Document Title: Applied Research and Development of a Three-Dimensional Topography System for Firearm Identification using GelSight

Author(s): Ryan Lilien

Document No.: 248639

Date Received: February 2015

Award Number: 2012-DN-BX-K058

Final Technical Report - Cover Page

Federal Agency and Organization Element: Department of Justice, Office of Justice Programs
Federal Grant or Other Identifying Number: 2012-DN-BX-K058
Project Title: Applied Research and Development of a Three-Dimensional Topography System for Firearm Identification using GelSight
PD/PI Name, Title, Contact Info: Ryan Lilien, Chief Scientific Officer, Cadre Research Labs; 68 Robbins St; Acton, MA 01720
Name of Submitting Official: Ryan Lilien
Submission Date: Jan 31, 2014
DUNS: 034448780 **EIN:** 27-2528667
Recipient Organization: Cadre Research Labs, LLC (small business)
Recipient Identifying Number (if any): N/A
Project/Grant Period: Start: 1/1/2013, End: 12/31/2013
Signature of Submitting Official:

1 Abstract

The proposed work is to investigate and develop a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge casings and to demonstrate the system's potential for increasing the quality and reducing the cost of forensic analyses. Despite the importance of tool mark analysis in the forensic sciences, the imaging and comparison of tool marks remains a difficult and time consuming endeavor. Cartridge case comparison is based on the observation that microscopic firearm imperfections (such as those on a breech-face) can be transferred to a fired cartridge case. Therefore, two cases with highly similar breech-face impressions are likely to have been fired through the same firearm. The ability to certify two cases as highly similar is therefore a function of both the ability to capture a high-resolution measurement of each case and the ability to identify and match relevant structural features.

The next generation of methods for the forensic analysis of tool marks should improve accuracy, reduce acquisition and operational costs, and shorten analysis time. Our platform utilizes the recently developed GelSight surface topography imaging system and custom feature-based image comparison software. Compared to competing technologies, our 3D GelSight-based system is fast, inexpensive, and not sensitive to the optical properties of the material being measured. In collaboration with forensics experts, we proposed to improve our hardware and software, conduct several moderate scale experimental benchmarks, and deploy system prototypes. These are critical steps towards the development of a low-cost, fast, and accurate next generation system for cartridge case comparison and database search.

Through this award we completed several important goals. We completed the research and design process and advanced our prototype scanner and casing analysis algorithms to the level where they can now be used in larger research studies and can be deployed to collaborating labs. We demonstrated that our GelSight-based imaging system is able to capture the three-dimensional surface topographies of cartridge cases at high resolution. We tested our imaging and analysis system using several experimental datasets. The results demonstrate the system's current performance on real-world casings (including well and poorly marked casings). We achieve excellent performance on good marking casings and surprisingly good performance on extremely challenging casings, casings that a firearms expert claims would be difficult and time consuming to match. Most importantly, there are no false-positives across approximately 200,000 comparisons. We also evaluated our system using a test set of clean, well marked test fires. We achieve even better performance with these clean casings.

In summary, we have redesigned the base scanner, designed and machined a fully functional cartridge case mount, achieved excellent performance across three experimental datasets, created software which allows 3D visualization of cartridge casings and database search, deployed our hardware and software system to several forensics labs, and we assembled a large experimental set of test fires. These results all strongly support the goals of the proposed study. We have demonstrated a novel technique for the imaging and analysis of cartridge casings and we have shown excellent performance compared to the state-of-the-art alternatives.

Table of Contents for Draft Final Technical Report

Main Body

2 Executive Draft Summary

Synopsis of Problem and Purpose: Despite the importance of tool mark analysis in the forensic sciences, the imaging and comparison of tool marks remains a difficult and time consuming endeavor. Cartridge case comparison is based on the observation that microscopic firearm imperfections (such as those on a breech-face) can be transferred to a fired cartridge case. Therefore, two casings with highly similar breech-face impressions are likely to have been fired through the same firearm. The ability to certify two casings as highly similar is therefore a function of both the ability to capture a high-resolution measurement of each casing and the ability to identify and match relevant structural features.

The next generation of methods for the forensic analysis of tool marks should improve accuracy, reduce acquisition and operational costs, and shorten analysis time. During this award, we investigated and developed a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge casings and demonstrated the system's potential for increasing the quality and reducing the cost of forensic analyses. Our platform utilizes the recently developed GelSight surface topography imaging system and custom feature-based image comparison software. Compared to competing technologies, our 3D GelSight-based system is fast, inexpensive, and not sensitive to the optical properties of the material being measured.

Project Design: In collaboration with forensics experts, we proposed to improve our hardware and software, conduct several moderate scale experimental benchmarks, and deploy system prototypes. These are critical steps towards the development of a low-cost, fast, and accurate next generation system for cartridge case comparison and database search. We aimed to utilize large experimental test sets and in-the-field testing to evaluate the performance of our system in a real-world environment.

Previous Work: The human examiner's ability to compare large number of cases was augmented approximately 15 years ago with the introduction of the 2D IBIS system from Forensic Technology (Quebec, Canada). This system combined a traditional 2D light microscopy image with software for image comparison and database search. The imaging component of these systems is a low powered light microscope. A digital camera is attached to the microscope optics and images can easily be acquired. When the system is used in a database search, an image of the query object is compared to a stored library of previously collected images. Hits are ranked by match score and presented in a rank list. A forensics expert sequentially considers each match and when possible, may take both pieces of evidence to a light microscope for manual comparison and confirmation of a 'hit'. Unfortunately, current 2D systems suffer from several disadvantages, including the unwanted effects of lighting variation, which often result in low match accuracy.

The specific details of the automated image matching algorithms behind these systems are proprietary and are not typically disclosed. It is believed that several methods are employed. Images are typically calibrated to a known standard. The image may be down-sampled, reducing the number of pixels in the original image by a factor of 4 or 16. The image intensities are often normalized. Outlier and drop-out pixels are identified and removed. The images are filtered (possibly in Fourier space, by removing high and low frequency terms). Finally, the region of interest (*i.e.,* the breech-face impression) is masked. The resulting image contains a normalized and cleaned region of interest that can now be compared to a second image.

One common method of image comparison is computing a cross-correlation between the corresponding pixel values in the two images. The underlying assumption is that once two images are normalized, they should contain the same structural features and thus similar pixel values. This method suffers from a few shortcomings. First and most importantly, non-informative features can adversely affect the match score. The entire breech-face is noisy and contains both informative and non-informative structural features. That is, while some of the structural features are similar between the two images, other features are not; because the cross-correlation typically considers the entire masked image, both informative and non-informative regions are compared and the quality of the match can be negatively affected. Second, the multiple pre-processing steps have the potential to eliminate relevant information in each image. Finally, imaging artifacts and shadows can adversely affect the match score. These problems are generally less severe in 3D systems.

Over the past few years, researchers have started to explore a second generation of technology for tool mark imaging. These techniques produce 3D images of tool marks. Several technologies have been considered, including: focus-variation microscopy, confocal microscopy, point laser profilometry, and scanning interferometry. Of these, confocal microscopy and focus-variation microscopy were recently identified as the most promising despite the limitations of dealing with steep slopes (confocal), artifacts of surface reflectivity (confocal), acquisition speed (focus-variation), and cost (confocal).

Our Scanning Technology: Our system utilizes advanced three-dimensional imaging algorithms (*e.g.*, shape from shading and photometric stereo) and the retrographic sensor of Johnson and Adelson to measure an object's three dimensional surface topography. The retrographic sensor is a block of clear elastomer with a thin layer of elastic paint on one side. When an object is pressed into the elastomer, the layer of paint conforms to the shape of the surface. The layer of paint removes the influence of the optical properties of the surface on shape estimation. Every material, such as glass, metal, plastic, or human skin, appears the same when pressed into the retrographic sensor. In contrast to confocal microscopy and focus-variation microscopy, this important feature of our system removes the influence of surface reflectivity on the measured topography. Recent work has improved this basic system to enable measurement of micron-scale geometry (sub-micron depth resolution and lateral (xy) resolution of less than $1\mu m$/pixel). Although firearms examiners typically consider tool marks ranging from tens to hundreds of microns in diameter, they do occasionally consider features as small as 10-15μm. The resolution of our system is capable of capturing and representing these small marks. In its current form, image acquisition requires 1 minute and three-dimensional surface reconstruction requires an additional 1.25 minutes.

In our existing prototype, cartridge case analysis is performed using modern computer vision feature-based matching techniques. Automatically identified distinctive features (corresponding to parts of traditional tool marks) are used to match and align two casings. The detected features correspond to peaks, ridges, basins, and valleys on the surface topography of the breach face impression. These features are detected automatically at multiple scales based on their distinctiveness and magnitude by looking at the local gradient structure of the surface representing the breach face impression. The score of the match is a function of the number and quality of matched features. 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 image identified as informative (*i.e.,* the features).

Results: Through the duration of this award we achieved several important goals. We developed a base scanner and custom cartridge case mount, both of which incorporate feedback from our forensic collaborators. We developed software which allows 3D visualization of cartridge casings as well as database search. The hardware and software has been advanced to the point where it can be used in larger research studies and can be deployed to collaborating labs. We deployed our hardware to four collaborating forensics labs and are assembling a large experimental test set of test fires.

We designed and machined a fully functional prototype cartridge case mount. The casing mount consists of a removable holder on a sliding drawer mechanism. The holder can secure a range of cartridge calibers (from 22 Short to 45 Auto) using three radially positioned rubber tipped nylon fingers (larger calibers will fit in an exchangeable second holder). Several adjustment levers allow the operator to properly position the cartridge for scanning. Once the mount is set for a particular caliber, scanning is extremely simple. The operator uses a single lever to open the holder's fingers, the casing is inserted, the holder is slid under the light-plate, a second lever is used to raise the holder thereby pushing the casing firmly into the gel. This process takes less than 5 seconds. We also designed a simple back enclosure. The enclosure provides increased physical stability while protecting both the z-axis camera positioning motor and all equipment cables.

We implemented a faster and stand-alone version of our core feature identification and casing matching algorithms. There are two parts to the computation 1) Image capture, topography computation, and feature detection and 2) Casing-to-casing comparison matching. Note that step (1) only needs to happen once for each case (*i.e.*, at the time of scan acquisition). Once step (1) is complete, the casing matching of step (2) can be performed as many times as the user chooses. Speed improvements have been realized for step (1): image capture time is approximately 30 seconds; surface topography computation has been reduced from 120 seconds to 75 seconds; and feature detection has been reduced from 12 to 8 seconds. Therefore the entire scan acquisition and 3D surface computation for a single case requires approximately 2 minutes. Larger performance gains have been realized in the more important step (2): single-thread case-to-case comparison has been reduced from 400 seconds to 2.5 seconds (this represents a greater than 150x speed improvement over the original method). By employing multiple processing threads we have further reduced a complete full resolution ($1.4\mu m$/pixel) case-to-case comparison to 0.21 seconds ($\sim$1800x faster than our initial implementation). The realized improvements were obtained via simple algorithmic implementation optimizations and not through major reworking. There is therefore more opportunity to further reduce the runtime.

We designed an auto-masking algorithm which generates an initial estimate of the location of the breech-face impression on the primer. During the import of the 3D scan, the user is asked to click (select) two points on the breech-face impression. The user simply clicks the mouse twice and the algorithm extends the mask to include the entire breech-face impression. The masked region appears in color, as a painted surface, in the 3D viewer. The user can fine tune the initial mask with the mouse.

We completed design and implementation of our primary 3D casing viewer. This software allows the user to view and interact with the 3D surface topography as rendered using the computer's graphics card. The user can zoom, rotate, and pan the rendered surface. Because the model is a true 3D rendering, we allow the user to change the position of the light source. This functionality is important for visualization as the direction of the light significantly affects the features visualized. It is important to note that the lighting is only for the user and is **not** used by our case matching algorithm. Our algorithm matches the surface topography independent of the light source. We have found that firearms examiners very much like the ability to position the light as it feels very natural to them. We completed a prototype version of our primary user interface. This program provides the firearms examiner a simple graphical user interface through which they can collect 3D scans, automatically mask the region of the breech-face impression, enter casing, firearm, and incident data, visualize and annotate the 3D scans, conduct searches, and visualize the matches (with accompanying match score). The software has been deployed to our test sites. We continue to improve the features of this software and are issuing regular updates.

We've deployed scanners to several collaborating labs. Our collaborator at the Oakland Police Department, Todd Weller, received our first scanner in April 2013. Since that time, Todd has been scanning casings and providing valuable feedback with respect to hardware and software design. In September 2013 we setup scanners and performed in person hands-on training for approximately 12 firearms examiners at the San Francisco Police Department (Andy Smith) and at a second government laboratory. Both labs are now using our scanners and casing analysis software. As described above, the machines have the prototype software which allows the examiner to collect 3D scans, enter casing, firearm, and incident data, conduct database searches, and visualize the results. We have setup regular communication with these facilities to obtain usability feedback. In mid-November we relocated the San Francisco machine to the Contra Costa Country lab (Chris Coleman). All participating labs have scanned casings that are now part of our test set and have provided excellent usability feedback to our research group.

We demonstrated that our GelSight-based imaging system is able to capture the three-dimensional surface topographies of cartridge cases at high resolution. We tested our imaging and analysis system using several experimental datasets. The first two datasets demonstrate the system's current performance on real-world casings (including well and poorly marked cases). We selected 47 firearms including: 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 without preference to their ability to mark cartridge casings. The intent was to select firearms that would represent real-world conditions. The third dataset consists of the Miami-Dade Test Set and demonstrates our performance on clean, well marked test fires. We achieve excellent performance on good marking cases and surprisingly good performance on extremely challenging casings, casings that a firearms expert claims would be difficult and time consuming to match.

