Peer Reviewed
Received 30 January 2015
Received in revised form 7 Martch 2016
Accepted 14 April 2016
Available online 27 December 2017

Forensic Science Seminar
ISSN 2157-118X
Volume 7 Number 2
31 December 2017

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

Ryan Lilien, *Ph.D.* [a, *]

a Cadre Research Labs, LLC, Chicago, IL 60654

* CORRESPONDING AUTHOR. Chief Scientific Officer. Add: Cadre Research Labs; 212 W Superior St; Suite 503; Chicago, IL 60654

ABSTRACT This research has investigated and developed a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge cases. The project, named Top-Match, combines the recently developed GelSight high-resolution surface topography imaging system with state-of-the-art algorithms for matching structural features. This project aims to extend the system to measure and compare striated toolmarks (e.g., aperture shear), to integrate these marks into the scoring function, and to investigate matching algorithms for comparing 3D surface topographies captured using different imaging modalities. Compared to competing technologies, it is fast, inexpensive, and not sensitive to the optical properties of the material being measured.

KEY WORDS 3D, topography, tool marks, firearm identification, GelSight, forensic science

I PROJECT PURPOSE AND BACKGROUND

In the described work, we investigated and developed a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge cases. We demonstrated the system's potential for increasing the quality and reducing the cost of forensic analyses. Several recent studies have called for improved imaging technology and matching algorithms to support firearm identification. Our project, named Top-Match, combines the recently developed GelSight high-resolution surface topography imaging system with state-of-the-art algorithms for matching structural features. Compared to competing technologies, our GelSight based system is fast, inexpensive, and not sensitive to the optical properties of the material being measured. This project aims to extend the system to measure and compare striated toolmarks (e.g., aperture shear), to integrate these marks into the scoring function, and to investigate matching algorithms for comparing 3D surface topographies captured using different imaging modalities (e.g., GelSight vs. confocal microscopy).

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 theMIT 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 also worked with colleagues at NIST and at the International Forensic Science Laboratory & Training Centre in Indianapolis (Dr. James Hamby). We continue to work with Andy Smith (San Francisco PD), Chris Coleman (Contra Costa County Office of the Sheriff), and Karl Larsen (U. Illinois at Chicago). These collaborators continue to be excellent partners and provide both scans and constructive feedback. The results described below made use of a large set of new and previously collected test fires.

II PROJECT DESIGN

Our one year project has three aims. First, we extended our base system so that it can capture, extract, and compare striated toolmarks starting with Aperture Shear (aka Primer Shear) marks. Second, developed a statistical confidence scoring model incorporating breech-face impression and primer shear marks. Finally, investigated algorithms for Cross-Modality Matching between GelSight and Confocal Microscopy scans. All 3 proposed aims were successfully completed during the project period.

2.1 Materials and Methods

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 focusvariation microscopy, the use of a painted elastomeric gel removes the influence of surface reflectivity on the measured topography. The scanner contains a linear xy-stage that allows fine positioning control (Fig 1F). The setup contains an 18-megapixel Canon digital camera with a 65mm macro lens. The setup supports up to 0.9μm/pixel; 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 1.4μm/pixel lateral resolution with submicron depth resolution. This resolution is appropriately matched to firearms forensics as firearms examiners typically consider toolmarks ranging from tens to hundreds of microns in diameter. To facilitate scanning, we redesigned the casing holder to more easily hold potentially damaged casings. The new holder utilizes a similar sliding and raising mechanism but now holds the casing using its strong extractor groove (in a manner similar to a kinetic puller) (Fig 1A). The operator retracts the holder insert using a single motion and places the casing's extractor groove on an elevated notch of the insert. The insert is replaced into the holder, slid under the light-plate, and a lever is used to raise the holder and casing firmly into the gel (Fig 1B-E). The process is reversed to remove the case. The design continues to use the angled flexure we implemented last year to minimize the risk of air trapping. The flexure holds the casing at 2-degrees off level when the holder makes initial contact with the gel. The flexure is pliable and gives way to a level orientation when the holder meets the resistance of the gel. The result is no air-trapping and a level case. The new holder was used for most of the project's data acquisition. Scan acquisition requires ~2 minutes per casing.

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) [1]. 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. That is, a file from our scanner can record the date, calibration information, objective, microns per pixel, and site-specific data while a confocal microscope scan can record parameters specific to that technology. Our group, Cadre, has created and distributed free software for reading and writing X3P to the forensic community. This software is currently being used by NIST in preparing their firearms forensics database (part of their current year NIJ award). We are encouraging all equipment manufacturers to adopt the reading and writing of X3P. We believe the work of the OpenFMC group is an important step forward for the field.

[1] The OpenFMC group includes members of Cadre Research Labs, NIST, academia, and another commercial vendor.

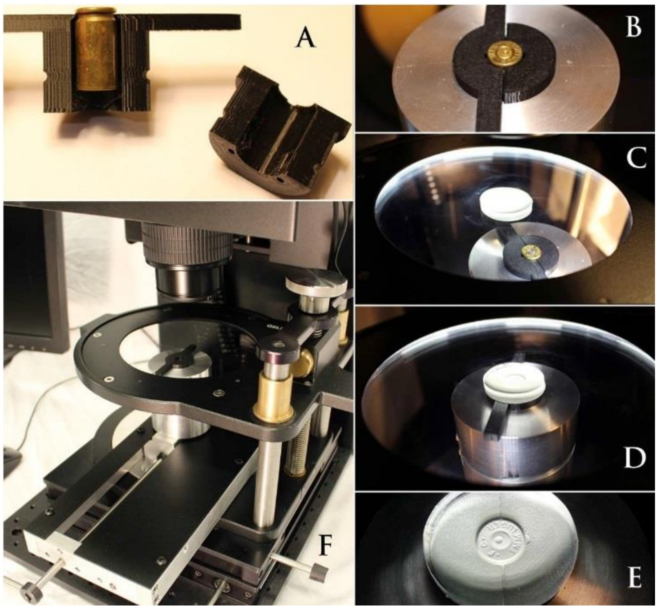

Fig.1 Scanning System and Holder. A new prototype mount is shown with a cartridge casing. The casing is held at the extractor groove. Two halves of the mount completely surround the casing. The mount assembly is inserted into a metal holder (B) that is slid under a piece of gel attached to the lightplate (C). A raising lever raises the holder and mount assembly into the gel (D). The pigmented surface of the gel contours to the casing and is ready for scanning (E). The entire scanning assembly (lightplate, objective lens, holder, and adjustment levers) is shown in panel (F).

