Developing Methods for Comparison of Cartridge Breechface Images

Xiao Hui Tai
Advisors: William F. Eddy, Xiaoyu Alan Zheng

February 13, 2017

Abstract

When a gun is fired, it leaves marks on cartridge cases that are thought to be unique to the gun. In current practice, firearms examiners inspect two cartridge cases using a comparison microscope, and if there is "sufficient agreement," they conclude that they come from the same gun, and testify in courts as such. A recent publication (2016) by the Presidents Council of Advisors on Science and Technology questioned the scientific validity of such analysis [1]. One recommendation was to convert firearms analysis to an objective method, through the use of fully automated methods. We propose one such method for comparing breechface marks on cartridge cases using 2D optical images. We improve on existing methodology by automating the selection of marks, as well as removing the effects of circular symmetry. Lastly we propose a method of computing a "random match probability" empirically, given a known database.

Keywords: Forensic science, firearms identification, cartridge cases, circular symmetry

Contents

1 Background and Purpose

When a gun is fired, it leaves marks on cartridge cases that are thought to be unique to the gun. In Figure 1 notice that the bottom surface of the cartridge is in contact with the breech block of the gun and the firing pin. During the firing process the cartridge is hit by the firing pin, which causes it to break up into two components, the bullet which goes out the barrel, and the cartridge case that is ejected from the side. This process leaves two kinds of marks on the bottom surface of the cartridge case. The first is the firing pin impression caused by the firing pin hitting the cartridge, and the second are breechface marks that are caused by the bottom surface of the cartridge pressing against the breech block of the gun. The marks can be seen more clearly in Figure 2. Breechface marks are marks impressed on the primer of the cartridge by the breech block, which is made of harder material than the primer. Any microscopic patterns or imperfections on the breech block may hence be reproduced in the breechface impression, which is what is thought individualizes each gun (see e.g. [4]).

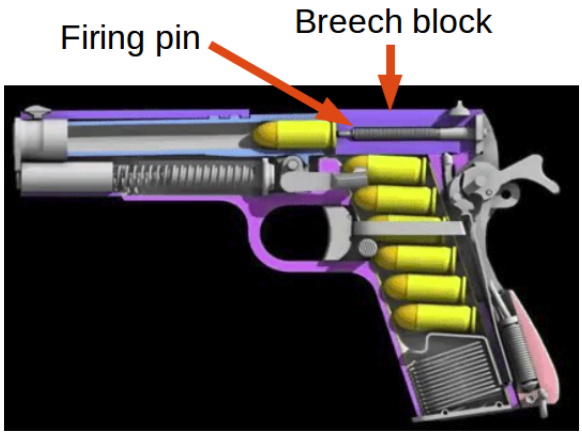

Figure 1: Gun that is about to be fired, showing the internal parts.

Law enforcement agencies routinely collect guns and cartridge cases from crime scenes, because of their potential usefulness in investigations (for example, if it is determined that cartridges found at separate crime scenes came from the same weapon, evidence can be combined, aiding investigations). In current practice, firearms examiners inspect two cartridge cases using a comparison microscope, and if there is "sufficient agreement" between the marks

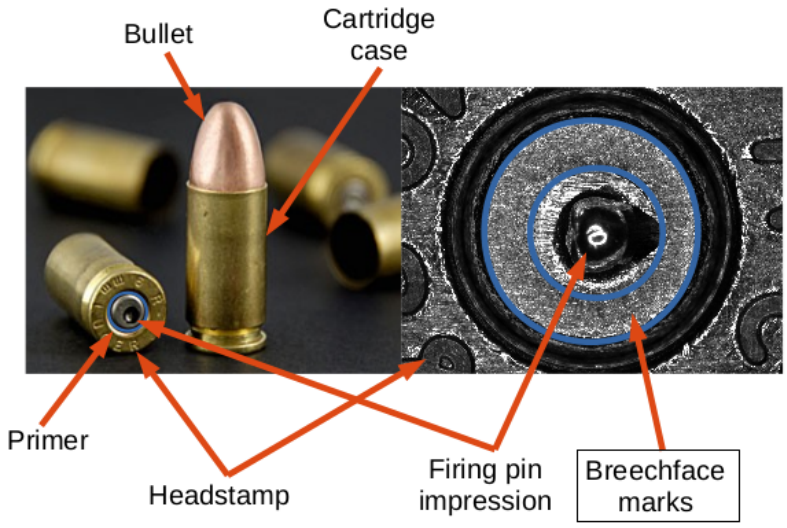

Figure 2: Physical cartridge and cartridge case on the left, and an image of the bottom of the cartridge case on the right, taken using a reflectance microscope.

on the casings, they conclude that they come from the same gun [3], and testify in courts as such. A recent publication by the President's Council of Advisors on Science and Technology [1] questioned the scientific validity of firearms analysis, making two recommendations for the path forward. The first is to continue improving firearms analysis as a subjective method, by conducting additional black-box studies to determine the reliability of firearms examiners' conclusions, as well as by introducing more rigorous proficiency testing. The second is to convert firearms analysis from a subjective method to an objective method, through the use of fully automated methods.

Here we primarily consider the second recommendation. There has been research by various groups with regards to this. In recent years, the National Institute of Standards and Technology (NIST) has been advocating the use of 3D topographies for measurement [9], coupled with algorithms that are largely automated [12, 8]. 3D topography measurements are in contrast to 2D optical (grayscale) images, the latter being older technology that has been in use by law enforcement agencies for over a decade. The comparison of physical cartridge cases using a comparison microscope (as described earlier) is often preceded by a search of the National Integrated

Ballistics Information Network (NIBIN), a national database of casings and test fires of guns from crime scenes. To perform such a search, examiners use a platform developed by Forensic Technology WAI, Inc. (FTI), of Montreal, Canada, which captures a 2D optical image of the "new" or "questioned" casing, and then returns a list of top ranked potential matches from NIBIN using a proprietary search algorithm. The examiner then uses this list and the corresponding images to determine which physical casings to focus their efforts on. In recent years FTI has developed a system capable of acquiring both 2D images and 3D topographies, and some laboratories have upgraded to this system,[1] but majority of images in local and national databases are still in 2D.

