Virtual and simulated striated toolmarks for forensic applications

Martin Baiker a,*, Nicholas D.K. Petraco d, Carol Gambino c, René Pieterman a, Peter Shenkin b, Peter Zoon a

a Department of Mircrotraces, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497CB Den Haag, The Netherlands
b Department of Mathematics and Computer Science, John Jay College of Criminal Justice, City University of New York, 524 West 59th Street, New York, NY 10019, United States
c Department of Sciences, Borough of Manhattan Community College, City University of New York, 199 Chambers Street, New York, NY 10007, United States
d Department of Sciences, John Jay College of Criminal Justice, City University of New York, 524 West 59th Street, New York, NY 10019, United States

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 28 December 2015
Accepted 21 January 2016
Available online 1 February 2016

Keywords:
Virtual and simulated toolmarks
Objective comparison
Striated toolmarks
3D tool models
Wavelet based multiscale analysis
Toolmark variability

ABSTRACT

Large numbers of experimental toolmarks of screwdrivers are often required in casework of toolmark examiners and in research environments alike, to be able to recover the angle of attack of a crime scene mark and to determine statistically meaningful properties of toolmarks respectively. However, in practice the number of marks is limited by the time needed to create them.

In this article, we present an approach to predict how a striated mark of a particular tool would look like, using 3D surface datasets of screwdrivers. We compare these virtual toolmarks qualitatively and quantitatively with real experimental marks in wax and show that they are very similar. In addition we study toolmark similarity, dependent on the angle of attack, with a very high angular resolution of 1°. The results show that for the tested type of screwdriver, our toolmark comparison framework yields known match similarity scores that are above the mean known non-match similarity scores, even for known match differences in angle of attack of up to 40°. In addition we demonstrate an approach to automatically recover the angle of attack of an experimental toolmark and experiments yield high accuracy and precision of 0.618 ± 4.179°. Furthermore, we present a strategy to study the structural elements of striated toolmarks using wavelet analysis, and show how to use the results to simulate realistic toolmarks.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Tools like screwdrivers and crowbars are often used during the commission of a crime and therefore striated toolmarks can regularly be found at a crime scene. In case a tool can be seized from a suspect afterwards the question arises, whether the marks were created with that particular tool. To tackle this question forensic toolmark examiners generate experimental test marks with the suspect tool in the laboratory and subsequently compare them to the questioned marks found at the crime scene.

The traditional method to compare questioned and test marks is to use 2D microscopy. The examiner puts both marks under a comparison microscope and manually illuminates the toolmarks with oblique light, such that the striations become visible as a light(ridges)-shadow(furrows) pattern. Subsequently, the examiner has to assess the possibility of (dis-)similarities between the marks, assuming that they are made with the same tool vs. assuming that they are made with different tools.

The traditional approach of toolmark examination relies on manual illumination and comparison of the marks and therefore includes subjective judgments. Therefore a report of the US National Academy of Sciences [2] asks for more objective ways to assess toolmark evidence and in recent years, the interest in the use of surface metrology for objective data acquisition and automated approaches for objective data analysis and comparison has been growing [1,3–12]. For quantitative toolmark comparison however, the statistical properties of toolmarks have to be known. Several parameters including the angle of attack, the substrate material, the axial tool rotation and the toolmark depth influence the toolmark formation process and the degree of similarity between toolmarks (Fig. 1, left). For objective toolmark comparison, it is important to determine the influence of the various parameters statistically. This can be done by creating experimental toolmarks and varying one particular parameter like the angle of
attack. Ideally, the range of the varied parameter should be chosen as large as possible with as high resolution as possible to obtain robust statistical estimates of the toolmark variability. In practice however this requires producing a huge number of experimental toolmarks. This is very time consuming and therefore prior studies limited the amount of a tool’s angle of attack to three [4,5,11] or five [3] angles. An alternative approach is to employ an approach that uses a 3D surface dataset of a particular tool, transforms, i.e. translates (shifts), rotates and scales, the dataset with a computer and subsequently predicts virtual marks that the tool would leave, depending on a given parameter. This offers the possibility to generate a large number of marks to study statistical properties of toolmarks theoretically. But also in daily practice, virtual toolmarks can play an important role. Typically, toolmark examiners have to create multiple test marks to compare with an unknown mark. However, even relatively soft substrate materials like lead may alter the state of a tool during toolmark creation [1]. If virtual toolmark generation software could predict, say, the angle of attack, with high accuracy (in case the suspect toolmark was indeed created with the suspect tool), a toolmark examiner would only have to create one experimental toolmark at that particular angle for comparison with the suspect mark.

