

Chapter 10 

Emerging Technology in Comparisons*

T.V. Vorburger¹, J. Song¹, N. Petraco², and R. Lilien³

¹ National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA

² City University of New York, John Jay College and The Graduate Center, New York, NY 10019, USA

³ Cadre Research Labs, Chicago, IL 60654, USA

Keywords: ballistics, breech face, bullet, cartridge case, exclusion, firearm, identification, individual characteristics, standards, surface, tool mark, topography

* Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

List of Acronyms:

ACCF	areal cross correlation function
AFTE	Association of Firearm and Toolmark Examiners
ASCLD/LAB	American Society of Crime Laboratory Directors/Laboratory Accreditation Board
ATF	Bureau of Alcohol, Tobacco, Firearms, and Explosives
CCF	cross-correlation function
CMC	congruent matching cells
CMS	consecutive matching striae
CSI	coherence scanning interferometry
CV	cross validation
DNA	deoxyribonucleic acid
DSCM	disk scanning confocal microscopy
IBIS	integrated ballistics identification system
ISO	International Organization for Standardization
LEA	land engraved area
LED	light emitting diode
LSCM	laser scanning confocal microscopy
NA	numerical aperture
NIBIN	National Integrated Ballistic Information Network
NIST	National Institute of Standards and Technology
PAM	programmable array microscope
PCA	principal component analysis
rms	root mean square
QC	quality control
SI	the international system of units
SRM	Standard Reference Material
SVM	support vector machine
VIM	International Vocabulary of Metrology, 3 rd edition (2012)

1. Introduction

This chapter is designed to introduce the reader to emerging technologies relating to 3D surface topography measurement and comparative analysis. While the chapter includes several technical terms and a few equations, effort has been made to omit content that only would be of interest to a metrologist. We encourage readers to push through the entire chapter. Technology is developing quickly and we hope that the content will provide a roadmap and some important waypoints to this emerging discipline. Note that this chapter has been adapted from a longer and slightly more technical paper. The interested reader is pointed towards the original content (Vorburger et.al. 2016).

This review describes some of the methods practiced and results accomplished thus far in the field. Gerules et al. published a broad review of methods for firearms analysis in 2013 (Gerules et.al. 2013). The current review focuses on topography methods with some illustrative examples and on recent work. The remainder of Section 1 provides a short history of ballistics identification systems. Section 2 describes surface topography measurement, and Section 3 describes analysis procedures and parameters, especially those to quantify similarity between surface topography images. Subsection 3.5 discusses the all-important issue of error rate estimation. Section 4 describes standards, notably physical standards and the concept of traceability. Information on the transition of these technologies into the crime lab, including interoperability, the X3P file format, and virtual comparison microscopy, is given in Section 5. Section 6 highlights a few ongoing issues and opportunities.

Since the early 1990s, commercial automated ballistics identification systems, such as the Drugfire (Roach 1997) and the Integrated Ballistics Identification System (IBIS) (Braga and Pierce 2004), have been developed, producing a revolution in the speed at which microscope inspections can proceed. These early systems served to partially replace physical open case files for database search and were not intended to be used for the reaching of conclusions regarding common origin. These systems typically include a digitized optical microscope to acquire 2D images of bullets and cartridge case surfaces, a signature analysis station, correlation software, and access to a large database where accumulated images reside. With such a system a large number of comparisons can be performed automatically. When a suspect image is input into the database, it is correlated with the images in the database, and a list of possible leading matches is output for further analysis by firearms examiners. These initial systems utilized 2D images, and therefore confirmation of an identified hit required examination of the original materials under a traditional comparison microscope.

Most of these original systems were based on comparisons of the optical intensity images acquired by the microscope. The quality of these 2D optical images is largely affected by lighting conditions, such as the type of light source, lighting direction, intensity, color and reflectivity of the material, and the image contrast. Since each of the images is acquired alone and not in a comparison microscope, the systems are more susceptible to slight variations in the alignment and lighting conditions. The significant effect of lighting conditions on the optical image has been discussed by Song et al. (Song, et.al. 2012) and A. Chu et al. (Chu, et al 2014)

Accurate identification also depends on the capability of the correlation software to identify the related correlation regions and to eliminate the unrelated regions from correlation. Most commercial systems

use proprietary correlation parameters and algorithms to quantify image similarity. These proprietary correlation methods often lack objective open tests of their parameters and algorithms. This may pose difficulty for laboratory assessments and inter-comparisons among different systems.

It was stated in the *“Theory of Identification”* issued by the Association of Firearm and Tool mark Examiners (AFTE) that *“...the comparison of tool marks...”* are to be made on the *“...unique surface contours...”* and *“surface contour patterns comprised of individual peaks, ridges and furrows. Specially, the relative height or depth, width, curvature...”* (SWGgun 2005) (Hamby 1999). Because ballistics signatures are geometrical micro-topographies by nature, direct measurement and correlation of the 3D surface topographies have been proposed for ballistic identification (DeKunder and Bonfanti 1999) (Weller et.al 2015) (Song and Vorburger 2006) (Vorburger et.al 2007). Such methods can avoid the confusing effects of variable lighting conditions and shadowing, and should likely improve correlation accuracy of automated systems. Several different types of optical instruments have been developed, which are capable of precise measurement of 3D surface topography. These methods will be discussed in Section 2. They are making it possible to use quantitative topography measurements for firearm evidence identifications, in addition to traditional methods based on conventional image comparisons. Development of ballistics identifications is therefore facing a likely evolution from qualitative image comparisons to quantitative topography measurements. Not all modern ballistic identification systems are making the transition to quantitative 3D measurements. Some systems are shifting their role from laboratory instruments to investigative tools for generating leads. These systems appear to be shifting away from quantitative 3D measurements. The reader is advised to evaluate each system for its intended use.

2. Topography Measurement

A number of different methods have been developed to measure surface topography. They may first be classified into three categories—line-profiling, area-integrating, and areal-topography—as described in an ISO standard (ISO 25178-6 2010). In this review we will emphasize areal topography, as it is the technology employed in all current ballistics identification systems. The surface topography features of cartridge cases and bullets, which are of interest to firearms experts, are generally in the micrometer to millimeter lateral range with heights in the submicrometer to hundred micrometer range. Even in this relatively narrow range there at least four different methods, most of them optical, which are useful and available as commercial instruments. Most of these methods are the subject of international documentary standards that outline the key properties and describe influence factors that are potential sources of uncertainty and error. Calibration procedures are currently the subject of further standardization efforts. In the following subsections, each method is presented with a bit of technical detail and references from which significantly more information can be obtained. If the reader skips these sections, note that there are multiple different methods for obtaining 3D surface topographies, that different ballistic identification systems utilize different methods, and that not every ballistics identification system utilizing the same method will obtain the same quality scans.

The first two described methods, confocal microscopy and coherence scanning interferometry are capable of collecting scans at high resolution, perhaps higher than is required for firearm and toolmark examination. They also require long scan acquisition times. Therefore while they may have been used by

academic research groups in the transition to 3D scanning technology, they may not end up in the crime laboratory.

2.1 Confocal microscopy

Confocal microscopy (Hamilton and Wilson 1982) is widely used, not only for fluorescence microscopy and 3D sectioning of transparent materials, but for the measurement of surface topography when used in reflection mode. A standards document, which describes confocal microscopy and its influence quantities has recently completed an ISO ballot as a final draft international standard (ISO FDIS 25178-607 2018). A schematic diagram of a typical confocal microscope is shown in Fig. 1 (ASME B46-2009 2010) (Weller et al. 2012). Most examples of this method rely on the use of pinholes for height discrimination. Incident light is focused through a pinhole, refocused onto the surface and reflected from it, then refocused through a conjugate pinhole placed before the detector. A strong signal through the pinhole will be detected only when the surface point is at the focusing height. This discrimination enables the tool to detect variations in surface height and topography when the surface is vertically scanned (Fig. 1) along the optical axis of the microscope. Variations in the method include laser scanning confocal microscopy (LSCM), disc scanning confocal microscopy (DSCM), and programmable array microscopy (PAM). (ISO FDIS 25178-607 2018) Different confocal microscopes have been used in a number of firearms and tool mark research studies (Vorburger et al. 2007) (Petraco et al. 2013) (Brinck 2008) (Bachrach et al. 2010). The vertical noise resolution and lateral resolution improve with the numerical aperture (NA) of the microscope. With a 50X objective, having a numerical aperture (NA) of 0.5, the vertical resolution can reach a few nanometers and the lateral resolution is on the order of a micrometer or less. A topography image of a fired cartridge case obtained with confocal microscopy is shown in Fig. 2 (Vorburger et al. 2007).

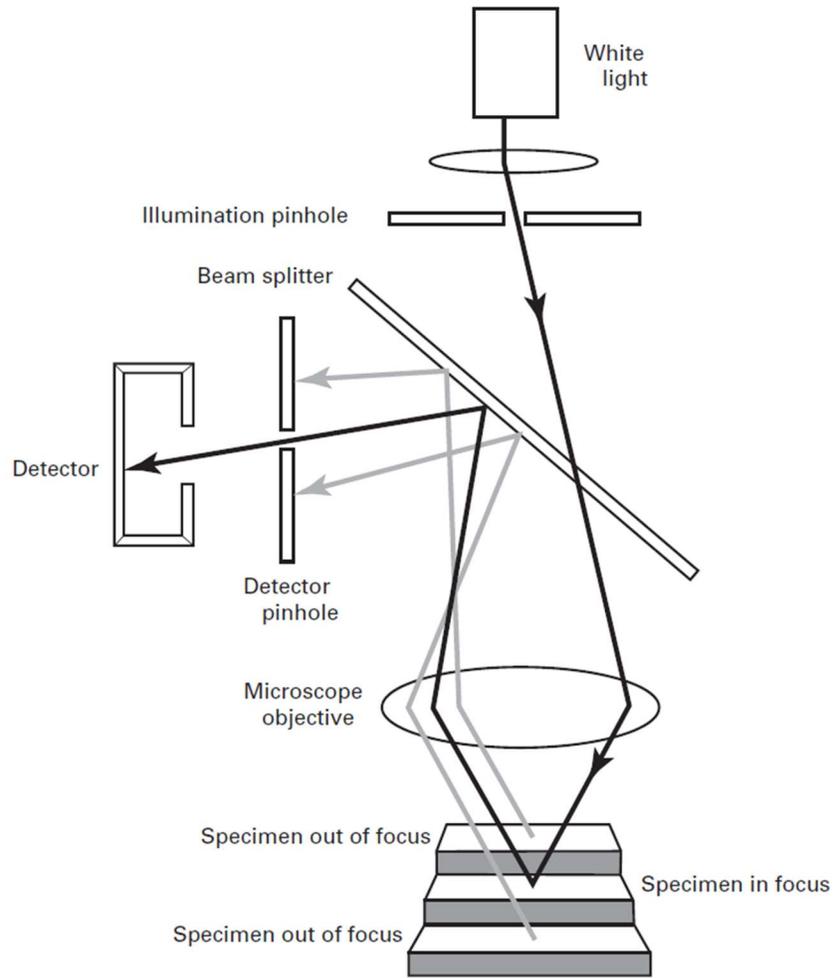


Figure 1. Schematic diagram of a confocal microscope for measuring surface topography (ASME B46-2009, 2010) (Weller et al. 2012).

