Learning-based Ballistic Breech Face Impression Image Matching

Joseph Roth[1], Andrew Carriveau[2], Xiaoming Liu[1], and Anil K. Jain[1]

[1]Department of Computer Science and Engineering, Michigan State University
[2]Forensic Science Division-Bridgeport Lab, Michigan State Police
[1,2]*rothjos1@cse.msu.edu, CarriveauA@michigan.gov, {liuxm,jain}@cse.msu.edu*

Abstract

Ballistic images of a cartridge case or bullet carry distinct "fingerprints" of the firearm, which is the foundation of widely used forensic examination in criminal investigations. In recent years, prior work has explored the effectiveness of correlation-based approaches in matching ballistic imagery. However, most of these studies focused on highly controlled situations and used relatively simple image processing techniques, without leveraging supervised learning schemes. Toward improving the matching accuracy, especially on operational data, we propose a learning-based approach to compute the similarity between two ballistic images with breech face impressions. Specifically, after a global alignment between the reference and probe images, we unroll them into the polar coordinate for robust feature extraction and global registration. A gentleBoost-based learning scheme selects an optimal set of local cells, each constituting a weak classifier using the cross-correlation function. Experimental results and comparison with state-of-the-art methods on the NIST database and a new operational database, OFL, obtained from Michigan State Forensics Laboratory demonstrate the viability of our approach.

1. Introduction

This paper presents an algorithm to quantitatively evaluate whether a *fired cartridge case* was fired from a specific *firearm* purportedly belonging to the suspect. As an important forensics evidence, *a toolmark is created when a harder surface, the tool, comes into contact with a softer surface and causes plastic deformation* [19]. Among all surfaces with toolmarks, fired cartridge cases and bullets are the two most commonly used in criminal investigations and court proceedings. Due to the random variations in the manufacturing of firearms, individual characteristics or "fingerprints" are left on the cartridge cases and bullets – the foundation of firearm examination. In many crime scenes, typical biometric signatures such as fingerprint or DNA may
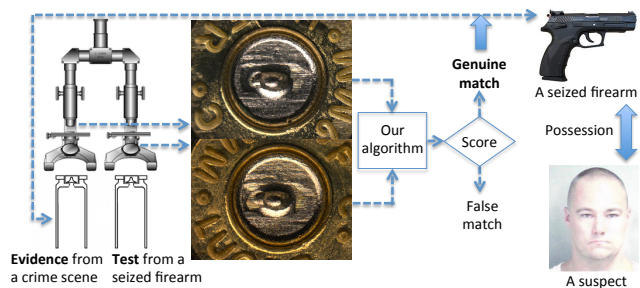

Figure 1. Given two ballistic images (evidance & test) captured by a comparison microscope, our algorithm determines whether the two cartridge cases were fired with the same firearm, which may link the evidence to the firearm, and subsequently to the suspect.

not be available. In such cases, linking the bullet to the gun owned by a suspect provides an important clue to apprehend and prosecute the suspect.

Normally, a firearm examiner fires a *test shot* from a questioned firearm and compares this shot with an *evidence* collected from the crime scene, using microscope-based ballistic images; a genuine match established by this visual examination of the two images will link the firearm (and hence the suspect in possession of the firearm) to the evidence. This manual examination is not only laborious, but also dependent on the examiner's expertise. It is estimated that $1,000+$ firearm examination cases per month are handled by the Michigan State Forensics Lab and each case can take about 2 hours. From 2001 to 2011, there were approximately $500,000$ firearms related crimes in the United States each year [14]. Therefore, there is an urgent need to fundamentally change the paradigm of firearm examination by developing an automatic and robust method, that is scientifically sound, for comparing ballistic images.

The National Research Council (NRC) has published comprehensive reports stating that existing solutions to automatic ballistic image matching are limited [13, 12]. The main commercial solution developed by Forensic Technology WAI Inc. [1] searches for the most similar images in a
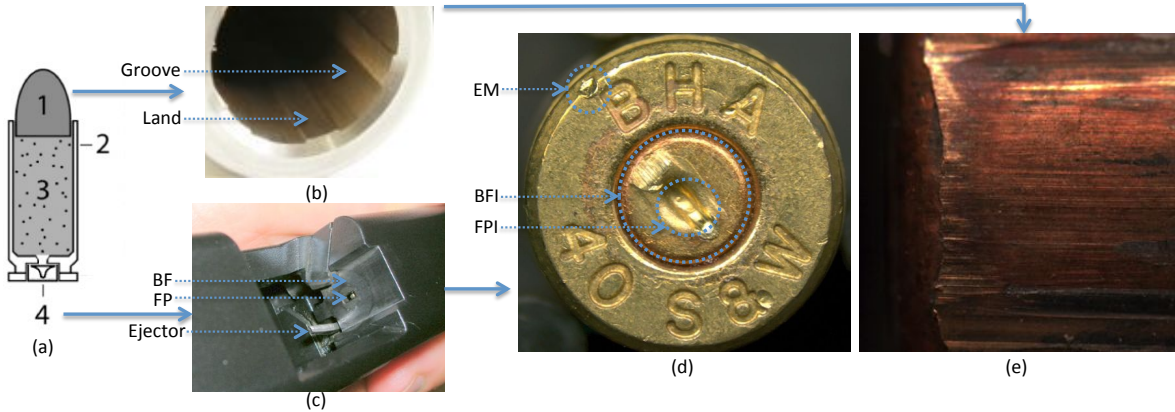
Figure 2. The *formation* and *taxonomy* of ballistic toolmarks[2]. (a) When the firing pin (FP) hits the primer (4), the powder (3) is ignited and propels the bullet (1) out of the barrel. At the same time, the explosive force pushes the cartridge case (2) against the breech face (BF) of the firearm. (b) The lands and grooves fabricated in the barrel leave striated toolmarks on the bullet when it passes through the barrel. (c) The FP, BF and Ejector of a firearm, each is responsible for one type of toolmarks on the cartridge case. (d) Three types of impressed toolmarks on the cartridge case: FP impression, BF impression and Ejector mark (EM). (e) Striated toolmarks on the bullet.

