Forensic Science International 271 (2017) 98–106

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests

Daniel Ott*, Robert Thompson, Junfeng Song

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:
Received 16 September 2016
Received in revised form 6 December 2016
Accepted 13 December 2016
Available online 21 December 2016

Keywords:
Firearms identification
Pattern recognition
Proficiency testing
Congruent matching cells

ABSTRACT

In order for a crime laboratory to assess a firearms examiner's training, skills, experience, and aptitude, it is necessary for the examiner to participate in proficiency testing. As computer algorithms for comparisons of pattern evidence become more prevalent, it is of interest to test algorithm performance as well, using these same proficiency examinations. This article demonstrates the use of the Congruent Matching Cell (CMC) algorithm to compare 3D topography measurements of breech face impressions and firing pin impressions from a previously distributed firearms proficiency test. In addition, the algorithm is used to analyze the distribution of many comparisons from a collection of cartridge cases used to construct another recent set of proficiency tests. These results are provided along with visualizations that help to relate the features used in optical comparisons by examiners to the features used by computer comparison algorithms.

Published by Elsevier Ireland Ltd.

1. Introduction

1.1. Background

The practice of tool mark comparison has a long history in the criminal justice system [1]. The goal of these examinations is to answer the question of common origin between a mark found at a crime scene (questioned mark) and a mark known to be produced by a suspected tool (reference mark). Such comparisons are a part of the broader field of pattern recognition and comparison which includes fingerprint, shoe print, tire tread, tool marks, and firearms identification. Until recently most of these disciplines have been practiced by conducting visual inspections (with the aid of a comparison microscope for firearms and tool mark examiners), with final determinations of common source being determined by a trained examiner. This practice has come under criticism in the United States, notably in a US National Academies report [2] due to the subjective nature of the examiner conclusions, the lack of an objective measurement of similarity, and the lack of a quantifiable error rate for these comparisons. The approaches to addressing the issues raised by the report [2] generally fall into two categories. The first is to develop measurement systems and computer algorithms that are able to quantify the degree of similarity between tool marks [3–7]. Using these systems and an appropriately selected, representative database of firearms evidence, it is possible to estimate error rates for comparisons and obtain consistent, objective similarity scores. The second approach is to attempt to quantify the ability of trained examiners. The advantage of this approach is that trained examiners are already integrated into the criminal justice system and these studies merely quantify the expected error rates in work that is already being performed. In some instances, examiners use codified language to express similarity in an attempt to achieve consistent results across practitioners [8]. Other approaches utilize standardized data sets to compare examiner's competency and, in some instances, provide estimates of error rates [9]. The purpose of this study is to apply computer algorithms to the test sets used to understand examiner proficiency and build a connection between the two approaches to address the recommendations of the National Academies for forensic science.[1]

1.2. Proficiency test sets

The focus of this article is on two firearms examiner proficiency

* Corresponding author.
E-mail addresses: daniel.ott@nist.gov (D. Ott), robert.m.thompson@nist.gov (R. Thompson), jun-feng.song@nist.gov (J. Song).

http://dx.doi.org/10.1016/j.forsciint.2016.12.014
0379-0738/ Published by Elsevier Ireland Ltd.

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

tests which utilize sample sets of fired cartridge cases. Over the years, several proficiency tests exist of this type, each with different test procedures and different methods for ensuring distribution of consistent tests to each participant [10]. The sets of cartridge cases can be produced by the test manufacturer by either multiple firings from known firearms or by replicating a single set of cases many times. The proficiency tests used in this study are from the Collaborative Testing Services firearms examination 526 test which uses many firings from, in this case, three firearms. The test kits are assembled from the firings such that each participant receives the same number of questioned and known reference firings from the same firearms, although each individual test will have some differences based on the consistency of the markings produced by the firearms used to create the test samples. Cartridge cases from both the 2010 (CTS 10-526) and 2015 (CTS 15-526) tests are used. The results from all participants of the test are available online [11]. Both tests include three reference cartridge cases fired from a mock suspect's firearm and four questioned cartridge cases from the mock scene of the crime. The objective of this test is to determine which, if any, of the questioned cases were fired from the suspect's firearm. In both instances, the test was constructed so that there was one questioned case which matched the suspect's firearm, two other questioned cases fired by a second firearm which matched each other, and a final questioned case which was fired from a third firearm. It is important to note that the published results include responses obtained from anyone that took the test which may include individuals who are not fully trained, certified, or otherwise expert examiners. Therefore, these results should not be considered representative of a global error rate for laboratories or examiners performing casework.

The first data set, is a single proficiency test from CTS 10-526 containing a set of seven fired 180 grain Federal® American Eagle® .40 S&W FMJ cartridge cases. The set contains three known reference cartridge cases fired from a .40 S&W caliber Smith & Wesson Springfield Armory XD40 handgun and four unknown casings. One unknown is fired from the same XD40 as the known cartridge cases. Two unknown cartridge cases were fired from a .40 S&W caliber Smith & Wesson Springfield Armory XD compact handgun. One unknown cartridge case was fired from a .40 S&W caliber Sig Sauer P226 handgun. Two of the known matching cartridge cases fired from the XD40 will be used as visual examples in the following sections to explain the comparison algorithms. After presenting the details of the algorithm, each pairwise combination of cartridge cases from this data set will be compared to demonstrate the ability of the algorithm to pass a single proficiency test in a manner analogous to how an examiner is tested.