In experiment one, the 282 Known Matches (KM) have significantly larger scores than the 19,458 Known Non-Matches (KNM). Although we do not currently define a score threshold for calling matches, we note that no KNM has a score above 100. In fact, no KNM has a score above 50 (and less than 1% are above 40). Known Non-Matches always have a score below 50. Similar trends hold for experiment two (47 firearms with 3 ammo types) and for our Large Test Set (which includes these 47 firearms as well as over 100 additional casings scanned by our collaborating labs). In the Large Test Set, 81% of casings have a correct match as the highest ranked casing. For the remaining 19% of casings, the algorithm is not wrong. There are no false positives. For these casings, the score of the top ranked casing is extremely low, indicating that the casing was poorly marked and that the algorithm is unable to find a match. We have not observed a false positive match across 193,732 known non-matches. The data for experiment three comes from competency test where all casings are strongly marked. This is reflected in our match score results. All questioned casings are correctly identified with their matching known pairs. All Known Non-Matches have scores below 38 (average: 13.8) and no Known Match has a score less than 75 (average: 266).

We completed a number of reproducibility and persistence studies. These experiments show that repeated scanning of a single casing does not demonstrate an effect on the number of features detected or match score. The persistence studies show that repeated scanning of the same casing does not result in a significant transfer of the casing's features to subsequent scans collected using the same gel.

We have begun to disseminate our research results. We gave an oral presentation at the 2013 AFTE meeting in Albuquerque, NM. This was an excellent event. We were able to share many of the results described in this report. We were also able to speak with and solidify collaborations with several forensics labs. We are presently writing our first research paper summarizing our system, matching algorithm, and case matching experiments. Our abstract for AFTE 2014 has been accepted and we will present the latest results this May.

Implications for 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: it increases the quality and efficiency of forensic analysis, it develops new instrumentation systems, and it provides a novel approach to enhancing the analysis and interpretation of forensic data derived from physical evidence.

The criminal justice system and forensic laboratories are now aware of our work. We are developing analytic techniques, grounded in mathematical science that are able to provide accurate quantitative sample comparison and database search. This benefits the criminal justice system and their ability to present firearm identification and tool mark evidence in the courtroom. It further benefits firearms examiners who can now more efficiently and more confidently search large databases to identify connections between crimes and known firearms. Additional impact will be made when our systems become available to all crime labs sometime in 2014.

Conclusion: We have successfully completed all proposed aims. The results obtained all strongly support the goals of the proposed study. We have developed a novel technique for the imaging and analysis of cartridge casings and we have demonstrated excellent performance compared to the state-of-the-art alternatives. Our results also suggest several next steps for the research. In 2014 under a new NIJ award we will pursue three additional aims 1) Striated Mark Imaging and Analysis. We will develop methods of imaging and comparing striated tool marks (*i.e.*, primer shear and land engraved areas). 2) Development of an Extended Statistical Significance Model. This model will extend the statistical model currently under development for impressed marks and will incorporate both impressed and striated marks. 3) Cross-Modality Matching. We will develop and test algorithms for comparing 3D surface topographies collected using different scanning modalities (*e.g.*, GelSight and Confocal Microscopy). These three aims all build off the successful work completed in 2013.

3 Introduction

3.1 Statement of Problem:

The proposed work is to investigate and develop a novel, accurate, and low-cost system for structural 3D imaging and comparison of cartridge casings and to demonstrate the system's potential for increasing the quality and reducing the cost of forensic analyses. The proposal directly addresses several aims of the NIJ's Applied Research and Development in Forensic Science for Criminal Justice Purposes program: it increases the quality and efficiency of forensic analysis, it develops new instrumentation systems, and it provides a novel approach to enhancing the analysis and interpretation of forensic data derived from physical evidence.

3.2 Statement of Rationale for the Research:

Despite the importance of tool mark analysis in the forensic sciences, the imaging and comparison of tool marks remains a difficult and time consuming endeavor. To the best of our knowledge there does not exist a 3D imaging and analysis platform for firearm forensics that is fast, inexpensive, and whose accuracy has been demonstrated on a large real-world test set. The next generation of methods for the forensic analysis of tool marks should improve accuracy, reduce acquisition and operational costs, and shorten analysis time. Our platform utilizes the recently developed GelSight surface topography imaging system and custom feature-based image comparison software. The proposed system has a per-pixel resolution of 0.8 microns. Compared to competing technologies, our GelSight-based system is fast, inexpensive, and not sensitive to the optical properties of the material being measured. Our system's preliminary results, collected prior to this award, are extremely promising. In collaboration with forensics experts, we proposed to improve our hardware and software, conduct several moderate scale experimental benchmarks, and deploy system prototypes. These are critical steps towards the development of a low-cost, fast, and accurate next generation system for cartridge casing comparison and database search.

3.3 Literature Citations and Review:

Cartridge case comparison is based on the observation that microscopic firearm imperfections (such as those on a breech-face) can be transferred to a fired cartridge case. Therefore, two casings with highly similar breech-face impressions are likely to have been fired through the same firearm. The ability to certify two casings as highly similar is therefore a function of both the ability to capture a high-resolution measurement of each casing and the ability to identify and match relevant structural features. Therefore, the development of analytic techniques, grounded in mathematical science and able to provide accurate quantitative sample comparison would be welcomed by the forensic community.

Our work directly addresses methods for tool mark analysis with a specific focus on firearm forensics. For over 90 years, cartridge case tool marks have been manually compared. The traditional approach utilizes light-microscopy to image the casings and human expertise to identify similarities between objects. While extremely successful, this approach is time consuming and lacks an interpretable quantitative measure of similarity resulting in an increasing number of courtroom admissibility challenges.

Several recent reports, including two from the National Academy of Sciences, have called for additional research [5, 6, 7]. Development of new instrumentation and comparison algorithms can help address the identified challenges.

2D Imaging Systems The human examiner's ability to compare large number of cases was augmented approximately 15 years ago with the introduction of the 2D IBIS system from Forensic Technology (Quebec, Canada). This system combined a traditional 2D light microscopy image with software for image comparison and database search. The imaging component of these systems is a low powered light microscope. A digital camera is attached to the microscope optics and images can easily be acquired. When the system is used in a database search, an image of the query object is compared to a stored library of previously collected images. Hits are ranked by match score and presented in a rank list. A forensics expert sequentially considers each match and when possible, may take both pieces of evidence to a light microscope for manual comparison and confirmation of a 'hit'. Unfortunately, current 2D systems suffer from several disadvantages which often result in low match accuracy [8].

The specific details of the automated image matching algorithms behind these systems are proprietary and are not typically disclosed. It is believed that several methods are employed. Images are typically calibrated to a known standard. The image may be down-sampled, reducing the number of pixels in the original image by a factor of 4 or 16. The image intensities are often normalized. Outlier and drop-out pixels are identified and removed. The images are filtered (possibly in Fourier space, by removing high and low frequency terms). Finally, the region of interest (*e.g.,* the breech-face impression) is masked. The resulting image contains a normalized and cleaned region of interest that can now be compared to a second image.

One common method of image comparison is computing a cross-correlation between the corresponding pixel values in the two images. The underlying assumption is that once two images are normalized, they should contain the same structural features and thus similar pixel values. This method suffers from a few shortcomings. First and most importantly, non-informative features can adversely affect the match score. The entire breech-face is noisy and contains both informative and non-informative structural features. That is, while some of the structural features are similar between the two images, other features are not; because the cross-correlation typically considers the entire masked image, both informative and non-informative regions are compared and the quality of the match can be negatively affected. Second, the multiple pre-processing steps have the potential to eliminate relevant information in each image. Finally, imaging artifacts and shadows can adversely affect the match score [2]. These problems are generally less severe in 3D systems.

3D Imaging Systems Over the past few years, researchers have started to explore a second generation of technology for tool mark imaging. These techniques produce 3D images of tool marks. Several technologies have been considered, including: focus-variation microscopy, confocal microscopy, point laser profilometry, and scanning interferometry. Of these, confocal microscopy and focus-variation microscopy were recently identified as the most promising despite the limitations of dealing with steep slopes (confocal), artifacts of surface reflectivity (confocal), acquisition speed (focus-variation), and cost (confocal)[4]. The researchers [4] note that while only a small number of labs may be able to afford confocal-based 3D imaging machines, there are many laboratories and research facilities that would benefit from a lower-cost three-dimensional imaging solution should one become available. Several 3D systems are commercially available, including the Alias System (Pyramidal Technologies) [3], IBIS BRASSTRAX-3D (Forensic Technology), and Balistika (Scientific and Technical Research Council of Turkey). To the best of our knowledge, none of these systems have wide deployment in the US, nor has a gold-standard technology emerged. As a point of comparison, confocal-based scanning and analysis systems are typically priced at or above $400,000 which make them unrealistic choices for most (if not all) law enforcement facilities.

4 Methods: Technology Overview

4.1 Scan Acquisition Hardware - GelSight

Our system utilizes advanced three-dimensional imaging algorithms (*e.g.*, shape from shading and photometric stereo) and the retrographic sensor of Johnson and Adelson [10] to measure an object's three dimensional surface topography. The retrographic sensor is a block of clear elastomer with a thin layer of elastic paint on one side. When an object is pressed into the elastomer, the layer of paint conforms to the shape of the surface (Fig. 3). The layer of paint removes the influence of the optical properties of the surface on shape estimation. Every material, such as glass, metal, plastic, or human skin, appears the same when pressed into the retrographic sensor. In contrast to confocal microscopy and focus-variation microscopy, this important feature of our system removes the influence of surface reflectivity on the measured topography. Recent work has improved this basic system to enable measurement of micron-scale geometry (sub-micron depth resolution and lateral (xy) resolution of less than $1\mu m$/pixel) [11]. Although firearms examiners typically consider tool marks ranging from tens to hundreds of microns in diameter, they do occasionally consider features as small as 10-15μm(Figure 1). The resolution of our system is capable of capturing and representing these small marks.

The sensor can be made from almost any clear rubber that is both strong and elastic, such as silicone rubber or thermoplastic elastomer (TPE). The reflective paint is typically made by dissolving some of the elastomeric material in a solvent and adding reflective pigments. This process creates a paint with strength and elasticity similar to the body of the sensor. For high-resolution measurements, the pigments are often applied directly to a thin gel coated in an adhesive. This process reduces the thickness of the layer of paint since no binder is used to hold the pigments.

The elastomeric sensor is mounted on a sheet of glass and a camera views the reflective skin through the clear gel (Fig. 3). A set of lights sequentially illuminates the sensor to reveal an initially shaded image. Calibration images of a grid of spheres are used to build a photometric model relating image intensity to surface orientation. Images of the object pressed into the sensor are collected and converted to surface normals using nonlinear least-squares optimization. A surface normal is the vector perpendicular to the object surface at the specified (x, y) location. The surface normals are then integrated to obtain a three-dimensional surface. We will refer to the three-dimensional surface topography obtained by GelSight as an *image* or *heightmap*. The scalar value recorded at each pixel is the surface height of the object at the corresponding location. A representative GelSight scan of a Glock fired cartridge case is shown in

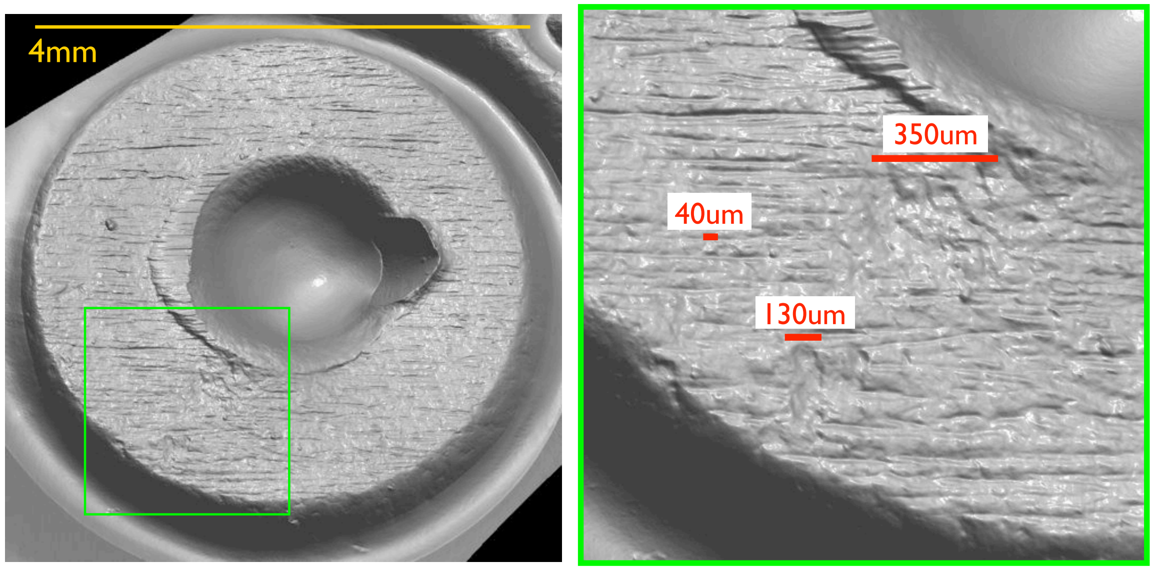
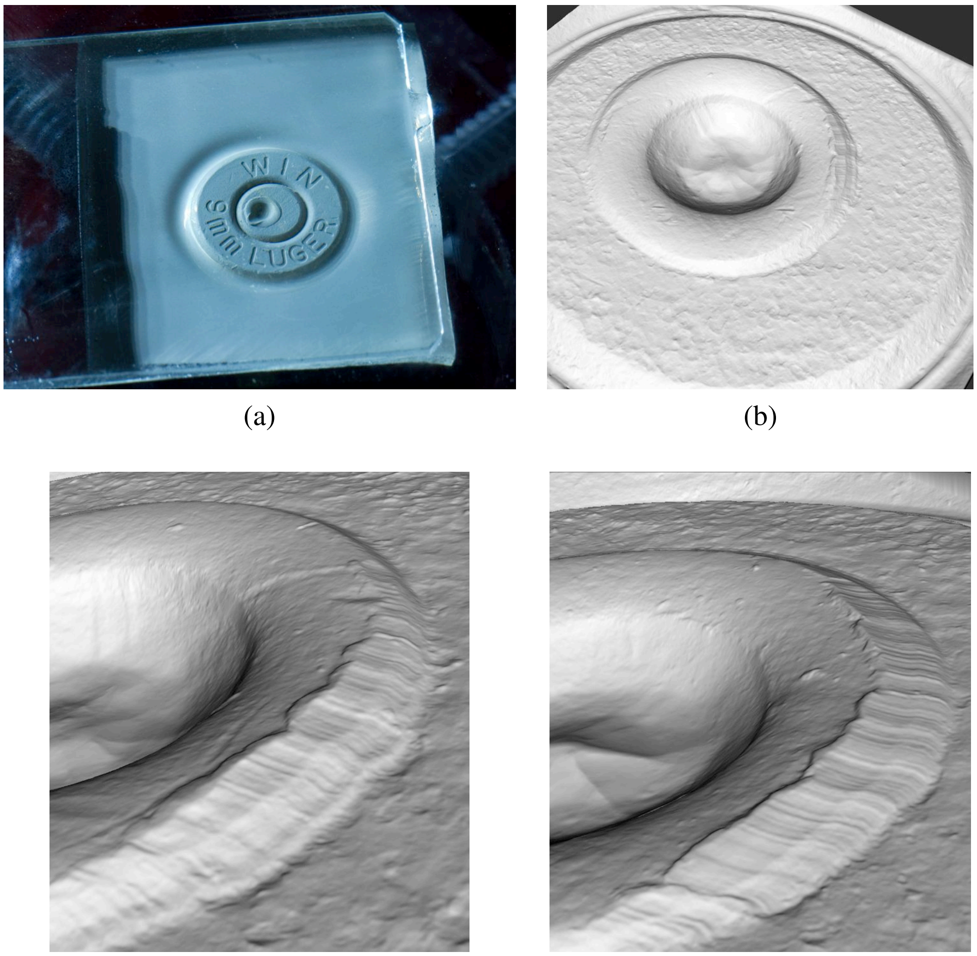


