JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2012, Vol. 57, No. 4
doi: 10.1111/j.1556-4029.2012.02115.x
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Nicholas D. K. Petraco,[1,2] Ph.D.; Peter Shenkin,[3] Ph.D.; Jacqueline Speir,[4] M.S.; Peter Diaczuk,[1,5] B.S.; Peter A. Pizzola,[6] Ph.D.; Carol Gambino,[7] M.S.; and Nicholas Petraco,[1,8,9] M.S.

Addressing the National Academy of Sciences' Challenge: A Method for Statistical Pattern Comparison of Striated Tool Marks

ABSTRACT: In February 2009, the National Academy of Sciences published a report entitled "Strengthening Forensic Science in the United States: A Path Forward." The report notes research studies must be performed to "…understand the reliability and repeatability…" of comparison methods commonly used in forensic science. Numerical classification methods have the ability to assign objective quantitative measures to these words. In this study, reproducible sets of ideal striation patterns were made with nine slotted screwdrivers, encoded into high-dimensional feature vectors, and subjected to multiple statistical pattern recognition methods. The specific methods employed were chosen because of their long peer-reviewed track records, widespread successful use for both industry and academic applications, rely on few assumptions on the data's underlying distribution, can be accompanied by standard confidence levels, and are falsifiable. For PLS-DA, correct classification rates of 97% or higher were achieved by retaining only eight dimensions (8D) of data. PCA-SVM required even fewer dimensions, 4D, for the same level of performance. Finally, for the first time in forensic science, it is shown how to use conformal prediction theory to compute identifications of striation patterns at a given level of confidence.

KEYWORDS: forensic science, *Daubert*, *Frye*, National Academy of Sciences, tool marks, multivariate, pattern recognition, error rates

Over the last two decades, DNA profiling has become one of the most widely applied techniques for the identification of biological samples in forensic science. This is largely because of the volumes of allele frequency data and the clear applicability of simple statistical methods. The unparalleled success of DNA profiling is likely responsible for the recent National Academy of Sciences' report on the raising of standards for scientific examination of all forms of physical evidence (e.g., tool marks, soils, dust, questioned documents, shoe prints, fire debris, fingerprints, gunshot residue, tire tracks, hairs, and fibers) (1). Tool mark impression evidence, for example, has been successfully used in courts for decades, but its examination has lacked scientific, statistical proof that would independently corroborate conclusions based on morphology characteristics (2–7). In our study, we will apply methods of statistical pattern recognition (i.e., machine learning) to the analysis of tool mark impressions.

The process of associating tool mark evidence to a specific tool involves classifying the object into groups of similar objects. To conclude that two striation patterns have a high likelihood of being associated, a tool mark examiner usually states that the "quantity and quality" of the markings were such as to allow them to properly draw their conclusion.

The Association of Firearms and Toolmark Examiners, the *de facto* group that set the standards for firearm and tool mark examination, states the theory of identification as it relates to tool marks as (2,8):

> "A. The theory of identification as it pertains to the comparison of tool marks enables opinions of common origin to be made when unique surface contours of two tool marks are in sufficient agreement."
> "B. This sufficient agreement is related to significant duplication of random toolmarks by correspondence of pattern or combination of patterns of surface contours."

"a. Significance is determined by comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows."

Impressions and striations made by tools and firearms can be viewed as mathematical patterns composed of peaks, ridges, and furrows, which we will refer to as features. Numerical classification methods (9–11) are of particular interest because they have the potential of assigning objective quantitative measures to the words "sufficient agreement" and "comparative examination." Unfortunately, only a few numerically based studies for tool mark and firearm

[1]Department of Sciences, John Jay College of Criminal Justice, City University of New York, 899 10th Avenue, New York, NY 10019.
[2]Faculty of Chemistry, Graduate Center, City University of New York, 365 5th Avenue, New York, NY 10016.
[3]Department of Mathematics and Computer Science, John Jay College of Criminal Justice, City University of New York, 899 10th Avenue, New York, NY 10019.
[4]Department of Chemical and Physical Sciences, Forensic Science Program, Cedar Crest College, 100 College Drive, Allentown, PA 18104.
[5]PEDICO Research, RR2 Box 62, Waymart, PA 18472.
[6]Special Investigations Unit, 421 East 26th Street, New York, NY 10016.
[7]Department of Sciences, Borough of Manhattan Community College, City University of New York, 199 Chambers Street, New York, NY 10007.
[8]Petraco Forensic Consulting, 73 Ireland Place, Amityville, NY 11701.
[9]New York City Police Department Crime Laboratory, 150-14 Jamaica Avenue, Jamaica, NY 11432.

Received 25 Dec. 2009; and in revised form 25 April 2011; accepted 4 June 2011.

impression/striation comparisons (mostly firearms) have appeared in the literature (12–41). Nichols's reviews give excellent background and insightful overviews for most of these studies (42–44).

In 1959, Biasotti recognized the need for empirically based statistical studies on tool mark patterns (20). His study recorded the number of matching "striations" between bullets fired from the same gun and bullets fired from different guns. He parameterized his model (now known as consecutively matching striae or the CMS model) in terms of groups of matching "lines." Biasotti (20) defined striation as "... engravings or striations appearing on the bullet as a result of being engraved by the individual irregularities on characteristics of the barrel plus any foreign material present in the barrel capable of engraving the bullet." He identified groups of matching lines between bullets in phase; matching occurred "…when the lines appear to be similar in contour and of common origin." Building on Biasotti's work, Neel and Wells (19) have recently published a large-scale statistical study, where they examined 4188 striated tool mark comparisons for various sized runs of consecutive matching striations. They found very low empirical frequencies for CMS runs larger than 4× if two different tools were known to have generated the striation patterns.

One of the first efforts in the computational comparison of striated firearm data was proposed by Gardner in the late 1970s (13). His system attempted to use the same striation data a ballistics examiner would use and echoed much of the same ideas that were implemented two decades later in the Integrated Ballistics Identification System (IBIS) system marketed by Forensic Technology Inc. Using an SEM, Gardner obtained multiple scans of land-and-groove areas from .38 caliber bullets fired by four different revolvers. Striations were quantified by comparing derivative peak heights/depths to the derivative mean line. Gardner then employed a heuristic probability-based formulation to generate similarity scores between all possible comparisons of engraved areas (land-to-land and grove-to-grove). Reasonable, correct classification rates were obtained on the small test set used—two bullets fired from each of three different guns.

Databases for tools and tool marks mostly came into existence in the early 1990s (32,45–59). The IBIS system is probably the most widely used database system for semiautomated firearm identification and functions like a computer-aided comparison microscope (47). The version in widest use employs 2D gray scale images of various striation and impression patterns on bullets and casings and converts them into a "signature." This signature is compared against a large database of signatures collected by local, state, and national law enforcement agencies. The image of a pattern from questioned firearms evidence is "scored" by a system "correlation" server for "similarity" against entries in the database. The reason the words "signature," "scored," "correlation," and "similarity" appear in quotes is because Forensic Technology Inc. does not scientifically define these very important technical terms and because the comparison algorithms are considered trade secrets of a private company. In 2005, a committee of the National Research Council was assembled to assess the feasibility, accuracy, and technical capability of a national ballistics database. The committee noted that IBIS is a tool for search, not verification (48). If a tool mark comparison system is to evolve to meet the needs of forensic science and be able to stand up to the *Daubert* challenge, the inner workings absolutely must be public knowledge.

With the advent of confocal microscopy and laser scanners, the acquisition of the entire 3D surface of a tool mark can be obtained. Very recently, an excellent study by Bachrach et al. (52) appeared where confocal microscopy was used to digitally record the surfaces of striated tool marks made by screwdrivers and tongue and groove pliers. The surfaces were then filtered and averaged to form a surface "signature." Similarity scores for all possible pairs of signatures were generated based on the cross-correlation function and used to produce matching and nonmatching distributions (histograms). Algorithm-generated identifications were found to be highly reliable so long as the screwdrivers' angles of attack were consistent (angle of attack was obviously not an issue for tongue and groove pliers). A National Institute of Standards and Technology (NIST) study on an automated bullet signature identification system has also recently been published (53). The NIST scheme used the same principles as the previous study to establish likely bullet–gun associations. This new system is reported to have a 10% increase in accuracy over current commercially available systems which is a marked increase.