The first data set provides a straightforward look at how a comparison algorithm would manage a single proficiency test, but it only allows a very small overall number of comparisons to be made. For a rigorous validation and understanding variability of scores, it is necessary to repeat similar tests many times. In the original proficiency test for firearms examiners, this was achieved by having many individuals take nearly the same test. For the computer algorithm presented here, a similar large dataset was necessary. Numerous samples that went unused by CTS during the administration of the CTS 15-526 proficiency test were used to build this second, larger dataset. The collection contains 44 firings fired by a .40 S&W caliber Ruger P94DC used to create the known samples and the matching, questioned samples. This will be referred to as Firearm 1. There were also firings from the other questioned firearms which will be referred to as Firearm 2 which had 18 firings and Firearm 3 which had 12 firings. The brand and model of these firearms is .40 S&W caliber Ruger P91DC for Firearm 2 and .40 S&W caliber Smith & Wesson SW40VE for Firearm 3. This particular proficiency test was constructed such that each examiner received 3 known firings from Firearm 1, one questioned firing from Firearm 1, two questioned firings from Firearm 2 and one questioned firing from Firearm 3. The ammunition used for all firings was 180 grain Federal® American Eagle® .40 S&W FMJ. Analysis of this large collection of samples will demonstrate the consistency with which the algorithm can pass a common proficiency test as well as provide insight into the variability of firings used in creating the proficiency test.

1.3. 2D Measurements with a comparison microscope

The traditional method for completing a proficiency test, as it is with evidence in an examination for a criminal case, is to use a comparison microscope to observe two pieces of evidence simultaneously. An optical bridge is used to combine the two microscope paths into the eyepiece for viewing. The operator is able to control the relative position and rotation of the two objects under observation as well as the comparison edge between images, that is, the location where one image transitions to the next. An example of a comparison microscope image is shown in Fig. 1. On the left there is a comparison of a breech face impression and on the right there is a comparison of a firing pin impression. The two images were obtained from the same pair of known source samples from the CTS 10-526 proficiency test. These cartridge case samples were fired from the Smith & Wesson Springfield Armory XD40 handgun described above. The samples are illuminated obliquely using a fluorescent light bar positioned at the top of the images. These magnifications and lighting conditions are indicative of what an examiner might observe while comparing cartridge case evidence.

The operator of the comparison microscope is then able to adjust the location where the transition from one objective to the other objective occurs. To demonstrate this concept, images of each breech face impression and firing pin impression were captured using a CCD camera integrated into the comparison microscope. Consistent illumination between samples was achieved by using a ring of white light LEDs around the microscope objective. Magnification of 20× was used to capture the breech face impressions and 40× was used to capture the firing pin impressions. The low 20× magnification for the breech face impression is typically used in the initial stages of manual examinations for coarse alignment and was employed here so that the entire image could be captured in a single field of view. The images were converted to grayscale and the breech face impression and firing pin impression were manually cropped. The images were then aligned automatically in position and rotation such that the

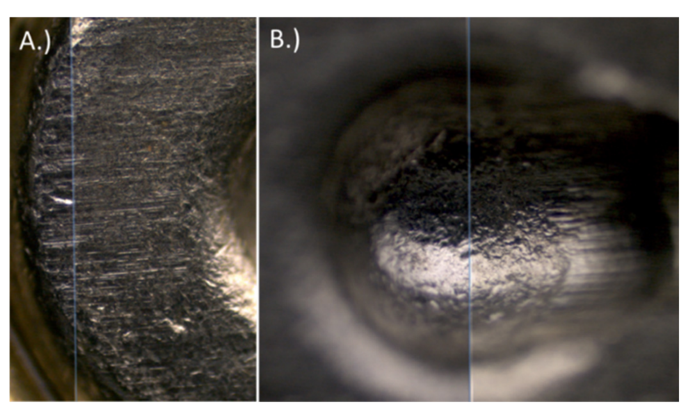

Fig. 1. Comparison microscope images using oblique bar lighting between two known matching cartridge cases. A.) Breech face comparison B.) Firing pin comparison. The transparent blue line separates the image of the first known firing on the left from image of the second known firing on the right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

areal cross correlation function (ACCF) between each image was maximized. The ACCF is the cross correlation function computed over an area as opposed to a 2D profile. The value of the ACCF for a given alignment indicates the similarity over the entire area of one image or topography to another. The ACCF is calculated using Eq. (1) where A and B represent a two dimensional matrix of points (either intensity values or surface heights) with dimensions $M \times N$, i and j are the pixel indices, μ is the mean value of the matrix, and σ is the standard deviation of the matrix.

$$ACCF = \frac{1}{MN}\frac{\sum_{j=1}^{N}\sum_{i=1}^{M}[(A_{ij}-\mu_A)\cdot(B_{ij}-\mu_B)]}{\sigma_A\sigma_B} \quad (1)$$

With the two compared images oriented in the position of greatest similarity as defined by the ACCF, a video can be constructed which shows the transition from one image to another in a similar way that an examiner would align surfaces and compare features across the surface by transitioning from one objective to the other using a comparison microscope. The same cartridge cases from Fig. 1 are shown again in Fig. 2 for breech face impressions and firing pin impression. A video showing the transition is provided in each figure.

1.4. 3D Measurements with a confocal microscope

The greyscale images in Figs. 1 and 2 depict the intensity of light reflected off the cartridge case surface. The image obtained will therefore depend on the illumination arrangement and does not directly relate to the actual surface height [12]. To provide consistent measurements that are directly related to the surface topography, it is necessary to utilize a so-called 3D measurement system. That is, a measurement system that is able to determine the height across the entire surface using a SI traceable unit of measure. Several techniques exist to achieve 3D measurements such as: stylus profilometry, interferometry, focus variation, confocal microscopy, and scanning electron microscopy. A review of the various techniques along with advantages and disadvantages of each system is given in [13]. Amongst these techniques is the disk scanning confocal microscope. The advantages of this system are vertical resolution in the range of several nanometers and horizontal resolution of several micrometers which is well suited for measuring features relevant to tool mark examination. The primary disadvantage of confocal microscopy is that the signal can become unreliable for highly sloped surfaces which can limit the surface area collected in the firing pin impression. This was the system chosen to measure the cartridge cases from the proficiency test. Regardless of the measurement system used, topographical surface maps are extremely useful for computer comparison algorithms as the data is collected using a consistent and repeatable unit of measure that is directly related to the features present on the surface.

