JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2015, Vol. 60, No. 2
doi: 10.1111/1556-4029.12642
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Ryan Spotts,[1] B.S.; L. Scott Chumbley,[1] Ph.D.; Laura Ekstrand,[1] M.S.; Song Zhang,[1] Ph.D.; and James Kreiser,[2] B.S.

Optimization of a Statistical Algorithm for Objective Comparison of Toolmarks*

ABSTRACT: Due to historical legal challenges, there is a driving force for the development of objective methods of forensic toolmark identification. This study utilizes an algorithm to separate matching and nonmatching shear cut toolmarks created using fifty sequentially manufactured pliers. Unlike previously analyzed striated screwdriver marks, shear cut marks contain discontinuous groups of striations, posing a more difficult test of algorithm applicability. The algorithm compares correlation between optical 3D toolmark topography data, producing a Wilcoxon rank sum test statistic. Relative magnitude of this metric separates the matching and nonmatching toolmarks. Results show a high degree of statistical separation between matching and nonmatching distributions. Further separation is achieved with optimized input parameters and implementation of a "leash" preventing a previous source of outliers—however complete statistical separation was not achieved. This paper represents further development of objective methods of toolmark identification and further validation of the assumption that toolmarks are identifiably unique.

KEYWORDS: forensic science, toolmark, algorithm, statistical comparison, pliers, striae, quasi-striated, shear cutter

In recent history, the legitimacy of scientific testimony has been questioned in several court cases—specifically *Daubert v. Merrell Dow Pharmaceuticals, Inc.* This challenge has had profound implications in the field of firearm and toolmark examination and resulted in many studies conducted to validate the practice of comparative forensic examination. The primary validation needed is the assumption of toolmark examination—every tool has its own unique surface that will leave a unique mark.

Screwdriver marks are among the most studied due to their uniform and continuous striae. They have been previously characterized using stylus profilometry and confocal microscopy in various attempts designed for the identification of matching and/or nonmatching toolmarks (1–4). The results from these types of studies typically show that striae may be successfully objectively compared using a computer algorithm with relatively high accuracy. For example, in a previous study by the authors using a statistical algorithm, marks from fifty sequentially manufactured screwdriver tips were successfully separated between matching and nonmatching pairs to a reasonable degree of accuracy (2).

Studies of other tools that produce striations have also been conducted. Pliers are another type of tool that can create a variety of marks. Cassidy, one of the first to study sequentially manufactured pliers (5), found toolmarks produced by plier teeth—such as when a burglar would twist off a door knob to enter a building—to be unique because the broaching process used to manufacture the plier teeth was performed in a direction perpendicular to the striae it would create. This study established the uniqueness of marks created by plier teeth; however the analysis was based on logical reasoning and not backed by mathematical analysis. More recently Bachrach et al. (4) studied tongue and groove pliers marks created on brass pipe, galvanized steel pipe, and lead rope. Test marks were made using a singular tooth from the pliers to create a striated mark. This study found that marks could be compared when made on different material but with less accuracy than marks made on the same material.

Petraco et al. (3) has studied striated chisel marks. The test marks created by the chisels were striated but discontinuous, resulting in patches of striations. Unfortunately, the nature of the created marks was too difficult for the employed software to analyze.

While regularly striated marks have received the majority of research attention, the extension of mathematically based studies to other forms of toolmarks is also highly desirable. The results discussed in this paper investigate the applicability of the algorithm employed in (2) to quasi-striated marks created by slip-joint pliers. This type of plier mark was chosen for two reasons. First, the type of mark produced, termed a shear cut, presents a more difficult pattern for identification than a fully striated mark. Second, pliers such as these and other tools that produce shear cut marks are routinely used by criminals to steal copper from construction sites. A July 30, 2013 report on CNBC stated that copper theft in the U.S.A. has become a 1 billion dollar industry (http://www.cnbc.com/id/100917758). Thus, objective examination and

[1]Ames Laboratory, Iowa State University, 2220 Hoover, Ames, IA 50011.
[2]Illinois State Police, Retired, 3112 Sequoia Dr., Springfield, IL 62712.

*Supported by the National Institute of Justice and the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division. The research was performed at the Ames Laboratory, which is operated for the U.S. DOE by Iowa State University under contract number DE-AC02-07CH11358. Funding was provided by award number 2009-DNR-119 from the National Institute of Justice.

Received 7 Oct. 2013; and in revised form 13 Feb. 2014; accepted 28 Feb. 2014.

identification of shear cut toolmarks are becoming increasingly important in law enforcement.

Results from an initial investigation on slip-joint pliers were conducted by Grieve (6). These results having shown promise, this paper presents results on shear cut marks made by 50 sequentially manufactured slip-joint pliers. While the initial results revealed the algorithm could correctly separate a large majority of matching and nonmatching pairs, some algorithm parameter values and options that work well for regularly striated marks are not optimal in the present setting. Two distinct deficiencies hindering algorithm operation were noted. The goal of this study was to investigate optimization of the parameters best suited for analysis of marks described as quasi-striated.

The first deficiency addressed involved parameters that affect the degree of statistical separation in the results. While separation was seen using the parameters employed for fully striated markings, better results could be obtained by changing the operational parameters of the algorithm. The second deficiency noted was concerning what the authors have termed the "Opposite End Problem". This problem manifests itself when, in a small number of cases, the algorithm declares a "match" from two data sets which are known to be nonmatching. Observation of the raw data files shows that the opposite ends of the two sets of toolmarks being compared are identified as the matching region. Such a match is physically impossible and results due to the inability of the algorithm to successfully complete the validation procedure, which is integral to the operation of the algorithm, when confronted with similar topography at opposite ends of the data sets. This possibility was first noted during research on regularly striated screwdriver toolmarks (1,2).

This study involves complete analysis of fifty pliers using various parameter values and an option that accounts for the "Opposite End" problem. The results of the study, including a brief description of the statistical algorithm, are discussed below.

Experimental Methodology

Fifty sequentially manufactured slip-joint pliers were obtained from Wilde Tool Co., Inc. It is common knowledge within the field of study that the manufacturing process significantly affects the toolmarks that are created (7,8). Thus, although the manufacturing process and test sample creation were previously described (6), it deserves restatement.

The pliers start as pieces of steel that are hot forged into half blanks. Each half blank was then cold forged once again using the same die for every piece. After forging, the first difference between half-pairs was introduced. Fifty halves were punched to create a small hole, while the other fifty halves received a double-hole punch—allowing future users to better hold a wider dimension range of objects. The gripping teeth and shear surface were next created with a broaching process. Two broaching machines were used in the production of the pliers. Plier halves with the double-hole punch went to one machine, while pliers with a single-hole punch went to the other. During this separation time, the manufacturer stamped the numbers 1–50 on the plier halves, so the correct sequence could be ensured. The broaching process on the shear surface created the characteristic nature that is of interest for this study.

After broaching, both halves of each plier were given the same heat treatment and shot peened to strengthen the material and increase the surface hardness. The flat side regions were next polished, and the double-hole punch half was branded with the company logo. The plier half with the double-hole punch and company logo was labeled as the "B" side for every plier pair, and the other side "A". An overview of pliers from unfinished to finished states is shown in Fig. 1.