database given a query image of the evidence item. While this retrieval system reduces the manual comparison, it does not directly address the major demand in forensic practice– a comparison of two ballistic toolmarks. In contrast, the academic efforts have either focused on formulating the image matching in the context of non-visual knowledge [3], or matching surface topography [19]. Their main limitation is that only the basic image processing techniques are applied, without any classifier learning. Further, there is a complete lack of exploration on ballistic images collected in real-world operational environment that contain more variations than images obtained in controlled lab environments.

Computer vision and machine learning have made great strides in both their individual capabilities and their confluence. It is generally agreed that a supervised learning-based approach is superior to a non-learning-based approach. Hence, we aim to take advantage of the advances in these two fields to systematically address the limitations of prior work. To this end, we propose a *learning-based* approach to compute the similarity (score) between two ballistic images and classify between the genuine match and false match (a.k.a. known match and known non-match) (Fig. 1). We focus on the breech face impression (BFI) among various ballistic toolmarks given its extensive use in forensics.

Given two 2D microscope images of cartridge cases, we first estimate the center of the BFI circular region of each image, followed by a global rotation estimation between two images. In order to build a global feature correspondence among all training images – a prerequisite for any learning method, we transform an image from the Cartesian coordinate space to the polar coordinate space, similar to how the iris code is generated [5]. Given the two unrolled images, one can use one or more local cells (local image region) with arbitrary size and location to compute

the similarity score. To achieve the maximum discrimination between the genuine and false scores, we employ a gentleBoost-based learning scheme to select a discriminative subset of local cells in the spatial domain, where each cell constitutes a weak classifier by using the classic cross-correlation function (CCF) score. Experiments are conducted on the NIST Ballistics Toolmark Database [2] and our newly collected operational dataset (OFL), with comparison to state-of-the-art methods.

2. Background and Prior Work

In this section, we first provide necessary background on how a toolmark is generated during the firing of a cartridge (bullet and cartridge case), and the various types of ballistics toolmarks. Then we review relevant work on manual and automatic firearm identification via toolmarks.

During the firing of a cartridge, different components (i.e., tools) of a firearm will leave *various* types of toolmarks that have individual characteristics of the fiream and can be used for firearm identification. As shown in Fig. 2 (a), a typical cartridge has four parts: the primer, the powder, the casing (or cartridge case), and the bullet. When a firearm is fired, the firing pin hits the pressure-sensitive primer and ignites the powder, which then propels the bullet out of the barrel. In order to increase the accuracy and range of the bullet by creating spin, the barrel is designed with a slight right or left spiral pattern. These surfaces are imparted with unique microscopic features due to random variations in the manufacturing process. Therefore, when a bullet travels through the barrel, such surface features will leave *striated toolmarks* on the body of the bullet.

Furthermore, there are three areas where *impression*

---

[2]Some of the images are downloaded from the Internet: (a) http://goo.gl/idcCFb; (b) http://goo.gl/O66Ffo; (c) http://goo.gl/TzrRz5; and (d) http://goo.gl/6nnnbk.
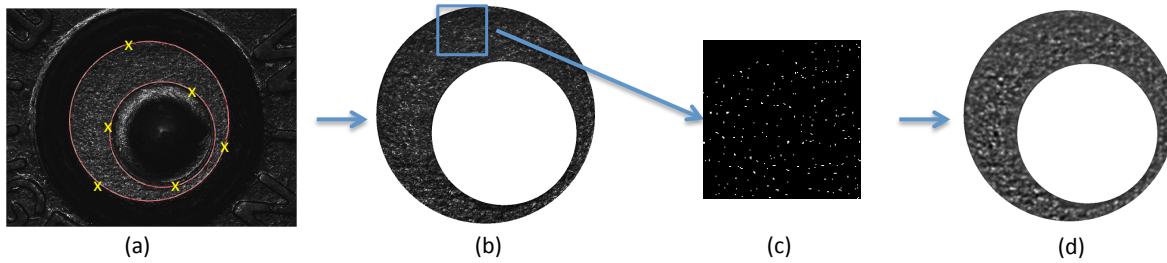
Figure 3. The pipeline of pre-processing a ballistic image: (a) BFI segmentation based on six labeled points, (b) BFI after compensating for planar bias, (c) detected outlier pixels (∼1% of all BFI pixels), and (d) filtered image – the input for all three approaches in Sec. 3.

Table 1. Ballistic image datasets (we use datasets with boldface).