All of the proficiency test cartridge cases were measured using a Nanofocus μSurf confocal microscope. For the breech face impressions, a 10× microscope objective was used with a nominal lateral pixel spacing of 3.1 μm. A total of 9 fields of view were stitched together (250 μm overlap) to generate a surface map of the impression. The firing pin impressions were measured independently using a 20× objective with nominal lateral pixel spacing of 1.6 μm and a single field of view was sufficient to capture the impression. Firing pin and breech face impressions were measured separately since the depth range for each impression varies significantly. In general, there is no need to measure the two impressions simultaneously because the marks originate from two distinct processes during the firing sequence involving two different parts of the firearm. Therefore, the relative orientation of the impression marks is not guaranteed and ultimately they are analyzed independently.

After measurement of the samples, the topographical maps of the surfaces were processed by trimming, removing outliers, and filtering to remove the low spatial frequency form and the high spatial frequency noise from the topography map. The Gaussian band pass filter used for the breech face impressions had cutoff wavelengths of 10 μm and 200 μm and the firing pin impressions were filtered using cutoff wavelengths of 1.6 μm and 110 μm. The high pass filter used for the firing pin impressions is a spline filter with modified boundary conditions intended to remove edge effects that occur at the perimeters of a domain [14].

With the processed topographical surface maps, it is possible to align the maps in position and rotation such that the ACCF is maximized and show transitions between surface maps that are analogous to Fig. 2. Fig. 3 shows the transition between the surface maps of the breech face and firing pin impressions previously analyzed in Fig. 2. Associated videos are provided in the figure.

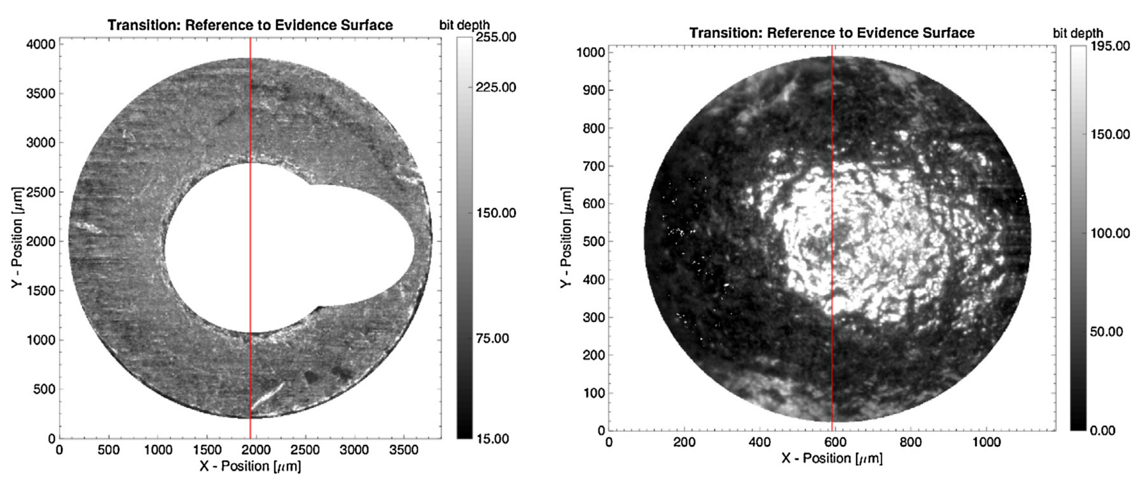

Fig. 2. A transition between digitized and processed images of the breech face (left) and firing pin (right) impressions. These are the same two known matching cartridge cases from Fig. 1. (Video available online).

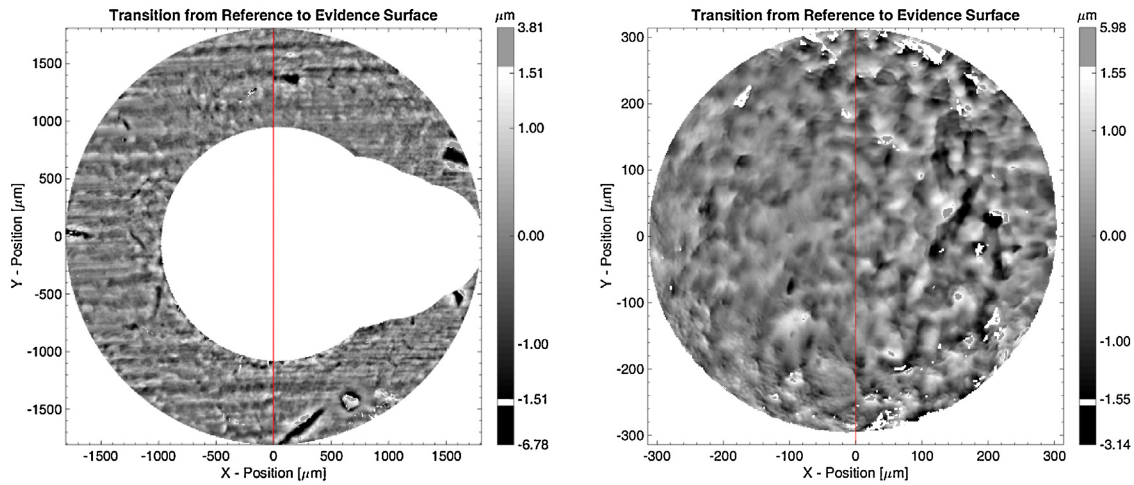

Fig. 3. A transition between digitized and processed topographical maps of the surface of the breech face impressions (left) and firing pin impressions (right). These are the same two known matching cartridge cases from Fig. 1. (Video available online).

The objective of this research is to draw a connection between the current methods employed by firearms examiners, which are illustrated in Fig. 1, and the analysis that a computer algorithm would apply to the data presented in Fig. 3. While the results do not directly compare the aptitude of examiners to computer algorithms, the use of proficiency tests provides a common ground for validating objective comparison algorithms. The key goal is to demonstrate that the algorithm discussed below is able to consistently reach correct conclusions when given comparisons of known ground truth. The analysis methods will be presented along with visuals that relate features an examiner uses to the regions used in objective similarity calculations which quantify the level of agreement between pieces of evidence. These connections will be useful tool to aid examiners in verifying and explaining their conclusions as well as for understanding and confirming similarity metrics.