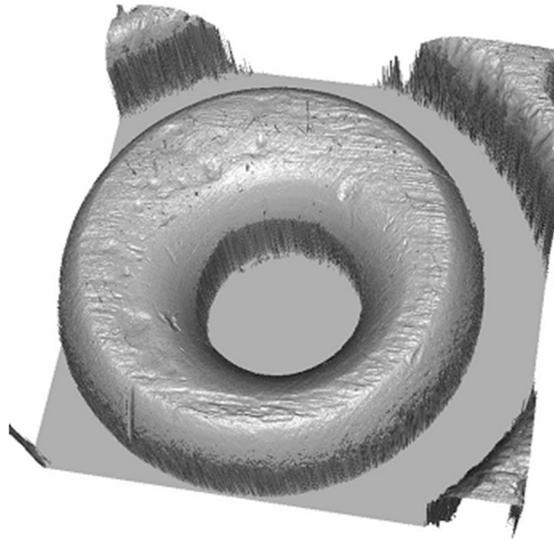


Figure 2. Topography image of the breech face impression of a fired 9 mm cartridge case obtained with disk scanning confocal microscopy (Vorburger et al. 2007). The field of view is roughly 4 mm on a side.

2.2 Coherence scanning interferometry

Coherence scanning interferometry (CSI) relies on interference between a beam of light reflected from the surface under study and a beam of light reflected from a reference surface. This method is the subject of a published standard (ISO 25178-604 2013) and other reviews (DeGroot 2011). A schematic diagram is shown in Fig. 3 (Bennett 1985). When the optical paths reflected from the reference surface and the test surface are equal, an interference pattern of bright and dark fringes is formed on the camera detector, but as either optical path is changed by distances larger than the coherence length of the light, the fringe contrast disappears. One can move the surface or the microscope vertically to observe a maximum in the signal modulation in order to locate the height of a surface point relative to its neighboring points. This depth localization process is repeated for all surface points. The vertical noise resolution is routinely a few nanometers but under some conditions can be as small as about 0.1 nm. The lateral resolution scales with the NA of the microscope in a manner similar to the confocal method.

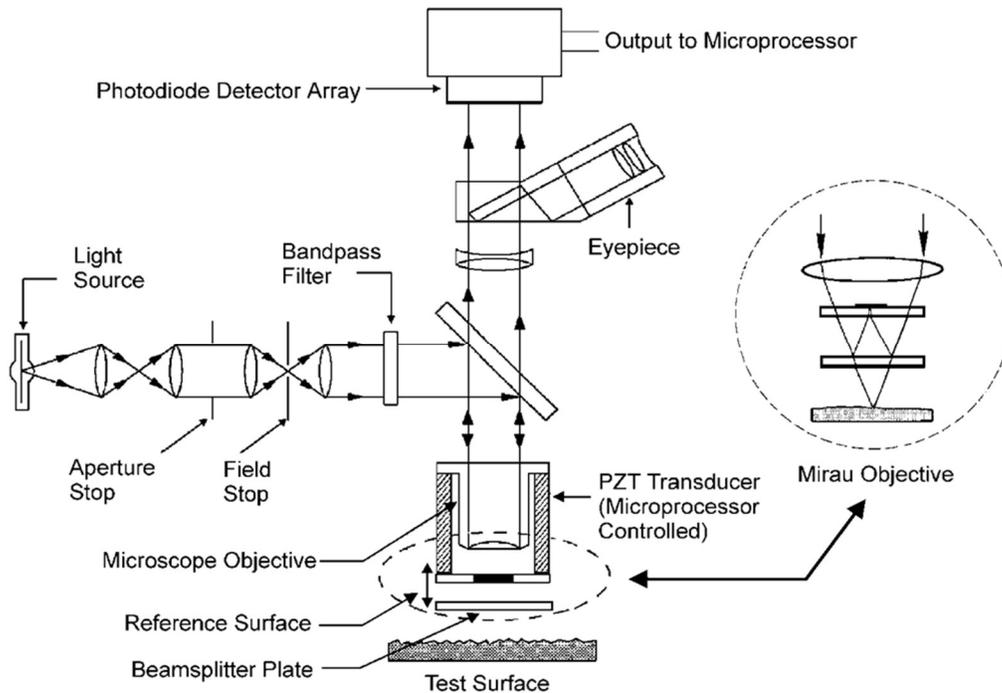


Figure 3. Schematic diagram of a coherence scanning interferometric microscope in the Mirau configuration (reprinted with permission) (Bennett, 1985).

2.3 Focus variation

Both confocal and CSI methods involve some manipulation of the light travelling through a microscope, either with pinholes or beam splitters. This leads to a cost in signal-to-noise. Focus variation (Fig. 4) is conceptually simpler (Helmlí 2011). The height sensing function derives from locating the surface at its sharpest, best focus position in the microscope. The peaks and valleys of the surface are focused at different positions as the surface scans vertically with respect to the microscope, a mode of operation similar to those of confocal and CSI. Focus variation is the subject of an International Standard (ISO 25178-606 2015). The method is capable of measuring steeply sloped surfaces, up to nearly 90° (Helmlí 2011). Because the method relies on contrast in images resulting from peaks and valleys of surface features, averaging of individual pixels is required to provide the height sensitivity, which involves a collective response from neighboring pixels as illustrated in Fig. 5 (ISO 25178-606 2015). This implies that both the lateral resolution and vertical resolution of the focus variation method may be more limited than those for confocal or coherence scanning. Focus variation has been favorably reviewed by Bolton-King et al. (Bolton-King et al. 2010).

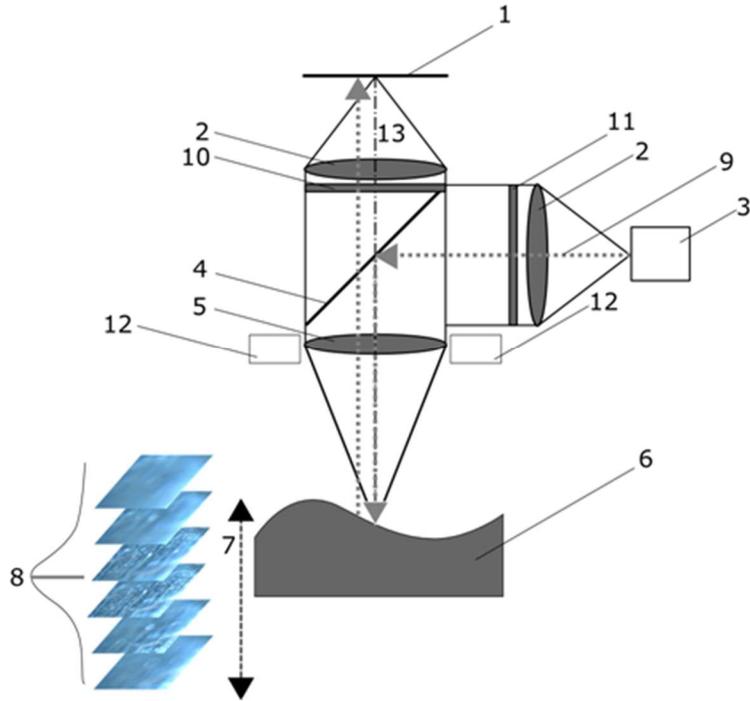


Figure 4. Schematic of a focus variation microscope. (1) camera sensor, (2) lenses, (3) light source, (4) semi-transparent mirror, (5) objective lens with limited depth of field, (6) sample, (7) vertical movement with drive unit, (8) contrast curve calculated from the local window, (9) light rays from the white light source, (10) analyzer, (11) polarizer, (12) ring light. Items 10-12 are optional (©ISO. This material is reproduced from ISO 25178-606:2015 with permission of the American National Standards Institute (ANSI) on behalf of ISO. All rights reserved.).

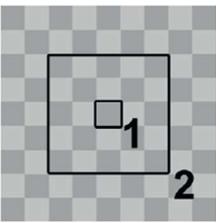
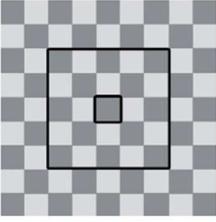
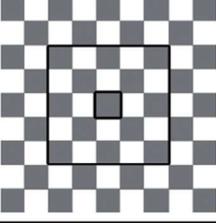
Scan position	Surface image	Standard deviation
Out of focus		10
Almost in focus		20
In focus		50
<p>1 Point of interest for which the focus information is calculated.</p> <p>2 5x5 neighbourhood of points used to calculate the focus information (standard deviation).</p>		

Figure 5. Calculation of focus information at a position of interest (1) using the contrast from a neighborhood of points (2). The contrast may be quantified by the standard deviation of the intensities of the neighboring points (©ISO. This material is reproduced from ISO 25178-606:2015 with permission of the American National Standards Institute (ANSI) on behalf of ISO. All rights reserved.).

2.4 Photometric stereo

Photometric stereo, also called shape from shading, involves the decoding of illumination patterns on surfaces cast by multiple light sources to produce a surface topography measurement. Depending on the number and directions of the light sources, this method can have different manifestations (Johnson 2011) (Sakarya et al. 2008). One of these is shown in Fig. 6 (Johnson 2011). Six light sources evenly spaced azimuthally illuminate the surface in turn at a grazing angle. The illumination patterns are analyzed and produce a surface topography image. The method illustrated here includes an additional technique, called GelSight, to reduce the sensitivity to variations in surface optical properties and to emphasize the surface topography. Integral to the setup is a soft, transparent gel with a pigmented film that directly contacts the surface. The film has uniform optical properties and a small grain size, which helps to minimize the effects of non-uniform surface reflectance properties (e.g., specularities). The microscope above the gel observes the illumination of the gel surface, which is itself reproducing the underlying surface topography. A topography image of the breach face impression of a unit of National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 2461 (Vorburger et al. 2014) obtained with GelSight photometric stereo is shown in Fig. 7 (Weller et al. 2015) (Lilien 2016).

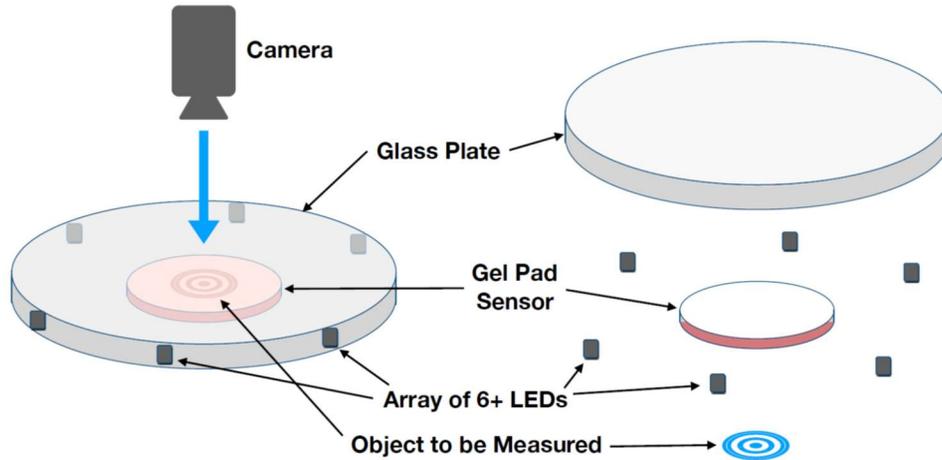


Figure 6. Schematic detail of a photometric stereo tool for measuring surface topography (assembled (left), separated (right)) (Weller et al. 2015) (Johnson et al. 2011) (Lilien 2016). Six or more LED light sources illuminate the rough surface of the object in turn at near grazing incidence angle. The Gel Pad sensor is a soft material with uniform optical properties that replicates the rough surface topography of the object when pressed down against it. The microscope between the glass plate and the camera is not shown.