Howitt et al. (60) recently published a model for the computation of correspondence probabilities (i.e., "matching") between striation patterns imparted on bullets fired by the same gun versus bullets fired from different guns. Again, the authors use the same functional definition of a "line" in a striation pattern as was given by Biasotti. This definition makes their theory very general and equally applicable to striation patterns imparted by actual tools, and striation patterns found on firearms evidence. Their theory can also take into account arbitrary magnification levels (which would be a parameter of a comparison microscope), and the number of lines found in a striation pattern. Output of the approach is a probability that various CMS runs on striation patterns generated by different sources would match purely by chance. The entire model rests on the assumption that the possible patterns, which the lines can form, are probabilistically independent of each other and are identically distributed. Their study shows some evidence for this assumption. The Howitt–Tulleners computed probability of random correspondence between 2× CMS runs on bullets (called doublets in the study) from different sources is between 0.1 and 0.16 at 20 mm resolution (i.e., at 10–16% chance) and 0.14–0.24 at 30 mm resolution. Biasotti's empirically derived probabilities for the same situation are 0.2–0.46 depending on whether or not the bullet is jacketed. Results, however, are in less agreement if 3× CMS runs (called triplets in the study) are considered (20,60). The Howitt–Tulleners computed probability of random correspondence for this situation is 0.003–0.005 at 20 mm resolution and 0.007–0.01 at 30 mm resolution. Biasotti's empirically derived probabilities, however, are 0.01–0.1 depending on jacketing (20,60).

Despite these studies, there are no standard methods for the application of probability and statistics to the analysis of tool mark evidence. Our work is intended to help in the establishment of standard protocols. We believe that when applying statistical pattern recognition methods to legal issues, they should rely on as few underlying assumptions about the data (i.e., the evidence) as possible. For these reasons, we wanted to choose methods which: (i) make few assumptions about the form of underlying statistical distributions of the data, (ii) function adequately with small or limited data sets, and (iii) have been shown to work well on real-world problems (e.g., in industry and medicine). Unlike past studies, we take an entirely multivariate approach to the statistical discrimination of striated tool marks. This approach has been successfully exploited numerous times in the forensic science literature (61–68). We utilize two of the most successful multivariate pattern recognition methods: partial least squares discriminant analysis (PLS-DA) and support vector machines (SVM). Both make no assumptions about underlying probability densities, are designed to work well with small sample sizes, and enjoy an extensive record of performance in the literature (69,70). Standard confidence intervals (at the 95% level) for the classification algorithms' output (tool mark I.D.s) are computed from conformal prediction theory (CPT) and reported as well. This study focuses on striated tool marks made by screwdrivers, which can be applied in practice quickly. It is expected that the methodology developed in this study will translate to any kind of striated tool mark or firearm evidence.

Methodology

Nine identical high-quality Craftsman® (Sears Holding Corp., Hoffman Estates, IL) quarter-inch slotted screwdrivers were purchased at a local hardware store. The screwdrivers, shown in Fig. 1, were brand new and came in packages of three. The screwdriver shafts were manufactured by drop forging blanks followed by grinding off of resulting flashing. Subclass characteristics on the tool's working surface are a possibility with some manufacturing methods, but grinding is not one of them. In general, statistical pattern comparison algorithms would detect both similarities (possible subclass characteristics) and dissimilarities (individual characteristics) between sets of features for many different tool marks. However, even though some of the features making up a tool's working surface may be subclass characteristics, it is unlikely that the majority of the multitudes of features used by a statistical pattern recognition scheme are subclass in nature. If subclass characteristics were a significant issue for a set of tools, the problem would manifest itself in a study of this design by producing relatively high error rates in the testing phase of the algorithm. We, in fact, did not encounter this problem. However, note that any statistical pattern recognition scheme should be thoroughly tested before being put into practice. With proper validation, it is unlikely that subclass characteristics would cause major problems with an algorithmic discrimination process and go unnoticed.

The screwdrivers were used to generate multiple reproducible striation pattern standards. For this pilot study, the striation patterns were made by hand while holding the tip of the screwdriver parallel to the striation medium surface. This was performed to introduce some "casework" realism into patterns to test how well the classification algorithms hold up to the variability expected from the pattern reproduction process. A jig to hold the screwdriver could also be used to make even more uniform patterns if desired.

Number 2 Roma Plastina modeling clay (http://www.Sculpture-House.com, last accessed April 2010) was used as impression medium. Figure 2 illustrates the process. Further details of striation pattern generation are available in reference (71). The striation patterns made by each of the nine screwdrivers were digitally photographed with a Nikon DSFi1 digital camera under a Nikon SMZ1000 stereo-microscope at 20× magnification (Nikon Corporation, Tokyo Japan). The positions and widths of a small number of striation lines/grooves were measured with a stage micrometer. The measurements were later used to calibrate the digital image processing program ImageJ (72). Striation patterns were generated for only one side of each screwdriver's working surface. A total of 75 actual patterns were recorded. Using these real striation patterns, 732 more striation patterns were simulated for testing purposes. See the statistical methods section for simulation details.

FIG. 1—*Nine quarter-inch standard slotted screwdrivers used in this study.*

Striation Pattern Quantification

The surface topography of a striated tool mark obviously contains a tremendous amount of physical information. Fortunately, the human mind, in particular the professional tool mark examiner's mind, can efficiently handle such a voluminous amount of information. Unfortunately, computers and mathematics are not as efficient as the human brain and they require "models" of physical reality to execute their orders in a reasonable amount of time.

For the striation patterns examined in this study, the positions of striation lines/groves measured with the stage micrometer were used to calibrate the image processing program ImageJ (72). Distances of each line or grove from the left edge of each striation pattern were measured to the nearest 0.05 mm. Each striation pattern is no more than 7 mm (*c.* 0.25 in) wide. For each pattern, a list of 140 pieces of information (7 mm/0.05 mm slots) is created. Each piece of information is a 1 or 0, that is, a "bit." The list thus consists of "slots" for information and is superimposed over each striation pattern. A 1 is recorded in a slot of the list if a line or grove is present or spans the slot. A 0 is recorded otherwise. The procedure yields a 140-dimensional (140D) binary feature vector for each pattern, which is reminiscent of a "bar-code" (cf., Fig. 3). In this study, it was found that of the 140 components in the feature vector, 19 always had value 0 across all recorded striation patterns. These nonvarying components were excluded and thus feature vectors of 121D (140D−19D = 121D) were used in the statistical analyses described later.

It should be noted that from the stack of striation patterns in Fig. 2 (rightmost picture in the figure), some of the patterns required alignment (registration) because a striation pattern was not entirely complete. Figure 2 (rightmost picture in figure) depicts nine registered (aligned) striation impression patterns collected using a single screwdriver. Registration was performed by aligning the left edge and/or an obvious groove shared between patterns generated by a single tool. As can be seen from Fig. 2, exemplar impressions generated by a single tool contain inherent variability. Although this can be controlled to some degree during exemplar pattern reproduction in a controlled laboratory setting, it must be considered an unavoidable consequence of patterns generated during the commission of a crime, and therefore should be represented in an exemplar database. All data measurements made with ImageJ were preprocessed and registered using a program written in Mathematica 7 (73). Our Mathematica notebooks are available upon request. A more automated registration algorithm is in development in our laboratory.

Statistical Methods

The binary feature vectors generated from the striation patterns were arranged into an $n \times p$ data matrix ($\mathbf{X}$) for analysis:

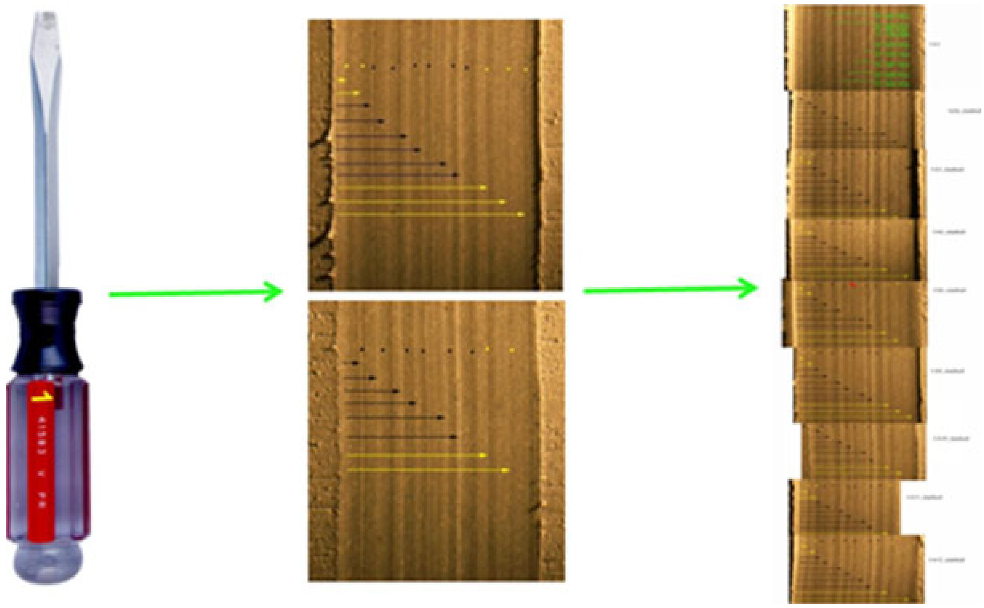

FIG. 2—*Process to generate striation pattern standards. As one can see from the figure, screwdriver number 1 was used to make nine striation pattern standards.*

