Forensic Science International: Synergy 5 (2022) 100272

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: www.sciencedirect.com/journal/forensic-science-international-synergy

Forensic comparison of fired cartridge cases: Feature-extraction methods for feature-based calculation of likelihood ratios

Nabanita Basu [a], Rachel S. Bolton-King [a,b], Geoffrey Stewart Morrison [a,c,*]

[a] Forensic Data Science Laboratory, Aston University, Birmingham, UK
[b] School of Justice, Security & Sustainability, Department of Society, Crime & Environment, Staffordshire University, Stoke-on-Trent, UK
[c] Forensic Evaluation Ltd, Birmingham, UK

ARTICLE INFO

Keywords:
Firearm
Cartridge case
Likelihood ratio
Feature
Calibration
Validation

ABSTRACT

We describe and validate a feature-based system for calculation of likelihood ratios from 3D digital images of fired cartridge cases. The system includes a database of 3D digital images of the bases of 10 cartridges fired per firearm from approximately 300 firearms of the same class (semi-automatic pistols that fire 9 mm diameter centre-fire Luger-type ammunition, and that have hemispherical firing pins and parallel breech-face marks). The images were captured using Evofinder®, an imaging system that is commonly used by operational forensic laboratories. A key component of the research reported is the comparison of different feature-extraction methods. Feature sets compared include those previously proposed in the literature, plus Zernike-moment based features. Comparisons are also made of using feature sets extracted from the firing-pin impression, from the breech-face region, and from the whole region of interest (firing-pin impression + breech-face region + flowback if present). Likelihood ratios are calculated using a statistical modelling pipeline that is standard in forensic voice comparison. Validation is conducted and results are assessed using validation procedures and validation metrics and graphics that are standard in forensic voice comparison.

1. Introduction

1.1. Outline

When firearms are fired at a crime scene and cartridge cases are ejected, these fired cartridge cases may later be recovered. Forensic practitioners may then compare two fired cartridge cases recovered from the crime scene with each other – a comparison of a fired cartridge case which bears markings of questioned source with another fired cartridge case which bears markings of questioned source (hereinafter we refer to this as "Scenario 1"). Forensic practitioners may also compare a fired cartridge case recovered from the crime scene with cartridge cases that they fire from a firearm seized from a suspect – a comparison of a fired cartridge case which bears markings of questioned source with fired cartridge cases which bear markings of known source (hereinafter we refer to this as "Scenario 2").

The evaluation in Scenario 1 could be conducted for investigative purposes, but could also be used for evidential purposes if no relevant firearms are available for comparison but the question of how many firearms were fired during the commission of a crime is relevant for legal decision making.

For simplicity, in the present paper we assume exactly two recovered cartridge cases in Scenario 1 and exactly one recovered cartridge case in Scenario 2. Real casework may involve larger numbers of recovered cartridge cases, but these can be dealt with via expansion or repetition of the methods described in the present paper.

For brevity, we will use the terms "questioned-source cartridge case" and "known-source cartridge case" as abbreviations for "cartridge case bearing marks of questioned source" and "cartridge case bearing marks of known source" respectively.

In the remainder of the introduction:

- We describe the anatomy of a fired cartridge case and the processes by which firearms leave marks on cartridge cases (§1.2).
- We describe current casework practice for comparison of fired cartridge cases (§1.3).
- We provide a summary of published research on feature-extraction methods and statistical-modelling methods that have previously been applied to forensic comparison of fired cartridge cases (§1.4).

* Corresponding author. Forensic Data Science Laboratory, Aston University, Birmingham, UK.
E-mail address: geoff-morrison@forensic-evaluation.net (G.S. Morrison).

https://doi.org/10.1016/j.fsisyn.2022.100272
Received 12 February 2022; Received in revised form 21 May 2022; Accepted 24 May 2022
Available online 27 May 2022

In the remainder of the paper:

- We describe the hypotheses, including the specification of the relevant population, that we have adopted for calculating likelihood ratios in the context of the present research (§2).
- We describe a feature-based system that we have developed for calculation of likelihood ratios from images of fired cartridge cases (§3). The system includes:
  - o a database of 3D digital images of the bases of fired cartridge cases (§3.2)
  - preprocessing of images (§3.3)
  - feature-extraction methods (§3.4)
  - a statistical modelling pipeline that calculates likelihood ratios (§3.5)
- We describe validation procedures (§4), and present and discuss the validation results (§5 and §6).

The focus of the present paper is on comparing the performance of different feature-extraction methods. The best-performing feature-extraction method will be used in planned future research using a larger database and Deep Neural Network (DNN) embeddings.

The research reported in the present paper is part of a wider programme of research which is outlined in Morrison [1].

1.2. Anatomy of a fired cartridge case

Fig. 1 shows an example of an image of the base of a fired cartridge case. The *head-stamp region* includes text indicating the manufacturer and calibre of the cartridge case. We assume that this information is factual, and that it narrows the "class" of the cartridge case without the need for interpretation. The other regions together constitute the *region of interest* and each of these regions is italicized on its first mention in the paragraph below.

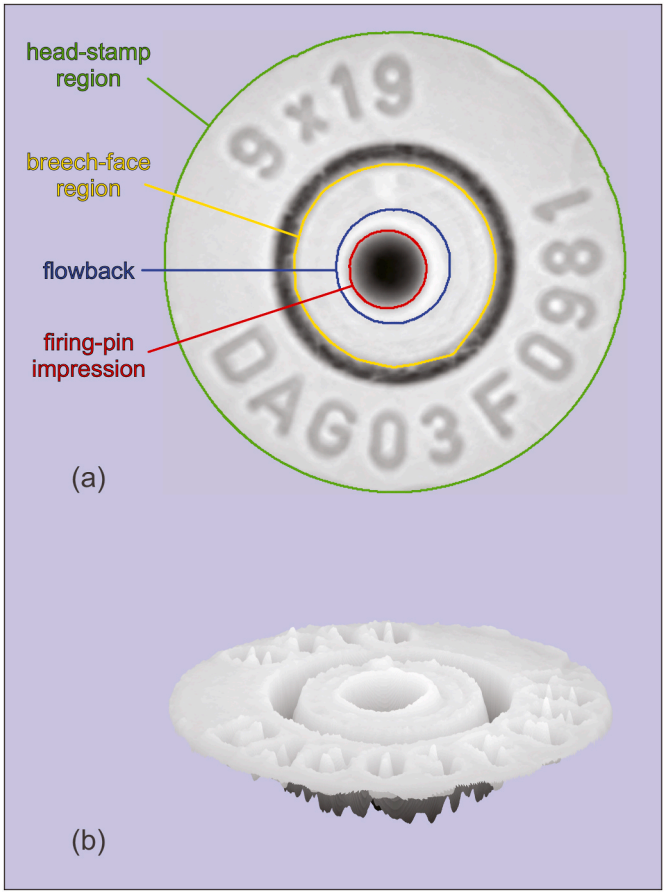

Fig. 1. Graphical representations of an example of the base of a fired cartridge case (9 mm diameter Luger-type ammunition). (a) Perpendicular view. (b) Oblique view with z scale exaggerated by a factor of 5.

An unfired cartridge of ammunition consists of a cartridge case, a bullet, and explosives. The cartridge case is a metal tube that is sealed at its base and is plugged at the other end (its mouth) by the bullet. Between the base of the cartridge case and the bullet are explosives. A cartridge is loaded into the chamber of a firearm. When the firearm is fired, the firing pin strikes the primer cup on the base of the cartridge case. This deforms the primer cup creating a concave *firing-pin impression*.[1] This kinetic action initiates an explosion within the cartridge case which forces the bullet forward out of the mouth of the cartridge case and along the barrel of the firearm. The explosion also forces the cartridge case backward until its base impacts the breech of the firearm.[2] This creates an impression of the breech face on the base of the cartridge case. The region of the base of the cartridge case where this impression is made is called the *breech-face region*. The explosion can also push outward the area around the firing-pin impression leading to convex deformation known as *flowback*. After the firearm has been fired, the cartridge case is (manually or automatically) ejected so that a new unfired cartridge can be loaded into the chamber. Typically, ejected cartridge cases fall to the ground, and they can potentially be recovered at a later time.

Breech faces are not perfectly smooth. They have irregularities due to the manufacturing process and potentially due to later wear or damage. These irregularities vary across firearms. A breech-face impression on the base of a fired cartridge case will reflect the irregularities of a breech face. The transferred patterns of irregularities can include, but are not limited to, parallel series of peaks and troughs. Differences in the irregularities of the breech faces of different firearms will cause variability in the breech-face impressions on cartridge cases fired from different firearms. Differences in the transfer of the irregularities of a breech face to cartridge cases will cause variability in the breech-face impressions on cartridge cases fired from the same firearm. Similarly, the location, shape, and surface details of firing-pin impressions can vary both across fires from the same firearm and across fires from different firearms.

Considering a firearm as the source of breech-face and firing-pin impressions, inferences with respect to which firearm fired a cartridge case can be drawn if the between-source variability in breech-face and firing-pin impressions is greater than their within-source variability.[3]

1.3. Current casework practice

For reviews of current casework practice in firearm examination see Bolton-King [2] and Nichols [3]. In current widespread practice, the analysis is a human-perception process and the interpretation of the extracted information is a subjective-judgement process. The forensic practitioner visually compares fired cartridge cases, viewing them side by side through a comparison microscope.[4] Properties that practitioners report taking into consideration include the position and shape of the firing-pin impression, and the heights, widths, and distances between parallel peaks and troughs on the firing-pin impression and on the breech-face region (Tobin & Blau [4]; Tai & Eddy [5]). Current

[1] For simplicity, we assume centre-fire cartridges. Some firearms use rim-fire cartridges, and a firing-pin impression appears on the edge of the base of the fired cartridge case rather than on a central primer cup.

[2] Breech designs vary, but a common design is for there to be a breech block, i.e., a block of metal, that halts the backward motion of the cartridge case.

[3] The design of many firearms allow firing pins and breech faces to be replaced, but for simplicity the present paper does not address scenarios involving such changes.

[4] A comparison microscope allows images of two different objects to be juxtaposed and rotated and aligned relative to one another.

widespread practice is to report the conclusion as a categorical decision, i.e., as "identification", "inconclusive", or "elimination" (or as "unsuitable for analysis"). Existing validation studies of practitioner performance have tended to use a small number of test trials, and have seldom reflected real casework conditions (Smith et al. [6]; Mattijssen et al. [7], [8]; Scurich et al. [9]).

For Scenario 2, practitioners typically fire 3 cartridges from the firearm of interest and compare the fired cartridge cases with the cartridge case recovered from the crime scene.

Images of fired cartridge cases for comparison with an image of a questioned-source cartridge case recovered from a crime scene may be selected via an automated database search. The automated search returns a set of candidates for comparison with the questioned-source cartridge case, i.e., cartridge-case images in the database that the automated system determines to be the most similar to the questioned-source cartridge-case image. Thereafter, the comparison between the questioned-source image and the known-source images in the candidate set becomes a variant of Scenario 1 or Scenario 2: the evaluation is conducted using the human-perception and human-judgement processes described above.

In a survey of practitioners presented in Scurich et al. [9], ~7% of respondents reported that a typical fired-cartridge-case comparison took less than 30 min, ~24% that it took 30–60 min, ~28% that it took 1–2 h, ~26% that it took 2–4 h, and ~15% that it took more than 4 h.