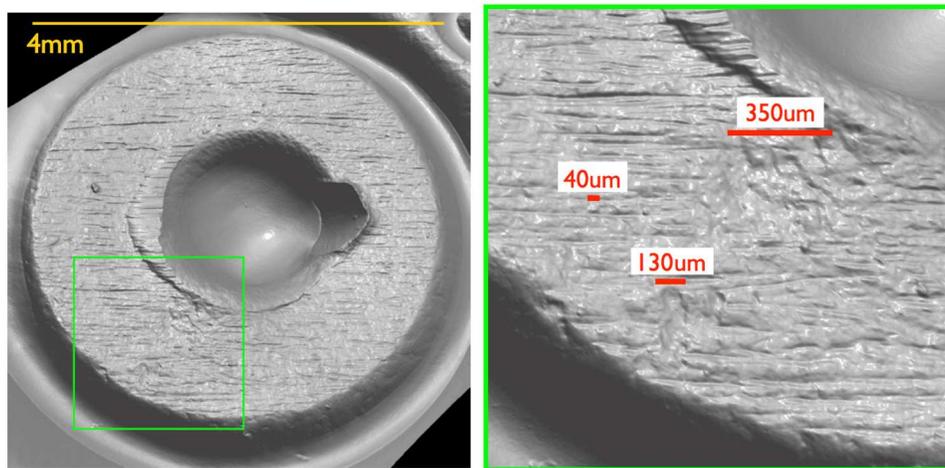


Figure 7. Topography image of the breech face impression of a unit of SRM 2461 Cartridge Case obtained with a photometric stereo tool (Lilien 2015a).

3. Analysis and Parameters

3.1 *The importance of similarity as a surface property in this field*

The function of establishing whether or not two bullets or cartridge cases were fired through the same firearm depends on obtaining some assessment of similarity between them. More specifically, we want to derive a quantitative measure of geometric similarity that will lead to identification or exclusion of them as being fired by the same firearm. To accomplish this task, the firearms examiner applies his/her expert judgment in a way that is difficult to quantify. An automated system, by contrast, must be programmed to produce a quantitative measurand for similarity, which the expert can use. Hence, much research in firearms identification is concentrated on finding algorithms and parameters that emphasize the individualized characteristics of surfaces and their similarity to those of other surfaces. The two most common ways of accomplishing this are to match a large section of one surface to that of the other, or to identify individual features on one member of a pair and look for similar features on the other. (Zitova and Glusser 2003) This section presents a vast simplification of each described analytic method. Analytic methods are rapidly evolving and the methods described below are continually being improved. Textbooks have been written on the component parts of each method described below. Additional details can be found in the provided references. This chapter is by no means a complete list of analytic methods. The intent of this chapter is not to make the reader an expert in comparison algorithms but rather to provide an overview of the types of methods currently in development.

Scoring functions can be developed to address two main comparison tasks. 1) a sorting function which accepts a single reference surface and a set of candidate surfaces and which sorts the candidate set to rank the surface scans from most to least similar as compared to the reference. 2) a scoring function which, given two surfaces, computes a statistically meaningful quantified measure of comparison. The quantified statistical comparison can take the form of a likelihood, an odds ratio, or an absolute probability, not merely a ranking. However, each of these statistical measures presents their own challenges.

In Section 3.2 we describe an important required step prior to surface comparison. In Section 3.3 we introduce the basic cross correlation function. In Sections 3.4 and 3.5 we move onto more sophisticated comparison approaches. While these methods continue to evolve, the presentation in this section should provide a framework onto which other, new, methods can be understood.

3.2 *Preprocessing*

Preprocessing is an important part of any surface measurement and analysis methodology. The details of each preprocessing step involve significant math and are beyond the scope of this chapter. Preprocessing may include the following steps. Decimation (or downsampling) may be performed to reduce the number of data points, for example, to reduce file size and to speed potential calculations. Bad data in the form of dropouts (i.e., unmeasured points) and outliers (i.e., mismeasured points) must be recognized and ignored or repaired. Filtering is the process of removing select structures (e.g. mismeasured baseline drift) and is often performed to emphasize individual characteristics. Two common filters typically operate to remove very coarse or very fine features. These coarse and fine

features are often referred to as low and high spatial frequencies respectively. Any preprocessing utilized by a scanning system should be done in a way that it does not affect the useful information within the measured surface. In other words, preprocessing should not change the underlying toolmark information content and should not push the toolmarks into erroneous identification or elimination. It is important that practitioners utilize established preprocessing techniques supported by the literature. The rest of this section describes several details of the filtering process.

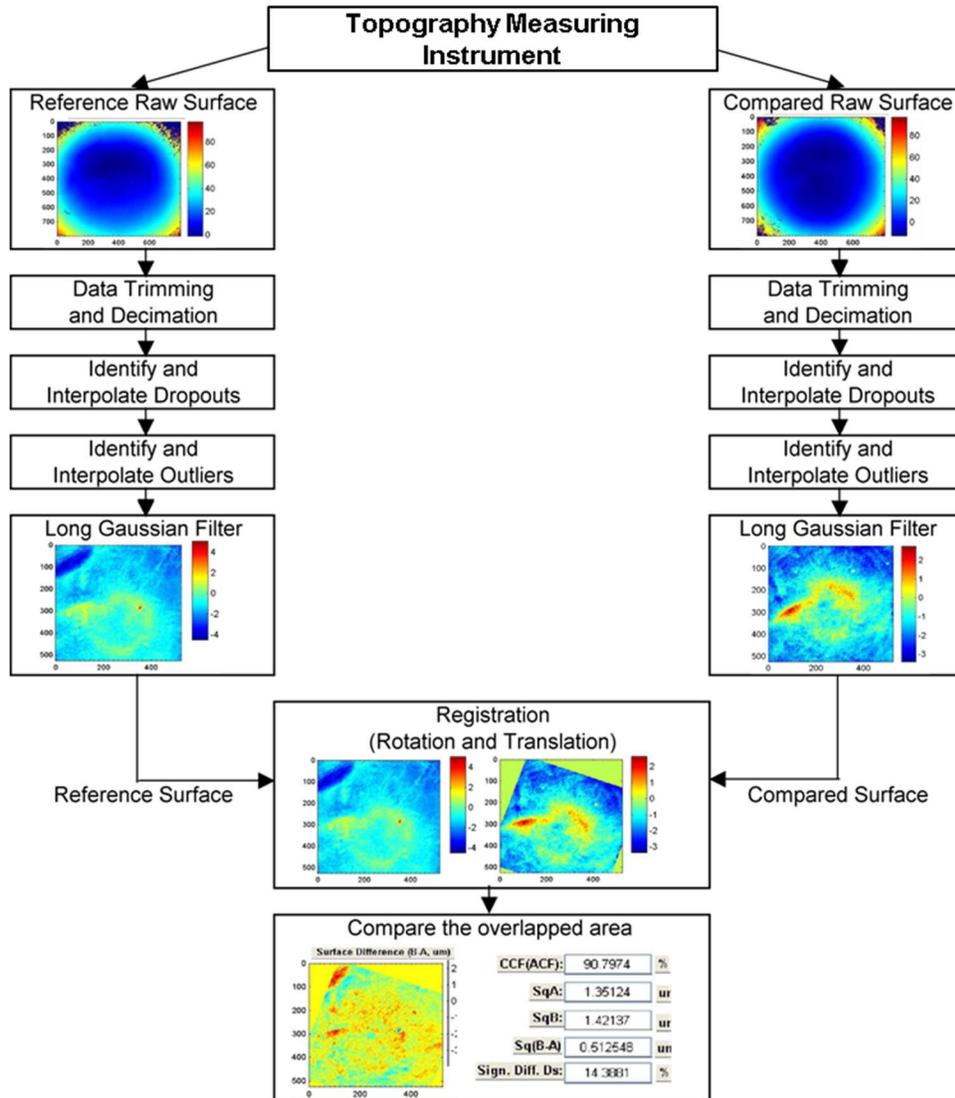


Figure 8. Illustration of a procedure for assessing the similarity of two topography images: dropout and outlier detection, filtering, registration, analysis and parameters. In addition to the long scale filtering operation (shown) a short scale or smoothing filter may also be applied (not shown). (Vorburger et al. 2007)

One of the most common filtering approaches is the digital Gaussian filter (ASME B46-2009 2010) (ISO 11562 1996). The Gaussian filter is a kind of moving-average, smoothing filter, where the moving average

window uses a Gaussian weighting function. The smoothed profile that results can be subtracted from the original profile to produce a profile where the long wavelength features are diminished. The scale of features that are diminished or eliminated is given by the long cutoff wavelength (ISO 25178-2 2012). Conversely, if one wants to remove short wavelength (high frequency) noise, the Gaussian smoothing filter may be applied with a short cutoff wavelength. Combining these two processes gives us a desired Gaussian bandpass filter defined by long and short cutoffs. Figure 9 illustrates how a filtered profile might appear. Figure 9a shows a segment of a longer profile containing the sum of three sinusoidal components: a waviness component with a wavelength of 1000 μm , a roughness component with a wavelength of 100 μm and a noise component with a wavelength of 4 μm . We wish to emphasize the 100 μm roughness component and attenuate the other two. Applying a Gaussian filter with a short wavelength cutoff of 25 μm attenuates the noise component by about 94 % while leaving the roughness and waviness components attenuated by less than 0.5 % (Fig. 9b). Applying a second Gaussian filter with a short wavelength cutoff of 250 μm attenuates the roughness component by about 98.7 % but attenuates the waviness component by only 4.2 % (Fig. 9c). Subtracting Fig 9c from Fig 9b yields a relatively unattenuated roughness component while severely attenuating the waviness component (Fig. 9d). It is of course important to not “throw out the baby with the bathwater” when filtering. Filters must preserve structural elements relevant to identification.

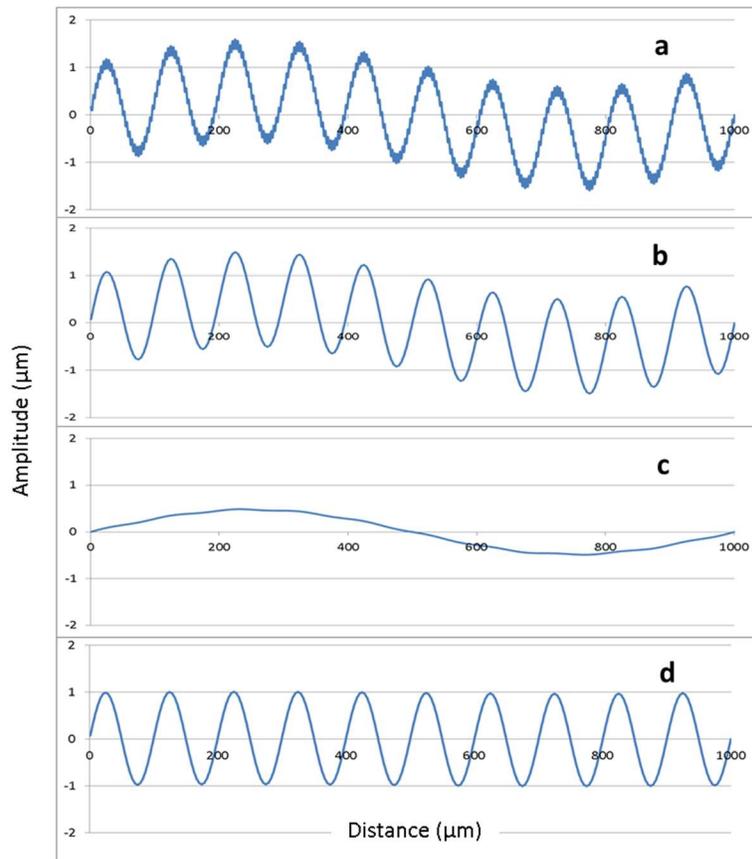


Figure 9. Illustration of a bandpass Gaussian filter; (a) segment of original profile with three sinusoidal components; (b) 25 μm short wavelength filter attenuates the noise component; (c) 250 μm short wavelength filter attenuates the roughness component; (d) subtracting c from b emphasizes the roughness component and attenuates the waviness component.