Wire test samples were created using bolt cutters to cut 2" samples from wire spools. The bolt cut ends were marked using a permanent marker so they could not be confused with the plier shear cut surfaces. Diameters of the wire used were 0.1620" for the copper and 0.1875" for the lead. Test marks were made by shear cutting the copper and lead wire. Shear cutters are defined by the Association of Firearm and Toolmark Examiners as "opposed jawed cutters whose cutting blades are offset to pass by each other in the cutting process" (9). As the shear face was used on the pliers to make the samples, by definition the created marks are shear cutting marks. Figure 2 pictorially shows the

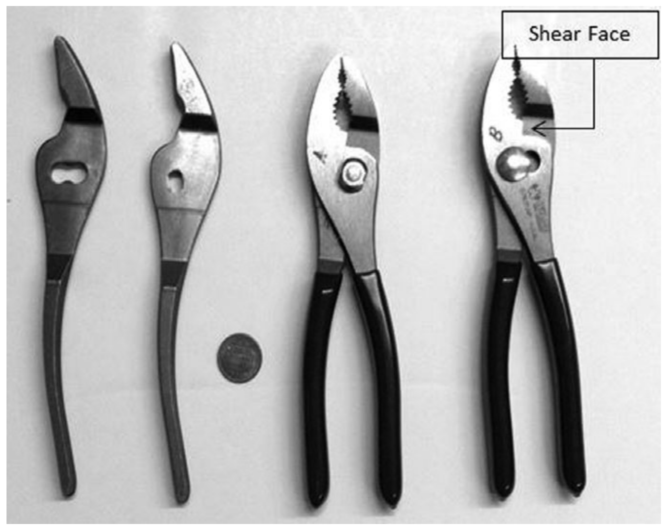

FIG. 1—*Slip-joint pliers in finished and unfinished states. The "A" and "B" sides were labeled as shown for every plier, with "B" appearing on the branded side.*

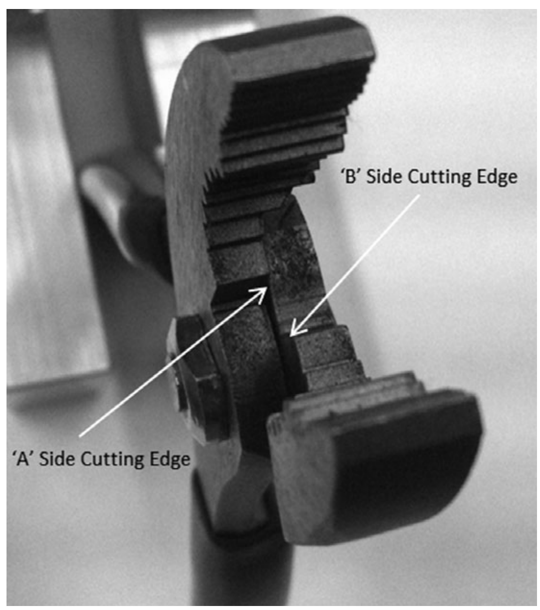

FIG. 2—*Shows the exact location on the pliers used to shear cut wire samples.*

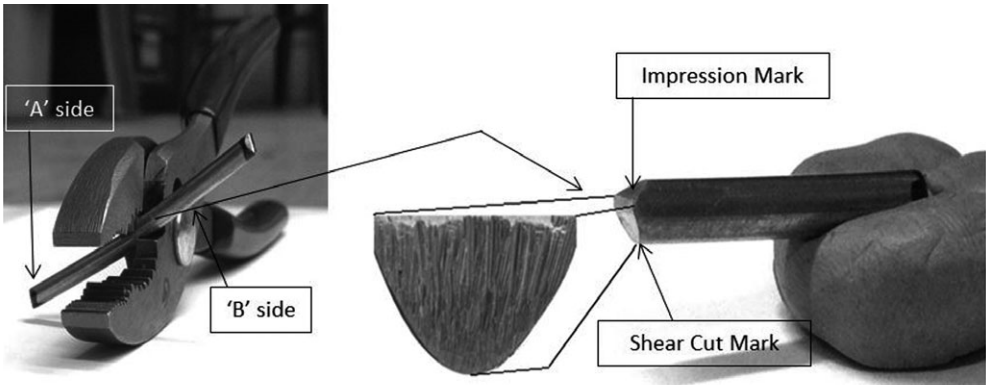

FIG. 3—*The left photograph shows an example wire sample mid-shear cut, revealing how the toolmark gains its angle. The right photograph shows the "B" side sample, also revealing why the created mark is not completely circular. A similar "A" side sample also exists but is not shown in the right photograph.*

exact location used on the pliers to create the toolmarks. Each shear cut was made by placing the sample between the shear surfaces—with the "B" surface always facing downward—marking the sides "A" and "B" corresponding to which plier shear surface would be acting on that section of the wire. Thus, two samples, one "A" and one "B", were created with each shear cut. The samples were shear cut alternating between copper and lead until ten of each sample type were created. This resulted in 2000 total samples: 1000 samples for both copper and lead with half of each coming from each side of the pliers. For consistency, every sample was made by the author who is a retired forensic examiner.

When the wire is mechanically separated, the two surfaces of the shear edges move past each other. The resultant action is therefore a combination of both cutting the surfaces and a shearing action of the edges as they move through the material. The result is two surfaces being created on each half of the separated wire sample, comprising both shear cut and impression markings, roughly at 90° to each other with both being ≈45° to the long axis of the wire. Only the shear cut surfaces on the "A" and "B" sides of the sample were scanned and analyzed. A schematic showing the process is shown in Fig. 3.

The scope of this study included only the copper samples, leading to a total sample size of 1000. To obtain the surface data from the samples, each piece was scanned using an Infinite Focus Microsope G3 (Alicona). Scans were completed at 10× magnification with a two micron vertical resolution. An example image obtained using the IFM is shown in Fig. 4. Similar to the initial study (6), the data were taken from two locations. The long edge (solid line in Fig. 4) is near where the shear cut began and the short edge (dashed line) is nearer where the shear cut ended. Striae near the beginning of the shear cut are longer and more regular than striae near the end of the shear cut, so it is important to observe the results at both locations. It is clear from viewing the figure that the pliers created a quasi-striated surface—a surface consisting of groups of parallel striae that are not continuous along the length of the mark.

An example of the scanned data from the infinite focus microscope prior to and post noise reduction is shown in Fig. 5. The cut surface is embedded in irregular spiky noise, which arises from the sample's edge and the background generated when making an IFM scan. This spiky noise must be removed so that it does not interfere with the statistical analysis. To remove this noise, the authors used a combination of automated cleaning algorithms and manual cleaning. The automated cleaning algorithms are described in more detail in (10), but a brief description of these follows. First, the 2D image texture and quality map from the IFM operating software were used to remove those points that were too dark or had a poor quality value. These points cause spikes in the visual output. Next, a seventh-order polynomial fitting was generated for each row of the data. For each point, the discrepancy between the measured and predicted depth was computed, and points with a discrepancy of 100 microns or greater were discarded. This process was repeated for the columns. Finally, small holes (<20 pixels in diameter) were filled through linear interpolation. To remove

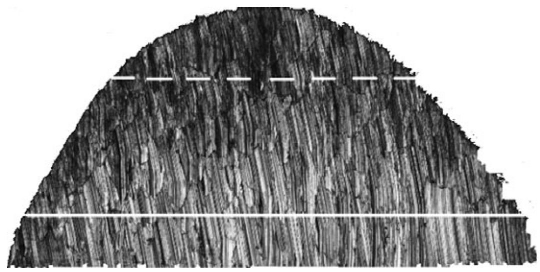

FIG. 4—*Example scan of a sheared copper wire. The dashed line is the short edge while the solid line is referred to as the long edge.*

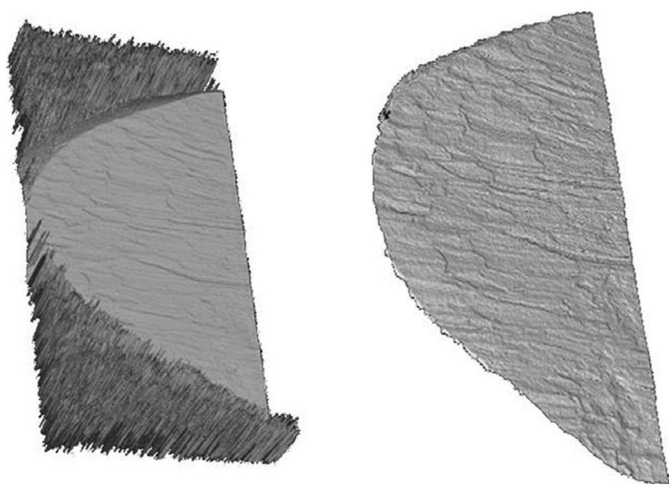

FIG. 5—*Uncleaned data are on the left while the detrended and cleaned data are shown on the right.*

any remaining spikes and the sides of the cut wire, the authors used a visual painting program to paint over noisy regions. The computer algorithm then interpreted the painted areas as data points to exclude from the analysis.

