Model-Based Similarity Scores for the Comparison of Cartridge Case Impressions

Joseph Zemmels,[1] Heike Hofmann,[1] and Susan VanderPlas[2]

[1] *Iowa State University*[a]

[2] *University of Nebraska - Lincoln*

(Dated: 14 September 2023)

1   Put your abstract here.

———
[a] jzemmels@iastate.edu; other info

I. INTRODUCTION

A *cartride case* is the part of firearm ammunition that houses the projectile and propulsive device. When a firearm is discharged and the projectile travels down the barrel, the cartridge case moves in the opposite direction and slams against the back wall, the *breech face*, of the firearm. Markings on the breech face are "stamped" into the surface of the cartridge case leaving so-called *breech face impressions*.

In a traditional examination, forensic examiners use these impressions analogous to a fingerprint to determine whether two cartridge cases were fired from the same firearm. First, two cartridge cases are collected - perhaps one is from a crime scene and the other is collected from a suspect's gun. An examiner places the two cartridge cases beneath a "comparison microscope" that merges the views of two compound microscopes into a single split view. The examiner assesses the "degree of similarity" between the markings on the cartridge cases and reaches either an *identification*, meaning the cartridge cases were fired from the same firearm, an *elimination*, meaning they were fired from different firearms, or an *inconclusive*, meaning the evidence is insufficient to make an identification or elimination (AFTE Criteria for Identification Committee, 1992).[1]

Critics of traditional forensic examinations cite a lack of "foundational validity" under-lying the procedures used by firearm and toolmark examiners (National Research Council, 2009; President's Council of Advisors on Science & Technology, 2016). In particular, ex-aminers rely largely on their subjective findings rather than on a well-defined procedure to measure similarity. PCAST (2016) pushed for "developing and testing image-analysis

2

algorithms" to objectively measure the similarity between cartridge cases. Some recently proposed methods focus on measuring similarity using binary rules [**better way to say that?**] [**CITATIONS**]. These methods have the benefit of being interpretable, although recent work has demonstrated that they can be highly sensitive to parameter choice (Zemmels *et al.*, 2023).

In this paper, we propose a model-based procedure for measuring the similarity between two digital scans of cartridge cases. Our procedure measures similarity using a set of numerical features rather than binary rules. The result is a continuous score obtained by evaluating a trained statistical model, which adds nuance to the similarity measure past concluding that the cartridge cases did or did not originate from the same firearm.

In the following sections, we first review recently proposed algorithms to compare firearm evidence. We introduce our similarity scoring pipeline and share results of applying the pipeline to a data set of cartridge case scans available at [**data repo citation**]. We discuss how our proposed method builds upon previously proposed methods to obtain nuanced, informative, and robust similarity measures.

A. Previous Work

Recent proposals for automatic cartridge case scoring algorithms borrow from image processing and computer vision techniques. For example, Vorburger *et al.* (2007) proposed using the cross-correlation function (CCF) to compare images or scans of cartridge case surfaces. The CCF measures the similarity between two matrices for all possible translations of one matrix against the other. Calculating the CCF while rotating one of the scans

therefore allows for estimation of the optimal translation and rotation, together referred to as the *registration*, between the two scans; simply choose the rotation/translation at which the CCF is maximized. Hare *et al.* (2017) used the CCF, among other features, to compare scans of bullets. Tai and Eddy (2018) developed an open-source cartridge case comparison pipeline that compared cartridge case images using the CCF.

Song (2013) noted that two matching cartridge cases often share similar impressions in specific regions, so calculating the CCF between two full scans may not highlight their similarities. Instead, Song (2013) proposed partitioning one cartridge case scan into a grid of "cells" and calculating the CCF between each cell and the other scan. If two cartridge cases are truly matching, then the maximum CCF value between each cell and the other scan, particularly the cells containing distinguishable breech face impressions, should be relatively large. Furthermore, the cells should "agree" on the registration at which the CCF is maximized. Song (2013) outlined the "Congruent Matching Cells" algorithm to determine the number of cells that agree on a particular registration. A cell is classified as a Congruent Matching Cell (CMC) if its estimated registration is within some threshold of the median registration across all cells and its CCF value is above some threshold. A number of follow-up papers proposed alterations to the the original CMC method (Chen *et al.*, 2017; Tong *et al.*, 2015). Zemmels *et al.* (2022) introduced an open-source implementation of the CMC method in the `cmcR` R package. As an alternative to defining Congruent Matching Cells, Zhang *et al.* (2021) proposed using a clustering algorithm from Ester *et al.* (1996) to determine the number of cells in agreement on a specific registration.

Currently, none of these papers have proposed rigorous procedure for comparing differ- ent cartridge case comparison algorithms. This includes selecting optimal parameters for a specific algorithm. Zemmels *et al.* (2023) proposed an optimization criterion to select parameters for the CMC algorithm. Analogously, Hare *et al.* (2017) developed a validation procedure to select parameters for a bullet comparison algorithm. In this work, we introduce a novel cross-validation procedure to learn and test optimal parameters for this cartridge case algorithm.