An important limitation of the basic Gaussian filter is the sensitivity of the filtered result to peaks and valleys in the data, which may not be of interest to the user. (Blunt and Jiang 2003) For these and other reasons, a wide number of other filtering methods have been developed and defined in documentary standards.

Both impressed and striated marks can be preprocessed using filtering. In addition, striation patterns are often summarized by mean profiles, which are often averaged along the surface in the direction of tool travel (Bachrach et al. 2010) (Faden et al. 2007) (Chumbley et al. 2010) (Chu et al. 2010) (Bachrach 2002). These mean profiles represent the cross-sectional linear profile of the striated mark.

3.3 Basic Similarity parameters for topography measurements

One of the oldest and most common similarity measures is the cross-correlation function (CCF) and the areal cross-correlation function (ACCF). The CCF can be used when comparing linear striated profiles (e.g., a bullet's land area) and the ACCF can be used when comparing areal impressed toolmarks (e.g., a cartridge cases's breech-face impression). Unless specified, when we mention the CCF we are referring to both the CCF and ACCF. The basic idea of these methods is to identify the best alignment between two surfaces and to quantify the degree of corresponding pixel-to-pixel similarity at that orientation. For a given alignment of two surfaces the CCF compares each corresponding (overlapping) pixel. If the surfaces are the same (e.g., identical) we would expect the same values at each corresponding pixel. That is, where one surface has a height of 3.2 micrometers, the second surface should also have a height of 3.2 micrometers, and where one surface has a height of 1.2 micrometers, the second surface should have a height of 1.2 micrometers. Mathematically, the equations below measure this similarity. A maximum CCF score occurs where each corresponding pixel has the same value. The use of averages (e.g., means) and variances simply ensures that the maximum value of CCF for perfectly correlated surfaces be 1, and the maximum value of CCF for similar but not identical surfaces, should be near 1. Dissimilar surfaces have a CCF value near 0. During the development of NIST's Standard Reference Material (SRM) bullets (NIST 2013), Song et al. used the CCF to quantify the similarity of bullet signatures (Ma et al. 2004) (Song et al. 2004). The cross correlation function between two surface profiles $z_A(x)$ and $z_B(x)$ may be calculated by

$$CCF(A,B,\tau) = \lim_{L \rightarrow \infty} \left(\frac{1}{L} \int_{-L/2}^{L/2} z_A(x) z_B(x + \tau) dx \right) / [R_q(A) R_q(B)], \quad (1)$$

Where τ is a shift distance between the profiles, and $R_q(A)$ and $R_q(B)$ are the root mean squared (rms) roughness values of the two profiles in the region of overlap. The cross correlation function for areal topographies may be calculated by

$$ACCF = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{[(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)]^{1/2}}, \quad (2)$$

where the two arrays A_{mn} and B_{mn} here are the digitized surface topography images $z_A(m,n)$ and $z_B(m,n)$, and m and n represent indices in the x and y directions. Equation 2 is the discrete form of Eqn. 1 extended to three dimensions. The CCF is computed as a function of displacement of two profiles, and

the ACCF is computed as a function of alignment (displacement or rotation or both) of two surfaces. A curve is often generated to show the ACCF or CCF score as a function of alignment position, such as the relative displacement between two striation profiles. The terms, CCF_{max} and $ACCF_{max}$ are used to represent the largest correlation achievable for any alignment of the profiles or surfaces.

The CCF parameter is not a unique parameter for topography comparison because CCF is not sensitive to vertical scale differences. If two profile signatures A and B have exactly the same shape but different vertical scales, their CCF_{max} is still 1. A parameter, called the signature difference, D_s , is useful for quantifying both scale and shape differences between profile or topography signatures A and B. It may be calculated as the normalized rms amplitude of the difference profile or difference topography image. For example,

$$D_s = R_q^2(B - A) / R_q^2(A), \quad (3)$$

where $R_q^2(A)$ is the mean square roughness of the reference signature $z_A(x)$, used here as a comparison reference. When two compared profile signatures are exactly the same, $D_s = 0$. In this way, D_s is a complementary parameter to CCF .

Weller et al. used the CCF_{max} parameter to compare topography images to identify spent cartridge cases from the same firearm slides (Weller et al. 2012). They started with ten 9 mm Luger caliber slides that were consecutively manufactured and that revealed both subclass characteristics and individual characteristics. This set of slides should be especially difficult to distinguish one from another. They obtained nine test fires from each slide, measured the topography of the breech face impression of all 90 cartridge cases, and performed cross-correlation calculations for the 8010 combinations of pairs. There were 7290 non-matching pairs, i.e., fired from different guns and 720 matching pairs. A graph of their results is shown in Fig. 10. Although this set of consecutively manufactured slides potentially contains subclass characteristics, which could persist from one firearm to another, there is good separation between the cross-correlation values for the matching pairs and the non-matching pairs.

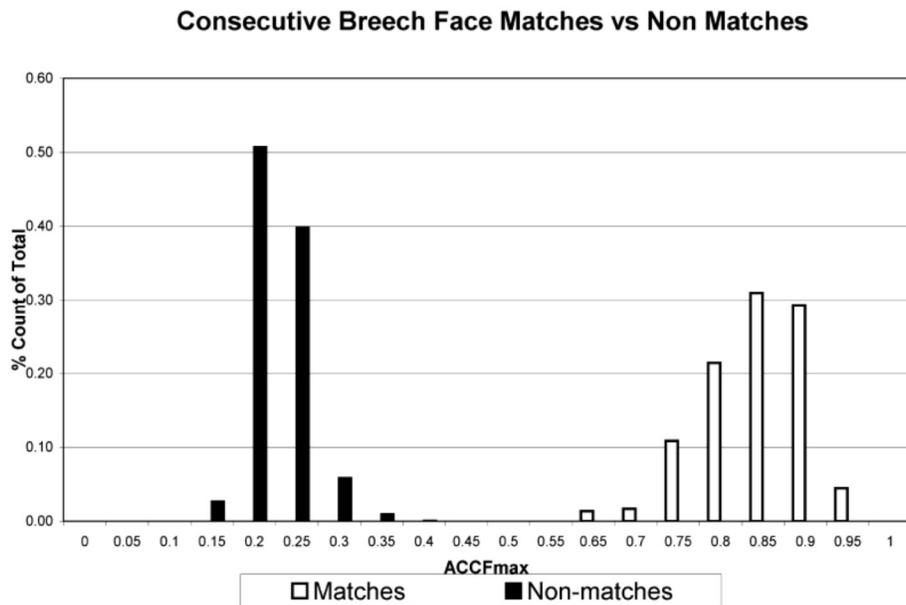


Figure 10. Data from Weller et al. (Weller et al. 2012) showing cross-correlation comparisons using the CCF_{max} parameter among 90 test fires from ten consecutively manufactured breech faces. No overlap of data was observed between matching (same breech face) and nonmatching (different breech face) comparisons (©2012 American Academy of Forensic Sciences, reproduced with permission of John Wiley and Sons).

Several additional comparison methods have been developed many of which involve sophisticated mathematics and computational algorithms. One method, closely related to CCF_{max} , which has been proposed for quantitative comparison is Chumbley et al.'s "T1" statistic. (Chumbley et al. 2010) Their method takes pairs of striated tool mark profiles and searches for a region of best agreement (as measured by a correlation coefficient) within a user-defined window.

Methods like the CCF and the difference profile which rely on comparison of the entire surface may work well when the entire surface is reliably reproduced from test fire to test fire. However, when reliable individual marks only appear on a portion of the measured surfaces these methods may have difficulty identifying this similarity. That is, if only a small portion of two surfaces are similar, then a score dependent on the entire surface will be lower than desired.

3.4 Advanced Similarity Methods

3.4.1 Congruent matching cells (CMC)

Song has developed an analytical method that seems to improve on the basic approach of correlating entire images (Song 2013). The method systematically divides measured 3D forensic images into "correlation cells", and uses cell correlation instead of correlation of the entire images. This is done because a firearm often produces characteristic marks, or individual characteristics, on only a portion of the surface. If a quantitative measure of correlation is obtained from the entire areas of a pair of images, the correlation accuracy may be relatively low because some *invalid regions* may be included in the correlation (Chu et al. 2010) (Chu et al. 2013). If instead, the correlation areas are divided into cells, the *valid regions* can be identified and the invalid regions can be eliminated. The use of a sufficiently large number of cells may provide a statistical foundation for estimating error rates from a well characterized population. Typically, there may be 7 x 7 cells in an image of a breech face impression and on the order of ten thousand pixels in a cell.

The CMC method works as follows. If topographies A and B originating from the same firearm are registered at their position of maximum correlation (Fig. 11), the registered cell pairs located in their common valid correlation regions, as shown by the solid cell pairs located in (A_1, B_1) , (A_2, B_2) , and (A_3, B_3) , are characterized by:

- 1) High pairwise topography similarity as quantified by a high value of the cross correlation function maximum CCF_{max} ;
- 2) Similar registration angles ϑ ; and
- 3) Similar x-y spatial distribution patterns.

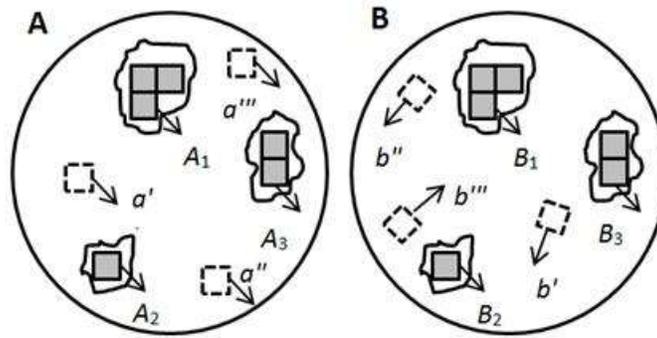


Figure 11. Schematic diagram of topographies A and B originating from the same firearm and registered at the position of maximum correlation. The six solid cell pairs are located in three valid correlated regions (A_1, B_1), (A_2, B_2), and (A_3, B_3). The three dotted cell pairs (a', b'), (a'', b''), and (a''', b''') are located in the invalid correlation region.

On the other hand, if the registered cell pairs are located in the invalid correlation regions of A and B, such as the dotted cells (a', a'', a''') and (b', b'', b''') in Fig. 11, or if they originate from different firearms, their maximum cross correlation value CCF_{max} would be relatively low, and their cell arrays would show significant differences in x - y distribution patterns and registration angles ϑ .

Congruent matching cell (CMC) pairs are therefore determined by four criteria, which must be satisfied simultaneously. The correlation value CCF_{max} must be larger than a chosen threshold T_{CCF} and the registration angle ϑ and x, y registration positions are within the chosen threshold limits T_ϑ, T_x and T_y , respectively.

A fifth criterion is the number of matching cell pairs required to satisfy the above criteria in order to decide that two images are truly matching overall. W. Chu et al.'s initial results for a set of breech face impressions suggested that a pattern of six matching cells was a sufficient identification criterion for pairing up the breech face impressions that were studied (Chu et al. 2013). Thus, when the number of CMC pairs of the correlated topographies A and B is equal to or greater than 6, A and B are concluded to be a match. Significant work has gone into developing a family of methods based on the CMC method. These approaches all utilize the same idea of identifying small patches of similarity, comparing them using a cross-correlation, and then quantifying the number of sufficiently similar regions.