1.4. Previous work using data, quantitative measurements, and statistical models

1.4.1. Introduction

In this subsection, we summarize published research on quantitative-measurement and statistical-modelling methods that have been applied in forensic comparison of fired cartridge cases. We first describe existing databases of images of the bases of fired cartridge cases (§1.4.2). We then summarize feature-extraction methods (§1.4.3), and statistical models that have been applied to features and to similarity scores (§1.4.4). Finally, we summarize the results of research on practitioners' attitudes toward the use of statistical models (§1.4.5).

1.4.2. Databases

Data of interest consist of 2D or 3D digital images of cartridge-case bases. 2D photographic images capture reflected light. 3D images capture surface topography, including depth information. In the present paper we focus on 3D images. There are several commercially marketed 3D imaging systems. The two most commonly used in operational forensic laboratories are Evofinder® and IBIS®. Research using such systems has the advantage of potentially being more quickly applicable to casework.

Published research on statistical models for comparison of fired cartridge cases has made use of training and validation datasets that are relatively small. Some existing datasets consist of a large number of fires from a small number of firearms, e.g., 10–60 test-fires from each of 1–5 firearms (Thumwarin [10]; Liong et al. [11]; Ott et al. [12]; Addinall et al. [13]), and others consist of a small number of fires from a somewhat larger number of firearms, e.g., 1–4 test-fires from each of 10–90 firearms (Xin et al. [14]; Legrá et al. [15]; Fadul et al. [16]).[5] In addition, only a subset of the datasets used in published research have themselves been published and made available to other researchers and practitioners. Published datasets include those in the NIST Ballistics Toolmark Research Database (NBTRD).[6] Some of the more commonly used datasets are described in: Lightstone [19]; LaPorte [20]; Fadul et al. [16].

In order to train a forensic-evaluation system that outputs likelihood ratios, one has to model both within-source and between-source variability. In order to do this, a dataset would be needed that includes a relatively large number of fires from each of a relatively large number of firearms of the same class. Datasets with a large number of firearms consisting of a small number from each of multiple classes would not be suitable for addressing "individualization", as opposed to "class", questions. To our knowledge, there are no existing datasets accessible for research purposes that contain images of a sufficient number of cartridge cases fired from each of a sufficient number of firearms of the same class to satisfy our requirements for training and validating a likelihood-ratio system.

1.4.3. Feature extraction

In published research, features have typically been extracted from the firing-pin impression and from the breech-face region. Flowback has usually been excluded from analysis (Ott et al., [12]; Song et al., [21]). Many features have been based on quantifications of what forensic practitioners report they pay attention to (see §1.3), but others have been based on functions fitted to image data without regard for interpretability of those features by humans. We will refer to the former as "human-inspired features" and the latter as "functional features".

Human-inspired features that have been extracted from firing-pin impressions include those based on the impression's location (Legrá et al. [15]), overall shape (Zhou et al. [22]; Li [23]; Thumwarin et al. [10]), and surface texture (Legrá et al. [15]). Human-inspired features that have been extracted from the breech-face region include those based on low-frequency undulations of parallel peaks and troughs (Gambino et al. [23]; Petraco et al. [24]), in the literature this is termed "waviness", and higher-frequency irregularities/residuals in those undulations (Petraco et al. [25]; Pan et al. [26]), in the literature this is termed "roughness". Most of these features have been extracted from manually-selected parts of the firing-pin impression or of the breech-face region.

Functional features that have been extracted from firing-pin impressions include values of central geometric moments (Ghani et al., [27]) and of Legendre moments (Chuan et al., [28]). From the whole cartridge-case base (including the headstamp region), Leng & Huang [29] extracted as features the values of circle-moment invariants (a modified version of central moments). From the whole region of interest (firing-pin impression + any flowback + breech-face region), Thumwarin et al. [10] extracted as features the magnitude-coefficient values from Fourier series fitted independently to each member of a set of concentric circles.

1.4.4. Statistical models

Statistical models applied in the published research have primarily been classification models rather then likelihood-ratio models. These classification models have included k nearest neighbors (Fischer & Vielhauer [30], [31]; Morris et al. [32]), linear discriminant analysis (Thumwarin et al. [10]; Ghani et al. [27]; Chuan et al. [28]), support vector machines (Zhou et al., [22]), bagged decision trees (Morris et al. [32]), and neural networks (Li [33]; Leng & Huang [29]; Morris et al. [32]; Ghani et al. [34]; Giudice et al. [35]; Razak et al. [36]).

Other statistical models used for classification or for database search have skipped extraction of features and have been based on similarity scores calculated as the correlation between pairs of digital images, i.e., the correlation between the z values (the intensities for 2D images, or the heights for 3D images) at the corresponding x and y points of the two images. Similarity scores are calculated for pairs of cartridge cases known to come from the same source and for pairs of cartridge cases known to come from different sources, and statistical models are fitted to these two sets of scores (Roth et al. [37]; Song [38]; Ott et al. [12]; Tai & Eddy [5,39]; Zhang [40]). This approach has been applied to the whole of the firing-pin impression or the whole of the breech-face region (Song et al. [41]; Roth et al. [37]). Prior to calculating the correlation coefficient, the firing-pin impressions or breech-face regions from the two

[5] In Zhang & Luo [17], 3070 test fires were produced from a total of 5 firearms. In Law et al. [18], 100 test fires were produced from each of 30 firearms.
[6] https://tsapps.nist.gov/NRBTD.

cartridge cases must be registered (rotated and aligned) relative to each other or to a common target. Rather than calculating the correlation over the whole of the firing-pin impression or the whole of the breech-face region, a commonly used approach is "congruent matching cells" (CMC) which calculates correlations over smaller areas which are called "cells". The cells are, for example, squares of predetermined size defined by a grid superimposed on the image. Each cell from the questioned-source image is independently rotated and aligned relative to the known-source image in order to find the cell on the latter that is maximally correlated with the former.[7] If the maximum correlation coefficient achieved exceeds a predefined threshold, these are designated CMCs. The number of CMCs between a pair of fired cartridge cases can be used as a similarity score. Variants of the CMC approach are described in: Zhang et al. [42], [43]; Chen et al. [44]; Tong et al. [45], [46].

To our knowledge, there is no published research describing calculation of likelihood ratios using statistical models applied to features separately extracted from each cartridge case, but there are a number of papers that describe calculation of likelihood ratios based on similarity scores. The most commonly used similarity score has been a correlation coefficient between pairs of digital images, calculated over the whole of or selected portions of the firing-pin impression or of the breech-face region (Riva & Champod [47]; Dong et al. [48]; Mattijssen et al. [7]; Riva et al. [49]). Other similarity scores used have been based on Euclidian distance between pairs of digital images, and on instantaneous angles on the surfaces of pairs of 3D images (Riva & Champod [47]; Riva et al. [49]). The most commonly used models have fitted kernel density distributions (Riva & Champod [47]; Dong et al. [48]; Mattijssen et al. [7]; Riva et al. [49]). Song et al. [50] used counts of the number of CMCs as similarity scores and fitted beta-binomial models to the count data. Similarity scores, however, do not take account of typicality with respect to the relevant population, and are therefore not an appropriate basis for calculating meaningful likelihood ratios in a forensic context (Morrison & Enzinger [51]; Neumann & Ausdemore [52]; Neumann et al. [53]). In the present paper we will therefore describe and validate a feature-based system for calculation of likelihood ratios.

1.4.5. *Practitioners' attitudes toward the use of statistical models*

In a survey of practitioners presented in Scurich et al. [9], some respondents had skeptical (or even hostile) attitudes toward the use of statistical models for comparison of bullets and comparison of fired cartridge cases, but others had more positive attitudes. One of the respondents with a more positive attitude emphasized the need for developers of statistical models to have a thorough understanding of firearms examination, and another emphasized the need for improved performance and for larger databases.

2. Hypotheses and relevant population

2.1. Introduction

In this section, we restate the two casework scenarios of interest, and state the hypotheses that we have adopted with respect to each of these scenarios, including specifying the relevant population. For both scenarios, the hypotheses define a common-source question.[8]

2.2. Scenario 1

One or more firearms are fired at a crime scene and the cartridge cases are ejected. Crime-scene investigators later recover two fired cartridge cases. A forensic practitioner compares the two questioned-source cartridge cases with one another and draws an inference with respect to whether they were fired by the same firearm or not.

H_s: The two cartridge cases were fired by the same firearm.
H_d: The two cartridge cases were fired by different firearms from the same population.

2.3. Scenario 2

A firearm is fired at a crime scene and the cartridge case is ejected. Crime-scene investigators later recover the fired cartridge case. Police investigators seize a firearm from a suspect. A forensic practitioner fires multiple cartridges from the seized firearm and collects the ejected cartridge cases. The forensic practitioner then compares the fired cartridge case recovered from the crime scene (the questioned-source cartridge case) with the cartridge cases fired from the suspect's firearm (the known-source cartridge cases) and draws an inference with respect to whether the questioned-source and known-source cartridge cases were fired by the same firearm or not.

H_s: The cartridge case bearing marks of questioned source and the multiple cartridge cases bearing marks of a single known source were fired by the same firearm.
H_d: The cartridge case bearing marks of questioned source and the multiple cartridge cases bearing marks of a single known source were fired by different firearms from the same population.

We will test two versions of Scenario 2, one in which the practitioner fires 3 cartridges from the seized firearm, and one in which they fire 9 cartridges.

2.4. Relevant population

In casework, the practitioner would first examine the questioned-source cartridge case in order to assess the class of firearms from which the cartridge case may have been fired. For the purposes of the research reported in the present paper, the relevant population of firearms that we have adopted is semi-automatic pistols that fire 9 mm diameter centre-fire Luger-type ammunition, and that have hemispherical firing pins and parallel breech-face marks. Examples of firearms in this class are Browning Hi-Power, CZ 75, Beretta 92FS, and Ruger P85. This particular class was chosen as the relevant population for the present research because it is commonly encountered in casework [55].[9]

The evaluation of the class of the firearm is generally considered to be the easiest step in the forensic comparison of fired cartridge cases due to gross differences in geometric form between classes (Bolton-King [2]; Nichols [3]). The present paper is not concerned with evaluation of class-level hypotheses.

[7] "Cells" on the known-source image can be of any orientation in any location and do not have to tessellate with each other.

[8] See Ommen & Saunders [54] on the distinction between specific-source and common-source likelihood ratios.

[9] It is not always the case that a firearm that has parallel breech-face marks will clearly transfer those marks to the breech-face region of the cartridge case. If a questioned-source cartridge case has clear parallel marks on its breech-face region, then the class of firearms can be restricted to those with parallel breech-face marks. If the questioned-source cartridge case does not have a clear pattern of marks on its breech-face region, then the class of firearms could be those with parallel breech-face marks, or with circular, cross-hatch, arc, or granular breech-face marks, or with smooth breech faces. In the present research, we have simply used cartridge cases fired from the class of firearms that have parallel breech-face marks without checking whether the cartridges playing the part of questioned-source cartridge cases actually have clear parallel marks. Including this step is something we leave for potential future research. Likewise adopting a broader population for cartridge cases including those without clear patterns of breech-face marks (and collecting data from that broader population) is something we leave for potential future research.