Figure 1: **Feature Sizes.** A 3D scan of the NIST Standard Casing illustrating several distinct features with their approximate sizes. (Left) The NIST standard casing. (Right) a close-up of the boxed region of the left image. Firearms examiners often rely on features $10\mu m$and larger.

Figure 3. A particular strength of our technology is its ability to capture surfaces with significant slope. This provides GelSight an advantage over confocal microscopy whose signal can become unreliable for slopes greater than 15-degrees [13]. Figure 5 shows a rendered heightmap containing a primer shear laying at a slope close to 45-degrees. Improved sensors would not affect the hardware and could easily be swapped into the existing setup. Representative GelSight scans of cartridge case are shown in Figure 4.

In its current form, image acquisition requires less than 1 minute and three-dimensional surface reconstruction requires an additional 1.25 minutes. The current hardware setup is built mostly from off-the-shelf optics and illumination and there are no components that require calibration by an expert (*e.g.,* no lasers or adjustable optical elements). If system components are replaced or upgraded, the simple photometric calibration procedure can be performed by the end user.

Note About Gel Formulation. Along with our partners at GelSight, Inc we spent a lot of time attempting to optimize the gel formulation. We are trying to optimize several different objectives, including the optical clarity of the gel, the firmness of the gel, the tackiness (or stickiness) of the clear side of the gel, and the tackiness of the painted side. Unfortunately, a granularity issue arose about halfway through the year. Many sheets of gel were manufactured with a very rough painted surface. Use of this gel results in a granularity in the scanned surface and a reduction in the scanning resolution. Unfortunately, most of the scans collected during our deployment study were affected by the granularity issue. The issue did not result in any false positives; however, the overall matching performance was a bit lower on the affected casings. We have now resolved the granularity issue and are focusing on adjusting the formulation to minimize the occurrence of air-trapping and wrinkles. Trapped-air and wrinkles occasionally occur during scanning requiring the user to reseat the gel. None of the scans used in the experiments below had trapped air or wrinkles; however, we are working to minimize the chance of their occurrence to make scanning easier for examiners.



(a) (b) (c) (d)

Figure 2: **Sensor and Sample Images.** (a) An image of a 9mm Luger cartridge case breech-face pressed into the elastomeric sensor. The surface topography of the breech-face is captured using our method, based on [10]. (b) The rendering of a captured surface. (c) and (d) Closeups of the surface topography of two different cartridge cases that were fired through the same firearm. The images show both primer shear marks and breech-face impressions. Note that although the proposed work is not yet addressing the automated comparison of primer shear marks, the the two images reliably capture the similarity between the two tool marks. Note that images were downsampled for presentation. In the original images, a single pixel corresponds to 1.7 microns. In the results described below, images are captured at 1.4 microns per pixel. The system has a maximum resolution of 0.8 microns per pixel.

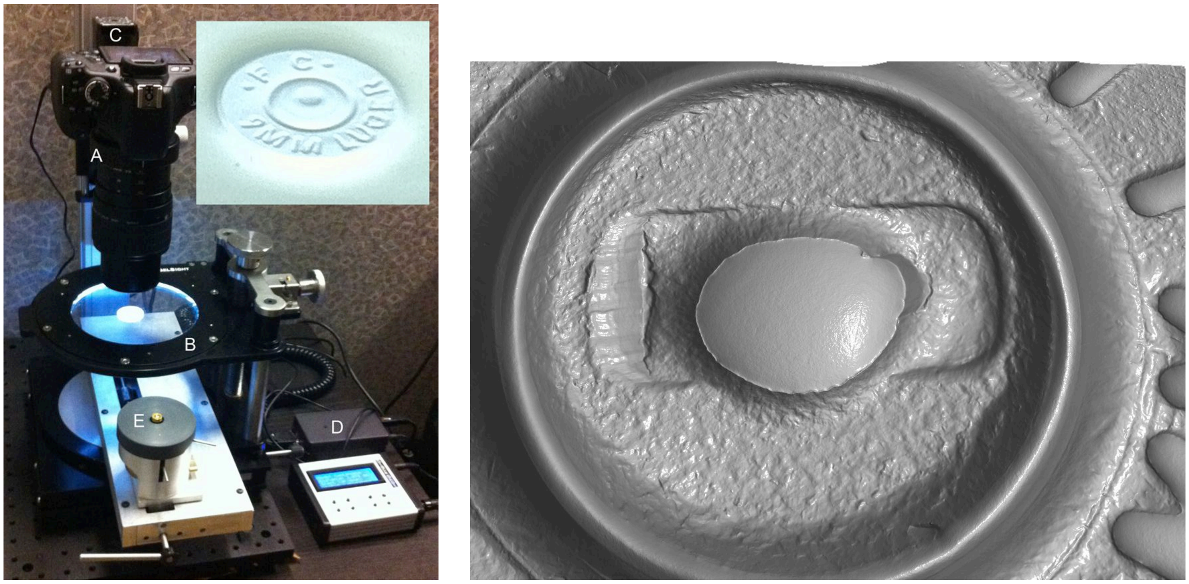


Figure 3: **TopMatch-GS 3D Scanner.** (Left) The current capture setup contains an 18-megapixel Canon EOS Rebel T3i digital camera with a 65mm macro lens (A), GelSight elastomeric sensor (appears as a white disk through glass), glass light-plate with embedded electronics (B), motorized camera control (C), electronics control box (D), cartridge case holder and mount (E). (Inset) View of case as seen through gel. (Right) Surface rendering of the 3D measured surface of a Glock fired casing. The impressed breech-face impression and striated primer shear marks are clearly present. Note that the firing pin impression was not imaged and may thus look unusual to a firearms examiner. Images were downsampled for presentation. In the original images, a single pixel corresponds to 1.4μm.

4.2 Scan Comparison and Matching - Feature-Based Image Matching

The GelSight images have a number of favorable properties that make them amenable to image matching using Computer Vision techniques. Computer Vision is the subarea of Computer Science concerned with the development of algorithms for image matching and analysis. The resolution of the GelSight images (approximately 1.4 microns per pixel) accurately captures the types of tool marks known to be informative to the forensics community.

In our existing prototype, cartridge case analysis is performed using modern computer vision feature-based matching techniques. Automatically identified distinctive features are used to match and align two casings. The score of the match is a function of the number and quality of matched features. 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 image identified as informative (*i.e.,* the features). Features correspond to regions of the image with nonzero gradients in both the x and y dimensions. In other words, we are looking for 'corner-like' regions of the height-map, of any size, which correspond to the same types of ridges, peaks, gouges, and concavities that a trained firearms examiner would identify.

Feature-based matching techniques are ideal for supporting database search and rapid query. For example, once an object is scanned the image features only need to be computed once. After the features have been detected, an efficient hash-based indexing scheme can be used to organize the images and to rapidly identify matches for a query image. Our method supports these extensions. The nature of our feature extraction and matching algorithms suggests that speed improvements will scale well (*i.e.,*

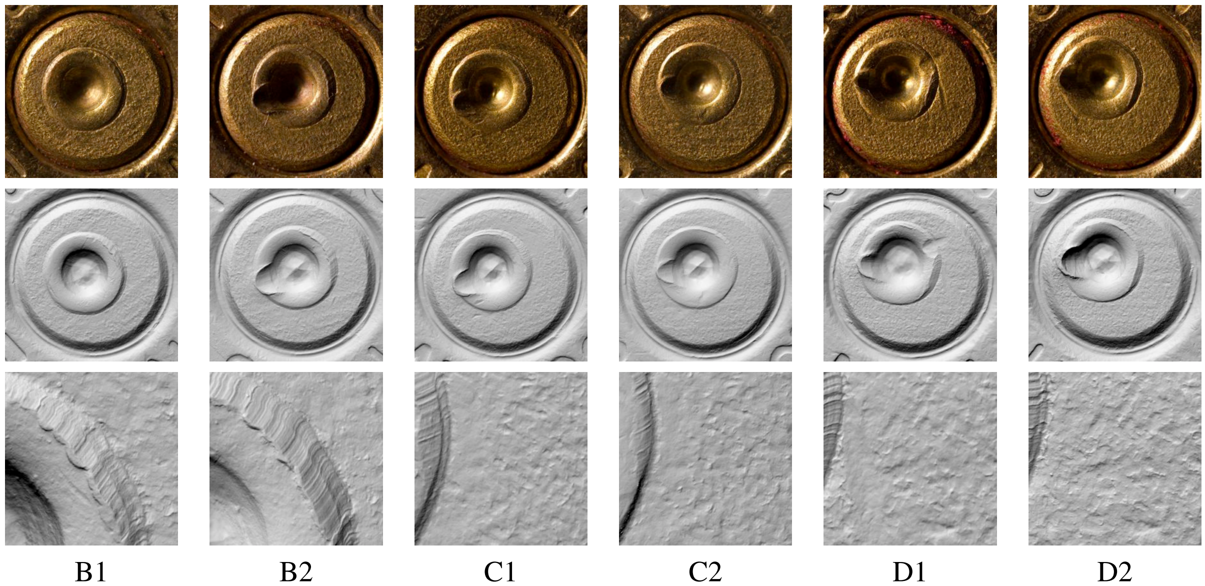


Figure 4: **Case Scans.** Six cartridge cases, two from each of three firearms (denoted B, C, and D). (Top) Traditional light microscope color images of the six cartridge cases. (Middle) Renderings of the measured geometry of each casing. (Bottom) Close-up view from each scan showing the detail that can be measured using our system. Note that these images are not at full resolution. The actual scans capture features beyond 2 micron resolution.

linearly) with multi-thread or GPU-enabled implementations [1, 9, 12]. It is reasonable to expect that these techniques could provide at least a 10-fold improvement in runtime.

The matching of cartridge casings through GelSight images can be thought of as an instance of the general object recognition and matching problem in Computer Vision (*e.g.,* looking for bicycles in travel photos). However, several aspects of the cartridge case matching problem make it easier than the general object matching task. For example, we control the scale of the casing in the image, we know the orientation of the objects up to a single in-plane rotation, and we do not experience significant skewing or distortion. Further, because the GelSight images describe the 3D shape of the tool marks (*i.e.,* the GelSight images are height-maps), it is not necessary to cope with appearance variations (*e.g.,* shadows) that arise due to variations in lighting conditions.

Note on subclass marks. Because our system measures the three dimensional structure of the masked region of the primer, we will capture subclass and system marks (*e.g.*, marks possibly left from several firearms of the same model). These marks may complicate comparison by any method. To date, we have not seen these marks result in false positives. That is, if the subclass and system marks are present, they do not appear significant enough to influence the classification. We note that ten of the Rugers in our forty-seven firearm experiment below had consecutively manufactured breech-faces. We did not observe any false-positives for these firearms. A reviewer commented that it might be worth allowing an examiner to exclude features that should not be used for correlation. Our current system allows the examiner to edit the mask of the included region of the breech-face impression. While this is slightly different from allowing the user to explicitly exclude certain features, it does enable the user to exclude potentially problematic regions of the primer. We note that allowing the operator to include and exclude specific features could introduce some undesired non-objectivity into the matching process.

5 Results

In this section we describe the progress made during the twelve month project period. We begin by reviewing the primary goals and objectives of the proposal and then expand on each aim to discuss what was accomplished. We note that there have been no significant changes in the project goals or direction since the project start. Our one year project has three primary aims.

1. Finalize hardware design and create two units for testing within Cadre and with external labs. One unit will be created at project start and a second unit approximately halfway through the year.

2. Complete Software Prototype. Software to include basic user interface functionality, a simple database interface, the ability to collect 3D scans of cartridge casings, the ability to mask and match these scans.

3. Conduct 50 Firearm Experiment. Test fires will be collected for at least 50 different firearms. This dataset will be used in an all-vs-all matching experiment to help evaluate and validate our matching algorithms.

The following accomplishments and results were obtained.

- In collaboration with GelSight Inc, we redesigned the base scanner (Fig. 3) to be stable and easy to use. The glass light-plate is now embedded in a circular enclosure with embedded electronics and directional LED lights (Fig. 3B). The electronics are enclosed in a black project box (Fig. 3D). A linear xy-stage allows fine positioning control. The setup contains an 18-megapixel Canon digital camera with a 65mm macro lens. The setup supports up to $0.9\mu m$/pixel; a 9mm primer and breech-face impression can be measured using a single image frame (*i.e.*, without stitching multiple images) at approximately $1.4\mu m$/pixel.