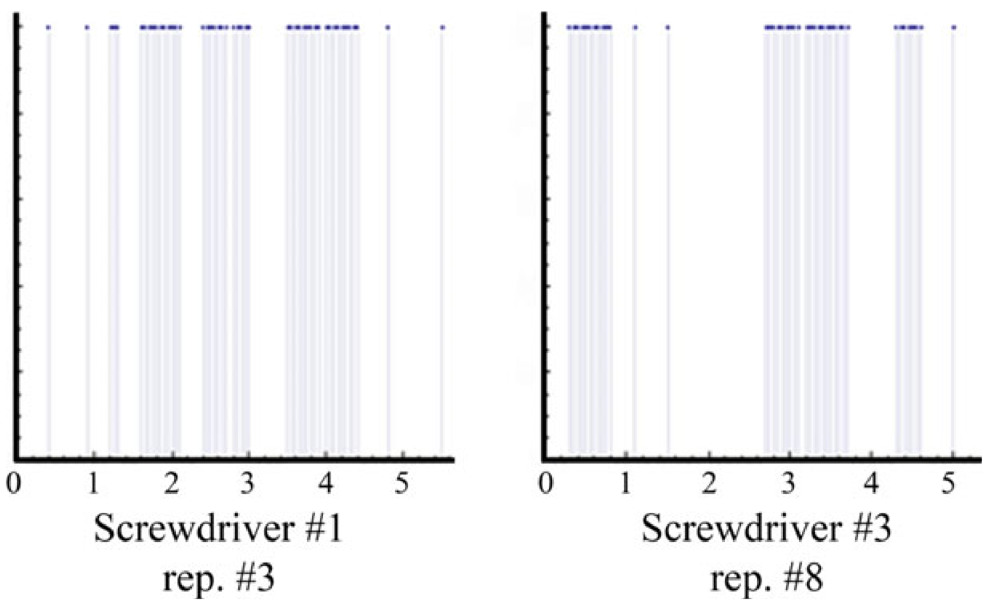

FIG. 3—*Graphical representations of feature vectors for two striation patterns. They are replicate pattern 3 for screwdriver #1 and replicate pattern 8 for screwdriver #3.*

$$\mathbf{X} = \begin{bmatrix} X_{11} & .. & X_{1j} & .. & X_{1p} \\ \vdots & & \vdots & & \vdots \\ X_{i1} & .. & X_{ij} & .. & X_{ip} \\ \vdots & & \vdots & & \vdots \\ X_{n1} & .. & X_{nj} & .. & X_{np} \end{bmatrix}$$

where $n = 75$ is the number of striation patterns and $p = 121$ is the number of components in each feature vector. For each X_{ij}

$$X_{ij} = \begin{cases} 1 \text{ if a line/groove falls in slot } j \text{ of striation pattern } i \\ 0 \text{ if a line/groove does not fall in slot } j \text{ of striation pattern } i \end{cases}$$

The symbol $\mathbf{X}_i$ designates row i of $\mathbf{X}$ and is a vector of data representing striation pattern i. Multivariate statistical methods were used to transform the data set ($\mathbf{X}$) into a new data set ($\mathbf{Z}$). These transformed variables may be used in classification algorithms (supervised or unsupervised) to discriminate between different screwdrivers. In this study, $k = 9$ different screwdrivers of the same brand were used. The multivariate transformation analyses of data set ($\mathbf{X}$) undertaken in this study were PLS-DA and principal component analysis (PCA). The multivariate statistical discrimination methods used in this study were PLS-DA (i.e., it is both a transformation and a discrimination method), one-versus-one multiclass SVM, and CPT utilizing PCA-SVM and 3-nearest neighbor (3-NN) classifiers. Further details on the properties of PCA- and PLS-derived variables as well as the above-stated discrimination methods are available in references (10,70,74,75). We note, however, that none of the methods used in this study depend on the data being Gaussian.

The raw data were mean-centered and variance scaled before all transformation/discrimination computations. All statistical computations were performed with the statistical software R (76).

Principal Component Analysis

PCA is a multivariate procedure that is used to reduce the dimensionality of a data set (**X**) to a new data set ($\mathbf{Z}_{PC}$) of "derived variables," which account for successively decreasing amounts of variance (74,77–79). The variance order of the variables in **Z** provides guidance for the reduction in the data's dimensionality while retaining an adequate representation. See references (74) and (79) for a fuller account of the details of PCA. The native R PCA program prcomp was used in this study (76).

Support Vector Machines

Small sample sizes are inevitable for many statistical studies of tool marks. Using statistical learning theory and its practical application, the SVM was developed in response to the need for reliable statistical discriminations within small sample studies (75). SVMs seek to determine efficient decision rules in the absence of any knowledge of probability densities for the data by determining maximum margins of separation (cf., Fig. 4) (69,75). This procedure produces an algorithm, which determines linear decision rules with (typically) large margins for error. In this study, a linear kernel was used with the SVM algorithm. Also, the penalty parameter, C, was found to be 10 using a standard line search (80).

The SVM methodology was originally designed to separate two groups, but it does have several incarnations which can handle multigroup (multicategory) problems. The most popular, because it works well in practice, is to consider all possible pairs of groups of striation patterns and determine a two-group (binary) SVM for each pair. Thus, if there are k samples of striation patterns generated from k screwdrivers, $k(k–1)/2$ binary SVMs are computed and group identity of a pattern is determined by voting of the decision rules (81). This multicategory approach to SVMs is called the one-versus-one method and is what we use in this study. The R package kernlab was used to perform all SVM computations (82).

Partial Least Squares Discriminant Analysis

Multivariate linear regression is a method to find linear relationships between dependent "response" variables and a data matrix of "predictor" variables **X** by exploiting the covariance between these entities. If the response variables are coded as class labels, the method can be used for supervised classification (i.e., discriminant analysis) (83). PLS is a method to determine the relationship between predictor and response variables when there are many more predictors than samples and/or when many of the predictors are correlated. The method determines the PLS "latent vectors" (LVs) analogous to principal components (PCs) from PCA. Also, just as in PCA, the first two or three LVs can be use to make an approximate graphical representation of the data. In this study we used the "softmax" function to map output of the PLS procedure to class assignments (i.e., striation pattern–screwdriver identifications) (84). The R packages pls and caret were used to perform PLS-DA computations (80,84). For more of the computational details of PLS see references (70,84).

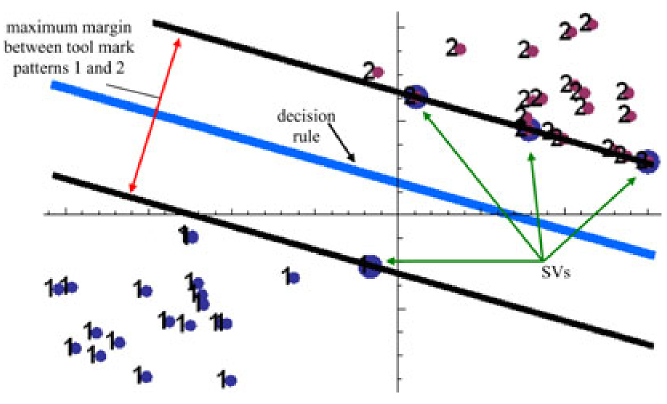

FIG. 4—*Graphical representation of support vector machine output for the separation of two data sets.*

Error Rate Analysis–Hold-One-Out Cross-Validation

An error for the computational learning methods used in this study is defined as a misidentification of a striation pattern. There are many methods that can be used to assess error rates for pattern comparisons. All methods produce estimates of error rates based on samples. The estimation is of "global" error rate, or how often an algorithm would make a misidentification on a population of striation patterns it was not trained on. The simplest method to empirically estimate the error rate is resubstitution (85). This is the application of the computed classification rules to the set of data used to derive them. The percentage of misclassifications via the resubstitution method is called the apparent error rate and is simply the empirical risk, R_{emp}. This is a biased estimate and tends to be overly optimistic and should be corrected. One option to improve the estimate of error rate is to use hold-one-out cross-validation (HOO-CV) (78). This method computes the decision rules using all but one of the tool mark patterns in the data set. The hold-one-out procedure is repeated for each tool mark pattern in the data set, and the results are averaged to compute an estimated error rate (86).