Breech-Face Impressions: In the developed prototype (ver 0.9), automatically identified distinctive features are used to match and align two casings. By requiring spatial coherence of matched features, the methodology is able to strongly indicate when two casings were fired through the same firearm. 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). Good feature points include the same types of ridges, peaks, gouges, and concavities that a trained firearms examiner would identify. The intuition behind the characterization of these geometric feature points is illustrated in Figure 2. By measuring the Hessian (that is, the matrix of second derivatives) the surface can be locally classified in terms of its shape. Peaks and saddle points are considered well localizable because, no matter which direction one moves, the surface changes. On the other hand, edges and plateaus do not have this property and are therefore less desirable. These two instances can be distinguished, by looking at the determinant of the Hessian which is non-zero for peaks and saddle points and zero (or small) for plateaus and edges. When a typical casing is entered into the system, TopMatch identifies thousands of peak and saddle point like features at multiple scales. The complex features identified by a human examiner are composed of multiple smaller peaks and saddle points; therefore, by matching large numbers of automatically identified peaks and saddle points we can parallel the human feature-based matching process. Feature detection takes place during scan acquisition and only happens once per casing. That is, once the breechface impression features are extracted they do not need to be extracted again. This important detail makes feature-based comparison inherently faster than cross-correlation methods.

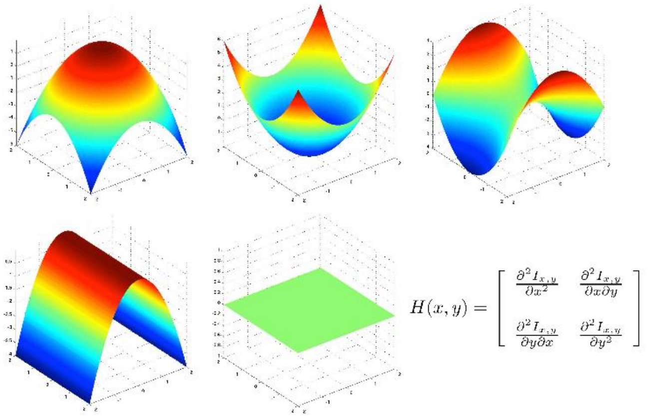

Fig.2 Representative Feature Patches. Representative shapes of local surface patches. (Top) Peaks and Saddle Points are good features and have a Hessian near 1. (Bottom) Edges and Plateaus are less informative and have a Hessian near 0. (Bottom Right) The Hessian matrix (H) defined in terms of the 3D surface I. The determinant of this matrix is a scalar value often referred to as the Hessian.

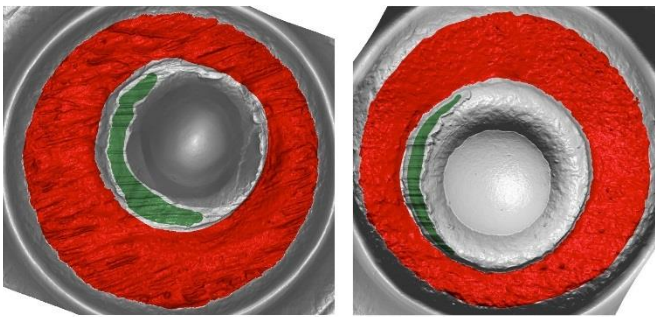

Fig.3 Breech-Face and Aperture Shear Impression Mask. The auto-masked region containing the breechface impression (red) and the aperture shear (green) for two casings (Norinco (left), Ruger (right)). In most cases, the breech-face impression mask identified by the auto-masker can be used directly, on occasion the user will want to manually tweak the mask around the edges. Because the aperture shears of these two casings do not lie in a rectangular region traditional extraction algorithms may have difficulty extracting the underlying linear profiles. Our extraction algorithm is able to extract the linear profile from both casings (see Figure 4).

To compare the breech-face impressions of two casings, the TopMatch matching algorithm identifies a maximal set of the detected features that are geometrically consistent between the two casings. 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. We have developed a heatmap visualization method to illustrate the location and density of matched features (Fig 9). The heatmap provides interpretability to complement the numerical match score.

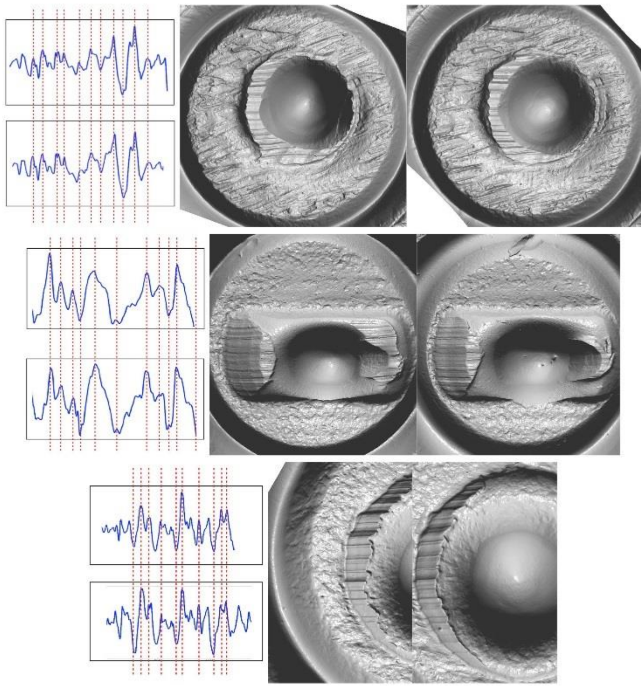

Fig.4 Extracted Aperture Shear Marks. Extracted profiles from each of two casings for each firearm are shown (Norinco (top), Glock (middle), and Ruger (bottom)). Shears can lie on various baselines and can be curved, arc'd, or flat. Red vertical lines indicate the positions of corresponding peaks and troughs. The corresponding extracted profiles are similar and can be matched using the developed algorithms.