3. Fired-cartridge-case-comparison system

3.1. Introduction

In this section, we describe the system we have developed for feature-based calculation of likelihood ratios from images of fired cartridge cases. First, we describe the construction of a database of 3D images of fired cartridge cases that was used for training and validating the algorithmic stages of the system (§3.2), then we describe the algorithmic stages of the system (§3.3–§3.5).

The image-preprocessing, feature-extraction, and statistical-modelling stages of the system are outlined in Fig. 2. In the initial stages, information from a known-source cartridge case and information from a questioned-source cartridge case are processed in parallel. In the final stages, the known-source and questioned-source information are combined.

In the first stage, images are preprocessed prior to feature extraction (§3.3). In the next stage, feature vectors are extracted from the images. In §3.4, we provide details of the multiple feature-extraction methods that we have tested. These methods include those that have previously been proposed and applied in the research literature (see §1.4.3), plus an additional method (Zernike moments) that *a priori* we expected to be effective.

The last three stages in the system are: dimension reduction, calculation of uncalibrated likelihood ratios, and calibration. In §3.5, we describe the statistical models used in each of these stages. The use of this statistical modelling pipeline is standard for backend modelling in state-of-the-art forensic-voice-comparison systems (Morrison et al. [56], [57]; Weber et al. [58]).

Matlab® code implementing the algorithms described in §3.3–§3.5 is available from: https://forensic-data-science.net/firearms/.

3.2. Database

The data for the present research were taken from the E^3 Database of Fired Cartridge Cases (release 1), that we built as part of the present research. This database is available from NBTRD.[10] A link to the database is also provided at: https://forensic-data-science.net/firearms/.

This database consists of 3D images of the bases of cartridge cases fired from firearms that were in the possession of a number of operational forensic laboratories, law-enforcement agencies, military units, and private individuals in Barbados, Canada, France, Germany, UK, and USA. The cartridges used were taken from whatever each provider had available, with the condition that they have brass primer cups. 10 cartridge cases were fired from each firearm (on occasion, one or more of these were missing). The original aim was to collect 3D images of the bases of cartridge cases fired from 1000 firearms, but progress toward this target was slowed by the COVID 19 pandemic. We plan to continue building the database and, in the future, release additional data. The research reported in the present paper makes use of data from cartridge cases fired from 297 firearms. This was the number available after excluding any firearms for which we received fewer than 8 fired cartridge cases.

The bases of the fired cartridge cases were digitally imaged using Evofinder® (software version 6.6.1.17), which uses a mixture of photometric stereo imaging and focus variation to capture 3D surface topography. The base of each cartridge case was digitally imaged, and the resulting data were exported as a matrix of values $z(x,y)$ in x3p format[11] with a resolution of 280 samples per mm in each of the x direction and the y direction (3.6 μm between samples). The resolution in the z direction was able to capture differences in height of less than 1 μm.

For the present research, the dataset was divided into two parts using a $^2/_3$ versus $^1/_3$ split: Data from 198 firearms (hereinafter the "training set") were used to train all the models up to and including calculation of uncalibrated likelihood ratios, and data from the remaining 99 firearms (hereinafter the "calibration/validation set") were used for cross-validated training of the calibration model and for validation.

3.3. Preprocessing

Prior to feature extraction, we applied the following commonly-used preprocessing steps:

1. Segmentation: Separation of the firing-pin impression and the breech-face region from the remainder of the image and from each other.

2. Illumination correction: Correction for non-uniformities in illumination, including planar-bias correction.

3. Noise removal: Removal of imaging artifacts.

4. Registration.
   Rotation and alignment
   Details of commonly-used preprocessing procedures are provided in Tai & Eddy [5]. Preprocessing is not a focus of the present paper, so we do not provide details here. For segmentation, whereas Tai & Eddy [5] uses thresholds based on individual pixel values with predetermined threshold values, we used adaptive thresholds based on smoothed contours.[12]

The following regions were segmented:

(a) the whole of the region of interest including flowback if present
(b) the whole of the region of interest excluding flowback if present
(c) the firing-pin impression alone
(d) the breech-face region alone

Fig. 3 shows examples of each of these segmented regions.

Although flowback has usually been excluded from analysis (Ott et al., [12]; Song et al., [21]), we hypothesized that the flowback region would contain useful information related to the firearm that fired the cartridge.

The output of preprocessing were matrices of values $z(x,y)$ with a resolution of 56 samples per mm in each of the x direction and y the direction (the downsampling procedure included anti-aliasing low-pass filtering). Within each matrix, the x and y values were centred by subtracting their means (calculated over the whole of the segmented region), and were scaled such that the entire segmented region fell within a unit circle: $x^2 + y^2 \leq 1$. This resulted in x and y values in the range $-1 \ldots 1$ with 0 in the centre. z values that corresponded to x and y combinations that fell outside the segmented region did not contribute to the calculation of the feature values (these z values were coded in Matlab as "not a number, NaN"). z values were scaled in millimetres, and were shifted so that the origin (zero value) was set to the plane fitted to the breech-face region during planar-bias correction (planar-bias correction was derived from the breech-face region only and applied to each of the segmented regions).

Because of the preprocessing, all data matrices had the same scale and the same location. Some of the features extracted for the present

[10] https://tsapps.nist.gov/NRBTD/Studies/Studies/Details/a023199a-b9f3-4a1a-89e8-c94054a7cf61.

[11] ISO 25178–72:2017/AMD 1:2020 Geometrical product specifications (GPS) — Surface texture: Areal — Part 72: XML file format x3p — Amendment 1.

[12] We plan to publish details of these modified procedures elsewhere, along with comparisons of results of segmentation using the original Tai & Eddy [5] procedures and our modified procedures.

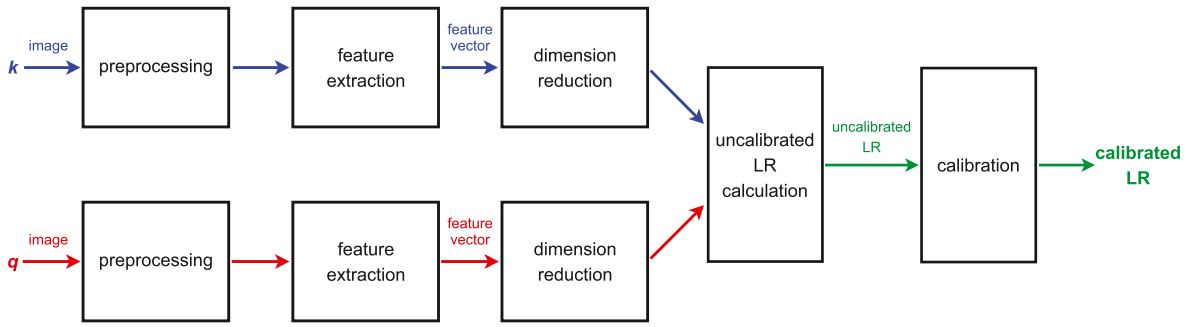

Fig. 2. Schematic of the feature-extraction and statistical-modelling stages of the system. Abbreviations: k = known source; q = questioned source; LR = likelihood ratio.

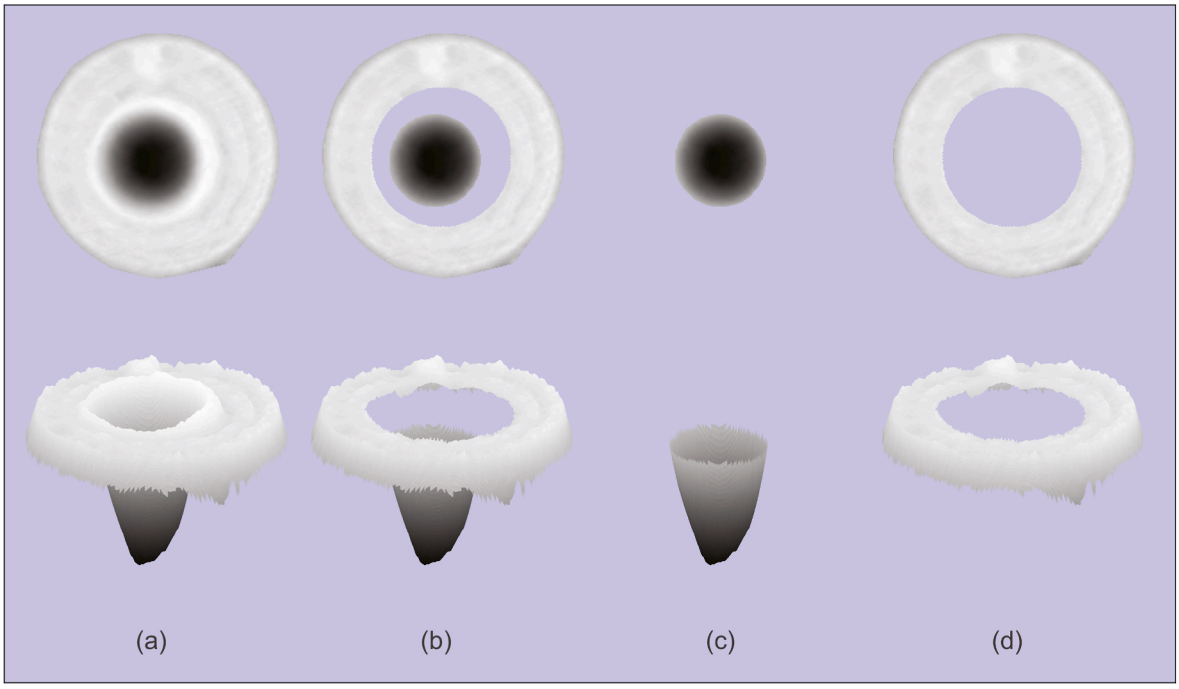

Fig. 3. Examples of segmented regions of a cartridge case (using the same example image as in Fig. 1): (a) whole region of interest including flowback; (b) whole region of interest excluding flowback; (c) firing-pin impression alone; (d) breech-face region alone. For the oblique views, the z scale is exaggerated by a factor of 5.

research are rotation invariant, but others are not. As part of pre-processing, we therefore rotated the data matrices. An ideal rotation procedure would use the questioned-source cartridge case in Scenario 2/one of the questioned-source cartridge cases in Scenario 1 as the target and rotate all other data matrices used for training and testing to that target. This, however, would require each data matrix in the entire dataset to be independently rotated to each questioned-source cartridge case/to each cartridge case used in validation as if it were a questioned-source cartridge case. This would be prohibitive in terms of processing time. We therefore arbitrarily selected one cartridge case from our dataset (the one shown in Fig. 1 and Fig. 3), rotated data from all other cartridge cases to this arbitrary target, then used this single rotated dataset for training and validation. The cost of rotation, especially ideal rotation, is a reason to prefer rotation-invariant features, but if rotation leads to substantial improvement in performance that cost may be justified.

3.4. Feature extraction

3.4.1. Introduction

We extracted and tested the same sets of functional features that have previously been proposed and applied in the published literature on forensic comparison of fired cartridge cases (see §1.4.3). We also extracted and tested Zernike moments (Zernike [59]; Teague [60]; Khotanzad & Hong [61]). Zernike moments have been widely used in many fields, including optometry, photonics, astronomy, and facial-expression analysis (e.g., Iskander et al. [62]; Sun et al., 2014 [63]; Pinhasi et al., [64]; Vretos et al. [65]). They are orthogonal and rotation invariant and have been found to outperform other moment-based approaches in terms of noise resilience, information redundancy, reconstruction capability, and classification accuracy (e.g., Teh & Chin [66]; Khotanzad & Hong [61]; Belkasim et al. [67]). We hypothesized that using Zernike moments as part of a fired-cartridge-case comparison system would result in better performance than using any of the previously proposed functional features.[13]

Below, we provide details of the extraction of:

- central moments (§3.4.2)

[13] We also tested central-moment invariants (Hu [68]; Flusser [69]; Flusser & Suk [70]), but they did not perform as well as Zernike moments.