Dataset	Toolm.	Sensor	# Firearms	# Cases	Environment
NIST-Weller	BFI	3D	10	90	Lab
	BFI	**2D**	**10**	**50**	**Lab**
NIST-Fadul	BFI	2D/3D	10	40	Lab
NIST-Light.	BFI	2D/3D	10	30	Lab
NIST-Hamby	Bullet	3D	10	70	Lab
OFL	**BFI**	**2D**	**73**	**146**	**Operational**

marks may be found on a fired cartridge case. First, a FP impression (FPI) is left on the primer as the FP strikes. Second, when the powder is ignited, the explosive force pushes the cartridge case against the breech face (BF) of the firearm, which impresses the surface topography of the BF on the soft surface of the primer. Third, when the spent cartridge case is pulled back and ejected out of the firearm, the ejector impacts an ejector mark on the edge of the case.

When an evidence (cartridge case and/or bullet) is found at a crime scene, it is of great interest to determine whether it was fired from a specific seized firearm. To make this determination, known test fired bullets and cartridge cases are obtained from the firearm. A trained examiner visually compares known test shots to the recovered evidence using a microscope. Based on the toolmarks, either striated, impressed or both, the examiner makes his conclusions.

While the aforementioned examination practice and the resultant forensic evidence have been used extensively in legal proceedings, concerns have been raised regarding the reliability and validity of toolmark-based forensic evidence. For example, in the two comprehensive reports published by NRC [13, 12], it was indicated that no framework exists to describe the uncertainty in judgment when linking an evidence to a specific firearm.

Unlike biometrics [9] or other computer vision applications, there is less concentrated research work in ballistic image matching. Gerules et al. [8] present an overview of the image processing, toolmark identification, and matching approaches. Two research groups, in particular, have obtained impressive results on 3D topographical measurements. Scientists from NIST proposed the Congruent Matching Cells (CMC) method [16, 18, 4]. Ravi and Champod propose three similarity scores for aligned 3D topographies [15]. The authors in [3] fuse the non-visual knowledge into the image matching procedure. However, in most

prior work, the focus was more on low-level image processing while less on learning, and the performance was not evaluated on real-world operational ballistic data. Our goal is to develop a learning-based approach towards filling a gap between research and practice. For this work, BF impression in 2D reflectance microscopy images are selected as our target toolmark because it is the most frequently examined ballistic toolmark in forensic laboratories.

NIST released the first open-access database of ballistics toolmarks (Table 1). It includes data from four different studies, all captured under controlled laboratory environments. The number of bullets for which ballistic images are available in the NIST databases vary from 30 to 90. In contrast, the Operational Forensic Lab (OFL) database that we collected in collaboration with Michigan State Forensic Lab is the first one collected in an operational forensic facility and consists of ballistic images for 146 fired cartridge cases; it will be released in the public domain.

3. Ballistic Image Matching Algorithms

This section presents the details of comparing two ballistic images with breech face impressions (BFI). We start with the pre-processing step for the ballistic images (Fig. 3). We then present existing methods for comparing ballistics images. Finally, we describe the proposed approach for globally registering an image pair and learning a matcher.

3.1. Pre-Processing Pipeline

Since the BFI occupies only a small region on the back of the cartridge, it is important to segment this area of interest from the remainder of the cartridge image. Not only will this help the algorithmic efficiency by avoiding computation on the irrelevant parts of the image, it will also increase the accuracy since the lettering and other parts of the cartridge may differ even for the genuine match. Although an automated BFI segmenter (i.e., detecting two circles) could be learned, since a firearm examiner is already manually capturing the microscope imagery, for better accuracy we rely on three points, marked by the examiner, on each of the two circles, to define the boundary of the BFI region.

We observe that there is an overall planar *bias* in the image intensity across the BFI region, for two potential reasons: (i) the cartridge surface may not be in a perfect par-

allel plane with the imaging plane of the camera, especially given the high magnification of the microscope, and (ii) a point light source may introduce additional intensity bias in the spatial domain. This bias can dominate the image matching over the BFI. To address this, we fit each image $\mathbf{I}(u, v)$ by an optimal planar function in the least-squares sense, i.e., $\operatorname{argmin}_{\mathbf{a}} \sum_{u,v} \|(u, v, 1)\mathbf{a} - \mathbf{I}(u, v)\|_2^2$, where $\mathbf{a}$ is a 3-dim vector defining the plane. After fitting, the compensated image is $\mathbf{I}' = \mathbf{I} - (u, v, 1)\mathbf{a}$.

During the 2D reflectance microscope imaging, the specular reflections or shadows may result in *outliers* in certain local pixels, whose intensities are very different from their neighborhood. These outliers inevitably have a negative impact on the score computation. Hence, we define a pixel as an outlier when its intensity is three times the standard deviation from the mean of its 21×21 pixel local patch, and replace it with the interpolation of neighboring pixels.

Finally, all available images are filtered in order to emphasize the discriminative features of the firearms. Since the imaging procedure may inevitably create noise in the image, a low-pass Gaussian filter is used to remove the high-frequency noise. On the other hand, there are low-frequency components that are shared across all images regardless of which firearm fired the bullet. To attenuate these non-discriminative characteristics, we also apply a high-pass Gaussian filter to each image.

3.2. NIST Techniques for Ballistic Matching

Given the pre-processed images, we now present the two related approaches that were proposed by NIST and are considered as the state of the art in ballistic image matching.