The main reason for NIST's recommendation to use 3D topographies is that optical images are sensitive to lighting conditions, whereas 3D topographies are a direct measurement of surface contours, which are reproducible and traceable to the SI unit of length. As part of a report by NIST in 2007 assessing the feasibility of developing a national ballistics imaging database [11], an algorithm for comparing 3D topographies of cartridge cases was developed. The accuracy of this method was compared to FTI's proprietary algorithm, concluding that 3D topographies produced more accurate results. Following NIST's lead, other groups have also developed algorithms for comparing 3D topographies (see e.g. [5]).

Despite these developments, there remains interest in 2D optical images, for the previously stated reason that majority of images in local and national databases remain in 2D. Recent work includes a paper by NIST applying their Congruent Matching Cells method, originally developed for 3D topographies, to 2D [10], as well as work by Roth et al. on images involving operational data sets taken from a police department, in addition to a data set of controlled test fires, as is usually used by researchers [6].

Here we extend the literature on 2D comparisons by improving on existing algorithms for comparing breechface marks, introducing a method that is both fully automated and open source.[2] The original comparison of performance of 2D with 3D [11] was made using FTI's proprietary algorithms on the 2D images, using two data sets (named the NBIDE and DeKinder studies). For both, 3D was an improvement over 2D, but the 3D results for

[1]For example, the Allegheny County Medical Examiner's Office installed the new system in 2016.

[2]The code is available at `https://github.com/xhtai/cartridges`.

DeKinder breechface marks remained relatively poor, while for the NBIDE data, methods using 3D were a marked improvement over 2D, and could almost perfectly distinguish matches from non-matches. Here we revisit the NBIDE data set to determine if the performance using 2D images remains as poor in comparison. To our knowledge, this data set has not been studied since.

There has also been ongoing research as to how to compute likelihood ratios from cartridge case comparisons, so that they may be incorporated with other evidence (see e.g. [7, 5]). Here we propose a method to compute "random match probabilities" empirically, given a known database. This method was developed for 2D comparisons but is also applicable to 3D.

2 Data and Methodology

2.1 Data

The casings used are the same ones collected and analyzed in NIST's 2007 study, and a detailed account of the experiment can be found in [11]. The casings were re-imaged by NIST at their Gaithersburg, Maryland campus, and the data were made available as part of the Ballistics Toolmark Research Database (`https://tsapps.nist.gov/NRBTD`), an open-access research database of fired bullet and cartridge reference data. Briefly, the casings were collected from test fires of 4 Ruger P95D, 4 Smith & Wesson 9VE and 4 Sig Sauer P226 pistols. Each was fired 3 times using each of 3 different brands of ammunition: PMC, Remington and Winchester. The total number of test fires was 108. Each of these casings were then imaged using a Leica FS M reflectance microscope with ring light illumination. The objective was 2X and the resolution was 2.53 μm, and images are 1944×2592 pixel grayscale files in png format. As in [11], we refer to these images as the NBIDE (NIST Ballistics Imaging Database Evaluation) data set.

To analyze these data, we will treat each image in turn as the new image, and the remaining 107 images as the known database. We perform 107 pairwise comparisons for each new image, for a total of 107*108 or 11,556 pairwise comparisons. For each pairwise comparison, we will compute a similarity score and the probability of obtaining a higher score by chance.

2.2 Steps for one pairwise comparison

The following are proposed steps for making one pairwise comparison. Steps 1 through 4 are pre-processing steps, 5 computes a similarity score and 6 a "random match probability."

1. Automatically select breechface marks by removing the non-primer areas and the firing pin impression
2. Level image
3. Remove circular symmetry
4. Outlier removal and filtering
5. Maximize correlation by translations and rotations
6. Compute probability of obtaining a higher score by chance.

These steps build on methodology published in [11] and implemented in [6]. There are three main improvements: steps 1, 3 and 6 above. Each step is explained in detail in the following subsections.

2.2.1 Automatically select breechface marks

The first step is to automatically select the breechface marks. In NIBIN ([11]) and other research ([10, 6]), this step is done manually by adjusting a larger circle to pick out the primer region, and a smaller circle for the firing pin impression. Here, we select the breechface marks automatically by first finding the primer region and then removing the firing pin impression. We do not constrain either region to be circular, and this is especially important for the firing pin impression, which could have different shapes depending on the make and model of the gun. A rough schema for finding the primer region is in Figure 3.

To remove the firing pin impression, we use use a similar set of steps, but we start with an edge detector to first identify the firing pin region. We use a Canny edge detector because it allows the specification of two thresholds, which enables weaker edges to be detected, if they are connected to a strong edge. This is very useful in detecting the firing pin impression, because not the entire border may be prominently marked. The steps are in Figure 4.

