Document Title:	**Applied Research, Development, and Method Validation for a Statistically Based Comparison of Tool Marks using GelSight-Based Three Dimensional Imaging and Novel Comparison Algorithms for Firearm Forensics**
Author(s):	**Ryan Lilien**
Document Number:	**250538**
Date Received:	**January 2017**
Award Number:	**2014-DN-BX-K012**

Research Performance Progress Report - Cover Page

Federal Agency and Organization Element: Department of Justice, Office of Justice Programs

Federal Grant or Other Identifying Number: 2014-DN-BX-K012

Project Title: Applied Research, Development, and Method Validation for a Statistically Based Comparison of Tool Marks using GelSight-Based Three Dimensional Imaging and Novel Comparison Algorithms for Firearm Forensics

PD/PI Name, Title, Contact Info: Ryan Lilien, Chief Scientific Officer, Cadre Research Labs; 212 W Superior St; Suite 503; Chicago, IL 60654

Name of Submitting Official: Ryan Lilien

Submission Date: March 28, 2016

DUNS: 034448780 **EIN:** 27-2528667

Recipient Organization: Cadre Research Labs, LLC (small business)

Recipient Identifying Number (if any): N/A

Project/Grant Period: Start: 1/1/2015, End: 12/31/2015

Reporting Period End Date: 12/31/2015

Report Term or Frequency: Twice a year

Signature of Submitting Official:

Note that this report aims to follow the NIJ deliverable "10 page final summary overview" format. This new format is a 10 page double spaced document. We therefore created to a 10 page report not including this cover page and not including figures (which appear at the end of this report). We also included a single page appendix that addresses "Implications for Criminal Justice Policy" and "Additional Accomplishments".

1 Project Purpose and Background

The described work represents a critical next step in investigating the overall performance characteristics of a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge casings, TopMatch. In January 2013, we began development of a 3D surface topography imaging and analysis system for firearm forensics based on the GelSight imaging technology and custom feature-based image comparison algorithms. The technology shows excellent (and continually improving) match accuracy. However, to fully establish the base credibility of the platform it is necessary to establish best scanning practices and to demonstrate that the method meets the quality control criteria of other forensic instruments. This proposal aims to enhance this novel, low-cost system for 3D imaging and comparison of cartridge casings by establishing best practices and conducting a set of methodology studies.

The research work was completed by Cadre Research Labs, a scientific computing contract research organization, working in collaboration with GelSight Inc, a company formed by the MIT researchers who developed the GelSight surface topography imaging technology. The two companies collaborate closely with Todd Weller, a firearms identification specialist and Criminalist in the Oakland Police Department. We continue to collaborate with colleagues at NIST, Andy Smith (San Francisco PD), Chris Coleman (Contra Costa County Office of the Sheriff), Nancy McCombs (DoJ Fresno Crime Lab), and Karl Larsen (U. Illinois at Chicago). These collaborators continue to be excellent partners and provide both scans and constructive feedback.

2 Project Design

The aims of our one year project follow the intent of the Scientific Working Group on DNA Analysis Methods (SWGDAM) Validation Guidelines for Analysis Methods document. That document describes seven validation criteria for new scientific methods: Accuracy, Stability, Precision, Repeatability, Reproducibility, Contamination Risk, and Performance Checks. In this project we complement our previous studies on Accuracy and evidence Stability by addressing the remaining issues of Precision, Repeatability, Reproducibility, Contamination Risk, and Performance Checks. In addition, we will study the effects of Cleaning Protocols, Focus Variation, and Ambient Lighting Variation. In Aim 1 we setup three datasets and develop five objective measures of scan quality. In Aim 2, we established a set of Best Scanning Practices, explored Cleaning Protocols, Focus Variation, and Ambient Lighting effects, investigated Precision, Repeatability, and Contamination Risk, and established Reproducibility via an

Inter-Operator Validation study. In Aim 3 we developed System Performance Checks. All 3 proposed aims were successfully completed during the project period.

3 Materials and Methods

A significant portion of the proposed work entailed the development of materials and methods for use in the remainder of the proposed project. These details are provided in this section. Many methods have been abbreviated to conform to the maximum page limit.

Base Scanner: The scan acquisition system uses advanced three-dimensional imaging algorithms (*e.g.*, shape from shading and photometric stereo) and the retrographic sensor of Johnson and Adelson [1, 2] to measure an object's three dimensional surface topography. In contrast to confocal microscopy and focus-variation microscopy, the use of a painted elastomeric gel removes the influence of surface reflectivity on the measured topography. The setup contains an 18-megapixel Canon digital camera with a 65mm macro lens. A small-pistol primer (*e.g.*, 9mm) and breech-face impression can be measured using a single frame (*i.e.*, without stitching multiple images) at approximately $1.4\mu m$/pixel lateral resolution with submicron depth resolution. Our newly revised casing holder secures the casing using its strong extractor groove. Scan acquisition requires ∼2 minutes per casing.

Dataset Assembly: The initial plan was to collect test fires from ten firearms and then to split this set into subsets for different parts of the experiments. In collaboration with Mr. Weller we created three sets from these initial ten casings. Two **medium** sets and one **small** set. The medium sets contain six pairs of test fires (four 9mm, two 40 caliber) and include three casings with aperture shear, three casings with granular marks, one with filing marks, one with milled marks, and one casing that is poorly marked. The two medium sets share four firearms in common. The small set contains three pairs of test fires and includes one Glock with strong a strong aperture shear, one casing with granular marks, one with filing marks, and one with milled marks. Thumbnail views of casings from each firearm are shown in Fig. 2.

Metric Development: A series of evaluative metrics were required for completion of the proposed aims. The following metrics were created. **Flatness**: A numerical measure of flatness is determined by first robustly fitting planes to each of the four corners of the scan, the entire scan, and the masked region of the breech face impression. Then for each computed plane we determine the maximum angle difference between it and the other five planes. We return the average of these six angles as our flatness measure. This is a very sensitive measure of flatness. A smaller number indicates a flatter scan; scores below 20 are considered within flatness tolerance (see Fig. 4 Bottom). **Number of Features**: In the experiments below, the Number of Detected Features is often normalized for easier comparison (see Supplementary Information). Casings with significant surface structure typically have more detected features. **Match Scores**: Most results below include the 01 Match Score (explained as the 01 Confidence Score below). Some results also include a Normalized Raw Match Score which is simply the raw count of the number of matched features. This raw match score is one component of the 01 match score. **Surface Contamination**: A four point qualitative scale was devised based on the observed contamination on the breech-face impression only (Fig. 3): (1) Perfect or near perfect scan. No large granules. Few to no small granules. (2) Some granules (small and/or large). Would not necessitate a rescan in a production setting. Fewer than 5 large granules. (3) Moderate number of granules. Would likely rescan in production setting. At least 5 large granules. (4) Significant contamination across entire surface. At least one giant granule or at least 10 large granules.

Scan Comparison Algorithm: Our scan comparison method is not based on cross-correlation. Instead, our approach is a feature-based iterative surface matching algorithm. In contrast to cross-correlation based methods, feature-based techniques compute the match score using only the portions of the scan identified as informative (*i.e.,* the features). **Breech-Face Impression:** Features correspond to automatically identified regions of the scan with nonzero gradients in both the x and y dimensions. To compare the breech-face impressions of two casings, the TopMatch iterative matching algorithm identifies a maximal set of the detected features that are geometrically consistent between the two cases. A set of matches is considered geometrically consistent if the matched features of two casings can be spatially aligned after a single rotation and translation of one scan. In other words, similarly shaped features are arranged in a similar geometric layout. The score of the match is a function of the number and quality of matched features. **Aperture Shear:** Whereas the breech-face impression is an impressed mark formed by direct force, the aperture shear is a striated mark formed by a shearing or scraping force. The unique information in a striated mark is contained in its linear profile and not in its overall surface. When using a traditional comparison scope, examiners arrange a split-screen side-by-side view where the linear striation profiles are shown intersecting the dividing line. The examiner then evaluates similarity between the observed stria. Therefore, the linear profile of the aperture shear should be considered separately from the breech-face impression. Most previous methods and commercial systems group the breech-face impression and aperture shear together. These previous systems incorrectly treat the aperture shear as an impressed mark which may result in reduced matching accuracy. Our system asks the user to indicate the presence and location of a shear mark during the masking process (green region in Fig. 1).The software utilizes a series of linear and nonlinear baseline correction, unwarping, and alignment methods to extract a single robust linear profile from the identified aperture shear (Fig. 1 Profile). Once extracted, the linear profiles are analyzed to identify peaks and valleys corresponding to traditional striae. A typical profile will have 10-30 detected peaks/valleys. Each peak is parameterized based on its location, width, and height.