II. METHODS

We now discuss the methods behind the comparison algorithm. We divide the methods into three stages:

1. **Pre-processing**: prepare cartridge case scans for comparison

2. **Comparing**: compare two cartridge cases and compute similarity features

3. **Scoring**: measure the similarity between the two cartridge cases using a trained classifier

The following sections detail each of these stages. Throughout, we treat "surface matrix" and "scan" synonymously.

After taking a topographical scan of the cartridge case surfaces, we manually annotate the breech face impression region (shown in red). We automatically pre-process and compare the scans resulting in a similarity score, either a binary classification or class probability, derived from a classifier model. As we pointed out in Chapter 2, prosaic descriptions like

the following are insufficient to reproduce an algorithm. We refer the reader to https://github.com/jzemmels/jdssvSubmission/tree/main/supplementary-code for the source code used to implement the method and derive the results presented.

A.   Pre-processing

We first use the open-source FiX3P web application (https://github.com/talenfisher/fix3p) to manually annotate the breech face impression region. The FiX3P software includes functionality to "paint" the surface of a cartridge case using a computer cursor and save the painted regions to a *mask.* A mask is a 2D array of hexidecimal color values of the same dimension as its associated surface matrix. When initialized, every element of a mask is a shade of brown (#cd7f32) by default. Any elements painted over by the user will be replaced with the user's selected color value.

We pre-process the raw scans by applying a sequence of functions available in the R packages x3ptools (Hofmann *et al.*, 2022) and cmcR (Zemmels *et al.*, 2022). Figure 1 shows the effect that each function has on the scan surface values. Gray pixels in each plot represent missing values in the surface matrix. The x3p_delete function removes values in the scan based on the associated mask. Next, the preProcess_removeTrend function subtracts a fitted conditional median plane from the surface values to "level-out" any global tilt in the scan. The preProcess_gaussFilter() function applies a bandpass Gaussian filter with wavelength cutoff values 16 and 500 microns to remove small-scale noise and other large-scale structure, which better highlights the medium-scale breech face impressions. Finally, the preProcess_erode() function applies the morphological operation of erosion with a

6

circular structure element of radius 12 on the edge of the non-missing surface values (Haralick *et al.*, 1987). This has the effect of shaving off values on the interior and exterior edge of the surface, which are often extreme "roll-off" values that unduly affect the comparing stage if not removed. The final result is a cartridge case surface matrix with emphasized breech face impressions.
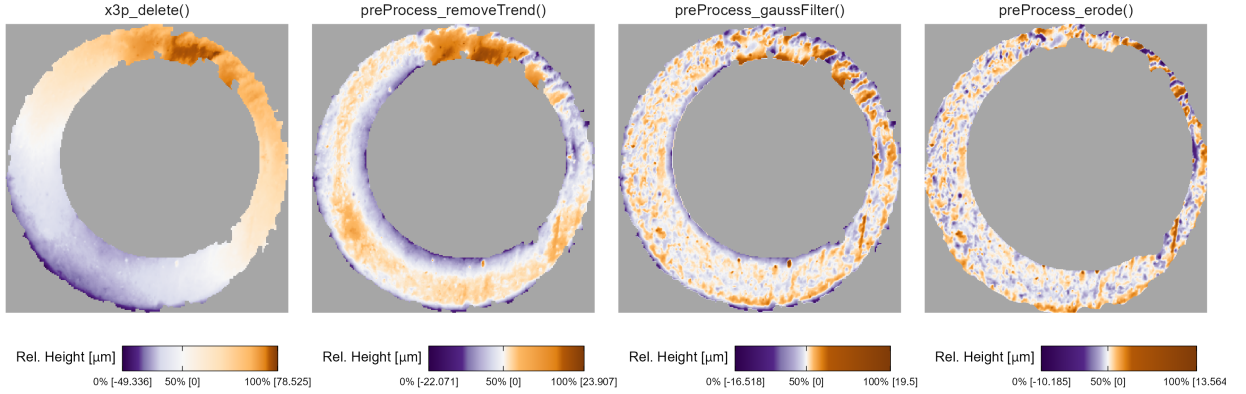

FIG. 1. We apply a sequence of pre-processing functions to each scan. Each pre-processing step further emphasizes the breech face impressions in the scan.

Next, we compute a set of similarity features for two pre-processed cartridge case scans.

B. Comparing

In this section, we introduce a set of similarity features for two cartridge case scans. We calculate features at two scales: between two full scans and between individual cells. Analogous to how a forensic examiner uses a comparison microscope with different magnification levels, this allows us to assess the similarity between two scans at the macro and micro levels.

1. Notational Conventions

First, we introduce notation that will be used to define the features. Let A and B denote two surfaces matrices that we wish to compare. For simplicity, we assume $A, B \in \mathbb{R}^{k \times k}$ for a positive integer k.[2] We use lowercase letters and subscripts to denote a particular value of a matrix: a_{ij} is the value in the i-th row and j-th column, indexed starting from the top-left corner, of matrix A. In the following sections, we will use the two known-match cartridge cases in Figure 2 as example matrices A and B.