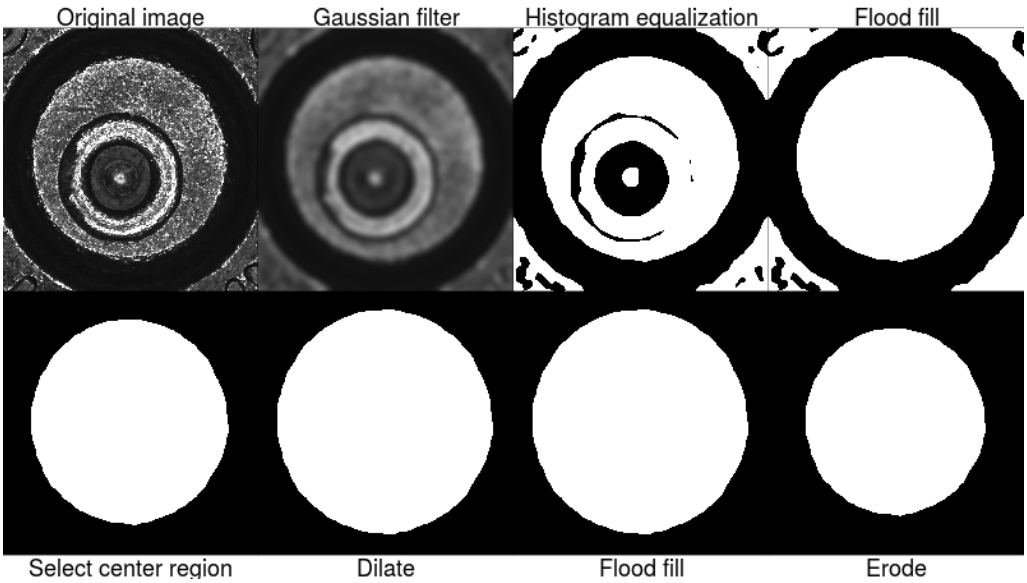

Figure 3: Series of steps to find the primer region for an example image. This image is from a Ruger gun, firing a PMC cartridge. In some images the firing pin impression is so close to the edge of the primer that the center region selected is not circular, and the last three steps (dilating, filling and eroding) morphologically closes the image.

Next we perform a second pass where we apply an edge detector again with slightly different parameters, to try to remove any remaining marks. This is necessary for some images where parts of the firing pin impression might not be as highly contrasted with the surrounding breechface impression, resulting in them being missed the first time. This different set of parameters picks up such edges that are connected to the previously identified firing pin impression. The steps are in Figure 5.

2.2.2 Level image

The second step is to level the image. As explained in [11] and [6] this step is necessary because the base of the cartridge case may not be level, and may instead be tilted slightly on a plane. As a result, images of such a surface may have differences in brightness that are planar in nature. We fit a plane that captures these differences, and then take the residuals, which ensures that the resulting image is free from planar differences in brightness. An

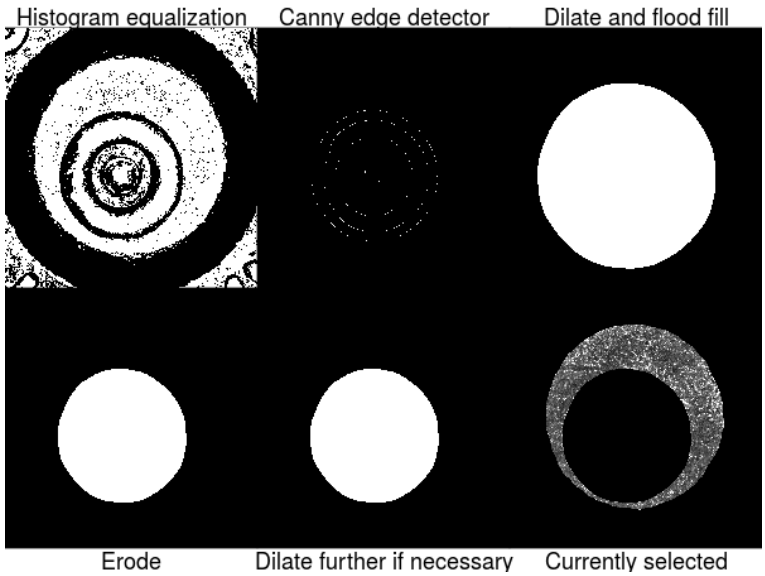

Figure 4: Series of steps to find the firing pin impression of the example image.

example is in Figure 6.

2.2.3 Remove circular symmetry

The next step is to remove the effects of circular symmetry. Analogous to step 2, the base of the cartridge case could have differences in depth that are circular in nature, for example the surface may slope inwards towards the center. This would cause corresponding differences in brightness such as the center of the image being darker than the edges. Like in the previous step, we fit a model that captures this circular symmetry, and then take the residuals. The residuals will be free from any circular symmetry.

The model that we fit is a linear combination of circularly symmetric basis functions [13]. Briefly, this model assumes that pixels located the same distance from the center of the image take the same value, and each basis function corresponds to a unique distance from the center. The coefficient for each basis function is the mean of pixel values for pixels with the corresponding distance from the center.

The details are as follows. We consider a square matrix of dimension $m \times m$, where m is odd. The center entry is $\left(\frac{m+1}{2}, \frac{m+1}{2}\right)$, and we say that

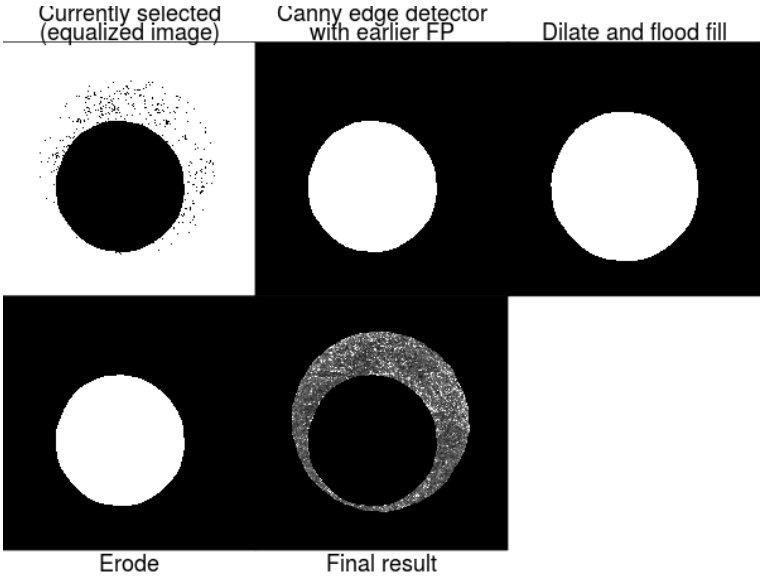

Figure 5: Here we run the edge detector a second time. In this particular example the entire firing pin impression has already been removed, so these steps do not produce any effect and the image remains unchanged.