2. Methods

2.1. Identifying regions of visual similarity

To a trained examiner, the aligned images in Fig. 3 may show clear signs of similarity and the ACCF value provides an independent objective similarity metric but there is still no clear connection between these comparison methods. By deconstructing the ACCF calculation in Eq. (1) it is possible to visually demonstrate the regions in two surface maps that contribute to the similarity score in the ACCF value. This can be thought of, intuitively, as analogous to an examiner mentally highlighting the regions of similarity in the transition videos of Fig. 3. The calculation is achieved simply by removing the averaging operation being performed in Eq. (1) such that the individual contribution of each pixel (ij) to the ACCF is retained as shown in Eq. (2). This calculation generates a map indicating the magnitude and sign of the similarity at that location. More simply, this map is a pointwise multiplication of two surface maps so that two aligned peaks (or valleys) on the surfaces will generate high positive similarity and a peak on one surface aligned with a valley on the second surface will generate a large negative similarity, i.e., a dissimilarity. The compared maps are centered and normalized before pointwise multiplication so that in featureless locations (points where either surface is at the mean height) there is neither a contribution of similarity or dissimilarity. The ACCF can be obtained by calculating the mean of the similarity map. Therefore, this map is a direct representation of the pointwise contribution to the overall similarity of the aligned surface maps.

$$Similarity\ \ Map_{ij} = \frac{A_{ij} - \mu_A}{\sigma_A} \cdot \frac{B_{ij} - \mu_B}{\sigma_B} \tag{2}$$

An example similarity map calculated using the known matching surface maps from the samples used throughout the previous figures is shown in Fig. 4. This figure captures the similarity that is qualitatively observed in the videos of Fig. 3 using a single image. Calculating the average of the similarity map will give the ACCF of the two surfaces used to construct the similarity map. In this case the ACCF for the breech face impression is 23% and the ACCF for the firing pin impression is 49%. To further illustrate the regions of similarity, a threshold is used to determine the most similar regions between the two surfaces. A simple threshold was defined for this example such that all areas of the similarity map that exceed 0.5σ (where σ is the standard deviation of the similarity map) from the mean pixel similarity value are considered 'highly similar,' whereas all of the areas that are more than 0.5σ below the mean are considered 'highly dissimilar.' Furthermore, the direction of the surface maps, above or below the mean is noted in the highly similar areas so that it is possible to distinguish similar peaks from similar valleys. These regions are then highlighted in the right half of Fig. 4. The highlighted similarity maps show regions which do not contribute strongly to matches as grey, similar peaks are highlighted red and similar valleys are highlighted green, and the dissimilar regions (overlap of a peak and valley) are highlighted yellow. This method for identifying the prominent similarity features of a surface is useful for human examiners to understand the regions which contribute to similarity metrics such as ACCF.

2.2. Comparisons using the CMC method

Many other metrics for quantifying pattern similarity for firearms evidence exists [4–7]. At the National Institute of Standards and Technology (NIST) a method called Congruent Matching Cells (CMC) has been developed [3]. This algorithm relies on dividing a topography from a reference image into a grid of cells. Each cell is independently compared to the entire evidence surface