Ekstrand et al. have developed a virtual toolmark generator [13] where a dataset of a tool’s working surface is acquired using 3D microscopy (focus variation data was specifically reported, but the system can utilize data from any 3D microscopy). The geometry of the working surface is projected in the direction of tool travel. This identifies the highest points on that projection which scrape the deepest into the substrate material. A novel implementation scheme using graphical processing units (GPUs) was employed to significantly speed up the procedure. The technique developed by the Iowa group can simulate a toolmark at arbitrary twist of the tool, and angles of attack. An experiment showed that automated detection of the angle of attack of a tool during toolmark creation could be done with a precision of ±5° to 1°. Bachrach et al. have recently reported a an approach, which exploits wavelet analysis of bullet Land Engraved Area (LEA) signatures [14] to generate new signatures with similar properties, i.e. simulate LEA signatures. Long wavelength shape and ‘brand’ (class) characteristics are extracted through the wavelet coefficients. The software uses fractal analysis to include local ‘randomness’ components (i.e. surface roughness) into the simulated signatures. This allows the random portions of the signatures to be generated by predetermined parametric probability distributions. The system is also capable of producing 2D LEA images and 3D bullet surfaces.

1.1. Contributions

In the previously described approaches that have been published, either the influence of a particular toolmark formation parameter has been studied by generating virtual toolmarks with 3D surface datasets, or realistic toolmarks were simulated based on existing experimental, hence limited, data. In this paper we describe a methodology that can both, generate a large amount of virtual toolmarks for studying one particular parameter like the tool angle of attack or axial rotation angle and use this data to simulate realistic (but non-existing) toolmarks that can be created with the same tool. More specifically, we present an approach to acquire 3D surface datasets of tools and to use them to predict virtual toolmarks over a wide range of angles of attack and axial rotation angles. In addition we show a way to analyze the geometrical features of the virtual toolmarks using wavelet decomposition and subsequently simulate realistic toolmarks. To demonstrate the usefulness of our framework, we study the impact of the angle of attack on toolmark similarity with very high resolution, we recover the true angle of attack of known-matching (KM) toolmarks, we compare true and simulated toolmark variability and assess qualitatively simulated toolmark profiles and toolmarks.

2. Methodology

2.1. Tools

The tools for creating the surface datasets were new standard off-the-shelf slotted screwdrivers model Gedore 150 S-8-175 [15] with blade dimensions of about (8 mm×1 mm). During manufacturing, all four sides and the front face of the blade have been ground manually, resulting in the grinding patterns visible in Fig. 1 (right).

2.2. Tool surface acquisition and pre-processing

The screwdrivers were put in a holder, in a position equivalent to 45° angle of attack α, and acquired using an Alicona Infinite Focus Microscope [16]. The working principle is based on focus variation, a non-contact, high resolution surface metrology approach. The acquisition parameters were set to VR = 200 nm (vertical resolution), HR = 2 μm (horizontal resolution) (with a sampling distance of ~438 nm), and M = 20× (objective magnification). The surface data was provided on a regular grid (typically around 3k × 20k pixels) with precision of 32 bit and was exported in the Alicona file format (al3d). For each tool, two datasets were
acquired, one with 0° and one with 180° axial rotation γ. Therefore the first dataset included the front face and one lateral face (as depicted in Fig. 1, right) and the second dataset the front face and the other lateral face. An example of a resulting dataset is shown in Fig. 2 (middle).

After the acquisition, the data was imported into Matlab 2012b [17] using our Alicona data import routine (available online [18]) and cropped manually with respect to the global x-axis (as depicted in Fig. 2, middle). The measurement system labeled invalid measurement points, which were substituted by averaging the ten nearest valid measurement points. Stitches in the data were detected using an elliptic variance filter (size = 10) and substituted as mentioned above (parameters were determined empirically). Invalid measurement points and stitches rarely occurred however, as demonstrated in Fig. 3.