3.2.1 Global CCF Approach

The widely accepted distance metric for matching BFIs is based on cross-correlation proposed by NIST [17]. Since then it has been used in other ballistic image matching techniques [18, 20]. The normalized cross-correlation function of a reference image $\mathbf{I}_1$ and a probe image $\mathbf{I}_2$ is defined as:

$$\text{CCF}(\mathbf{I}_1, \mathbf{I}_2, d_u, d_v) = \frac{\sum_{u,v} \mathbf{I}_1(u, v)\mathbf{I}_2(u + d_u, v + d_v)}{\sqrt{\sum_{u,v} \mathbf{I}_1(u, v)^2}\sqrt{\sum_{u,v} \mathbf{I}_2(u, v)^2}},$$
(1)

where d_u and d_v represent the translation of the probe image. The CCF score ranges from -1 to 1, where 1 indicates a perfect match of two images, 0 would arise from two random images, and -1 from inverse images.

To match a probe image with a reference image, the probe is rotated across a range of angles θ, and shifted by d_u and d_v pixels at each rotation. The optimal values of θ, d_u and d_v are found when the CCF score is maximized. Since all images are typically captured under the same magnification of the microscope, no scale variation is present in the data. This brute force procedure of finding optimal matching seems computationally intensive, but
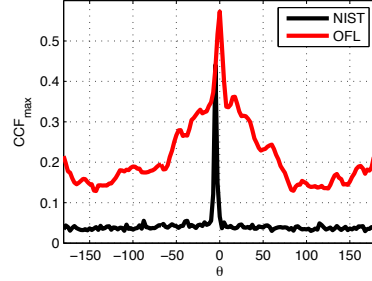

Figure 4. Example CCF$_{\text{max}}$ searches for the best registration angle. The NIST dataset example has a sharp peak at $\theta = -5$ where the probe and reference images best align. The OFL dataset is more challenging, but still displays a clear peak at $\theta = 0$.

fortunately there are techniques for an efficient implementation. First, similar to convolution, cross-correlation satisfies $\mathcal{F}\{\mathbf{I}_1 \star \mathbf{I}_2\} = (\mathcal{F}\{\mathbf{I}_1\})^* \cdot \mathcal{F}\{\mathbf{I}_2\}$, where $\mathcal{F}$ denotes the Fourier transform and an asterisk indicates the complex conjugate. This means all possible translations can be evaluated efficiently at once via the Fast Fourier Transform (FFT). Note that in the FFT computation, both images must be padded with zeros to avoid the affects of circular cross-correlation. Second, we use a coarse-to-fine search for θ (see Fig. 4). Initially, we search half-resolution images with $\theta \in \pm 180°$ in $2.5°$ increments. We then refine the search using the full-resolution images at $0.5°$ increments within $\pm 2.5°$ from the coarse optimum.

This global approach is able to perfectly distinguish genuine matches from false matches in a recent evaluation on the 3D topological BF datasets [20]. We will demonstrate its effectiveness on 2D reflectance microscopy images.

3.2.2 Congruent Matching Cells (CMC) Approach

One limitation of the global approach is that the entire BFI region may not be useful for comparing two ballistic images. This is because the FP may strike at different positions or certain areas of the primer may experience more deformation creating a stronger mark. Therefore, the CMC method was proposed to improve the correlation accuracy by first identifying invalid correlation areas and then discarding them from further consideration [4].

Instead of producing a single CCF score, the CMC method computes a number of *valid* matching cells as follows. The probe image is divided into a grid of 7×7 non-overlapping cells. Due to the ring-like shape of the BFI region, many cells are immediately discarded because of an insufficient number of BFI pixels. Each remaining cell is then rotated and translated across the reference image to identify its matching cell by maximizing the cross-correlation function CCF$_{\text{max}}$. If CCF$_{\text{max}} < T_{\text{CCF}}$, the cell is discarded since no high correlation match is found, where T_{CCF} is a preset threshold. Finally, we select the largest set of remaining cells whose variances of θ, d_u and d_v estimates are less than preset thresholds T_θ, T_u, and T_v, re-

spectively. The cardinality of this set is the CMC score.

Intuitively, if the two BFIs are from a genuine match (i.e., a pair of cartridge cases originated from the same firearm), there will be a large number of valid matching cells. While for a false match, their estimated parameter values CCF_{max} will be low, and even if cells find individual matches in the reference image, their θ, d_u and d_v values will not be consistent. NIST determined appropriate thresholds by selecting values that best separate the known genuine scores from known false scores.

3.3. Proposed Approach

We propose a learning-based approach to identify specific local parts of the BFI region that are most discriminative for matching. While a global approach may decrease the genuine scores due to the inclusion of low-correlation local regions in the score calculation, we also recognize a number of limitations of the local approach, namely CMC. First, CMC forces each cell to make equal contribution to the similarity score via hard thresholds (T_{CCF}, T_θ, T_u, T_v), despite that cells with high correlation should have a stronger influence in computing the score. Second, the requirement of *sufficient* data in a CMC cell arises due to the ring-like BFI being broken into square cells. We apply a polar transformation to the image to keep the cells within the confines of the valid BFI. Third, there is no *global registration* of the images in CMC, and therefore, the same pair may have *different* similarity scores under different initial rotations since the cells will be in different relative positions. We desire a global registration and the resultant score that is independent of the cartridge rotation by the examiner under the microscope. Such a global registration is also crucial for developing a learning-based approach.