such a matrix is circularly symmetric if entries located the same distance from the center entry take the same value. Since images matrices of pixel values, pixels located the same distance from the center of a circularly symmetric image take the same value. Such an image can then be decomposed into a linear combination of the matrices in a circularly symmetric basis. The first few matrices in the basis are given in Figure 7, where each image in the panel represents one matrix. Each matrix takes the value 1 for pixels that are the same distance from the center, and zero otherwise. Bases are enumerated from center outwards.

Now, to represent these matrices as basis functions, we define functions f_k for each matrix k, taking the ij-coordinates as inputs, and returning the value 0 or 1 depending on whether the input pixel is the required distance from the center, i.e. if it is white or black in the pictorial representation in Figure 7. An example of a basis function is in Equation (1). The decomposition of an image can then be represented as Equation (2), where K is the number of basis functions, f_k is the kth basis function, β_k is the basis function coefficient for f_k, and ϵ is the error.

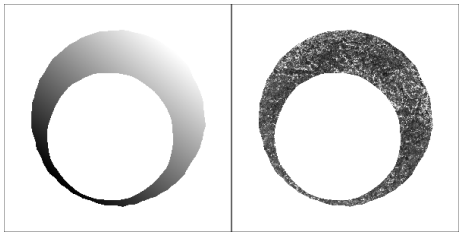

Figure 6: We have the fitted plane on the left and the residuals on the right. In this example the original image is slightly darker in the bottom left corner and brighter on the top right. The residuals are free from any such effects. We take the residuals for further processing.

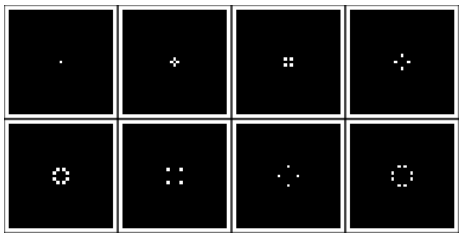

Figure 7: Illustration of the first eight matrices in a circularly symmetric basis. Each matrix takes the value 1 for pixels that are colored white and 0 otherwise.

$$f_5(i,j) = \begin{cases} 1 & \text{if } d((i,j), center) = d_5 = \sqrt{5} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$$Image(i,j) = \sum_{k=1}^{K} \beta_k f_k(i,j) + \epsilon(i,j), \tag{2}$$

The number of basis functions required, K, depends on the resolution of the image. For example, a 3×3 image has three possible distances from the center (including the corners), and can be represented using three basis functions. A 1769×1769 image (this is the size of our images after cropping) requires 237,569 basis functions. The coefficient for each basis function is the mean of pixel values for pixels with the corresponding distance from the center, and this is also the maximum likelihood estimate.

Since we have only selected the breechface impression for analysis, some fraction of the coefficients will not be computed, but the total number of

coefficients is still relatively large. There are large local fluctuations in the model coefficients since many of them are estimated only using a small number of pixels (many of the pixels are also missing due to the selection of the breechface area). Hence we fit a loess regression to the basis coefficients to reduce the variance. We use a locally weighted quadratic function, specifying the proportion of points to be included in the fit to be 0.75. This results in an effective degrees of freedom of less than 5 for each image. The results are in Figure 8.

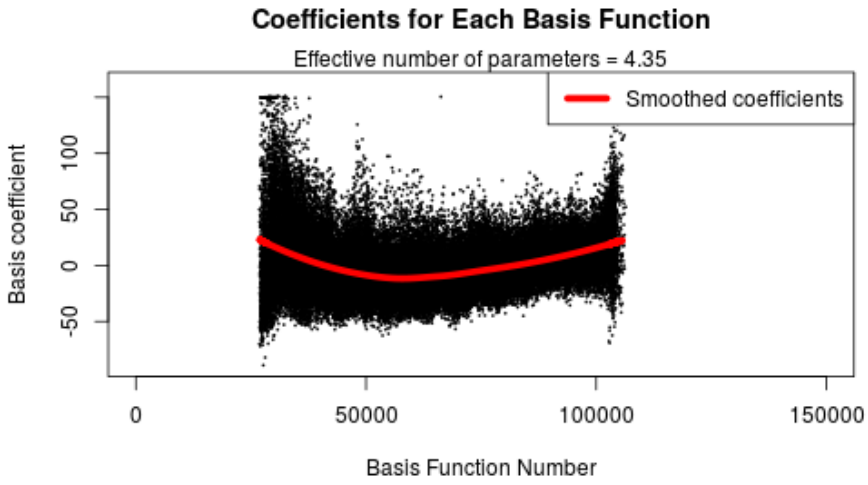

Figure 8: Fitted coefficients for our example image.

The fitted model and the residuals are in Figure 9. These residuals are free from both planar bias from the previous step, and circular symmetry.

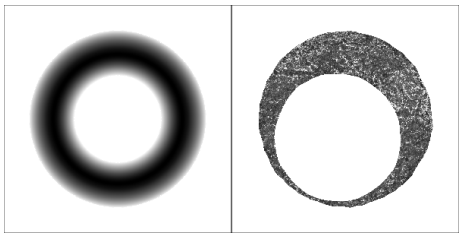

Figure 9: Fitted circularly symmetric model and resulting residuals.