2.3. Virtual toolmark generation

After pre-processing, the surface dataset was rotated such that the edge of the blade was outlined to the global x-axis and the front face was outlined to the xy-plane (Fig. 2). This step was fully automated and was based on minimizing the standard deviation in z-direction of all surface points, projected onto the yz-plane, within a nonlinear optimization framework with two parameters: rotation with respect to the y-axis and rotation with respect to the z-axis. Subsequently a polygon with two line segments was fitted to the projected points to determine the rotation around the x-axis and the exact angle between the lateral face and the front face. To finally predict a profile at a given angle of attack, the surface dataset was rotated with respect to the global x-axis. Subsequently, the x-coordinates of all surface points were binned (bin size = pixel size of the original data) and the maximum z-coordinates of the individual bins yielded the profile (Fig. 4). Some examples for various angles (10°, 30° and 50°) are presented in Fig. 2 (right). Based on such a 1D profile, a striated toolmark can be generated by repeating the profile multiple times and adding realistic noise to the data.

The procedure described above, is generic and can deal with large angles with respect to the y- and z-axes (refer to Fig. 2). The determination of one profile at full resolution takes ≈2 s (sampling distance ≈438 nm). However, if the model is placed such that the rotation around the y- and z-axes is relatively small (several degrees) during the data acquisition, an alternative analytic approach for profile determination can be employed, which only takes ≈0.1 s (please refer to the appendix for details).

2.4. Simulation of realistic toolmarks

Our approach to simulate toolmarks from the same source tool is to separate the structural components of the surfaces into a number of spatial scales. For the stochastic portion of the simulation process, we have chosen a wavelet decomposition for scale, primarily for speed. The particular wavelet filtering
algorithm we chose scales linearly with the size of the toolmarks, and thus can significantly speed up the simulation process.

Toolmarks can be thought of as 1D (profiles, Fig. 10) or 3D (surfaces, Fig. 5) signals. Specifically, we wish to simulate the striation patterns left by the tool. Striation patterns are collections of visually dominant line impressions with a specific direction.

Imagine the 3D striation pattern shown in Fig. 5 as a rectangular matrix of points. It’s easy to see that the pattern is largely homogeneous along the direction of the dominant line texture, i.e. each ‘row’ (a profile) that makes up the data matrix is visually similar to all the other rows of the data matrix. From this observation we assume that a 3D striation pattern can be modeled as a collection of 1D profiles, stacked as rows. The 1D discrete wavelet transform (DWT) is an orthogonal transformation which decomposes a length $N = 2^l$ signal into $J$ levels. The levels are ordered by the length scale of the detail they capture and correspond to a unit-less standardized scale $r_j = 2^{l-j}$, with $j = 1, 2, \ldots, J$. That is, $j$ indexes the detail level. The approximate physical bandwidth of each level can be found by taking the product of the sampling interval $\Delta x$ and $r_j$. We typically measure $\Delta x$ in physical units of $\mu$m for forensic toolmark applications. We also note that the signal length must be a power of 2. If it is not, the user can pad the profile with zeros, chop the profile or resample it. Any padding is discarded in the final step. For example, say we have a toolmark that consists of 9784 points. Let us first reduce this to $2^{12} = 8192$ points, either by interpolation/resampling-sampling, or just simple truncation of the signal. Thus, there are $J = 13$ scale ‘levels’ to examine, which corresponds to sampling physical features at length scales in the range between 0.18 $\mu$m (smallest features, most detail) and 757.15 $\mu$m (largest features, least detail) for a sampling distance of $\Delta x = 0.18 \mu$m (Table 1). Fig. 6 shows a wavelet level decomposition (down to detail level 7) of an example striation pattern profile with $N = 8192$ points.

If, say, five striation patterns are collected from the same tool, assume that each derived profile will be of slightly different length. The longest profile is chosen as the reference. The remaining profiles are aligned to this reference with a cross correlation function and then stacked into a matrix. One can expect to find missing $z$-values where there was overhang between the profiles after alignment (Fig. 7, left). To remedy this, the aligned profiles are averaged down the columns (i.e. with respect to the toolmark length dimension in Fig. 5) to form what we call the ‘group grand mean profile’. Any missing values in the profile matrix will be filled in with the corresponding portions of the group grand mean profile (Fig. 7, right).