Error Rate Analysis–Bootstrapping

Another improved error rate estimate is the refined bootstrap (86,87). First B sets of bootstrap data, $\mathbf{X}^*$ are generated by randomly selecting (with replacement) n striation pattern feature vectors from the original data set **X**. It should be noted that each bootstrap data set contains the same number of elements (striation pattern feature vectors) as the original data set, thus some patterns may be repeated. The decision rules, g^*, are recomputed for each bootstrap sample and an average error rate is computed using them on the original data as well as the bootstrapped data used to compute them (86). An alternative to the refined bootstrap is the .632 bootstrap error rate estimate (87). The prediction error is more likely to be larger for test patterns not contained in a given bootstrap sample. Thus to give a more conservative error rate estimate, the .632 bootstrap focuses on these larger errors. All cross-validation and bootstrap error rate estimates were computed either with native functions in the kernlab and pls packages or with the ported R package, bootstrap (87).

Error Rate Analysis–Conformal Prediction and Confidence Regions

Solomonoff's and Kolmogorov's algorithmic theory of randomness is a mathematically sophisticated way to gauge the amount of true information in a string of symbols (88). Recently, a method that gives confidence levels to the identification of unknown patterns and control over error rates has arisen from the study of algorithmic randomness (89). This method, called conformal prediction, can be applied to any statistical pattern comparison algorithm and holds a great deal of potential if applied to tool mark analysis. Prediction regions (confidence intervals) produced by conformal prediction can give a judge or jury an easy to understand measure of reliability for tool mark pattern identification because the method yields confidence on a scale of 0–100%.

The way the method works is actually very simple (89,90). Given a training set of striation patterns with known identities (called a bag) and at least one striation pattern of unknown identity, an estimate of randomness is computed for the bags containing the unknown striation pattern with all possible labels for its identity. The only assumption is that the striation patterns of the training set are drawn independently from the same, but unknown probability distribution.

Randomness of the bag is tested in a way analogous to what is performed in traditional hypothesis testing (89,91,92). The null hypothesis is that unknown striation pattern **x** with assigned identity label y [i.e., the pair (x,y)] belongs to the bag and does not significantly decrease the bag's randomness. The alternative hypothesis is that the pair (x,y) does not belong to the bag and thus y must be a different label than the one assigned. p-Values are computed for randomness estimates. Thus, conformal prediction regions for tool mark pattern identities can be thought of as generalizations of confidence intervals known from textbook hypothesis testing. Traditionally, confidence intervals are computed for population parameters (e.g., a sample average) to give an indication of the regions where their true values may fall. Technically, the Neyman–Pearson interpretation of $(1-\alpha)\times100\%$ confidence interval for an estimated population parameter (here, striation pattern identities) constructed from a random sample of a given sample size will contain the true population parameter $(1-\alpha)\times100\%$ of the time (90–93). The value α is called the level of significance and is the probability that any given confidence interval constructed from a random sample will *not* contain the true population parameter.

It should be noted that the null hypothesis can be accepted for multiple label prediction regions of the striation pattern's identity. In such cases, the identity assignment (i.e., the prediction region) at the $(1-\alpha)\times100\%$ confidence level is ambiguous. While multilabel output is not wholly uninformative, ideally, the prediction region will contain only one label with a p-value ≤0.05. This means that the conformal prediction algorithm has produced a prediction region with only one label and a confidence level of at least 95%.

p-Values for striation pattern test identities (I.D.s) are found by computing nonconformity scores. The nonconformity score, α_i, for the ith striation pattern using one-versus-one multiclass SVMs was computed as

$$\alpha_i = \frac{1}{k-1}\sum_{j=1}^{k(k-1)/2}\lambda_{i,j}$$

where $\lambda_{i,j}$ is a matrix element of an n row by $k(k–1)/2$ column matrix of Lagrange multipliers (89). The formula just sums all the columns in this matrix and weights the resulting n-dimensional vector by $1/(k–1)$. The Lagrange multipliers were all computed by the binary SVM function of kernlab (82). Because a suitable nonconformity measure has not yet been developed for PLS, we did not use this method with CPT. Instead 3-NN classification, which has been shown to perform well in a number of machine learning tasks, was used (10,89,94,95). A 3-NN nonconformity score for striation pattern i is computed by first finding the distance matrix between all striation patterns and selecting three closest to i with the same I.D. and the three closest to i with a different I.D. The actual score is then computed as ratio

$$\alpha_i = \frac{\sum_{j=1}^{3}\mathrm{dist}(i,j)^{\text{same I.D.}}}{\sum_{j'=1}^{3}\mathrm{dist}(i,j')^{\text{different I.D.}}}$$

A p-value for each possible labeling $\mathrm{tlab}_i \in \{1,2,...,k\}$ of a test striation pattern is computed as

$$p_{\mathrm{tlab}_i} = \frac{\#\{j\in\{1,2,...,n\}:\alpha_j^{\mathrm{tlab}_i}\geq\alpha_{\text{test}-\text{pattern}}^{\mathrm{tlab}_i}\}}{n}$$

where $a_j^{\mathrm{tlab}i}$ is the nonconformity score of the jth pattern when the test pattern is labeled as screwdriver tlab_i, $a_{\text{test}-\text{pattern}}^{\mathrm{tlab}i}$ is nonconformity score of the test pattern labeled as screwdriver tlab_i, and $p_{\mathrm{tlab}i}$ means the p-value of the test pattern labeled as screwdriver tlab_i (89). A set of k p-values is computed for each test pattern, one for each possible screwdriver identity. For a chosen significance level ε (i.e., level of confidence $1-\varepsilon$), a $1-\varepsilon$ confidence region of labels is determined by selecting those labels of the test pattern with p-values $\geq \varepsilon$. Ideally, the output confidence region contains only one label. If a multilabel confidence region is output, it counts as a correct I.D. if it contains the true label of the striation pattern, although obviously it is less informative. Empty regions can also be output if the CPT algorithm cannot confidently identify the striation pattern. Empties automatically count as errors (89).

Results and Discussion

Partial Least Squares Discriminant Analysis

Figure 5 shows all 75 striation patterns on the basis of the first two LVs. The advantage of a 2D or 3D plot of the data is to give some insight into the overall data structure, which can be instructive for judges and juries. One must also bear in mind that a good deal of information may have to be discarded to make such plots. This is the case for Fig. 5. The first two LVs only account for 33% of the data's overall variance. It should be noted, however, that a good deal of discrimination (large intergroup separation, small intragroup separation) is already apparent between the groups of striation patterns.

Examining Fig. 5 further, one can see that two screwdriver #3 patterns (at close to coordinate [−4, −4.5]) are clearly separated from the other screwdriver #3 patterns. They appear as outliers in this 2D PLS space and it is reasonable to assume that almost any statistical discrimination algorithm will mistakenly identify them as screwdriver #7 patterns. Their departure from the other screwdriver #3 patterns can be explained by examining the actual physical striation patterns, which are labeled patterns 6 and 8 in Fig. 6. The rightmost groove imparted by screwdriver #3 is highlighted in the figure. It spans from where the right edge of the screwdriver made contact with the clay, and across to the left about 1.2 mm (cf., Fig. 6, pattern 3). Replicate patterns 6 and 8 are clearly aligned with replicate pattern number 7 in Fig. 6 (note all the common lines vertically down the patterns). Note how the remaining 0.7 mm of the grove in question is missing in patterns 6 and 8, likely caused by striking the clay surface at a slight angle. This

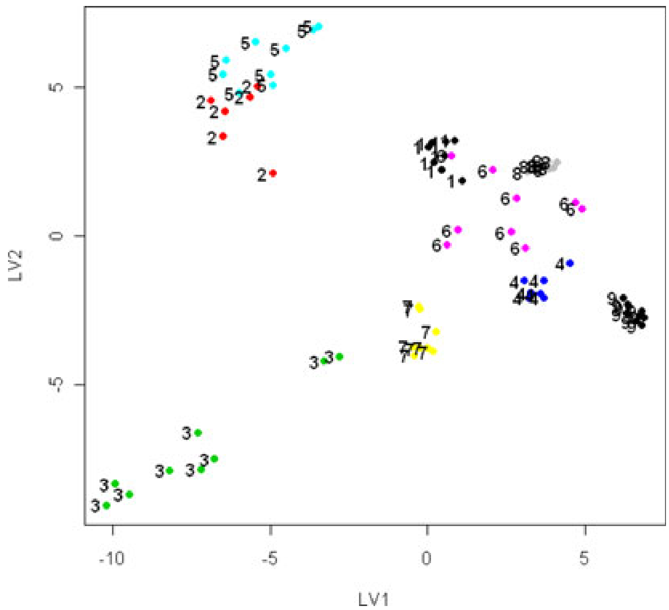

FIG. 5—*All striation patterns projected into the space of the first two PLS latent vectors (33% variance retained). Each point represents a striation pattern. The boldfaced numbers to the left of the points tell which screwdriver generated the pattern.*