3.4.2 Principal component analysis (PCA)

An example of an alternative to the cross-correlation approach is the multivariate *machine learning* scheme discussed by Petraco et al. (Petraco 2013) (Petraco 2012) (Petraco 2011). A tool mark surface contains a tremendous amount of information. Most of the information is lost in summarizing the surface with a single number (i.e. a single univariate similarity metric). Instead, the machine learning approach derives a set of values to characterize surfaces. These vectors of *features* can be standard surface parameters or any other numerical or categorical values that potentially discriminate one surface from another, assuming that the surfaces are generated from different sources (ASME B46-2009 2010) (ISO 4287 1997).

For the system constructed by Petraco et al., pre-processing first involves dropout/outlier interpolation. The surfaces are filtered into roughness and waviness components via the methods and standards outlined in section 3.2. Registration with a quick cross-correlation between pairs of profiles is performed to find translations that yielded maximum, though not necessarily high, similarity (areas of overhang are padded with zeros) (Petraco et al. 2013) (Petraco et al. 2012) (Gambino et al. 2011). Next, feature extraction is performed to produce feature vectors of the surfaces. Petraco then automatically extracts a set of features by applying principal component analysis (PCA) to a set of mean profiles. PCA is a mathematical method for analyzing points to identify ways in which they co-vary. It is often used to reduce the spatial dimensionality of points while maintaining desired mathematical properties. Intuitively, PCA transforms mathematical representations of the surface features to simplify their representation and to make these representations amenable to further analysis.

Once a feature set is chosen, the data is split randomly into training and testing sets. Machine learning algorithms are “trained” to recognize tool marks in the training set with a high probability. The training is essentially a model fitting procedure with many methodologies to choose from. When a machine learning scheme is selected and fit, the discrimination functions are applied on the test set in order to estimate an overall error rate.

Choices must be made concerning the discrimination algorithm to be used and the method to assess intermediate error during the training/fitting process. Petraco et al. have found that the support vector machine (SVM) discrimination algorithm combined with PCA and hold-one-out cross-validation is a balanced machine learning scheme for forensic tool mark discrimination (Petraco et al. 2013) (Petraco et al. 2012) (Petraco 2011) (Gambino et al. 2011). An SVM is a mathematical method for determining efficient decision rules in the absence of any knowledge of probability densities (Vapnik 2013). Two surfaces can be compared using these rules to reach a ‘decision’ regarding common origin. This procedure produces linear decision making rules for identification, while minimizing the risk of error.

3.4.3 Feature-Based Methods

Recently, Lilien completed a development study of a commercial firearms identification system comprised of 1) a photometric stereo system with Gelsight imprinting for measuring the surface topography of breech face impressions and 2) a feature based system for characterizing the surface signatures and identifying matches (Weller et al. 2015) (Johnson et al. 2011) (Lilien 2016). The basis of the comparison algorithm is the automated identification of 3D geometric features on each cartridge case surface. These features range in size from a few micrometers to hundreds of micrometers in diameter. When comparing the measured 3D surfaces of two cartridge cases, the algorithm looks for similar arrangements of similar features. The system was tested in cooperation with the Oakland and San Francisco police departments. One of the tests involved 47 firearms of the 9 mm Luger type, and three test fires for each firearm. A round robin comparison of all test fires should produce 141 different matches among more than 19,000 possible combinations. Lilien’s software found 111 correct matches at the number one ranked position and with the criterion that the match score be greater than a certain threshold (i.e., that the algorithm should be confident). Notably, there were no false positives among the chosen matches. Lilien also developed a procedure to calculate a confidence level for these matches and claimed confidence levels of 99.99 % or higher for 102 of the matches found. Figure 12 shows a “confusion matrix” that plots the match scores as shades of gray for all comparisons. The overall array

shows 141 x 141 comparisons. Cartridge cases fired by the same gun form close-knit 3 by 3 arrays straddling the central diagonal. Roughly 23 firearms stand out as highly identifiable, such as the one indicated by the blue arrow. Roughly nine firearms are much more difficult to identify, such as the one indicated by the red arrow, where the comparison of different cartridge cases from the same gun appear to give results that are indistinguishable from non-matches in this chart. Entries exactly along the diagonal are trivial cases where a single image is compared with itself.

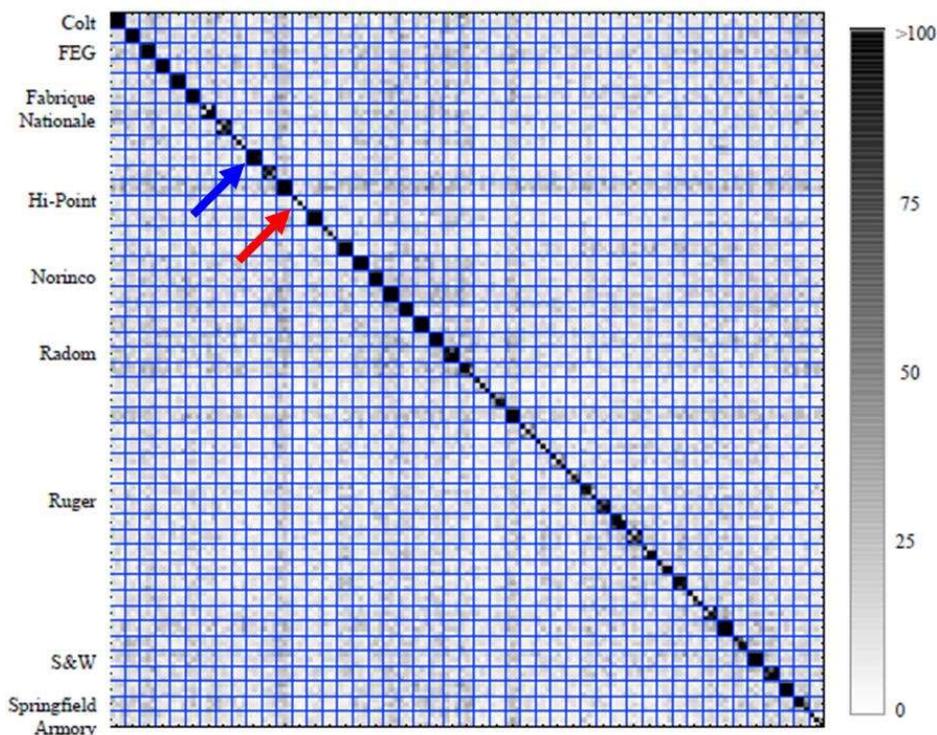


Figure 12. Results by Lilien of 19 981 comparisons among 141 cartridge cases (3 each from 47 firearms). “Each cell in the matrix corresponds to the match score between two casings (specified by the involved row and column).” The firearms are separated in the matrix by blue lines. All cartridge cases fired by the same firearm are grouped into 3x3 cells along the main diagonal. The blue arrow indicates an example where the separation of matches is well differentiated from non-matches. The red arrow indicates an example where very little differentiation of matches from non-matches is occurring (originally published by the National Institute of Justice, U.S. Department of Justice) (Lilien 2015a).

3.5 Statistical Error Rate Estimation

Reporting an error rate for firearm identification—that is, the probability that an identification is actually a false positive or the probability that an exclusion is actually a false negative—has been singled out as a fundamental challenge in forensic science (National Research Council 2008) (National Research Council 2009) (President's Council of Advisors on Science and Technology (PCAST) 2016). Coincidentally, there is much debate in the community about the desired output of a comparison algorithm. For example, there

is growing consensus that a comparison algorithm should output a statistical measure quantifying the support for common origin and not a discrete conclusion of Identification, Inconclusive, or Elimination. In this scenario the algorithm would report a likelihood ratio, probability, or coincidental match rate.

Central to any statistical model is the assembly and analysis of toolmark surfaces from a complete range of toolmark types. In other words, it is important to study and quantify the degree of geometric similarity and difference seen among known matches and known non-matches for different firearm makes/models, calibers, manufacturing methods (e.g., granular, milled, broached, filed), ammunition type, ammunition material, and ammunition condition. Assembling such data is not an easy challenge. It is estimated that test fires from more than ten thousand firearms would need to be examined to establish this information. Work towards this critical research goal is underway.

3.5.1 Bayesian Statistics

Bayesian statistics are based on Bayes' rule which defines the probability of a hypothesis in the presence of observed data,

$$p(H|D) = p(D|H) * p(H) / p(D)$$

Where H is the hypothesis and D is the data. p(H) is the probability of the hypothesis being true in the absence of any data, p(D) is the probability of observed the data under any condition, p(D|H) is the probability of observing the data given that the hypothesis is true, and p(H|D) is the probability of the hypothesis being true given the observed data. We note that Bayes' rule is a mathematical truth and is not up for debate. However, what is unsettled is the numerical value of each term in the context of firearm forensics. Let us describe each of the four terms in additional detail. Consider that H is the hypothesis that two test fires were fired through the same firearm and that D is the computed (algorithm) match score between the two 3D topographies of these test fires.

- p(H) is also known as the "prior" probability and represents the probability (between 0% and 100%) that the test fires came from the same firearm before you've even looked at the scans and before we know the similarity score D.
- p(D) represents the probability (between 0% and 100%) that any pair of cartridge cases have a similarity score of D regardless of whether the two test fires came from the same or different firearms.
- p(D|H) is also known as the "likelihood" and represents the probability (between 0% and 100%) that a pair of cartridge cases fired through the same firearm obtain a similarity score of D. It is called the likelihood because it measures how likely one is to see the data D given hypothesis H.
- p(H|D) is also known as the "posterior" probability and represents the final desired probability (between 0% and 100%) that the two test fires came from the same firearm after (e.g., posterior) you have observed (e.g., computed) the similarity score of D.

Of these four terms we want to get values for the three on the right side of the equation so that we can compute the value p(H|D) on the left. Unfortunately, it's very difficult to compute some of these terms.

For example, $p(D|H)$ is computable given a large population of firearms. We would simply assemble an extensive set of firearms, collect pairs of test fires from each firearm, and compute the distribution of similarities we observe. We can then see how often a score of D arises. Unfortunately it is not so easy to estimate $p(D)$ and $p(H)$. Remember that $p(H)$ represents the probability that the two test fires came from the same firearm before you've looked at the evidence. If the total population of firearms of the target class is say 50 million, is $p(H)$ simply 1 in 50,000,000? This may provide a lower bound; but in reality not all 50 million firearms were in the same city where the crime took place so $p(H)$ is probably a bit higher than that. Similarly, $p(D)$ depends on the number of firearms of the target class.

Despite the difficulty of estimating these terms, all is not lost. Conservative estimates can be made for each term, which provides a lower estimate (a lower bound) on the desired $p(H|D)$. Another approach is to compute a likelihood ratio which represents the ratio of probabilities between the two competing alternatives. In our example the two alternatives are, (H) that the two test fires came from the same firearm and $(\text{not } H)$ that the two test fires came from different firearms. The likelihood ratio is thus $P(D|H) / P(D|\text{not } H)$. In other words, the probability of observing similarity score D given that the two test fires were fired through the same firearm divided by the probability of observing similarity score D given that the two test fires were fired through different firearms. Recall that both these likelihoods can be estimated using a sufficiently large sample set of test fires.

The use and implications of Bayesian probabilities and likelihood ratios goes beyond the scope of this book. Because they can be estimated with less difficulty, these quantities are generally used instead of $p(H|D)$ to quantify the degree of evidence and support for a hypothesis, but both require careful application and appropriate data estimation to be utilized correctly. We expect that researchers will build these models into software that can be used by the firearms examiner. The models should be supported by large peer-reviewed studies capable of validating their accuracy. In the future it's possible that an examiner will compare two pieces of evidence, reach their conclusion, and then submit the scans to an automated scoring function to provide a quantified degree of statistical confidence which can be included in their final report.