Each toolmark profile can be considered as a linear combination of wavelet basis functions. The expansion coefficients are known as wavelet coefficients and are essentially uncorrelated. Also, most coefficients have near zero magnitude. Therefore the DWT provides an excellent ‘compressed’ representation of the toolmark from which to base a simulation method. Thus, as the next step, each profile in the group is decomposed into its corresponding wavelet coefficient series with DWT. For each scale, vectors of wavelet coefficients are collected together into separate matrices, one matrix per level. So for example, consider the $j = 8$th level of detail. There will be $2^7 = 128$ wavelet coefficients at that level for each of the five profiles which will be collected into a 5 row by 128 column matrix. This process is repeated for each level $j$. Each column of each ‘level-matrix’ represents how the coefficients vary for a local region across the $5$ real profiles. Non-parametric kernel density estimates (KDEs) are fit to the coefficients in each column of each level-matrix. Samples of ‘simulated’ level coefficients are then drawn from probability densities level-by-level, and assembled into ‘simulated’ level-matrices. So, for example, say the user requests 50 simulated profiles. The simulator will generate 13 (there are 13 levels in our running example) level matrices. The matrices will each have 50 rows but the number of columns varies: $r_j = 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096$, depending on the level $j$. A row from each of the 13 level matrices is then selected and assembled into a wavelet coefficient vector. Finally, the inverse DWT is executed on the simulated wavelet coefficient vector yielding a simulated profile.

Each simulated profile is evaluated for its acceptability by computing its correlation to each of the real profiles used to simulate it. The user chooses a value for a correlation score below which a simulated profile will be discarded if it does not reach this magnitude when correlated with its real counterparts. By allowing the user to choose the threshold, he has control over how much variation there will be within the total set of real and simulated profiles. For example if the user chooses to set the threshold high, averaged down the columns (i.e. with respect to the toolmark length dimension in Fig. 5) to form what we call the ‘group grand mean profile’. Any missing values in the profile matrix will be filled in with the corresponding portions of the group grand mean profile (Fig. 7, right).

Each toolmark profile can be considered as a linear combination of wavelet basis functions. The expansion coefficients are known as wavelet coefficients and are essentially uncorrelated. Also, most coefficients have near zero magnitude. Therefore the DWT provides an excellent ‘compressed’ representation of the toolmark from which to base a simulation method. Thus, as the next step, each profile in the group is decomposed into its corresponding wavelet coefficient series with DWT. For each scale, vectors of wavelet coefficients are collected together into separate matrices, one matrix per level. So for example, consider the $j = 8$th level of detail. There will be $2^7 = 128$ wavelet coefficients at that level for each of the five profiles which will be collected into a 5 row by 128 column matrix. This process is repeated for each level $j$. Each column of each ‘level-matrix’ represents how the coefficients vary for a local region across the $5$ real profiles. Non-parametric kernel density estimates (KDEs) are fit to the coefficients in each column of each level-matrix. Samples of ‘simulated’ level coefficients are then drawn from probability densities level-by-level, and assembled into ‘simulated’ level-matrices. So, for example, say the user requests 50 simulated profiles. The simulator will generate 13 (there are 13 levels in our running example) level matrices. The matrices will each have 50 rows but the number of columns varies: $r_j = 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096$, depending on the level $j$. A row from each of the 13 level matrices is then selected and assembled into a wavelet coefficient vector. Finally, the inverse DWT is executed on the simulated wavelet coefficient vector yielding a simulated profile.

Each simulated profile is evaluated for its acceptability by computing its correlation to each of the real profiles used to simulate it. The user chooses a value for a correlation score below which a simulated profile will be discarded if it does not reach this magnitude when correlated with its real counterparts. By allowing the user to choose the threshold, he has control over how much variation there will be within the total set of real and simulated profiles. For example if the user chooses to set the threshold high.

### Table 1

Relationship between the $J=13$ scale levels $r_j$ and the physical bandwidth $r_j \times \Delta x$.

<table>
<thead>
<tr>
<th>$r_j$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
<th>2048</th>
<th>4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_j \times \Delta x$ (rounded)</td>
<td>0.18</td>
<td>0.37</td>
<td>0.74</td>
<td>1.48</td>
<td>2.96</td>
<td>5.92</td>
<td>11.83</td>
<td>23.66</td>
<td>47.32</td>
<td>94.64</td>
<td>189.29</td>
<td>378.57</td>
<td>757.15</td>
</tr>
</tbody>
</table>
say 0.99, the total set of real and simulated profiles will ‘look’ very similar, i.e. have high intra-group correlation values. On the other hand, if the user decides to set the correlation threshold low, say 0.4, there will be a wide variation among the ‘looks’ of the profiles. That is to say, though the simulated profiles were intended to be representative of toolmarks made by the same tool, their intra-group correlations will be low. One can think of the former case as an ‘easy to identify’ set of toolmarks, as the profiles all ‘look’ similar. This is juxtapose to the latter case which can be thought of as a ‘hard to identify’ set of toolmarks because there will be a wider variation among the profiles. The user also has the option to feed the simulated profiles back into the simulation loop and treat them as if they were real profiles, thereby increasing the sample size for the KDE fits of the DWT coefficients. Only the real profiles are ever used in judging whether or not to keep a simulated profile. An example of typical 1D simulated profiles along with the real profiles used to simulate them appears in Fig. 8.