Aperture (Primer) Shears: The aperture shear and breech-face impression are very different type of toolmarks. Whereas the breech-face impression is an impressed mark formed by direct force, the aperture shear is a striated mark formed by a shearing force. The unique information in a striated mark is contained in its linear profile and not in its overall surface. That is, it is not important if the shear is short or extended or if it rides on a sloped baseline vs a plateau baseline. This is demonstrated in that firearms examiners compare aperture shears using a split-screen side-by-side comparison of the linear striae. Therefore, the linear profile of the aperture shear should be considered separately from the breechface 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 approach treats them separately.

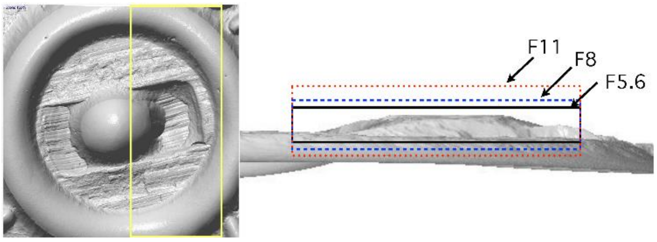

Fig.5 Visualization of Focal Depth. (Left) Scan of a casing from a Glock firearm with a large (high) flowback. (Right) Side view of boxed region showing height profile of the flowback and the focal depths for three f-stop settings. The flowback is 113μm in height, an F5.6 (black box) achieves a depth of 176μm, F8 (blue dashed box) achieves 249μm, and F11 (red dotted box) achieves 352μm. All three can capture an image with both the breech-face impression and top of the flow-back in focus; however, larger f-stop values make this easier.

We have developed a robust advanced aperture shear extraction algorithm (Aim 1). The system asks the user to indicate the presence and location of a shear mark during the masking process (green region in Fig 3). The masking takes place in our 3D viewer and the user can zoom and rotate the surface in three-dimensions. The user can also move the virtual light to graze the shear marks and bring out strong shear lines. We note that the algorithm extracts the shear profile directly from the measured 3D surface and is thus independent of the position of the virtual light source. 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. Once extracted, the linear profile is analyzed to identify peaks and valleys corresponding to traditional striae. A typical profile will have 10-30 detected peaks/valleys. Each detected extrema is parameterized based on its location, width, and local shape. Extrema are considered plausible matches if their aligned positions, size and shape match to within a threshold. Unlike older methods our approach can extract profiles from curved, sloped, and warped shears like those shown in Figure 3. Overall, our masking and extraction approach seems to work well (Fig 4).

Depth of Field: To better capture steep slopes and potentially raised aperture shears, we modified our scanning protocol to use an aperture of F8 and ISO 200. This is different from the older protocol which utilized F5.6 and ISO 100. The new settings increase the depth of focus from 176μm to 249μm while maintaining the same image brightness. Fortunately, most toolmarks on the primer are extremely shallow (fewer than 20μm). Figure 5 shows a Glock casing with a large flowback that measures 113μm in height.

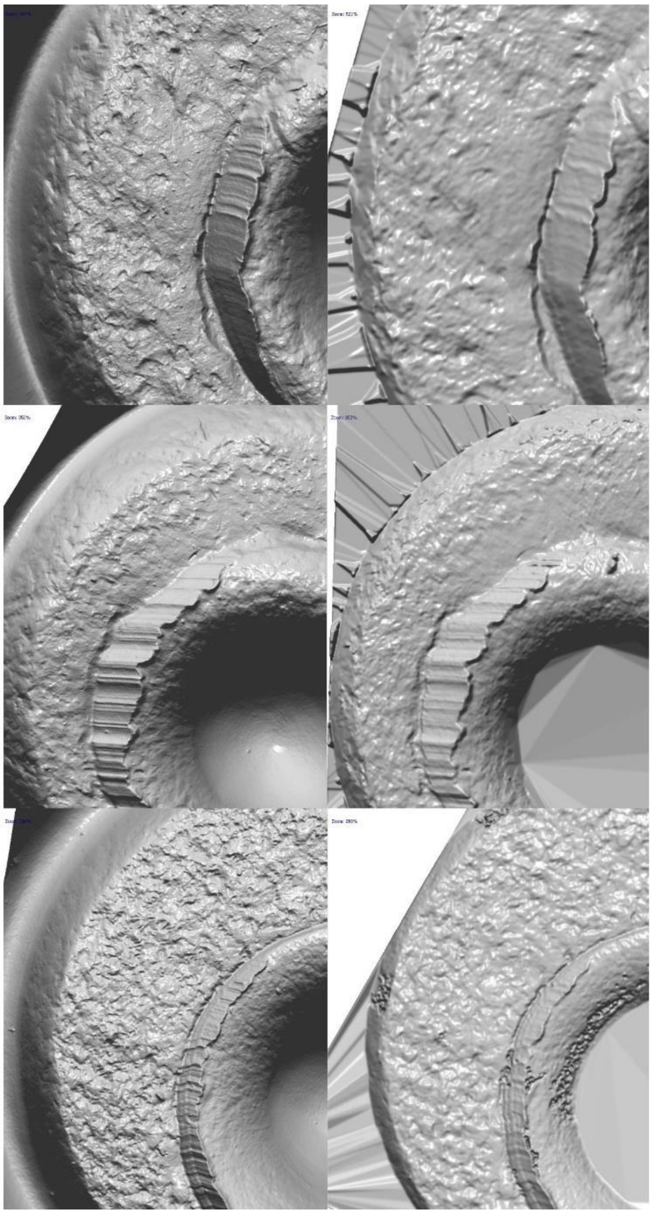

Fig.6 Cross-Modality Scans. GelSight scans (left) vs Confocal scans (right) for three casings from Sets 5 and 6. (Top) Ruger from Set 5, (Middle) Ruger from Set 5, (Bottom) Ruger from Set 6. GelSight scans are acquired at 1.4μm/pixel, Confocal scans are acquired at 3.1μm/pixel.

