the other hand, the white light interferometer sensor may be able to achieve better depth resolution than the white light confocal microscope for relatively flat surfaces. Given that the lateral and depth resolution of the confocal microscope was more than sufficient for the current application, that the slopes associated with tool mark topography are often significant, and that the cost of the confocal microscope was less than the white light interferometer, the NanoFocus MicroSurf white light confocal sensor was selected for our particular application. The operating conditions used in this study are shown in Table 2. The NanoFocus MicroSurf white light confocal microscope proved to be accurate, robust to vibration, and easy to use.

Data Analysis Software—The automated comparison of data requires two main software components: the signature generation component, and the correlation component. The main purpose of the signature generation component is to isolate those features that are characteristic of the specimen under consideration (individual characteristics) from those that are common to all specimens of the same type (class characteristics). Consider, for example, the case of a group of screwdrivers of the same make and model. As these screwdrivers are manufactured to the same specifications, the overall geometric shape of the tool marks created by them is very similar. On the other hand, as no two manufactured parts are ever identical, there are microscopic variations specific to each screwdriver blade. The key premise to be validated in this study is whether the process through which the blade features are transferred to a tool mark captures these specific features (most likely together with class characteristics features) in a repeatable manner. The challenge associated with the development of an effective automated tool mark comparison system is, therefore, to separate class characteristics from individual characteristics, and to treat them in the appropriate manner.

TABLE 2—*Main performance parameters of NanoFocus Microsurf.*

	Objective Lens	
	20×L	50×L
Numerical aperture	0.4	0.6
Single patch fov (μm)	800 × 800	320 × 320
Lateral resolution (μm)	1.5	0.6
Vertical resolution (nm)	20	10
Standoff (mm)	12	10.6

Signature Generation Component—Figure 3 shows the main algorithmic modules of the signature generation component. These modules are:

Preprocessing: The unprocessed data obtained from the acquisition hardware is referred to as "raw data." Raw data often includes inaccurate or questionable data points. We refer to such points as *unreliable* data points. The preprocessing module is responsible for the identification and preliminary handling of unreliable data points. Two types of unreliable data points are considered: drop-offs and outliers.

Drop-off points are points corresponding to regions of the specimen where the acquisition system has been unable to acquire data. In the case of optical systems, this limitation is generally because of insufficient light being collected by the optical system due to either low reflectivity or a steep slope on the specimen's surface. Such points are usually identified by the acquisition system as

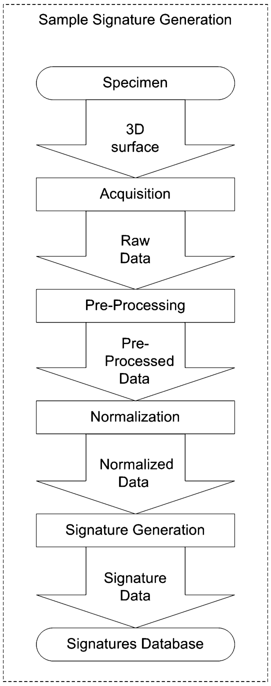

FIG. 3—*Signature generation steps.*

having insufficient light reflection intensity. As these points are identified by the optical system, there is no need to develop algorithms to recognize them. Nevertheless, the preprocessing software developed for this application keeps track of drop-off points for later data handling.

"Outliers" are those data points that are inaccurately measured by the imaging system, but which are not recognized as such via the intensity of reflection information (in other words, the intensity of reflection associated with such points is within the nominal range). For this reason, these points are much more difficult to identify, and specific algorithms had to be developed for this purpose. Two approaches were used to identify such outliers. The first approach was based on the estimation of the slope between a point and its neighbors. Any point for which the local slope is above a preestablished threshold is identified as an outlier. The second approach was based on the statistical distribution of the data in the vicinity of the point under consideration. Any point which deviates beyond a predetermined number of standard deviations with respect to the local mean is considered an outlier. Once all unreliable points are identified, they are recorded in a "mask" which is then used for the remainder of the signature generation process. Figure 4 shows a raw topographical image of the same data as in Fig. 2 and its corresponding mask, where unreliable points are shown as dark points.

Normalization: The normalization module is responsible for compensating for the variations in the topographical images that result from inconsistencies during the acquisition process. A comprehensive presentation of the normalization process is beyond the scope of this paper. However, we consider a simple illustrative example. Let us assume that a given tool mark sample is acquired twice, but in each case, the tool mark surface is oriented differently. Figure 5 represents this situation, where a single cross section of data is considered for ease of presentation. Data 1 represents the data acquired the first time, while Data 2 represents the data acquired the second time. While these two sets of data correspond to the same tool mark (and should therefore be identical if one ignores instrument noise), they appear different due to a different relative orientation between the sample surface and the microscope during the acquisition process. If left uncorrected, these two data sets may be erroneously judged to be dissimilar by the correlation algorithms. The purpose of the normalization process is to bring these two data sets to a "level playing field." In the case of this example, the first step in the normalization process is to identify a baseline or a reference horizon. Let us assume that an appropriate baseline for the type of data under consideration is a linear function (in fact, the baseline could be a shape corresponding to a class characteristic). Once the baseline is identified, the purpose of the normalization is to apply a transformation to compensate for the fact that the tool marks under consideration were not acquired in a uniform manner. For the example under consideration, the simplest such transformation would be the rotation of the data.