3. Experimental results and discussion

3.1. Validation of the virtual toolmark generation framework

To demonstrate that the proposed framework produces realistic results, the virtual toolmarks were compared to experimental toolmarks, which were created in wax at angles of attack of 15°, 30°, 45°, 60° and 75°, using the same screwdrivers (SDs) that were also used for creating the surface datasets. In total, 10 screwdrivers (in fact both blade sides of 5 screwdrivers) were used to create 10 surfaces and experimental toolmarks. Subsequently the experimental toolmarks were cast using gray Forensic Silicone [19], imaged with the Alicona IFM and pre-processed. Then 1D profiles (signatures) were determined by averaging the toolmarks along the direction of toolmark creation (with respect to the toolmark length in Fig. 5). To be able to compare an experimental profile with a virtual profile, these were aligned first using a multi-scale registration framework, which besides shift takes moderate scaling between profiles into account. Subsequently, toolmark similarity was determined using cross correlation, which measures the linear dependency between two toolmark profiles. Cross correlation yields values between −1 and 1, indicating negative and positive dependency respectively. A cross correlation value of 0 means no dependency. In other words, in case two profiles are similar (e.g. known matches), the cross correlation is close to 1 and if they are clearly different (e.g. known non-matches), it will be close to 0. For details regarding toolmark creation, casting, alignment and comparison, please refer to [3].

Qualitative results of the comparison of virtual and experimental (real) toolmarks are shown in Fig. 9 for angles of attack of 15°, 45° and 75°. Virtual toolmark creation only yields one toolmark profile for each angle. Therefore for visualization purposes, the profile was stretched horizontally. The virtual toolmark contains more fine details compared to the experimental toolmarks. This effect gets clearer with increasing angle of attack as the experimental toolmark at 15° appears to be of much better quality compared to the toolmark at 75° (Fig. 9). With increasing angle of attack, more and more material remains in the mark during toolmark creation and fine striations are filled with substrate material. This negatively influences the toolmark quality [1]. However the fact that toolmark profiles were created by averaging along the direction of striations, partially compensates for this. This can be seen in Fig. 10, were quantitative results of comparing 1D profiles are shown. The cross correlation between experimental and virtual profiles only slightly decreases with increasing angle of attack.

To determine statistically meaningful distributions of known match similarities, the virtual and experimental toolmark profiles of all ten screwdrivers at all five angles were compared, hence 10 × 5 = 50 comparisons were made. The known non-matching

---

**Fig. 7.** After alignment to the reference profile, profiles in a group are stacked into a matrix. Portions where there is overhang are filled in initially with missing value place holders (left) but can be filled in with group grand mean z-values (right). Both profiles are now the same length.

**Fig. 8.** Three real and five corresponding simulated profiles.
distributions were determined using virtual profiles at 15° and 75°, by comparing them to all virtual profiles at 15° and 75° of the remaining 9 screwdrivers. This was done for all SDs, yielding in total 10 × 9 × 2 = 180 comparisons. We chose a large angle difference between the profiles, 15° and 75°, to ensure that the profiles have minimal similarity. The results are shown in Fig. 11 (left) as KM Real vs. Virtual (black) and KNM Real vs. Virtual (green). To determine how realistic these distributions are, two experimental toolmarks were generated for each scenario. In other words, for SD1 at 15° angle of attack, two physical toolmarks were created. Subsequently, the second set was compared to the first set as described above for the comparison between virtual and real toolmarks. The results are shown in Fig. 11 (left) as KM Real vs. Real (red) and KNM Real vs. Real (blue). Qualitatively, the KM and KNM distributions for the two comparison cases, real vs. virtual and real vs. real are very similar, with most KM cross correlation values above 0.85. In addition, the KM and KNM distributions are well separated. For the KM Real vs. Virtual comparisons there exist a few similarity values that are lower than 0.85. As can be seen in Fig. 11 (right), these stem from the comparisons of profiles at the angles of attack of 60° and 75° and can be explained by the decreasing quality of the experimental toolmarks with increasing angle of attack. The mean similarity value including only toolmarks at small angles of attack (15°, 30° and 45°) is 0.946 ± 0.042. Note that pilot experiments (results not presented here) in which virtual profiles of geometrical primitives like spheres and pyramids were created and compared to the mathematically defined ‘ideal’ profile did not yield any dependency between the rotation angle of an object and profile similarity.