2.2.4 Outlier removal and filtering

The last pre-processing step is outlier removal and filtering. Again this follows the methodology of [11] and [6]. Outliers are removed and filled in using the methodology in [2] so as not to affect the similarity scores being computed, and filtering highlights certain features of the image. The resulting image after all pre-processing is on the left in Figure 10.

2.2.5 Maximize correlation by translations and rotations

After pre-processing, step 5 computes a similarity metric. Here we use the maximum cross-correlation function (CCF_{max}), used in [12] and [6]. For each rotation angle, we compute

$$CCF(I_1, I_2) = \frac{\sum_{i,j} I_1(i,j) I_2(i+dy, j+dx)}{\sqrt{\sum_{i,j} I_1(i,j)^2} \sqrt{\sum_{i,j} I_2(i,j)^2}}, \tag{3}$$

where I_1 and I_2 are the two images, i indexes the rows and j indexes the columns, and dx and dy represent translations. The CCF is a matrix of correlation values, where each entry corresponds to a particular translation, and we store the maximum correlation. We repeat for rotation angles 2.5° apart, and then .5° apart in the neighborhood of the highest correlation. The maximum correlation between two images is known as the CCF_{max}. An example is in Figure 10.

2.2.6 Compute probability of obtaining a higher score by chance

In the final step, we convert each similarity score into a statement of probability, specifically a "random match probability." [7] and [5] propose methods for computing likelihood ratios, which are a ratio of the densities of similarity scores using non-match versus match distributions. [7] proposes parametric distributions applicable to all guns, while [5] proposes case-specific distributions, determined empirically by conducting test fires for each make and model of gun.

Here we focus on the denominator of the likelihood ratio, and propose a method to compute the probability of obtaining a higher score by chance empirically given a known database. This method can be applied to all guns. These probabilities attach meaning to the scores, and also serve as a measure of uncertainty for this comparison procedure.

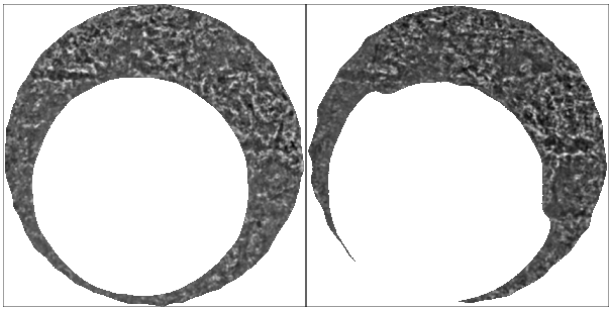

Figure 10: We compare our example image (on the left) against another image from the NBIDE study, which was obtained using the same gun. We obtain a similarity score of .36, with a rotation angle of $-15°$, meaning that the second image is rotated 15° counter-clockwise. Plotting the two images with the second correctly rotated, we notice that the breechface marks are now lined up nicely.

We assume that all CCF_{max} values for non-matches are drawn from the same distribution, and given such a distribution, we compare each newly computed score against this distribution, and compute the right tail proportion. This value is the probability of observing a higher CCF_{max} by chance. In reality, we do not have access to such a distribution, but what we might have instead is a known database, where we are able to compute all pairwise non-matching scores. These form a sample from the unknown distribution. An illustration is in Figure 11. For example, using the NBIDE data set, for each new image we have a database of 107 images, and doing all pairwise comparisons within this database, we have a total of 10,494 non-match scores, which would form a sample from the unknown population of non-match scores. We can then compare .36 against this distribution.

3 Results

3.1 All pairwise comparisons

As described in Section 2.1, we do all pairwise comparisons for each of the 108 images in the data set, for a total of 11,556 comparisons. Since we know the true identity of all the images, we are able to group the comparisons into true matches and non-matches. The distributions of CCF_{max} values split

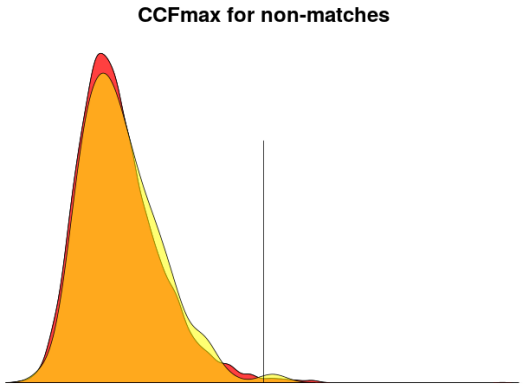

Figure 11: The red distribution is the theoretical distribution of CCF_{max} values for non-matches, which we do not know. The yellow distribution are values computed using a known database, and can be treated as a sample from the red distribution. The vertical line represents a new CCF_{max} value. To compute the probability of observing a higher score than this, we take the proportion of the yellow distribution to the right of the vertical line.

by true matches and non-matches are in Figure 12. The two distributions overlap, meaning that this method is unable to perfectly separate the true matches from the non-matches.

We also look at the probabilities obtained in step 6, and again we group the results by true matches and true non-matches. The results are in Figure 13. For the true matches, ideally the probabilities should be small, indicating that the CCF_{max} values are large enough that the probability of getting a larger score by chance is small. Here for about 80% of the observations the probabilities are actually zero, because some of the similarity scores for the true matches are much larger than the largest observed value for true non-matches, as seen in Figure 12. There are several probabilities close to 1, which correspond to the points in the left part of the overlapping regions in Figure 12.

3.2 Comparison with other results

As described in Section 2.2, we have added two pre-processing steps: the automatic selection of breechface marks and the removal of circular symmetry.