Based on this simple example, we can articulate the purpose of the normalization process as it applies to any tool mark data of interest. The normalization process consists of the application of a geometric transformation to the preprocessed data in an effort to compensate for any inconsistencies resulting from the acquisition process. In other words, the goal of the normalization process is to ensure that the data is represented in a consistent way regardless of variations which may have taken place during the acquisition process.

It is important to note that the normalization process would be significantly more challenging—if not impossible—if the data under consideration were photographical data. While processes similar to normalization can be developed for photographical data, it would be significantly more difficult to achieve the same level of accuracy as that achievable with topographical data. Also, it is worth noting that in order to perform the normalization process accurately, it is necessary to have knowledge of which points can be considered reliable. For the example under consideration, only reliable points are used in the estimation of the baseline. Otherwise,

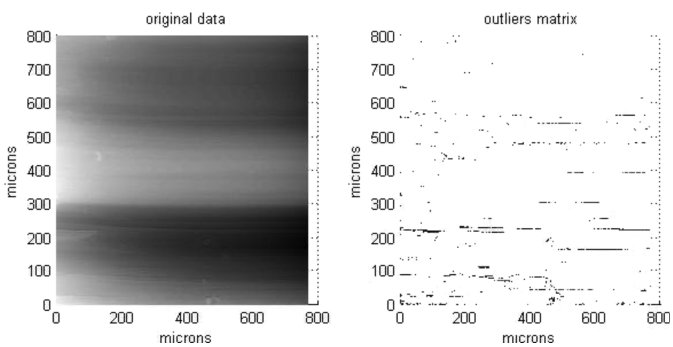

FIG. 4—*Tongue and groove pliers tool mark (left) and corresponding mask (right). Unreliable points are indicated as dark in the mask.*

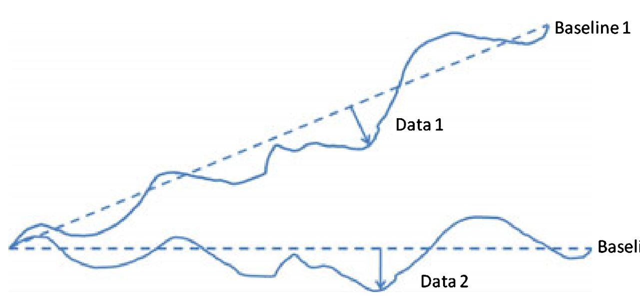

FIG. 5—*Conceptual example of normalization process in the case of different orientations.*

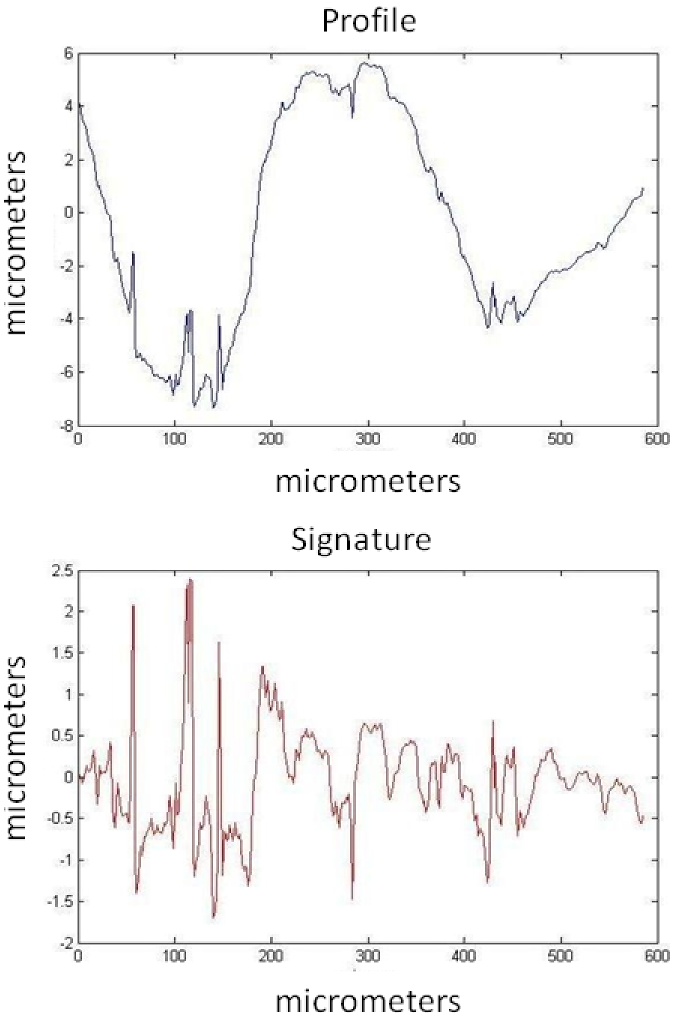

FIG. 6—*Profile (top) and signature (bottom) of the tool mark.*

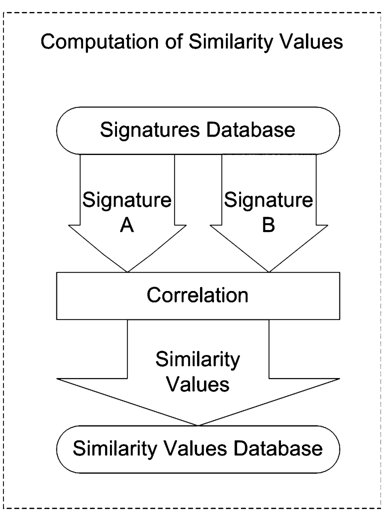

FIG. 7—*Signature correlation steps.*

the result of the normalization process is not consistent between different tool marks. For this reason, identification of unreliable points precedes the normalization process.