3.2. Angle of attack dependent variation of toolmark similarity

Virtual toolmarks enable examiners to study the effect of a particular toolmark generation parameter on a mark with high resolution and over a large range, without requiring a large amount of experimental toolmarks. In previous work we studied the effect of the angle of attack on toolmark similarity with experimental toolmarks at 15°, 30°, 45°, 60° and 75° (thus with a resolution of 15°). The results showed that known matching toolmarks with 15° angle difference are still highly similar and that even toolmarks with 30° angle difference are often still more similar than known non-matches [3]. In this work, we studied the effect of the angle of attack with a resolution of 1° using virtually generated toolmarks. A qualitative example of the resulting profiles in a large range of the angle of attack is given in Fig. 12. The figure illustrates in a powerful way, how structures in the tool surface influence a toolmark and demonstrates the gradual change of a toolmark with the angle of attack.

For quantitative analysis, we chose an angle of 45° as the reference and assessed a range of ±40° (i.e. 81 profiles) with respect to the reference. We used 9 screwdrivers for this experiment. One screwdriver was discarded, because the virtual profiles at very low angles of attack were not realistic as a result of the blade geometry. The results are shown in Fig. 13. On the left, the angle difference dependent cross correlation is shown for all nine screwdrivers separately and on the right, the mean and standard deviation of all nine graphs are shown in form of an error bar (blue). In previous work [3], we have shown that using only geometrical details of toolmarks in the wavelength range between 100 μm and 300 μm instead of using the entire range (20 – 2000 μm) for comparing toolmarks (of unused tools) leads to a slightly better separation of known match and known non-match distributions. Therefore, an additional error bar is shown in the figure, the similarity between profiles with structures in the
range 100 – 300 μm (red). The mean and standard deviation of the KNM similarities is shown for comparison.

Several observations can be made by examining Fig. 13. Most importantly, the similarity between profiles is high, even for large differences in angle of attack, and in many cases clearly above the similarities found for the known non-matches. In four out of nine cases, the similarity graphs are entirely above the KNM range for the complete range of angles. In seven cases, the KM similarities are above the KNM similarities down to roughly –35° angle difference and in eight cases, the KM similarities are above the KNM similarities up to roughly +28° angle difference. These results are in line with the results presented earlier [3], but in this work the angular resolution is much higher (1° vs. 15°). In general, profile similarity in the structural range between 100 and 300 μm is slightly lower, except of the area between –40° and –30°. While the mean similarity between the full structural range profiles falls within one standard deviation of the KNM similarity values, the mean similarity of the restricted range still lies clearly above.

Note that the amplitudes of toolmark profiles are decreasing with increasing angle of attack and that structures become less distinct. This can be observed in Fig. 10 but gets particularly clear in Fig. 12. The explanation for this is that the angle between the lateral face of the blade and the substrate material is small for high angles of attack (as defined in Fig. 1). However, the similarity values are high over a relatively large range of difference in the angle of attack. The reason is that cross correlation, used to calculate toolmark similarity, is relatively insensitive to amplitude changes, as long as the width of a structure is not affected.

Fig. 11. Left: Histograms of the cross correlation values of known matching and known non-matching toolmarks. Given are comparisons between two sets of real experimental toolmarks (KM = red, 0.958 ± 0.027 and KNM = blue, 0.227 ± 0.130) and between real and virtual toolmarks (KM = black, 0.929 ± 0.060 and KNM = green, 0.205 ± 0.128). Note that the KM distributions were scaled with factor 0.5 for presentation. Right: Histograms of the cross correlations between known matching real and virtual toolmarks, dependent on the angle of attack. (For interpretation of the references to color, the reader is referred to the web version of this article.)