01 Confidence Score: The breech-face impression and aperture shear similarity scores are then combined (using regression) into an overall similarity score. The combined score takes into consideration that if a firearm is likely to produce an aperture shear on one casing that it is likely to produce a shear on a second casing. The output is a confidence score between 0 and 1. The confidence score is not a strict probability; that is, a score of 0.9 does not indicate 90% chance of a match. However, because the score is directly correlated with toolmark similarity, because it is on a common scale and comparable between two pairs of cases, because it is numeric, and because it is bounded, it can be modeled with an underlying probability distribution to assign a more interpretable "probability of match" for each pairwise correlation. Additional detail regarding the comparison algorithm is provided in our recent paper [3]. A

Normalized Metrics: The 'best' Number of Detected Features and Match Scores are different for each casing. That is, firearm X may inherently mark better than firearm Y. Therefore, to facilitate numeric reporting for the above experiments we typically normalize the Number of Features and Match Scores to lie between 0 and 1. To compute a normalized metric for a casing we find the maximum value for that casing across all experimental variants and then divide all raw values by this maximum. For example, in the cleaning experiments consider a casing X which has 5000, 6000, and 5500 features for cleaning protocols (1), (2), and (3). We divide each raw value by 6000 (the maximum) to obtain 0.83, 1.00, and 0.92 as the normalized number of features score. This allows direct comparison with casing Y which has 4000, 4600, and 4300 which has a maximum value of 4600 and normalizes to 0.87, 1.00, 0.93. Using this method, normalized scores can be compared within an experiment (*e.g.*, cleaning result to cleaning result) but not across experiments (*e.g.*, cleaning result to persistence result). The 01 Match Score does not need to be normalized.

Open File Format: To support the free exchange of topography data, we led the creation of a new consortium named the OpenFMC (Open Forensics Metrology Consortium). The group's first accomplishment was the adoption of a new file format, X3P, for the storage and exchange of three-dimensional surface topography data. The X3P format contains a number of data fields in which we can store evidence specific, scan specific, and hardware specific data. Our group, has created and distributed free beta software for reading and writing X3P to the forensic community. This software is currently being used by NIST in preparing their firearms forensics database. We've demonstrated the portability of the X3P format by successfully reading and comparing scans from NIST (NanoFocus confocal) and Alicona.

Cleaning Method Experiments: Several cleaning protocols were evaluated. These included various solvents, mechanical cleaning agents, and drying agents. The six casings of the Small test set were used in this experiment. Each casing was placed in a small zip-top bag of potting soil and shaken. The casings were removed from the bag, tapped on the tabletop (to remove large clumps of visible dirt), subjected to one of the cleaning protocols, and then scanned. The gel pad was cleaned with tape between each scan. A scan of flat glass was collected between each different cleaning condition to verify no dirt remained on the gel. Each scan was evaluated using the Number of Features, Match Scores, and Surface Contamination measures. Additional details in Supplementary Material.

Lighting Variation Experiments: The effects of lighting on surface measurement were evaluated. Conditions evaluated included light from 0 to 8,000 lux (a typical office is 500-1000 lux). Three bulb types (fluorescent and incandescent) were used. The light source was positioned slightly to the left of the scanner to maximize the chance of any asymmetry related issues. Each scan was evaluated using the Flatness, Number of Features, and Match Score measures. Additional details in Supplementary Material.

Aperture Variation Experiments: The effects of different apertures were evaluated. The advantage of a smaller aperture (or larger f-number) is a larger depth of field. The disadvantage is a darker image which requires a longer exposure time or higher ISO gain to obtain the desired image brightness. For each aperture setting, the exposure time and ISO gain were adjusted to maximize the range of observed brightnesses while not saturating the collected image. For each scan the camera was focused on the breech-face-impression. Each scan was evaluated using the Flatness, Number of Features, and Match Scores measures. Additional details in Supplementary Material.

Repeatability Experiments: The ability of a single operator and a single instrument to produce similar measurements over time was assessed by having one operator (Lilien) scan the same Medium test set five times over a two month period. Each scan was evaluated using the Flatness, Number of Features, and Match Scores measures. Additional details in Supplementary Material.

Precision Experiments: The ability of the scanner to collect a consistent scan several times in succession was evaluated by scanning the Medium test set once; however, each casing was scanned five replicate times without removing the sample from the scanner. Each scan was evaluated using the Flatness, Number of Features, and Match Scores measures. Additional details in Supplementary Material.

Persistence Experiments: Persistence contamination refers to the potential 'memory imprint' that could be temporarily left on the scanner by a previous casing. A complex scanning schedule was designed where a primary casing was pushed into the gel for a set amount of time (up to 30 minutes). After the imprint time, the same gel was used to scan a) a known non-match to the original, b) a known-match to the original, and c) a piece of flat glass. Each scan was evaluated using the Flatness, Number of Features, and Match Scores measures. Additional details in Supplementary Material.

System Performance Checks: We developed a quality control system performance check which utilizes the Rubert nickel-boron plated sinusoidal reference standard 525E. This specimen has a wavelength of 135μm and an amplitude of 19μm with a reported R_a of 6.25μm. A custom measurement algorithm was developed to analyze a scan of this specimen. The algorithm applies a baseline correction and then computes the measured wavelength, amplitude, and R_a value. These numbers can be used as a quality control check on the scanning process. A typical result is shown in Fig. 5. It shows a measured wavelength of 134.6μm (actual: 135μm), amplitude of 19.86μm (actual: 19μm), and R_a of 6.23 (actual: 6.25) thus demonstrating micron lateral and depth resolution. This reference standard is being integrated into the quality control workflow of the scanner.

Interoperator Study: A small interoperator scanning study was completed to determine if different examiners could each obtain high quality scans after having been trained on the scanner. Ten individuals participated in the study (seven novice operators and three experienced operators). Novice operators received approximately two hours of training (including seminar and hands-on). Each operator received consistent instructions and scanned one of two different (but highly similar) Medium test sets. Based on the location of the operators, four different scanners were used. Eleven scan sets were collected (one experienced operator collected each test set). Additional details in Supplementary Material.

Match Prediction: A small investigation was conducted to determine if a quick assessment of a scan could determine the likelihood that a match could be identified. Our large hundred firearm dataset (described in Weller [3]) was used in this investigation. None of the scans in this study were warped, out-of-focus, or dirty. We investigated a number of scan variables to determine if any correlated with likelihood of successful match. None of the considered variables were strongly correlated with match likelihood. Because all scans were sufficiently flat, there was no correlation between flatness and match likelihood. The most highly correlated variables were the Masked Surface Area (larger surface area increased likelihood of match), Number of Surface Features (more features increased likelihood of match), and Number of Large Features Detected (more large features increased the likelihood of match) (Fig. 6).

4 Data Results and Analysis

Cleaning Method Experiments: The results of the cleaning experiments (Sec. 3) are provided in Table 1. The results demonstrate that while all cleaning methods reduce the number of visualized dirt particles, those involving physical contact with a soft brush were generally superior to both non-contact and alternative contact methods. There is no statistical difference in the Normalized Number of Detected Features nor the 01 Match Score for any of the cleaning protocols. Interestingly we notice that even the dirty casings have the same Number of Detected Features and the same 01 Match Score. While there was no clear best protocol among several equally successful approaches, we selected the protocol of isopropyl alcohol, a soft bush, and compressed air as the best practice because it combines a solvent and physical contact which may be useful under a wider range of contamination conditions.

Lighting Variation Experiments: The results of the lighting variation experiments (Sec. 3) are provided in Table 2. All scans are appropriately flat (Flatness Scores less than twenty) (Fig. 4). There is no trend observed between the flatness and light intensity. While the Normalized Number of Detected Features is consistent across all lighting conditions there is a slight trend where the raw match score drops slightly for very bright directional lighting. This is primarily observed for 8000 lux directional lighting. We note that these conditions should never occur within the lab setting. The 01 Match Score does not show this trend. Best practices are thus to use ambient room lighting (not in direct sun) or a black-out curtain. In summary, normal lighting variation has minimal to no effect on scan acquisition.

Aperture Variation Experiments: The results of the aperture variation experiments (Sec. 3) are provided in Table 3. All scans are appropriately flat (Flatness Scores less than twenty). The data supports the hypothesized trend that longer exposure times and larger ISO gains will result in a slightly less sharp scan. This manifests as a decrease in the Normalized Number of Features. F5.6 has 1.00 and 0.99; F8 has 0.95 and 0.95; and F11 has 0.89 and 0.90. Thus there is a roughly 5% drop in Number of Detected Features from F5.6 to F8 and another 5% drop from F8 to F11. The Normalized Raw Match Score follows the same trend. This indicates that the 5% of "lost" features come equally from matched and non-matched features. We note that the 01 Match Score is not affected by the camera settings and shows consistently strong results across all conditions. The results support the expected tradeoff where increasing the aperture increases the depth of field but potentially results in a lower resolution image. We selected F8 1.3s ISO400 as the best practice as it provides an improved depth of field (relative to F5.6) without an excessive drop in the number of features detected.