As it is impossible to scan every sample at precisely the same angle relative to the equipment, it is necessary to correct for this sample angle using a process called detrending. To detrend the data, linear least squares were used to fit a plane to the data. To make this process faster and less sensitive to noise, only 80 points were used in the plane fitting. These points were selected in an "X" pattern that evenly covered the majority of the sample surface. Once the plane fitting was obtained, the plane was subtracted from the surface data to remove the global surface angle.

When employed in the initial study (6), the comparative algorithm used was discovered to have the same limitation that prevented it from operating effectively in certain instances in (2). For a more complete discussion of the algorithm, the reader is referred to (2). Briefly, the algorithm works in two major steps, the optimization and validation steps. An iterative "search" window of user-determined size (in pixels) is held stationary on Trace 1, while the correlation to a same size window is calculated over the entirety of Trace 2. The window is then shifted one pixel over on Trace 1 and the process is repeated. This is performed until the two regions of best correlation are found. Figure 6 schematically shows this process.

Once the region of highest correlation is found during the optimization step, two shifts are applied and compared—random shifts and rigid shifts. This is the validation step. The size of the "validation" windows that are shifted are user determined. During the rigid shift step, a user-defined window is moved a set distance from the best correlation window on each trace and the correlation at that point is calculated. For the random shift step, the same size window is moved randomly calculated distances from the best correlation window for the two comparison scans and the correlation is again calculated. An example of rigid and random shifts is shown in Fig. 7. The number of rigid and random shifts employed is also user defined; for the purposes of this study, the number was set at 50. Comparison of the rigid shift correlation values to the random shift correlation values by the algorithm produces the statistical values to be mentioned.

In the initial investigation (6), outlier data points were observed to stem from the algorithm misidentifying the opposite ends of marks as a positive match. One example of this is shown in Fig. 8. The solid line orthogonal to the shear cutting direction is the location the profiles were compared along and the dashed lines represent coordinate axes.

As Fig. 8 illustrates, through random chance, opposite ends of a mark are occasionally selected as having the regions of highest correlation between marks for the selected window size. Clearly,

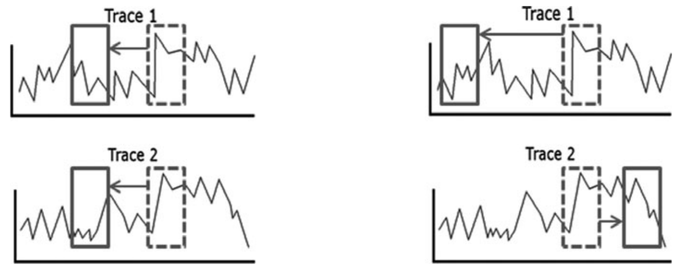

FIG. 7—*A rigid shift is shown on the left and a random shift is shown on the right.*

given that the shape of the shear cut wire specifies a definitive left and right side of the shear cut, it is physically impossible for this pair of windows to correspond. In investigating the cause for the false match, it was discovered that in cases where the region of highest correlation between two marks occurs at the end of the scan profile, the validation routine used by the algorithm to ascertain the quality of the comparison cannot function properly. When a "match" is found near the end of a scan profile, the space needed to successfully accomplish the rigid shifts and complete the validation step does not exist. This results in an incorrect validation, and a "match" being declared when in fact a nonmatch may exist.

To address this problem, a "leash" was applied to the search window of the original algorithm (2) during the optimization step, the purpose being to limit the comparison distance between profiles. In this case, the comparative correlation is no longer calculated over the entirety of Trace 2 for each iteration of the search window, but only to a certain percentage of the entire distance. Figure 9 shows schematically an example of how the leash limits the search range for the region of highest correlation. Leashing the search window makes it impossible for the algorithm to identify regions far from each other on the real surface as matching. Where contextual information exists concerning the shape of a mark (such as exists for a distinctive mark like those used) this in no way affects the objective performance of the algorithm.

The current version of the leash is set as a percentage of the total length of the trace. The leash was set at 80% for this analysis. Figure 10 shows the same plier comparison as Fig. 8 but after the leash was implemented. The algorithm clearly finds a reasonable location for best correlation but now computes a low value for the Wilcoxon rank Sum test statistic (T1), indicating a nonmatching pair.

A Wilcoxon rank Sum test statistic (centered and scaled to have a nominal SD 1) is calculated during the validation step and is what is returned by the algorithm. The T1 statistic is determined by comparing the results of rigid and random shifts. Matching marks should have relatively high correlation after a rigid shift if they are truly similar and lower correlations during random shifts. The magnitude of the T1 statistic is affected by how much the rigid and random shifts differ. High rigid shift correlation and low random shift correlation would result in a high T1 value—indicating a matching pair—while the opposite scenario would result in a low or negative T1 value indicating a nonmatching pair. The reason many shifts are applied is because random chance may allow a few random shift windows to have a high correlation. As more shifts are applied to a matching pair, the probability of observing a small T1 statistic will decrease. As more shifts are applied to a nonmatching pair, the expected