- We designed and machined a fully functional prototype cartridge case mount (Fig. 3E). The casing mount consists of a removable holder on a sliding drawer mechanism. The holder can secure a range of cartridge calibers (from 22 Short to 45 Auto) using three radially positioned rubber tipped nylon fingers (larger calibers will fit in an exchangeable second holder). Several adjustment levers allow the operator to properly position the cartridge casings for scanning. Once the mount is set for a particular caliber, scanning is extremely simple. The operator uses a single lever to open the holder's fingers, the casing is inserted, the holder is slid under the light-plate, a second lever is used to raise the holder thereby pushing the casing firmly into the gel. This process takes less than 5 seconds. The process is reversed to remove the case. The current version is actually a revision of the initial (January 2013) design. The initial design was machined 'too perfectly', it raised the case perfectly straight into the gel which caused occasional air-trapping (*i.e.*, small air bubbles trapped between the gel and the concavity on the case where the flow-back meets the breech-face impression). Our redesign uses a robust flexure to position the casing holder at 2-degrees off level. This slight angle minimizes the risk of air-trapping. The flexure is pliable and gives way to a level orientation when the holder meets the resistance of the gel. This results in an overall level casing with a lower risk of air-trapping. All current systems have this flexure incorporated design. We are currently further improving this design to more easily switch between calibers, to accept partially damaged casings, and to accept bullets.

- We designed a simple back enclosure (not seen in Fig. 3). The enclosure provides increased physical stability while protecting both the z-axis camera positioning motor and all equipment cables. We found that unbound cables had the tendency to get in the operator's way; hiding the cables eliminates this problem. These enclosures are installed on all current systems.

- We improved the calibration process by implementing an automatic calibration routine. The operator still needs to calibrate the scanner once per sheet of gel; however, the process is now much easier. Before the auto-calibration, the user was required to place a series of calibration circles on the calibration image. This was a time consuming process. The accuracy with which the user placed the circles influenced the quality of the 3d surface measurement. Our new method automatically places the circles on the calibration image. This reduces the time required for calibration and improves the calibration accuracy resulting in a higher quality scan.
- We implemented a faster and stand-alone version of our core feature identification and casing matching algorithms. There are two parts to the computation 1) Image capture, topography computation, and feature detection and 2) Casing-to-casing comparison matching. Note that step (1) only needs to happen once for each casing (*i.e.*, at the time of scan acquisition). Once step (1) is complete, the casing matching of step (2) can be performed as many times as the user chooses. Speed improvements have been realized for step (1): image capture time is approximately 30 seconds; surface topography computation has been reduced from 120 seconds to 75 seconds; and feature detection has been reduced from 12 to 6 seconds. Therefore the entire scan acquisition and 3D surface computation for a single case requires approximately 2 minutes. Some of this processing now takes place in the background, which allows the user to continue working without waiting for a computation to finish. Larger performance gains have been realized in the more important step (2): single-thread casing-to-casing comparison has been reduced from 400 seconds to 2.3 seconds (this represents a greater than 150x speed improvement over the original method). In the matching results described below we employ multiple processing threads and complete a full resolution ($1.4\mu m$/pixel) case-to-case comparison every 0.21 seconds ($\sim$1800x faster than our initial implementation - not 1800% faster but 1800 times faster). The realized improvements were obtained via algorithmic implementation optimizations and not through major reworking.
- We began exploring the use of GPU computing (the use the the Graphical Processing Unit) to further improve runtime. Computing problems must have a very specific form to take advantage of GPU computing and significant effort and multiple trials are typically required to benefit from this new processing paradigm. Our first attempt at converting part of our code to use the GPU did not provide a runtime advantage. Learning from that experience we now know which parts of our method make heavy use of matrix manipulations and are thus the best candidates for GPU optimization. It is these parts of the method that can best take advantage of the GPU's many processor cores. We also invested some effort in modifying the matching algorithm to better handle slightly warped casings. This new approach will use the same features that have been successful for us, but will better handle the matching process. This new approach is more amenable to the GPU framework as the method relies more heavily on large matrix manipulations. We tested some of these core matrix operations in a GPU framework and observed a several orders of magnitude speedup. This suggests that we will be able to design a highly efficient GPU-based implementation of our next algorithm. We plan to return to this problem in early 2014. All deployed systems have GPU capable cards, most are modest in power, two have more powerful cards.
- We started to explore automatic labeling of casings as being poorly-marked or sufficiently-marked. In our experiments, if there were fewer than 1000 features identified in the masked region of the breech-face impression the casing was considered poorly marked otherwise the casing was considered sufficiently marked. We report performance of these categorizations below.
- We designed an auto-masking algorithm which generates an initial estimate of the location of the breech-face impression on the primer. During the import of the 3D scan, the user is asked to click

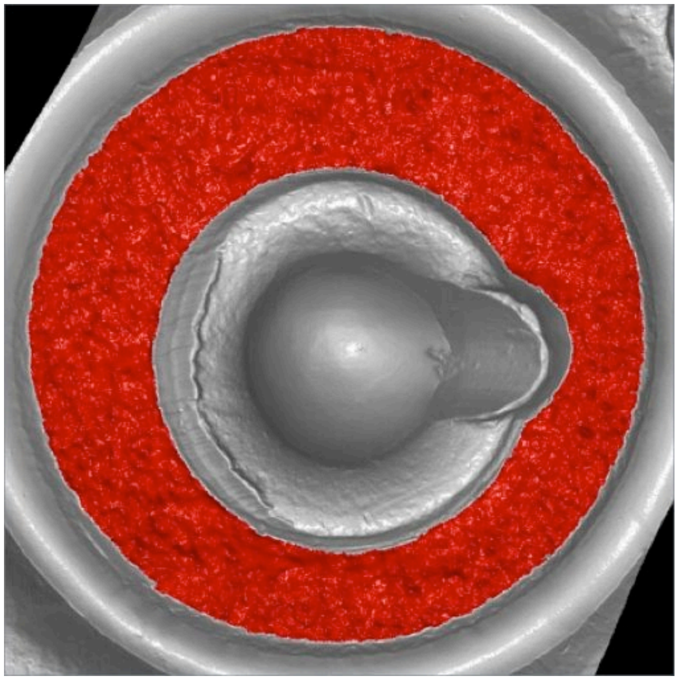


Figure 5: **Breech-Face Impression Mask.** The auto-masked region containing the breech-face impression is shown in red. Note that the firing pin impression, primer shear, and ejector marks are not included in the current matching algorithm. In most cases, the mask identified by the auto-masker can be used directly, on occasion the user will want to manually touch-up the mask around the edges.

(select) two points on the breech-face impression. The user simply clicks the mouse twice and the algorithm extends the mask to include the entire breech-face impression. The masked region appears in color, as a painted surface, in the 3D viewer (Figure 5). The user can fine-tune the initial mask with the mouse. Because the masking takes place in our 3D viewer, the user can zoom and rotate the surface in three-dimensions. The auto-masker seems to work very well and was used in the matching results below. It is incorporated into our core software.

- We completed design and implementation of our primary 3D casing viewer. This software allows the user to view and interact with the 3D surface topography as rendered using the computer's graphics card. The user can zoom, rotate, and pan the rendered surface. Because the model is a true 3D rendering, we allow the user to change the position of the light source. This functionality is important for visualization as the direction of the light significantly affects the features visualized. It is important to note that the lighting is only for the user and is **not** used by our case matching algorithm. Our algorithm matches the 3D surface topography independent of the light source. We have found that firearms examiners very much like the ability to position the light as it feels very natural to them. The user can save snapshots of the view displayed on the screen. Finally, the user has the ability to annotate the surface with a dragged shape (*e.g.*, circle) to indicate regions of interest.

- We have completed the first version of our primary user interface (now version 0.9.4). This program provides the firearms examiner a simple graphical user interface through which they can collect 3D scans, automatically mask the region of the breech-face impression, enter casing, firearm, and incident data, visualize and annotate the 3D scans, conduct database searches, and visualize the matches (with accompanying match score) (Figure 6). We have implemented several search

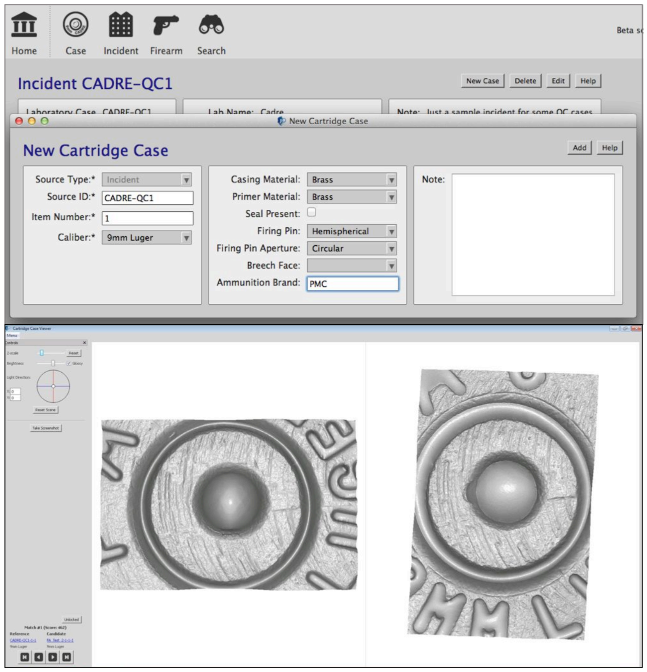


Figure 6: **User Interface Software Images.** Our prototype user interface software allows the firearms examiner the ability to visualize, database, search, and score casings. (Top) Example window showing the primary object types (*i.e.*, Casings, Incidents, and Firearms) and data entry for the creation of a new Casing. (Bottom) Example side-by-side 3D visualization of two matching casings. These two casings have a high match score and have been automatically identified and aligned by our software. The user can lock the two casings such that they rotate together and can rotate the casings both in the plane and in full 3D.

types. A user can compare a casing, firearm, or incident against either the entire database or another casing, firearm, or incident. The ability to restrict search in this way lets the operator conduct fast queries where, for example, test fires from a single firearm could be compared to only the incident of interest. The results of such as search are typically ready within a minute. We continue to improve the features of this software and are issuing regular updates to test sites.

- We've deployed scanners to four collaborating labs. Please see section, "Deployment Studies" below.

- We have started to disseminate our research results. We gave an oral presentation at the 2013 AFTE meeting in Albuquerque, NM. This was an excellent event. We were able to share many of the results described in this report. We were also able to speak with several forensics labs and solidified the involvement of several local, state, and federal crime labs. Todd Weller also presented our work at the California Associate of Criminalists Fall Seminar. Our abstract for AFTE 2014 has been accepted for oral presentation (May 2014). We are currently writing our first research paper summarizing our system, matching algorithm, and case matching experiments.

- We collected test fires from 50 different 9mm Luger caliber firearms using three ammunition types. For each firearm we collected three test fires of PMC Brand (115GR bullet, brass casing, brass primer), two test fires of Remington Brand (115GR bullet, brass casing, nickel primer), and two test fires of RWS Brand (125GR, brass casing, nickel primer). Three of these firearms are Glocks and so we hold them aside when considering our all-vs-all tests. This is because an examiner would never compare a case with a rectangular firing pin aperture to a case with a round aperture.

- We collected 75 test fires from 25 additional 9mm Luger caliber firearms using Winchester brand ammunition (brass primer, brass casing). Unfortunately the primer on these casings was coated with a lacquer that complicates tool mark capture. It is known that when a lacquer covers the primer that the breech-face impression is not efficiently transferred to the primer. That is, the force of the primer against the breech-face shatters the lacquer. The shattered lacquer falls off the casing as is not recovered. The underlying brass of the primer typically does not pick-up the impression of the breech-face. The result is a relatively smooth breech-face impression, or simply a breech-face impression with non-informative tool marks. Unfortunately this means that the 75 casings collected for these 25 firearms are not useful for our analysis. We have kept the casings in hope that we can revisit casings with lacquer sealant. show figure
- Using the scanned test fires, we masked the breech-face impression and conducted an all-vs-all casing comparison (experiment fully described below).
- We completed a small fingerprint data collection experiment. See section "Fingerprint Scans" below.
- We collected a small number of scans of 22 Long, 25 Auto, 32-Auto, 380 Auto, 40 S&W, and 45 Auto caliber cartridges. Unfortunately these scans were collected by our collaborators using a defective batch of gel. As such, the scans were highly granular in appearance and were not good candidates for matching. A few of the results are presented below.
- We collected data for the reproducibility and persistence experiments. In these experiments we scanned the same three casings 50 times (one run of 30, two runs of 10). This experiment produced two results. First, we observe little to no signal degradation over the 50 scans. In other words, the act of scanning does not appear to alter the casings. Second, we observe no noticeable persistence (or memory) within the gel. In other words, if we use a single piece of gel to scan casing X and then casing Y, the resulting scan of casing Y does not look like casing X. That is, the gel has no memory of casing X when we use it to scan casing Y.

5.1 Development of Statistical Scoring Function

We developed a simple statistical scoring function. Matches with our original algorithm are ranked according to the number of features that are identified to be consistent. Towards the end of 2013 we had collected a sufficient number of casings to begin construction of a statistical scoring function to complement the initial number of matched features score. This statistical scoring function can assign a more interpretable "probability of match" to each pairwise correlation. There are two goals in doing this. First, we wish to begin putting breach-face impression matching on a firmer scientific footing by associating a statistically meaningful probability of match. Second, we want to improve the quality of the match ranking provided with this function, as it is able to account for not only the number of matched features but other quantities that measure the quality of those matches.