Signature Generation: The signature generation module is responsible for emphasizing those features which are specific to the tool mark under consideration (individual characteristics), while minimizing the features which may be common to all tool marks of the same type (class characteristics). For the tool marks under consideration, this process consists of two steps. The first step involves the conversion of the topographical tool mark data (in the form of a 2D array) into a single data vector that corresponds to a cross section of the tool mark. The second step involves using a Gaussian band pass filter to eliminate the low frequency component corresponding to the class characteristics of the tool mark. Figure 6 shows an example of the signature generation process applied to normalized data profile. Notice that as a result of the signature generation process, all low frequency components are discarded, while the high frequency components are left intact.

Signature Correlation Component—The signatures generated by the signature generation module are stored in a database, and are accessible to the signature correlation component (see Fig. 7). Given a pair of signatures, the purpose of the signature correlation component is to evaluate a metric indicative of their degree of similarity. We refer to the value achieved by such metric as a *similarity measure*. Let us denote a pair of signatures corresponding to two different striated tools mark by:

$$z_i(n), z_j(n);\ n = 1, \ldots, N. \tag{1}$$

where the mean value of z_k (denoted by $\bar{z}_k$) is equal to zero for both $k = i,\ j$. We define the *relative distance* between two signatures of the same number of points as:

$$\mathrm{rdist}_{i,j} = 1 - \frac{\sum\limits_{n=1,\ldots,N} (z_i(n) - z_j(n))^2}{\sum\limits_{n=1,\ldots,N} (z_i(n) + z_j(n))^2} \tag{2}$$

The relative distance metric is a time-domain similarity metric (as opposed to frequency-domain, wavelet-domain, etc.), and it offers advantages in terms of being well suited to handle signatures of different lengths and signatures with missing data points (unreliable data points). The relative distance defined in (2) is upper-bounded by 1, where a relative distance of 1 indicates that the two signatures satisfy $z_i(n) \equiv z_j(n)\ \forall n = 1, \ldots, N$. In other words, the two signatures are identical. On the other hand, a similarity metric value close to zero indicates that there is only a subtle (or insignificant) relationship between the two signatures.

As discussed with reference to the normalization process, it is reasonable to assume that there will be differences in the area imaged for each tool mark. For this reason, while computing the similarity measure between two signatures, it is necessary to allow a pre-established degree of relative lateral displacement or "shift" between them. However, as one signature is "shifted" with respect to its counterpart, the number of points of comparison decreases. For this reason, a slight modification of Equation (2) is necessary. Let us consider the case where signature *j* is shifted to the right by Δ points with respect to signature *i*. In such a case, the number of overlapping points between the two signatures decreases to $N - \Delta$, and the region of overlap between the two signatures becomes:

$$z_i(n + \Delta),\ z_j(n);\ n = 1, \ldots, N - \Delta \tag{3}$$

The relative distance between the two shifted signatures is computed by:

$$\mathrm{rdist}_{i,j}(\Delta) = 1 - \frac{\sum\limits_{n=1,\ldots,N-\Delta} (z_i(n+\Delta) - z_j(n))^2}{\sum\limits_{n=1,\ldots,N-\Delta} (z_i(n+\Delta) + z_j(n))^2} \tag{4}$$

A similar computation can be made in the case of a left shift, which is denoted by a negative value of Δ. Based on this definition, the similarity measure between two signatures is defined by:

$$s_{i,j}(\Delta_{\max}) = \max_{|\Delta|<\Delta_{\max}} \mathrm{rdist}_{i,j}(\Delta) \tag{5}$$

The maximum relative shift $\Delta_{\max}$ in Equation (5) is selected so as to reflect the inconsistencies inherent to the acquisition process. The properties of the similarity metric defined by Equation (5) are inherited from the properties of Equation (2).

Tool Selection and Sample Tool Marks Preparation

While both striated and impressed tool marks were considered as part of this study, this paper only discusses striated tool marks (the results obtained for impressed tool marks are in preparation). In particular, we consider two types of tools: screwdrivers and tongue and groove pliers. The screwdrivers used in this study were Craftsmen Professional screwdrivers (model # 47441) while the tongue and groove pliers used in this study were Cooper Tools Crescent pliers (model # R210C).

Tool Marks Sample Preparation—For each of the tool types under consideration, 10 sample tools of the same manufacturer and model number were purchased. For each sample tool, 10 tool mark samples were created under the same conditions for each medium of interest. We refer to each such group of 100 tool marks created on the same medium and under the same conditions as a *set*. Table 3 summarizes the different sets of tool mark samples created as part of this study. As shown in Table 3, seven different sets of tool marks, totaling 700 individual specimens were used for this study.