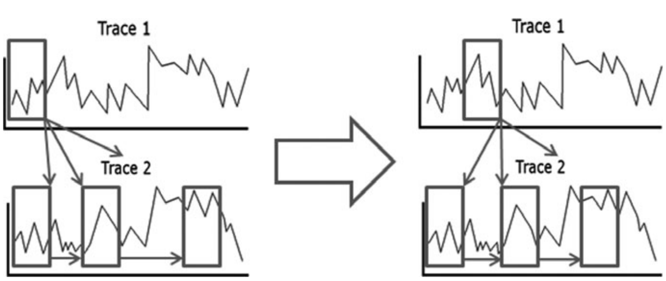

FIG. 6—*Generalized example of the iterative optimization step.*

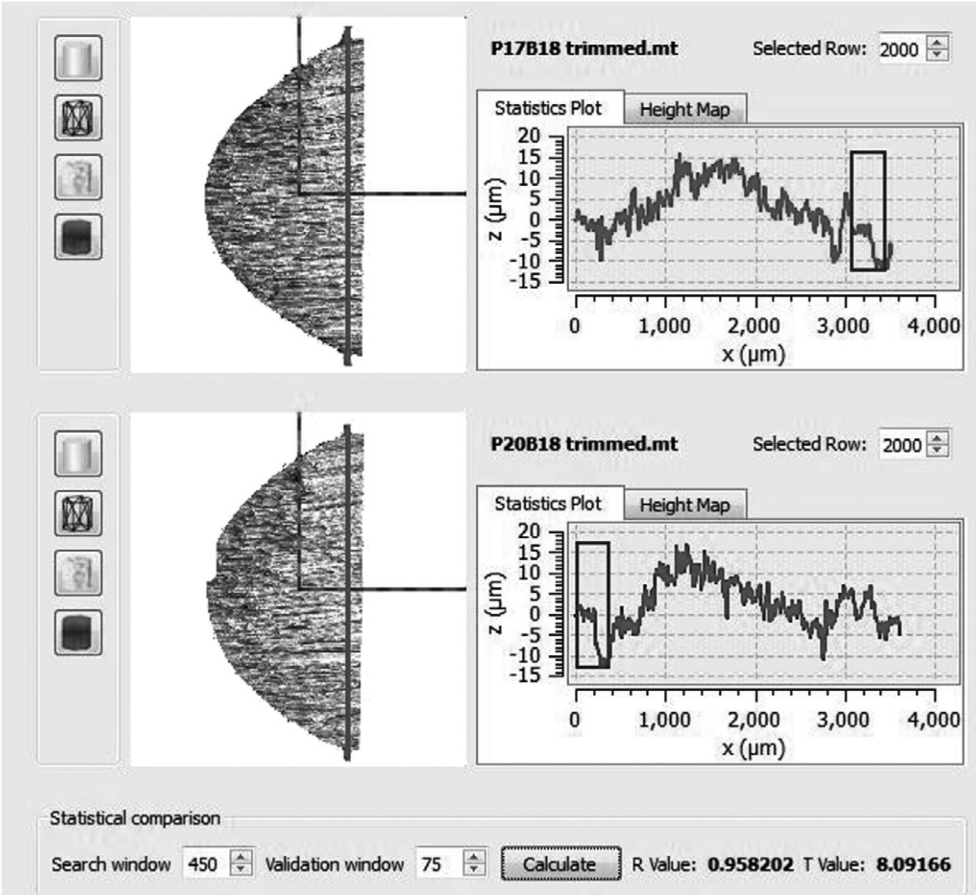

FIG. 8—*Example of computer algorithm false positive (T1 = 8.09). Opposite ends of the toolmark were misidentified as a positive match—indicated by the "high" value for the T1 statistic (See text).*

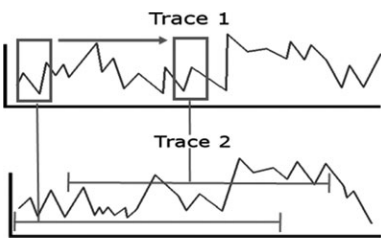

FIG. 9—*Generalized example showing possible ranges for the iterative search window.*

trend would be average rigid and random shift correlations that become closer in value—resulting in a T1 statistic value near zero.

Currently, there is not a definitive T1 value that perfectly defines when a match or nonmatch pair has been confirmed. This is due to the nature of the data—the comparisons are not independent events. Even if a definitive value was created for this study, it would likely not be applicable to other toolmark comparisons due to inherent differences in toolmark variability between different tools. Statements of correct or incorrect "identification" in this paper are qualitative—when the matching pairs consistently have significantly larger T1 values than nonmatching pairs, it is fair to state that the algorithm is correctly separating ("identifying") the majority of the pairs. A more advanced statistical argument is necessary to truly state whether an individual comparison was correctly identified.

Results

The data from the fifty sequentially manufactured pliers were compared using three different types of comparisons resulting in three sets of data. All three comparison types were performed using data from both the long and short edges as defined in Fig. 3.

Set 1: Comparing known matching pairs. Data for Set 1 were created by comparing marks made by the same side of the same pliers. Comparisons were made between marks 2 and 4, as well as marks 6 and 8 for both sides of each plier. An example of the methodology for comparisons in Set 1 is best described in a tabular format; an example of the comparison order through two pliers is shown in Table 1.

Set 2: Comparing known nonmatching pairs. Data for Set 2 were created by comparing marks made by different sides of the same pair of pliers. This set could confirm that both sides leave a unique mark. Comparisons were made between sides "A" and "B" for marks 10, 12, and 14. An example of the methodology for comparisons in Set 2 is shown in Table 2.

Set 3: Comparing known nonmatching pairs. Data for Set 3 were created by comparing marks from the same side of different pliers. Marks 16, 18, and 20 were compared between different pliers for both sides. An example of the methodology for comparisons in Set 3 is shown in Table 3.

Search and validation window sizes of 200 and 100 pixels, respectively, were used as part of the initial analysis. These window sizes had been previously used for successful matching of

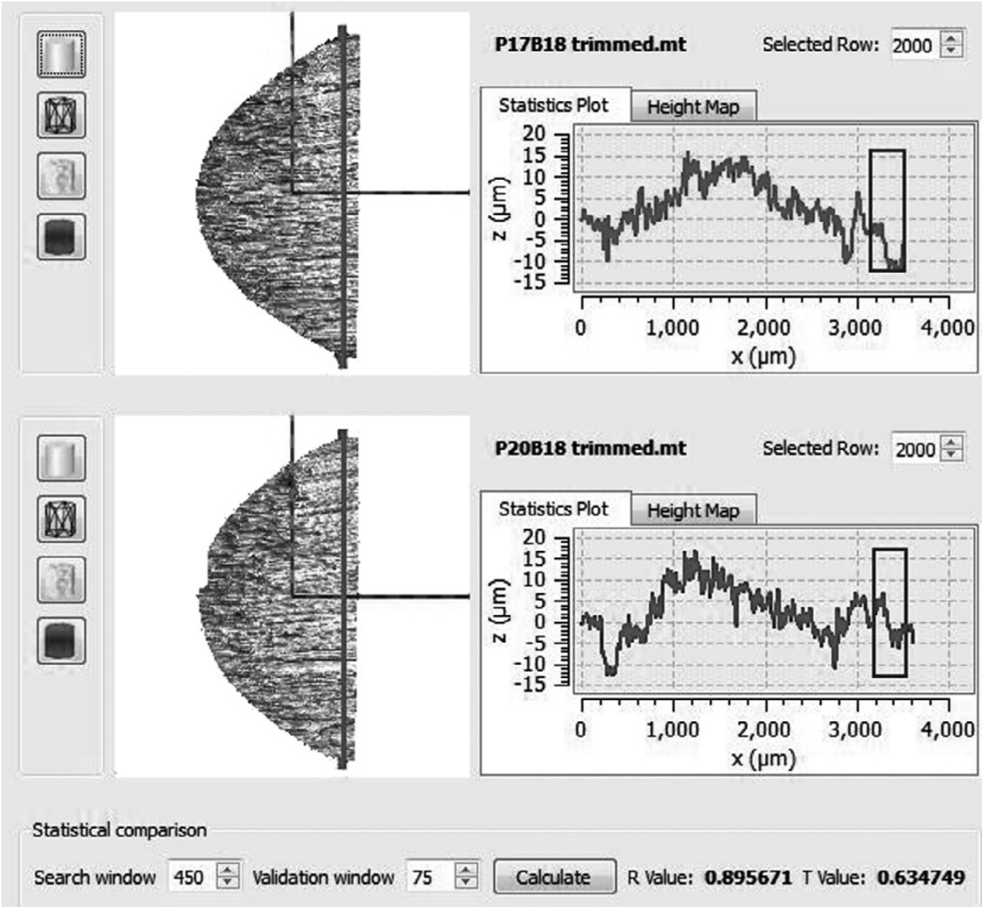

FIG. 10—*Plier comparison from Figure 8 after implementation of the leash. Comparison is now consistent with the expected result for nonmatching toolmarks (T1 = 0.63).*

TABLE 1—*Example methodology for Set 1.*

Comparison	Plier Number	Side	Mark Number	Plier Number	Side	Mark Number
1	1	A	2	1	A	4
2	1	A	6	1	A	8
3	1	B	2	1	B	4
4	1	B	6	1	B	8
5	2	A	2	2	A	4
6	2	A	6	2	A	8
7	2	B	2	2	B	4
8	2	B	6	2	B	8

TABLE 2—*Example methodology for Set 2.*

Comparison	Plier Number	Side	Mark Number	Plier Number	Side	Mark Number
1	1	A	10	1	B	10
2	1	A	12	1	B	12
3	1	A	14	1	B	14
4	2	A	10	2	B	10
5	2	A	12	2	B	12
6	2	A	14	2	B	14

TABLE 3—*Example methodology for Set 3.*

Comparison	Plier Number	Side	Mark Number	Plier Number	Side	Mark Number
1	1	A	16	2	A	16
2	3	A	16	4	A	16
3	1	A	18	4	A	18
4	2	A	18	3	A	18
5	1	A	20	3	A	20
6	2	A	20	4	A	20

screwdriver toolmarks (2). The results for all three data sets are shown in Fig. 11 presented as box and whisker plots. The solid black line represents the median value of the comparisons. The upper and lower bounds of the box represent quartiles, and the whiskers are within one and a half times the difference of the quartiles. Outliers are denoted by circular dots. A T1 statistic close to zero indicates little or no correlation between the data sets (i.e., a nonmatching pair), while a larger positive value would indicate a correlation exists between the two data sets being compared.