Repeatability Experiments: The results of the repeatability experiments (Sec. 3) are provided in Table 4. The results demonstrate that repeat scans are consistent in their quality. There is no statistical difference between the Flatness Scores, Normalized Number of Detected Features, and 01 Match Score. The dataset contains one outlier Scan 4 Firearm 3; while there is no obvious reason why this match was missed we note that Firearm 3 is the most weakly marking of those used in the test set. We conclude that the scanner and scanning protocol have excellent repeatability.

Precision Experiments: The results of the precision experiments (Sec. 3) are provided in Table 5. The results demonstrate that replicate scans are extremely similar. As expected, the replicate scans of the Precision Experiments are more self-consistent than those of the Repeatability Experiments (where the casings were physically removed from the scanner). Removing the casings caused them to potentially collect dust/dirt, they were recleaned, the casings were remounted, and the scanner was refocused for each replicated Repeatability scan; this was not the case of the Precision scans. The results confirm that the scanner has high precision, the replicate scans are highly consistent, and there is no feature drift.

Persistence Experiments: The results of the persistence experiments (Sec. 3) are provided in Table 6. All scans are appropriately flat (Flatness Scores less than twenty). The Normalized Number of Detected Features is consistent across all experiments. We do not observe an increase in Number of Detected Features for the 30 minute impression scans (which might suggest that some of the features from the original casing are persisting and appearing on the target casing). All Known Matches (KM) correctly identify to their sister casing. All Known Non-Matches (KNM) are correctly not identified. Best practice is to remove the casing from the gel after the scan; however, these results indicate that leaving the casing in the gel for up to 30 minutes will not induce a false positive match. There is no memory effect[1].

Interoperator Study: The results of the interoperator study (Sec. 3) are provided in Tables 7 and 8. All scans are appropriately flat (Flatness Scores less than twenty). There was no observable difference in the Normalized Number of Detected Features between the Novice and Experienced operators. There was no significant difference in the 01 Match Scores obtained for Novice Test Set A, Novice Test Set B, and Experienced Test Set B (0.868, 0.897, and 0.891); we note that there was only a single experienced set collected for Test Set A which had a high mean 01 Match Score (0.980) which may have been an outlier. The only observed difference between the experienced and novice operators came in the cleanliness of the scans collected by the different operators (Table 8)[2]. The experienced operators had a mean cleanliness score of 1.08 compared to the novices who scored 1.23 (lower is better); 93% of the scans collected

[1]Although the data demonstrates that no structural features persist as a memory in the gel, we did notice that after a long impression time the gel did initially have a palpable slight depression where the casing pushed into the gel. This depression was short lived and the gel returned to its original non-depressed shape after a minute or two.

[2]Surface Contamination scores are rated 1, 2, 3, or 4 as described in the Methods section (1 is cleanest, 4 is dirtiest).

by the experienced operators scored a 1 while 77% of the scans collected by novices scored a 1. No scan from any operator scored worse than a 2. These results show that all operators are able to collect high quality scans and there was little observable difference between most operators. Novice operators may be slightly less focused on casing cleaning; however, this result may be more a reflection of the fastidiousness of the experienced operators. We note that the result of the Cleaning Experiments (above) show that even dirty casings (Surface Contaminant score of 4) are typically successfully matched.

5 Scholarly Products Produced

The primary product of the proposed research is the presentation of our results and progress. We delivered oral presentations on our system and experimental results at the 2015 AAFS meeting (Orlando, FL), the 2015 AFTE meeting (Dallas, TX), the 2015 Eastern Regional AFTE Meeting (FBI, Quantico, VA), and the 2015 International Association of Forensic & Security Metrology Meeting (San Diego, CA). We've submitted to speak at the 2016 AFTE meeting (May 2016). These presentations are shared with our NIJ Program Manager Gregory Dutton (either in person or via teleconference). Building from our research results, we will be running a Virtual Microscopy workshop at the 2016 AFTE Meeting. Our first full-length publication on our results was submitted and accepted to the AFTE Journal in 2015 [3]. We are now preparing our best practice investigation paper (2015 grant results) for publication.

6 Summary

We successfully completed the proposed aims during the project period. We established carefully selected casing test sets (Aim 1A), we developed a series of scan quality and scanner performance metrics (Aim 1B), we established scanning best practices (cleaning, lighting, camera aperture) (Aim 2A), we evaluated scanner performance (repeatability, precision, and persistence contamination) (Aim 2B), we completed a small inter-operator validation study (Aim 2C), and we designed a system and scan performance checks (Aim 3). Through the year we continued collaboration with academic, industry, and government collaborators. We presented our work at four academic conferences and had a full-length research paper accepted for publication in the AFTE Journal.

In addition to the above aims we've improved our overall matching and analysis algorithm, continued deployment studies with a number of crime labs and academic labs, led implementation and adoption of the X3P open file format, demonstrated cross-modality matching (using X3P to share data between different imaging hardware), and evaluated several hardware improvements. Through this work we established best practices (cleaning, lighting, camera settings), demonstrated excellent repeatability and precision, eliminated the concern of persistence contamination, and showed that both novice and experienced operators can collect high quality scans. The completion of these studies sets a solid foundation on which our scanning methodology and comparison algorithms can build.

Appendix

Additional Accomplishments: In addition to the results described above we were able to accomplish the following during the project period: First, we significantly improved the quality of the painted layer of the gel which reduces the measurement noise of captured scans. An improved gel utilizes a new iron oxide pigment which appears red in contrast to our earlier silver gel. The iron oxide gel has a more uniform pigment layer which improves measurement fidelity. We've combined the iron oxide pigment with variations in gel firmness. Firmer gel is typically easier to manipulate and is less likely to become damaged while maintaining the ability to conform to surface geometry. In Fall 2015, we collaborated with Nina Jefferson (Boston Police Department) who spent time at our lab to scan approximately 500 casings from her set of Boston Police Department Glock 22 (40 caliber) test fires. We are currently working with Ms. Jefferson to analyze these casings as part of her Master's thesis (Boston University). The scans will be made available to Alan Zheng for inclusion in his NIST Research Ballistics Toolmark Database. We completed initial development of a rudimentary stand-alone casing viewer for X3P files that is freely available. Viewer beta testing began in August 2015 and we anticipate a full release in the near future. This viewer allows examiners to view 3D scans acquired using either our system or any commercial system that supports the X3P format. We've received great interest in this software. Alan Zheng (NIST) will link to this software from his NIST Reference Ballistic Toolmark Database.

Implications for Criminal Justice Policy and Practice

Our primary impact has been the development of a novel 3D imaging and analysis system with reduced cost and improved accuracy compared to existing solutions. Our work directly addresses several aims of the NIJ's Applied Research and Development in Forensic Science for Criminal Justice Purposes program.

Through direct collaboration, networking, talks, seminars, and publications we have made many forensic labs (local, state, and federal), practitioners, and policy makers within the criminal justice system aware of this work. We are developing measurement and analytic techniques, grounded in mathematical science that are able to provide accurate quantitative sample comparison and database search. We're establishing best scanning practices (cleaning, lighting, camera settings) and operational statistics (repeatability, precision, persistence contamination, system performance checks) for our system. This work benefits the criminal justice system and their ability to present firearm identification and toolmark evidence in the courtroom. Additional impact will be made as more crime labs become aware of the work and as we continue to disseminate results (*e.g.*, upcoming AAFS and AFTE meetings). In 2015, two team members, Lilien and Weller were selected to the Firearms subcommittee of NIST's new OSAC initiative. Through their work on the OSAC, Weller and Lilien are creating guidelines and standards for emerging forensic technologies.

At least five crime laboratories have had access to our technology. This would not have been possible prior to receiving this award. For labs that currently have 2D imaging systems, our 3D system provides a significant improvement in imaging and match accuracy. For labs that currently have competing 3D imaging systems, we feel our system offers more flexibility and transparency with respect to how the scanner works, increased resolution, improved visualization, and interpretable match score.

Figures and Tables

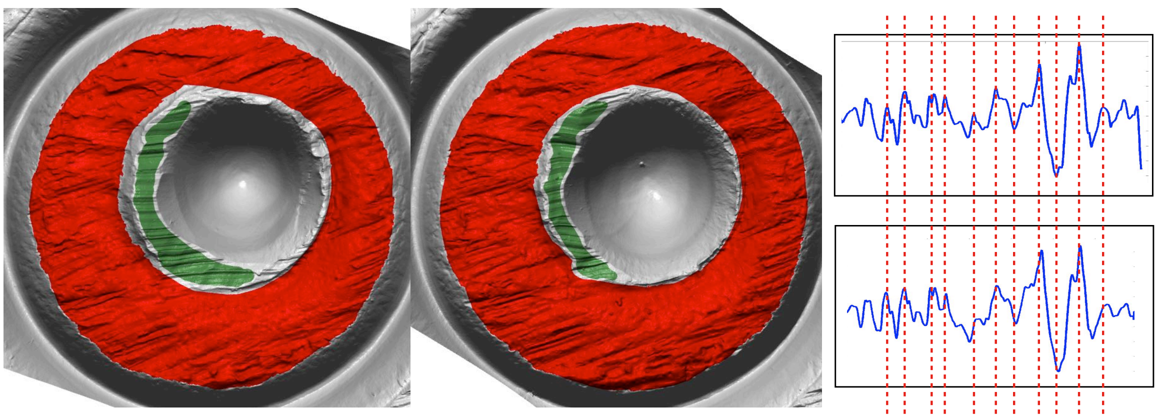


Figure 1: **Aperture Shear Extraction** (Left) A pair of Norinco test fires with breech-face impression masked in red and aperture shear masked in green. The analysis algorithm extracts a set of informative surface features from the red region. The algorithm extracts a linear striation profile and corresponding set of peaks and valleys (corresponding to traditional striae) from the green region. (Right) The two extracted linear profiles. Corresponding striae are indicated by vertical red dashed lines.