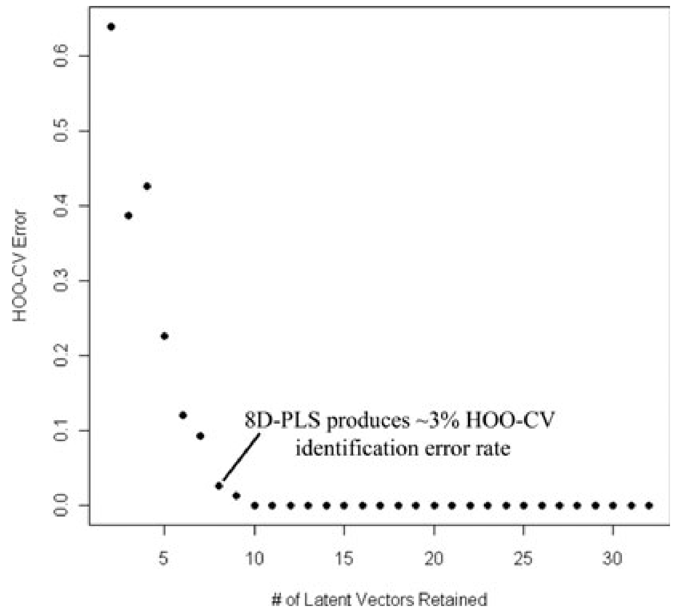

FIG. 7—*Hold-one-out cross-validation (HOO-CV) error rates versus PLS dimension (i.e., number of latent vectors retained). As expected, classification error generally decreases as dimension of the space is increased.*

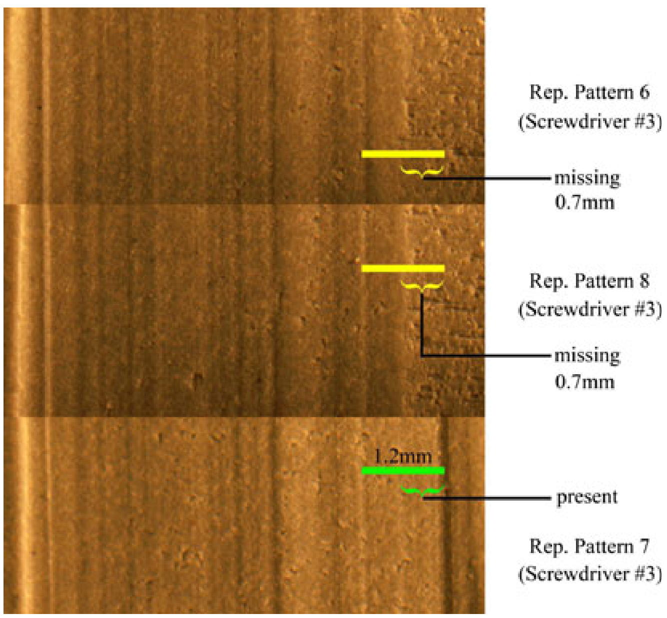

FIG. 6—*Three screwdriver #3 patterns with the rightmost groove highlighted. The first two, pattern 6 and pattern 8, appear as outliers in this 2D PLS space.*

TABLE 1—*PLS correct classification rate estimates.*

	Data Dimension*			
	2D (%)	3D (%)	8D (%)	38D (%)
Method of estimation†				
Apparent	48	68	100	100
Hold-one-out CV	36	61	97	100
Refined bootstrap	78	78	98	99
.632 bootstrap	37	53	98	99
Random test set‡	47	71	100	100

*Data dimension is the number of latent vectors retained to represent the data.

†The method's estimated error rate = (100 − correct classification rate estimate)%.

‡Preprocessing transformation computed using corresponding random training set.

grove is apparent on all of the striation patterns generated by screwdriver #3 with the exception of patterns 6 and 8 (photographs available on request) and its absence in these patterns is the cause of their distance from the other screwdriver #3 patterns in Fig. 5.

Such a phenomenon may be of concern to the reader because obviously a perpetrator of a crime will not take the time to make sure all the striation patterns they leave are made at the same angle! This was not a problem for the discrimination methods used in this study. With the addition of a few more dimensions to the data, these patterns were easily classified as stemming from screwdriver #3 (i.e., the classification holds up to cross-validation and bootstrapping.). Figure 7 shows that by retaining 8D PLS space (72% variance retained), the HOO-CV error rate can be dramatically reduced. Thus, 8D PLS space has more than enough discrimination power to identify most of the striation patterns in this data set. Table 1 shows the HOO-CV, refined bootstrap and .632 bootstrap methods all estimate the global correct identification rate at 97–98%. A 100% correct classification rate was achieved on a randomly generated test set of consisting of 25% of the data. Scaling up to 38D PLS space (95% variance retained), the estimated global correct identification rates are all nearly perfect. It should be noted that this is still a relatively low-dimensional model considering that the raw data set is 121D.

Principal Component Analysis with Support Vector Machines

Figure 8 shows the data in the basis of the first two PCs. There is a strong resemblance to the 2D PLS plot (Fig. 5), which is typical. The advantage of PCA, however, is that it reduces the dimension of the data set in a completely unsupervised way. Thus, if natural clustering between groups of data exists in 2D or 3D PCA space, there is strong evidence that the differences are real. PCA, however, requires that a separate classification method be used to numerically identify to which group a data point belongs. While PLS is a totally supervised method, it is its own classification method. The classification method we chose for PCA in this study is SVMs. An SVM is a supervised identification method; however, the procedure assumes very little about the distribution of the data. This is quite unlike most other classification methods in machine learning; the dropping of a distributional assumption is in the authors' minds, a major advantage in forensic applications.

Figure 9 shows a plot of HOO-CV classification error rate with increasing dimension of PC space. To make this figure, a test pattern is first held out, PCA followed by SVM is performed on the remaining data, and then the held out pattern is projected and classified. This retains some noise in the test pattern to stress the method and thus yields a higher HOO-CV error rate estimate. It is also computationally more expensive.

First note that the error rates are lower than for PLS-DA. An HOO-CV error rate of *c.* 3% is achieved with 4D for PCA (52% variance retained) as opposed to 8D for PLS. Table 2 shows the estimates of apparent and global correct identification rates using PCA-SVM. For all of the global correct classification rate procedures used to construct this table, PCA is performed first on the entire data set and then hold-out or bootstrapping is applied (This is how it would be performed in practice once an identification model is fit). Thus, some variance in the test observations is initially removed, which is likely why the 4D HOO-CV error is slightly lower in Table 2 compared to that seen in Fig. 9.

For reference, Fig. 10 shows the data on the basis of the first three PCs. Even though only 33% of the variance is retained with three PCs, the estimated correct classification rates are very good. By 4D, they are nearly perfect. For reference, the 31D (95% variance retained) estimated global correct classification rates are

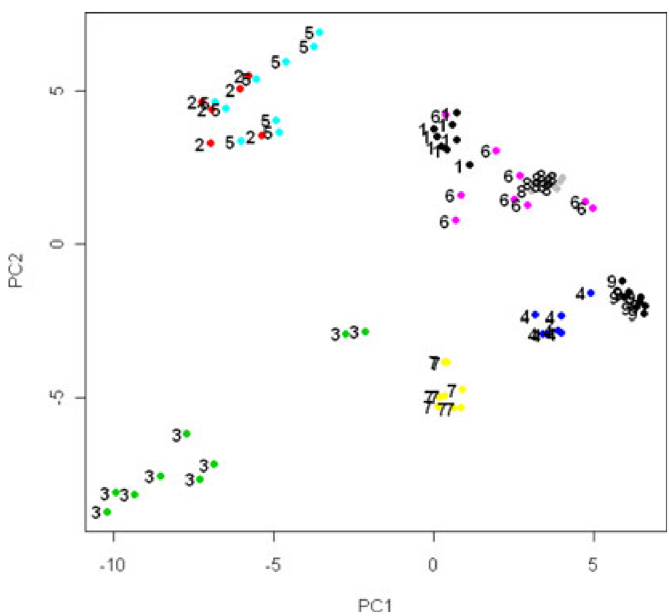

FIG. 8—*All striation patterns projected into the space of the first two principal components (33% variance retained). Each point represents a striation pattern. The boldfaced numbers to the left of the points tell which screwdriver generated the pattern.*