3.3.1 Unrolled Image and Global Registration

As stated above, we seek to transform the images so that rectangular cells have more valid BFI regions as well as create a global registration among different image pairs. To achieve both these goals, motivated by the classic iris code [5], we generate an *unrolled image* by applying a polar transformation. To convert a Cartesian BFI image $\mathbf{I}$ to a polar BFI image $\mathbf{P}$, we use the polar transformation,

$$\mathbf{P}(r, \theta) = \mathbf{I}(r\cos(\theta) + u_c, r\sin(\theta) + v_c), \qquad (2)$$

where (u_c, v_c) is the BFI region center. The center of the reference image is simply the center of the annotated BFI circle (circle fitted to the three annotated points). For the probe image, we first register it to the reference through the global correlation approach. Before converting it to polar coordinates, we rotate it by the optimal θ value and translate it by d_u and d_v. The height of the unrolled image is equal to the radius of the BFI and the width is equal to the circumference. We call the polar image the unrolled image since it unrolls the circular BFI into a rectangle.
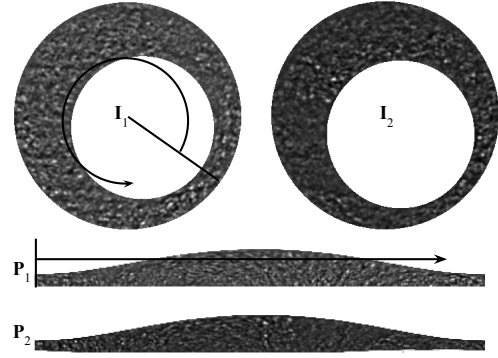

Figure 5. The process of creating unrolled images. A probe $\mathbf{I}_2$ is aligned to a reference $\mathbf{I}_1$ in Cartesian space. Both images are then transformed into polar space $\mathbf{P}_1$ and $\mathbf{P}_2$, and the line connecting the BF and FP centers on the reference is defined as $\theta = 0$.

At this point, both the probe and reference images are aligned with each other in the polar coordinates, but there is no global registration between arbitrary pairs; a different initial rotation of the reference image would produce a different unrolled image. To address this issue, we note that the FPI is off center on the BFI. Typically the side with the FPI has only a small strip of valid BFI while the other side has a large swatch of valid BFI. We leverage this observation to globally align the reference and probe by circularly shifting the reference to have the smallest usable part on the left boundary, which makes the largest usable part in the center of the unrolled image. The probe is shifted by the same amount as the reference to maintain the alignment within the pair. Fig. 5 presents a visual demonstration of the unrolling and global registration.

3.3.2 GentleBoost-based Classifier Learning

Given the unrolled images of the reference and probe, different local cells may have different abilities to separate genuine matches from false matches. Hence it is essential to select local cells that are most discriminative in this two-class classification. We use a boosting-based method since it merges feature selection and classifier learning in a single step. While different variants of boosting have been proposed [11], we use the GentleBoost algorithm [7] for two reasons. First, unlike the common AdaBoost [6], the weak classifier in GentleBoost is a soft classifier with continuous output. This addresses the issue of hard thresholding in CMC. Second, GentleBoost is superior to other boosting variants in object detection due to its robustness to noisy data and resistance to outliers [10].

GentleBoost iteratively selects weak classifiers to form a strong classifier using summation: $F = \sum_{m=1}^{M} f_m$, as shown in Algorithm 1, where $F(x)$ is the strong classifier and $f_m(x), m = 1, ..., M$ are the M weak classifiers. Given a ballistics database, the unrolled images of genuine matches are the positive samples with labels $y_i = 1$, and

---

Algorithm 1: The GentleBoost-based matcher learning.

Input: Training data $\{\mathbf{P}_1^i, \mathbf{P}_2^i; i = 1, 2, ..., K\}$ and their
corresponding class labels $\{y_i; i = 1, 2, ..., K\}$.

Output: A strong classifier $F(\mathbf{P}_1, \mathbf{P}_2)$.

1. Initialize weights $w_i = \frac{1}{K}$, and $F(\mathbf{P}_1, \mathbf{P}_2) = 0$.
2. **for** $m = 1, 2, ..., M$ **do**

    (a) Fit the regression function $f_m(\mathbf{P}_1, \mathbf{P}_2)$ by weighted least-squares (LS) of y_i to $\mathbf{P}_1^i, \mathbf{P}_2^i$:

$$f_m(\mathbf{P}_1, \mathbf{P}_2) = \operatorname*{argmin}_{r \in \mathcal{R}} \epsilon(f), \qquad (3)$$

    where $\epsilon(f) = \sum_{i=1}^K w_i(y_i - f(\mathbf{P}_1^i, \mathbf{P}_2^i))^2$.

    (b) Update $F(\mathbf{P}_1, \mathbf{P}_2) = F(\mathbf{P}_1, \mathbf{P}_2) + f_m(\mathbf{P}_1, \mathbf{P}_2)$.

    (c) Update $w_i = w_i e^{-y_i f_m(\mathbf{P}_1^i, \mathbf{P}_2^i)}$.