Confidence Score: A useful scoring algorithm reports a meaningful numeric score indicating the degree of similarity between two casings (Aim 2). A meaningful score allows a threshold to be set where a pair of casings with a score above the threshold is highly likely to have a common origin. A meaningful score allows an examiner to stop looking at a ranked list of candidate matches once the match scores fall below a threshold. The TopMatch system implements a meaningful match score achieving these goals using a logistic regression of several objective measures of similarity. The TopMatch scoring method compares the breech-face impression independent from the aperture shear and then unifies these two scores in an overall match score. The breech-face impression logistic regression utilizes the total number of matched features, the individual similarity (scale, pose) between corresponding features, the masked region of the breech-face impression, and the percent of masked region covered by the matched features. The aperture shear regression utilizes the total number of matched extrema and the total number of unmatched features. Casing-to-casing comparison requires approximately 1 second of compute time (on a single core). gobbleIt is possible to decrease the comparison time.

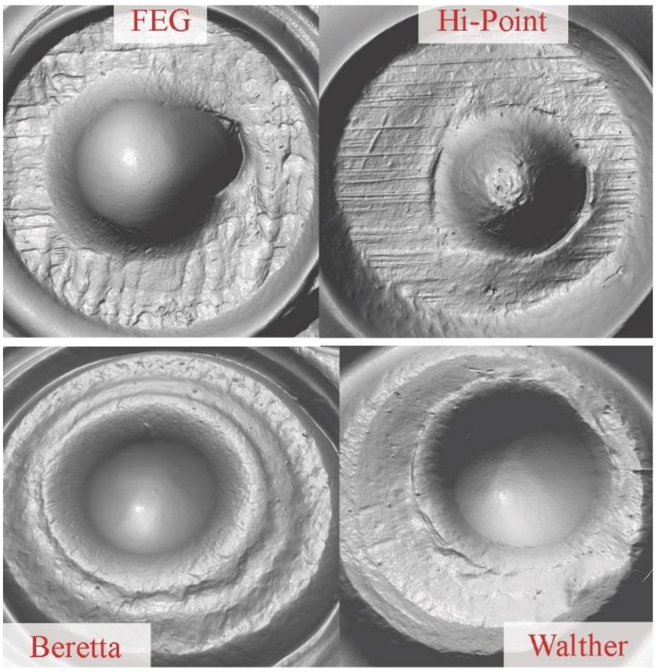

Fig.7 Sample Scans. (Top) Two casings (FEG and Hi-Point) that are well marked and for which the correct matching casing was found at the top ranked position. (Bottom) Two casings that are poorly marked. The Beretta has a very small breech-face impression and theWalther has very few surface features. We are able to find a correct match for the Beretta whereas no casings were deemed statistically significant for theWalther. Therefore, although the Walther did not find its match there were no false positives.

The confidence score lies between 0 and 1, but it is not a probability; that is, a score of 0.9 does not mean 90% chance of a match. However, because the score is directly correlated with toolmark similarity, because it is on the same scale and comparable between two comparisons, because it is numerical, 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. This probability is inherently conditioned on the underlying dataset which demonstrates the importance of including several firearm, toolmark, and ammunition types. Our regression model is fit using the collected data. It therefore includes over 24 firearm manufacturers and 5 brands of ammunition. The model will be extended as we collect additional scans. As we incorporate more data, we will be able to determine if the new data deviates in some unexpected way from the initial data.

Cross-Modality Matching: The developed analysis algorithms can compare any two 3D surface topographies (e.g., a GelSight scan and a confocal microscopy scan). In our datasets, the confocal scans contain drop-out pixels (locations where the scanner was unable to measure the surface topography) and high-frequency noise. We discovered that it was necessary to apply two preprocessing steps prior to matching. We first use linear interpolation to replace drop-out pixels and then apply a low-pass filter (σ=0.0062mm) to remove high frequency noise. We are grateful to Alan Zheng (NIST) for providing the confocal scans. Datasets 5 and 6 contain the cross-modality scans. Figure 6 shows the difference in resolution between the GelSight and Confocal scans.

Partner Labs: We continue deployments with four labs. Three labs in the Bay area including Todd Weller (Oakland) and Chris Coleman (Contra Costa County). We also have a deployment with Dr. Karl Larsen at the U. of Illinois. All labs are collecting scans of test fires and are providing useful feedback.

2.2 Datasets

The accuracy of the TopMatch scanning hardware and software was assessed through a series of experiments using six 9mm Luger datasets. The sets are chosen to evaluate breech-face impression matching, aperture shear matching, well marked casing matching, and cross-modality matching. Unless otherwise noted, all TopMatch scans were acquired at a spatial resolution of ~1.4μm/pixel. Prior to acquisition all casings were cleaned with a mild solvent (isopropyl alcohol) and a soft brush.

Note that we collected scans for all casings in each set. This approach differs from the way many labs use their scanning and database systems. In a typical scenario, an examiner will select and enter only the best test fire among all those available. If all test fires are extremely poorly marked then an examiner may elect to not enter any of them. In a laboratory setting, the motivation for this exclusion is to prevent the database from acquiring too many bad scans that may later result in false positive matches. In contrast to this exclusion approach, in the datasets below, the test fires were not prescreened. As a result, many of the casings that are included in the matching accuracy studies below would not have been deemed high enough quality for entering into a traditional database system.

(Set 1) Forty Seven Firearm Set: The first dataset includes 47 firearms: 2x Colt, 5x Hi-Point, 7x Fabrique Nationale, 5x S&W, 5x Radom, 16x Ruger (including 10 with consecutively manufactured breech-faces), 5x Norinco, 1x FEG, 1x Springfield Armory. The firearms were selected from two police department's reference collections without preference to their ability to mark cartridge casings. The intent was to select firearms that would represent real-world conditions in terms of toolmark quality and type (e.g., filing marks, granular marks, milled marks). In this set, the toolmarks left on a casing range from being extremely reliable and interpretable (Fig. 7 top row) to being unreliable, irreproducible, and barely present (Fig. 7 bottom row). A trained firearms examiner, Todd Weller (Oakland), manually examined a number of the collected casings and made a note that at least ten of the 47 firearms did not mark well. Most of the selected firearms came from the Oakland reference collection; some of the firearms came from the San Francisco reference collection courtesy of our collaborator Andy Smith. In Set 1, there are three test fires from each firearm (PMC Brand, 115GR bullet, brass casing and primer).