- circle-moment invariants (§3.4.3)
- Legendre moments (§3.4.4)
- Coefficients of Fourier series fitted to concentric circles (§3.4.5)
- Zernike moments (§3.4.6)

§3.4.7 provides, for each feature-extraction method, the number of features that we extracted.

3.4.2. Central moments

Central moments were previously applied to forensic comparison of fired cartridge cases in Ghani et al. [27].

Raw geometric moments have the general form given in Equation (1), in which m and n are non-negative integers that specify the orders of the moment, and $f(x,y)$ is an arbitrary function of x and y. Equation (2) provides the form applicable for digital data. N_x and N_y are the number of discrete x and discrete y values respectively. Equation (3) provides the formula for calculating central moments. Since we centred our data in x and y during preprocessing, $\bar{x}=0$ and $\bar{y}=0$ and there will be no difference between our $\mu_{m,n}^{\text{raw}}$ and $\mu_{m,n}^{\text{central}}$ values.

$$\mu_{m,n}^{\text{raw}} = \iint x^m y^n f(x,y)\,dx\,dy \tag{1}$$

$$\mu_{m,n}^{\text{raw}} = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} x_i^m y_j^n z(x_i,y_j) \tag{2}$$

$$\mu_{m,n}^{\text{central}} = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} (x_i-\bar{x})^m (y_j-\bar{y})^n z(x_i,y_j) \tag{3}$$

Fig. 4 shows the products of the power functions on x and on y up to order 4 plotted over a disc of unit radius. All panels in Fig. 4 through Fig. 8 are plotted with $-1\ldots1$ as the range for each of the x, y, and z axes. Because the x and y values are in the range $-1\ldots1$, as m and n increase, the magnitudes of the outputs of the power functions decrease toward zero. For visualization purposes, in each panel of Fig. 4 we have scaled the product of the power functions so that the maximum magnitude on the z axis is 1. The plot in each panel represents a function that, if sampled at the same x and y values as the matrix of data values $z(x,y)$, produces a matrix of values that can be pointwise multiplied with a matrix of data values $z(x,y)$ and the products summed to extract a scaled central moment that can be used as a feature value.

Central moments are not orthogonal and are not rotation invariant.

3.4.3. Circle-moment invariants

Circle-moment invariants were previously applied to forensic comparison of fired cartridge cases in Leng & Huang [29].

Circle-moment invariants are a modified version of central moments. They have the form given in Equation (4). Whereas central moments use the signed values of the power functions on x and on y, circle-moment invariants use the absolute values. Fig. 5 shows the products of the absolute power functions on x and on y up to order 4 plotted over a disc of unit radius. For visualization purposes, we have scaled the product of the absolute power functions so that the maximum magnitude in each panel is 1.

Circle-moment invariants are rotation invariant, but not orthogonal.

$$\mu_{m,n}^{\text{circle}} = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \left|(x_i-\bar{x})^m\right|\left|(y_j-\bar{y})^n\right| z(x_i,y_j) \tag{4}$$

3.4.4. Legendre moments

Legendre moments were previously applied to forensic comparison of fired cartridge cases in Chuan et al. [28].

The previously considered moments have used a power function of x and a power function of y, but moments can be generalized to use other functions. Legendre moments have the form given in Equation (5), in which $\mathscr{L}_m(\cdot)$ is a Legendre polynomial of order m. Legendre polynomials up to order 4 are given in Equation (6). After the specification of the zeroth and first Legendre polynomials, higher orders in the series can be generated using Equation (7). Fig. 6 shows the scaled products of Legendre polynomials on x and on y up to order 4 plotted over a disc of unit radius. Legendre moments are orthogonal, but not rotation invariant.

$$\mu_{m,n}^{\text{Legendre}} = \frac{(2m+1)(2n+1)}{4}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \mathscr{L}_m(x_i)\mathscr{L}_n(y_j) z(x_i,y_j) \tag{5}$$

$$\mathscr{L}_0(x) = 1 \tag{6}$$

$$\mathscr{L}_1(x) = x \tag{7}$$

$$\mathscr{L}_2(x) = \frac{1}{2}\left(3x^2-1\right)$$

$$\mathscr{L}_3(x) = \frac{1}{2}\left(5x^3-3x\right)$$

$$\mathscr{L}_4(x) = \frac{1}{8}\left(35x^4-30x^2+3\right)$$

$$\mathscr{L}_m(x) = \left(2-\frac{1}{m}\right)x\mathscr{L}_{m-1}(x) - \left(1-\frac{1}{m}\right)\mathscr{L}_{m-2}(x)$$

3.4.5. Concentric-circle features

Fourier series fitted to concentric circles were previously applied to forensic comparison of fired cartridge cases in Thumwarin et al. [10]. The coefficient values from the Fourier series were used as features. For brevity, we refer to these features as "concentric-circle features".

Imagine a circle of radius ρ and a function $z(\rho,\theta)$ where θ specifies the angle in radians around the circumference of the circle. Fix ρ, and fit a Fourier series to the function $z(\theta)$ with the first-order component being a cosine with a period of 2π radians. All non-zeroth components will be cosines whose periods are $2\pi/n$ radians where n is a positive integer. Each component will therefore complete an integer number of periods as it travels around the circumference of the circle and will meet itself exactly in phase. The function $z(\theta)$ can be reconstructed to order N using a Fourier series as in Equation (8), in which A_0 is the mean value of $z(\theta)$, and A_n is the magnitude coefficient and φ_n the phase coefficient of component n of the series. Cosine functions up to $N=4$ with zero phase ($\varphi_n=0$ for all n) fitted to a unit-radius circle ($\rho=1$) are plotted in the top row of Fig. 7. Other rows of Fig. 7 show cosine functions up to successively lower N values fitted to successively smaller circles. Across the rows of Fig. 7, the period (ρ_m/N_m) of the highest order cosine function is the same.

$$\hat{z}_N^{\text{Fourier}}(\theta) = \frac{A_0}{2} + \sum_{n=1}^{N} A_n \cos(n\theta-\varphi_n) \tag{8}$$

As in Thumwarin et al. [10], we only extracted concentric-circle features from the whole region of interest including flowback. In order to fit Fourier series covering the segmented region of interest, we specified the radius ρ_m of each member of a series of concentric circles. ρ_m values were selected such that circles fell entirely within the segmented region of interest. We transformed the Cartesian-coordinate data matrices, $z(x,y)$, to polar coordinates, $z(\rho(x,y),\theta(x,y))$. We then selected the $z(\rho(x,y),\theta(x,y))$ data points that were closest to each concentric circle. Fourier series were fitted independently to each circle.

As the radii ρ_m of the circles decrease, so do the lengths of their circumferences. As ρ_m decreased, we decreased the order N_m of the Fourier series so that, across all circles, when measured in millimetres, the period of the highest order component was the same. The reduction in N_m with reduction in ρ_m is illustrated in Fig. 7. Details of the values of N_m and ρ_m used for feature extraction in the present research are

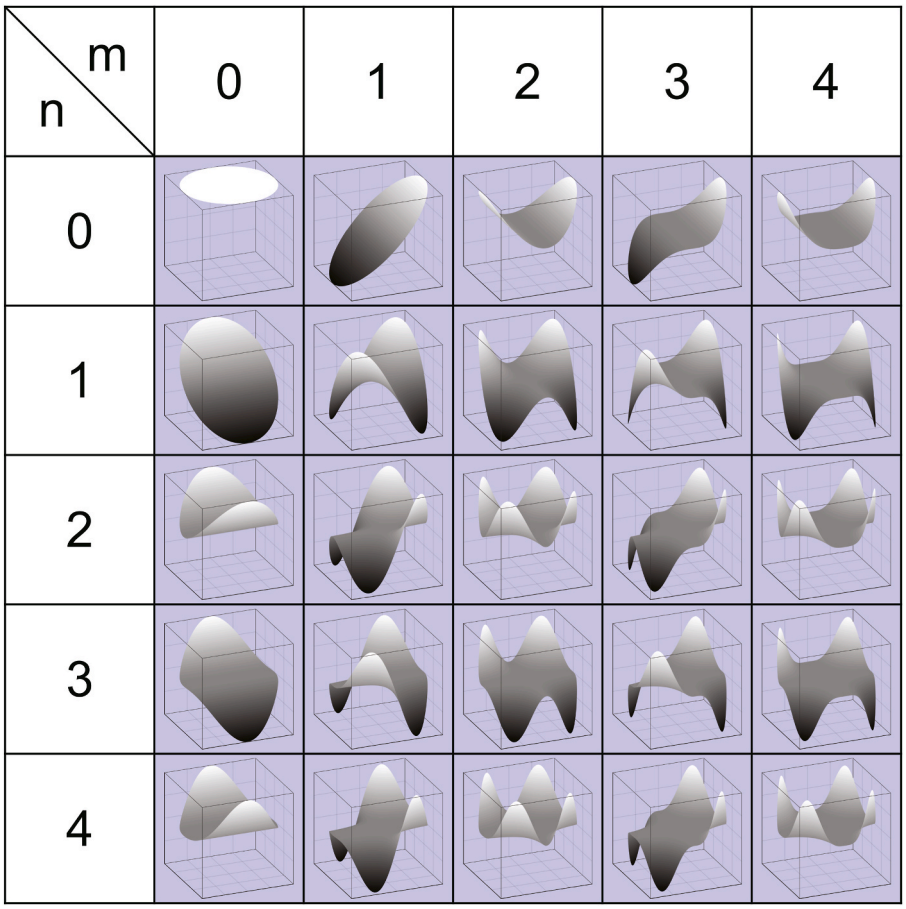

Fig. 4. Plots of scaled products of power functions used in the calculation of central moments up to order 4.

provided in §3.4.7 below.

The magnitude coefficients of a Fourier series can be used as rotation-invariant features. We henceforth refer to these features as "concentric-circle magnitude features". We also extracted features that took account of both magnitude and phase. We henceforth refer to these features as "concentric-circle magnitude and phase features". Phase *per se* is inconvenient as a feature because of discontinuity of values at $\varphi_n = 0 = 2\pi$. An alternative representation of a component of a Fourier series, given in Equation (9), makes use of weighted cosine and sine functions. We will use the weights a_n and b_n as paired features that together capture both magnitude and phase information.

$$A_n \cos(n\theta - \varphi_n) = a_n \cos(n\theta) + b_n \sin(n\theta) \tag{9}$$

$$a_n = A_n \cos(\varphi_n)$$
$$b_n = A_n \sin(\varphi_n)$$

$$A_n = \sqrt{a_n^2 + b_n^2}$$

$$\varphi_n = \begin{cases} \cos^{-1}\left(\frac{a_n}{A_n}\right) & \text{if } b_n \geq 0 \\ -\cos^{-1}\left(\frac{a_n}{A_n}\right) & \text{if } b_n < 0 \end{cases}$$

3.4.6. Zernike moments

Zernike polynomials were described in Zernike [59], and Teague [60] and Khotanzad & Hong [61] provide introductions to Zernike moments. To our knowledge, Zernike moments have not previously been applied to forensic comparison of fired cartridge cases.