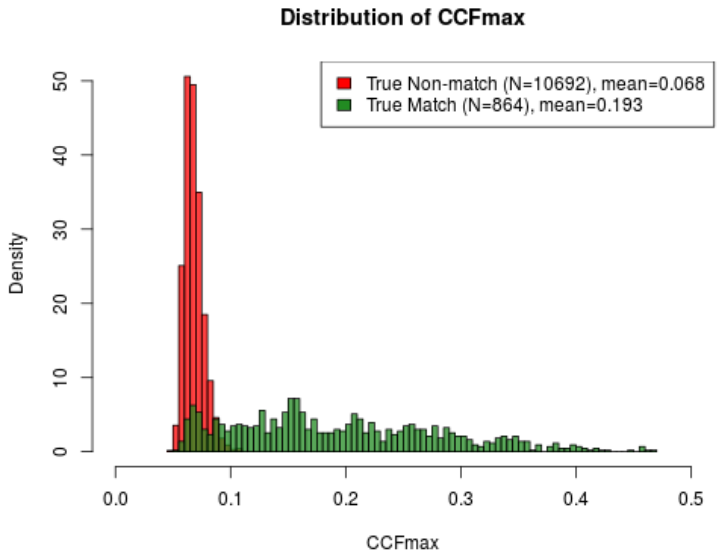

Figure 12: The true non-match CCF_{max} distribution has a much smaller mean and variance than the distribution for true matches, but there is not a clear separation between the two. This means that some true matches do not produce very high similarity scores.

We compare the similarity scores with those without the addition of these steps, and the results are in in Figure 14. From the plot in the center, it is fairly obvious that removing circular symmetry has on average little impact on the similarity scores for true matches. For the true non-matches, there are more points below the 45-degree line (in fact, over 2/3 of the points), indicating that removing circular symmetry reduces scores for non-matches. This is illustrated more clearly in the toy example in Figure 15. For the true matches, the individual marks are similar enough that even after any circular symmetry is removed, the images remain highly correlated.

To determine the effect of automatically selecting the breechface marks, we consider the first plot in Figure 14. There is a linear relationship between the two sets of results, but one noticeable difference is that each score on the horizontal axis corresponds to a range of observed scores on the vertical axis, and the positive correlation is not as strong. The scores in general are higher since they are above the 45-degree line. This is especially true for the true matches, and if we were to draw a best fit line for only the true

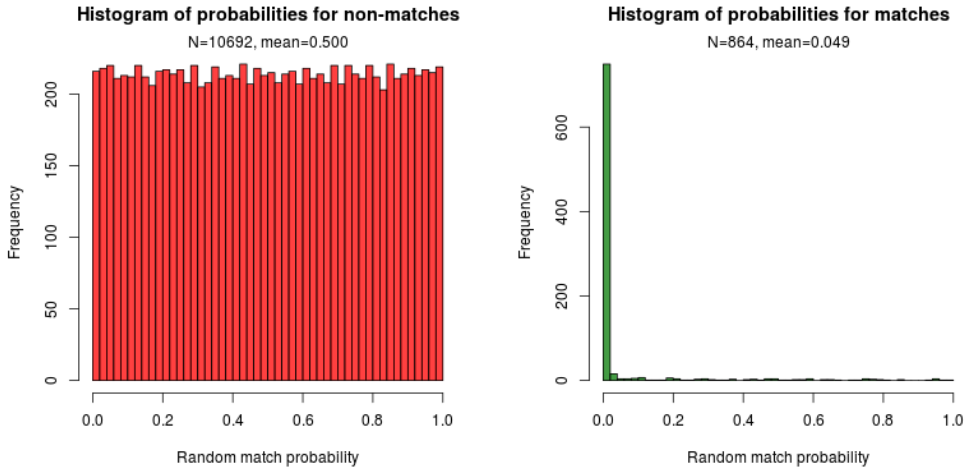

Figure 13: The true non-match probabilities follow a roughly uniform distribution, which is to be expected.[3] The true matches have low probabilities of obtaining a higher score by chance, which is a good result.

matches, we notice that this is above the 45-degree line. The effect on the true non-matches is less obvious, and an argument can be made for there being little to no effect.

Putting the two together, we get the last plot in Figure 14. We see a weaker linear relationship, higher scores for the true matches (on average), and slightly lower scores for the true non-matches, although the effect is not as pronounced as in the second plot. Note that although these observations hold on average, it need not be the case that there is better separation between the true matches and non-matches for each new image, since there is some variability in the scores especially due to the automatic removal of breechface marks, as previously noted.

As a final step, we compare the results to those in [11]. The metric that was used to evaluate performance on the 2D algorithm was the number of matches in the "top 10 list." For each new image, there are 107 images in

[3]We can think of the random match probabilities in the context of the hypothesis test, with the null being that the images do not come from the same gun, and the alternative being that they do. The calculated probabilities are p-values in the context of this hypothesis test. The CCF_{max} values of true non-matches should then come from the null distribution, which tells us that the p-values follow a uniform distribution.

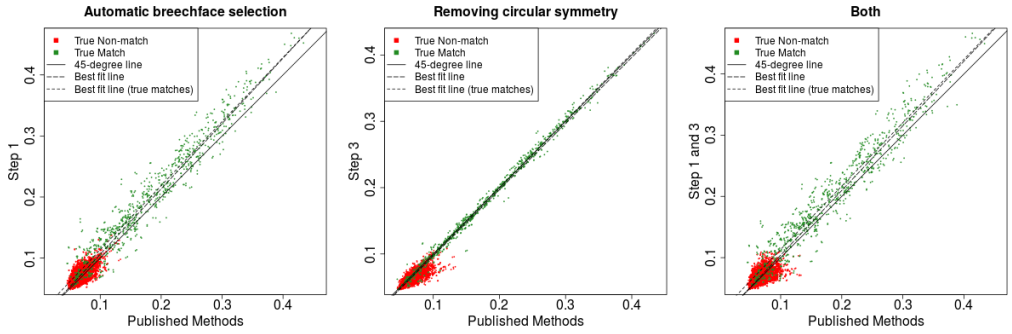

Figure 14: CCF_{max} values for all pairwise comparisons using different sets of pre-processing steps are plotted. Published methods on the horizontal axis refer to the pre-processing steps being a manual selection of breechface marks, leveling, outlier removal and filtering.

the known database, and the top 10 list is the list of 10 images that have the highest similarity scores when compared against the new image. The maximum possible number of true matches in the top 10 list is 8, and this represents the best result. We produce this same metric using the similarity scores we calculate after pre-processing using different methods, and the results are in Table 1.