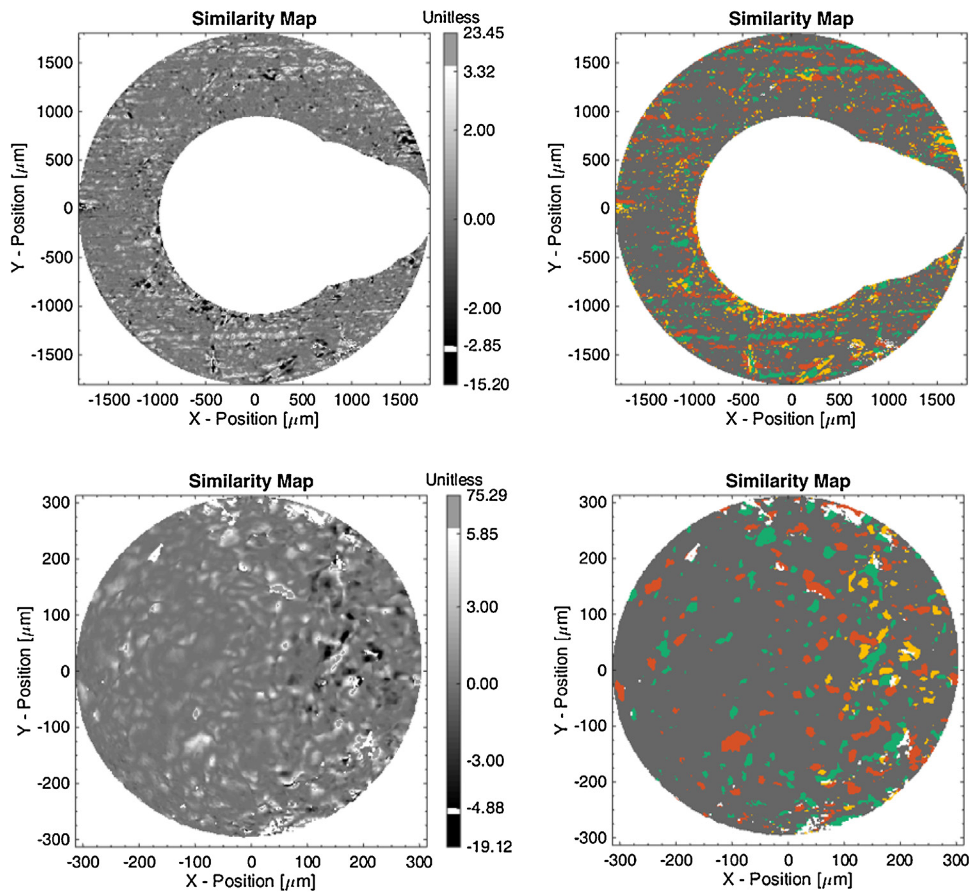

Fig. 4. The similarity map for the aligned 3D topographical surface maps from the breech face impressions (top) and the firing pin impressions (bottom). The left figures show a grayscale representation with white indicating highly similar areas and black highly dissimilar areas. The figures on the right are highlighted based on a threshold which identifies the most prominent similar peaks (red), prominent similar valleys (green), highly dissimilar regions (yellow), and regions which do not contribute significantly to the ACCF (gray). These similarity maps correspond to the known matching cartridge cases shown in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to determine the spatial transformation that produces maximum ACCF between the cell and the evidence surface. This is repeated for each cell and then the registration locations of the cells are analyzed to find collections of cells that required the same relative rotation and translation (within some tolerances) from reference surface to evidence surface. Each collection of cells is congruent between the compared surfaces. The collection containing the largest number of cells is indicative of the maximum achievable similarity between two surfaces and so this number serves as the similarity metric in this algorithm. While the algorithm still uses ACCF in the calculation of similarity, it is able to independently analyze discrete regions on the surface. This allows the algorithm to effectively ignore dissimilar regions between two same source marks where, perhaps, the tool did not impress well or the surface is otherwise damaged in a way unrelated to the tool mark. In contrast, the ACCF calculation between entire surfaces is significantly affected by these dissimilar regions even if the rest of the surface matches very well. The CMC method has been successfully applied in various tests [15,16] and produced clear separation between the scores of known matching (KM) and known non-matching (KNM) samples making it a promising algorithm for ballistic evidence comparisons. The surfaces from Fig. 1 have been compared using the CMC method and the results are shown in Fig. 5. This figure demonstrates the regions where the congruent cells were defined on the first surface and the corresponding region where high physical similarity was found on the second surface.

The regions bounded by the CMCs in Fig. 5 indicate areas that are considered similar according the CMC method. While the localization (cell size) of the similar areas is quite coarse, this can be compared to the highlighted similarity maps in Fig. 4. These highlighted areas indicate the highly similar regions according to the ACCF calculation and also generally correspond to the visually perceived similarity observed in the transition videos of Fig. 3. The third column of Fig. 5 shows the CMCs overlaid on the similarity map. It is important to note the differences between the similarity map and the locations counted by the CMC method. This is due to the discretization used in the CMC algorithm where each cell from the reference topography searches the evidence topography independently. So while the similarity map may show high similarity in a region, the corresponding cell may find a location of even greater physical similarity in a location that is not congruent. Such regions might include very common surface patterns that are not unique in and of themselves. It is informative to understand both methods of visualization when analyzing how two surfaces are similar.

The analysis methods presented in this section were provided with consistent examples in order to explain and compare various methods for assessing similarity. These tools for visualizing and quantifying similarity are useful in developing an understanding of objective similarity but may also find use amongst examiners in practice. The following section will address the validity of these methods by demonstrating the application of the CMC method to numerous firearms impression comparisons from the remainder of the CTS test samples. Along with previous validation studies and on-going validation research, these results will further elucidate

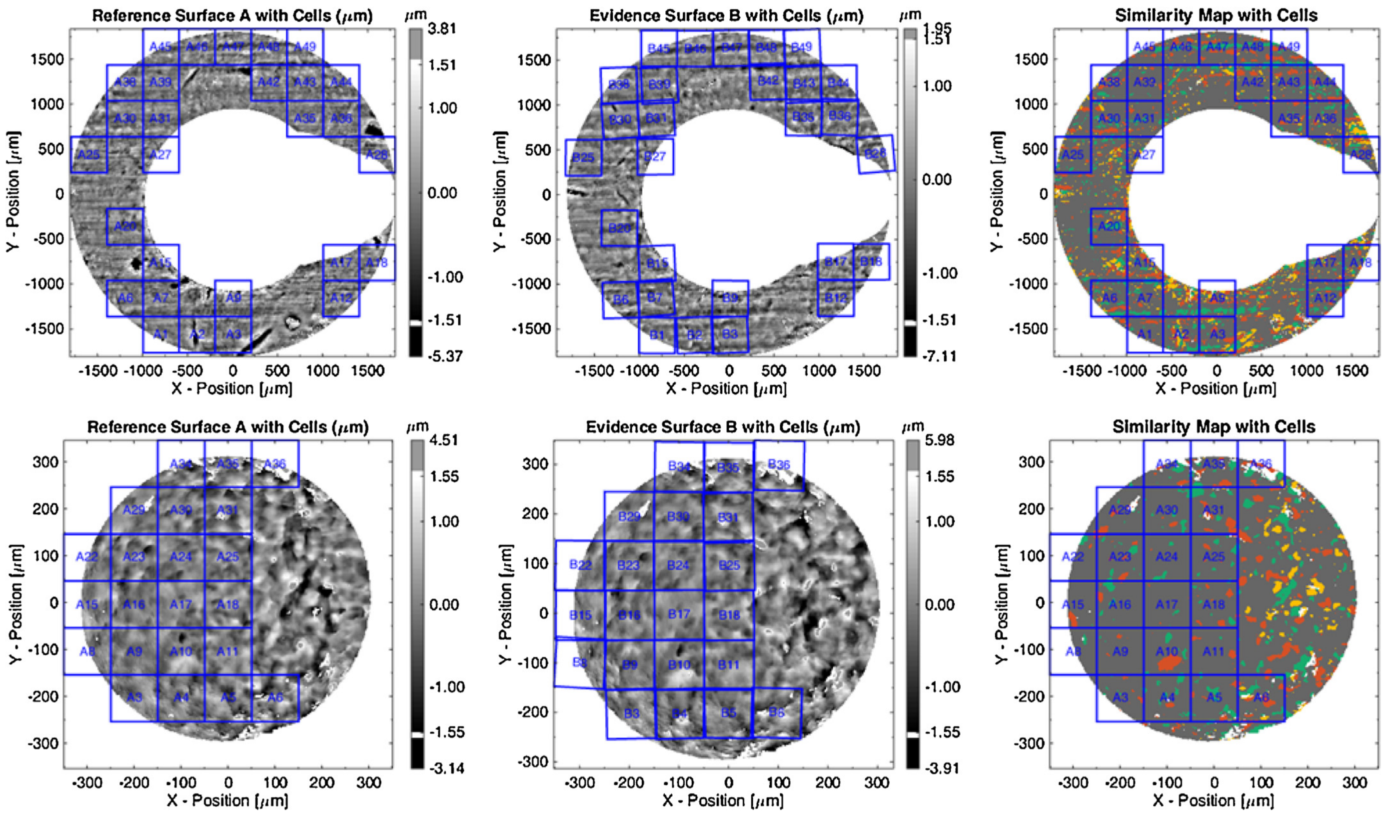

Fig. 5. The CMC method applied to the known matching breech face impressions (top) and firing pin impressions (bottom). On the left is the grid of cells defined on the reference surface. Only the cells which were determined to be congruently matching are shown. In the middle is the registration location for each of the cells on the evidence surface. At the right is the similarity map between the two surfaces with the CMCs locations overlaid and threshold applied.

the efficacy of the CMC method in the context of established proficiency tests.