    (d) Normalize the weights such that $\sum_{i=1}^K w_i = 1$.

end

3. Output the classifier $F(\mathbf{P}_1, \mathbf{P}_2) = \sum_{m=1}^M f_m(\mathbf{P}_1, \mathbf{P}_2)$.

---

those corresponding to the false matches are negative samples with labels $y_i = -1$. Since there are substantially more negative than positive samples, we apply a horizontal shift to the unrolled probe within ± 2 pixels in 1 pixel increments, i.e., one genuine match will create five positive samples.

The key in designing a boosted classifier is the set of M weak classifiers. Given that the CCF has demonstrated its effectiveness in prior work, we decide to use CCF to construct our weak classifiers, defined as:

$$f_m(\mathbf{P}_1, \mathbf{P}_2) = \frac{2}{\pi}\operatorname{atan}(g(\mathrm{CCF}(\mathbf{P}_1(r_m), \mathbf{P}_2(r_m), 0, 0) - t_m)),$$
$$(4)$$

where t_m is a two-class threshold, atan() function serves the purpose of generating a continuous output within $[-1, 1]$, g is a constant that controls the smoothness of atan(), and r_m specifies the location and size of one local cell. Basically, this is a regression function that converts the CCF score $\mathrm{CCF}(\mathbf{P}_1(r_m), \mathbf{P}_2(r_m), 0, 0)$ of a local cell r_m to a value within $[-1, 1]$. Since global correlation has been used to register two images, we set d_u and d_v to be zero for efficient CCF computation. Note that the collection of all local cells at various locations and of different sizes forms the hypothesis space $\mathcal{R}$. During Step 2(a), for each r within $\mathcal{R}$, we calculate the CCF scores for all positive and negative pairs, estimate the optimal threshold t_m by binary search, and compute the weighted error $\epsilon(f)$. The local cell r with the minimum $\epsilon(f)$ is chosen as the mth weak classifier. After learning is completed, our classifier can be compactly represented as a $M \times 5$ matrix, $\{r_m, t_m\}_{m=1}^M$.

4. Experiments

4.1. Datasets

Table 1 shows the dataset collected in the Weller study of the NIST Ballistics Toolmark Database. It used a Ruger P-series 9mm caliber pistol with 10 consecutively manufactured slides, and images were captured via a Leica FS M comparison microscope. Details about the gun and manufacturing process may be found in [20]. This dataset consists of both 3D topographical maps as well as 2D imagery. We use the 2D images since most forensic examiners do not currently acquire 3D maps, and 2D image comparison is more challenging than 3D maps. There are 5 test fires from each slide; we use the side light images given their proven performance for BF comparisons.

We also present a new operational forensic laboratory (OFL) dataset collected to reflect the real-world practices in a forensics laboratory. A Leica FS C comparison microscope was used, equipped with a Leica model DFC320 digital camera. OFL consists of BFI images of 2 shots from 73 Glock semiautomatic handguns in 9mm Luger, .357 SIG, .40 S&W and .45 Auto caliber. Unlike the NIST dataset that has 10 slides of the same handgun, OFL has a variety of different guns. While the different guns may appear to make the matching task easier, the imaging conditions in OFL are more typical of challenging conditions that a forensic examiner may encounter. The BFI area of both NIST and OFL datasets has a resolution of $1,500 \times 1,500$ pixels.

4.2. Experimental Setup

We evaluate four different approaches in this work: 1) the global CCF approach outlined in Sec. 3.2.1, 2) the state-of-the-art CMC approach, 3) the global CCF approach applied to the unrolled images, and 4) the proposed learning-based approach. Due to the small number of available images in the two datasets, we use 50%-hold-out evaluation for all experiments where we randomly select half of the available *guns* for training while keeping the other half for testing. The hold-out procedure was repeated five times.

The NIST dataset has been shown to achieve perfect classification using both the CCF method on the 3D topographical maps, and the CMC method on both the 3D and raw 2D images [4, 18]. We implemented both the CCF and CMC methods and evaluated them on the 2D side light images. We use the same experimental parameters for CMC as in [18]: $T_{\mathrm{CCF}} = 0.25$, $T_\theta = 3°$, and $T_u = T_v = 150$ μm. To evaluate the results on the NIST dataset, which is known to be separable, we show histograms of the genuine and false scores. On each test set, there are 100 genuine matches and 500 false matches.

For the more challenging OFL dataset, we only run the global CCF approach and the proposed approach, since we find that the global CCF performs better than CMC and takes substantially less time to evaluate. Because the OFL dataset actually produces classification errors, we show the performance using the Receiver Operating Characteristic (ROC) curve, a plot of False Positive Rate (FPR) vs. the True Positive Rate (TPR). On each test set, there are 72 genuine matches and $1,440$ false matches.