We have explored two main approaches. First we model the distribution of the number of matched features for known matches and known non-matches. From this distribution we derive the probability that a known non-match could generate the specified number of matched features. This represents the 'traditional' approach. Second, we explored a technique known as Logistic Regression which operates on a set of meta-features extracted about the match. Meta-features include, for instance, the number of matched features, the average difference in feature appearance, differences in feature scale, the size of the masked region, and the overall fraction of the masked region covered by matched features. We note that we have only begun to investigate these two approaches. For example, the current approach is fit

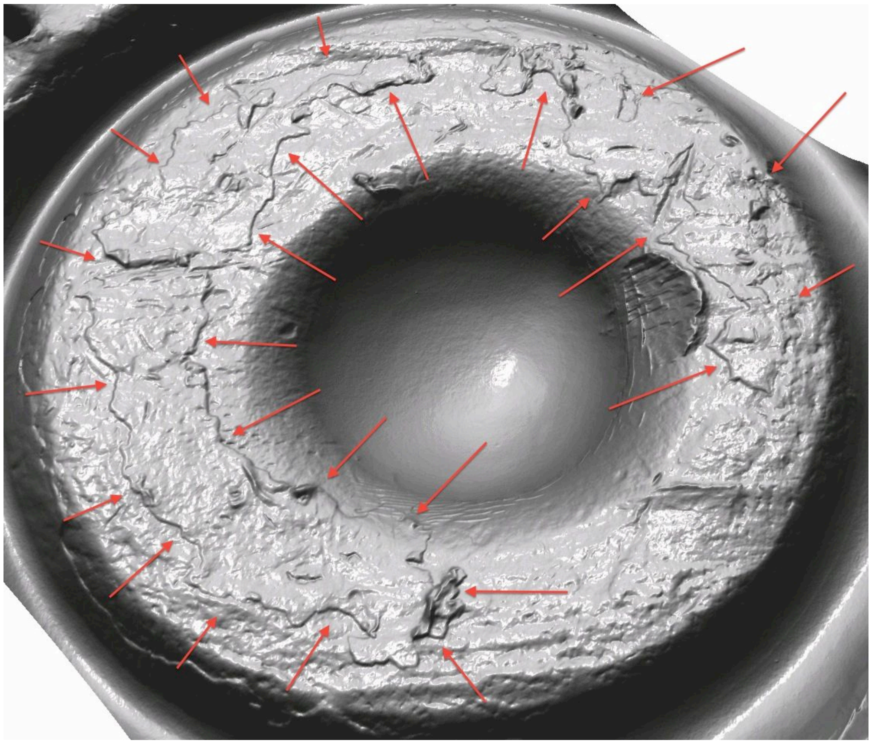


Figure 7: **Winchester Ammunition with Lacquer.** A rendered view of the 3D measured surface of a primer that had been covered with a protective lacquer prior to firing. The lacquer can be seen as a skin which lies on the surface of the primer and bunches in some locations (see arrows). The lacquer can obstruct the tool mark transfer process. In this example, the identifiable features of the breech-face impression have not transferred and the casing is not identifiable with sister casings from the same firarm.

using data from three ammunition types for seventeen of our initial 47 firearm test set. We will extend our model with additional training data as we accumulate more scans. The nature of our formulation allows us to observe if newly collected casings are consistent with the previously seen casings. As we incorporate more training data, we will be able to determine if the new data deviates in some unexpected way from the initial data.

The results of the Logistic Regression-based significance score are presented in many of the following result sections. We believe that the current model is a bit conservative. That is, at the 1 in 10,000 (99.99%) confidence level we might expect 20 of 200,000 known non-matches to appear significant (false positives); however, we see no false positives in 200,000 comparisons. The results are very promising and we will be extending this function in 2014.

5.2 Forty Seven Firearm Experiments

This series of experiments was designed as the first test of our redesigned scanning hardware, casing mount, and image analysis and matching algorithms. We selected 47 firearms including: 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 without preference to their ability to mark cartridge casings. The intent was to select firearms that would represent real-world conditions. In reality, the tool marks left on a casing range from being extremely reliable and interpretable (Fig. 8) to being unreliable, irreproducible, and barely present (Fig. 9). Our dataset includes casings of all types. Our collaborator Todd Weller, an expert firearms examiner, manually examined a number of the 350 collected casings and made a note that at least ten of the fifty firearms did not mark well. For example, for one pair of test fires from a poorly marking Hi-Point he noted, “This is a Hi-Point with an unfiled breach face. Overall shape of BF impression is similar, but a lot of ‘noise’ or non-reproduced marks are present under high magnification. Very difficult ID that would require a considerable amount of time to find and document agreement of marks. BF is mostly smooth.” (Fig. 9 Top-Left). Most firearms came from the Oakland reference collection. Some firearms and test fires were collected by our collaborator Andy Smith (SFPD).

For each firearm we collected three test fires of PMC Brand (115GR bullet, brass casing, brass primer), two test fires of Remington Brand (115GR bullet, brass casing, nickel primer), and two test fires of RWS Brand (125GR, brass casing, nickel primer). Our analysis was complicated by the fact that the RWS brand ammunition was found to have a manufacturing defect (confirmed by speaking with the manufacturer, Ruag Ammo). They confirmed that a minor contaminant likely got onto the battery cup before plating. The contaminant prevents the nickel plating from bonding and it can therefore peel or bunch during firing. This peeling results in a paper or ‘skin’ like appearance and is clearly visible in Figure 9-TopRight. This defect does not affect the functioning of the ammunition, but it does introduce non-firearm specific geometry onto the primer. This peeling is visible on at least 12 of the imaged casings. It is likely that lesser degrees of peeling are present on many of the RWS casings. Despite this peeling, sufficient identifiable tool marks are present on these casings to allow our case matching algorithm to find the correct match.

Finally, we note that in the following experiments the only tool marks used in matching are those of the breech-face impression. Our matching algorithm does not consider the firing pin impression, the primer shear (aperture shear), or the ejector mark. Note that many other systems and algorithms group the true breech-face impression and aperture shear together. We do not. In 2014 we will be incorporating analysis of aperture shear. We note that the excellent performance described throughout this report is based only off the comparison of breech-face impressions.

5.2.1 Experiment One: Single Ammunition Type

The first experiment included three test fires of PMC brand (115GR bullet, brass casing, brass primer) from each of the 47 firearms. Each casing was cleaned with rubbing alcohol and a soft brush prior to scanning. The 3D surface topography was measured using our system at an image resolution of 1.4μm/pixel. For each casing, the breech-face was masked using our auto-masking algorithm and automatically compared to all other casings using our feature-based matching algorithm. The entire all-vs-all comparison (19,881 comparisons) required $\sim$3.75 hours of compute time[1]. Our iterative matching algorithm identifies a maximal set of self-consistent matched features corresponding to geometric parts of traditional tool marks. 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). A set of matches is considered self-consistent if the matched features of two casings can be spatially aligned after a single rotation and translation of one image (modulo the scale). The score of the match is a function of the number and quality of matched features. Note that we are now supplementing these numerical matches scores with a statistical significance score (*i.e.*, the probability that a match of specified magnitude would occur due to random chance), see section below.

The score matrix (often referred to as a ‘confusion matrix’) for this experiment is shown in Figure 10. The 282 Known Matches (KM) have significantly larger scores than the 19,458 Known Non-Matches (KNM). Although we do not currently define a score threshold for calling matches, we note that no KNM

[1] All experiments were performed on a single high-end desktop workstation.

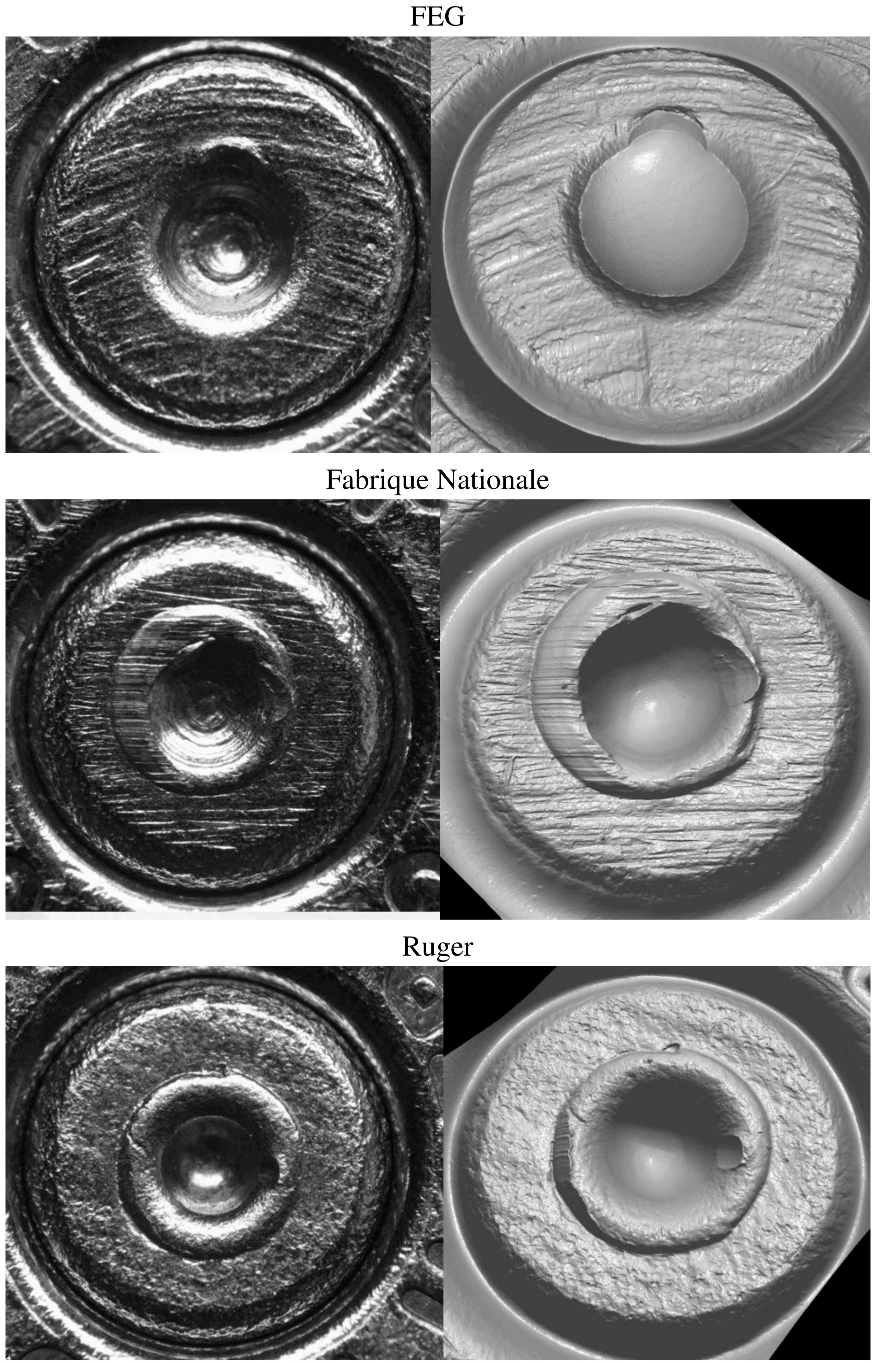


Figure 8: **Three Examples of Good Scans.** Three scans using the PMC brand ammunition (brass case, brass primer). In each pair, a traditional light microscope image is shown on the left while a rendering of the TopMatch-GS 3D collected surface is shown on the right. All three casings match correctly. (Top) FEG brand firearm showing milled marks. (Middle) Fabrique Nationale brand firearm showing filed type marks. (Bottom) Ruger brand firearm showing granular impressed marks.

has a score above 50 and less than 1% of KNM have a score above 40. This phenomenon holds across all experiments performed (see Large Dataset Experiment section below). Known Non-Matches always have a score below 51. This means that if two casings have a match score greater than 50 we can be