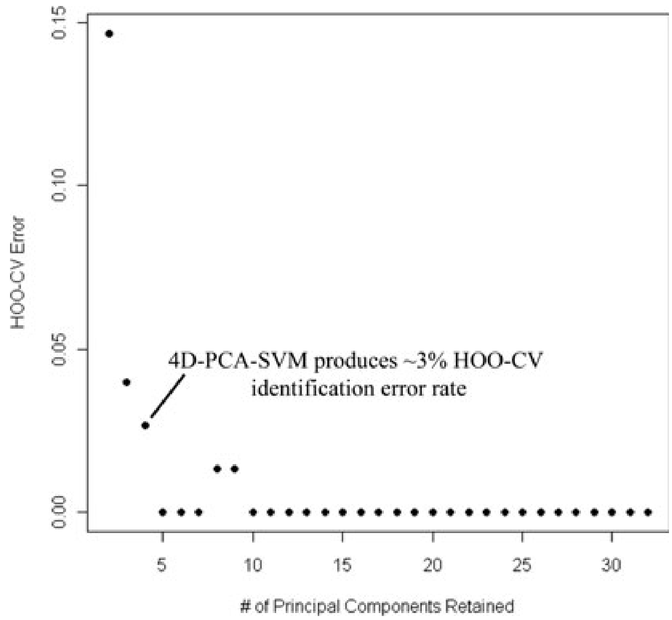

FIG. 9—*Hold-one-out cross-validation (HOO-CV) error rates versus PCA dimension (i.e., number of principal components retained). As expected classification, error generally decreases as dimension of the space is increased. However, the decrease is faster than for PLS.*

TABLE 2—*PCA-SVM correct classification rate estimates.*

	Data Dimension*			
	2D (%)	3D (%)	4D (%)	31D (%)
Method of estimation†				
Apparent	87	100	100	100
Hold-one-out CV	81	97	99	100
Refined bootstrap	88	96	98	99
.632 bootstrap	81	97	98	99
Random test set‡	76	94	100	100

*Data dimension is the number of principal components retained to represent the data.

†The method's estimated error rate = (100 − correct classification rate estimate)%.

‡Preprocessing transformation computed using corresponding random training set.

shown. All dimensions shown in Table 2 performed well on a randomly chosen test set of 25% of the data.

Conformal Prediction with PCA-SVM

Table 3 gives the results for CPT utilizing PCA-SVM for classification in both the on-line and the off-line modes at the 95% level of confidence. Columns two and three show CPT results when applied to the 75 real striation patterns when reduced to 4D (indicated as adequate for good discrimination by HOO-CV; cf., discussion above) and 121D (not reduced in dimension at all, i.e., the full data set). In the on-line mode, immediately after I.D. prediction regions are generated by the CPT algorithm, the true I.D. of the striation pattern being tested is fed into the computations and an I.D. prediction region is generated for the next striation pattern to be tested. This "on-line" process of continually updating the data set used to make I.D. predictions guarantees that the CPT algorithm

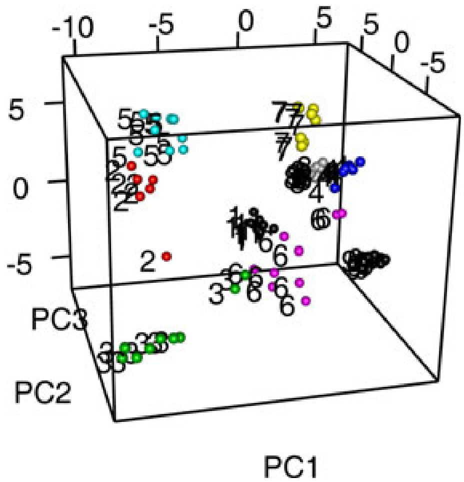

FIG. 10—*All striation patterns projected into the space of the first three principal components (44% variance retained). Each point represents a striation pattern. The boldfaced numbers to the left of the points tell which screwdriver generated the pattern. Cf. Table 2 for correct identification rates of striation patterns in this space.*

TABLE 3—*95% Confidence prediction results using PCA-SVM classification.*

	Data Dimension*	
	4D†,‡ (%)	121D‡ (%)
On-line mode		
% Error	6	6
% Unique and correct I.D. produced	94	94
% Efficiency	100	100
% Empty intervals	6	6
Off-line mode		
% Error	0	6
% Unique and correct I.D. produced	100	88
% Efficiency	100	94
% Empty intervals	0	6

PCs, principal components.

*Data dimension is the number of PCs retained to represent the data. 121D is the full dimensionality of the data set.

†Estimated minimal dimension to adequately represent the striation pattern data.

‡Results based on a random *c.* 75%:25% randomized split of the real striation pattern data into training/test sets (58 training patterns/17 test patterns).

yields erroneous prediction regions at a level of less than or equal to $\alpha \times 100\%$, up to small statistical fluctuation (89).

As is apparent from columns two and three of Table 3, application of CPT to the real striation patterns produced confidence regions containing the correct identity of the test pattern to within a reasonable statistical fluctuation of the theoretical error rate. An error rate of 6% was found for both 4D and 121D, which is within statistical fluctuation of the expected 5% error rate. In other words, incorrect 95% confidence intervals (empty or not containing the true I.D.) were produced only about 5% of the time (89,95).

Also, prediction regions produced by CPT can be multilabel. Such multilabel results are not counted as incorrect if they contain the true I.D. of the striation pattern being tested. Obviously, though, while not totally useless, correct but multilabel prediction regions are not maximally informative. Thus, an important measure of the CPT algorithm's performance is also the rate at which it produces regions with one unique I.D. and that I.D. is correct. Table 3 shows that on-line CPT produced unique and correct I.D.s of the striation patterns well above 90% of the time at the 95% level of confidence. Such results are very rigorous numerical identifications of striation patterns and are accompanied by a definitive level of confidence. They can also be stated easily in a courtroom situation.

The original developers of CPT acknowledge that strict on-line classification is not always realistic in practice (89,90). That is, one cannot expect that the true label for each test striation pattern can be produced immediately after classification occurs. To circumvent this problem, there are three solutions. The first is to repeat the CPT identification sequence on each test pattern, building up to it by first classifying a fairly large sequence of randomly selected striation patterns with known labels. Essentially, this is on-line prediction using many independent random samples from a large data set augmented with one unknown striation pattern as the last pattern to be tested. In this way, a valid confidence level is maintained for the label set produced for the unknown pattern. A drawback to this approach is that it is very computationally intensive. A second possibility is to use CPT in "slow teacher" mode, where the true identities of unknown test patterns are eventually presented to the algorithm. This also produces a confidence level "as advertised" but larger statistical fluctuations in the error rate can occur for small data sets (89,90).

The last possibility is what we pursue in this study; use the CPT algorithm in the "off-line" mode, where the same sequence of striation patterns with known identities is used to predict identities of a set of unknown patterns. CPT in the off-line mode can no longer theoretically guarantee that the $(1-\alpha)\times 100\%$ confidence regions produced are erroneous $\alpha\times 100\%$ of the time. However, despite this fact, it has been shown empirically that off-line mode CPT confidence region error rates actually do remain close to $\alpha\times 100\%$ (96). Table 3 shows the off-line performance of CPT with PCA-SVM classification. Results are indeed commensurate with CPT run in the on-line mode, that is, only about 5% of the striation patterns were incorrectly identified. Also correct, unique label prediction regions were produced *c.* 90% of the time or higher. Computations, however, were much faster in off-line mode.

Conformal Prediction with k-nearest Neighbors

While a nonconformity measure for PLS-DA which can be used with CPT is imminent, none has yet been released in the literature. Thus, we employed an alternative classification method which has been extensively shown in the literature to perform well when combined with CPT, k-nearest neighbors (89,90,95). Table 4 shows the 95% confidence CPT results using 3-NN for classification in both the on-line and the off-line modes. Again, the error rates are all on or close to 5%. It should be noted that a 0% error rate for 4D 3-NN CPT classification on the real striation patterns is a bit optimistic (cf., Table 4 column two). The uniqueness and efficiency of the prediction regions produced by the algorithm run in both modes are consistently *c.* 90% or more as was the case for PCA-SVM. Again, this means that most of the time CPT produced a 95% confidence interval with only one I.D. for the test pattern and that I.D. was correct.