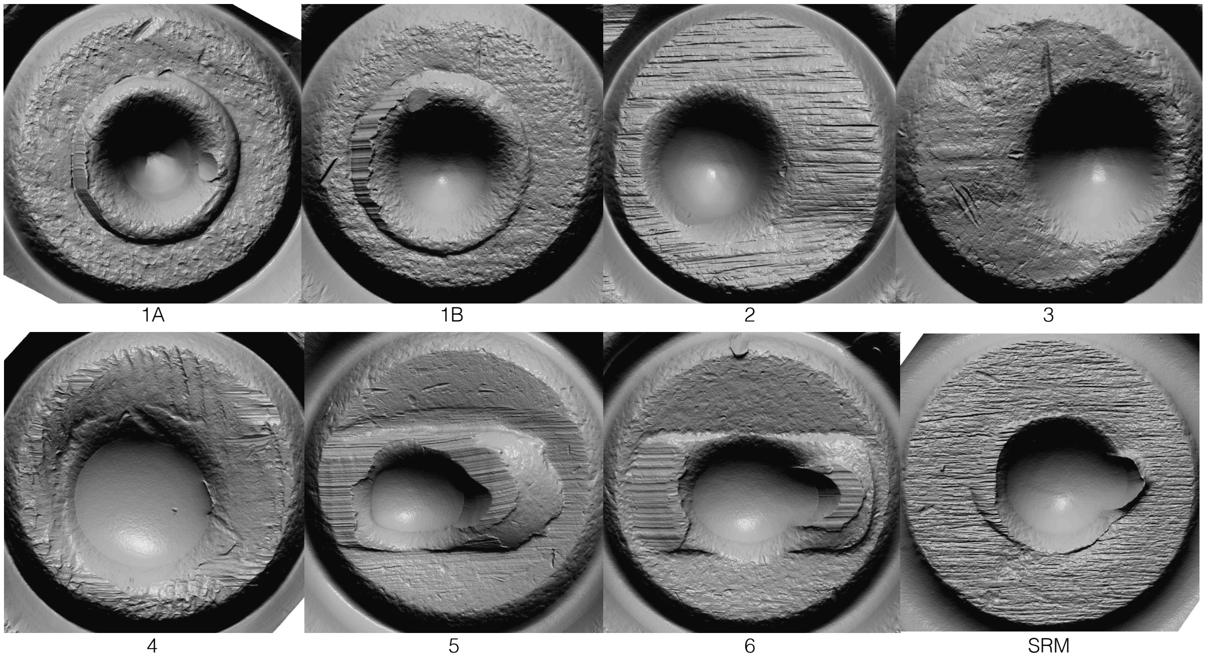


Figure 2: **Test Casings.** The two medium test sets included test fires from firearms represented by casings 2-6 and either 1A or 1B. Firearm 1A/1B are granular with aperture shear, firearm 2 has filing marks, firearm 3 is minimally marked, firearm 4 has milled marks, firearms 5 and 6 have strong aperture shears, the NIST-SRM casing is included in a few experiments. The Small Set includes Firearms 2, 5, and 6. Firearms: 1A: Fabrique Nationale 9mm, 1B: Ruger 9mm, 2: Hi-Point 9mm, 3: Hi-Point 9mm, 4: Norinco 9mm, 5: Smith & Wesson 40S&W, 6: Hi-Point 40S&W.

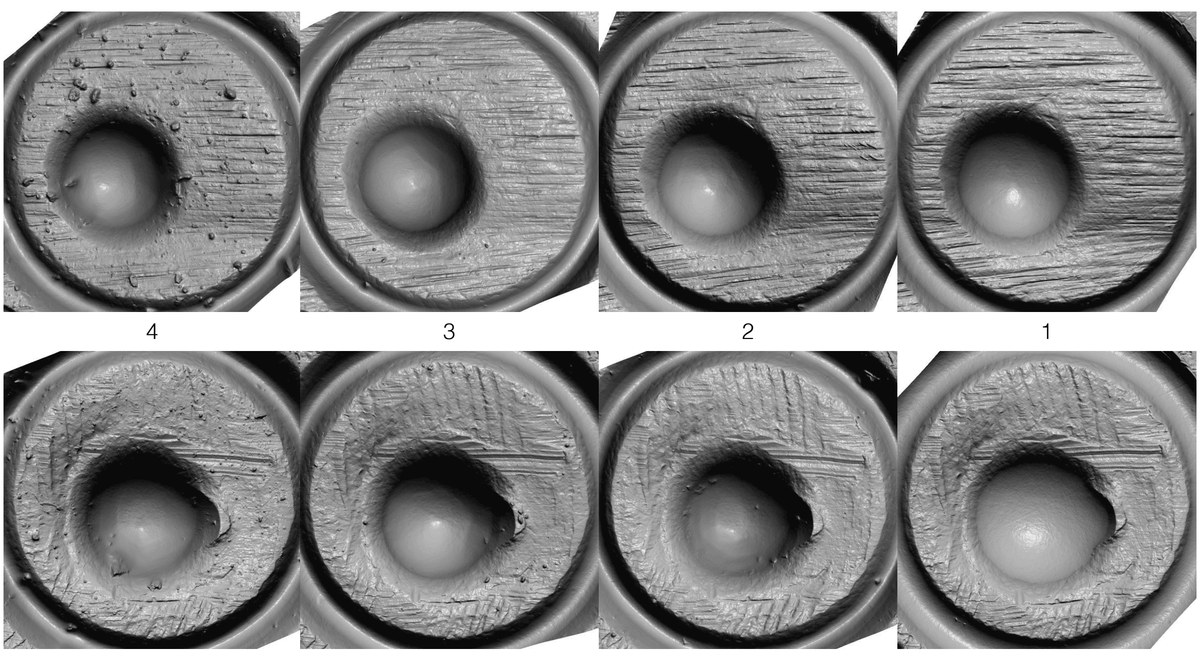


Figure 3: **Surface Contamination Score.** Each casing was assigned a value from 1-4 based on a human assessment of surface contamination (see text). (Top Row) Hi-Point 9mm; (Bottom Row) Norinco 9mm

Cleaning Protocol	Surf. Contam.	Norm. Num Feat	01 Match Score
Dirty	4.0 (0.0)	0.98 (0.03)	0.997 (0.006)
IPA + Manual Air	3.3 (0.5)	0.95 (0.04)	0.995 (0.008)
IPA + Compressed Air	3.0 (0.6)	0.96 (0.05)	0.998 (0.003)
IPA + Brush	1.2 (0.4)	0.95 (0.03)	0.997 (0.005)
IPA + Brush + Manual Air	**1.2 (0.4)**	**0.94 (0.05)**	**0.999 (0.001)**
IPA + Swab + Manual Air	3.8 (1.0)	0.95 (0.05)	0.999 (0.002)
IPA + SpBrush + Manual Air	2.3 (0.5)	0.95 (0.04)	0.969 (0.062)
WD + Brush	1.5 (0.6)	0.90 (0.06)	0.999 (0.001)
Tape	2.0 (0.0)	0.96 (0.05)	0.988 (0.023)
Tape + IPA + Manual Air	1.2 (0.4)	0.92 (0.05)	0.999 (0.001)

Table 1: **Cleaning Protocol Results.** Scan metrics are presented for ten test conditions on the Small test set. Abbreviations Used: Compressed Air: Can of commercial compressed air, IPA: Isopropyl Alcohol, Manual Air: Manual hand-squeeze air blowing device, SpBrush: Sponge brush, Surf Contam: Surface Contamination Metric, Swab: Cotton-tipped swab, WD: WD40 lubricant and solvent. The details of each protocol are listed in the order they are applied. For example, "IPA+Brush+Manual Air" means that the casing heads were first dipped in isopropyl alcohol, then cleaned with a soft brush, and finally dried with a few puffs of air. Normalization for each firearm took place across all cleaning protocols. Additional details on each protocol are provided in Supplementary Material. The Surface Contamination score, Normalized Number of Detected Features, and 01 Match Scores are described in Section 3. The Match Score is the average (standard deviation) of those computed between corresponding pairs of matched test fires for the specified cleaning protocol. Each column lists the mean value with standard deviation in parentheses. The IPA+Brush+Manual Air protocol (bold) was selected as the best practice.