Observation of Fig. 11 shows that the algorithm performs reasonably well for the quasi-striated plier marks. Both the long and the short edge comparisons show significantly higher T1 values for the known matches of Set 1 than for the known nonmatches of Sets 2 and 3.

Results from the initial investigation that used the original algorithm are shown in Fig. 11 for comparison. The same source data were used in both experiments. Note that the leash included as a fix to the opposite end problem has resulted in a substantial

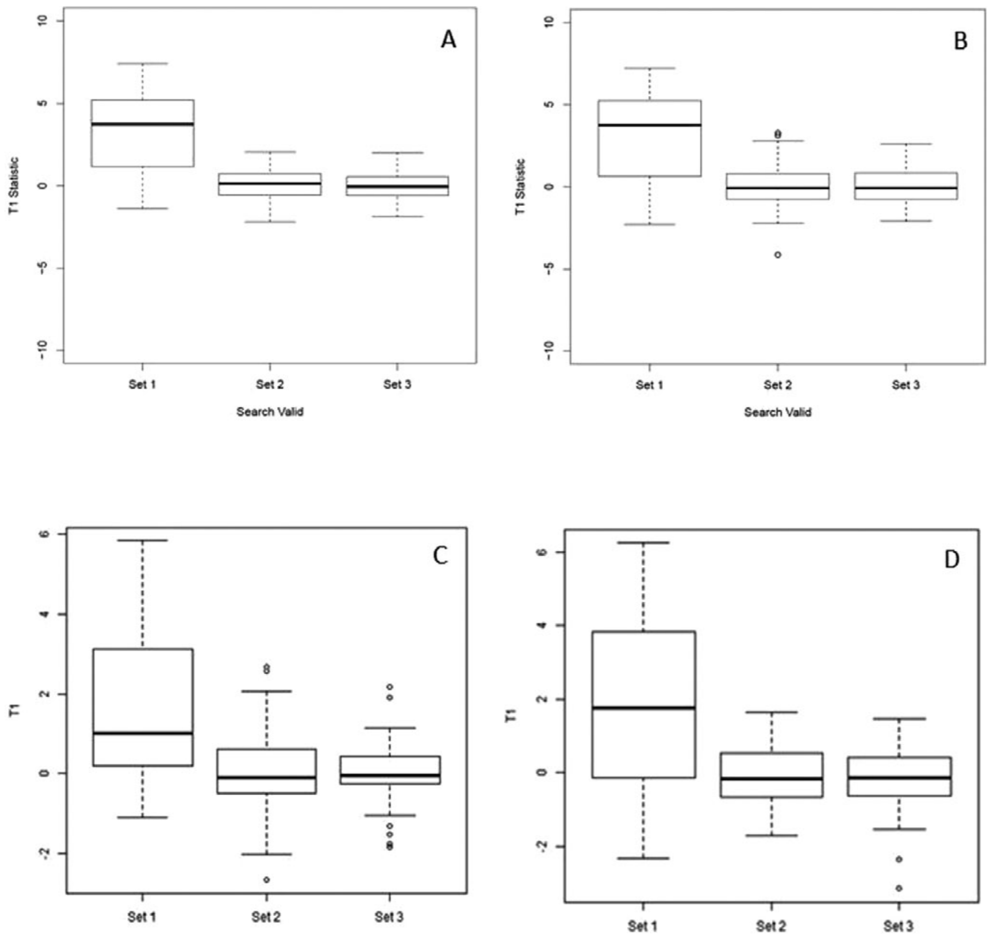

FIG. 11—*Results for search and validation window sizes of 200 and 100 pixels. (A) Current long edge results (B) current short edge results (C) prior long edge results and (D) prior short edge results (6). Note the difference in scale between the current data and early published results.*

improvement of algorithm performance over the original version. The median value for Set 1 has been increased from a T1 value of *c.* 1.5–2 to a T1 value of 4, and even more importantly there is now statistical separation of the quartiles for the long edge data. Data Sets 2 and 3 are still centered approximately on zero, and there is a net reduction in the total number of outliers.

As the quasi-striated marks produced by the plier shear cuts are far less regular than the previous screwdriver marks studied, experiments were conducted to determine the effect window size (i.e., search and validation) may have on the results. The 2:1 window size ratio was maintained for this second round of analysis with the search windows set to 1000, 500, 200, and 100 pixels with corresponding validation window sizes of 500, 250, 100, and 50 pixels. The results of these experiments for the short and long edges are shown in Figs 12, 13, and 14. Figure 12 shows that except for the smallest window size (100–50), Set 1 always has a median value well above zero, with the median increasing as window size increases from *c.* 4.0 (200–100) to 7.5 (1000–500). In contrast, the median values for Sets 2 and 3 always hover near zero as expected for nonmatching pairs, regardless of the window size. An apparent increase in data spread and number of outliers is also observed with increasing window sizes. Results from both long and short edges were similar to each other.

In some cases during the analysis, the algorithm would not return a result for every comparison. This is because the algorithm does not allow validation windows to overlap. Thus, as larger and larger window sizes are used it becomes more likely that the algorithm will run out of profile length, especially on short edges with large window sizes, and not return a T1 value. For a 2:1 ratio, the algorithm did not return 6 values for the Set 1 short edge, 9 values for the Set 2 long edge, 13 values for the Set 2 short edge, and 19 values for the Set 3 short edge. These numbers should be compared to the total of 3965 data comparisons that did return a result for the 2:1 ratio analysis. The algorithm returned a result more than 98% of the time.

With a clear trend in the results due to window size, the effect of size ratio was next analyzed using both 4:1 and 6:1 search to validation window size ratios. Search windows were set to 800, 600, 400, and 200 pixels with corresponding validation window sizes of 200, 150, 100, and 50 pixels used for the 4:1 ratio experiment. Search windows were set to 750, 600, 450, and 300 pixels with corresponding validation window sizes of 125, 150, 75, and 50 pixels for the 6:1 ratio experiment. The results from these analyses are shown in Figs 15–20 for the short and long edges.

Observation of Figs 15 and 16 shows that the clear trend of increasing T1 value with increasing window size observed for the 2:1 ratio holds true for both the 4:1 and the 6:1 ratios for Set 1 comparisons. An increasing number of outliers were also observed with increasing window size. The median values for the Set 1 data ratios were still qualitatively significantly above a zero value, with medians approaching a T1 value of 6.