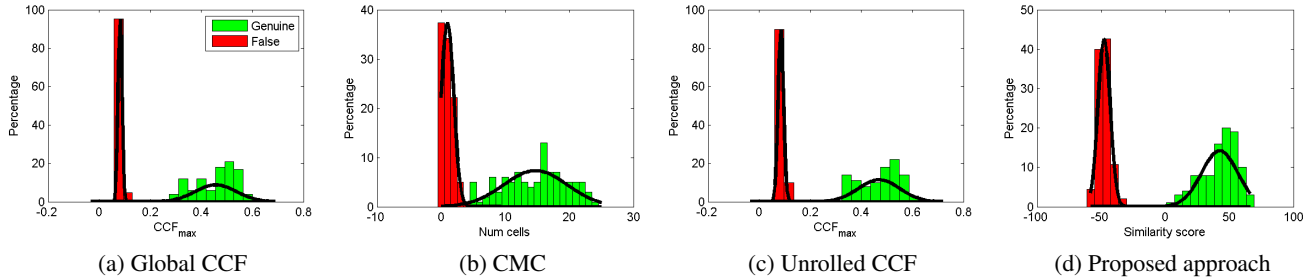(a) Global CCF   (b) CMC   (c) Unrolled CCF   (d) Proposed approach

Figure 6. Comparison of score histograms of four approaches on the NIST dataset, overlaid with fitted Gaussian distributions.

Table 2. Mean and s.d. of the expected errors on the NIST dataset.

Approach	Mean	S.d.
Global CCF	$1.2e^{-4}$	$1.3e^{-4}$
CMC	$3.4e^{-2}$	$1.7e^{-2}$
Unrolled CCF	$1.7e^{-4}$	$1.8e^{-4}$
Proposed approach	$2.3e^{-6}$	$2.8e^{-6}$

4.3. Results

NIST dataset Fig. 6 compares the results of the four approaches on the NIST dataset. Even though the NIST techniques were developed for topographical data, they still maintain perfect separation on the 2D images, but the similarity scores for the genuine pairs are lower than the ones reported on the 3D data. The proposed learning-based approach also achieves perfect classification. In order to quantify the error, we fit Gaussian distributions to the score histograms, and use the overlapping area of two distributions as the expected errors of these approaches. With five hold-out evaluations, the means and standard deviations of the expected errors are reported in Table 2.

There are a few things to note about these results. First, the unrolled CCF method performs nearly identical to the global CCF method on the Cartesian images, which indicates that the unrolling process does not substantially effect the performance but gives us the benefit of alignment among image pairs. Second, the CMC method appears to perform the worst of any of the reported methods on the 2D images. Third, the fact that NIST techniques achieve excellent results indicates we have a good-quality implementation of pre-processing steps and NIST techniques. Fourth, despite using a relatively small dataset, our proposed algorithm has a substantial lower expected error compared to the NIST techniques, which can also be observed from the wider separation of two histograms in Fig. 6 (d).

OFL dataset The OFL dataset contains real-world variations in ballistic images due to gun types and lighting conditions. As such, we expect the performance to be worse on OFL. Fig. 7 reports the ROC curves comparing the proposed method with the global CCF method. The proposed method performs better, especially at typical operating points of low FPR. The performance comparison of OFL vs. NIST datasets indicates that ballistic image matching on operational data is very challenging, and warrants future research.
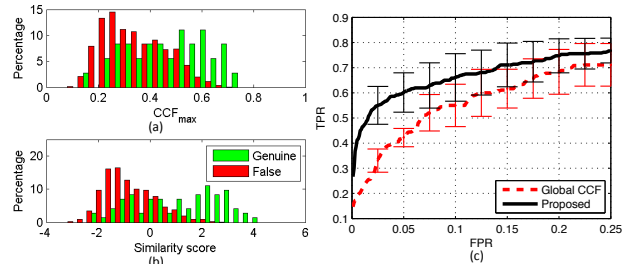

Figure 7. Results on OFL: (a) histogram of global CCF scores, (b) histogram of scores from the proposed approach, and (c) comparison in terms of ROC, based on five hold-out evaluations.
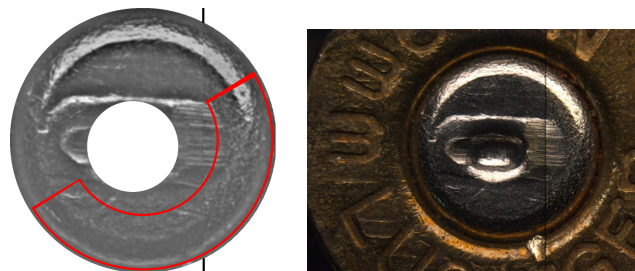

Figure 8. The same two OFL images with automatic alignment by cross-correlation (left) and with manual alignment by a forensic examiner (right). Each image is comprised of two images stitched together along a vertical seam. The top selected cell from the proposed approach is overlayed on the left.

Fig. 8 shows examples of aligned two OFL images.

We hypothesize that a learning-based method could identify local regions of the BFI that are highly discriminative. In Fig. 9 we overlay the selected cells onto sample BFI images. The selected cells tend to avoid the FPI and boundary of the BFI while concentrating on the valid portion of the BFI with a variety of different size regions.