4. Standards, traceability, and uncertainty for topography measurements

Instrument calibration and measurement traceability is important when performing comparisons of surface topography, especially if the topography images to be compared are generated by different instruments. For example, the CMC method, which divides images into small cells and uses pairwise cell correlations, quantifies both the topography similarity of the correlated cell pairs and their pattern congruency. Both metrics are based on geometrical topography measurement with traceability to the SI unit of length.

According to the *International vocabulary of metrology (VIM)*, metrological traceability is defined as “property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (BIPM 2012) (see also ASCLD/LAB 2013).

In light of the above definition, three key steps for establishing metrological traceability and quality assurance for the topography measurements and imaging correlations of ballistics signatures have been proposed (Song et al. 2010):

- The establishment of reference standards for topography measurements,
- A chain of comparisons relating topography measurements of the reference standards to topography measurements of bullets, cartridge cases, and toolmarks, and
- The estimation of uncertainty in the measured quantities and/or the estimation of error rates in classifications and firearms identifications based on topography measurements.

We confine the discussion of these issues primarily to topography profiles and images.

4.1 Physical standards

Physical and documentary standards are critical for maintaining control in surface topography measurements.

Over the years, crime laboratories have implemented quality control (QC) bullets and cartridge cases for testing the accuracy and reproducibility of their surface imaging systems. These are bullets and cartridge cases fired from a single firearm, which is kept in the central laboratory as a reference and which may be typical of firearms recovered during investigations. This firearm could be used successively over time to provide artifacts (QC bullets) for different laboratories or at different times. However, the QC bullets could have problems with uniformity and traceability. In the late 1990s, the ATF expressed the need for physical standards that would be more stable over time and more reproducible. In response, NIST developed Standard Reference Material (SRM) bullets and cartridge cases, SRMs 2460 and 2461 (NIST 2013), respectively (Fig. 13). These highly reproducible standards enable users of optical imaging and topography measuring systems to test the quality and stability of their systems from time to time and from one place to another.

For topography measuring systems, master profiles and topography images of the standard bullets and cartridge cases, respectively, are available online for downloading and correlation with users' own topography measurements (Bui and Vorburger 2007). For crime labs participating in the ATF's NIBIN with IBIS optical imaging systems (Ultra Electronics, accessed 2019), the ATF maintains *Golden Images* of bullets and cartridge cases, acquired with IBIS workstations, to which NIBIN users can correlate their own acquired images (Song et al. 2012) (Vorburger et al. 2014). Users with other types of optical systems may develop their own Golden Images using the SRMs as well.

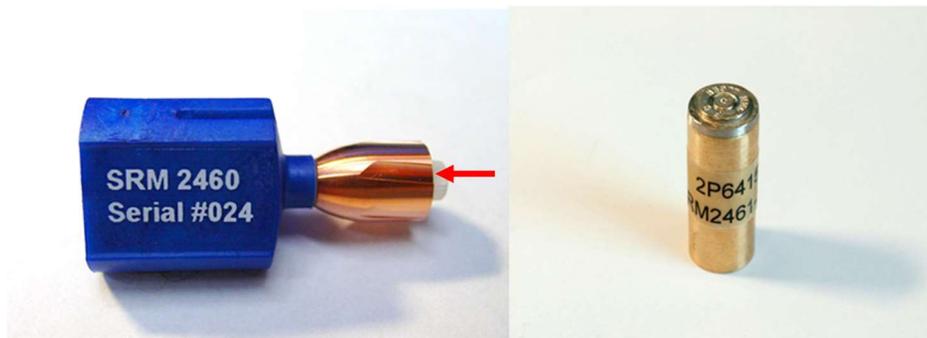


Figure 13. A SRM 2460 Standard Bullet (left) and a SRM 2461 Standard Cartridge Case (right). The red arrow indicates one of six land engraved areas around the periphery of the standard bullet.

4.2 A chain of comparisons

An example flow diagram for the establishment of a Traceability and Quality System using the SRM materials is shown in Fig. 14 (Vorburger et al. 2014). For topography profiles and images, we emphasize the right side of the chart. The topographies of the SRM bullets are nearly identical to one another as are the topographies of the SRM cartridge cases. These similarities are quantified by the cross correlation maximum and the fractional difference parameters quoted on the SRM certificates of calibration. Most of the units of the SRMs are made available to industry, and a few are held at NIST as check standards for NIST’s own topography measurement quality control. Since 2003, one of them, SRM 2460, Serial No. 001, land engraved area (LEA) 1, has been routinely measured and correlated with a NIST master topography image more than 35 times and has demonstrated high measurement reproducibility: all the correlation values CCF_{max} are higher than 99% (Song et al. 2012).

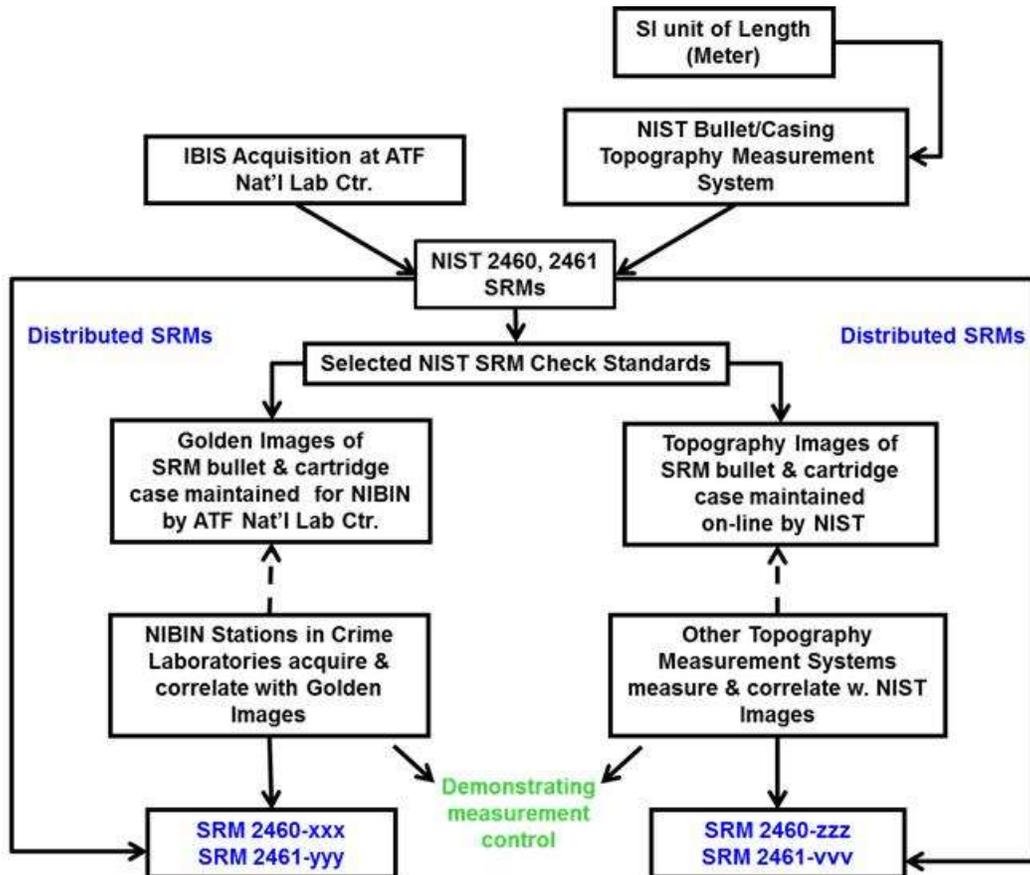


Figure 14. Establishment of a Traceability and Quality System for NIBIN acquisitions and correlations.

Topography images of the master surfaces are available online and may be downloaded for correlation. These include the profiles of all six LEAs of SRM 2460 Standard Bullet masters and master topography

images of the breech face impression, firing pin impression, and ejector mark of the SRM 2461 Standard Cartridge Case. By correlating measurements of the user's own SRM with the master profiles or images, the user can provide evidence that his/her topography measurements are accurate and that the user's system can measure bullet and cartridge case surfaces similar to those of the SRM standard. Control charts can be used to further demonstrate that the system is stable over time (Song et al. 2012) (Vorbürger et al. 2014).

4.3 Uncertainty and error rate

The issue of uncertainty in topography measurements of bullets and cartridge cases largely amounts to the specific task of calculating an error rate for making identifications and exclusions, about whether there is a common origin for a pair of surfaces, using topography data and software analysis. The usual approaches to calculating uncertainties in the measured properties of a single object do not apply when two surfaces are compared for their similarity. Quantifiers of similarity between them need to be established as well as uncertainties in those quantifiers. For conventional, open parameters of similarity, such as cross correlation and relative difference, the results are unitless and traceability to SI units is not relevant (Ma et al 2004). Calculation of uncertainty and error rate for ballistic evaluations is still an evolving research issue.

We make the following observations about uncertainty using cross correlation as an example of a similarity metric. Sources of measurement error are likely to reduce the calculated cross correlation between two measured topography images, not increase it. If two topographies are measured by the same instrument, systematic sources of error are likely to cancel out. If so, they would not change the accuracy of the result. If they do not cancel out, the resulting errors in a series of correlations is likely to lead to variations in the results that can be recognized as statistical uncertainty. If two topographies are measured by different instruments and even more so by different methods, errors in either measurement lead again to reduced correlation values. Since errors of measurement generally lead to reduced correlation, we do not expect these errors to cause a decision error when a *positive identification* between two surfaces is made based on correlation results. However, if the correlation results suggest a choice of *exclusion* or *inconclusiveness*, the probability of error should be estimated.

5. Ballistics Identification Systems in the Crime Lab

5.1 Virtual Comparison Microscopy

The analysis of the microscopic features of a measured 3D surface topography on the computer and without physical access to the original specimen is referred to as Virtual Comparison Microscopy (VCM) or Virtual Microscopy (VM) (Senin et al. 2006). A full discussion of VCM is beyond the scope of this chapter; however, we will point out a few advantages and novel uses of VCM as compared to traditional light comparison microscopy (LCM).

First, VCM allows instant access to remote or historic data without the need to requisition or physically transfer the items to your possession. This is a significant advantage when casework requires the analysis of specimens from a different physical location (e.g., requesting a specimen from a different laboratory within the same state system). All that is required for comparison is the digital data file which can be transferred without the risk of damage to the original specimen.

Second, VCM offers advantages for training, proficiency testing, and validation studies. It allows for training examples to be shared worldwide, providing trainees at any location the opportunity to see the most useful examples of specific or rare phenomena. Proficiency tests and validation studies can be conducted using a core set of scanned samples, thereby eliminating test-set to test-set variability which is inherent to physical tests and studies. In addition, tests can more easily be 'injected' into an examiner's digital casework thereby eliminating the potential effects of knowing that one is being tested.

Third, VCM can improve lab efficiency of verifications and blind verifications. Using VCM, verifications can be conducted remotely by examiners at another location. VCM verification can be truly blind in that all information from the first examiner's analysis can be hidden from the verifier.

Current studies are validating the use of VCM for firearm and toolmark analysis. Duez et al. conducted a study with 46 qualified firearms examiners and 3D surface topographies collected on a Cadre TopMatch 3D scanning system (Duez et al. 2018). The study design included two test sets each with three knowns and four unknowns. Participants were asked to complete a comparison worksheet and to annotate the surfaces of provided topographies to indicate regions of similarity and difference (Fig. 15). This annotation map highlights the regions of the surface that were marked as similar while making an identification. The reported conclusions of all 46 firearms examiners (368 of 368 conclusions) were correct.