Zernike moments have the form given in Equation (10), in which $V_{m,n}(\rho,\theta)$ is a Zernike polynomial parameterized in polar coordinates. Equation (11) provides the form for calculating Zernike moments from digital data. The constraint $x_i^2 + y_j^2 \leq 1$ was already enforced by the preprocessing of our data.

$$\mu_{m,n}^{\text{Zernike}} = \frac{m+1}{\pi} \int_{\rho=0}^{1} \int_{\theta=0}^{2\pi} V_{m,n}(\rho,\theta) f(\rho,\theta) d\rho d\theta \tag{10}$$

$$\mu_{m,n}^{\text{Zernike}} = \frac{m+1}{\pi} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} V_{m,n}\left(\rho(x_i,y_j), \theta(x_i,y_j)\right) z(x_i,y_j) \Big|_{x_i^2+y_j^2 \leq 1} \tag{11}$$

Zernike polynomials are calculated as in Equation (12), which consists of a function $R_{m,n}(\rho)$ dependent on distance ρ from the centre of a disc of unit radius, and a function $\cos(n\theta)$ or $\sin(n\theta)$ dependent on angle θ around the disc. The angle function is also dependent on n, which can be specified as a positive or a negative integer, or as zero. For $n \geq 0$ the angle-dependent function is a cosine function, and the notation uses n as a subscript. For $n < 0$ the angle-dependent function is a sine function, and the notation uses $-n$ as a subscript.

$$V_{m,n} = R_{m,n}(\rho)\cos(n\theta) \tag{12}$$

$$V_{m,-n} = R_{m,n}(\rho)\sin(n\theta)$$

The $R_{m,n}(\rho)$ are a series of orthogonal polynomial functions dependant on the values of m and n, see Equation (13).

$$R_{m,n}(\rho) = R_{m,-n}(\rho) = \sum_{i=0}^{\frac{m-|n|}{2}} (-1)^i \frac{(m-i)!}{i!\left(\frac{m+|n|}{2}-i\right)!\left(\frac{m-|n|}{2}-i\right)!} \rho^{m-2i} \tag{13}$$

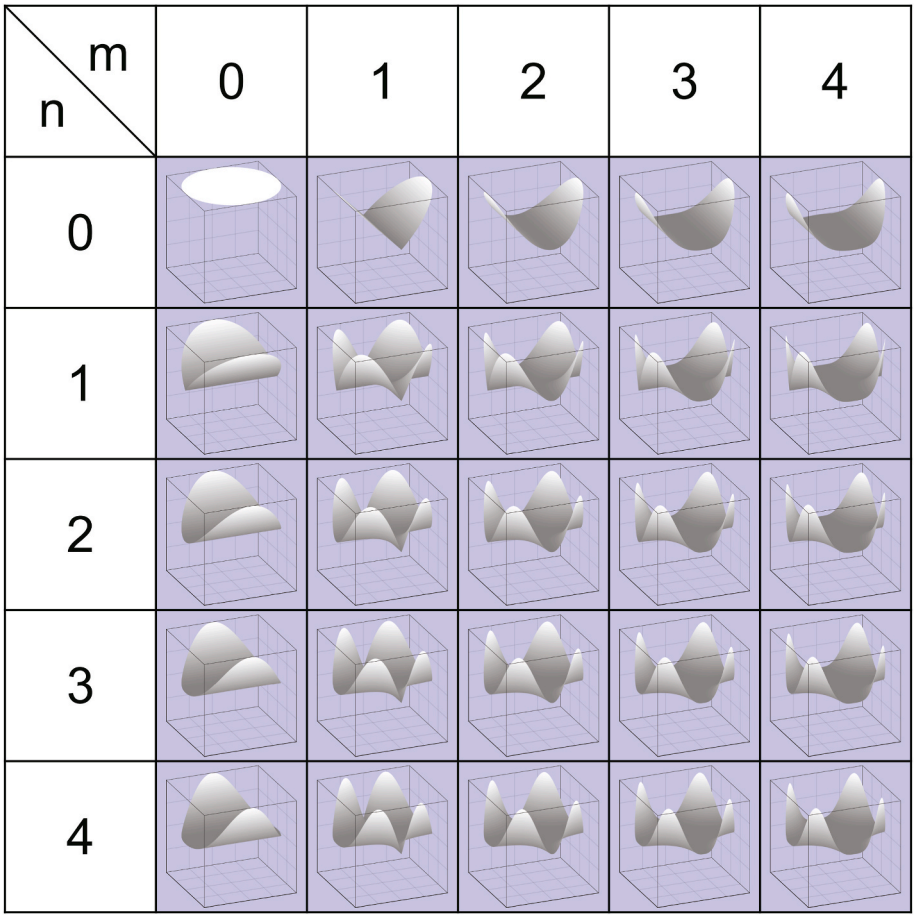

Fig. 5. Plots of scaled products of absolute power functions used in the calculation of circle-moment invariants up to order 4.

Zernike moments are defined for even values of $m-|n|$ with the constraint that $|n| \leq m$. The $R_{m,n}(\rho)$ up to order 4 are given in Equation (14). Fig. 8 shows Zernike polynomials up to order 4 plotted over a disc of unit radius. The output of Zernike polynomials intrinsically fall in the range $-1 \ldots 1$.

$$R_{0,0}(\rho) = 1 \tag{14}$$

$$R_{1,1}(\rho) = \rho$$

$$R_{2,0}(\rho) = 2\rho^2 - 1$$

$$R_{2,2}(\rho) = \rho^2$$

$$R_{3,1}(\rho) = 3\rho^3 - 2\rho$$

$$R_{3,3}(\rho) = \rho^3$$

$$R_{4,0}(\rho) = 6\rho^4 - 6\rho^2 + 1$$

$$R_{4,2}(\rho) = 4\rho^4 - 3\rho^2$$

$$R_{4,4}(\rho) = \rho^4$$

To calculate Zernike moments, we used the method described in Iskander et al. [62] and given in Equation (15), in which: z is a data matrix $z(x,y)$ rearranged into a column vector; $\boldsymbol{V}$ is a matrix in which each column is a matrix of Zernike polynomial values $V_{m,n}(\rho(x,y), \theta(x,y))$ rearranged into a column vector in the same way as for z, and for which the number of columns equals to the number of Zernike moments to be extracted[14]; superscript T indicates the transpose of the matrix; and $\widehat{\boldsymbol{\mu}}$ is a column vector of estimated Zernike moments. This method is a least-squares fit assuming a model in which the data are the product of the Zernike polynomials and the moments, plus a random error, i.e., $z = \boldsymbol{V\mu} + \boldsymbol{\varepsilon}$.

$$\widehat{\boldsymbol{\mu}} = \left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right)^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{z} \tag{15}$$

As discussed at the end of §3.4.5 and shown in Equation (9), a pair of cosine and sine functions capture both the magnitude and phase of a component of a Fourier series. Likewise a pair of Zernike polynomials $V_{m,n}$ and $V_{m,-n}$ with the same m and $|n|$ values capture both magnitude and phase information, therefore Zernike moments $\mu_{m,n}^{\text{Zernike}}$ and $\mu_{m,-n}^{\text{Zernike}}$ with the same m and $|n|$ values can be used as paired features that capture both magnitude and phase information.

Theoretically, Zernike moment magnitude and phase features are not rotation invariant, but Zernike moment magnitude features, $\left\|\mu_{m,\pm n}^{\text{Zernike}}\right\| = \sqrt{\left(\mu_{m,n}^{\text{Zernike}}\right)^2 + \left(\mu_{m,-n}^{\text{Zernike}}\right)^2}$, are rotation invariant.

[14] Subject to the previously stated constraints regarding m, and n, we extracted Zernike moments for all negative, zero, and positive n for each m up to the maximum order of m used.

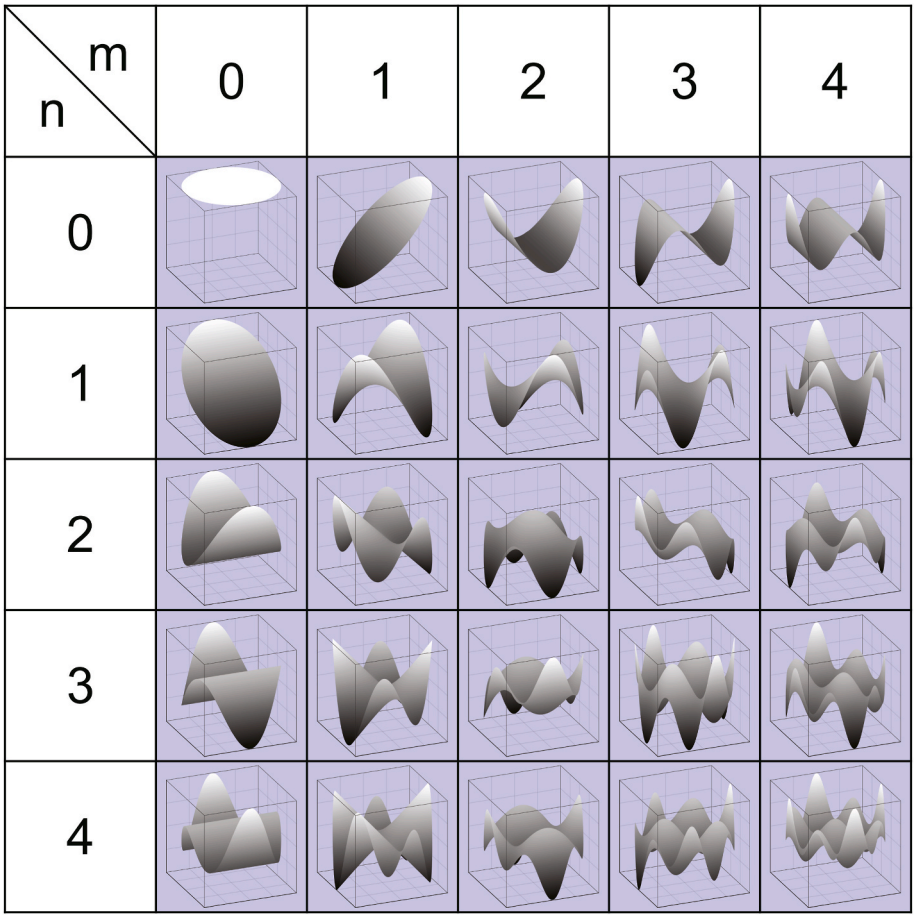

Fig. 6. Plots of scaled products of Legendre polynomials used in the calculation of Legendre moments up to order 4.