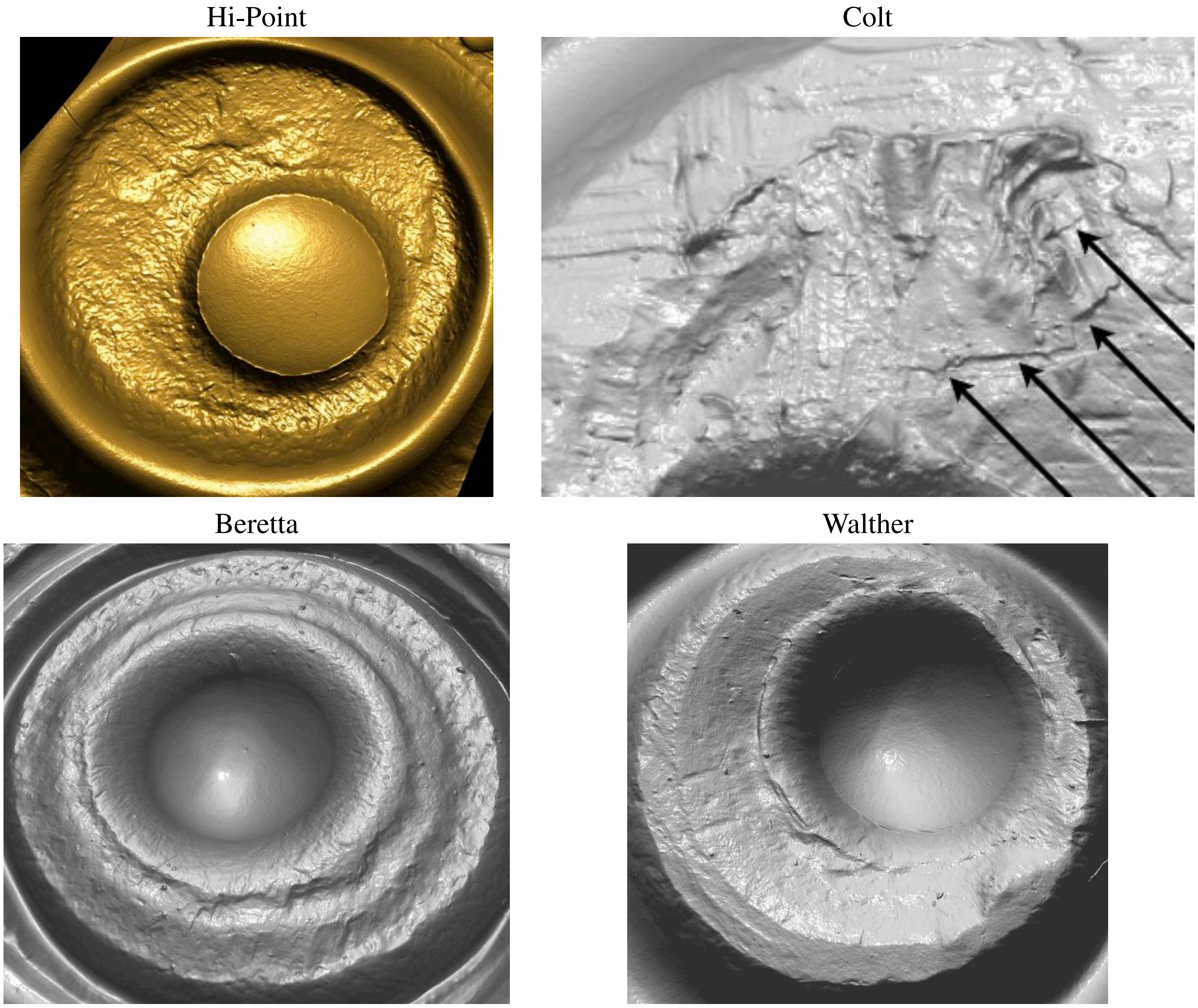


Figure 9: **Examples of Difficult Scans.** (Top-Left) Rendering (in brass color) of TopMatch-GS 3D scan of a Hi-Point (PMC brand ammunition) that was flagged as difficult to match by a firearms examiner but which was correctly matched using our algorithms. (Top-Right) Close-up of the ‘peeling’ effect seen in a number of RWS brand ammunitions. The edge of the peeling is highlighted with arrows. This represents an ammunition manufacturing defect confirmed by the manufacturer. (Bottom-Left) Rendering of a casing from a 9mm Beretta (Winchester brand ammunition) showing a very small region of breech-face impression. Despite the fact that most of the field-of-view is taken up by the flow-back our algorithm is able to correctly identify the known matching casing. (Bottom-Right) Rendering of a casing from a 9mm Walther (Remington brand ammunition) showing a poorly-marked and sloped breech-face impression. The scarcity of features on this casing prevented a match from being found. The top two casings were included in the 47-firearm experiment, all four were included in the Large Data Set Test.

very confident that they are a correct match. In reality, most KM have scores well above 100. We note that some of the poorly marked casings are difficult to match. These difficult matches can have very low match scores, scores that overlap with the Known Non-Matches. These matches would be considered missed hits. However, it’s most important to stress that we have no false positives across all experiments performed. That is, we never call a Known Non-Match as a match. Stated another way, a KM involving two poorly marked casings may have a small match score, but a KNM never has a large match score.

Results for All Casings (both Sufficiently-Marked and Poorly-Marked). Each casing has a ranked list of candidate matches. For 79% of the casings, the top scoring candidate match has a score above 50 and 100% of these matches are correct[2]. The remaining 21% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 21% of casings has a score less than 50. In this situation, the algorithm considers the casings inconclusive and does not claim a match. There are no false positives.

Results for Sufficiently-Marked Casings. In this experiment the system automatically ignored casings considered poorly marked (*i.e.*, casings with fewer than 1000 identified features). Because the PMC brand ammunition marked well, 91% (128 of 141) of casings were considered sufficiently marked. For 82% of the casings, the top scoring candidate match has a score above 50 and 100% of these matches are correct[3]. The remaining 18% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 18% of casings has a score less than 50. In this situation, the algorithm considers the casings inconclusive and does not claim a match. There are no false positives.

Statistical Significance of Matches. We computed our statistical significance score for each pair of casings. The results of the statistical match score are shown on the right of Table 2. For 72% of the casings, the top ranking casing was a match at 99.99% confidence (1 in 10,000). For 60% of the casings, the top ranking casing was a match at 99.9999% confidence (1 in 1,000,000). Many of the remaining casings have correct matches at the top of their ranked list (see previous two paragraphs); however, they might not be statistically significant at the specified level. Most importantly, none (0%) of the KNM appear as a match at either the 99.99% or 99.9999% match confidence level.

5.2.2 Experiment Two: Three Ammunition Types

The second experiment included three test fires of PMC Brand (115GR bullet, brass casing, brass primer), two test fires of Remington Brand (115GR bullet, brass casing, nickel primer), and two test fires of RWS Brand (125GR, brass casing, nickel primer) for the same 47 firearms examined in the previous section. The 3D surface topography of each casing was measured, masked, and matched using the same protocol described in Experiment One. The entire all-vs-all comparison (108,241 comparisons) required $\sim$7 hours of compute time. Qualitatively the Remington and RWS brand ammunitions did not pick up the same quality of tool marks as the PMC brand ammunition. Part of this may be explained by the nickel plating manufacturing defect of the RWS brand ammunition. As a result only 233 of 329 (71%) of the casings were considered sufficiently marked (compared with 91% of the PMC brand ammunitions). As in experiment one, the 1974 Known Matches (KM) have significantly larger scores than the 105,938 Known Non-Matches (KNM). Once again, no KNM has a score above 50 (less than 1% are above 40) and there are no false-positives.

Results for All Casings (both Sufficiently-Marked and Poorly-Marked). Each casing has a ranked list of candidate matches. For 73% of the casings, the top scoring candidate match has a score above 50 and 100% of these matches are correct[4]. The remaining 27% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 27% of casings has a score less than 50. In these cases, the algorithm does not claim a match. There are no false positives.

[2]Most casings score significantly better than 50, for 61% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

[3]Most casings score significantly better than 50, for 67% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

[4]Most casings score significantly better than 50, for 47% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

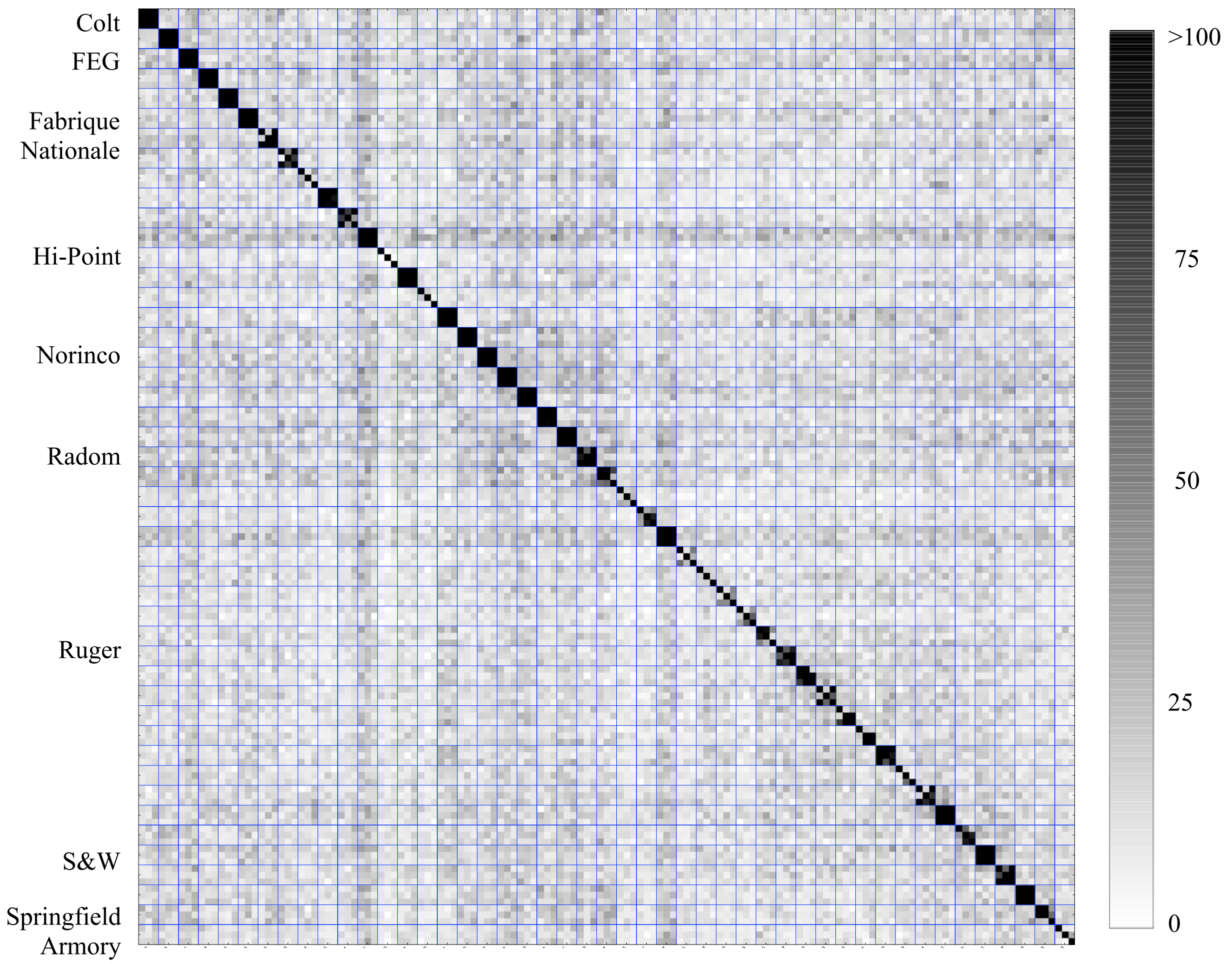


Figure 10: **Experiment One (47 Firearms, 141 Cases) Confusion Matrix.** The result of 19,881 comparisons. Each cell in the matrix corresponds to the match score between two casings (specified by the involved row and column). The firearms are grouped by manufacturer and separated in the matrix by thick blue lines. Match scores go from white (indicating a score of zero) to black (indicating a score of at least 100). The three test fires from each firearm are presented in order. Thus we expect a series of small 3x3 boxes of correct matches along the diagonal of this graph (each 3x3 box captures the match scores of the 3 test fires from each firearm with themselves, *i.e.*, correct matches). Most firearms match extremely well, with a small tight black 3x3 patch of known matches. The KNM are in the 'off-diagonal' position, they are all white or very light grey indicating a small match score. A clearer view of a similar confusion matrix appears in Figure 11.

Results for Sufficiently-Marked Casings. In this experiment the system automatically ignored casings considered poorly marked (*i.e.*, casings with fewer than 1000 identified casings). For 83% of the casings, the top scoring candidate match has a score above 50 and 100% of these matches are correct[5]. The remaining 17% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 17% of casings has a score less than 50. In these cases, the algorithm does not make claim a match. There are no false positives.

Statistical Significance of Matches. We computed our statistical significance score for each pair of casings. The results of the statistical match score are shown on the right of Table 2. For 57% of the

[5] Most casings score significantly better than 50, for 63% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

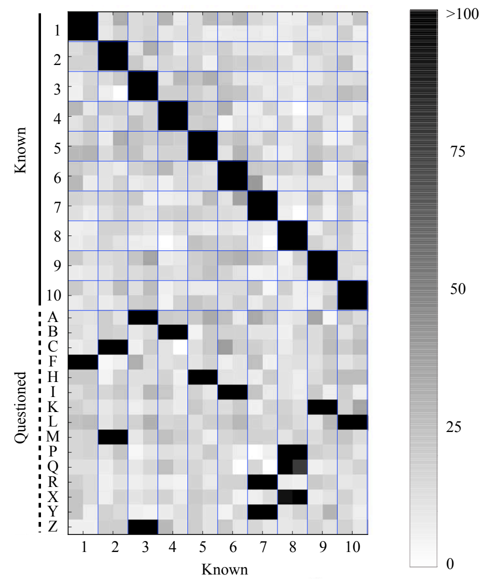


Figure 11: **Experiment Three (Miami-Dade Study) Confusion Matrix.** The confusion matrix shows performance attained when the casings are all well-marked. The distributions of KM and KNM score distributions are extremely distinct. The top of the matrix shows the match scores for the matched knowns. The lower half of the matrix shows the scores between the 15 questioned casings and the matched knowns. The correct answers (KM) have very large scores and are easy to spot. The algorithm achieves 100% accuracy on this 'clean' test set.

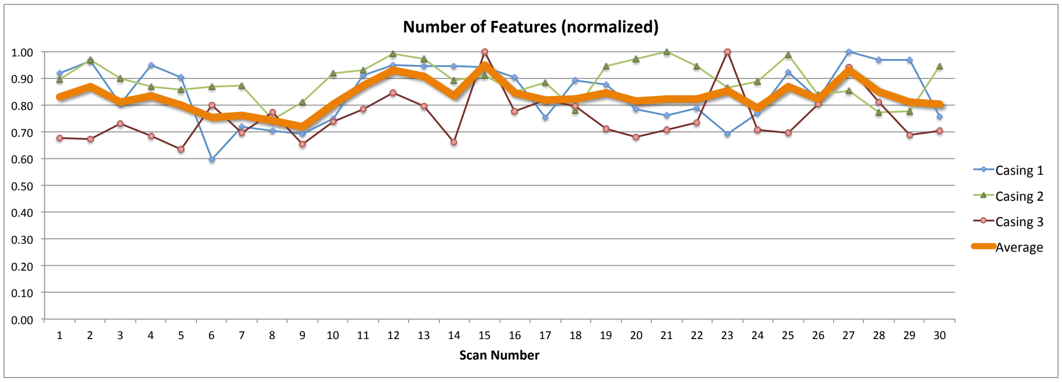
casings, the top ranking casing was a match at 99.99% confidence (1 in 10,000). For 46% of the casings, the top ranking casing was a match at 99.9999% confidence (1 in 1,000,000). Many of the remaining casings have correct matches at the top of their ranked list (see previous two paragraphs); however, they might not be statistically significant at the specified level. Most importantly, none (0%) of the KNM appear as a match at either the 99.99% or 99.9999% match confidence level.