While creating the sample tool marks, care was taken to minimize the likelihood of damaging the working surface of the tool. For this reason, the first set of tool mark samples was created on lead for both tool types. Once the repeatability and individuality of these tool marks was evaluated, we proceeded to harder media. In the case of screwdrivers, sample tool marks at three angles of attack (15°, 30°, and 45°) were created in lead, and an additional set was created at an angle of attack of 30° on aluminum. The cross-sectional width of these striated tool marks was 5.0 mm. In the case of the tongue and groove pliers, the creation of tool mark samples in lead rope was followed by the creation of samples on brass and galvanized steel pipe. The cross-sectional width of these tool marks was *c.* 7.4 mm.

TABLE 3—*Tool mark sets.*

Set	Tool Type	Conditions	Media
SD01	Screwdriver	45 deg	Lead
SD02	Screwdriver	30 deg	Lead
SD03	Screwdriver	15 deg	Lead
SD04	Screwdriver	30 deg	Aluminum
TG01	Tongue and groove pliers		Brass
TG02	Tongue and groove pliers		Galvanized steel
TG03	Tongue and groove pliers		Lead

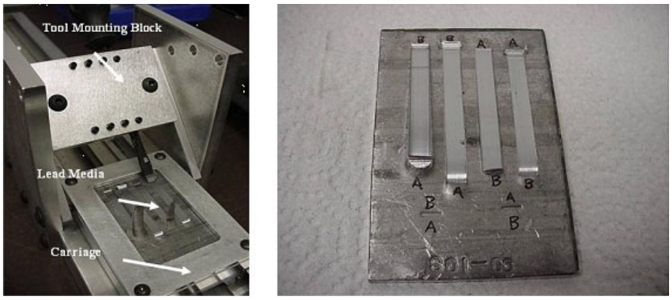

FIG. 8—*Device to create screwdriver striated tool marks (left) and an example of such a tool mark (right).*

Creation of Tool Marks from Screwdrivers—In the case of screwdrivers, we found that it was very difficult to create uniform striations manually. For this reason, a device to assist in the creation of these tool marks was designed and built (see left side of Fig. 8). The main components of this device are a carriage and a tool-mounting block. The carriage was designed so that a 5.08×7.62 cm piece of metal sheet could be rigidly affixed to it. The tool-mounting block was designed such that a screwdriver could be rigidly mounted at a variety of predetermined angles with respect to the medium affixed to the carriage. The carriage could then be translated using a lead screw, allowing for the displacement of the medium with respect to the blade of the screwdriver. Furthermore, the lead screw was motorized using a conventional electric drill, producing a constant speed displacement of the sample medium with respect to the screwdriver blade. This device enabled the creation of very clean and uniform tool marks. With the assistance of this device, the sample tool mark sets were created on both lead and aluminum sheets in an identical fashion. An example of the types of sample tool marks created with this device can be seen on the right side of Fig. 8.

Prior to making the test samples, each screwdriver was labeled with an identifying number between 01 and 10. Also, both sides of the screwdriver blade were labeled using the conventional A–B labeling (see Fig. 9). The screwdriver tool mark samples were created on lead and aluminum. $30.48 \times 30.48 \times 0.32$ cm sheets were cut into $5.08 \times 7.62 \times 0.32$ cm rectangles using a metal shear. Prior to labeling the medium, each sample was flattened by impacting it with a dead blow hammer. The medium was then labeled "SXX-YY" along the bottom edge, where "XX" refers to the screwdriver's label (01 through 10) and "YY" refers to the tool mark sample number (01 through 40). The sample was placed in the aluminum frame and held in place by securing the upper frame plate with four screws. It was then positioned so that the screwdriver, when fixtured, contacted the upper region of the sample. The screwdriver was placed into the tool-mounting block with the

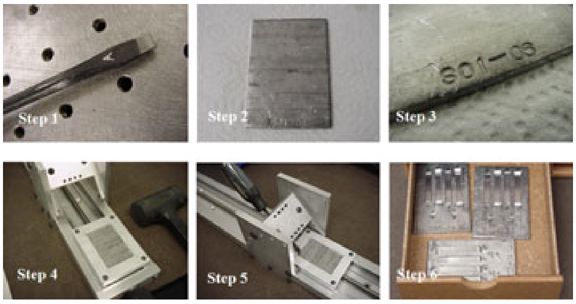

FIG. 9—*Steps involved in the creation of tool marks from screwdrivers.*

FIG. 10—*Steps involved in the creation of tool marks from tongue and groove pliers.*

blade labeled "A" facing upwards. The end of the screwdriver was then gently tapped until the tip just impacted the surface of the medium sample. It was then secured with two set screws. By operating the drill, the sample was moved toward the rear of the device for about an inch, creating a "push" tool mark on the sheet. Push and pull marks were made with each side of the blade and two impression marks were made using the tip of the screwdriver as seen in the right side of Fig. 8. For this study, only one push mark from each tool impression sample was used. After the creation of the tool mark, the samples were stored in a container. Special care was taken at every step of the process to avoid contact with skin or moisture so as to minimize the oxidation rate.

To study the effect of variation of screwdriver angle of attack on striated tool marks, a total of 300 striated tool marks were created on lead at screwdriver angle of attack of 15°, 30°, and 45° to the lead sheet (see Table 3). To study the effect of media, 100 tool mark samples were created at a screwdriver angle of attack of 30° on aluminum sheet using the same procedure. Topographical images of all of the prepared tool marks were acquired with a lateral resolution of 1.52 μm. For processing purposes, these data sets were decimated to a lateral resolution of 4.56 μm.