3.4.7. Numbers of features extracted

Given the relatively small size of our dataset, we did not want to extract a very large number of features, but we wanted to extract sufficient information to obtain reasonably good performance on the cartridge-case-comparison task. We initially focussed on extracting Zernike-moment magnitude and phase features for Scenario 1, and *a priori* believed that extracting up to 10th order moments for the firing-pin impression and up to 20th order for the breech-face region would be a reasonable compromise. We chose a lower order for the firing-pin impression because its gross shape is usually considered an important source of information, whereas we chose a higher order for the breech-face region in order to capture finer details of surface irregularities. In preliminary tests, we also tested up to 5th and up to 15th order for the firing-pin impression, and up to 10th and up to 30th order for the breech-face region, but up to 10th and up to 20th order for the firing-pin impression and breech-face region respectively gave better or no worse results. Up to 10th order Zernike moments (up to $m = 10$) result in a total of 66 magnitude and phase features, and up to 20th order (up to $m = 20$) result in a total of 231 magnitude and phase features. In addition to fitting models to features extracted from the firing-pin impression alone and to features extracted from the breech-face region alone, we also fitted models to the concatenation of these two sets of features. The concatenation of firing-pin plus breech-face features contained a total of 297 features. When extracting features from the entire region of interest (either including or excluding flowback), we used up to 23rd order Zernike moments (up to $m = 23$), resulting in a total of 300 magnitude and phase features. These choices as to number of features to extract are somewhat arbitrary, however, we will treat them as specifications for the system and then validate the performance of that system.

For the Zernike-moment magnitude-only features, using the same orders as stated above, 36 features were extracted from the firing pin impression, 121 from the breech-face region, and 156 from the whole region of interest.

For the other moment-based feature sets (central moments, circle-moment invariants, and Legendre moments), we extracted approximately the same number of features as we had Zernike-moment magnitude and phase features: up to 7th order (up to $m = n = 7$) from the firing-pin impression, a total of 64 features; up to 14th order (up to $m = n = 14$) from the breech-face region, a total of 225 features; and up to 16th order (up to $m = n = 16$) from the whole region of interest, a total of 289 features.[15]

As in in Thumwarin et al. [10], for the concentric-circle features, we only extracted features from the whole region of interest. A 23rd order Fourier series was fitted to the outermost circle (circle $m = 1$ with order $N_1 = 23$), matching the order of the Zernike moments. Based on measurements from the cartridge case which was used as the target for rotation, the radius of the outermost circle (the largest circle that could be drawn within the segmented region of interest) was $\rho_1 = 1.671$ mm. The circumference of that circle was therefore $c_1 = 2\pi\rho_1 = 2\pi \times 1.671 = 10.497$ mm, and the period of the highest order component of the Fourier series was therefore $\tau_{1,N_1} = c_1/N_1 = 10.497/23 = 0.456$ mm. This specifies the smallest wavelength of repetitive surface irregularities in the region of interest from which these features can extract information. Additional circles were then drawn, concentric to the outermost circle but with smaller radii. Moving from the outermost to the

[15] The calculation of the number of features includes moments for which $m = 0$ or $n = 0$, e.g., up to 7th order is 8^2 features.

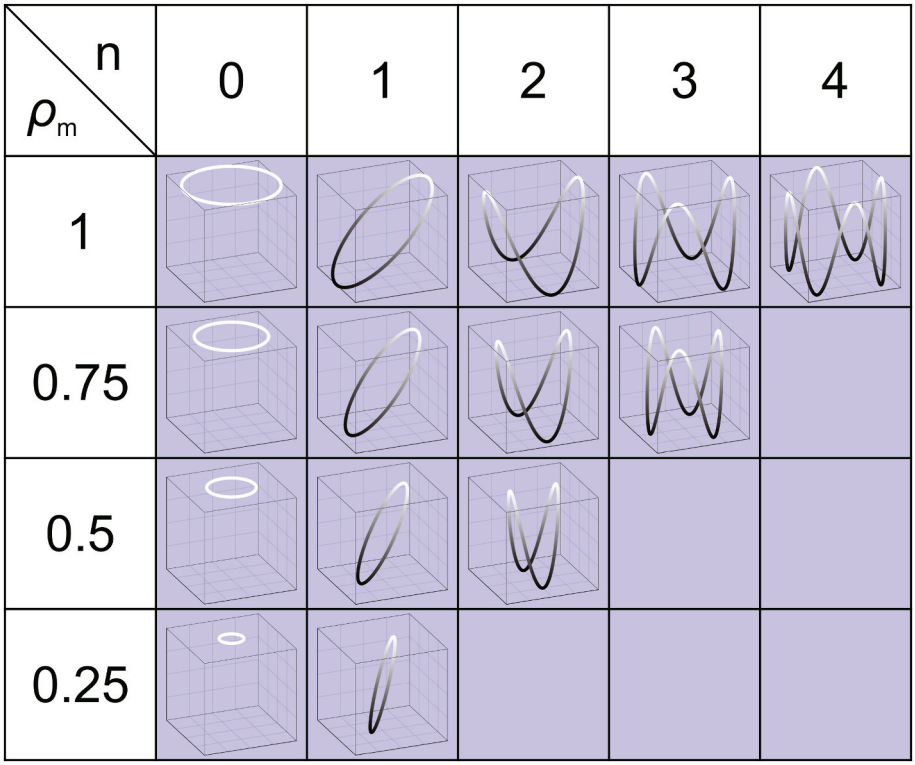

Fig. 7. Cosine components of a Fourier series up to component 4 ($N = 4$) fitted to a unit-radius circle ($\rho = 1$), and cosine components up to successively lower N fitted to successively smaller circles.

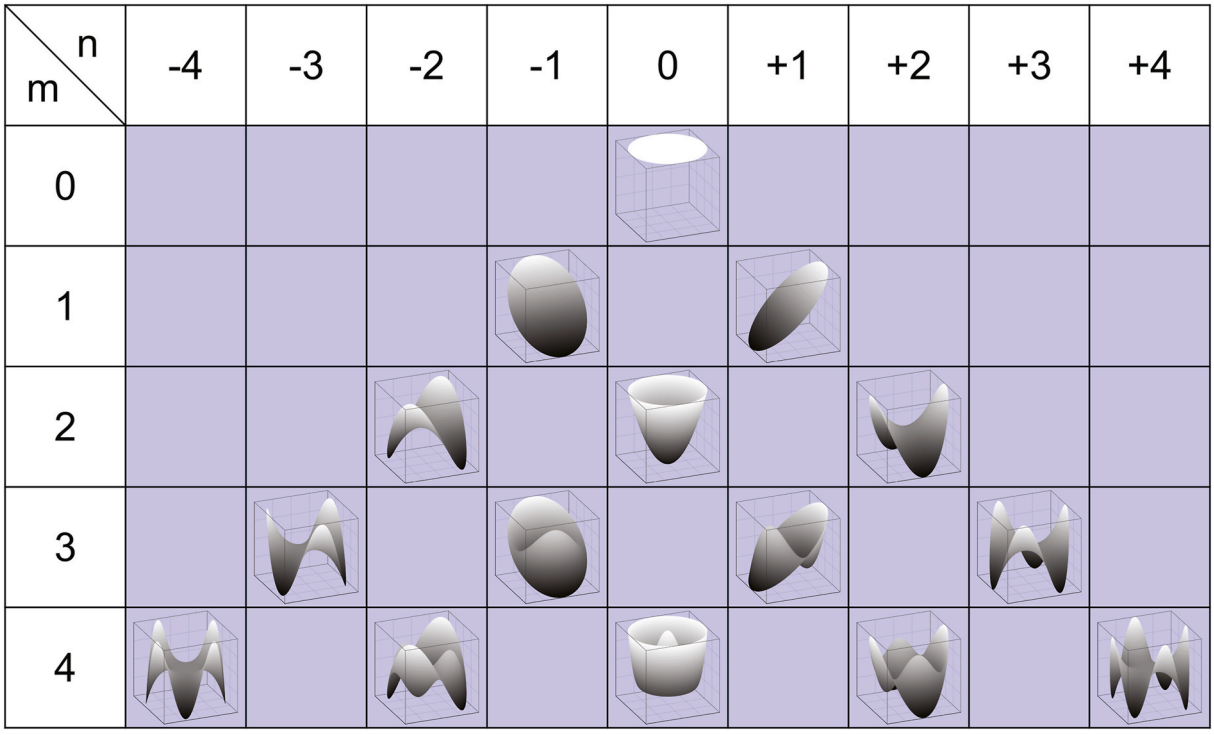

Fig. 8. Plots of Zernike polynomials used in the calculation of Zernike moments up to order 4.

innermost circle, the order of the Fourier series for each circle was set to be two less than that of the previous circle: $N_{m+1} = N_m - 2$, i.e., the orders were 23, 21, 19, …, 5, 3, 1 (a total of 12 circles). The radius of each circle was then calculated such that the period of the highest order component of the Fourier series fitted to that circle τ_{m,N_m} was equal to τ_{1,N_1}, i.e., $\rho_m = N_m \tau_{1,N_1}/2\pi$. The resulting radii were 100, 91, 83, 74, 65, 57, 48, 39, 30, 22, 13, and 4% of the radius of the outermost circle. The pattern of reduction in ρ_m with reduction in N_m is illustrated in Fig. 7, but with orders 4, 3, 2, 1. Starting with $N_1 = 23$ and reducing in steps of 2, the total number of concentric-circle magnitude and phase features

was 300, and the total number of concentric-circle magnitude-only features was 156.

When extracting moments from the whole of the region of interest excluding flowback, data within the flowback region were not used in the calculations (in Matlab, data within the flowback region were coded as "not a number, NaN").

3.5. Statistical models

3.5.1. Introduction

This subsection provides details of the statistical modelling pipeline previously outlined in §3.1 and Fig. 2, i.e., dimension reduction (§3.5.2), calculation of uncalibrated likelihood ratios (§3.5.3), and calibration (§3.5.4).

3.5.2. Dimension reduction

Given the relatively small size of the dataset, in order to reduce the number of parameter values to be estimated in the next stage of modelling and in order to reduce potential redundancy of information among features, we reduced the number of feature dimensions using principal component analysis (PCA; Pearson [71]; Hotelling [72]). For each feature set, for the firing-pin impression we reduced the number of dimensions to 10, for the breech-face region to 20, and for the whole region of interest (including or excluding flowback, and including the concatenation of features from the breech-face-region and the firing-pin-impression) to 30. These values were chosen as a compromise between trying not to discard potentially useful information and trying not to have too many dimensions relative to the number of firearms available for training the between-source covariance matrix in the next stage of modelling.