Bulb Type	Brightness	Flatness Score	Norm. Num Feat	Norm. Raw Match Score	01 Match Score
Blackout	0	7.8 (3.2)	0.94 (0.03)	0.95 (0.03)	0.999 (0.00)
Typical Office	400	6.7 (3.8)	0.95 (0.02)	0.88 (0.11)	0.999 (0.00)
Standard Incandescent	750	4.7 (2.0)	0.93 (0.02)	0.87 (0.15)	0.999 (0.00)
	1200	5.8 (2.5)	0.94 (0.03)	0.81 (0.07)	0.999 (0.00)
	3000	6.9 (2.6)	0.93 (0.03)	0.79 (0.14)	0.999 (0.00)
	8000	8.4 (3.1)	0.98 (0.02)	0.45 (0.28)	0.999 (0.01)
Standard Fluorescent	750	7.7 (1.5)	0.96 (0.02)	0.93 (0.05)	0.999 (0.00)
	1200	11.8 (4.1)	0.91 (0.03)	0.78 (0.02)	0.999 (0.00)
	3000	6.5 (1.4)	0.95 (0.03)	0.86 (0.08)	0.999 (0.00)
	8000	8.6 (3.9)	0.97 (0.03)	0.78 (0.05)	0.999 (0.00)
Full-Spectrum Fluorescent	750	5.8 (1.2)	0.96 (0.01)	0.81 (0.14)	0.999 (0.01)
	1200	6.2 (2.1)	0.98 (0.02)	0.84 (0.11)	0.999 (0.00)
	3000	7.1 (1.9)	0.95 (0.02)	0.85 (0.12)	0.999 (0.00)
	8000	5.5 (1.8)	0.98 (0.02)	0.71 (0.10)	0.999 (0.01)

Table 2: **Lighting Variation Results.** Scan metrics are presented for the tested lighting conditions (Standard Incandescent (2990K), Standard Fluorescent (5000K), and Full-Spectrum Fluorescent (6500K)) and indicated light intensities on the Small test set. Normalization for each firearm took place across all three bulbs and four lightness values. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Detected Features, Normalized Raw Match Score, and 01 Match Score are described in Section 3. The Match Score is the average (standard deviation) of those computed between corresponding pairs of matched test fires for the specified lighting condition. The flatness scores are also shown in Fig. 4. Each column lists the mean value with standard deviation in parentheses. Analysis is provided in Section 4.

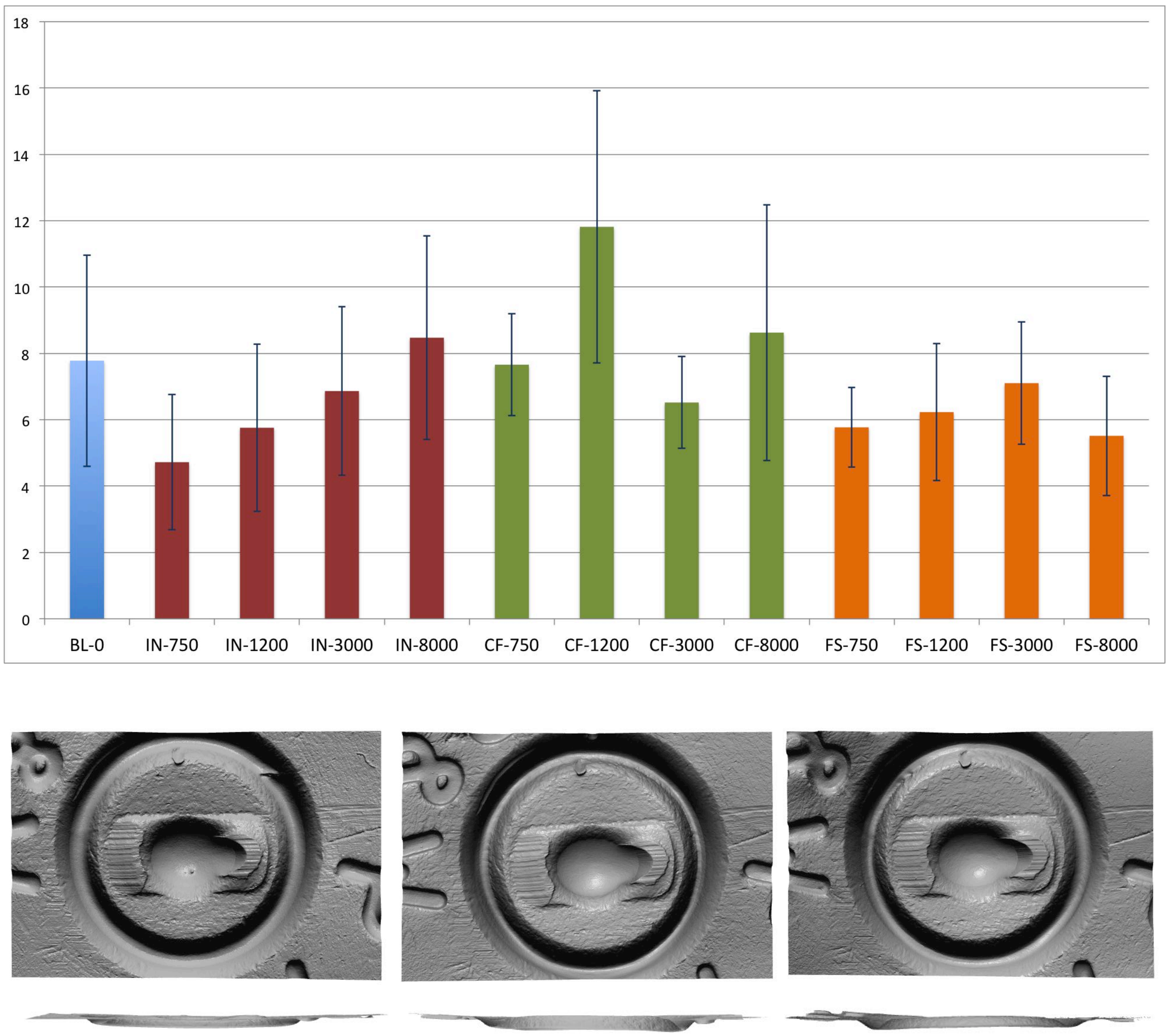


Figure 4: **Lighting Experiments - Flatness.** (Top) The mean flatness for the scans collected under each lighting condition are shown with one standard deviation lines shown. Each lighting condition is indicated by the light source and lux level. BL: Blackout conditions, IN: Standard Incandescent, CF: Standard Compact Fluorescent, FS: Full-Spectrum Fluorescent. No statistical difference is observed between the different conditions. Scans with a flatness score below 20 are considered perfectly flat. (Bottom) Three examples of flatness scores from our dataset. Shown are top-down and side on views for each scan. Note that the side view often cuts through the annulus and/or headstamp lettering causing the appearance of slight deviations (left to right: flatness scores of 3.2, 4.6, and 8.9).

Aperture/ Camera	Firearm	Flatness Score	Mean Feature Size	Norm. Num Feat	Norm. Raw Match Score	01 Match Score
F5.6 1s ISO400	SRM	6.7	25.6	1.00 (0.01)	-	-
	2	8.4	26.6		0.969 (0.028)	0.999 (0.001)
	5	6.0	24.2			
	6	7.2	26.3			
	Mean	7.1	25.7			
F5.6 1s ISO200	SRM	12.0	25.6	0.99 (0.01)	-	-
	2	10.0	26.7		0.972 (0.044)	0.999 (0.001)
	5	6.5	24.3			
	6	7.4	26.3			
	Mean	9.0	25.7			
F8 1s ISO800	SRM	10.7	26.0	0.95 (0.01)	-	-
	2	9.5	26.9		0.938 (0.050)	0.997 (0.006)
	5	6.5	24.4			
	6	7.6	26.4			
	Mean	8.6	25.9			
F8 1.3s ISO400	SRM	7.8	25.9	0.95 (0.01)	-	-
	2	8.8	25.7		0.931* (0.018)*	0.999* (0.001)*
	5	6.0	24.4			
	6	7.2	26.5			
	Mean	7.5	25.6			
F11 1.3s ISO800	SRM	9.5	26.2	0.89 (0.02)	-	-
	2	9.1	27.1		0.896 (0.058)	0.999 (0.001)
	5	5.9	24.5			
	6	7.0	26.8			
	Mean	7.9	26.2			
F11 2s ISO400	SRM	11.0	26.1	0.90 (0.02)	-	-
	2	9.6	27.2		0.893 (0.035)	0.999 (0.001)
	5	6.2	24.5			
	6	7.4	26.7			
	Mean	8.6	26.1			