Creation of Tool Marks from Tongue and Groove Pliers—For the creation of striated tool marks from tongue and groove pliers, it was decided that the tool mark of interest would correspond to the striated tool mark created by a single predetermined tooth on the jaw of the tongue and groove pliers. Each tool mark was created by the rotation of the tool as the jaws firmly grip the curved surface of a cylindrical sample of the medium in a uniform fashion. A description of the steps involved in the creation of three sets of 100 tool marks on brass pipes, steel pipes, and lead rope follows (see Fig. 10).

Prior to making the tool marks, each pair of tongue and groove pliers was labeled with an identifying number between 01 and 10. Both sides of both jaws of each pair of the tongue and groove pliers were labeled using the traditional a–b and A–B labeling convention. The tooth of interest on the jaw of the tongue and groove pliers, which on contact with the pipe/rope would create the striated tool mark, was identified and marked. The indication of the tooth of interest was made with a punch. Appropriate care was taken to ensure that this process did not physically alter the tooth in any way. The media to be used for the creation of the striated tool marks from tongue and groove pliers were brass pipes, galvanized steel pipes, and lead rope. The pipes/ropes were selected to have a 1.27 cm internal diameter, which facilitated the contact of the same tooth of the tool's jaw for all media of interest. The lead rope was cut into equal pieces of 5.08 cm, while the brass and galvanized steel pipes were cut into equal pieces of 25.40 cm using a band saw. The pipe/rope was rigidly mounted on to a vise affixed to a work bench and was tightly clamped to eliminate movement during the creation of the tool mark. The jaw of the tongue and groove pliers was brought in contact with the pipe/rope to identify the region on the pipe/rope where the tooth of interest would make contact. Once the tooth was aligned satisfactorily over the surface of the medium, the region of contact on the pipe/rope was identified by marking it with a line drawn with a soft tip marker. The purpose of drawing this line was to indicate the position on the pipe/rope where the tooth of interest would be initially placed to create the tool mark. The tongue and groove pliers were brought in contact with the surface of the pipe/rope so that the tooth of interest was in alignment with the line drawn in the previous step. While holding the tongue and groove pliers firmly with both hands, it was slowly rotated around the surface of the pipe/rope such that only the tooth of interest was in direct contact with the pipe/rope over the region of interest. This rotational movement resulted in the creation of a tool mark consisting of striations imparted from the movement of the tooth of interest over the pipe's/rope's surface. Once a sufficiently long striated tool mark was created (about half to 1 cm in length), the tongue and groove pliers were carefully withdrawn from the pipe's/rope's surface. After the tool mark had been created, a soft brush was used to clean the jaws of the tongue and groove pliers before the creation of the subsequent tool mark. Each tool mark was labeled as "TSXX-YY" where "XX" referred to the tongue and groove pliers (01 through 10) and "YY" referred to the tool mark sample number (01 through 30). The brass and galvanized steel pipes were then cut into two pieces by a band saw such that each half had five tool marks and then stored in a container.

Three-dimensional images of each of the prepared tool marks were acquired with a lateral resolution of 1.52 μm. For processing purposes, these data sets were decimated to a lateral resolution of 4.56 μm.

Statistics

Tables 4 and 5 summarize the sets of data which were compared, and the number of matching (i.e., same tool) and nonmatching (i.e., different tool) comparisons performed for screwdrivers and tongue and groove pliers tool marks, respectively. Each set of comparisons shown in Tables 4 and 5 corresponds to one of the three scenarios discussed in the Methods section. As an example,

TABLE 4—*Numbers of comparisons of matching/nonmatching pairs of screwdriver tool marks.*

Matching/ Nonmatching Pairs	SD01	SD02	SD03	SD04
SD01	450/4500	1000/9000	1000/9000	X
SD02		450/4500	1000/9000	1000/9000
SD03			450/4500	X
SD04				450/4500

TABLE 5—*Numbers of comparisons of matching/nonmatching pairs of tongue and groove pliers tool marks.*

Matching/ Nonmatching Pairs	TG01	TG02	TG03
TG01	450/4500	1000/9000	1000/9000
TG02		450/4500	1000/9000
TG03			450/4500

consider the comparison of set SD01 against itself. Such comparison resulted in 450 matching similarity measure values and 4500 nonmatching similarity measure values of tool marks created onto the same medium, under the same conditions. The comparison of set SD01 against itself corresponds to Scenario (a). Set SD01 was also compared against set SD02, resulting in 1000 matching similarity measure values and 9000 nonmatching similarity measure values of tool marks created onto the same medium, under different angle of attack. This set of comparison corresponds to Scenario (c). By analyzing the differences between the distributions obtained from the comparisons of SD01 versus SD01, SD02 versus SD02, and SD01 versus SD02 it is possible to isolate and evaluate the effect of screwdriver angle of attack on the created tool mark. In a similar manner, the effect of different media was analyzed (Scenario [b]).