3. Results

3.1. Evaluating a single proficiency test (CTS 10-526)

The CMC method was applied to the seven samples from the CTS 10-526 proficiency test. Every pairwise combination of cartridge case comparison was evaluated using the CMC method. The seven correct identifications and fourteen exclusions were correctly concluded for each of the 21 possible comparisons using the CMC method on both the breech face and firing pin impressions. This single proficiency test examination serves as a proof of concept to show that this computer algorithm is able to pass a proficiency test in much the same way that a single examiner would assess their proficiency using a single test. In and of itself, this result is limited without the context provided by many trials or many other participants in the test. The published CTS test results provide some context in that the algorithm would have been amongst the 95% of participants that reached correct conclusions for this test [11].

3.2. Distribution of CMC scores (CTS 15-526)

To provide a more appropriate context for the CMC method passing the single CTS 10-526 proficiency test, it is necessary to validate the algorithm using a much larger but similarly structured set of data. This validation was achieved by analyzing the collection of samples from CTS 15-526 as an entire data set by applying the CMC method to pairwise combinations of the cartridge cases. For purposes of this section, the firings from each firearm were analyzed as a collection of samples rather than by constructing numerous, separate proficiency tests for individual analysis. This allows us to make conclusions regarding the ability of the CMC method to pass any proficiency test randomly constructed from this particular set of cartridge cases. Same source score distributions were created by using the CMC method to determine similarity between different firings from the same firearm. These distributions will illustrate the relative difficulty of comparisons with each firearm and quantify uncertainty in the comparison scores from the perspective of the CMC algorithm. The total number of same source comparisons was 946 for Firearm 1, 153 for Firearm 2, and 66 for Firearm 3.

Different source score distributions were also analyzed to demonstrate the magnitude of similarity (or dissimilarity) between the known non-matching samples in this test. This set of samples can be used to generate three different source distributions using comparisons between Firearm 1 and Firearm 2, Firearm 1 and Firearm 3, and Firearm 2 and Firearm 3. A total of 1536 different source comparisons were conducted. The comparison scores using the CMC method on the breech face impressions are shown in the top row of Fig. 6 and for the firing pin impressions in the bottom row of Fig. 6. Each row contains several plots which detail the comparisons specific to a particular firearm.

4. Analysis

For the breech face impression comparisons in Fig. 6, it is clear that Firearm 1 and Firearm 2 are extremely difficult for the CMC algorithm to identify correctly. Further investigation into the cause of the low scores for these firearms shows that the nature of the marks is inconsistent between firings, consisting of large regions