To accommodate structurally missing values, we adapt standard matrix algebra by encoding the notion of "missingness" into the space of real values as follows: if an element of either matrix A or B is missing, then any element-wise operation including this element is also missing. Standard matrix algebra holds for non-missing elements. For example, the addition operator is defined as:

$$A \oplus_{NA} B = (a_{ij} \oplus_{NA} b_{ij})_{1 \leq i,j \leq k} = \begin{cases} a_{ij} + b_{ij} & \text{if both } a_{ij} \text{ and } b_{ij} \text{ are numbers} \\ NA & \text{otherwise} \end{cases}$$

Other element-wise operations such as $\ominus_{NA}$ are defined similarly. For readability, we will use standard operator notation $+, -, >, <, I(\cdot), \dots$ and assume the extended, element-wise operations as defined above. Note that this definition of dealing with missing values is consistent with a setting of `na.rm = FALSE` in terms of calculations in R (R Core Team, 2017).
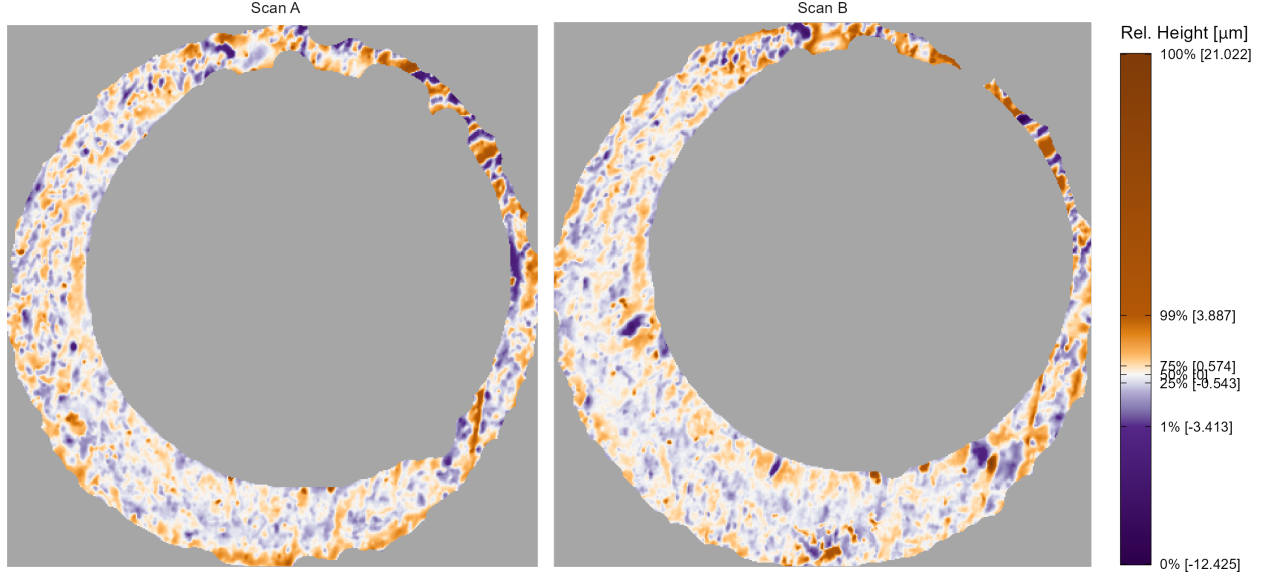
8

FIG. 2.   A matching pair of processed cartridge case scans. We measure the similarity between these cartridge cases using the distinguishable breech face impressions on their surfaces.

2. *Registration Estimation*

A critical step in comparing A and B is to find a transformation of B such that it aligns best to A (or vice versa). In image processing, this is called *image registration.* Noting that A and B are essentially grayscale images with structurally missing values, we rely on a standard image registration technique (Brown, 1992).

In our application, a registration is composed of a discrete translation by $(m, n) \in \mathbb{Z}^2$ and rotation by $\theta \in [-180°, 180°]$. To determine the optimal registration, we calculate the *cross-correlation function* (CCF) between A and B, which measures the similarity between A and B for every possible translation of B, denoted $(A \star B)$. We estimate the registration by calculating the maximum CCF value across a range of rotations of matrix B. Let B_θ

9

145   denote B rotated by an angle $\theta \in [-180°, 180°]$ and $b_{\theta_{mn}}$ the m, n-th element of B_θ. Then

146   the estimated registration (m^*, n^*, θ^*) is:

$$(m^*, n^*, \theta^*) = \arg \max_{m,n,\theta} (a \star b_\theta)_{mn}.$$

147   In practice we consider a discrete grid of rotations $\Theta \subset [-180°, 180°]$. The registration

148   procedure is outlined in algorithm 1. We refer to the matrix that is rotated as the "target."