5.3 Experiment Three: Miami-Dade Study

We next explored the Miami-Dade Study (Test Set 8) test fires (materials kindly provided by Dr. Thomas Fadul). We included the Miami-Dade Study to demonstrate the performance of our system on a set of well marked casings (in contrast to the real-world test sets of the previous sections). The Miami-Dade set consists of ten pairs of matched knowns and fifteen individual 'questioned' (or unknown) cases. 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 our first two experiments, the cases in the Miami-Dade Study are all strongly marking. This is reflected in our match score results. The average known match has a match score of 266 (range: 75-545, only two of fifty KMs scores below 100). All questioned cases are correctly identified with their matching known pairs. The average known non-match has a score of 13.8 (range: 0-38, less than 10% score above 30). The score matrix for this experiment is shown in Figure 11.

5.4 Experiment Four: Reproducibility and Persistence Studies

The reproducibility and persistence experiments had two aims. First, the reproducibility studies aimed to determine if there are any noticeable degradative effects caused by repeatedly scanning the same casing. Second, the persistence studies aimed to determine if a casing leaves a memory or imprint in the gel



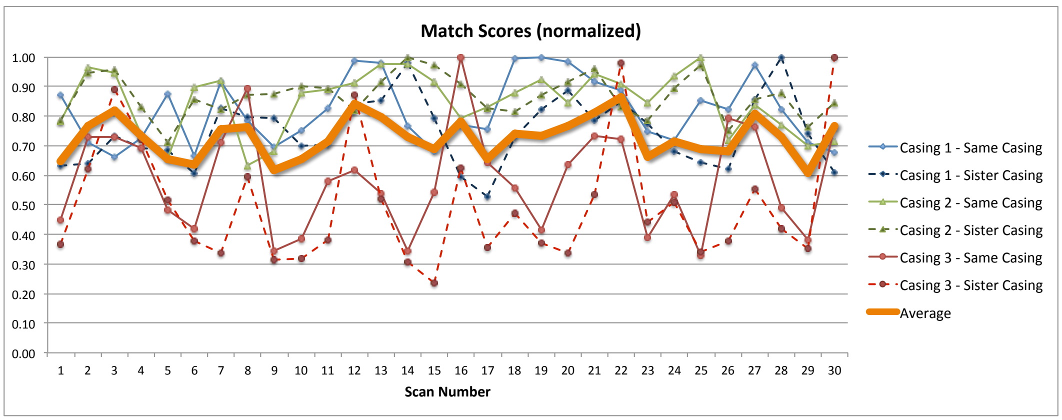


Figure 12: **Replicate Studies.** (Top) The Number of features detected on each replicate scan for the three test casings. The number of features are normalized so that the maximum number of features for each casing have a value of 1.0. The average (thick orange line) shows that there is no downward trend. The number of detected features is fairly consistent over time. (Bottom) The Match scores for each of the three casings to a second scan of itself (Same Casing, solid line) and to a different, sister, casing from the same firearm (Sister Casing, dashed line). The scores for each of the six experiments are normalized so that the largest score for each experiment has a value of 1.0. The average (thick orange line) shows that the average match scores do not decrease with repeat scans.

which can impact the scan quality of additional casings acquired using the same piece of gel.

Reproducibility. For the reproducibility part of the experiment, the following protocol was followed for each of the three different casings (one from each of three different firearms):

- Each casing was scanned 30 times using a single piece of gel. Between scans, the casing was physically removed from the scanner, rotated, and reinserted.
- For each run, we computed the number of detected features (Fig. 12 Top).
- For each run, we computed the match score to a second, previously collected scan of the same casing (Casing x - Same Casing data) and to a previously collected scan of a different casing shot through the same firearm (Casing x - Sister Casing data). This data is shown in Figure 12 Bottom.

Overall, although there is scan to scan variability in the number of detected features, there is no downward trend. Repeat scans do not have a lower number of features detected nor do they have lower match scores to the same or sister casings.

Persistence. For the persistence part of the experiment we extended the above protocol. Once again we used casings from three different firearms. After collecting 30 scans of a single casing a (from firearm

	Firearm A		Firearm B		Firearm C	
	Self	Target A	Self	Target B	Self	Target C
Persistent Scan of a after Target	-	-	0.99	20.0 (6.0)	0.98	16.8 (6.5)
Clean Scan of a	-	-	0.70	15.7 (4.6)	0.70	14.5 (3.8)
Persistent Scan of b after Target	0.60	18.3 (6.7)	-	-	0.88	12.4 (5.3)
Clean Scan of b	1.06	15.9 (7.0)	-	-	1.06	10.6 (4.2)
Persistent Scan of c after Target	1.12	16.3 (6.1)	0.64	12.2 (5.8)	-	-
Clean Scan of c	1.10	13.0 (6.3)	1.10	11.9 (3.8)	-	-

Table 1: **Persistence Results.** The three major columns represent the results comparing to Firearms A, B, and C. Each row corresponds to a different casing. The rows labeled 'Persistent Scan' are the casings scanned after thirty scans of a casing from a different firearm (as specified in the column header). The rows labeled 'Clean Scan' are the casings scanned on a clean piece of gel. The 'Self' column score is the average normalized match score of the specified casing to two sister casings from the same firearm. Therefore, a self-score close to 1.0 indicates that the match score of the specified casing to the sister casings is equal to the average match score of those casings.. The 'Target X' column lists the mean and standard deviation of the scores for the persistent or clean scan (row) vs thirty replicates of a casing from the specified target (column). All Target scores are low (as expected) and there are no false positive results. However, we note that the Target scores for the Persistent scans are typically slightly higher than the Target scores for the Clean scans. The effect is slight.

A) we used the same piece of gel to scan two different casings b and c (from two different firearms, B and C). In theory, if the casing from A left an impression in the gel, the final scan of the B (or C) casing would look more like A than would a clean scan of B (or C). We call the scan of casing b collected after thirty scans of casing a, the 'persistent scan of b after a'. The results are presented in Table 1. The details of the scoring are explained in the Table caption.

We note that unlike the Reproducibility studies above, the number of sample points averaged in computing Table 1 is small. Therefore, we expect a higher variance in the Persistence results. The Self scores for the Persistent and Clean scans are generally similar. There is no trend that the Clean casing scans have higher Self scores than the Persistent scans. All Target scores are low (as expected) and there are no false positive results. However, we note that the Target scores for the Persistent scans are typically slightly higher than the Target scores for the Clean scans. The effect is small, contributing perhaps between 1 and 4 additional features to the match score. These features could simply be noise, or they could be due to permanently embedded dust or structural damage to the piece of gel. The possibility is certainly something to note; however, because we do not anticipate anyone scanning the same casing thirty times we are confident persistence effects will not cause a false positive match. We plan on further persistence studies in 2014.

5.5 Deployment Studies

Our collaborator at the Oakland Police Department, Todd Weller, received our first scanner in April 2013. Since that time, Todd has been scanning casings and providing valuable feedback with respect to hardware and software design decisions. In September we setup scanners and performed in person hands-on training for approximately twelve firearms examiners at two test labs in the Bay Area (including the San Francisco Police Department (Andy Smith) and another government lab). In December, we moved the SFPD machine to the Contra Costa County Office of the Sheriff (Chris Coleman). All labs are now using our scanners and casing analysis software. As described above, the software allows the examiner to collect 3D scans, enter casing, firearm, and incident data, conduct database searches, and visualize the results. We have regular communication with all test sites to obtain usability feedback. The California labs have the same goals. We want the users to get familiar with using the new technology, to collect a number of scans, and to provide feedback on both the hardware and software design. The feedback from these labs has been invaluable. We have changed naming conventions, optimized workflow, and improved our training sessions / materials.

Most recently (December 2013), we deployed a machine to Professor Karl Larsen at the University of Illinois Forensic Science Training Program. Dr. Larsen will be working with two masters students on two projects. They will be scanning a number of standard Competency Tests and will be investigating the effects of lacquers on primer surfaces.

Across the five test sites we have now trained 22 individuals in the use of the scanner. During the training session, we have the operators scan three quality control test fires. Every examiner was able to collect a high-quality scan of the quality control test fires and match them against a prepopulated database. It typically takes a few sessions before an operator is comfortable with the machine and is able to capture high-quality scans.

As part of the deployment studies we asked our collaborators to collect a small number of scans of different caliber casings. We obtained a few scans (typically less than 10 pairs) of 22 Long, 25 Auto, 32-Auto, 380 Auto, 40 S&W, and 45 Auto caliber cartridges. Unfortunately these scans were collected using a defective batch of gel. As such, the scans were highly granular in appearance and are not good candidates for matching. Because these scans are defective we hesitate to list the results; however, for the 45-Auto caliber tests, we had ten firearms, each test fired 3 times. 87% of the casings had a correct match as their top scoring candidate casing. In the remaining 13% of casings, the top ranked candidate casing did not have a significant score and would not have been identified as a match. There were no false positives. We will revisit these different calibers in 2014 now that the gel formulation issue has been resolved.

5.6 Experiment Five: Large Test Set Experiments

As part of the deployment study, our collaborating labs scanned a number of casings. To create our Large Test Set we pulled all 9mm Luger casings[6] with a hemispherical firing pin and circular firing pin aperture. Approximately 100 casings were included from the deployment tests. Altogether over 100 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 (including 10 with consecutively manufactured breech-faces), Norinco, Sig Sauer, Springfield Armory, Star, Taurus, Uzi, and Walther. A total of 417 of these 431 casings have a known match in the test set, 14 casings do not have a known match. This set represents real-world casings. The firearms were simply pulled from a reference collection without regard to their ability to produce tool marks. We note that this approach differs from the way many labs currently use database systems. When an examiner enters test fires into one of the current database systems, they often select the best test fire among all those available. If all test fires are extremely poorly marked then an examiner may elect to not enter the scan. 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 approach, we did not prescreen test fires prior to entry into our system. Many of the casings that we've included in our set would not have been deemed high enough quality for entering into a traditional database system (*e.g.* IBIS).

Results for All Casings (both Sufficiently-Marked and Poorly-Marked). Each casing has a ranked list of candidate matches. For 69% of the casings, the top scoring candidate match has a score above 50

[6] We had to discard a few scans that were badly warped. Warping results from improper scanning technique and is more common among new operators of the system.

and 100% of these matches are correct[7]. The remaining 31% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 31% of casings has a score less than 50. In these cases, the algorithm does not claim a match. There are no false positives.

Results for Sufficiently-Marked Casings. In this experiment the system automatically ignored casings considered poorly marked (*i.e.*, casings with fewer than 1000 identified casings). For 77% of the casings, the top scoring candidate match has a score above 50 and 100% of these matches are correct[8]. The remaining 23% of casings do not mark well enough for a significant match to be identified by the algorithm. In other words, the top scoring candidate for these 23% of casings has a score less than 50. In these cases, the algorithm does not make claim a match. There are no false positives.

Statistical Significance of Matches. We computed our statistical significance score for each pair of casings. The results of the statistical match score are shown on the right of Table 2. For 52% of the casings, the top ranking casing was a match at 99.99% confidence (1 in 10,000). For 43% of the casings, the top ranking casing was a match at 99.9999% confidence (1 in 1,000,000). Many of the remaining casings have correct matches at the top of their ranked list (see previous two paragraphs); however, they might not be statistically significant at the specified level. Most importantly, none (0%) of the KNM appear as a match at either the 99.99% or 99.9999% match confidence level. It is interesting to note that we see no false positives at the 99.99% confidence level despite the strict (simple) interpretation that 1 in 10,000 KNM might score at this confidence. If 1 in 10,000 KNM are expected to match due to random chance at 99.99% confidence then we might expect 19 (193,732 / 10) KNM to falsely score as significant. The fact that this does not happen suggests that our confidence score may be conservative. In other words, what we're calling 99.99% might represent "at least 99.99%" but more likely 99.9999%.

5.7 Fingerprint Studies

Although not part of the original proposed aims, we were asked by our program manager if we could collect a few scans of fingerprints to demonstrate the quality of 3d reconstruction and to determine if fingerprint scans could be matched using our algorithm. A sample scan with resolution of 4.2 microns/pixel is shown in Figure 13. We collected a set of fingerprint scans from two individuals. Fingerprint scanning presents a different set of challenges than scanning cartridge casings. We created a simple finger holder which attached to our casing mount and holds the finger against the gel. Because the fingerprint is physically larger, we dialed the zoom back to 1x and collected the fingerprint scans at a resolution of 4.2 microns per pixel. Care was taken to hold the finger as still as possible. Although image acquisition time is only 12 seconds, any slight movement during the scan can potentially influence the measurement. This movement is obviously not a concern when scanning casings. Significant detail, including sweat pores, can be seen on the scans. We applied the same feature-based image matching to the collected scans. That is, we ran our matching code as-is, without modification. Approximately 20-times as many features were identified on the fingerprint than on a typical cartridge casing. This is likely the result of a larger region of interest (*i.e.*, a larger masked region) and almost no smooth locations. The large number of features slows the scan comparison time; closer to 15 minutes were required to compare two fingerprints. A number of modifications should be made to our algorithm if we decide to compare fingerprints. Most importantly is an automatic curation of detected features and an incremental comparison approach.

[7]Most casings score significantly better than 50, for 44% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

[8]Most casings score significantly better than 50, for 57% of the casings, the top scoring candidate match has a score above 100 and 100% of these matches are correct.