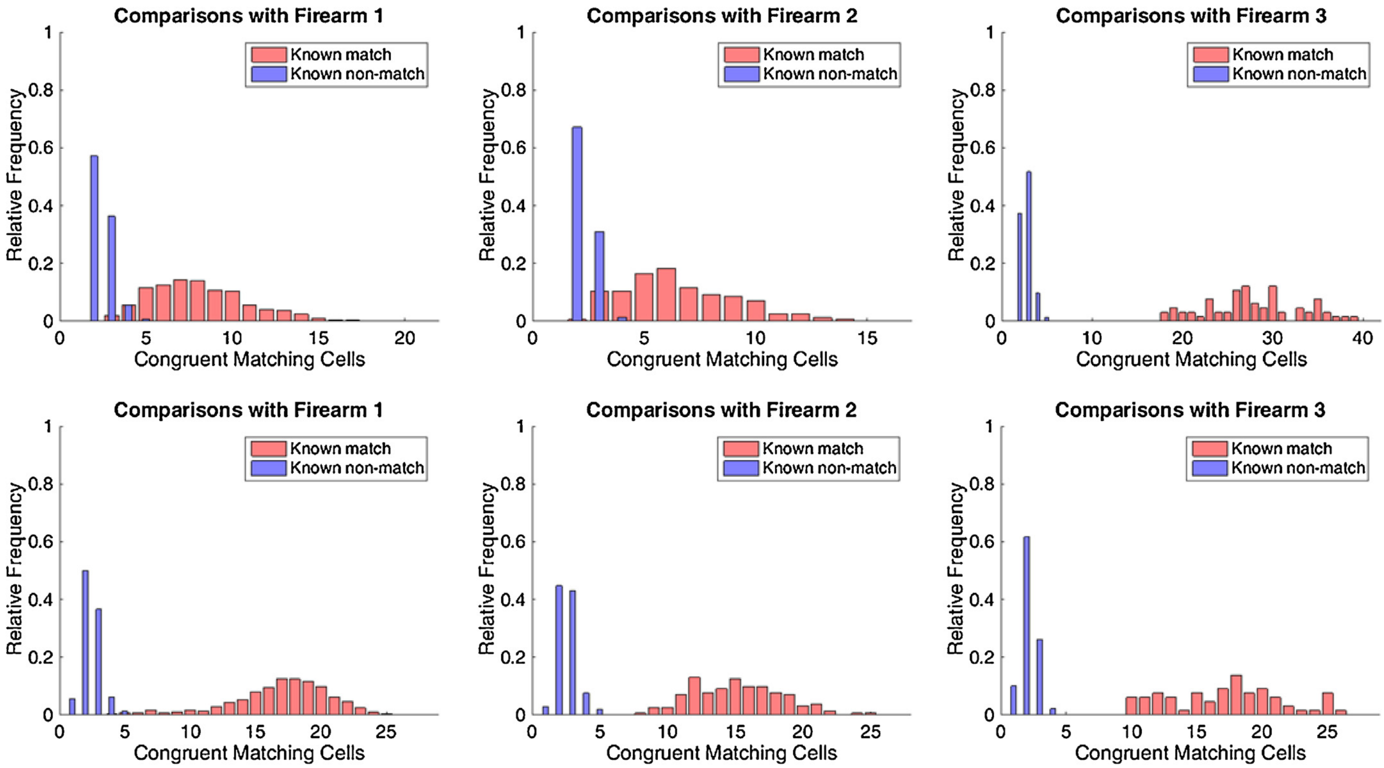

Fig. 6. The distribution of breech face impression CMC scores (top row) and firing pin impression CMC scores (bottom row) from the collection of cartridge cases fired from the three different firearms used to construct the CTS 15-526 proficiency test.

where poor contact was made with the breech face. When a comparison involved two such samples, the union of the poorly marked area could be sufficient to generate low CMC scores.

For the comparisons involving the firing pin impressions, 12 false negative conclusions were obtained out of the 946 known matching comparisons. All of the false negative conclusions are associated with a single firing pin impression which contained a large anomalous region through the center, possibly associated with a foreign particle present on the firing pin during impact with the primer. This reduced the effective number of cells for comparisons with this impression as well as obscuring some of the most reproducible features for this firing pin impression. The remaining 31 comparisons involving this impression were identified correctly. This highlights one limitation of the CMC method in dealing with large dissimilar regions in impression evidence. A more detailed histogram for Firearm 1 is shown in Fig. 7. This figure shows all comparisons using Firearm 1 but displays separate, normalized histograms for any comparison involving the damaged surface.

Using just the breech face or just the firing pin impressions with the CMC method would not yield perfect performance for any randomly selected proficiency test constructed from these samples. However, combining the scores from the two impressions can yield better separation of the distributions. In Fig. 8 the breech face CMC score is added to the firing pin CMC score for every comparison and the new combined score histograms are displayed for each firearm. This combined score leverages the independent nature of the two impressions to better discriminate the non-matches from the matches for these particular firearms. The mean and standard deviation for each of the distributions shown in Figs. 6 and 8 is compiled in Table 1. Using the combined CMC scores from the breech face and firing pin impression comparisons it is possible for the CMC method to pass any randomly selected proficiency test that could be generated from the CTS

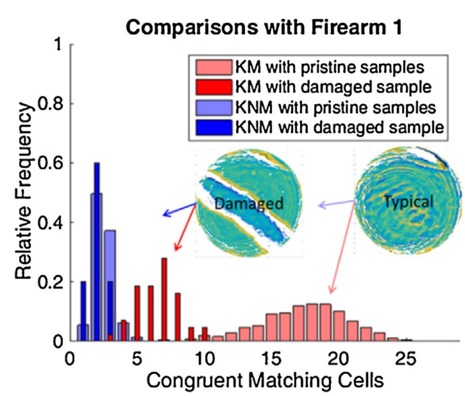

Fig. 7. A detailed histogram of the firing pin impressions from Firearm 1. The sample that had additional tool marks of unknown origin is displayed in a separate normalized histogram in bold colors with narrow bars. All other comparisons with samples from this firearm are shown in the histograms with wide bars and muted colors. The inlayed images demonstrate the filtered firing pin topography for the damaged surface as well as an example of a typical firing pin topography for this firearm.

15-526 samples that were measured. Therefore, for any test constructed from this data, the conclusions reached by the CMC method would be among the 99% of participants that reached correct conclusions for this test [11].

5. Conclusions

In conclusion, a number of visualization methods have been presented for comparing cartridge case firearms evidence. Using videos that show a transition from one image to another aligned

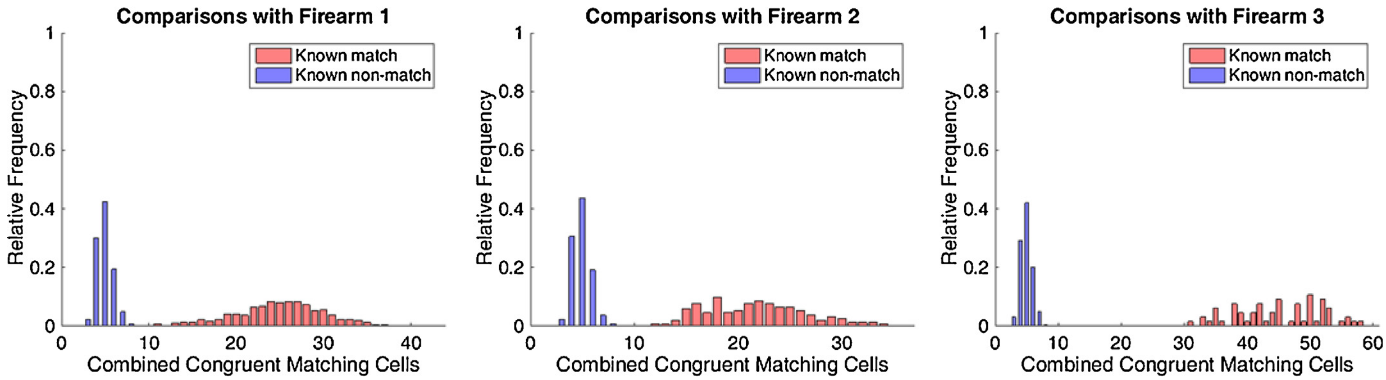

Fig. 8. The distribution of the combined CMC scores from the collection of cartridge cases fired by three different firearms used to construct the CTS 15-526 proficiency test. The combined score consists of the addition of the breech face impression score and the firing pin impression score.

Table 1
The moments (means and standard deviations, σ) of the distributions of CMC scores for the firearms used in the 2015 test.