149   The result is the estimated registration of the target matrix to the "source" matrix.

Data: Source matrix A, target matrix B, and rotation grid Θ

Result: Estimated registration of B to A, (m^*, n^*, θ^*), and cross-correlation function

maximum, $CCF_{\max}$

for $\theta \in \Theta$ **do**

    Rotate B by θ to obtain B_θ;

    Calculate $CCF_{\max,\theta} = \max_{m,n}(a \star b_\theta)_{mn}$;

    Calculate translation $[m_\theta^*, n_\theta^*] = \arg \max_{m,n}(a \star b_\theta)_{mn}$

end

Calculate overall maximum correlation $CCF_{\max} = \max_\theta \{CCF_{\max,\theta} : \theta \in \Theta\}$;

Calculate rotation $\theta^* = \arg \max_\theta \{CCF_{\max,\theta} : \theta \in \Theta\}$;

return *Estimated rotation θ^*, translation $m^* = m_{\theta^*}^*$ and $n^* = n_{\theta^*}^*$, and $CCF_{\max}$*

Algorithm 1: Image Registration Procedure

150   To accommodate missing values, we also compute the *pairwise-complete correlation* using

151   only the complete value pairs, meaning neither value is missing, between A and B.

3. *Registration-Based Features*

153 *a. Full-Scan Registration.* We first estimate the registration between two full scans A

154 and B using algorithm 1 with a rotation grid $\boldsymbol{\Theta} = \{-30°, -27°, ..., 27°, 30°\}$. This results

155 in an estimated registration (m^*, n^*, θ^*) and similarity measure $CCF_{\max}$. We also perform

156 algorithm 1 with the roles of A and B reversed, meaning the target scan A is aligned to

157 source scan B.

158 To accommodate these two comparison directions, we introduce a new subscript $d = A, B$,

159 referring to the source scan in Image Registration Algorithm. Consequently, we obtain two

160 sets of estimated registrations, $(m_d^*, n_d^*, \theta_d^*)$ and $CCF_{\max,d}$, for $d = A, B$.[3] For $d = A$, we

161 then apply the registration transformation $(m_A^*, n_A^*, \theta_A^*)$ to B to obtain B^* and compute

162 the pairwise-complete correlation, $cor_{\text{full},A}$, between A and B^*. We repeat this in the other

163 comparison direction to obtain $cor_{\text{full},B}$ and average the two:

$$cor_{\text{full}} = \frac{1}{2}\left(cor_{A,\text{full}} + cor_{B,\text{full}}\right).$$

164 We assume that the **full-scan pairwise-complete correlation** is large for truly match-

165 ing cartridge cases.

166 *b. Cell-Based Registration.* We next perform a cell-based comparison procedure, which

167 begins with selecting one of the matrices, say A, as the "source" matrix that is partitioned

168 into a grid of cells. The left side of Figure 3 shows an example of such a cell grid overlaid on

169 a scan. Each of these source cells will be compared to the "target" matrix, in this case B^*.

170 Because A and B^* are already partially aligned from the full-scan registration procedure, we

171   compare each source cell to B^* using a new rotation grid of $\mathbf{\Theta}'_A = \{\theta^*_A - 2°, \theta^*_A - 1°, \theta^*_A, \theta^*_A +$

172   $1°, \theta^*_A + 2°\}$.

173   We now extend the surface matrix notation introduced previously to accommodate cells.

174   Let A_t denote the t-th cell of matrix A, $t = 1, ..., T_A$ where T_A is the total number of cells

175   containing non-missing values in scan A (e.g., $T_A = 43$ in Figure 3) and let $(a_t)_{ij}$ denote the

176   i, j-th element of A_t.

177   The cell-based comparison procedure is outlined in algorithm 2.

Data: Source matrix A, target matrix B^*, grid size $R \times C$, and rotation grid $\mathbf{\Theta}'_A$

Result: Estimated translations and $CCF_{\max}$ values per cell, per rotation

Partition A into a grid of $R \times C$ cells;

Discard cells containing only missing values, leaving T_A remaining cells;

for $\theta \in \mathbf{\Theta}'_A$ **do**

    Rotate B^* by θ to obtain B^*_θ;

178

    for $t = 1, ..., T_A$ **do**

    Calculate $CCF_{\max,A,t,\theta} = \max_{m,n}(a_t \star b^*_\theta)_{mn}$;

    Calculate translation $[m^*_{A,t,\theta}, n^*_{A,t,\theta}] = \arg\max_{m,n}(a_t \star b^*_\theta)_{mn}$

    end

end

return $\boldsymbol{F}_A = \{(m^*_{A,t,\theta}, n^*_{A,t,\theta}, CCF_{\max,A,t,\theta}, \theta) : \theta \in \mathbf{\Theta}'_A, t = 1, ..., T_A\}$

Algorithm 2: Cell-Based Comparison Procedure

179   Rather than exclusively returning the registration that maximizes the overall CCF as in

180   algorithm 1, algorithm 2 returns the set $\boldsymbol{F}_A$ of translations and CCF values for each of the