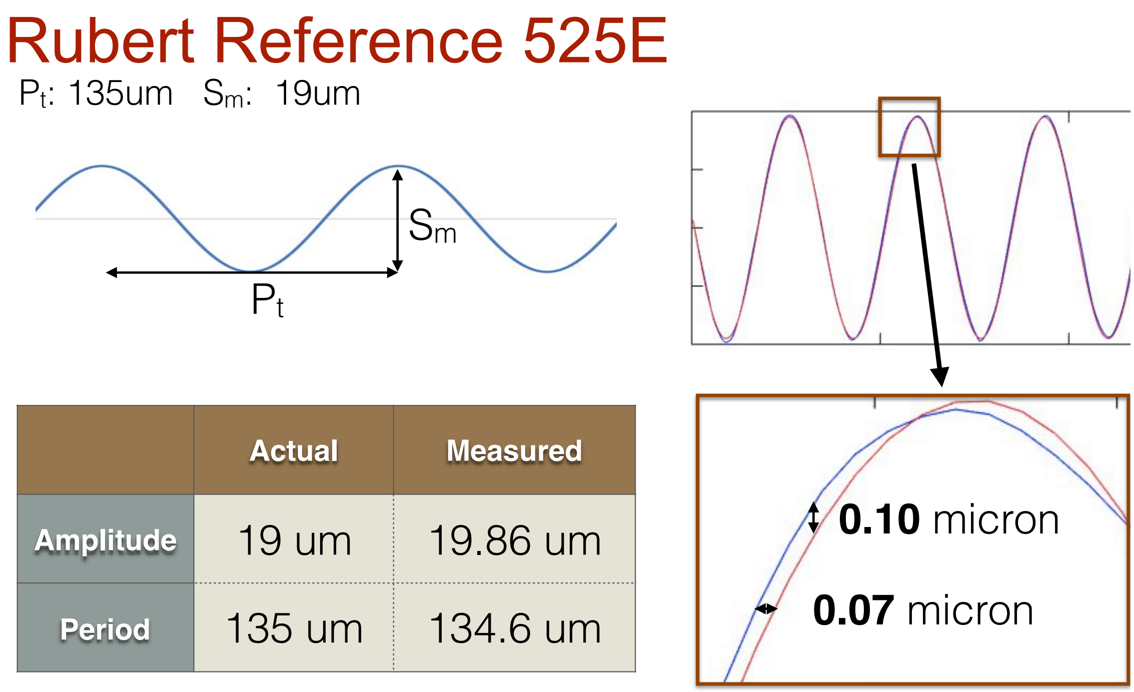
Table 3: **Aperture Variation Results.** Scan metrics are presented for the three different aperture settings on the Small test set (Three firearms and the NIST SRM casing). Firearm numbers correspond to Figure 2. F5.6 was evaluated at 1s:ISO400 (exposure time:ISO gain) and 1s:ISO200. F8 was evaluated at 1s:ISO800 and 1.3s:ISO400. F11 was evaluated at 1.3s:ISO800 and 2s:ISO400. Normalization for each firearm took place across all six camera settings. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Detected Features, Normalized Raw Match Score, and 01 Match Score are described in Section 3. The Match Score is the average (standard deviation) of those computed between corresponding pairs of matched test fires for the specified camera settings. For the three rightmost columns the mean values are reported with standard deviations in parentheses. *The inclusion of one outlier value causes these values to be lower than expected. The values shown are when the outlier has been removed. When the outlier is included the Normalized Raw Match Score becomes 0.889(0.104) and the 01 Match Score becomes 0.895(0.257).

Scan Iteration	Firearm	Flatness Score		Normalized Num Feat		01 Match Score
		Casing 1	Casing 2	Casing 1	Casing 2	
1	1A	6.7	8.3	0.95	0.96	0.987
	2	10.0	7.2	1.00	1.00	0.999
	3	5.3	10.5	1.00	0.96	0.991
	4	8.9	7.9	0.96	0.99	0.999
	5	4.8	4.8	0.96	0.98	0.999
	6	6.8	4.9	0.98	0.97	0.999
	Mean(Std)	7.2(2.0)		0.98(0.02)		0.996(0.007)
2	1A	11.6	7.9	0.92	0.96	0.932
	2	7.1	10.0	0.99	0.99	0.999
	3	7.9	11.0	0.98	0.98	0.999
	4	6.9	6.7	0.93	0.95	0.999
	5	6.9	9.3	1.00	0.96	0.999
	6	6.6	8.2	0.98	0.92	0.999
	Mean(Std)	8.3(1.8)		0.96(0.03)		0.988(0.026)
3	1A	6.7	6.2	0.97	0.98	0.891
	2	7.6	9.6	0.98	0.95	0.999
	3	3.7	8.6	0.98	1.00	0.988
	4	7.0	5.6	0.94	1.00	0.999
	5	3.3	6.5	0.94	0.98	0.999
	6	8.1	7.5	1.00	0.96	0.999
	Mean(Std)	6.7(1.9)		0.97(0.02)		0.980(0.042)
4	1A	5.7	7.4	0.98	0.94	0.698
	2	9.2	5.3	0.94	0.92	0.999
	3	3.8	5.2	0.96	0.96	0.004
	4	5.2	6.0	0.95	0.95	0.999
	5	2.4	2.5	0.97	1.00	0.999
	6	6.7	5.8	1.00	1.00	0.999
	Mean(Std)	5.4(1.9)		0.96(0.03)		0.784(0.382)
5	1A	6.2	6.4	1.00	1.00	0.887
	2	10.0	5.7	0.98	0.89	0.999
	3	5.6	8.3	0.93	0.95	0.996
	4	6.3	5.3	1.00	1.00	0.999
	5	2.0	2.7	0.94	0.98	0.999
	6	2.9	5.3	1.00	0.97	0.999
	Mean(Std)	5.6(2.3)		0.97(0.03)		0.981(0.044)

Table 4: **Repeatability Results.** Scan metrics are presented for five scan iterations (acquired over a two month period) of the Medium test set. Firearm numbers correspond to Figure 2. Normalization for each firearm took place across all five scan iterations. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Features, and 01 Match Score are described in Section 3. The Match Score is that computed between the pair of matched test fires for the specified iteration of the indicated firearm. The Mean(Std) rows list the mean and standard deviation of the above values. Small standard deviations indicate consistent results.

Firearm	Replicate Scan Number	Flatness Score		Normalized Num Feat		01 Match Score
		Casing 1	Casing 2	Casing 1	Casing 2	
1A	1	6.9	6.5	0.99	0.99	0.871
	2	6.7	6.5	1.00	0.99	0.810
	3	6.5	6.6	1.00	0.98	0.869
	4	6.7	6.6	0.99	0.99	0.898
	5	6.5	6.5	1.00	1.00	0.857
2	1	7.9	6.0	1.00	0.99	0.998
	2	8.1	6.3	0.99	1.00	0.993
	3	8.1	6.4	1.00	0.99	0.998
	4	8.2	6.1	1.00	1.00	0.994
	5	8.2	6.0	1.00	0.99	0.996
3	1	7.6	4.5	1.00	0.99	0.996
	2	8.0	4.5	1.00	0.99	0.988
	3	7.9	4.7	0.99	1.00	0.999
	4	8.0	4.7	0.99	1.00	0.997
	5	7.8	4.7	1.00	1.00	0.995
4	1	8.5	6.3	0.98	0.99	0.999
	2	8.5	6.5	0.99	1.00	0.999
	3	8.8	6.4	0.99	1.00	0.999
	4	8.5	6.3	1.00	1.00	0.999
	5	8.6	6.4	1.00	0.99	0.999
5	1	2.7	9.1	0.99	0.99	0.862
	2	2.6	9.4	1.00	1.00	0.965
	3	2.6	9.6	1.00	0.99	0.789
	4	2.6	9.6	1.00	0.99	0.991
	5	2.6	9.5	0.99	1.00	0.040
6	1	5.1	10.3	0.99	0.99	0.999
	2	5.4	10.4	0.98	0.99	0.999
	3	5.4	10.7	1.00	1.00	0.999
	4	5.5	11.0	1.00	1.00	0.999
	5	5.5	11.1	0.99	1.00	0.999

Table 5: **Precision Results.** Scan metrics are presented for the five replicate scans of each casing in the Medium test set. Firearm numbers correspond to Figure 2. The Flatness Score and Normalized Number of Features are listed for each of the two casings. Normalization for each firearm took place across all five scan replicates. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Features, and 01 Match Score are described in Section 3. The Match Score is that computed between the pair of matched test fires for the specified replicate of the indicated firearm. Averages and Standard deviations were not computed because of the consistency observed.



	Actual	Measured
Amplitude	19 um	19.86 um
Period	135 um	134.6 um

Figure 5: **Sinusoidal Reference Standard.** A system performance check was developed around a sinusoidal reference standard with wavelength $135\mu m$ and amplitude $19\mu m$. An automated quality check analysis demonstrates that lateral and depth resolution is typically better than $1\mu m$. See Section 3.