The purpose of performing the large number of correlations discussed above is to empirically estimate the distribution of matching and nonmatching similarity measure values for the scenarios of interest. An analysis of these distributions allows us to conclude whether the tool marks created by these tools under the conditions of interest display individualizing and repeatable features. If these distributions are distinct at a given level of significance, we can conclude that the individuality and repeatability criteria have been verified, or at least have not been disproven to that level of significance. Figure 11 shows the empirically estimated matching and nonmatching similarity measure distributions for screwdriver tool marks created on lead at a 30° angle of attack (set SD02). As seen in Fig. 11, the distributions of matching and nonmatching similarity measure values are quite distinct. The nonmatching distribution has a mean of .33 with a standard deviation of .07, while the matching distribution has a mean of .92 with a standard deviation of .07. Clearly, these empirical distributions indicate a high degree of similarity among marks from the same tool (repeatability) and differences between marks from different tools (individuality). The same behavior can be observed in the inter-comparison of sets SD02, SD03, and SD04 corresponding to the comparison of screwdriver tool marks, and TG01, TG02, and TG03 corresponding to the comparison of tongue and groove pliers tool marks. In all these cases, either no or minimal overlap can be seen between the distributions. As an example of these results in the case of tongue and groove pliers, Fig. 12 shows the matching and nonmatching similarity distributions for tongue and groove tool marks created on steel pipes (set TG02).

To summarize the behavior of each of the sets of comparisons shown in Tables 3 and 4, it is convenient to select a metric which quantifies the degree of overlap between the matching and nonmatching distributions. Such a simple and convenient metric is the *empirical error rate*. The empirical error rate is a simple metric which has the appealing feature of having an intuitive interpretation. A brief description of this metric follows:

Empirical Error Rate

Having the empirically generated distributions of matching and nonmatching similarity values, it is possible to compute an optimal threshold such that if a given pair of tool marks yields a similarity value above such threshold, it is assumed that the pair of tool marks under comparison match. Similarly, if a given pair of tool marks yields a similarity measure below the optimal threshold, it is assumed that the pair of tools marks under comparison does not match. The boundary or threshold value is selected to minimize the empirical error rate (defined as the mean of both false positive and

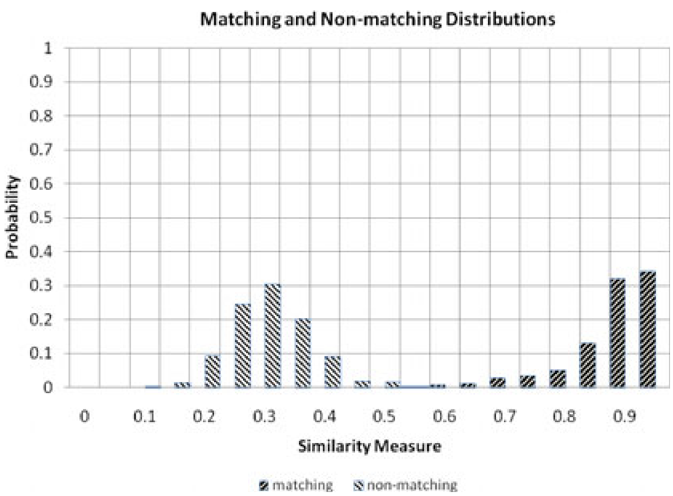

FIG. 11—*Matching and nonmatching distributions of similarity values for screwdriver striations on lead sheet at 30 degrees.*

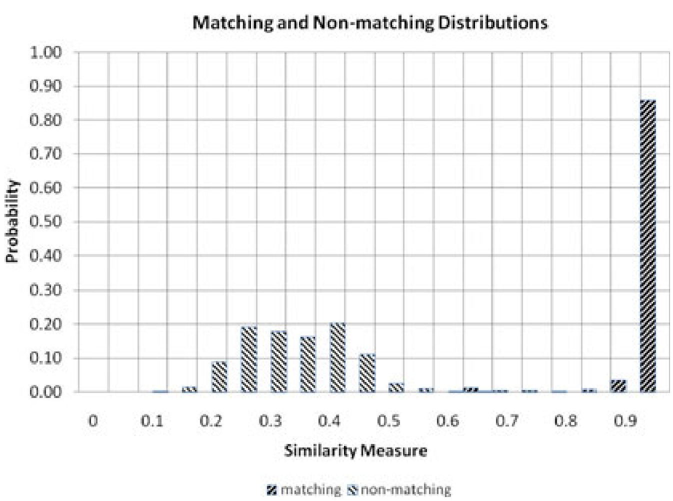

FIG. 12—*Matching and nonmatching distributions of similarity values for tongue and groove pliers striations on steel pipes.*

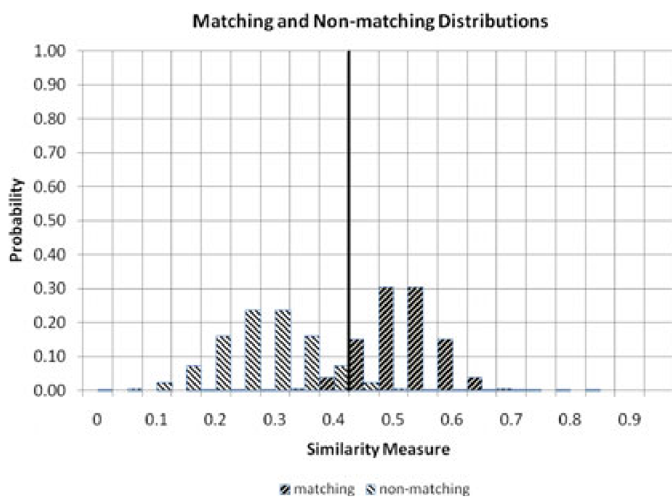

FIG. 13—*Empirical error rate estimation.*

false negative probabilities of error). We could have selected any other threshold value so as to shift the proportions of each type of error as desired. Figure 13 shows a graphical representation of this approach, where two distributions are shown—a matching distribution and a nonmatching distribution. Having identified the optimal threshold (vertical line), it is possible to estimate the probability of false positive and false negative identification. We use the empirical error rate as a metric of tool mark individuality, where a low empirical error rate is indicative of high specificity and repeatability.