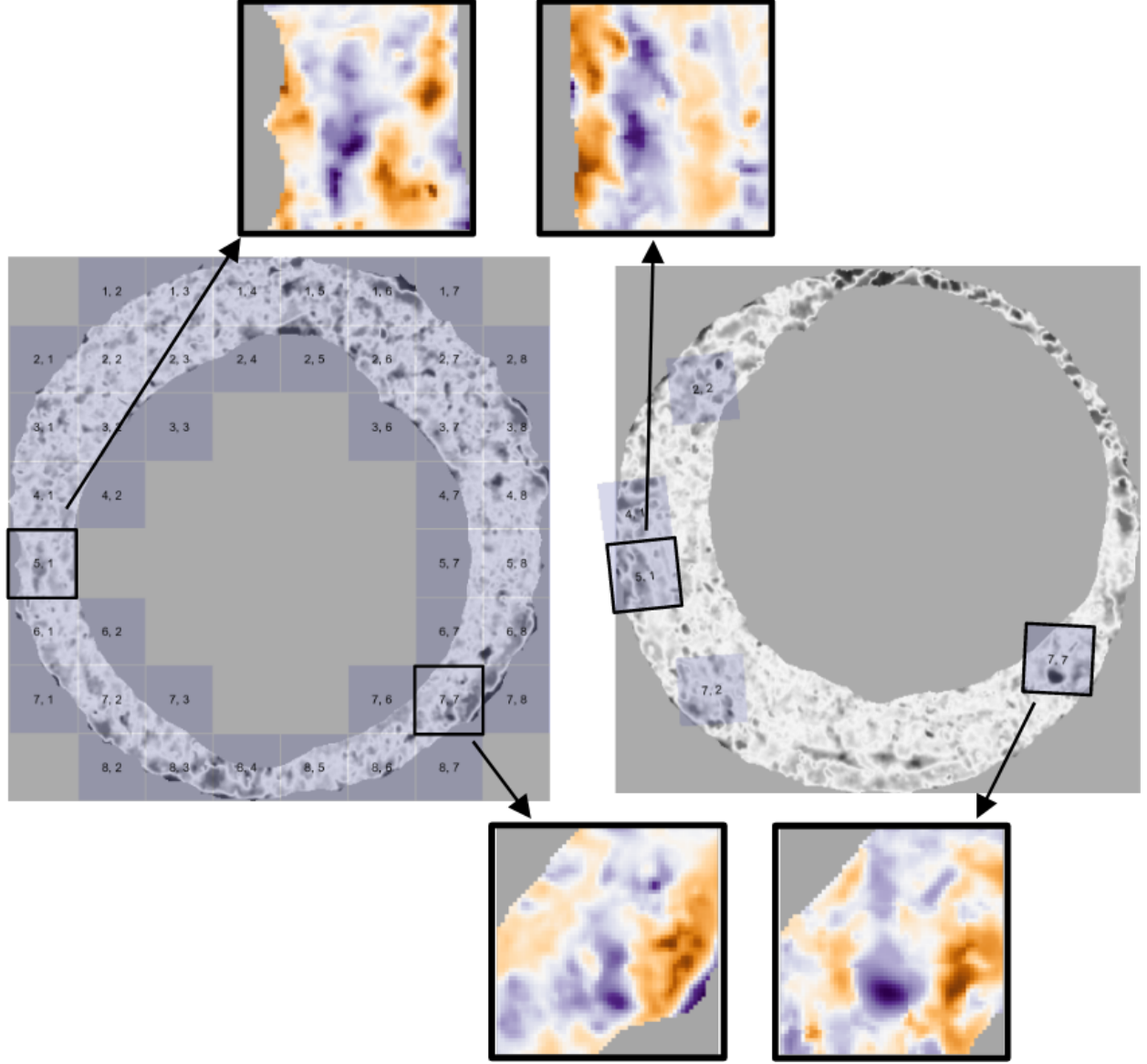181 T_A cells and each rotation in $\mathbf{\Theta}'_A$.

FIG. 3. Estimated registrations of cells from a non-match pair of cartridge cases. A source scan (left) is separated into an 8×8 grid of cells. We exclude cells containing only missing values (visualized here as gray pixels). Each source cell is compared to a target scan (right) to estimate where it aligns best. We show a handful of cells at their estimated alignment in the target scan and magnify the surfaces captured by cell pairs 5, 1 and 7, 7. Although the cartridge case pair is non-matching, we note that there are similarities in the surface markings for these cell pairs.

182   Figure 3 shows the estimated registrations of cells between two non-match cartridge cases.

183   We magnify the surface values captured by cell pairs 5, 1 and 7, 7 and note the similarities

184   in the surface values; for example, the dark purple region in the middle of the cell 7, 7 pair.

185   Just as with the whole-scan registration, we calculate the pairwise-complete correlation

186   between each cell A_t and a matrix $B^*_{\theta,t}$ of the same size extracted from B^*_θ after translating

187   by $[m^*_{A,\theta}, n^*_{A,\theta}]$. From this we obtain a set of pairwise-complete correlations for each cell and

188   rotation: $\{cor_{A,t,\theta} : t = 1, ..., T_A, \theta \in \mathbf{\Theta}'_A\}$.