Time	Orig Casing	Target Casing	Target		01 Match Score
			Flatness Score	Normalized Num Feat	Target→Orig
0 min	F2 C1	F2 C1 (ID)	8.0	1.00	-
	F2 C1	F2 C2 (KM)	7.2	1.00	0.999
	F2 C1	SRM (KNM)	8.9	1.00	0.001
	F5 C1	F5 C1 (ID)	3.0	1.00	-
	F5 C1	F5 C2 (KM)	3.5	1.00	0.999
	F5 C1	SRM (KNM)	9.6	0.99	0.001
	F6 C1	F6 C1 (ID)	3.5	1.00	-
	F6 C1	F6 C2 (KM)	6.6	1.00	0.999
	F6 C1	SRM (KNM)	7.1	1.04	0.001
5 min	F2 C1	F2 C1 (ID)	5.5	1.00	-
	F2 C1	F2 C2 (KM)	5.2	0.97	0.999
	F2 C1	SRM (KNM)	7.5	0.97	0.001
	F5 C1	F5 C1 (ID)	2.7	0.99	-
	F5 C1	F5 C2 (KM)	4.9	1.01	0.999
	F5 C1	SRM (KNM)	6.1	0.97	0.001
	F6 C1	F6 C1 (ID)	4.5	1.00	-
	F6 C1	F6 C2 (KM)	8.1	1.01	0.999
	F6 C1	SRM (KNM)	8.4	1.06	0.001
10 min	F2 C1	F2 C1 (ID)	5.4	0.97	-
	F2 C1	F2 C2 (KM)	4.4	1.00	0.999
	F2 C1	SRM (KNM)	15.2	0.97	0.001
	F5 C1	F5 C1 (ID)	3.0	0.97	-
	F5 C1	F5 C2 (KM)	7.0	1.01	0.999
	F5 C1	SRM (KNM)	4.7	0.98	0.001
	F6 C1	F6 C1 (ID)	3.4	1.01	-
	F6 C1	F6 C2 (KM)	6.9	0.99	0.999
	F6 C1	SRM (KNM)	14.0	1.00	0.001
30 min	F2 C1	F2 C1 (ID)	8.7	1.03	-
	F2 C1	F2 C2 (KM)	6.2	1.03	0.999
	F2 C1	SRM (KNM)	7.1	0.97	0.001
	F5 C1	F5 C1 (ID)	5.3	0.97	-
	F5 C1	F5 C2 (KM)	6.2	1.02	0.999
	F5 C1	SRM (KNM)	8.1	1.03	0.001
	F6 C1	F6 C1 (ID)	3.5	1.00	-
	F6 C1	F6 C2 (KM)	7.3	1.02	0.999
	F6 C1	SRM (KNM)	9.0	1.06	0.001

Table 6: **Persistence Results.** Scan metrics are presented for the three time durations. One test fire from each of the three firearms in the Small test set were used as the original casing. Firearm numbers correspond to Figure 2. Fx Cy indicates firearm x and casing number y. Therefore F2 C1 and F2 C2 are sister casings and constitute a known match. The original casings were placed into the gel for 5, 10, and 20 minutes. The casing was removed from the gel. A specified target casing (from a different firearm) was then scanned as normal. The table reports the Flatness Score and Normalized Number of Features for the target scan. The 01 Match Score is shown for the target scan compared to the original casing. Target casings include a known match and a known non-match. Normalization for each firearm is relative to the baseline results. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Features, and 01 Match Scores are described in Section 3. Averages and Standard deviations were not computed because of the consistency observed. Note that in these experiments the 01 Match Score only considered the Breech-Face Impression because we were focused on the persistence of impressed marks.

Operator Class	Firearm	Flatness Score	Normalized Num Feat	01 Match Score
Novice Set A	1A	7.2 (0.5)	0.98 (0.01)	0.786 (0.155)
	2	7.8 (1.0	0.98 (0.01)	0.999 (0.001)
	3	6.8 (1.2)	0.98 (0.01)	0.431 (0.490)
	4	6.4 (0.9)	0.97 (0.03)	0.999 (0.001)
	5	5.8 (2.5)	0.97 (0.03)	0.996 (0.007)
	6	6.9 (1.1)	0.98 (0.03)	0.999 (0.001)
	Mean	6.8	0.98	0.868
Experienced* Set A	1A	6.3	0.98	0.887
	2	7.9	0.93	0.999
	3	7.0	0.91	0.996
	4	5.8	0.92	0.999
	5	2.4	0.93	0.999
	6	1.1	0.97	0.999
	Mean	5.1	0.94	0.980
Novice Set B	1B	6.5 (0.9)	0.93 (0.04)	0.733 (0.225)
	2	6.9 (2.2)	0.93 (0.03)	0.999 (0.001)
	3	7.2 (1.8)	0.94 (0.06)	0.834 (0.322)
	4	6.8 (0.7)	0.89 (0.07)	0.999 (0.001)
	5	6.5 (1.4)	0.96 (0.04)	0.918 (0.131)
	Mean	6.8	0.93	0.897
Experienced Set B	1B	9.3 (4.4)	0.90 (0.09)	0.740 (0.271)
	2	7.8 (1.8)	0.94 (0.06)	0.999 (0.001)
	3	6.9 (1.7)	0.92 (0.06)	0.720 (0.466)
	4	6.7 (1.8)	0.95 (0.04)	0.999 (0.001)
	5	5.6 (1.2)	0.94 (0.04)	0.995 (0.002)
	Mean	7.3	0.93	0.891

Table 7: **Interoperator Scanning Study - Scan and Match Quality.** The mean and standard deviation of each metric for each of the six firearms in the two Medium test sets (A & B) separated by operator experience (Novice vs. Experienced). Firearm numbers correspond to Figure 2. For firearms that appear in both Sets A and B, different test fires appear in Sets A and B. The version of Set B distributed to participants did no include firearm 6 and thus Set B only has five firearms. Normalization for each firearm took place across all scan replicates. Additional details on each protocol are provided in Supplementary Material. The Flatness Score, Normalized Number of Features, and 01 Match Score are described in Section 3. The Match Score is that computed between the pair of matched test fires for the specified firearm. The Mean rows list the average of the corresponding values. *Note that only one expert operator collected test set A; there is therefore only a single value for each firearm and means and standard deviations are not reported.

Operator Class	Operator	% One	Surf. Contam. Mean	Surf. Contam. Stdev
Novice	1	100%	1.00	0.00
	2	92%	1.08	0.29
	3	100%	1.00	0.00
	4	30%	1.70	0.48
	5	40%	1.60	0.52
	6	90%	1.10	0.32
	7	90%	1.10	0.32
	Mean	77%	1.23	
Experienced	1	100%	1.00	0.00
	2	80%	1.20	0.42
	3	100%	1.00	0.00
	4	90%	1.10	0.32
	Mean	93%	1.08	

Table 8: **Interoperator Scanning Study - Cleanliness.** The two rightmost columns report the mean and standard deviation of the Surface Contamination scores for the indicated operator. The % One column lists the percentage of scans that had a Surface Contamination score of 1 (the cleanest score). The Surface Contamination score is described in Section 3.

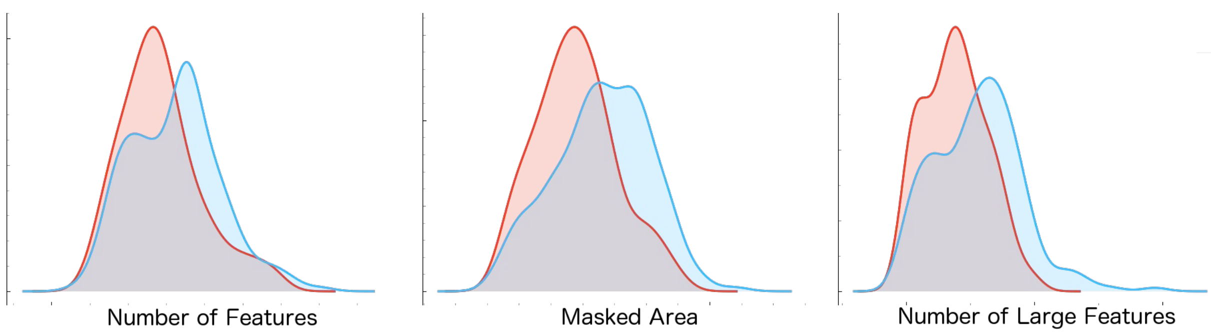


Figure 6: **Match Likelihood Correlation** Probability distribution functions for casings based on their best Known Match (KM) 01 Match Score. Casings whose best KM scores greater than 0.5 are shown in blue and casings whose best KM scores 0.5 or lower are shown in red. The distributions show that casings with an increased number of detected features, an increased masked area (*eg.*, the size of the masked breech-face impression), and an increased number of large detected features are more likely to have a higher scoring best KM (blue distribution vs. red distribution). The effect in this analysis is minor; however, a multivariate analysis (currently underway) may be more informative.