After dimension reduction using PCA, we calculated linear discriminant functions (LDFs; Fisher [73]; Rao [74]) on the training data and transformed all the data into the LDF space. For LDF training, each individual firearm (each source), constituted a category. We did not use LDFs for additional dimension reduction, but only to rotate the data into orthogonal dimensions that maximized between-source versus within-source variance ratios. In preliminary tests, using LDFs for dimension reduction led to worse results. If there are mismatches in conditions between the questioned-source item and the known-source item (as is common in forensic voice comparison), then the mismatch in conditions can be the cause of substantial within-source variability. In this circumstance, training LDFs on data that include the mismatch and using only the lower-order dimensions serves as a mismatch-compensation technique: the lower-order dimensions have higher ratios of between-source to within-source variance, including within-source variance due to mismatched conditions, than do the higher-order dimensions. Since all the cartridge cases in our dataset had brass primer cups, there was no within-source mismatch in conditions for the training data or for the calibration/validation data, hence no mismatch-compensation advantage to be gained from dimension reduction using LDFs.[16]

3.5.3. Calculation of uncalibrated likelihood ratios

Uncalibrated likelihood ratios were calculated using a common-source likelihood-ratio model known in the automatic-speaker-recognition literature as the two-covariance version of probabilistic linear discriminant analysis (PLDA; Prince & Elder [75]; Kenny [76]; Brümmer & de Villiers [77]; Sizov et al. [78]).[17] We used the implementation from Sizov et al. [78]. The form of the model is as given in Equation (16), in which λ is an uncalibrated likelihood ratio, $f(\boldsymbol{v}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is a multivariate Gaussian probability-density function, $\boldsymbol{v}_q$ and $\boldsymbol{v}_k$ are post-PCA-LDF questioned-source and known-source feature vectors respectively, $\widehat{\boldsymbol{\mu}}_r$ is the estimate of the mean vector for the relevant population, and $\widehat{\boldsymbol{\Sigma}}_w$ and $\widehat{\boldsymbol{\Sigma}}_b$ are, respectively, the within-source covariance matrix and the between-source covariance matrix estimates for the relevant population.

$$\lambda = \frac{f\left(\begin{bmatrix}\boldsymbol{v}_q\\ \boldsymbol{v}_k\end{bmatrix}\middle|\middle|\begin{bmatrix}\widehat{\boldsymbol{\mu}}_r\\ \widehat{\boldsymbol{\mu}}_r\end{bmatrix}, \begin{bmatrix}\widehat{\boldsymbol{\Sigma}}_w+\widehat{\boldsymbol{\Sigma}}_b & \widehat{\boldsymbol{\Sigma}}_b\\ \widehat{\boldsymbol{\Sigma}}_b & \widehat{\boldsymbol{\Sigma}}_w+\widehat{\boldsymbol{\Sigma}}_b\end{bmatrix}\right)}{f(\boldsymbol{v}_q|\widehat{\boldsymbol{\mu}}_r, \widehat{\boldsymbol{\Sigma}}_w+\widehat{\boldsymbol{\Sigma}}_b) f(\boldsymbol{v}_k|\widehat{\boldsymbol{\mu}}_r, \widehat{\boldsymbol{\Sigma}}_w+\widehat{\boldsymbol{\Sigma}}_b)} \quad (16)$$

For each segmented region, we trained three different PLDA models, which differed in their $\widehat{\boldsymbol{\Sigma}}_w$ values:

Model 1 v 1 corresponds to Scenario 1.

A pooled $\widehat{\boldsymbol{\Sigma}}_w$ was calculated using all feature vectors from all sources in the training data.

Model 1 v 3 corresponds to Scenario 2 and assumes the practitioner fired 3 cartridges from the seized firearm.

From the 10 feature vectors of each source (corresponding to the 10 cartridge cases from each firearm), there are $\binom{10}{3} = 120$ possible combinations of 3 feature vectors. 10 of these combinations were randomly selected, and the mean vector for each of these combinations was calculated. A pooled $\widehat{\boldsymbol{\Sigma}}_w$ was then calculated using the combination of all the original singleton feature vectors and all the three-mean feature vectors from all sources in the training data.

Model 1**v 9** corresponds to Scenario 2 and assumes the practitioner fired 9 cartridges from the seized firearm.

From the 10 feature vectors of each source (corresponding to the 10 cartridge cases from each firearm), all $\binom{10}{9} = 10$ possible combinations of 9 feature vectors were drawn, and the mean vector for each of these combinations was calculated.[18] A pooled $\widehat{\boldsymbol{\Sigma}}_w$ was then calculated using the combination of all the original singleton feature vectors and all the nine-mean feature vectors from all sources in the training data.

Model 1 v 1, Model 1 v 3, and Model 1 v 9 will have successively smaller-valued within-source covariance matrices, the latter two reflecting the size of the group of known-source cartridge cases that will be compared with the questioned-source cartridge case.

The mean vector for each source in the training data was calculated using, as applicable for each model, all the original singleton feature vectors from that source, or all the original singleton feature vectors from that source plus all the three-mean or all the nine-mean feature vectors belonging to that source. $\widehat{\boldsymbol{\mu}}_r$ and $\widehat{\boldsymbol{\Sigma}}_b$ were then calculated using all of the mean vectors from each source.

Prior to training the PLDA model, independently for each feature-vector dimension, the training data were centred to 0 and were scaled to a standard deviation of 1. These transformations, obtained from the mean and standard deviation of the training data, were subsequently applied to the calibration/validation data. Given this centring and scaling, for Model 1 v 1 and Model 1 v 9, should be a vector of zeros, the diagonal of should be a vector of ones, and the values of should be the same for both models. These values will differ slightly for Model 1 v 3 because of the random sub-selection of data used in training that model.

3.5.4. Calibration

Whereas the model used to calculate uncalibrated likelihood ratios requires the estimation of a large number of parameter values in a multivariate data space, a calibration model is a parsimonious model which requires the estimation of a small number of parameters in a

[16] In preliminary work, we tested several other dimension-reduction methods, but none outperformed the combination of PCA + LDF.

[17] In Aitken & Lucy [79], it is called the "multivariate normal (MVN) procedure".

[18] Occasionally, the number of fired cartridge cases available for a firearm was 8 or 9 rather than 10, in which case the number of feature vectors available was used.

univariate space. The ratio of parameter values to be estimated relative to the number of data points is therefore much smaller for the latter model than for the former.

We calibrated the uncalibrated likelihood ratios using a logistic-regression model. Logistic regression is commonly used as a calibration model in forensic voice comparison (González-Rodríguez et al. [80]; Morrison [81]). We used the regularized-logistic-regression model described in Morrison & Poh [82], with a regularization weight equivalent to a set of feature vectors from one firearm.[19]

For each segmented region and for each PLDA model (Model 1 v 1, Model 1 v 3, and Model 1 v 9), we trained different calibration models. Each calibration model was trained using a set of same-source scores and a set of different-source scores, where: a "score" is an uncalibrated log likelihood ratio, $\log(\lambda_{ij})$[20]; a same-source score, $\log(\lambda_{ij})|_{i=j}$, is the logged output of a PLDA model when the input is a pair of feature vectors originating from different cartridge cases fired from the same firearm ($\boldsymbol{v}_i$ versus $\boldsymbol{v}_j$ with $i = j$); and a different-source score, $\log(\lambda_{ij})|_{i\neq j}$, is the logged output of a PLDA model when the input is a pair of feature vectors originating from cartridge cases fired from different firearms ($\boldsymbol{v}_i$ versus $\boldsymbol{v}_j$ with $i \neq j$). To calculate scores for training the calibration model, $\boldsymbol{v}_i$ versus $\boldsymbol{v}_j$ pairs were entered into Equation (16), with $\boldsymbol{v}_i$ in place of $\boldsymbol{v}_q$ and with $\boldsymbol{v}_j$ in place of $\boldsymbol{v}_k$. Given a set of same-source scores and a set of different-source scores, a logistic regression model was trained using an iterative procedure (conjugate-gradient method; Hestenes & Stiefel [83]; Minka [84]) that estimated values for the intercept and slope coefficients β_0 and β_1 of Equation (17), which was then used to convert each uncalibrated log likelihood ratio, $\log(\lambda)$, to a calibrated log likelihood ratio, $\log(\lambda)$.[21]

$$\log(\lambda) = \beta_0 + \beta_1 \log(\lambda) \quad 17$$

Same-source pairs of feature vectors ($\boldsymbol{v}_i$ versus $\boldsymbol{v}_j$ with $i = j$) and different-source pairs of feature vectors ($\boldsymbol{v}_i$ versus $\boldsymbol{v}_j$ with $i \neq j$) for training each calibration model (and for cross-validation) were constructed as follows:

Model 1 **v 1:** To create same-source pairs of feature vectors, all $\binom{10}{2} = 45$ possible combinations of 2 feature vectors were drawn from the 10 feature vectors originating from a firearm. One of the feature vectors in each pair was assigned to $\boldsymbol{v}_i$ and the other to $\boldsymbol{v}_j$ (the Model 1 v 1 PLDA model is symmetrical so the order of assignment is irrelevant). This resulted in 45 pairs of same-source feature vectors from each firearm. To create different-source pairs of feature vectors, each feature vector from each firearm was compared with each feature vector from every other firearm. This resulted in 100 pairs of different-source feature vectors from each pair of firearms.

Model 1 **v 3:** To create same-source pairs of feature vectors, each of the 10 feature vectors originating from a firearm was selected in turn, and the selected singleton feature vector was assigned to $\boldsymbol{v}_i$. From the remaining 9 feature vectors of each firearm, using random selection without replacement, 3 non-overlapping combinations of 3 feature vectors were drawn, and the mean vector of each combination was in turn assigned to $\boldsymbol{v}_j$. This resulted in 30 pairs of same-source feature vectors from each firearm. To create different-source pairs of feature vectors, each feature vector from each firearm was compared with each of the mean vectors of 3 non-overlapping randomly selected combinations of 3 feature vectors from each of the other firearms. The combinations of 3 feature vectors were randomly selected without replacement from the total of 10 feature vectors from the second firearm (one of the feature vectors was not used). A different random selection from the second firearm was used for comparison with each of the singleton feature vectors from the first firearm. The singleton feature vector was assigned to $\boldsymbol{v}_i$ and each of the three-mean vectors was in turn assigned to $\boldsymbol{v}_j$. This resulted in 30 pairs of different-source feature vectors from each pair of firearms (with $\boldsymbol{v}_1$ versus $\boldsymbol{v}_2$ counted as a different pair to $\boldsymbol{v}_2$ versus $\boldsymbol{v}_1$).

Model 1 **v 9:** To create same-source pairs of feature vectors, each of the 10 feature vectors originating from a firearm was selected in turn, the selected singleton feature vector was assigned to , and the mean vector of the other 9 feature vectors was assigned to . This resulted in 10 pairs of same-source feature vectors from each firearm. To create different-source pairs of feature vectors, each feature vector from each firearm was compared with the mean of each of the possible combinations of 9 feature vectors from every other firearm. The singleton feature vector was assigned to and the nine-mean vector to . This resulted in 100 pairs of different-source feature vectors from each pair of firearms (with versus counted as a different pair to versus).

In addition to separately calibrating the scores from each of the firing-pin impression and the breech-face region, we also used a logistic-regression model to simultaneously fuse and calibrate scores from these two regions. The scores were parallel in that each firing-pin-impression score corresponded to a breech-face-region score that was calculated using the same combination of digital images (including for Model 1 v 3, the same random selections of images). Given a parallel set of same-source and different-source scores, a regularized-logistic-regression model was trained resulting in estimated values for the intercept β_0 and for two slope coefficients β_1 and β_2. These coefficient values were then used to fuse and calibrate a parallel pair of scores, $\log(\lambda_1)$ extracted from the firing-pin impression and $\log(\lambda_2)$ extracted from the breech-face region, as in Equation (18).

$$\log(\lambda) = \beta_0 + \beta_1 \log(\lambda_1) + \beta_2 \log(\lambda_2) \quad 18$$

Calibration and validation were performed together using cross-validation (see §4.2 for details).

4. Validation

4.1. Introduction

A system validation was conducted for each different feature-extraction method applied to each different segmented region. Validation was conducted according to the relevant recommendations in the *Consensus on validation of forensic voice comparison* (Morrison et al. [85]). In this section, we describe the validation procedures (§4.2), and the metric (log-likelihood-ratio cost, C_{llr}; §4.3) and graphic (Tippett plot; §4.4) used to represent the results.

4.2. Validation procedures

Calibration and validation were performed using cross-validation, comparing feature vectors from each firearm with other feature vectors from the same firearm and with feature vectors from all the other firearms in the calibration/validation set.