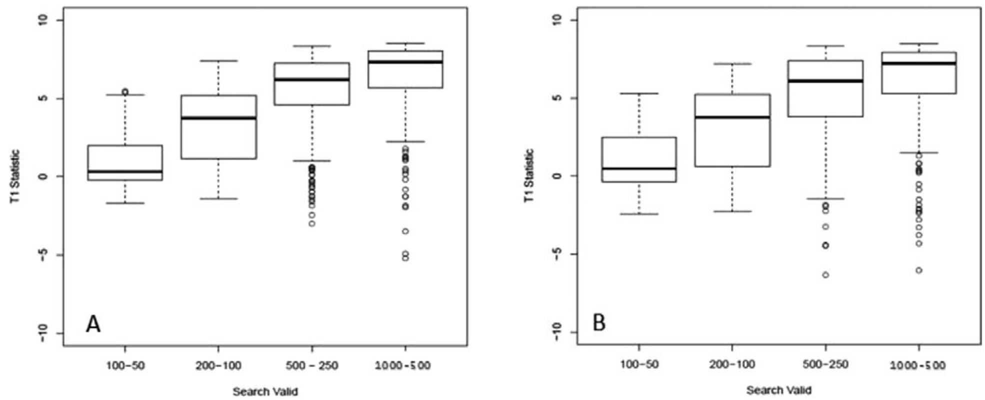

FIG. 12—*Data Set 1 results for a 2:1 window ratio. (A) Long edge. (B) Short edge.*

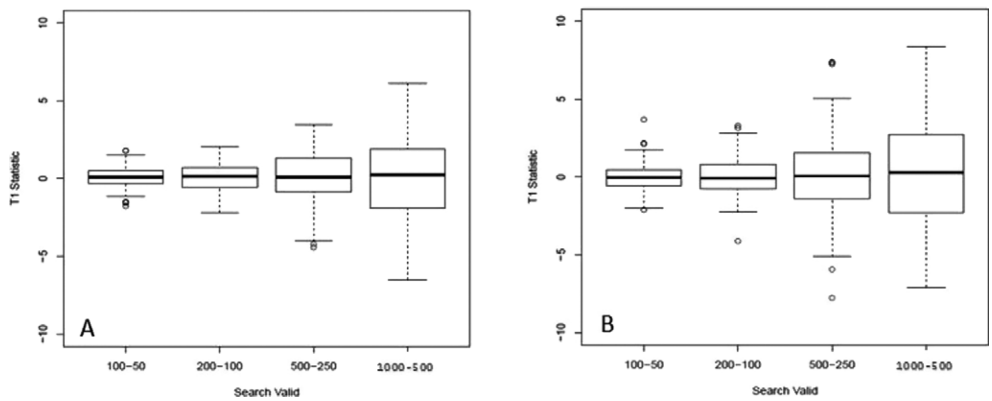

FIG. 13—*Data Set 2 results for a 2:1 window ratio. (A) Long edge. (B) Short edge.*

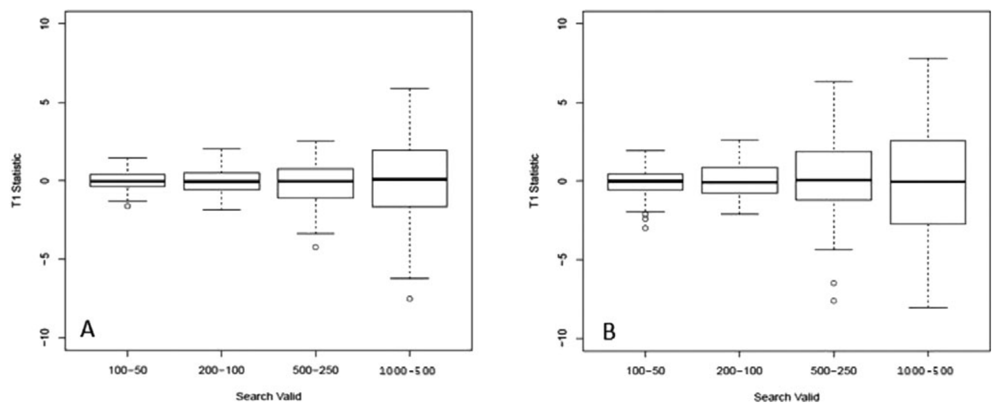

FIG. 14—*Data Set 3 results for a 2:1 window ratio. (A) Long edge. (B) Short edge.*

Set 2 and 3 comparisons, known nonmatches, are shown in Figs 17–20. Observations for Set 2 and 3 comparisons showed a median value near zero regardless of window size, increasing data spread with increasing window sizes and no clear trend in the number of outliers. In general, long edge comparisons had better results evidenced by the general decrease in data spread. The algorithm did not fail to return any results for the 4:1 and 6:1 ratios. This analysis contained 4012 comparisons for each ratio.

Discussion

The results presented add further credence to the basic assumption involved in toolmark identification, namely, that all manufactured tools are unique due to the machining processes used in their manufacture. This uniqueness is transferred to toolmarks as the tool is employed. Use of advanced characterization methods and computer algorithms can, to a large degree, allow objective comparison and identification of a series of toolmarks.

When the research transitioned from regularly striated to quasi-striated marks, it became apparent that a parameter optimization of the algorithm employed is necessary for different tools. This optimization led to improved results. The algorithm used in this research was optimized to provide better results for the current set of toolmarks by experimentally changing window sizes and utilizing an option that limits errors due to the opposite end problem. While the leash restriction is effective, it should be

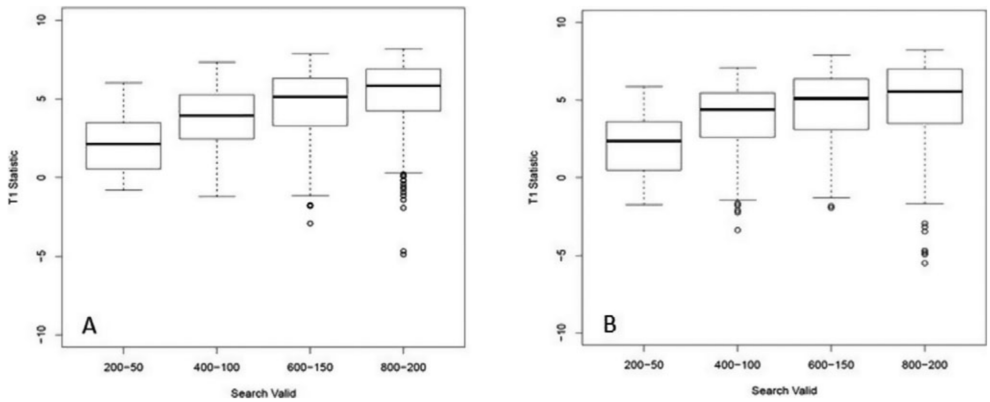

FIG. 15—*Data Set 1 results for a 4:1 window ratio. (A) Long edge. (B) Short edge.*

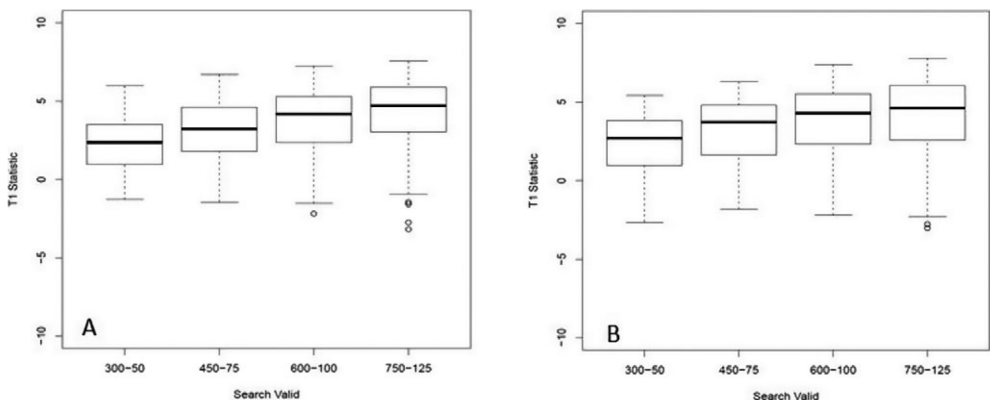

FIG. 16—*Data Set 1 results for a 6:1 window ratio. (A) Long edge. (B) Short edge.*