(Set 2) One Hundred One Firearm Set: This set includes test fires from the 47 firearms in Set 1 and from 64 additional firearms. The set includes 337 casings, 26 of the casings do not have a known match in the set and 311 of the casings have one or more known matches. The set includes two additional test fires from the 47 firearms in Set 1 – Remington Brand ammunition (115GR bullet, brass casing, nickel primer). Therefore some casings in the set have a single known match, some have multiple known matches, and some have zero known matches. A total of 101 firearms are represented among 431 casings from the following firearm manufacturers: Armi Fratelli, Baikal, Beretta, Browning Arms, Bryco Arms, Colt, Hi-Point, Fabrique Nationale, FEG, Heckler & Koch, Intratec, Kahr Arms, Keltec, S&W, Radom, Ruger, Norinco, Sig Sauer, Springfield Armory, Star, Taurus, Uzi, and Walther. At least seven ammunition manufacturers were represented: Federal, Fiocchi, PMC, RWS, RP, Speer, Winchester, and 'Unknown'. Scans were collected both at Cadre's lab and in three San Francisco area crime labs (including San Francisco and Oakland).

(Set 3) Glock Set: The third set includes 328 test fires from 164 Glock 9mm Luger caliber firerarms (all 9mm Luger models are represented: G17, G19, G26, and G34). There are two test fires from each firearm. These casings were obtained from Dr. Jim Hamby (International Forensic Science Laboratory & Training Centre, Indianapolis). Identification of Glock casings typically relies on matching their aperture shears. Therefore, this set serves as an excellent test of the aperture shear extraction and matching code. The set includes test fire ammunition from Blazer, CBC, CCI, FC, Geco, PROOF, RTAC, RWS, Speer, and WCC (brass, aluminum, or steel casings with brass or nickel primers). In most cases the two test fires were not collected on the same type of ammunition (e.g., one member of the pair may have used CBC ammunition while the corresponding known-match may have used CCI ammunition). Many casings have a lacquer present and many have an alphanumeric primer stamp. Primer stamps were not included in the masked regions.

(Set 4) Miami-Dade Study Set: The fourth test set includes the Miami-Dade Study (Test Set 8) test fires (provided by Dr. Thomas Fadul). These casings are included to demonstrate the performance of the system on a set of well marked casings (in contrast to the other real-world test sets). The Miami-Dade set consists of ten pairs of matched knowns and fifteen individual 'questioned' (or unknown) casings. The examiner is tasked with matching each questioned case with one of the known pairs. All test fires use Federal Cartridge ammunition (brass casing, nickel primer). In contrast to the first three test sets, the cases in the Miami-Dade Study are all strongly marking.

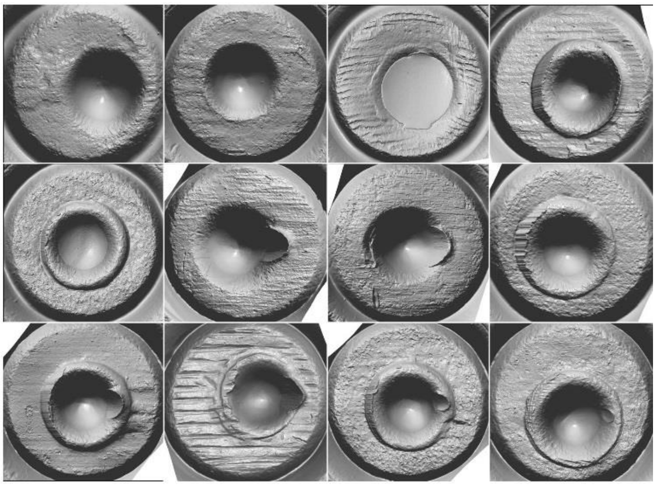

Fig.8 Core12 Set 6. 3D renderings of TopMatch scans from all 12 firearms in the Core12 set (1-12, left-toright top-to-bottom). The firearms produce casings with filing, granular, and milled marks. The set includes both well marked and poorly marked casings. A few of the casings have usable aperture shears. The firearms include: Fabrique Nationale, Hi-Point, Norinco, Radom, Ruger, and S&W.

(Set 5) JFS Set: The fifth and sixth sets were used in cross-modality matching experiments. Set 5 includes 50 casings, five from each of ten consecutively manufactured pistol slides (Ruger P-series 9mm Luger) as collected and described in previous work [3] (Winchester ammunition, 147 grain bullet). All casings are strongly marked. These casings were scanned using both the TopMatch system (at 1.4μm/px) and a Nipkow disk confocal microscope located at the NIST (at 3.1μm/px). The confocal scan data was converted to X3P format at NIST and transferred to Cadre for analysis. The confocal scans were processed as described above. Set 5 includes 50 confocal scans and 50 TopMatch scans.

(Set 6) Core12 Set: The sixth set is a selection of two test fires from each of twelve firearms from the Forty Seven Firearm set. The twelve firearms were selected to be representative of a range of toolmarks. They include test fires from: (Fig 8). They contain well and poorly marked casings, granular marks, filing marks, and milled marks. Some casings have aperture shear. Our collaborator Alan Zheng at NIST collected confocal scans of these 24 casings using their Nipkow disk confocal system at a resolution of 3.1µm/px. The confocal scan data was converted to X3P format at NIST and transferred to Cadre for analysis. The confocal scans were processed as described above. Set 6 includes 24 confocal scans and 24 TopMatch scans.