189   We repeat algorithm 2 and the pairwise-complete correlation calculation using B as the

190   source scan and A^* as the target, resulting in cell-based registration set $\boldsymbol{F}_B$ and pairwise-

191   complete correlations $\{cor_{B,t,\theta} : t = 1, ..., T_B, \theta \in \mathbf{\Theta}'_B\}$.

192   For $d = A, B$ and $t = 1, ..., T_d$, define the cell-wise maximum pairwise-complete correlation

193   as:

$$cor_{d,t} = \max_\theta \{cor_{d,t,\theta} : \theta \in \mathbf{\Theta}'_d\}.$$

194   We compute two features, the **average** and **standard deviation of the cell-based**

195   **pairwise-complete correlations**, using the correlation data:

$$\overline{cor}_{\text{cell}} = \frac{1}{T_A + T_B} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} cor_{d,t}$$

$$s_{cor} = \sqrt{\frac{1}{T_A + T_B - 1} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} (cor_{d,t} - \overline{cor}_{\text{cell}})^2}.$$

196   We expect $\overline{cor}_{\text{cell}}$ and s_{cor} to be large for truly matching cartridge case pairs relative to

197   non-matching pairs.

14

198   For $d = A, B$ and $t = 1, ..., T_d$, define the per-cell estimated translations and rotation as:

$$\theta_{d,t}^* = \arg \max_{\theta} \{ CCF_{\max,d,t,\theta} : \theta \in \boldsymbol{\Theta}_d' \}$$

$$m_{d,t}^* = m_{\theta_{d,t}^*,d,t}^*$$

$$n_{d,t}^* = n_{\theta_{d,t}^*,d,t}^*.$$

199   We compute the **standard deviation of the cell-based estimated registrations**

200   using the estimated translations and rotations:

$$s_{\theta^*} = \sqrt{\frac{1}{T_A + T_B - 1} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} (\theta_{d,t}^* - \bar{\theta}^*)^2}$$

$$s_{m^*} = \sqrt{\frac{1}{T_A + T_B - 1} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} (m_{d,t}^* - \bar{m}^*)^2}$$

$$s_{n^*} = \sqrt{\frac{1}{T_A + T_B - 1} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} (n_{d,t}^* - \bar{n}^*)^2}$$

201   where

$$\bar{m}^* = \frac{1}{T_A + T_B} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} m_{d,t}^*$$

$$\bar{n}^* = \frac{1}{T_A + T_B} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} n_{d,t}^*$$

$$\bar{\theta}^* = \frac{1}{T_A + T_B} \sum_{d \in \{A,B\}} \sum_{t=1}^{T_d} \theta_{d,t}^*.$$

202   We expect $s_{\theta^*}, s_{m^*}, s_{n^*}$ to be small for truly matching cartridge case pairs relative to

203   non-matching pairs.

15

From the full-scan and cell-based registration procedures, we obtain six features summarized in Table I.

cor_{full}	Full-scan pairwise-complete correlation
$\overline{cor}_{\text{cell}}$	Average cell-based pairwise-complete correlation
s_{cor}	Standard deviation of the cell-based pairwise-complete correlations
s_{m^*}	Standard deviation of the cell-based vertical translations (in microns)
s_{n^*}	Standard deviation of the cell-based horizontal translations (in microns)
s_{θ^*}	Standard deviation of the cell-based rotations (degrees)

TABLE I. Six similarity features based on registering full scans and cells.

4. Density-Based Features

We wish to identify when multiple cells agree on, or cluster around, a particular registration value. However, pursuant with the notion that only certain regions of matching cartridge cases contain distinctive markings, it is unreasonable to assume and empirically rare that **all** cells agree on a single registration. In fact, it is common for many cells to disagree on a registration. For example, the left scatterplot in Figure 4 shows the per-cell estimated translations $[m_{A,t,\theta}^*, n_{A,t,\theta}^*]$ when scan A is used as source and B^* as target rotated by $\theta = 3°$. The right scatterplot shows the per-cell estimated translations with the roles of

16

A and B^* reversed for $\theta = -3°$. We see distinctive clusters, the black points, in both plots among many noisy, gray points. The task is to isolate the clusters amongst such noise.

We use the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm proposed by Ester *et al.* (1996) to identify clusters. Compared to other clustering algorithms such as k-means (MacQueen, 1967), DBSCAN does not require a pre-defined number of expected clusters. Instead, the algorithm forms clusters if the number of points within an $\epsilon > 0$ distance of a point exceeds some pre-defined threshold, $minPts > 1$. If a point does not belong to a cluster, then DBSCAN labels that point as "noise." In Figure 4, we use DBSCAN with $\epsilon = 5$ and $minPts = 5$ to identify clusters of size 14 and 13, respectively, visualized as black points. These cluster sizes suggest that the scans match. Additionally, the mean cluster centers are approximately opposites of each other: $(\hat{m}_A, \hat{n}_A, \hat{\theta}_A) \approx (16.9, -16.7, 3°)$ when A is used as source compared to $(\hat{m}_B, \hat{n}_B, \hat{\theta}_B) \approx (-16.2, 16.8, -3°)$ when B^* is used as source. This provides further evidence of a match.