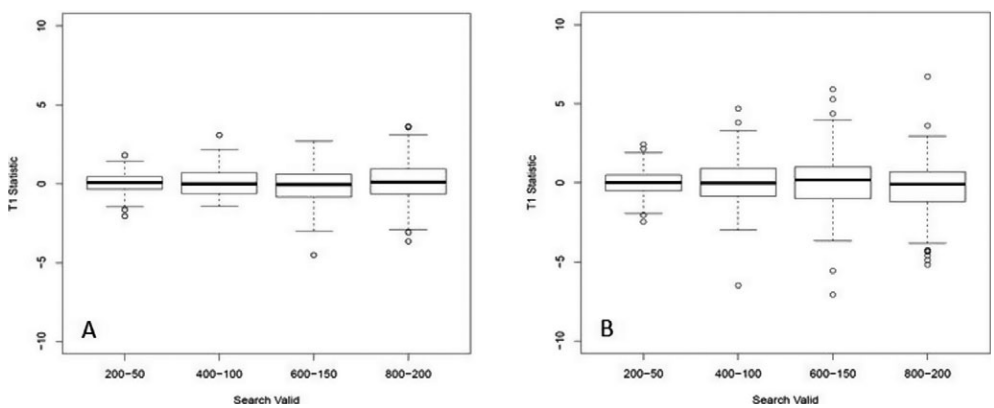

FIG. 17—*Data Set 2 results for a 4:1 window ratio. (A) Long edge. (B) Short edge.*

realized that its effectiveness is only made possible by the introduction of contextual knowledge into the analysis. For the plier marks, the nonsymmetric shape of the shear cut makes it easy to determine in which direction the scans should be analyzed. A more symmetric plier mark might be more difficult to orient properly to make use of the leash, involving a trained examiner to ensure the data were obtained correctly.

In the most ideal scenario, there would be complete data separation between known matching and nonmatching pairs, giving a clear indication of correlation, with no outliers in the data. Although ideal degree of separation has not been achieved, there is clearly a large majority of correctly identified toolmarks. Close examination of the outlying data points from both edges reveals that for these specific comparisons the algorithm produces a correct result for the vast majority of window combinations used. For example, consider Table 4 where several individual outlying points are shown. Incorrect matches, bolded and italicized, were found that were inconsistent with the algorithm results for other window sizes. Of the 12 different window sizes employed, the algorithm typically returns a "correct" answer for most of the window combinations. Note also that the majority of these outlier points stem from larger window sizes.

If the underlying hypothesis behind the application of the T1 statistic is that matching pairs will have more correlation than

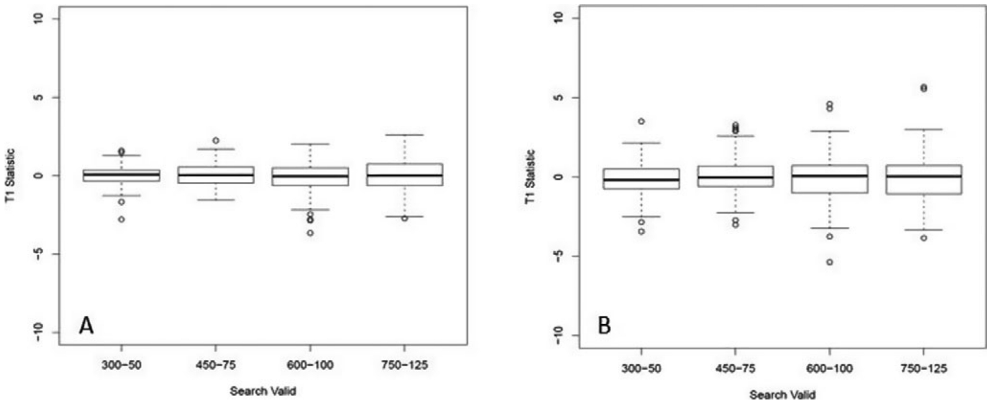

FIG. 18—*Data Set 2 results for a 6:1 window ratio. (A) Long edge. (B) Short edge.*

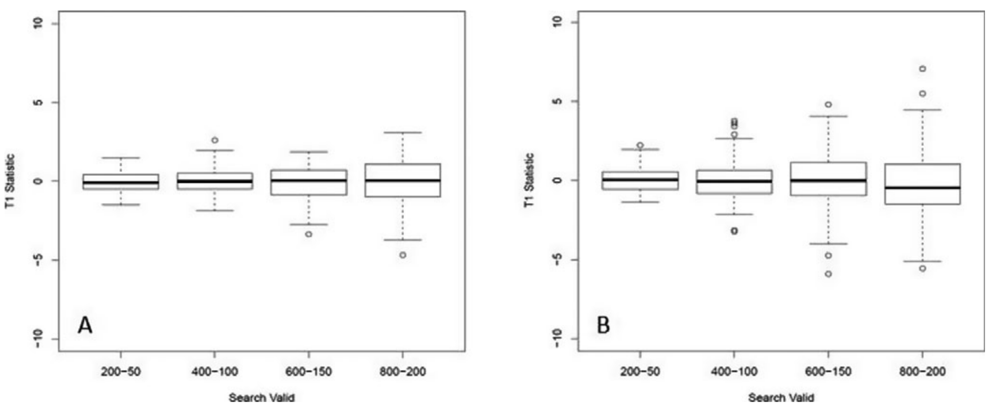

FIG. 19—*Data Set 3 results for a 4:1 window ratio. A) Long edge. B) Short edge.*

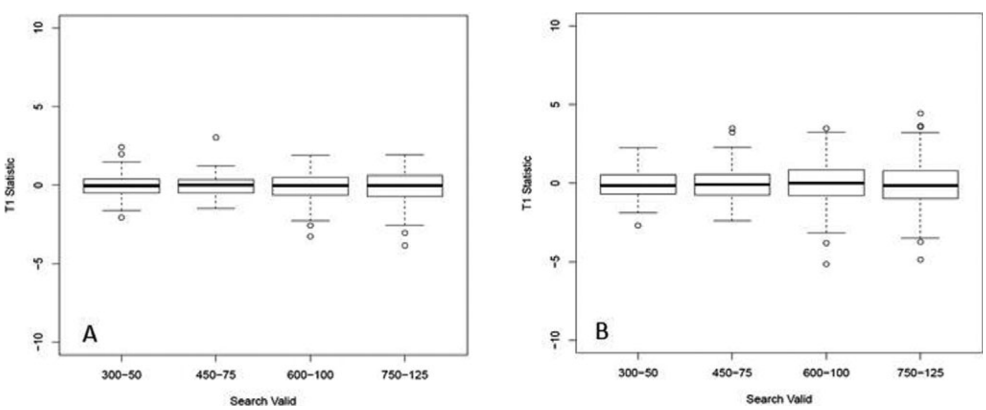

FIG. 20—*Data Set 3 results for a 6:1 window ratio. (A) Long edge. (B) Short edge.*

nonmatching pairs, one might assume that this also holds true if one uses more search and validation window combinations. A simple experiment was performed to see the effect of using multiple search and validation window combinations simultaneously could have on separating known matches from known nonmatches. The data from the *c.* 6000 possible long edge comparisons were used for this exploratory analysis. The average T1 values from all 12 search and validation window combinations were determined for each comparison. The results were organized by data set and are shown in Fig. 21. The average result returned by the algorithm when all window sizes are considered remains considerably above a zero T1 value for matching pairs and remains near zero for nonmatching pairs.