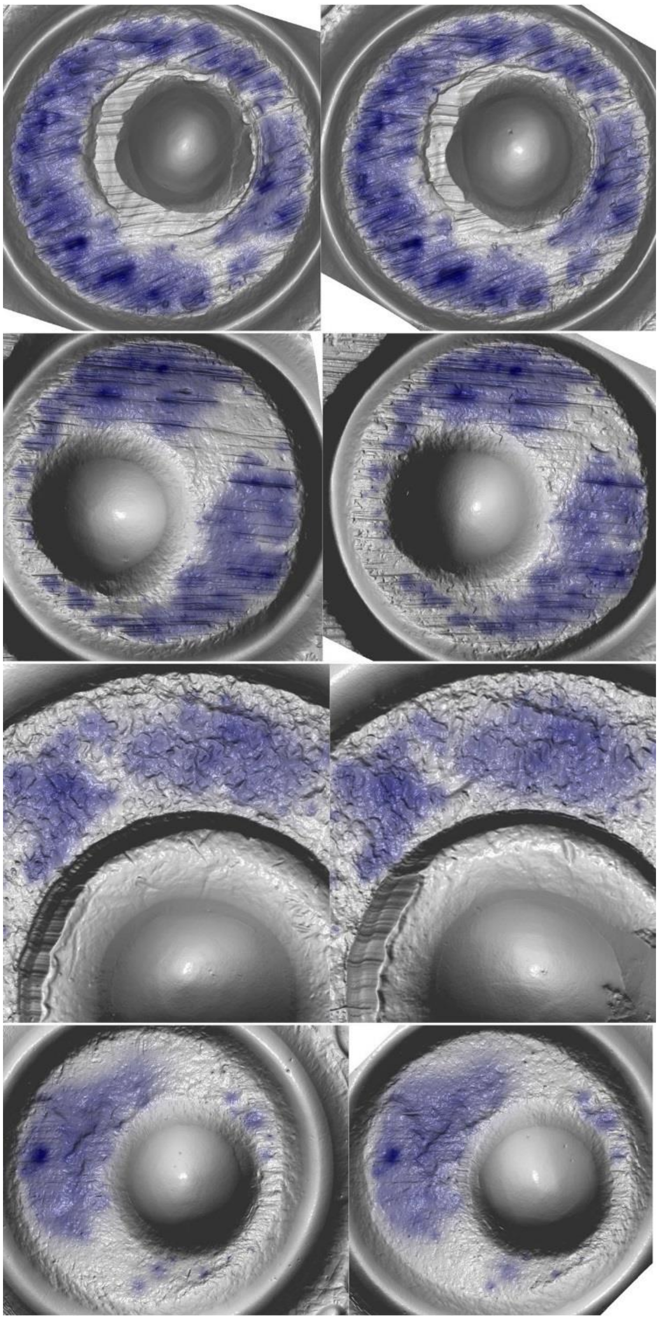

Fig.9 Heatmap. Our feature-based matching algorithm can visually indicate to the user the regions of breechface impression with similar geometric features. This heatmap provides interpretability to the examiner which complements the numerical match score. Darker shading appears where there is a higher density of matched features. Regions with light or no shading have less conserved structure. The top three rows show well marked casings. Note the region in the second row that is not colored and which has different surface geometry. The bottom row shows a weakly marked casing where the regions of similarity are smaller.

III DATA RESULTS AND ANALYSIS

All-vs-all comparisons were performed for each test set. In an all-vs-all comparison, each casing is compared to every other casing (except itself). Each pair of casings is either a Known Match (KM) if both casings were both fired through the same firearm or a Known Non-Match (KNM) if they were fired through different firearms. A False Positive occurs if a KNM is labeled as a match. All comparisons were performed on a multi-core high-end desktop workstation. The results for several score thresholds are shown in Tables 1- 3. The Table captions provide some interpretation of the match results. We can draw several conclusions. First, the breech-face impression plus aperture shear scoring function is able to match significantly more casings than the breech-face impression alone (see Set 3 performance). Second, performance on well marked casings (Sets 4 and 5) are all 100% with no false positives across all thresholds. Third, the cross-modality experiments show that the method is able to correctly match surface topographies from different scanning modalities. It is important that scans be properly preprocessed to remove drop-outs and high-frequency noise. The algorithm performed equally well matching GelSight scans as it did Confocal scans. On the well marked Set 5, perfect accuracy was achieved in matching each GelSight scan with its sister Confocal scan[2] Fourth, the

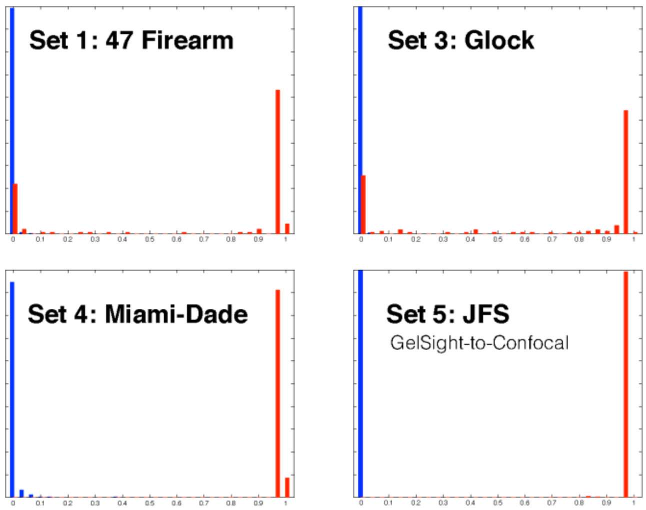

Fig.10 Score Histograms. The histograms of scores are shown for Sets 1, 3, 4, and 5. Known matches are shown in Red. Known Non-Matches are shown in Blue. The 0-1 confidence score shows excellent separation between KM and KNM.

[2] Note: While both casings in the pair were scanned on both imaging modalities, Table 3 does NOT include the comparison of a GelSight scan of a casing to the Confocal scan of the same identical casing. We did perform that experiment and for all cases a GelSight scan and a Confocal scan of the same identical casing achieved a match score near 1.

performance on real-world casings (Sets 1 and 2) is very strong. We are able to match approximately 80% of casings without identifying a false positive. Fifth, we see that Glock casings greatly benefit from the inclusion of aperture shear comparison in the scoring function. Finally, we note that for all datasets the scores for KM and KNM are strongly separated (Fig 10).