Fig. 12. The profiles the screwdriver tip shown in the middle (for presentation purposes, two surface datasets of the same screwdriver were stitched together) would produce in a substrate material, depending on a particular angle of attack, are shown at the bottom. In total, 71 1D profiles (10°, 11°, . . . , 80°) are shown, which are stacked vertically. The profile heights are represented by the intensity of the image with white areas being relatively higher than black areas. The red lines indicate selected structures in the blade causing prominent ridges and furrows in the substrate material. The damage in the middle of the blade (inset in the middle) causes a ridge at low angles (white) but a furrow at high angles of attack (black). At about 50°, neither a ridge nor a furrow is created. (For interpretation of the references to color, the reader is referred to the web version of this article.)
3.3. Automated identification of the angle of attack

In the previous section we showed that realistic toolmarks can be produced using 3D surface datasets of a tool and how these profiles can be used to study the influence of the angle of attack on toolmark similarity. For these experiments, the angle of attack was known. In actual casework, the angle of attack usually is not known. Therefore forensic toolmark examiners typically create experimental toolmarks at many different angles of attack and assess, which exemplar shows the highest similarity with the suspect (unknown) mark. The examiner can then decide whether or not more experimental toolmarks should be made with a more restricted range of angles of attack. This procedure is not only time consuming but also bears the risk of damaging the tool, if the marks are created in a hard substrate material [1]. An alternative approach is to mimic this procedure and determine the most likely angle of attack automatically. To study how accurate an unknown angle of attack can be recovered by our automated virtual profile determination and profile comparison method, we used the data described in Section 3.1. To each experimental toolmark at all available angles, hence 15°, 30°, 45°, 60° and 75°, we registered all available profiles of the corresponding screwdriver surface datasets and retained the angle that yielded the highest profile similarity.

A histogram of the differences in angle of attack is presented in Fig. 14. It can be seen that our framework is capable of recovering the actual angle of attack with high precision and accuracy with a mean difference in actual and estimated angle of $0.618 \pm 4.179^\circ$ (max: 9; min: $-8^\circ$). Theoretically, another iteration could be run with higher resolution, say 0.1°, in the range close to the angle with highest similarity. However even if the actual angle is estimated with an error of $\pm(5 \ldots 10)^\circ$, the profile similarity values are still significantly higher (see Table 2) compared to the known non-matching cases ($0.227 \pm 0.103$). Alternatively the angle of attack can be determined within an automated angle optimization framework. This is particularly interesting if the analytic approach described in the appendix is used for profile determination, as the optimization only takes several seconds.

### Table 2

<table>
<thead>
<tr>
<th>Angle estimation error</th>
<th>KM similarity</th>
<th>KNM similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-5^\circ$</td>
<td>0.973 ± 0.048</td>
<td>0.227 ± 0.103</td>
</tr>
<tr>
<td>$+5^\circ$</td>
<td>0.982 ± 0.016</td>
<td></td>
</tr>
<tr>
<td>$-10^\circ$</td>
<td>0.924 ± 0.010</td>
<td></td>
</tr>
<tr>
<td>$+10^\circ$</td>
<td>0.935 ± 0.032</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 13. Cross correlations between a virtual profile at 45° angle of attack and virtual profiles at angles between $5^\circ$ and $85^\circ$ (relatively $-40^\circ$ to $+40^\circ$), for 9 screwdrivers individually (left) and as error bars (right). For the error bars two scenarios are considered: correlation between profiles with structures between the full wavelength range 20 μm and 2000 μm (blue) and with structures between 100 μm and 300 μm (red). The black lines indicate the mean and standard deviation of the KNM similarities ($0.227 \pm 0.103$).

Fig. 14. Histogram of differences between the real angle of attack at which an experimental toolmark was created and the estimated angle. The mean difference is $0.618 \pm 4.179^\circ$.

In case the angle of attack during acquisition of the surface data and during experimental toolmark creation is not exactly the same, it will show up as a systematic error in the difference between real and estimated angle. But as the framework yields high profile similarity values for a large range around the correct angle, the influence of a possible tool positioning error is very limited. Note that the precision is not affected.