It is important to note that the empirical error rates obtained as part of this study depend not only on the repeatability and individuality of the tool marks under consideration but also on the algorithms developed as part of the automated comparison system. These algorithms are significantly less sophisticated than the pattern recognition capabilities of a well-trained human tool mark examiner. Therefore, while the results presented in this paper have the benefit of objectivity, they are not meant to provide an estimate of the probability of an erroneous identification by an experienced tool mark examiner.

Results and Conclusions

In this section, we present the results obtained in each of the three scenarios described in the Methods section.

Scenario (a) Same Medium, Same Conditions

The empirical error rates for all screwdriver tool mark comparisons are summarized in Table 6. Among these results, the ones that correspond to Scenario (a) are located along the diagonal of the table (i.e., comparisons of SD01 vs. SD01, SD02 vs. SD02, SD03 vs. SD03, and SD04 vs. SD04). In all but one case, the empirical error rate is 0.00%. The only exception corresponds to SD01 versus SD01, where the empirical error rate 0.11% corresponds to a false exclusion out of 450 matching comparisons and no false inclusions out of 4500 nonmatching comparisons. There are no incorrect matches of two different tools for any of the same medium, same angle comparisons. These results indicate that for the media and angles of attack under consideration, the resulting screwdriver tool marks are sufficiently repeatable and specific to allow for very reliable identification. It may be significant that the only observed errors are at the highest angle of attack.

In a similar manner, Table 7 summarizes the results for tongue and groove pliers. As in the case of screwdrivers, those which correspond to Scenario (a) are located along the diagonal of the table (i.e., comparisons of TG01 vs. TG01, TG02 vs. TG02, and TG03 vs. TG03). Once again, in all but one case, the empirical error rate is 0.00%. The only exception corresponds to TG01 versus TG01, where the empirical error rate 0.03% corresponds to no false exclusions out of 450 matching comparisons, and three false inclusions out of a total of 4500 nonmatching comparisons. As for the screwdriver marks, these results indicate that for the media under consideration, the tongue and groove pliers tool marks are sufficiently repeatable and specific to allow for very reliable identification. The effect of the metals studied does not appear to be significant, since all metals produce very low error rates and the only errors observed are on brass which has hardness intermediate between that of lead and steel.

TABLE 6—*Empirical error rate for screwdriver tool mark comparisons.*

			Lead			Aluminum
			45deg	30deg	15deg	30deg
Empirical Error Rate			SD01	SD02	SD03	SD04
Lead	45deg	SD01	0.11%	13.61%	49.50%	X
	30deg	SD02		0.00%	33.51%	8.36%
	15deg	SD03			0.00%	X
Aluminum	30deg	SD04				0.00%

TABLE 7—*Empirical error rate for tongue and groove pliers tool mark comparisons.*

		Brass	Steel	Lead
Empirical Error Rate		TG01	TG02	TG03
Brass	TG01	0.03%	0.23%	2.46%
Steel	TG02		0.00%	1.58%
Lead	TG03			0.00%

Scenario (b) Different Media, Same Conditions

Table 6 includes the empirical error rates resulting from the comparison of screwdriver tool marks created under the same conditions (30° of attack) but onto different media (lead vs. aluminum: sets SD02 vs. SD04). As discussed for Scenario (a), the empirical error rate is 0.00% when screwdriver tool marks created onto the same medium at an attack angle of 30° are compared for both aluminum and lead. As shown in Table 6, it increases to 8.36% when tool marks on lead are compared with tool marks on aluminum. This 8.36% error rate corresponds to 63 false exclusions out of 1000 matching comparisons and 938 false inclusions out of 9000 nonmatching comparisons.

In a similar fashion, Table 7 includes the empirical error rate resulting from the comparison of tongue and groove pliers tool mark samples created in different media (comparisons TG01 vs. TG02, TG01 vs. TG03, and TG02 vs. TG03). For striation marks produced by tongue and groove pliers the medium onto which the tool marks are created has a measurable effect on the tool marks. The empirical error rate for brass versus steel comparison is relatively low at 0.23%, corresponding to represent four false exclusions out of 1000 matching comparisons and six false inclusions out of 9000 nonmatching comparisons. The reasonable success rate for these two metals probably results from the fact that they do not differ greatly in hardness. In contrast, comparison of marks on either brass or steel with those on lead result in higher error rates, 2.46% and 1.58%, respectively.

Scenario (c) Same Medium, Different Conditions (Screwdrivers Only)

Table 6 also includes the empirical error rate resulting from the comparison of screwdriver tool marks created on the same medium (lead) but under different angles of attack. As shown in Table 6, the variation of the angle of attack has a significant effect on the resulting tool mark even if the medium is the same. The error rates for comparison increase as the difference between the angle of attack is increased. The total error rates are pronounced enough that comparison of tool marks created at 15° with those created at 45° is no better than random guessing, which would have an error rate of 50% (close to the obtained 49.5%). The likely reason for the inability to correctly match tool marks made by the same tool at different angles of attack is that the points of the tool surface that are in contact with the receiving surface are different at the two angles.