Boosting parameters There are two parameters that can be set in the proposed method: the number of weak classifiers, M and the weak classifier smoothness parameter, g. For selecting M, we plot the expected error in classifying the training samples in Fig. 10 and select M to avoid overfitting. The performance on the NIST dataset saturates with no negative impact and we select $M = 100$. But for OFL, which has fewer genuine matches, we select a small number of weak classifiers, $M = 15$, since the performance on
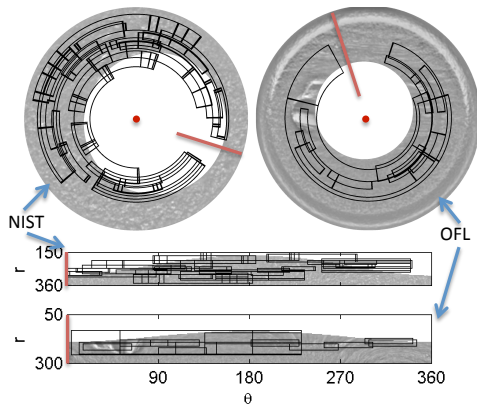
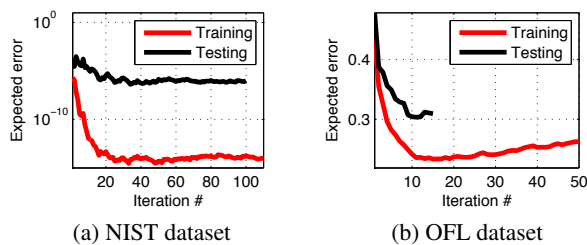Figure 9. Boosting-selected features/cells for NIST and OFL datasets displayed on Cartesian and unrolled images.

(a) NIST dataset   (b) OFL dataset

Figure 10. Expected error per boosting iteration on two datasets.

the training set decreases with larger M. For g, we select a value that produces the lowest expected error on the training set, which is $g = 5$ for NIST, and $g = 1$ for OFL.

The potential rectangular cells are set to the full size of the unrolled image, $\frac{1}{2}$ size, $\frac{1}{4}$ size, and $\frac{1}{8}$ size placed with an offset of 10 pixels from each other. There are a total of $7,973$ and $7,788$ potential features in the hypothesis spaces of NIST and OFL datasets, respectively.

Computational cost All algorithms are implemented in Matlab and run on a standard PC with an Intel i7-4770s 3.1GHz CPU. The time to pre-process each image is $\sim$4s. Comparing two images with the global CCF method takes $\sim$7s. The CMC method takes the longest time at $\sim$35s. The proposed method requires alignment via the global CCF method, but takes less than a second to extract features and compute the similarity score, with a total time of $\sim$8s.

5. Conclusions

We present a learning-based classifier for ballistic image matching of breech face impressions. The proposed approach is compared against state-of-the-art methods on both the NIST ballistics toolmark database and a newly collected operational forensic lab (OFL) dataset. While promising performances are observed on the controlled laboratory data available from NIST, our results show that additional research efforts are required for large-scale operational ballistic imaging matching. Our research also highlights the need for a large operational benchmark ballistic image database to develop probabilistic models for ballistics matching to address concerns about the "uncertainty in judgment".

References

[1] http://www.ultra-forensictechnology.com.
[2] http://www.nist.gov/forensics/ballisticsdb.
[3] A. Biedermann and F. Taroni. A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues. *Forensic Science International*, 163(1):18–33, 2006.
[4] W. Chu, M. Tong, and J. Song. Validation tests for the congruent matching cells (CMC) method using cartridge cases fired with consecutively manufactured pistol slides. *AFTE Journal*, 45(4):361–366, Jan. 2013.
[5] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE T-PAMI*, 15(11):1148–1161, 1993.
[6] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
[7] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 38(2):337–374, April 2000.
[8] G. Gerules, S. Bhatia, and D. Jackson. A survey of image processing techniques and statistics for ballistic specimens in forensic science. *Science and Justice*, 53(2):236–250, 2013.
[9] A. K. Jain and A. Ross. Bridging the gap: from biometrics to forensics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1674), 2015.
[10] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *Proc. 25th Pattern Recognition Symposium, Madgeburg, Germany*, pages 297–304, September 2003.
[11] R. Meir and G. Rätsch. *An introduction to boosting and leveraging*. S. Mendelson and A. Smola, Editors, Advanced Lectures on Machine Learning, LNAI 2600. Springer, 2003.
[12] National Research Council. *Ballistic Imaging*. The National Academies Press, Washington, DC, 2008.
[13] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, DC, 2009.
[14] M. Planty and J. Truman. Firearm violence, 2993-2011. Technical Report NCF 241730, U.S Department of Justice, May 2013.
[15] F. Riva and C. Champod. Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences*, 59(3):637–647, 2014.
[16] J.-F. Song. Proposed NIST ballistics identification system (NBIS) based on 3D topography measurements on correlation cells. *AFTE Journal*, 45(2):184–194, 2013.
[17] J.-F. Song and T. V. Vorburger. Proposed bullet signature comparisons using autocorrelation functions. In *NCSL*, Toronto, Canada, July 2000.
[18] M. Tong, J. Song, W. Chu, and R. M. Thompson. Fired cartridge case identification using optical images and the congruent matching cells (CMC) method. *Journal of Research of the National Institute of Standards and Technology*, 119:575–582, 2014.
[19] T. Vorburger, J. Yen, B. Bachrach, T. Renegar, J. Filliben, L. Ma, H. Rhee, A. Zheng, J. Song, M. Riley, C. Foreman, and S. Ballou. Surface topography analysis for a feasibility assessment of a national ballistics imaging database. Technical Report 7362, NISTIR, 2007.
[20] T. Weller, A. Zheng, R. Thompson, and F. Tulleners. Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides. *Journal of Forensic Sciences*, 57(4):912–917, 2012.