3.4. Realistic 1D toolmark simulation

Naturally the question arises, how close to reality simulated profiles are. In order to address this question, three ($N=3$) mean

Fig. 15. Histograms of KM correlation scores for profiles made at a 45° angle. The black curve outlines the correlation scores between the simulated and the real profiles used to simulate them (training set). The red curve outlines the correlation scores between the simulated and the real profiles not used to simulate them (test set). (For interpretation of the references to color, the reader is referred to the web version of this article.)
profiles were generated from experimental toolmarks made at 45°. For a given screwdriver, one profile was held out while the remaining two were used to simulate 25 profiles. The held out set is referred to as the ‘KM Real Test set’. Each mean profile in the KM Real Test set was cross correlated against each of the simulated profiles from its known source screwdriver. The inter-profile shifting in the cross-correlation process took place over ±2000 profile points (≈0.9 mm). The red line in Fig. 15 outlines the maximum cross correlation scores obtained. The black line in the same figure outlines the histogram of the scores obtained with the real profiles used to generate the simulants (i.e. the ‘KM Real Training set’). As can be seen, the scores produced with the training set overlap well with the scores from the test set, indicating the simulation algorithm is producing simulants that are plausible representations of mean profiles made with the same screwdriver.

Next, in order to examine if the simulated profiles produce the same kind of KM/KNM separation of cross correlation similarity scores as the real profiles do, KM/KNM cross correlation scores distributions of real and simulated profiles were determined for angles of attack of 15°, 45°, and 75°. Fig. 16 shows that the KMN score distributions (red, orange and green) are well reproduced.

Examining the KMN score histograms across all of the angles of attack (right hand distribution of Fig. 16), the simulant vs. simulant (cyan) and real vs. simulant (blue) comparisons tend to have longer lower (left) tails than the corresponding real vs. real (magenta) comparisons. This manifests due to the fact that the correlation score cut-off for the retention of the simulants profiles with their real counterparts was set to 0.75 within the simulation algorithm, at all angles of attack. However the KMN cross correlation scores are usually not less than 0.9 (the real vs. real histograms; magenta curves, Fig. 16). The real vs. real KMN score distributions are generated by a relatively small number of real profiles. We set the simulation algorithm correlation score at 0.75 and not 0.9 to take into account the uncertainty in the (unknown) true real vs. real KM distributions, hence the fatter lower tails. This is also a more conservative approach.

4. Summary and conclusions

In this article a method is proposed to generate virtual toolmarks using high resolution surface datasets of screwdrivers and show how the toolmark profiles change with angle of attack. We also demonstrate that virtual toolmarks of a screwdriver can be used to estimate the angle of attack with which an experimental toolmark was created with high accuracy and precision (0.618 ± 4.179). Automated angle detection can save time in casework, since typically the angle of attack of crime scene marks is not known. In addition, the comparison framework yields relatively high similarity values for known matches, even for large differences in angle of attack during toolmark creation. We further demonstrate that a set of real toolmarks can be used to simulate toolmarks with properties that are similar to those of the real ones. With qualitative and quantitative experiments we show that virtual toolmarks of a screwdriver are very similar to the experimental toolmarks made with that particular screwdriver and that the simulated toolmarks are realistic.

The presented methods can be employed to artificially generate toolmark datasets with very high resolution, but without the need to create a large amount of experimental toolmarks. In addition, damages to the tools that might occur if many toolmarks have to be made can be avoided.

In future work we plan to combine the virtual toolmark generation method with the statistical properties of various substrate materials often encountered in casework, to predict how a mark of a tool will look like in a particular substrate material.

Appendix

Analytic approach for time efficient toolmark profile determination

In case the rotation angles around the y- and z-axes are within several degrees (please refer to Fig. 2 for the coordinate system), which usually can be obtained if some care is taken during the
placement of the model under the microscope, the approach for profile determination can be simplified and, most importantly, sped up significantly. Like with the general approach, the surface dataset has to be rotated such that the edge of the blade is outlined to the global x-axis and the front face is outlined to the xy-plane, as described in Section 2.3. Afterwards however, the surfaces points of the model are not used directly to determine the profile points along the x-axis by binning, but the model surface is re-sampled first to a regular grid with equal sampling distance, e.g.

using Lanczos3 interpolation [20]. Given a regular grid with size \( n \times m \) (xy-plane) and an angle of attack \( \alpha \), the resulting profile points for each location \( i = 1, \ldots, n \) along the x-axis can be determined as \( \text{max} (z'_{ij}) \), with \( j = 1, \ldots, m \) and \( z'_{ij} = \cos(\alpha) \times (z_{ij} + \tan(\alpha) \times y_{ij}) \). A schematic is presented in Fig. 17.

References