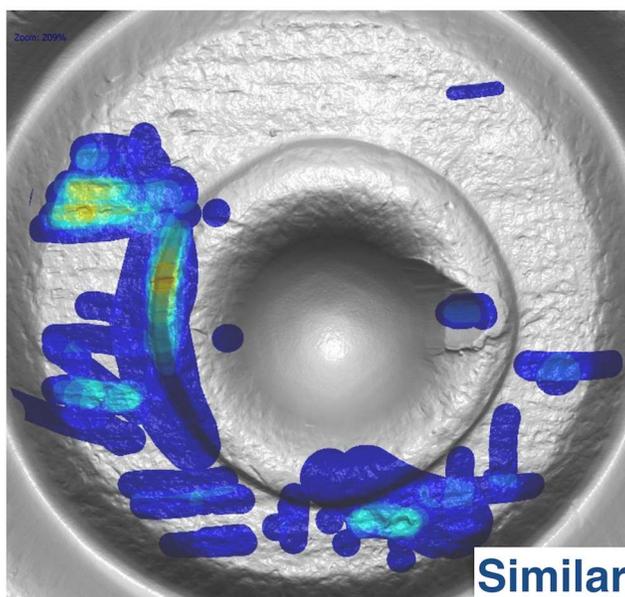


Figure 15. Similarity Annotation Map from the study of Duez et al. Regions in dark blue were marked as similar individual characteristics by a small number of participants. Regions in yellow and red were marked by almost all participants. Images like this can provide insight into the examiner’s decision making process and can serve as teaching points if mistakes are made.

5.2 Data Exchange

In theory, any system which measures surface topographies in standard units should be able to exchange data with any other system which does the same. In 2014 the Open Forensic Metrology Consortium (OpenFMC), a group of academic, government, and industry researchers agreed to support the use of the X3P file format for the exchange of 3D surface topography data in firearm and toolmark analysis. The X3P datafile is an ISO standard (ISO 25178-72:2017) that can be implemented by any vendor, researcher, government agency, or interested party. X3P is simply a container for the efficient and accurate transfer of 3D surface topographies. It has been used to exchange data between measurement systems within multiple disciplines (including firearm and toolmark analysis). X3P is intended to be extended for application-specific use. OpenFMC has created a firearm specific data record which allows an X3P file to contain firearm specific meta-data such as firearm make/model, imaged region of interest, caliber, cartridge case material, primer material, bullet diameter, bullet weight, etc. While it is not required to use this firearm specific data record, it is OpenFMC’s hope that vendors and labs will adopt it for data exchange. Vendors participating in OpenFMC have expressed this intent. Virtually all vendors making equipment for 3D imaging within firearm forensics are part of the OpenFMC group.

5.3 Algorithm Requirements

Comparison algorithms should meet three criteria to be used within a crime lab. First, the score should be statistically grounded in that numbers should be consistent where a score of k in two different runs should provide the same amount of confidence and support. Algorithms that are consistent in this way can form the basis of a statistical model. Second, the algorithm should be explainable in that a firearms examiner should be able to describe in a few sentences the basic principles on which it works. This explanation should not include PhD-level scientific detail nor should it include equations. Finally, the algorithm should be interpretable in that the algorithm should be able to explain, typically through visual representation, evidence to support the reported score. For example, algorithms have recently been developed for visually representing regions of geometric similarity through colored shading (Fig. 16) (Lilien 2015b) (Ott et al. 2017).

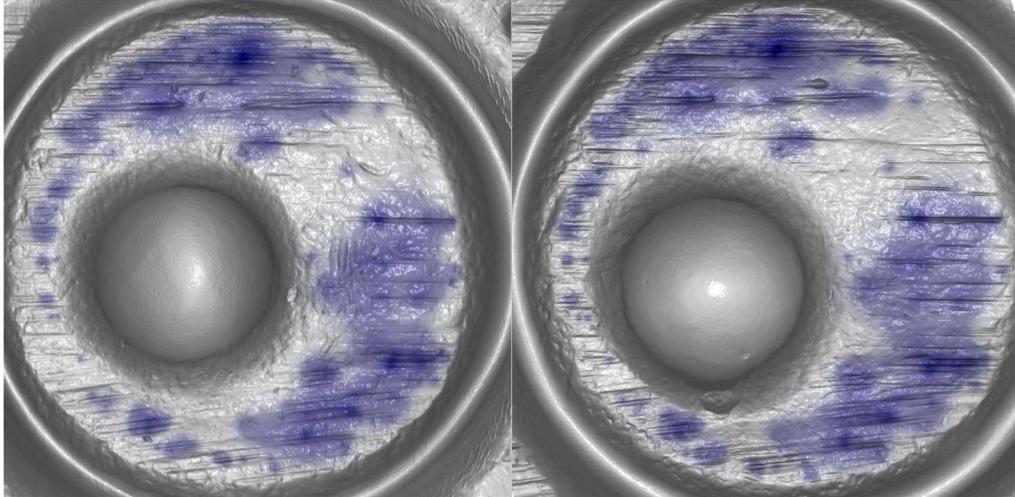


Figure 16. Heatmap visualization. The comparison algorithm can produce images like that shown here where the identified geometric similarity between two cartridge cases is highlighted in blue. Darker blue indicates more geometric similarity and unshaded regions indicate that the algorithm did not find similarity (Lilien 2015b).

5.4 Standards and Guidelines

In 2014 NIST and the Department of Justice formed the Organization of Scientific Area Committees (OSAC) for the establishment of standards and guidelines for forensic science. The Firearms and Toolmarks subcommittee organized within the Physics and Pattern Interpretation committee is creating standards to ensure accurate and high quality results for 3D topographic analysis including VCM. When these standards are published they will provide excellent guidance to those interested in utilizing this new technology. It will be highly recommended that labs comply with the approved standards.

5.5 Shift from the Lab

There is a recent shift where some systems are moving their focus from the crime laboratory to early analysis within police departments. That is, the vendor is shifting the use of the ballistics identification system from a high-end laboratory instrument to a rapid investigative tool. We advise readers which are part of crime labs to keep an eye on this transition and to consider 3D measurement systems which calibrate against known reference standards, which comply with OSAC and AFTE standards, and which support the X3P file format. Through the use of these standards one can guarantee the quality and accuracy of the measured 3D surface topography.

6. Ongoing Issues and Opportunities

This review has been largely concentrated on the emerging field of surface topography measurements and analysis for ballistic surfaces and tool marks and aimed to provide useful information to surface metrologists and ballistics examiners in their common field of interest. Whether virtual comparison microscopy methods come to rival and outstrip the usefulness of conventional optical microscopy will likely depend on several factors:

Outliers and Dropouts

Methods for measuring 3D surface topographies often produce dropouts (or non-measured points) and outliers. These erroneous point measurements must be identified as such by the measurement system software. A number of statistical methods have been used to discern and minimize the effect of these erroneous data points in the stored data. However, a standard approach for fired ballistics and tool marks may need to be defined in order to promote interoperability among topography images obtained with different optical methods.

Speed

Many of the optical topography methods discussed here require the collection of lots of images, perhaps 1000, as the surface is scanned through different heights relative to the microscope housing. Some of these systems therefore require significant time to measure an entire surface. One should inquire about and be aware of the imaging times of various systems.

Expense

Currently, most topography measuring systems cost significantly more than conventional microscopes—roughly speaking, the one costs more than \$100 000 and the other, less than \$100 000.

Measurement Accuracy and Resolution

As described above, not all 3D measurement systems collect data at the same sampling resolution, and pixel size is not always representative of the size of features which can be resolved. That is, not all 3 micron per pixel sampling scanners are the same. In addition, some systems may display a higher resolution 2D image on top of a 3D surface measured with lower resolution, providing the illusion of a 3D surface measured at higher resolution than is actually the case. Measuring the surface against known microscale references is one way to assess the resolution of these systems.

As of this writing it is not yet known what measurement resolution is required for accurate human and algorithmic based comparisons. As the topographic resolution decreases, the surface loses fine features which may be critical for analysis. For example, at lower resolution fine striated marks such as those on a bullet land or a cartridge case aperture shear may be lost. In place of fine striations one may only see a blurred region. Topographies should be captured at a resolution that allows human and algorithm based comparison at a level comparable to or better than traditional LCM.

Uncertainty

Topography methods coupled with advanced statistical analyses have finally provided an opportunity to address the question of uncertainty in firearm and tool mark identifications. Several case studies are beginning to calculate an error rate for identification and exclusion of matching surfaces. In some cases those error rates have been impressively small even for consecutively produced barrels or slides or tools. However, the most advanced work has so far been performed on small databases or on small collections of firearms or other materials. Scaling up the models to large databases like the NIBIN and

adjusting the statistical model to produce believable error rates for real criminal cases is a major challenge and a major opportunity for researchers. Once accomplished, such a development will pave the way to calculating error rates for firearms identification for real court cases, first as an independent approach to support conclusions drawn by firearms experts using comparison microscopes, and possibly afterwards, to stand on its own as admissible evidence in court in a manner similar to DNA evidence.

7.0 Conclusions

This chapter provided a high-level introduction to the emerging field of 3D topographic imaging and analysis within the firearm and toolmark discipline. Although the focus of this chapter was not on vendor specific ballistic identification systems, to the best of our knowledge all systems currently on the market utilize one or more of the described scan-acquisition and analysis methods. Rather than attempting to drill down into too many specific studies, which will likely be replaced by newer results by the time you read this chapter, we attempted to provide the framework for the technology which is less likely to change.

Labs will soon transition to the use of virtual comparison microscopy. The first step of this is already underway and crime labs are beginning to validate 3D virtual microscopy for use by qualified firearms examiners. In some cases, if a conclusion can be reached from the 3D topography then it is not necessary to go back to examine the physical evidence. The second step will be the development of statistical functions capable of providing an examiner a quantitative statistical measure to support reported conclusions. We do not see the role of the firearm examiner going away anytime soon as a human still needs to lead the analysis, perform the comparison, and reach a conclusion. The computational algorithm will support the examiner's report with a quantified number.

It is important that any 3D system utilized in forensic work conduct a series of performance checks designed to assess its accuracy both before and during active use. That is, all systems should be calibrated against known microscale reference standards as described in the previous section to ensure the accuracy of the measurement system.

To ensure interoperability and data exchange between labs, it is critical that all systems and all labs support the ISO standard X3P file format for the exchange of surface topography data when it has been collected in standard units.

Acknowledgements

The authors are grateful to B. Muralikrishnan, J.H. Yen, and M.E. Nadal for their careful review of the manuscript and to X.A. Zheng (Fig. 1), F. Helmlí (Figs. 4 and 5), and T.J. Weller (Fig. 10) for use of their illustrations. The work was supported in part by the NIST Special Programs Office.

References

ASCLD/LAB, "Measurement Traceability Policy American Society of Crime Laboratory Directors / Laboratory Accreditation Board," 2013, [Online]. Available: <https://anab.qualtraxcloud.com/ShowDocument.aspx?ID=6536>

ASME B46-2009, "Surface Texture (Surface Roughness, Waviness, and Lay)," American Society Mechanical Engineers, New York, 2010.

B. Bachrach, "Development of a 3D-Based Automated Firearms Evidence Comparison System," *Journal of Forensic Sciences*, vol. 47, pp. 1-12, 2002.

B. Bachrach, A. Jain, S. Jung and R. Koons, "A Statistical Validation of the Individuality and Repeatability of Striated Tool Marks: Screwdrivers and Tongue Groove Pliers," *Journal of Forensic Sciences*, vol. 55, pp. 348-357, 2010.