Table 1: Comparison of the number of true matches in "top 10 lists" for various methodologies. All rows except the last use 2D optical images.

Methodology	8	7	6	5	4	3	2	1	0
FTI in 2007 [11]	13	25	23	19	13	9	5	1	0
Published methodology [11, 6]	58	19	14	10	2	2	1	2	0
Proposed methodology	68	21	7	1	4	3	2	2	0
Only add step 1	67	19	7	4	4	3	2	1	1
Only add step 3	60	21	12	6	4	2	2	1	0
3D [11]	101	7	0	0	0	0	0	0	0

The performance of published methods is markedly better than FTI's algorithm, but we note that the latter were obtained in 2007 and might not be an accurate reflection of current FTI/NIBIN technology. Although we see that our proposed methodology is an improvement over other 2D methods, we do not achieve the accuracy of 3D topographies.

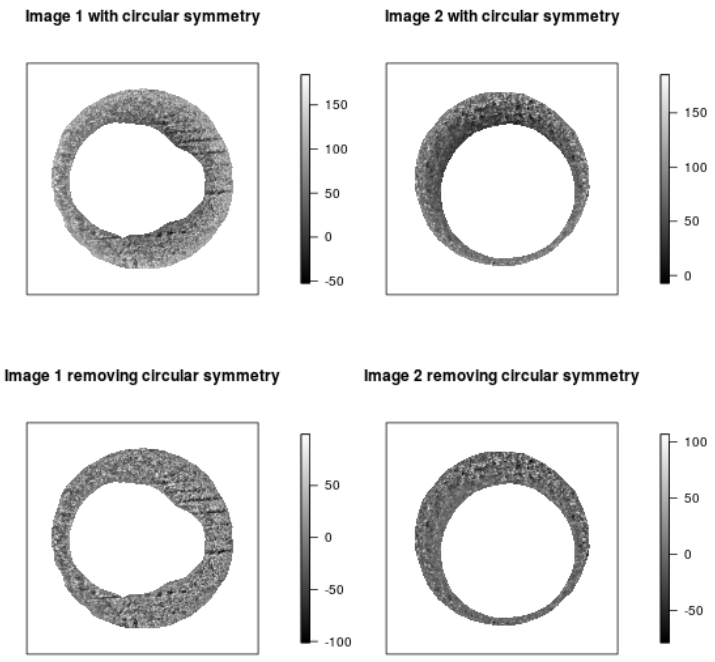

Figure 15: This example illustrates the effect of removing circular symmetry. We compare the images on the left respectively with those on the right. In the images on the first row, we added the same circular symmetry to both images, causing them to be darker in the center and getting brighter towards the edges. The similarity score in this case is .72. On the second row, this circular symmetry is removed and the score drops to just .04, because the individual marks are not very similar.

4 Discussion

4.1 Assumption for computing random match probabilities

In computing the probabilities of observing a higher CCF_{max} value by chance in step 6, we assumed that the all CCF_{max} values for non-matches are drawn from the same distribution. We also rely on this assumption in order to apply the proposed calculation to all makes and models of guns.

With the data in this data set, it is difficult to properly evaluate this

assumption. As a first step we consider the non-match distribution in Figure 12. We can break this up by gun brand, gun and cartridge brand, as well as by type of comparison, to see if there are differences in distributions. As an example, the distributions by type of comparison are in Figure 16. If we were to do formal tests, such as the Kolmogorov-Smirnov test, we do see significant differences in distributions. Although some of these differences might be statistically but not practically significant, it is possible that non-match distributions depend on both the image in the database, as well as the new image, and we consider this situation.

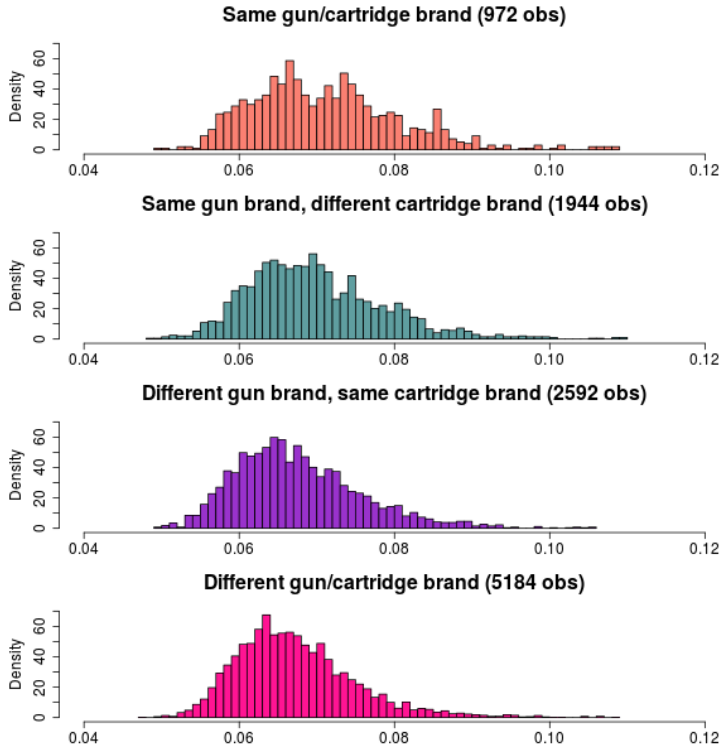

Figure 16: Breaking up the distribution of CCF_{max} values for non-matches by type of comparison. The distribution for comparisons involving the same gun and cartridge brand is shifted slightly to the right.