DataSet	# Casings	Number (Fraction) of Casings w Correct Top Result: Match Score Any	> 50	> 100	Confidence at x 99.99	99.9999
47 Firearm 1 Ammo	141	123 (0.87)	111 (0.79)	86 (0.61)	102 (0.72)	85 (0.60)
47 Firearm 3 Ammo	329	281 (0.85)	241 (0.73)	154 (0.47)	186 (0.57)	152 (0.46)
Large Test Set	417	337 (0.81)	286 (0.69)	182 (0.44)	218 (0.52)	180 (0.43)

DataSet	# KNM	Number (Fraction) of KNM that Satisfy: Match Score > 50	> 100	Confidence at x 99.99	99.9999
47 Firearm 1 Ammo	19,458	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
47 Firearm 3 Ammo	105,938	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Large Test Set	193,732	1 (0.000005)	0 (0.00)	0 (0.00)	0 (0.00)

Table 2: **All-vs-All Matching Results.** The performance is shown for the three real-world data sets (all casings included, no casings were excluded with the poorly marked casing criteria). (Top) Known-Matches. The Match Score columns show the number (and fraction) of casings whose best scoring candidate casing satisfies the listed criteria and is a correct match. For example, in the 47 Firearm, 1 Ammo test set, 111 casings have a best scoring casing that has a match score greater than 50 and it is a correct Known Match. The 'Any' column is the number of casings whose best ranked match is correct regardless of score (*e.g.*, if the top ranked match for a casing has a score of 45 and is correct it would be counted in the Any column, but not the > 50 column. The 'Confidence at x' column ranks matches based on their statistical significance score and counts the number of casings whose best scoring casing is significant at the specified confidence. For example, in the 47 Firearm, 1 Ammo test set, 102 casings have a best scoring casing that the algorithm finds significant at the 99.99% confidence level (1 in 10,000) and 85 casings have a best scoring casing that is significant at the 99.9999% confidence level (1 in 1,000,000) and these matches are correct. (Bottom) Known Non-Matches. Note the large number of KNM in these data sets. The table shows that there is 1 false positive (of 193,732) if using the low > 50 threshold. However, there are no false positives for any of the other thresholds.

The resulting confusion matrix is shown in Figure 13D and shows that among these two individuals (X and Y) the correct scan is identified every time. All correct known-matches have scores above 175. All known non-matches have scores below 19.

6 Conclusions

6.1 Discussion of Findings

Through the work of this grant we demonstrated that our GelSight-based imaging system, TopMatch-GS 3D, is able to accurately capture and compare high-resolution three-dimensional surface topographies of cartridge casings. We completed all proposed project aims. We found that our computationally-assisted auto-masking algorithm works well to identify the region of the breech-face impression and that our feature based comparison algorithms demonstrate excellence performance on real-world data.

We tested our imaging and analysis system using several experimental datasets. Three datasets demonstrate the system's current performance on real-world casings (including well and poorly marked casings). A final dataset demonstrates our performance on clean, well marked test fires. We achieve excellent performance on good marking casings and surprisingly good performance on extremely challenging casings, casings that a firearms expert feels would be difficult and time consuming to match using manual comparison. Most importantly, all KNM have low scores and are not considered a significant match at either the 99.99% or 99.9999% confidence levels. We observed no false-positives across

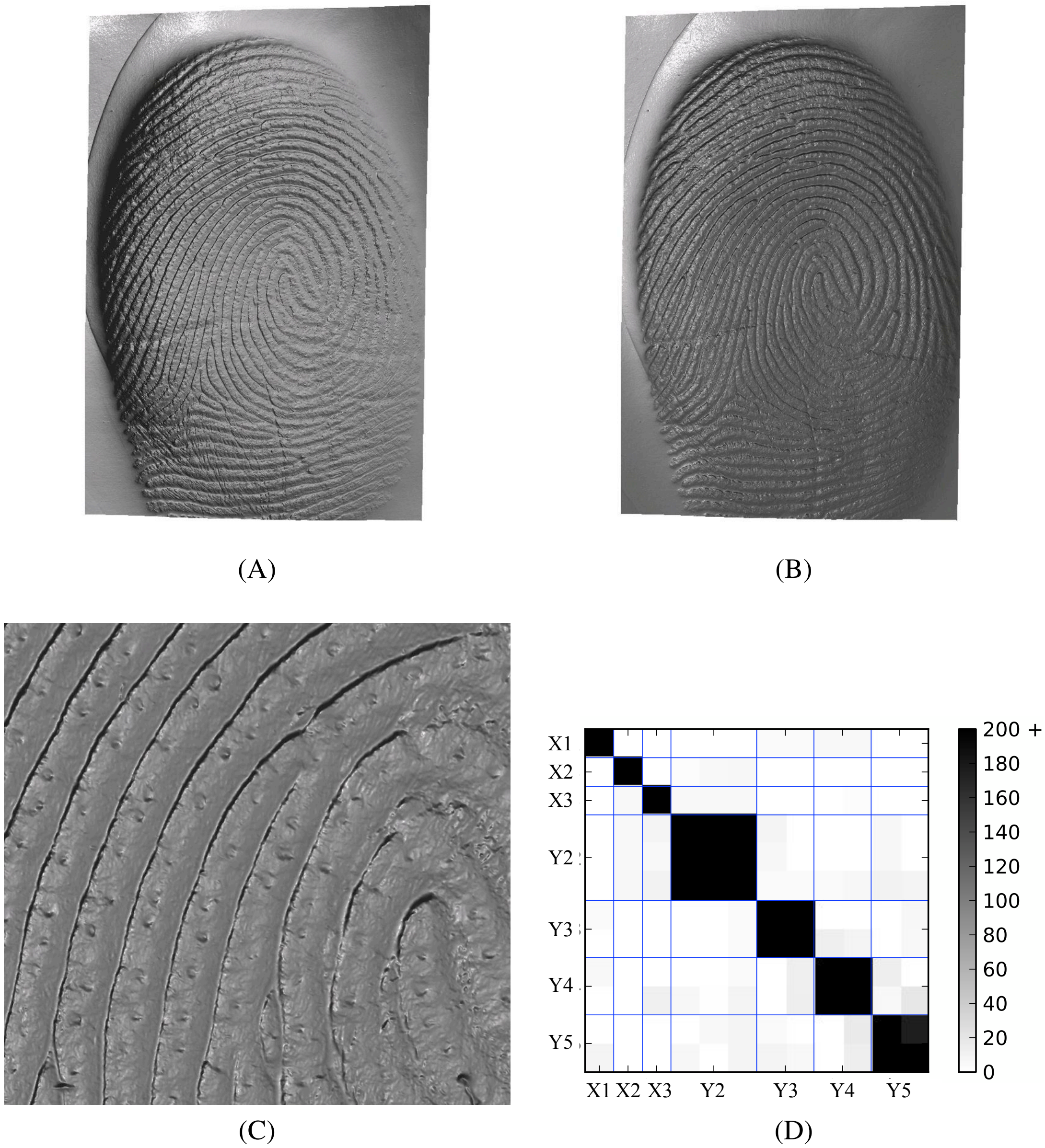


Figure 13: **Fingerprint Scans.** A scan of a fingerprint is shown. Figures (A) and (B) show a single scan lit from two different directions. Figure (C) is a zoomed view, sweat pores are clearly visible. Figure (D) shows the confusion matrix between two individuals, X and Y. Three fingers were scanned for individual X (X1, X2, and X3) and four fingers were scanned for individual Y (Y2, Y3, Y4, Y5). One scan was collected for fingerprints X1, X2, and X3. Two scans were collected for fingerprints Y3, Y4, and Y5. Three scans were collected for fingerprint Y2. All correct known-matches have scores above 175. All known non-matches have scores below 19. Elements on the diagonal correspond to comparison with itself. The scan was collected in a single image with a per-pixel resolution of 4.2 microns.

approximately 200,000 comparisons.

As described above, a small number of difficult Known Matches have a low match score. It is understandable that due to a paucity of surface features, some casings are inherently difficult to match. For these casings, our algorithm typically detects fewer surface features. Our software can automatically identify these casings and will soon be able to notify the examiner that it might be beneficial to collect another test fire.

We completed a series of reproducibility and persistence studies which suggest there is no noticeable effect of repeated scanning and gel use. Our short investigation into the use of our system for scanning fingerprints demonstrated proof-of-concept that our methods would work in that domain as well.

We have identified a number of critical areas for improvement. First, in 2014, we will extend our scoring function (and statistical model) to incorporate striated marks (starting with the aperture shear). Second, we will extend our approach for identifying poorly marked casings. Our initial method of thresholding the number of identified features does improve overall performance; however, we feel we can do better. Third, we will improve our statistical scoring function. Our current approach seems to be an excellent start. The method is somewhat conservative. For example, at a 99.99% confidence we would expect 1 in 10,000 known non-matches to be identified as a false positive. However, in the Large Data Set study we identify zero KNM as falsely positive at the 99.99% confidence level. We have plans to improve the statistical significance score by incorporating a generative model of feature matching and additional meta-features. We are currently working on these extensions. We believe our statistical scoring function to be an excellent step towards assigning an accepted statistical significance or confidence value to a pair of matched casings. Every firearms examiner we have spoken with agrees on the importance of this work with respect to their ability to make a confident identification as well as their ability to have their findings accepted in a courtroom.

Our deployment studies were extremely useful. Teaching firearms examiners to use the new technology taught us how to make the system more easy to use. Their feedback helped identify parts of the system that were working well and those areas that needed improvement. We did observe some differing of opinions. For example, some labs felt strongly that the firing-pin drag mark should be oriented at the 3-o'clock position whereas others said 12-o'clock and others said it didn't matter. We've compromised and the next version of the software will allow each user to select how the casing is displayed. We also completed an informal inter-operator scanning study. All operators trained on our equipment were asked to scan one of three quality control casings. They were able to obtain a statistically significant match score every time. We plan on extending this work.

In summary, we have redesigned the base scanner, designed and machined a fully functional cartridge case mount, improved our matching algorithm, achieved excellent performance across several experimental datasets, created software which allows 3D visualization of cartridge casings and database search, deployed our hardware and software system to forensics labs, and have assembled a large experimental set of test fires. These results all strongly support the goals of the proposed study. We have demonstrated a novel technique for the imaging and analysis of cartridge casings and we have shown excellent performance compared to the state-of-the-art alternatives.

6.2 Implications for 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: it increases the quality and efficiency of forensic analysis, it develops new instrumentation systems, and it provides a novel approach to enhancing the analysis and interpretation of forensic data derived from physical evidence.

The criminal justice system and forensic laboratories are now aware of our work. We are developing analytic techniques, grounded in mathematical science that are able to provide accurate quantitative sample comparison and database search. This benefits the criminal justice system and their ability to present firearm identification and tool mark evidence in the courtroom. It further benefits firearms examiners who can now more efficiently and more confidently search large databases to identify connections between crimes and known firearms. Additional impact will be made when our systems become available to all crime labs in 2014.

We were pleased to receive great interest following our presentation at AFTE. Several crime labs were excited about our system and in collaborating with its development. At least four crime laboratories (including Oakland, San Francisco, and Contra Costa County) now have access to our newly developed 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 is faster, is more accurate, and offers more transparency with respect to how the scanner works, the details of matching, and the interpretability of the match score.

6.3 Implications for Further Research

The progress made through the current award and the experimental results obtained lay the groundwork for additional research. Three aims will be addressed in 2014 under a newly awarded NIJ research grant. 1) Striated Mark Imaging and Analysis. We will develop methods of imaging and comparing striated tool marks (*i.e.*, primer shear and land engraved areas). 2) Development of an Extended Statistical Significance Model. This model will extend that currently under development for impressed marks and will incorporate both impressed and striated marks. 3) Cross-Modality Matching. We will develop and test algorithms for comparing 3D surface topographies collected using different scanning modalities (*e.g.*, GelSight and Confocal Microscopy). These three aims all build off the successful work completed in 2013.

7 Dissemination of Research Findings

We presented our system and experimental results at the 2013 AFTE meeting in Albuquerque, NM. Our presentation, "Development of a 3D-Topography Imaging and Analysis System for Firearm Identification using GelSight and Feature Based Case Matching" was given by Lilien and Weller during the technical session. The presentation was also given to our NIJ Program Managers Danielle McLeod-Henning and Gregory Dutton (July 2013). Todd Weller presented our work at the California Association of Criminalists Fall Seminar (October 2013). We are currently preparing a paper for submission to a peer reviewed journal. We anticipated submitting this by the end of 2013; however, we decided to wait for the results using a larger test set of casings that includes scans from our deployment studies. We now plan on submitting the paper in the first quarter of 2014. The publication will introduce our imaging and analysis system and will discuss the cartridge casing matching experiments described above. Finally, our work was accepted for oral presentation at the 2014 AFTE meeting in Seattle, WA. We will present the results of our 2013 work (as described in this report) and our preliminary work through the first half of 2014.

References

[1] P. Babenko and M. Shah. MinGPU: a minimum GPU library for computer vision. *J. Real-Time Image Processing*, 3:255–68, 2008.

[2] A. Banno, T. Masuda, and K. Ikeuchi. Three dimensional visualization and comparison of impressions on fired bullets. *Forensic Science International*, 140:233–40, 2004.

[3] M. Barrett and G. Warren. Portable forensic ballistics examination instrument: advanced ballistics analysis system (ALIAS). *AFTE Journal*, 43:74–8, 2011.

[4] R. Bolton-King, P. Evans, C. Smith, J. Painter, D. Allsop, and W. Cranton. What are the prospects of 3D profiling systems applied to firearms and toolmark identification. *AFTE Journal*, 42:23–33, 2010.

[5] D. Cork, J. Rolph, E. Meleran, and C. Petrie. *Ballistic Imaging*. Committee to Assess the Feasibility, Accuracy and Technical Capability of a National Ballistics Database, National Research Council. National Academies Press, 2008.

[6] National Research Council, editor. *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, 2009.

[7] J. DeCeuster, R. Hermsen, M. Mastaglio, and R. Nennstiel. A discussion on the usefulness of a shared European ballistic image database. *Science and Justice*, 52:237–42, 2012.

[8] J. DeKinder, F. Tulleners, and H. Thiebaut. Reference ballistic imaging database performance. *Forensic Science International*, 140:207–15, 2004.

[9] P. DeRuvo, A. Distante, E. Stella, and F. Marino. A GPU-Based vision system for real time detection of fastening elements in railway inspection. *16th IEEE International Conference on Image Processing*, pages 2333–6, 2009.

[10] 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.

[11] 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.

[12] K. Lillywhite, L. Dah-Jye, and Z. Dong. Real-time human detection using histograms of oriented gradients on a GPU. *Workshop on Applications of Computer Vision*, pages 1–6, 2009.

[13] T. Vorburger. Optical methods of surface measurement. Presented at the NIST Measurement Science and Standards in Forensic Firearms Analysis Conference, 2012.