Conclusion

The intention of this study was to show that it is possible to associate relatively flat ideal striation patterns with a "scraping"

TABLE 4—*95% Confidence prediction results using 3-nearest neighbor classification.*

	Data Dimension*	
	4D[†,‡] (%)	121D[‡] (%)
On-line mode		
% Error	0	6
% Unique and correct I.D. produced	100	94
% Efficiency	100	100
% Empty intervals	0	6
Off-line mode		
% Error	0	6
% Unique and correct I.D. produced	100	94
% Efficiency	100	100
% Empty intervals	0	6

PCs, principal components.
*Data dimension is the number of PCs retained to represent the data. 121D is the full dimensionality of the data set.
[†]Estimated minimal dimension to adequately represent the striation pattern data.
[‡]Results based on a random *c.* 75%:25% randomized split of the real striation pattern data into training/test sets (58 training patterns/17 test patterns).

type tool using objective numerical measures of similarity, error rate, and confidence. The criteria of selection for the statistical methods used in this study were that they have a peer-reviewed track record, a high rate of success in the fields in which they have been applied, and that they are relatively free of many of the underlying assumptions that typically underlie comparison methods in statistics (e.g., assumed parametric distributions, dichotomous decisions). PLS-DA was able to differentiate striation patterns made by screwdrivers at or higher than a 97% correct classification rate (≤3% error rate) with 8D feature vectors. PCA-SVM showed comparable high performance with only 4D feature vectors. CPT, which has grown out of Solomonoff's and Kolmogorov's algorithmic theory of randomness, can in fact be used to control identification error rates. For the first time, in forensic science, we used CPT to produce tool mark identifications (conformal prediction regions) at the 95% level of confidence. As is advertised by the theory, error rates were always 5% to within a small statistical fluctuation. Uniquely labeled and correct conformal prediction regions were produced at or greater than 90% of the time using CPT in both on-line and off-line modes.

Our sample size for this pilot study was necessarily small; however, the results strongly indicate the feasibility of using machine learning techniques to identify tool marks. Studies are currently under way to drastically increase the sample size, vary the angle at which the tool strikes the impression media, use incomplete striation patterns (more akin to what a tool mark examiner encounters in practice), and most importantly use 3D metrological instrumentation and software to carry out the same set of tasks. Also, the techniques presented in this study could be extended to striation patterns found on firearms evidence. Last, the authors would be happy to share the data set and the Mathematica and R software written by them for this study, upon request.

Acknowledgments

We are grateful to Professors Patrick Buzzini and Christopher Saunders for enlightening discussions of subjectively based Bayesian methods and valuable comments on our manuscript. We also thank Ms. Loretta Kuo, Officer Patrick MacLaughlin, Ms. Helen Chan, and Mr. Manny Chapparo, for many valuable discussions and for reading our manuscript.

References

1. Edwards HT, Gatsonis C, Berger MA, Cecil JS, Bonner-Denton M, Fierro M, et al. Strengthening forensic science in the United States: a path forward, 1st edn. Washington, DC: National Academies Press, 2009.
2. Moran B. A report on the AFTE theory of identification and range of conclusions for tool mark identification and resulting approaches to casework. AFTE J 2002;34(2):227–35.
3. Grzybowski RA, Murdock JE. Firearm and toolmark identification meeting the *Daubert* challenge. AFTE J 1998;30(1):3–14.
4. Bohan TL, Heels EJ. The new scientific evidence ''Standard'' and the standards of several states. J Forensic Sci 1995;40(6):1030–44.
5. Thompson E, Wyant R. Knife identification project (kip). Ann Firearm Toolmark Exam J (AFTE) 2003;35(4):366–70.
6. Tomasetti KA. Analysis of the essential aspects of striated tool mark examinations and the methods of examination. Ann Firearm Toolmark Exam J (AFTE) 2002;34(3):289–301.
7. Collins JM. The language of toolmarks. Ann Firearm Toolmark Exam J (AFTE) 1998;30(1):366–70.
8. AFTE. Theory of identification as it relates to toolmarks. AFTE J 1998;30(1):86–8.
9. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd edn. New York, NY: Wiley, 2001.
10. Theodoridis S, Koutroumbas K. Pattern recognition, 3rd edn. San Diego, CA: Academic Press, 2006.
11. Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. London, UK: Cambridge University Press, 2004.
12. Leon FP. Automated comparison of firearm bullets. Forensic Sci Int 2006;156:40–50.
13. Gardner GY. Computer identification of bullets. AFTE J 1979;11(2):26–33.
14. Taroni F, Champod C, Margot P. Statistics: a future in tool marks comparison? AFTE J 1996;28(4):222–32.
15. Champod C, Baldwin D, Taroni F, Buckleton JS. Firearm and tool marks identification: the Bayesian approach. AFTE J 2003;35(3):307–16.
16. Buckleton J, Nichols R, Triggs C, Wevers G. An exploratory Bayesian model for firearm and toolmark interpretation. AFTE J 2005;37(4):352–61.
17. Biedermann A, Taroni F. A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues. Forensic Sci Int 2006;163:18–33.
18. Katterwe H, Baldwin D, vanBeest M, Besler C, Birkett J, Girod A, et al. Conclusion scale for shoeprint and toolmark examinations. J Forensic Identification 2006;56(2):255–81.
19. Neel M, Wells M. A comprehensive statistical analysis of striated tool mark examinations. Part 1: comparing known matches to known nonmatches. AFTE J 2007;39(3):176–98.
20. Biasotti AA. A statistical study of the individual characteristics of fired bullets. J Forensic Sci 1959;4(1):34–50.
21. Miller J, Neel M. Criteria for identification of toolmarks part III: supporting the conclusion. AFTE J 2004;37(1):7–38.
22. Brackett J. A study of idealized striated marks and their comparison using models. J Forensic Sci Soc 1970;10(1):27–56.
23. Blackwell R, Framan E. Automated firearms identification systems AFIDS: Phase I. AFTE J 1980;12(4):11–37.
24. Deinet W. Studies of models of striated marks generated by random processes. J Forensic Sci 1981;26(1):35–50.
25. Stone R. How unique are impressed toolmarks? AFTE J 2003;35(4):376–83.
26. Collins E. How unique are impressed toolmarks? An empirical study of twenty worn hammer faces. AFTE J 2005;37(4):252–95.
27. Bunch S. Consecutive matching striation criteria: a general critique. J Forensic Sci 2000;45(5):955–62.
28. Banno A. Estimation of bullet striation similarity using neural networks. J Forensic Sci 2004;49(3):1–5.
29. Bachrach B. Development of a 3D-based automated firearms evidence comparison system. J Forensic Sci 2002;47(6):1–12.
30. Bachrach B. A statistical validation of the individuality of guns using 3D images of bullets. National Institute of Justice, Grant Report: 97-LB-VX-0008. Washington, DC: National Institute of Justice, 2006.
31. Geradts Z, Keijzer J. TRAX for toolmarks. AFTE J 1996;28(3):183–90.