Score type	Firearm 1 KM		Firearm 1 KNM		Firearm 2 KM		Firearm 2 KNM		Firearm 3 KM		Firearm 3 KNM	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Breech face	***8.1***	2.8	**2.5**	0.6	***6.6***	2.5	**2.3**	0.5	***27.8***	5.2	**2.8**	0.7
Firing pin	***16.9***	3.9	**2.5**	0.8	***15.2***	3.3	**2.6**	0.7	***17.2***	4.4	**2.2**	0.6
Combined	***24.9***	5.2	**5.0**	0.9	***21.7***	4.8	**4.9**	0.9	***45.1***	6.8	**5.0**	0.9

image, it is possible to display the qualitative similarity that firearms examiners are using to make determinations of similarity. This method can be extended to similar visualizations using measured topographical surface maps such that a qualitative comparison is also possible with 3D measurement tools. Using these aligned surface maps, it was possible to relate qualitative similarity to a quantifiable similarity metric (the ACCF) using a similarity map. Such maps may be useful in practice by highlighting regions of interest for further analysis by examiners, or by numerically supporting conclusions made independently by examiners. Additionally, the visualization methods can add to the crime laboratory's quality program when examination files are subjected to peer and administrative review. These maps were then used as a tool for relating visual comparisons to the results obtained by the CMC method. While the comparisons methods are based on fundamentally different algorithms, it is possible to explain certain conclusions made by the CMC method in a more intuitive fashion. Finally, the variability in a firearms proficiency test was examined by using the CMC method to compare a collection of samples used to generate the CTS 15-526 firearms test. These comparisons proved challenging for the algorithm but ultimately a correct set of conclusions was reached by considering both firing pin and breech face impressions jointly. This demonstrates the potential utility of considering more than one tool mark on a cartridge case when making conclusions using automated algorithms.

The further analysis of low performing samples, such as the damaged area shown in Fig. 7, suggests a need for knowledgeable operators to interpret anomalous results. These results also indicate the strength of the CMC algorithm at identification of well-marked samples which generate high CMC scores. Since the known non-matching scores are consistently very low, a comparison with a high CMC score can be expected to indicate a high degree of similarity with low uncertainty, whereas Fig. 6 shows that non-matches and poorly marked samples produce overlapping score distributions, so low scoring comparisons have a higher degree of uncertainty in the resulting conclusions. To quantify these uncertainties, larger and more inclusive data sets must be studied but the initial findings fit well into the framework of an adversarial judicial system where false positive conclusions should be minimized. Although further validation and testing is necessary before implementing these comparison methodologies, the ability to relate visual comparisons to quantifiable metrics will help to support firearms examiners in their current procedures, ease the transition into new comparison technologies, and provide juries with easy to interpret comparison results backed by repeatable and systematic analysis.

Acknowledgments

The funding for this project is provided by the Forensic Measurement Challenge Program (FMC2012) and the Special Programs Office (SPO) of NIST. The authors are grateful to Richard Hockensmith at Collaborative Testing Services for providing test samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.forsciint.2016.12.014.

References

[1] Scientific Working Group for Firearm and Toolmarks (SWGGUN), The Foundations of Firearm and Toolmark Identification. http://www.swggun.org. (Retrieved 9 June 2016).
[2] The National Research Council, Strengthening Forensic Science in the United States—A Path Forward, NRC, Washington, D.C, 2009 pp. 153–154, 184, 155.
[3] J. Song, Proposed congruent matching cells (CMC) method for ballistic identification and error rate estimation, AFTE J. 47 (3) (2015) 177–185.
[4] T. Weller, M. Brubaker, P. Duez, R. Lilien, Introduction and initial evaluation of a novel three-dimensional imaging and analysis system for firearm forensics, AFTE J. 47 (4) (2015) 198–208.
[5] F. Riva, C. Champod, Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases, J. Forensic Sci. 59 (3) (2014) 637–647.
[6] U.M. Leloglu, Characterisation of tool marks on cartridge cases by combining multiple images, IET Image Process. 6 (7) (2012) 854.

[7] C. Gambino, P. McLaughlin, L. Kuo, F. Kammerman, P. Shenkin, P. Diaczuk, N. Petraco, J. Hamby, N.D.K. Petraco, Forensic surface metrology: tool mark evidence, Scanning 33 (5) (2011) 272–278.
[8] C. Champod, A. Biedermann, J. Vuille, S. Willis, J. De Kinder, ENFSI guideline for evaluative reporting in forensic science, a primer for legal practitioners, Crim. Law Justice Wkly. 180 (March (10)) (2016) 189–193.
[9] D.P. Baldwin, S.J. Bajic, M. Morris, D. Zamzow, A study of false-positive and false-negative error rates in cartridge case comparisons, Technical Report #IS-5207, (2014) .
[10] ASCLD/LAB, Approved Providers and Tests. http://www.ascld-lab.org/approved-providers-and-tests/. (Retrieved 9 June 2016).
[11] Collaborative Testing Services, Summary Reports. https://www.ctsforensics.com/reports/default.aspx?F_CategoryId=22&pg=2. (Retrieved 9 June 2016).
[12] J. Song, W. Chu, T.V. Vorburger, R. Thompson, T.B. Renegar, A. Zheng, J. Yen, R. Silver, M. Ols, Development of ballistics identification—from image comparison to topography measurement in surface metrology, Meas. Sci. Technol. 23 (5) (2012) 054010.
[13] T.V. Vorburger, J. Song, N. Petraco, Topography measurements and applications in ballistics and tool mark identifications, Surf. Topogr. Metrol. Prop. 4 (1) (2016) 013002.
[14] M. Tong, H. Zhang, D. Ott, X. Zhao, J. Song, Analysis of the boundary conditions of the spline filter, Meas. Sci. Technol. 26 (9) (2015) 095001.
[15] W. Chu, M. Tong, J. Song, Validation tests for the congruent matching cells (cmc) method using cartridge cases fired with consecutively manufactured pistol slides, AFTE J. 45 (4) (2013) 361–366.
[16] M. Tong, J. Song, W. Chu, R. Thompson, Fired cartridge case identification using optical images and the congruent matching cells (CMC) method, J. Res. NIST 119 (2014) 575–582.