To calculate the density-based features, we first use a 2D kernel density estimator (Venables and Ripley, 2002) to identify the rotation $\hat{\theta}_d$ at which the per-cell translations achieve the highest density. Next, we compute clusters using the DBSCAN algorithm amongst the estimated translations $\{(m^*_{d,t,\hat{\theta}_d}, n^*_{d,t,\hat{\theta}_d}) : t = 1, ..., T_d\}$ like those shown in Figure 4.[4] Let $\boldsymbol{C}_d$ denote the set of cells in the DBSCAN cluster. We treat the mean cluster centers as the estimated translations $[\hat{m}_d, \hat{n}_d]$.

We calculate four features from the density-based clustering procedure: **average DB-SCAN cluster size** C, the **DBSCAN cluster indicator** C_0, and the **root sum of squares of the dens**ity-estimated registrations $(\Delta_\theta, \Delta_{\text{trans}})$ defined as:
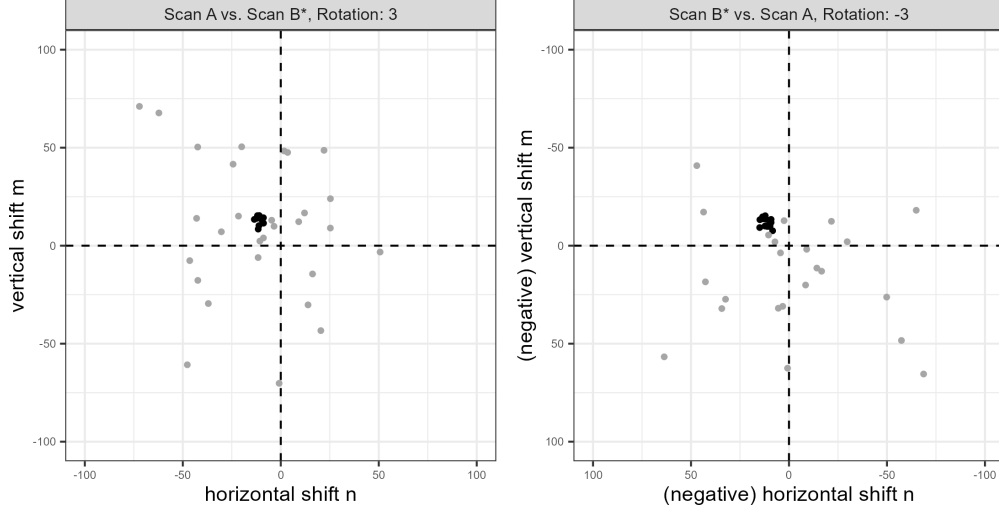
17

FIG. 4. Cluster assignments based on the Density Based Spatial Clustering with Applications to Noise (DBSCAN) algorithm for estimated translations in two comparison directions. Using scan A as source results in a cluster of size 14 (left) compared to 13 when scan B^* is used as source (right). Noting the reversed axes in the right plot, we see that the clusters are located approximately opposite of each other. Points are jittered for visibility.

$$C = \frac{1}{2}\left(|\boldsymbol{C}_A| + |\boldsymbol{C}_B|\right)$$

$$C_0 = I(|\boldsymbol{C}_A| > 0 \text{ and } |\boldsymbol{C}_B| > 0)$$

$$\Delta_\theta = |\hat{\theta}_A + \hat{\theta}_B|$$

$$\Delta_{\text{trans}} = \sqrt{(\hat{m}_A + \hat{m}_B)^2 + (\hat{n}_A + \hat{n}_B)^2}$$

where $|\boldsymbol{C}_d|$ denotes the cardinality of $\boldsymbol{C}_d$ and $I(\cdot)$ is the identity function equal to 1 if the predicate argument "$\cdot$" evaluates to TRUE and 0 otherwise. We use both C and C_0 because of potential missingness in the values of C if no cluster is identified. Missing C values

18

are imputed using the median non-missing value when fitting classifiers, so the missingness information is retained in C_0.

For truly matching cartridge case pairs, we expect C to be large, C_0 to be 1, and $\Delta_\theta, \Delta_\text{trans}$ to be small relative to non-matching pairs. We obtain four density-based features summarized in Table II.

C	Average DBSCAN cluster size
C_0	DBSCAN cluster indicator
Δ_θ	Absolute sum of the density-estimated rotations (degrees)
Δ_trans	Root sum of squares of the density-estimated translations (in microns)

TABLE II. Four similarity features based on the density-based clustering procedure.

III. RESULTS

IV. DISCUSSION

V. CONCLUSION

ACKNOWLEDGMENTS

19

sity, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln.

Additionally, we would like to thank the technicians and staff at the Roy J. Carver High Resolution Microscopy Facility for collecting the topographical scans used in this paper.