J. Bennett, "Comparison of Techniques for Measuring the Roughness of Optical Surfaces (figure adapted from)," *Optical Engineering*, vol. 24, pp. 380-387, 1985.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, "International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM)," 2012, 3rd Edition. [Online]. Available: <https://www.bipm.org/en/publications/>.

R. Bolton-King, J. Evans, C. Smith, J. Painter, D. Allsop and W. Cranton, "What are the Prospects of 3D Profiling Systems Applied to Firearms and Toolmark Identification?", *AFTE Journal*, vol. 42, pp. 23-33, 2010.

L. Blunt and X. Jiang, *Advanced Techniques for the Assessment of Surface Topography*, London: Kogan Page Science, 2003.

A. Braga and G. Pierce, "Linking Crime Guns: The Impact of Ballistics Imaging Technology on the Productivity of the Boston Police Department's Ballistics Unit," *Journal of Forensic Sciences*, vol. 49, pp. 701-706, 2004.

T. Brinck, "Comparing the Performance of IBIS and BulletTrax-3D Technology Using Bullets Fired Through 10 Consecutively Rifled Barrels," *Journal of Forensic Sciences*, vol. 53, pp. 677-682, 2008.

S. Bui and T. Vorburger, "NIST Surface Metrology Algorithm Testing System (SMATS)," *Precision Engineering*, vol. 31, pp. 218-225, 2007.

W. Chu, J. Song, T. Vorburger, J. Yen, S. Ballou and B. Bachrach, "Pilot Study of Automated Bullet Signature Identification Based on Topography Measurements and Correlations," *Journal of Forensic Sciences*, vol. 55, pp. 341-347, 2010.

W. Chu, M. Tong and J. Song, "Validation Tests for the Congruent Matching Cells (CMC) Method using Cartridge Cases Fired with Consecutively Manufactured Pistol Slides," *AFTE Journal*, vol. 45, pp. 361-366, 2013.

A. Chu, D. Read and D. Howitt, "A Computerized Database for Bullet Comparison by Consecutive Matching," 2014. [Online]. Available: <http://nij.gov/publications/Pages/publication-detail.aspx?ncjnumber=247771>.

L. Chumbley, M. Morris, M. Kreiser, C. Fisher, J. Craft, L. Genalo, S. Davis, D. Fadern and J. Kidd, "Validation of Tool Mark Comparisons Obtained Using a Quantitative, Comparative, Statistical Algorithm," *Journal of Forensic Sciences*, vol. 55, pp. 953-961, 2010.

P. DeGroot, "Coherence Scanning Interferometry," in *Optical Measurement of Surface Topography*, R. Leach, Ed., Berlin, Springer, 2011.

J. DeKinder and M. Bonfanti, "Automated Comparisons of Bullet Striations Based on 3D Topography," *Forensic Science International*, vol. 101, pp. 85-93, 1999.

P. Duez, T. Weller, M. Brubaker, R. Hockensmith and R. Lilien, "Development and Validation of a Virtual Examination Tool for Firearm Forensics," *Journal of Forensic Sciences*, vol. 63, no. 4, pp. 1069-1084, 2018.*

D. Faden, J. Kidd, J. Craft, L. Chumbley, M. Morris, L. Genalo and J. Kreiser, "Statistical Confirmation of Empirical Observations Concerning Tool Mark Striae," *AFTE Journal*, vol. 39, pp. 205-214, 2007.

C. Gambino, P. McLauchlin, L. Kuo, F. Kammerman, P. Shenkin, P. Diaczuk, N. Petraco, J. Hamby and N. Petraco, "Forensic Surface Metrology: Toolmark Evidence," *Scanning*, vol. 33, pp. 272-278, 2011.

G. Gerules, S. Bhatia and D. Jackson, "A Survey of Image Processing Techniques and Statistics for Ballistic Specimens in Forensic Science," *Science & Justice*, vol. 53, no. 2, pp. 236-250, 2013.

J. Hamby, "The History of Firearm and Tool Mark Identification," *AFTE Journal*, vol. 31, pp. 266-284, 1999 (Revised April 2008).

D. Hamilton and T. Wilson, "Three-Dimensional Surface Measurement Using the Confocal Scanning Microscope," *Applied Physics*, vol. 27, pp. 211-213, 1982.

F. Helmlí, "Focus Variation Instruments," in *Optical Measurement of Surface Topography*, R. Leach, Ed., Berlin, 2011.

ISO 4287, "Geometrical Product Specification (GPS) - Surface Texture: Profile Method - Terms, Definitions, and Surface Texture Parameters," International Organization for Standardization, Geneva, 1997.

ISO 11562, "Geometrical Product Specifications (GPS) - Surface Texture: Profile Method- Metrological Characteristics of Phase Correct Filters," International Organization for Standardization, Geneva, 1996.

ISO 25178-6, "Geometrical Product Specification (GPS) - Surface Texture: Areal: 6. Classification of Methods for Measuring Surface Texture," International Organization for Standardization, Geneva, 2010.

ISO 25178-2, "Geometrical Product Specification (GPS) - Surface Texture: Areal: Part - Terms, Definitions, and Surface Texture Parameters," International Organization for Standardization, Geneva, 2012.

ISO 25178-604, "Geometrical Product Specification (GPS) - Surface Texture: Areal: 604. Nominal Characteristics of Noncontact (Coherence Scanning Interferometry) Instruments," International Organization for Standardization, Geneva, 2013.

ISO 25178-606, "Geometrical Product Specification (GPS) - Surface Texture: Areal: Part 606 - Nominal Characteristics of Non-Contact (Focus Variation) Instruments," International Organization for Standardization, Geneva, 2015.

ISO FDIS 25178-607, "Geometrical Product Specification (GPS) - Surface Texture: Areal: Part 607 - Nominal Characteristics of Non-Contact (Confocal Microscopy) Instruments," International Organization for Standardization, Geneva, 2018.

ISO 25178-72, "Geometrical Product Specifications (GPS) - Surface Texture: Areal - Part 72: XML File Format X3P," International Organization for Standardization, Geneva, 2017.

M. Johnson, F. Cole, A. Raj and E. Adelson, "Microgeometry Capture Using an Elastomeric Sensor," *ACM Transactions on Graphics*, vol. 30, pp. 46:1-46:8, 2011.

R. Lilien, "Applied Research and Development of a Three-Dimensional Topography Analysis System for Firearm Identification Using GelSight," National Institute of Justice, Doc. No. 248639, Washington DC, 2015a.

R. Lilien, "Applied Research and Development of a Three-Dimensional Topography System for Imaging and Analysis of Striated and Impressed Tool Marks for Firearm Identification using GelSight", NIJ Grant Number 2013-R2-CX-K005, 2015b.

L. Ma, J. Song, E. Whinton, A. Zheng and T. Vorburger, "NIST Bullet Signature Measurement System for Reference Material (RM) 8240 Standard Bullets," *Journal of Forensic Sciences*, vol. 49, pp. 649-659, 2004.

National Research Council, "Ballistic Imaging," National Academies Press, Washington DC, 2008.

National Research Council, "Strengthening Forensic Science in the United States," National Academies Press, Washington DC, 2009.

NIST, "SRM 2460: Standard Bullet," 2013. [Online]. Available: https://www-s.nist.gov/srmors/view_detail.cfm?srm=2460

NIST, "SRM 2461: Standard Cartridge Case," 2013. [Online]. Available: https://www-s.nist.gov/srmors/view_detail.cfm?srm=2461

D. Ott, R. Thompson, and J. Song, "Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests", *Forensic Science International* vol. 271, pp 98–106, 2017.

N. Petraco, *Color Atlas of Forensic Tool Mark Identification*, Boca Raton: CRC Press, 2011.

N. Petraco et al., "Application of Machine Learning to Tool Marks: Statistically Based Methods for Impression Pattern Comparisons (NCJ 239048)," National Institute of Justice, Washington DC, 2012.

N. Petraco et al., "Estimates of Striation Pattern Identification Error Rates by Algorithmic Methods," *AFTE Journal*, vol. 45, pp. 235-244, 2013.

President's Council of Advisors on Science and Technology (PCAST), "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods," PCAST, Washington DC, 2016.

R. Roach, "Drugfire and IBIS Help Lawmen Fight Bad Guys," *Washington Business Journal*, 1997.

U. Sakarya, U. Leloglu and E. Tunali, "Three-Dimensional Surface Reconstruction for Cartridge Cases Using Photometric Stereo," *Forensic Science International*, vol. 175, pp. 209-217, 2008.

N. Senin, R. Groppetti, L. Garofano, P. Fratini and M. Pierni, "Three-Dimensional Surface Topography Acquisition and Analysis for Firearm Identification," *Journal of Forensic Sciences*, vol. 51, no. 2, pp. 282-295, 2006.

J. Song, E. Whinton, D. Kelley, R. Clary, L. Ma and S. Ballou, "SRM 2460/2461 Standard Bullets and Cartridge Cases Project," *Journal of Research of the National Institute of Standards and Technology*, vol. 109, pp. 533-542, 2004.

J. Song and T. Vorburger, "Topography Measurement and Applications," *Proceedings SPIE*, vol. 6, no. 280, pp. 1T1-1T8, 2006.

J. Song, T. Vorburger, S. Ballou, L. Ma, B. Renegar, A. Zheng and M. Ols, "Three Steps Towards Metrological Traceability for Ballistics Signature Measurements," *Measurement Science Review*, vol. 10, pp. 19-21, 2010.

J. Song, T. Vorburger, S. Ballou, R. Thompson, J. Yen, B. Renegar, A. Zheng, R. Silver and M. Ols, "The National Ballistics Imaging Comparison (NBIC) Project," *Forensic Science International*, vol. 216, pp. 168-182, 2012.

J. Song, "Proposed NIST Ballistics Identification System (NBIS) Using 3D Topography Measurements on Correlation Cells," *AFTE Journal*, vol. 45, pp. 184-194, 2013.

SWGgun 2005, "The Fundamentals of Firearm and Toolmark Identification," [Online]. Available: <http://swggun.org>.

Ultra Electronics Forensic Technology, Inc., [Online]. Available: <http://ultra-forensictechnology.com/pubs#bro>; accessed March 11, 2019.

V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

T. Vorburger et al., "Surface Topography Analysis for a Feasibility Assessment of a National Ballistics Imaging Database", NISTIR 7362 National Institute of Standards and Technology, Gaithersburg MD, 2007.

T. Vorburger, J. Yen, J. Song, R. Thompson, T. Renegar, A. Zheng and M. Tong, "The Second National Ballistics Imaging Comparison (NBIC-2)," *Journal of Research of National Institute of Standards and Technology*, vol. 119, pp. 644-673, 2014.

T. Vorburger, J. Song and N. Petraco, "Topography Measurements and Applications in Ballistics and Tool Mark Identifications," *Surface Topography: Metrology and Properties*, vol. 4, 013002, 2016; doi:10.1088/2051-672X/4/1/013002.

T. Weller, A. Zheng, R. Thompson and F. Tulleners, "Confocal Microscopy Analysis of Breach Face Marks on Fired Cartridge Case from 10 Consecutively Manufactured Pistol Slides," *Journal of Forensic Sciences*, vol. 57, pp. 912-917, 2012.

T. Weller, M. Brubaker, P. Duez and R. Lilien, "Introduction and Initial Evaluation of a Novel Three-Dimensional Imaging and Analysis System for Firearm Forensics," *AFTE Journal*, vol. 47, no. 4, pp. 198-208, 2015.

B. Zitova and J. Glusser, "Image Registration Methods: A Survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977-1000, 2003.