Glock Matching: The Glock dataset performance is significantly stronger when the scoring function includes the aperture shear (Table 1 vs 2). Manual examination of the extracted profiles shows that about 88% of the casings have well resolved linear aperture shear profiles (i.e. about 1 in 9 Glocks did not have a strongly resolved aperture shear). This suggests that the aperture shear matching would be successful on about 0.882 or 77% of the casings. We are able to match about 69% of the casings without a false positive which means there is still be room for improvement.

Well-Marked vs Real-World: The well marked casings (Sets 4 and 5) score extremely well and are significantly 'easier' for our algorithm to correctly identify. When comparing the performance of different matching algorithms it is vital to consider the casings used in the study. If the casings are well marked it is much easier for a simple algorithm to correctly match them. Algorithm evaluation must include a set of real-world casings. Via informal survey, firearms examiners estimate that about 70-80% of casings can be identified using only the breech-face impression. While we are glad to achieve 100% accuracy on the well marked casings, we are proud to achieve 80% on Set 1 with no false positives; however, we believe there is still room for improvement.

Scan Quality and Ammunition Type: Match accuracy on Set 2 is about 10% lower than on Set 1. This is likely due to two reasons. First, many of the scans in Set 2 were acquired early in the project with older, lower quality gel. This resulted in a noisier scan. Second, Set 2 contained at least 7 types of ammunition. It is well known that matching across ammunition types is more difficult and that some ammunition types mark poorly.

Table 1 Breech-Face Impression and Aperture Shear Results. The matching performance results for the first four datasets using both the breech-face impression and aperture shear comparison. Numerical classification results for four different thresholds (0.4, 0.7, 0.8, and 0.9). The performance on Set 3 is significantly better than when the shears are not included (Table 2). No false positives are seen in any dataset with a threshold of 0.6 or higher. There is a single false positive at a threshold of 0.4 in Set 1 and Set 2 are the same pair of casings. We are looking into why it scored 0.46. Note that Set 1 and Set 2 should in theory perform very similarly as they both contain a range of real-world test fires. A few issues make Set 2 more difficult. First, many of the scans were collected with older gel, resulting in a more dusty and noisy measurement. Second, Set 2 includes at least 7 ammunition types compared to Set 1 which contains only a single type.

Score Threshold	Casings with a Match	Num (%) Casings w/Top Score > Thresh and Correct	% Pairs > Thresh that are Correct	Known Non-Matches	% (Num) KNM > Thresh
Set 1: 47-Firearm Set					
0.4	141	117 (83%)	99.0%	9,729	0.001% (1)
0.6	141	116 (82%)	100%	9,729	0% (0)
0.8	141	116 (81%)	100%	9,729	0% (0)
0.9	141	113 (80%)	100%	9,729	0% (0)
Set 2: 101-Firearm Set					
0.4	311	223 (72%)	99.6%	56,108	0.002% (2)
0.6	311	213 (68%)	100%	56,108	0% (0)
0.8	311	211 (68%)	100%	56,108	0% (0)
0.9	311	206 (66%)	100%	56,108	0% (0)
Set 3: Glock Set					
0.4	328	226 (69%)	97.4%	53,464	0.006% (3)
0.6	328	216 (66%)	100%	53,464	0.004% (2)
0.8	328	206 (63%)	100%	53,464	0% (0)
0.9	328	196 (60%)	100%	53,464	0% (0)
Set 4: Miami-Dade Study Set					
0.4	35	35 (100%)	100%	549	0% (0)
0.6	35	35 (100%)	100%	549	0% (0)
0.8	35	35 (100%)	100%	549	0% (0)
0.9	35	35 (100%)	100%	549	0% (0)

Table 2 Breech-Face Impression Only Results. The matching performance results for the first four datasets using only breech-face impression comparison. Numerical classification results for four different thresholds (0.4, 0.7, 0.8, and 0.9). For Sets 1, 2, and 4 the performance is very similar to that when the aperture shear is included (Table 1) because these sets only contain a few casings with aperture shears (e.g., some Rugers and Norincos). The performance on Set 3 is lower here than when the shear is included (Table 1). The performance on the Miami-Dade set is perfect.

Score Threshold	Casings with a Match	Num (%) Casings w/Top Score > Thresh and Correct	% Pairs > Thresh that are Correct	Known Non-Matches	% (Num) KNM > Thresh
Set 1: 47-Firearm Set					
0.4	141	117 (83%)	100%	9,729	0% (0)
0.6	141	116 (82%)	100%	9,729	0% (0)
0.8	141	114 (81%)	100%	9,729	0% (0)
0.9	141	111 (79%)	100%	9,729	0% (0)
Set 2: 101-Firearm Set					
0.4	311	223 (72%)	100%	56,108	0% (0)
0.6	311	213 (68%)	100%	56,108	0% (0)
0.8	311	208 (67%)	100%	56,108	0% (0)
0.9	311	204 (66%)	100%	56,108	0% (0)
Set 3: Glock Set					
0.4	328	178 (54%)	100%	53,464	0% (0)
0.6	328	170 (52%)	100%	53,464	0% (0)
0.8	328	164 (50%)	100%	53,464	0% (0)
0.9	328	146 (45%)	100%	53,464	0% (0)
Set 4: Miami-Dade Study Set					
0.4	35	35 (100%)	100%	549	0% (0)
0.6	35	35 (100%)	100%	549	0% (0)
0.8	35	35 (100%)	100%	549	0% (0)
0.9	35	35 (100%)	100%	549	0% (0)

Score Thresholds: We selected thresholds of 0.4, 0.6, 0.8, and 0.9. Because our scoring function does a good job separating the KM and KNM, most tests identified no false positives even at a conservative threshold of 0.4. It may be possible to lower the threshold to pick up additional true positives.