[1]The AFTE range of conclusions also permits the examiner to decide that the evidence is *unsuitable* for examination, which can occur if evidence quality is poor; for example, a fragment of a cartridge case is recovered rather than a full cartridge case.

[2]This assumption of equally-sized, square matrices is easily enforced by padding the matrices with additional missing values. Due to the presence of (structurally) missing values around the breech face impression region, additional padding does not interfere with the structure of the scan.

[3]In reality, the true aligning registrations in the two comparison directions are opposites of each other. However, because we compare discretely-indexed arrays using a nearest-neighbor interpolation scheme, the estimated registrations may differ slightly.

[4]If more than one cluster is identified, we binarize the points based on whether they were assigned to any cluster or if they are a noise point and proceed as if there is only one cluster. We assume that two or more clusters form only because of the course rotation grid considered. Were a finer grid used, the points would coalesce into a single cluster around the true translation value. This assumption has empirical support through our experimentation.

AFTE Criteria for Identification Committee (**1992**). "Theory of identification, range striae comparison reports and modified glossary definitions," AFTE Journal **24**(3), 336–340.

Brown, L. G. (**1992**). "A survey of image registration techniques," ACM Computing Surveys **24**(4), 325–376, https://doi.org/10.1145/146370.146374, doi: 10.1145/146370.146374.

Chen, Z., Song, J., Chu, W., Soons, J. A., and Zhao, X. (**2017**). "A convergence algorithm for correlation of breech face images based on the congruent matching cells (CMC) method," Forensic Science International **280**, 213–223, https://doi.org/10.1016/j.forsciint.2017.08.033.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (**1996**). "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, AAAI Press, p. 226–231, doi.org/10.5555/3001460.3001507.

Haralick, R. M., Sternberg, S. R., and Zhuang, X. (**1987**). "Image analysis using mathematical morphology," IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-9**(4), 532–550, doi: 10.1109/tpami.1987.4767941 https://doi.org/10.1109/tpami.1987.4767941.

Hare, E., Hofmann, H., and Carriquiry, A. (**2017**). "Automatic Matching of Bullet Land Impressions," The Annals of Applied Statistics **11**(4), 2332–2356, http://arxiv.org/abs/1601.05788.

Hofmann, H., Vanderplas, S., Krishnan, G., and Hare, E. (**2022**). *x3ptools: Tools for Working with 3D Surface Measurements*, https://github.com/heike/x3ptools, r package version 0.0.3.9000.

MacQueen, J. B. (**1967**). "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. L. Cam and J. Neyman, University of California Press, Vol. 1, pp. 281–297.

National Research Council (**2009**). *Strengthening Forensic Science in the United States: A Path Forward* (The National Academies Press, Washington, DC), https://doi.org/10.17226/12589.

President's Council of Advisors on Science & Technology (**2016**). "Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods," https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

R Core Team (**2017**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Song, J. (**2013**). "Proposed "NIST Ballistics Identification System (NBIS)" Based on 3D Topography Measurements on Correlation Cells," American Firearm and Tool Mark Examiners Journal **45**(2), 11, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=910868.

Tai, X. H., and Eddy, W. F. (**2018**). "A Fully Automatic Method for Comparing Cartridge Case Images,," Journal of Forensic Sciences **63**(2), 440–448, http://doi.wiley.com/10.1111/1556-4029.13577.

Tong, M., Song, J., and Chu, W. (**2015**). "An Improved Algorithm of Congruent Matching Cells (CMC) Method for Firearm Evidence Identifications," Journal of Research of the

National Institute of Standards and Technology **120**, 102, https://doi.org/10.6028/jres.120.008.

Venables, W. N., and Ripley, B. D. (**2002**). *Modern Applied Statistics with S*, fourth ed. (Springer, New York), http://www.stats.ox.ac.uk/pub/MASS4, iSBN 0-387-95457-0.

Vorburger, T. V., Yen, J. H., Bachrach, B., Renegar, T. B., Filliben, J. J., Ma, L., Rhee, H. G., Zheng, A., Song, J. F., Riley, M., Foreman, C. D., and Ballou, S. M. (**2007**). "Surface topography analysis for a feasibility assessment of a national ballistics imaging database," Technical Report NIST IR 7362, https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7362.pdf, doi: 10.6028/NIST.IR.7362, edition: 0.

Zemmels, J., Hofmann, H., and VanderPlas, S. (**2022**). *cmcR: An Implementation of the 'Congruent Matching Cells' Method*, r package version 0.1.9.

Zemmels, J., VanderPlas, S., and Hofmann, H. (**2023**). "A study in reproducibility: The congruent matching cells algorithm and cmcR package," The R Journal **14**(4), 79–102, https://doi.org/10.32614/rj-2023-014, doi: 10.32614/rj-2023-014.

Zhang, H., Zhu, J., Hong, R., Wang, H., Sun, F., and Malik, A. (**2021**). "Convergence-improved congruent matching cells (CMC) method for firing pin impression comparison," Journal of Forensic Sciences **66**(2), 571–582, https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.14634, doi: 10.1111/1556-4029.14634.