32. Geradts Z, Keijzer J, Keereweer I. A new approach to automatic comparison of striation marks. J Forensic Sci 1994;39(4):974–80.
33. Geradts Z, Bijhold J, Hermsen R, Murtaugh F. Image matching algorithms for breech face marks and firing pins in a database of spent cartridges of firearms. Forensic Sci Int 2001;119:97–106.
34. Senin N, Groppetti R, Garofano L, Fratini P, Pierni M. Three-dimensional surface topography acquisition and analysis for firearm identification. J Forensic Sci 2006;51(2):282–95.
35. Faden D, Kidd J, Craft J, Chumbley LS, Morris M, Genalo L, et al. Statistical confirmation of empirical observations concerning tool mark striae. AFTE J 2007;39(3):205–14.
36. DeKinder J, Tulleners F, Thiebaut H. Reference ballistic imaging database performance. Forensic Sci Int 2004;140:207–15.
37. DeKinder J, Bonfanti M. Automated comparisons of bullet striations based on 3D topography. Forensic Sci Int 1999;101:85–93.
38. Song J, Vorburger T, Renegar T, Rhee H, Zheng A, Ma L, et al. Correlation of topography measurements of NIST SRM 2460 standard bullets by four techniques. Measurement Sci and Tech 2006;17(3):500–3.
39. Booker JL. Examination of the badly damaged bullet. J Forensic Sci Soc 1980;20:153–62.
40. Uchiyama T. A criterion for land mark identification. AFTE J 1988;20(3):236–51.
41. Uchiyama T, Igarashi N, Nagai M. The frequency of occurrence of individual characteristics of firearms on fired bullets and cartridge cases. AFTE J 1988;20(4):376–90.
42. Nichols RG. Firearm and toolmark identification criteria: a review of the literature. J Forensic Sci 1997;42(3):466–74.
43. Nichols RG. Firearm and toolmark identification criteria: a review of the literature-part 2. J Forensic Sci 2003;48(2):318–27.
44. Nichols RG. Defending the scientific foundations of the firearms and tool mark discipline: responding to recent challenges. J Forensic Sci 2007;53(3):586–94.
45. Silbert RW. Drugfire revolutionizing forensic firearms identification and providing a foundation for a national firearms identification network. USA Crime Lab Digest 1994;21:63–8.
46. Baldur R, inventor. Forensic Technology Wai Inc., assignee. Fired cartridge examination method and imaging apparatus. US patent 5654801 (file date: 1/3/1995; issue date: 8/5/97).
47. Baldur R, Barrett MR, inventors. Forensic Technology Wai Inc., assignee. Computer automated bullet analysis apparatus. US patent 5390108 (file date: 5/24/1991; issue date: 2/14/95).
48. Cork DL, Rolph JE, Meieran ES, Petrie CV, eds. Ballistic imaging. Washington, DC: National Academies Press, 2008.
49. Geradts Z. Content-based information retrieval from forensic image databases. Utrecht, The Netherlands: University of Utrecht, 2002.
50. Geradts Z, Keijzer J, Keereweer I. Automatic comparison of striation marks and automatic classification of shoe marks. Proceedings SPIE 1995, http://spie.org/x648.html?product_id=218471 (accessed March 8, 2012).
51. DeKinder J, Prevot P, Pirlot M, Nys B. Surface topology of bullet striations: an innovating technique. AFTE J 1998;30:294–9.
52. Bachrach B, Jain A, Jung S, Koons RD. A statistical validation of the individuality and repeatability of striated tool marks: screwdrivers and tongue and groove pliers. J Forensic Sci 2010;55(1):348–57.
53. Chu W, Song J, Vorburger T, Yen J, Ballou S, Bachrach B. Pilot study of automated bullet signature identification based on topography measurements and correlations. J Forensic Sci 2010;55(2):341–7.
54. Banno A, Masuda T, Ikeuchi K. Three dimensional visualization and comparison of impressions on fired bullets. Forensic Sci Int 2004;140:233–40.
55. Sakarya U, Leloglu UM, Tunali E. Three-dimensional surface reconstruction for cartridge cases using photometric stereo. Forensic Sci Int 2008;175:209–17.
56. Evans JPO, Smith CL, Robinson M. Validation of line scan imaging technique for imaging cylindrical forensic ballistics specimens. AFTE J 2004;36(4):275–80.
57. Zographos A, Robinson M, Evans J, Smith CL. Ballistics identification using line-scan imaging techniques. Proceedings of the Institute of Electrical and Electronics Engineers 31st Annual International Carnahan Conference on Security Technology; 1997 Oct 15–17; Canberra, Australia. IEEE 1997;82–7, doi 10.1109/CCST.1997.626243.
58. Smith C. Fireball: a forensic ballistic imaging system. Proceedings of the Institute of Electrical and Electronics Engineers 31st Annual International Carnahan Conference on Security Technology; 1997 Oct 15–17; Canberra, Australia. IEEE 1997;64–70 doi 10.1109/CCST.1997.626240.
59. vanBeest M, Zaal D, Hardy H. The forensic application of the Mikrocad 3D imaging system. Inform Bull Shoeprint/Toolmark Examiners 2000;6:65–72.
60. Howitt D, Tulleners F, Cebra K, Chen S. A calculation of the theoretical significance of matched bullets. J Forensic Sci 2008;53(4):868–75.
61. Doble P, Sandercock M, DuPasquier E, Petocz P, Rouxa C, Dawson M. Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. Forensic Sci Int 2003;132(1):26–39.
62. Sandercock PML, Pasquier ED. Chemical fingerprinting of unevaporated automotive gasoline samples. Forensic Sci Int 2003;134(1):1–10.
63. Sandercock PML, Pasquier ED. Chemical fingerprinting of gasoline 2. Comparison of unevaporated and evaporated automotive gasoline samples. Forensic Sci Int 2004;140(1):43–59.
64. Sandercock PML, Pasquier ED. Chemical fingerprinting of gasoline 3. Comparison of unevaporated automotive gasoline samples from Australia and New Zealand. Forensic Sci Int 2004;140(1):71–7.
65. Egan W, Galipo RC, Kochanowski BK, Morgan SL, Bartick EG, Miller ML, et al. Forensic discrimination of photocopy and printer toners. III. Multivariate statistics applied to scanning electron microscopy and pyrolysis gas chromatography/mass spectrometry. Anal Bioanal Chem 2003;376:1286–97.
66. Egan W, Morgan SL, Bartick EG, Merrill RA, Taylor HJ. Forensic discrimination of photocopy and printer toners. II. Discriminant analysis applied to infrared reflection absorption spectroscopy. Anal Bioanal Chem 2003;376:1279–85.
67. Morgan SL, Bartick EG. Discrimination of forensic analytical chemical data using multivariate statistics. In: Blackledge RD, editor. Forensic analysis on the cutting edge: new methods for trace evidence analysis. New York, NY: Wiley, 2007;331–72.
68. Ratle F, Gagne C, Terrettaz-Zufferey AL, Kanevski M, Esseiva P, Ribaux O. Advanced clustering methods for mining chemical databases in forensic science. Chemometrics Intelli Lab Sys 2008;90(2):123–31.
69. Vapnik VN. Statistical learning theory. New York, NY: Wiley, 1998.
70. Wold S, Eriksson L, Trygg J, Kettaneh N. The pls method-partial least squares projections to latent structures-and its application in industrial rdp (research, development and production). Proceedings of COMPSTAT 2004, 16th Symposium of IASC; 2004 Aug 23–27; Prague, Czech Republic. The Hague, The Netherlands: International Association for Statistical Computing, 2004;1–44.
71. Petraco N, Petraco NDK, Faber L, Pizzola PA. Preparation of tool mark standards with jewelry modeling waxes. J Forensic Sci 2009;54(2):353–8.
72. Rasband W. ImageJ [computer program]. National Institute of Health, http://rsb.info.nih.gov/ij/ (accessed April, 2010).
73. Wolfram Development. Mathematica [computer program], 7th edn. Champaign, IL: Wolfram Research, Inc., 2009.
74. Jolliffe IT. Principal component analysis, 2nd edn. New York, NY: Springer, 2004.
75. Vapnik VN. The nature of statistical learning theory. New York, NY: Springer, 2000.
76. R Core Development Team. R: a language and environment for statistical computing [computer program], version 2.9.1. Vienna, Austria: R Foundation for Statistical Computing, 2009.
77. Bertsekas DP, Tsitsiklis JN. Introduction to probability, 1st edn. Belmont, MA: Athena Scientific, 2002.
78. Rencher AC. Methods of multivariate analysis, 2nd edn. Hoboken, NJ: Wiley, 2002.
79. Petraco NDK, Gil M, Pizzola PA, Kubic TA. Statistical discrimination of liquid gasoline samples from casework. J Forensic Sci 2008;53(5):1092–101.
80. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;8(5):1–26.
81. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. London, UK: Cambridge University Press, 2000.
82. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab—an s4 package for kernel methods in R. J Stat Softw 2004;11(9):1–20.
83. Barker M, Rayenes W. Partial least squares for discrimination. J Chemometrics 2003;17(6):166–73.
84. Mevik BH, Wehrens R. The pls package: principal component and partial least squares regression in R. J Stat Softw 2007;18(2):1–24.
85. Smith CAB. Some examples of discrimination. Ann Eugenics 1947;18:272–83.
86. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 1983;78(382):316–31.

87. Efron B, Tibshirani RJ. An introduction to the bootstrap, 1st edn. Boca Raton, FL: Chapman & Hall/CRC, 1993.
88. Li M, Vitanyi P. An introduction to Kolmogorov complexity and its applications, 3rd edn. New York, NY: Springer, 2008.
89. Vovk V, Gammerman A, Shafer G. Algorithmic learning in a random world, 1st edn. New York, NY: Springer, 2005.
90. Shafer G, Vovk V. A tutorial on conformal prediction. J Mach Learn Res 2008;9:371–421.
91. Dekking FM, Kraaikamp C, Lopuhaa HP, Meester LE. A modern introduction to probability and statistics: understanding how and why, 1st edn. New York, NY: Springer, 2005.
92. Lehmann EL, Romano JP. Testing statistical hypotheses, 3rd edn. New York, NY: Springer, 2005.
93. Rosner B. Fundamentals of biostatistics. Belmont, CA: Thomson, 2006.
94. Bishop CM. Pattern recognition and machine learning, 1st edn. New York, NY: Springer, 2006.
95. Gammerman A, Vovk V. Hedging predictions in machine learning. Comp J 2007;50(7):151–77.
96. Vanderlooy S, Maaten L, Sprinkhuizen-Kuyper I. Off-line learning with transductive confidence machines: an empirical evaluation. In: Perner P, editor. Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition; 2007 July 18–20; Leipzig, Germany. Berlin, Germany: Springer-Verlag, 2007;310–23.

Additional information and reprint requests:
Nicholas D. K. Petraco, Ph.D.
Department of Sciences
John Jay College of Criminal Justice
899 10th Avenue
New York, NY 10019
E-mail: npetraco@jjay.cuny.edu