V SUMMARY

We successfully completed the proposed aims during the project period. We developed a robust algorithm for extracting the linear profile of aperture shears (Aim 1). This method is able to extract profiles from curved, flat, or arc'd shears. Manual examination of the extracted profiles shows that we can extract informative profiles for approximately 88% of Glock casings. These linear profiles can be matched as part of our matching algorithm (Aim 2). The matching results demonstrate a significant improvement in Glock matching ability when the shears are considered. We created an open file format (X3P) for the free exchange of 3D surface topography data. This format allowed collaboration with our colleagues at NIST. We demonstrated that cross-modality matching is possible and that in many cases it works extremely well (Aim 3). The cross-modality experiments demonstrate that the TopMatch scoring algorithm is agnostic to 3D imaging modality. To achieve these results the confocal scans required simple preprocessing (mainly interpolation of drop-outs and denoising with a low-pass filter). The system is able to accurately identify known matches when scans were acquired with GelSight or Confocal scanning systems. The algorithm is able to identify known matches where one scan is a GelSight scan and the other is a Confocal scan. We are now trying to better understand the situations that lead to strong cross-modality matching. Crossmodality matching allows labs with different technology to share 3D data for identification. In future, not all labs will have the same 3D scanning system; however, by using an open file format (such as X3P) any two labs can freely exchange data.

Table 3 Cross-Modality Match Results. Our matching algorithm is able to match Set 5 casings from either GelSight or Confocal scans with 100% accuracy. It also achieves 100% accuracy when one scan was collected with GelSight and the sister casing was collected with Confocal. There are no false positives. The Set 6 casings are less well marked. The GelSight scans are correctly matched for 83% of the casings, the Confocal scans are correctly matched for 67% of the casings. There are no false positives. The current confidence score struggled a little when one scan was GelSight and the sister casing was Confocal. Here, the scores were lower. At a threshold of 0.2, 50% of the casing pairs are correctly identified with 1 (of 252) false positive. Note that 0.2 is a low threshold (thus the understandable 1 false postive). We are investigating why these cross-modality scores are lower than for Set 5. Because the same-modality matching worked so well, we are optimistic that we will be able to improve this performance.

Score Threshold	Casings with a Match	Num (%) Casings w/Top Score > Thresh and Correct	% Pairs > Thresh that are Correct	Known Non-Matches	% (Num) KNM > Thresh
Set 5: JFS Set. GelSight-to-GelSight Comparison					
0.4	50	50 (100%)	100%	1,125	0% (0)
0.6	50	50 (100%)	100%	1,125	0% (0)
0.8	50	50 (100%)	100%	1,125	0% (0)
0.9	50	50 (100%)	100%	1,125	0% (0)
Set 5: JFS Set. Confocal-to-Confocal Comparison					
0.4	50	50 (100%)	100%	1,125	0% (0)
0.6	50	50 (100%)	100%	1,125	0% (0)
0.8	50	50 (100%)	100%	1,125	0% (0)
0.9	50	50 (100%)	100%	1,125	0% (0)
Set 5: JFS Set. GelSight-to-Confocal Comparison					
0.4	100	100 (100%)	100%	4,850	0% (0)
0.6	100	100 (100%)	100%	4,850	0% (0)
0.8	100	100 (100%)	100%	4,850	0% (0)
0.9	100	100 (100%)	100%	4,850	0% (0)
Set 6: Core12 Set. GelSight-to-GelSight Comparison					
0.4	24	20 (83%)	100%	264	0% (0)
0.6	24	20 (83%)	100%	264	0% (0)
0.8	24	20 (83%)	100%	264	0% (0)
0.9	24	20 (83%)	100%	264	0% (0)
Set 6: Core12 Set. Confocal-to-Confocal Comparison					
0.4	24	16 (67%)	100%	264	0% (0)
0.6	24	16 (67%)	100%	264	0% (0)
0.8	24	16 (67%)	100%	264	0% (0)
0.9	24	16 (67%)	100%	264	0% (0)
Set 6: Core12 Set. GelSight-to-Confocal Comparison					
0.2	48	24 (50%)	96%	252	0.4% (1)

In 2015 we will focus on establishing best practices and statistical performance for the system. We will conduct an inter-operator variability study. We will also continue to improve the matching algorithm and shear extraction algorithm. The results summarized in Tables 1-3 are an excellent start; however, there is still room for improvement. We know we can further improve the shear extraction and matching algorithm. This year we also improved the scan acquisition protocol and well as the gel formulation. These advances improve the quality of the scan. We would like to go back and rescan all casings with our new process and gel. We are also interested in getting our system in the hands of more examiners.

APPENDIX

Additional Accomplishments:

In addition to the results described above we were able to accomplish the following during the project period: First, we improved the quality of the painted layer of the gel which reduces the measurement noise of captured scans. An improved gel was used through the second half of the project period. A further improved gel was developed at the end of the year and was not used with any of the datasets above. Second, our collaborator Chris

Coleman (Contra Costa County) and his lab was able to scan approximately 500 casings that are part of the California Highway Patrol 40 caliber S&W test set. This set contains paired test fires and will be another great resource. The scanning of these casings is almost complete and we anticipate being able to utilize it in 2015. Finally, we developed a stand-alone casing viewer for X3P files that will be available for free from our website (Spring 2015). This viewer will allow examiners to view 3D scans acquired using either our system or any commercial system that supports the 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. 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 (i.e., upcoming AAFS and AFTE meetings). Most recently, two team members, Lilien andWeller were selected to the Firearms subcommittee of NIST's new OSAC initiative. Weller and Lilien will help create 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.

REFERENCE

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, A. Zheng, R. Thompson, and F. Tulleners. Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides. J. Forensic Science, 57:912–7, 2012.

POSTSCRIPT

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Award Number: 2013-R2-CX-K005

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this federally funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

PD/PI Name, Title, Contact Info: Ryan Lilien, Chief Scientific Officer, Cadre Research Labs. DUNS: 034448780 EIN: 27-2528667 Recipient Organization: Cadre Research Labs, LLC (small business). Recipient Identifying Number (if any): N/A. Project/Grant Period: Start: 1/1/2014, End: 12/31/2014. Reporting Period End Date: 12/31/2014. Report Term or Frequency: Twice a year.