Supplementary Material

This supplementary material section contains additional detail on each of the completed experiments and protocols.

Details of Cleaning Method Experiments: Abbreviations and definitions used below: **BRSH**: soft nylon brush (typically used to clean the casing head); **CMP**: Compressed Air (Endust brand multi-purpose duster compressed air spray can); **CTTN**: Cotton Swab (typically dipped in IPA and then used to clean the casing head); **IPA**: Isopropyl Alcohol (91%); **Manual Air**: Rocket Blower (a manual squeeze ball that emits a puff of air when squeezed (a manual version of compressed air); **SPBRSH**: Sponge Brush (a small foam tipped painting brush, used to clean the casing head); **Tape**: Clear polypropylene tape with acrylic adhesive (a clear tape used for cleaning with adhesive that will not leave a residue behind, when applied to a casing, the tape is gently applied to the surface and then removed by pulling at an angle); **WD**: WD-40 lubricant and solvent spray (WD-40 contains both a lubricant (which may minimize the risk gel wrinkling) and a proprietary solvent (which like IPA can serve to clean the casing surface)).

The listed of solvents, mechanical cleaning agents, and drying agents were evaluated in a series of experiments (see Section 3 and Table 1). Camera settings of F8 1.3s ISO400 were used.

- **Dirty:** Casings placed in bag of dirt, bag was shaken, casings removed and tapped on a hard surface to remove large clumps of dirt.
- **IPA + Manual Air:** Starting with dirty casings, the casing head was dipped in IPA, casing was shaken to remove bulk liquid, 15-20 blasts from the Rocket Blower were applied (approximately 1” from the casing).
- **IPA + Compressed Air:** Starting with dirty casings, the casing head was dipped in IPA, casing was shaken to remove bulk liquid, 6-8 short blasts of compressed air (approximately 2” from the casing)..
- **IPA + Brush:** Starting with dirty casings, the casing head was dipped in IPA, a soft brush was used to apply 10 circular brush strokes followed by 10 linear strokes. No explicit drying process was used.
- **IPA + Brush + Manual Air:** Starting with dirty casings, the casing head was dipped in IPA, a soft brush was used to apply 10 circular brush strokes followed by 10 linear strokes, 15-20 blasts from the Rocket Blower were applied (approximately 1” from the casing).
- **IPA + Swab + Manual Air:** Starting with dirty casings, the casing head was dipped in IPA, a cotton swab dipped in IPA was used to clean the casing head (approximately 10 strokes), 15-20 blasts from the Rocket Blower were applied (approximately 1” from the casing).
- **IPA + SpBrush + Manual Air:** Starting with dirty casings, the casing head was dipped in IPA, a sponge brush was used to clean the casing head (approximately 15 strokes), 15-20 blasts from the Rocket Blower were applied (approximately 1” from the casing).
- **WD + Brush:** Head dipped in a small amount of WD-40, a soft brush was used to apply 10 circular brush strokes followed by 10 linear strokes. No explicit drying process was used.
- **Tape:** Clear cleaning tape was pressed onto head of casing, a finger was used to make sure good contact was made, tape removed and reapplied, finger pressure applied, tape removed and reapplied a third time, finger pressure applied. A different part of the tape was used each time.

- **Tape + IPA + Manual Air:** Clear cleaning tape pressed onto head of casing and finger pressure applied a total of three times (see 'Tape' above), head was then dipped in IPA, finally 15-20 blasts from the Rocket Blower were applied (approximately 1" from the casing).

Each scan was evaluated using the Normalized Number of Detected Features, 01 Match Score, and Surface Contamination measures (Section 3).

Details of Lighting Variation Experiments: Scans for the lighting experiment were collected using the following materials and methods.

Light Sources

- **Standard Incandescent**: EcoSmart Energy Efficient Incandescent Bulbs (2990K Temperature Spectrum)
- **Standard Fluorescent**: EcoSmart Compact Fluorescent Bulbs (5000K Temperature Spectrum)
- **Full Spectrum Fluorescent**: LimoStudio Full-Spectrum Compact Fluorescent (6500K Temperature Spectrum)

Each of the above light sources was installed into one or more tripod mounted light sockets and positioned near the scanner to achieve a range of different light intensities. Lights were positioned to achieve lux measurements of 750, 1200, 3000, and 8000 lux as measured using a Digital Lux Meter LX1330B (range: 0-200,000 lux; accuracy: +/- 3% (<20,000 lux), +/- 5% (>20,000 lux)). In addition, scans were collected in a black-out baseline condition (0 lux) and a typical office lighting baseline condition (400 lux using office installed overhead lights). In all cases, a calibration was performed at the start of scanning under the same lighting conditions as the rest of the scan set. The one exception to this was a final test that collected scans in 880 lux Full Spectrum light but which used used a calibration set from the black-out condition.

For reference: 1 lux: full moon on clear night, 80 lux: office hallway, 300-500 lux: office lighting, 1000 lux: TV studio lighting, 10,000 lux: bright daylight.

Each scan was evaluated using the Flatness, Normalized Number of Detected Features, and Normalized Raw Match Score, and 01 Match Score measures (Section 3).

Details of Aperture Variation Experiments: The following aperture settings were considered. For each aperture both the exposure time and ISO gain were adjusted to ensure a full range of pixel intensities in the collected images. For example, a smaller aperture (larger f-number) increases the depth-of-field; however, as less light makes it through the aperture to the camera sensor it becomes necessary to increase the exposure time or ISO gain. It is generally known that increasing the exposure time or increasing the ISO gain both result in slightly less focused images. Therefore while an increase in depth-of-field may theoretically allow more field-of-view to be in focus this increase may be offset by the decrease in focus resulting from the exposure and ISO adjustments. DOF: Depth of Field.

- **F5.6 1s ISO400**: DOF $176\mu m$
- **F5.6 1s ISO200**: DOF $176\mu m$
- **F8 1.3s ISO400**: DOF $249\mu m$
- **F8 1s ISO800**: DOF $249\mu m$

- **F11 2s ISO400**: DOF $352\mu m$
- **F11 1.3s ISO800**: DOF $352\mu m$

Each scan was evaluated using the Flatness, Mean Feature Size, Normalized Number of Detected Features, Normalized Raw Match Score, and 01 Match Score measures (Section 3).

Details of Repeatability Experiments: Each casing was scanned five times over a two month period. The same sheet of gel was used for all scan sets while a different individual piece of gel was used for each set. The scanner was calibrated during the collection of each scan set. Camera settings of F8 1.3s ISO400 were used. Scans were collected in blackout conditions.

Each scan was evaluated using the Flatness, Normalized Number of Detected Features, and 01 Match Score measures (Section 3).

Details of Precision Experiments: Each casing was scanned five times consecutively without reseating the casing into the gel. The scanner was calibrated one time during data collection. Camera settings of F8 1.3s ISO400 were used. Scans were collected in blackout conditions.

Each scan was evaluated using the Flatness, Normalized Number of Detected Features, and 01 Match Score measures (Section 3).

Details of Persistence Experiments: The persistence experiments have the most complex data collection schedule. Three casings were selected as **Impression** casings. In each experiment, the Impression casing was pressed into the gel, scanned, and then left to sit for a specified amount of time (0, 5, 10, or 30 minutes). After the elapsed impression time, three additional scans were collected (a known match, a known non-match (the NIST SRM casing), and scan of flat glass).

The scanner was calibrated several times through data collection. Each impression casing used two gel punches (one for the 0, 5, and 10 min impressions and one for the 30 min gel impression). All gel punches came from a single gel sheet. Camera settings were F8 1.3s ISO400. Scans were collected in blackout conditions. Each scan was evaluated using the Flatness, Normalized Number of Detected Features, and 01 Match Score measures (Section 3).

Details of Interoperator Study: Ten operators participated in the interoperator scanning study. Participants were classified as experienced or novice. Experienced participants typically had at least six months of scanning experience. Novice participants were typically new to scanning (this being their first scanning experience). Novice participants received approximately two hours of training prior to data collection (presentation and hands-on). An experienced operator typically watched the novice through the first few casings and gave corrective comments. The operators were on their own for the remainder of data collection. Participants included four trained firearms examiners, two researchers, three forensic science students, and one layperson. Data was collected at four different sites (four different scanners were used). Two similar casing sets were utilized in the study (Fig. 2). Set A was scanned by three novice and one expert operator. Set B was scanned by four novice and three expert operators. Each scan was evaluated using the Flatness, Normalized Number of Detected Features, 01 Match Score, and Surface Contamination measures (Section 3).

References

[1] M. Johnson and E. Adelson. Retrographic sensing for the measurement of surface texture and shape. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1070–7, 2009.

[2] M. Johnson, F. Cole, A. Raj, and H. Adelson. Microgeometry capture using an elastomeric sensor. *ACM Trans. on Graphics, Proc. of SIGGRAPH*, 30:139–44, 2011.

[3] 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*, 47:198–208, 2015.