While the above observation is interesting, it is recognized that providing any statistical merit to such an analysis would require substantial further development of the arguments on which the algorithm is based. This is necessary as correlations based on different algorithm parameter values will have different distributions, both in matching and in nonmatching cases. Such a task is not anticipated at this time for this research group; however, it is suggested as a possible field of endeavor for the reader.

TABLE 4—*Comparison of returned T1 values for selected outlier data points.*

Sample	Data Variation for Specific Comparisons													
	Search and Validation Window Size (Pixels)												Data Set	Match
Plier Comparison	100–50	200–100	500–250	1000–500	200–50	400–100	600–150	800–200	300–50	450–75	600–100	750–125		
P31A16 – P32A16	0.60	0.75	−2.72	***5.91***	0.01	−0.54	−2.30	0.55	−0.01	−0.02	−1.74	−2.33		
P27B16 – P28B16	0.18	−0.09	***5.45***	####	0.33	0.01	1.27	0.62	−0.86	−0.88	−0.67	0.16	3	No
P34A18 – P35A18	0.03	−0.64	0.51	***5.45***	−0.86	−0.99	−0.16	1.03	−1.04	−1.09	−0.76	0.16	3	No
P2B18 – P3B18	0.59	2.08	***5.96***	0.03	1.71	−1.17	***4.08***	1.75	−1.78	1.47	3.08	0.70	3	No
P20A14 – P20B14	−0.53	−0.16	***3.99***	0.32	−0.17	−0.64	2.16	***6.71***	−1.04	1.10	0.92	1.72	2	No
P40A14 – P40B14	−1.16	−0.02	***7.36***	−4.22	−0.42	0.23	***4.38***	−2.27	0.62	−1.01	1.54	−2.20	2	No
P4A12 – P4B12	−0.66	−0.54	0.40	***5.66***	−0.36	−0.29	1.75	0.70	0.07	0.13	2.00	1.07	2	No
P18A12 – P18B12	1.02	−0.02	1.53	***5.53***	0.79	−0.24	0.46	−0.04	−0.49	−0.66	0.08	0.50	2	No
P30B2 – P30B4	0.14	1.20	3.62	6.11	0.66	1.55	3.15	***−5.52***	0.40	0.66	2.22	−1.42	1	Yes
P37B2 – P37B4	−0.51	0.37	3.33	***−4.31***	0.63	3.02	1.26	−0.31	2.55	0.87	−0.21	−0.55	1	Yes
P23B6 – P23B8	1.74	−1.15	−0.36	6.22	−0.54	3.51	−0.93	***−4.80***	2.02	−0.48	−0.58	***−3.02***	1	Yes

Note: '####' is an example of the algorithm not returning a value.

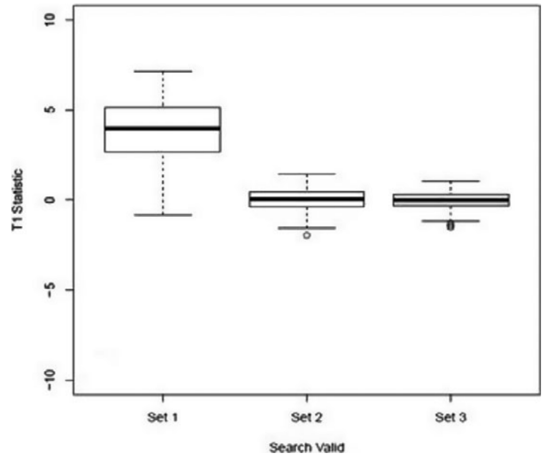

FIG. 21—*Average T1 values obtained when combining multiple window sizes for analysis.*

Observation of Table 4 suggests that each unique type of toolmark will most likely require a study to determine what the best set of operating parameters is for that particular mark and that any attempt to operate the code using a "one size fits all" mentality is insufficient for this algorithm. Future work will investigate whether an automated method to quickly evaluate multiple search and validation window combinations is possible. Such an enhancement would greatly speed analysis as new and more complicated toolmarks are examined.

Finally, while the quasi-striated marks examined in this study involve an added complexity when compared to the regularly striated marks previously examined, they are still less complex than, for example, impression marks. As the toolmark to be analyzed becomes more and more complex, it is becoming increasingly likely that development of a truly robust objective algorithm will involve moving from a linear pixel-to-pixel comparison of the data to one that involves an area comparison. While this would represent a major shift as it relates to the operation of the current algorithm, exploratory efforts using this line of approach are already underway by other research groups (11).

Summary and Conclusions

This study was completed using 1000 samples of copper wire shear cut into two pieces using 50 sequentially manufactured pliers. The resultant toolmark on the shear cut surfaces was quasi-striated in nature, consisting of groups of striations. Pairs of shear cut surfaces were objectively compared utilizing a statistical algorithm that had been previously successful in comparing regularly striated marks. The algorithm was optimized and applied using a leash option to the search and validation windows to prevent incorrect identification related to matching at opposite ends of the comparison pairs from occurring. This resulted in a noticeable improvement in the analysis. Known matching pairs had large T1 values for the majority of comparisons (indicating a match) and known nonmatching pairs had near zero values for the majority of comparisons (indicating a nonmatch). A high degree of separation in the data was observed although sufficient statistical separation was not achieved. While the results have improved, more work is needed to increase the robustness of the identification process. Future improvements to the analysis method may involve automated means to examine combinations of search and validation windows quickly or, more radically, changing the process to compare areas rather than single linear files.

Acknowledgments

The research group thank Adam Froeschl and Wilde Tool Co., Inc. for their aid in gathering sequentially manufactured slip-joint pliers. Their effort helped make this research possible. Also thank Dr. Max Morris for the critical review of this study and technical guidance.

References

1. Faden D, Kidd J, Craft J, Chumbley LS, Morris M, Genalo L, et al. Statistical confirmation of empirical observations concerning toolmark striae. AFTE J 2007;39(3):205–14.
2. Chumbley LS, Morris M, Kreiser J, Fisher C, Craft J, Genalo L, et al. Validation of toolmark comparisons obtained using a quantitative, comparative, statistical algorithm. J Forensic Sci 2010;55(4):953–61.
3. Petraco N, Chan H, De Forest P, Crim D, Diaczuk P, Gambino C, et al. Application of machine learning to toolmarks: statistically based methods for impression pattern comparisons. Rockville, MD: National Criminal Justice Reference Service, 2012 Jul; Document No.: 239048.
4. Bachrach B, Jain A, Jung S, Koons R. A statistical validation of the individuality and repeatability of striated toolmarks: screwdrivers and tongue and groove pliers. J Forensic Sci 2010;55(2):348–57.
5. Cassidy FH. Examination of toolmarks from sequentially manufactured tongue-and-groove pliers. J Forensic Sci 1980;25(4):796–809.
6. Grieve T. Objective analysis of toolmarks in forensics [master's thesis]. Ames, IA: Iowa State University, 2013.
7. Miller J. An introduction to the forensic examination of toolmarks. AFTE J 2001;33(3):233–48.
8. Monturo C. The effect of the machining process as it relates to toolmarks on surfaces. AFTE J 2010;42(3):264–6.
9. AFTE. Glossary of the association of firearm and toolmark examiners. 5th edn. Montreal, QC: Forensic Technology, Incorporated, 2007;182. CD.
10. Zhang S, Ekstrand L, Grieve T, Eisenmann DJ, Chumbley LS. 3D data processing with advanced computer graphics tools. In: Interferometry XVI: techniques and analysis proceedings of SPIE 8493. Bellingham, WA: SPIE, 2012;849310-849310-7.
11. Song John. Proposed "NIST Ballistics Identification System (NBIS)" based on 3D topography measurements on correlation cells. AFTE J 2013;45(2):184–94.

Additional information and reprint requests:
Ryan Spotts, B.S.
Ames Laboratory
Iowa State University
2220 Hoover
Ames
IA 50011
E-mail: rspotts@iastate.edu