In an actual comparison against a known database, the identity of the new image is unknown, so we are unable to determine which non-match distribution the computed CCF_{max} value is coming from. We propose two

options. The first is to construct different empirical non-match distributions, for each gun-cartridge combination found in the known database. For example, if we have Ruger-PMC images in the database, we construct a non-match distribution for Ruger-PMC based on all non-match comparisons in the database involving Ruger-PMC. Then when comparing a new image to any Ruger-PMC image, the similarity score produced will be compared to the Ruger-PMC non-match distribution, to determine the probability of obtaining a higher score by chance. This probability might be interpreted as an average over all possible identities of the new image. This first option is plausible if the known database is a large enough representative sample of the population of non-match images.

If we do not have a well-constructed database, we consider the following second option. Based on Figure 16, we might expect that the non-match similarity scores are the highest for comparisons where both the gun and cartridge brand are the same. A conservative approach would be to only use these comparisons to form the empirical non-match distribution. Every new comparison would be compared to this distribution, and the probabilities might then be interpreted as an upper bound: the probability of obtaining a larger similarity score by chance is less than p, where p is the probability computed in this fashion.

5 Conclusion

We have proposed an improvement of existing methodology for comparing 2D optical images of breechface impressions. This algorithm is fully automated and open source, and provides better accuracy. Accuracy however does not reach the level of 3D topographies for this data set. This could be because the marks left on the cartridge cases were insufficiently pronounced to be adequately captured using 2D ring light photography, whereas these were captured in the 3D topographies. Another point to note is that the methodology that we improved upon is not the only existing methodology, but it is arguably the most established. Our proposed changes are to the pre-processing steps and can be applied to different similarity metrics or registration procedures (e.g. [8, 6]). Similarly, our proposed empirical method of computing "random match probabilities" given a known database, can be adapted to different similarity metrics and also to 3D topographies.

Regarding future work, the methodology should be evaluated on 2D op-

tical images collected from other studies, and eventually modified to be used for 3D topographies. Based on this larger data set performance of 2D should again be compared to 3D, in order to say with more certainty if 3D is indeed superior. In order to properly evaluate a fully automated algorithm, we would need to test it on a large amount of data to measure its performance on different makes and models of guns and ammunition, as well as over a large number of fires per gun (known in the literature as persistence studies). Comparisons should also be made with the conclusions of human examiners. Non-match distributions should be looked at more closely to evaluate assumptions and to ensure that "random match probabilities" are computed in a sound manner. Finally, little work (to our knowledge) has been done on large database searches, for example how to compute a signature or hash value for each image, and this is a potential related area of research.

6 Acknowledgements

This work was supported by the Center for Statistics and Applications in Forensic Evidence (CSAFE). We would like to thank Xiaoyu Alan Zheng at NIST for advice and for collecting the data. We also thank Joseph Roth and his co-authors at Michigan State University for sharing the code they used in [6].

References

[1] President's Committee of Advisors on Science and Technology. *Report to the President on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Executive Office of the President, Sept. 2016.

[2] John D'Errico. *inpaint_nans*. 2004. URL: `https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans` (visited on 02/07/2017).

[3] AFTE Criteria for Identification Committee. "Theory of Identification, Range of Striae Comparison Reports, and Modified Glossary Definitions – An AFTE Criteria for Identification Committee Report". In: *AFTE Journal* 24.2 (1992), pp. 336–340.

[4] Laura Lightstone. "The Potential for and Persistence of Subclass Characteristics on the Breech Faces of SW40VE Smith & Wesson Sigma Pistols". In: *AFTE Journal* 42.4 (2010), pp. 308–322.

[5] Fabiano Riva and Christophe Champod. "Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases". In: *Journal of Forensic Sciences* 59.3 (2014), pp. 637–647. ISSN: 1556-4029. DOI: `10.1111/1556-4029.12382`. URL: `http://dx.doi.org/10.1111/1556-4029.12382`.

[6] J. Roth et al. "Learning-based ballistic breech face impression image matching". In: *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. Sept. 2015, pp. 1–8. DOI: `10.1109/BTAS.2015.7358774`.

[7] John Song. "Proposed "Congruent Matching Cells (CMC)" Method for Ballistic Identification and Error Rate Estimation". In: *AFTE Journal* 47.3 (2015), pp. 177–185.

[8] John Song. "Proposed "NIST Ballistics Identification System (NBIS)" Based on 3D Topography Measurements on Correlation Cells*". In: *AFTE Journal* 45.2 (2013).

[9] J Song et al. "Development of ballistics identification–from image comparison to topography measurement in surface metrology". In: *Measurement Science and Technology* 23.5 (2012), p. 054010. URL: `http://stacks.iop.org/0957-0233/23/i=5/a=054010`.

[10] Mingsi Tong et al. “Fired Cartridge Case Identification Using Optical Images and the Congruent Matching Cells (CMC) Method”. In: *Journal of Research of the National Institute of Standards and Technology* 119 (2014).

[11] T. Vorburger et al. *Surface topography analysis for a feasibility assessment of a National Ballistics Imaging Database*. Tech. rep. National Institute of Standards and Technology, 2007.

[12] Todd J. Weller et al. “Confocal Microscopy Analysis of Breech Face Marks on Fired Cartridge Cases from 10 Consecutively Manufactured Pistol Slides*”. In: *Journal of Forensic Sciences* 57.4 (2012), pp. 912–917. ISSN: 1556-4029. DOI: 10.1111/j.1556-4029.2012.02072.x. URL: http://dx.doi.org/10.1111/j.1556-4029.2012.02072.x.

[13] Lubov E. Zeifman and William F. Eddy. “Statistical Estimation of the Point Spread Function of the Hubble Space Telescope”. 2017.