Discussion

As stated at the beginning of this paper, the main goal of the study herein reported is to validate the basic premise of tool mark identification. As can be seen, the results obtained from this study provide substantial evidence to the validity of this basic premise of tool mark identification in the case of striated tool marks.

A number of important conclusions can be derived from the results, discussed in the previous section, as stated below:

- Striated tool marks produced by screwdrivers and tongue and groove pliers are both repeatable and specific enough to allow for reliable identification of the producing tool when they are created on the same medium and under the same conditions (for the media and tools used in this evaluation).
- When striated tool marks are created on different media but under the same conditions, the tool marks can still be identified with high reliability. In the case of tongue and groove pliers, it is interesting to note that the empirical error rate increases with an increase in the degree of dissimilarity in the hardness of the medium onto which the tool marks are created. This implies that while the practice of creating control tool marks in lead is a sound one from the perspective of avoiding damage to the tool's working surface, a higher degree of agreement may be achievable if tool marks are created onto media of similar hardness as that of the evidence tool mark.
- Screwdriver striated tool marks depend significantly on the angle of attack at which the tool mark is created (more so than with respect to the media). So much so, that tool marks created by the same screwdriver may appear completely different if created at drastically different angles of attack. Therefore, the comparison of an evidence screwdriver tool mark requires the creation of control tool marks at multiple angles of attack.
- It was observed that irrespective of the type of comparison (i.e., within the same sets such as TG01 vs. TG01 or between different sets such as TG01 vs. TG02, etc.), the nonmatching distributions obtained for a given tool type always had similar characteristics, in particular a low median and relatively low standard deviation. While this is not a surprising result, it has meaningful implications. First of all, it provides strong evidence to the premise that the probability of obtaining a high degree of similarity while comparing a pair of nonmatching tool marks is extremely low. If the behavior observed for the set of tools used in this evaluation can be considered as characteristic of all tools of the same type (which is likely to be the case at least for those tools manufactured by the same techniques), the probability of a pair of different tools having similar features is extremely low.
- It was observed that in some of the cases where both the conditions and media were the same (Scenario a) the empirical probability of error was not always zero. Upon inspection of the raw tool mark images, it was noticed that the nonzero probability of error was because of the presence of a very small number of "bad" tool mark images (where we loosely use the term bad to indicate that such tool marks display highly anomalous features as a result of the creation and/or acquisition process). These bad images resulted in a matching pair being erroneously classified as a nonmatching pair (and never the other way around). In other words, the probability of error originated from a faulty image, and not because the tool itself would not create repeatable and individual tool marks (as other tool marks created by the same tool resulted in perfectly good images). Given the low probabilities of error associated with these cases, even a single bad tool mark image can have a relatively significant effect.

Based on these observations, it is evident that the obtained results provide substantial evidence to the validity of the basic premise of tool mark identification. Furthermore, these results reinforce the validity of many current practices of tool mark examiners.

Acknowledgments

Funded by the Federal Bureau of Investigation under Contract Number J-FBI-02-128. Mention of trade names is for information purposes only and does not imply endorsement by the FBI or the federal government.

References

1. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
2. Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
3. Davis JE. An introduction to tool marks, firearms and the striagraph. Springfield, IL: Charles C Thomas Pub Ltd, 1958.
4. De Kinder J, Prevot P, Pirlot M, Nys B. Surface topology of bullet striations: an innovating technique. AFTE J, Spring 1998;30(2):294–9.
5. De Kinder J, Bonifanti M. Automated comparison of bullet striations based on 3D topography. Forensic Sci Int 1999;101(2):85–93.
6. Geradts ZJ, Zaal D, Hardy H, Lelieveld J, Keereweer I, Bijhold J. Pilot investigation of automatic comparison of striation marks with structured light. Proc SPIE 2001;4232:49–56.
7. Bachrach B. Development of a 3D-based automated firearms evidence comparison system. J Forensic Sci 2002;47(6):1253–64.
8. Banno A, Masuda T, Ikeuchi K. Three dimensional visualization and comparison of impressions on fired bullets. Forensic Sci Int 2004;140(3):233–40.
9. Miller J. An introduction to the forensic examination of toolmarks. AFTE J, Summer 2001;33(3):233–48.
10. Biasotti A, Murdock J. Criteria for identification or state of the art of firearm and toolmark identification. AFTE J 1984;16(4):16–34.
11. Miller J, McLean M. Criteria for identification of toolmarks. AFTE J 1998;30(1):15–61.
12. Nichols RG. Firearm and toolmark identification criteria: a review of the literature. J Forensic Sci 1997;42(3):466–74.
13. Nichols RG. Firearm and toolmark identification criteria: a review of the literature, part II. J Forensic Sci 2003;48(2):318–27.
14. Stoney DA. What made us ever think we could individualize using statistics? J Forensic Sci 1991;31(2):197–9.

Additional information and reprint requests:
Benjamin Bachrach, Ph.D.
Vice President
Director, Electro-mechanical Systems
Intelligent Automation, Inc.
15400 Calhoun Drive, Suite 400
Rockville, MD 20855
E-mail: bach@i-a-i.com