Considering a matrix of all possible combinations of two cartridge cases: Since Model 1 v 1 is symmetrical, the same-source comparisons were those on the diagonal of the matrix and the different-source comparisons were those on the upper right of the matrix (or those on the bottom left, but not both). Since Model 1 v 9 and Model 1 v 3 are not symmetrical, the same-source comparisons were those on the diagonal of the matrix and the different-source comparisons were those on both the upper right and the lower left of the matrix.

[19] In the notation of Morrison & Poh [82]: $w^{\prime\prime} = \kappa^{\prime\prime}/2N$, where $\kappa^{\prime\prime} = 1$, and N is the number of firearms that contributed to scores that were used to train the logistic-regression model. See Morrison & Poh [82] for further explanation.

[20] Use of the term "score" to refer to an uncalibrated log likelihood ratio is common in forensic voice comparison. Such scores, which take account of both similarity and typicality, should not be confused with similarity scores (see §1.4.4).

[21] Natural logarithms were used for the calculations.

Leave-one-source-out/leave-two-sources-out cross-validation was used: In a cross-validation loop in which the score to be calibrated was a same-source score, e.g., the result of comparing a cartridge case fired from firearm *A* with another cartridge case fired from firearm *A*, all scores that resulted from comparisons in which one or both members of the pair was a cartridge case fired from firearm *A* were excluded from the data used to train the calibration model (leave-one-source-out). In a cross-validation loop in which the score to be calibrated was a different-source score, e.g., the result of comparing a cartridge case fired from firearm *A* with a cartridge case fired from firearm *B*, all scores that resulted from comparisons in which one or both members of the pair was a cartridge case fired from firearm *A* or a cartridge case fired from firearm *B* were excluded from the data used to train the calibration model (leave-two-sources-out).

4.3. Validation metric: Log-likelihood-ratio cost (C_{llr})

Given a same-source input, a good output from a forensic-evaluation system would be a likelihood-ratio value that is much larger than 1, a less good output would be a value that is only a little larger than 1, a bad output would be a value less than 1, and a worse output would be a value much less than 1. *Mutatis mutandis*, given a different-source input, a good output would be a value much less than 1.

A metric that captures this gradient goodness is the log-likelihood-ratio cost (C_{llr}; Brümmer & du Preez [86]), which is calculated as in Equation (19), in which λ_s and λ_d are likelihood-ratio outputs corresponding to same-source and different-source input pairs respectively, and N_s and N_d are the number of same-source and different-source input pairs respectively.

$$C_{llr} = \frac{1}{2}\left(\frac{1}{N_s}\sum_{i=1}^{N_s} log_2\left(1+\frac{1}{\lambda_{s_i}}\right) + \frac{1}{N_d}\sum_{j=1}^{N_d} log_2\left(1+\lambda_{d_j}\right)\right) \quad (19)$$

Lower C_{llr} values indicate better performance. C_{llr} values cannot be less than 0. A system that always responded with a likelihood ratio of 1 irrespective of the input, and hence gave no useful information, would have a C_{llr} value of 1. A system with a C_{llr} of less than 1 is providing useful information. C_{llr} values substantially greater than 1 can be produced by uncalibrated or miscalibrated systems.

For further explanation of C_{llr} and its interpretation, see Appendix C of Morrison et al. [85].

4.4. Validation graphic: Tippett plot

Tippett plots (Meuwly [87]) consist of plots of the empirical cumulative probability distributions of the same-source log-likelihood-ratio values and of the different-source log-likelihood-ratio values. The tradition is to plot lines joining the data points rather than to plot the data points themselves. Tippett plots of some of the results of the present study are provided in Fig. 9 below. The *y*-axis values corresponding to the curves rising to the right give the proportion of same-source test results with log likelihood-ratio values less than or equal to the corresponding value on the *x*-axis. The *y*-axis values corresponding to the curves rising to the left give the proportion of different-source test results with log likelihood-ratio values greater than or equal to the corresponding value on the *x*-axis. In general, shallower curves with greater separation between the two curves indicates better performance. Tippett plots give an indication of the range of possible likelihood-ratio values that the system could generate under the test conditions, and can also reveal problems such as bias in the output.

For further explanation of Tippett plots and their interpretation, see Appendix C of Morrison et al. [85].

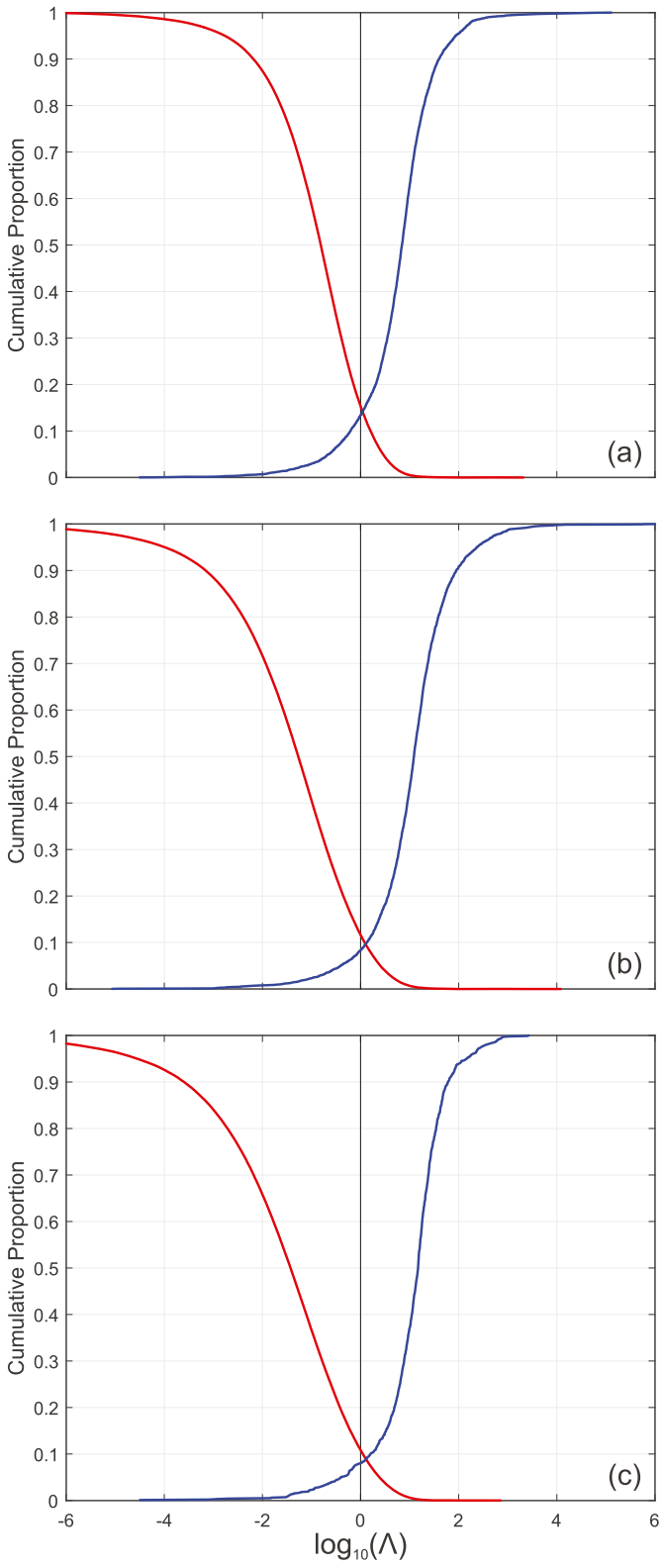

Fig. 9. Tippett plots of validation results obtained using Zernike moment magnitude and phase features extracted from the whole of the region of interest including flowback. (a) Model 1 v 1. (b) Model 1 v 3. (c) Model 1 v 9.

5. Results

5.1. Introduction

In this section, we present the validation results, including C_{llr} values (§5.2) and selected Tippett plots (§5.3).[22]

5.2. C_{llr} values

Table 1, Table 2, and Table 3 provide C_{llr} values obtained from the validations of Model 1 v 1, Model 1 v 3, and Model 1 v 9 respectively. Each table provides C_{llr} values from the factorial of combinations of feature set and segmented region (including using feature concatenation and score-level fusion to combine features extracted separately from the breech-face region and the firing-pin impression). Examining these tables, a clear pattern of results emerges:

1. The feature set resulting in best performance is the Zernike moment magnitude and phase feature set.
2. The segmented region resulting in best performance is the whole of the region of interest including flowback.
3. The larger the size of the group of known-source cartridge cases that was compared to the questioned-source cartridge case, the better the performance.

5.3. Tippett plots

Fig. 9 provides Tippett plots of validation results obtained using Zernike moment magnitude and phase features extracted from the whole of the region of interest including flowback. The results show good calibration, with the same-source and different-source curves crossing near $\log_{10}(\lambda) = 0$. For all models, likelihood-ratio values into the thousands in favour of same source would be supported, and likelihood-ratio values into the tens of thousands in favour of different source would be supported. The larger the number of known-source cartridge cases used, the more negative the limit of different-source log-likelihood-ratio values obtained (the Tippett plots are truncated at $\log_{10}(\lambda) = -6$, hence not all values are shown). Related to this pattern, substantial asymmetry is apparent for the same-source versus different-source log-likelihood-ratio values from Model 1 v 9. This increasing asymmetry is the expected and commonly observed pattern as within-source variability becomes much less than between-source variability.[23]

6. Discussion

6.1. Introduction

In this section, based on the results, we discuss:

- what we consider to be the best feature set (§6.2)
- the benefit of rotating the data matrices to a common target (§6.3)
- the benefit of using more known-source cartridge cases (§6.4)
- what we consider to be the best segmentation of the cartridge case base (§6.5)

6.2. Best feature set

In §3.4.1, on theoretical grounds and based on empirical results of applications in other fields, we hypothesized that using Zernike moments as features would result in better performance than using any of the other feature sets previously proposed in the literature on forensic comparison of fired cartridge cases.

Our results demonstrated that this was indeed the case with respect to other moment-based feature sets. For the whole region of interest including flowback, compared to Legendre moment features (the best performing non-Zernike moment-based feature set), C_{llr} values for Zernike moment magnitude and phase features were lower by 10%, 14%, and 16% for Model 1 v 1, Model 1 v 3, and Model 1 v 9 respectively.

Compared to concentric-circle magnitude and phase features, however, C_{llr} values for Zernike moment magnitude and phase features were only lower by 1%, 2%, and 3% for Model 1 v 1, Model 1 v 3, and Model 1 v 9 respectively. Although, the improvement is slight, Zernike moment magnitude and phase features have the advantage of being simpler to extract.

For future work, including ultimate application to casework, we therefore consider Zernike moment magnitude and phase features to be the best feature set to use.

6.3. Benefit of rotation

In §3.3 we noted that the cost of rotating the data matrices is a reason to prefer rotation-invariant features, but, if rotation leads to substantial improvement in performance, that cost may be justified.

After performing rotation, the theoretically non-rotation-invariant Zernike moment magnitude and phase features did not consistently result in better performance than the theoretically rotation-invariant Zernike moment magnitude-only features (see Table 1, Table 2, and Table 3). For all three models, for breech-face region alone and for score fusion (and for Model 1 v 9 for the whole region of interest excluding flowback), C_{llr} values were actually lower for Zernike moment magnitude-only features than for Zernike moment magnitude and phase features.

For all three models, for the whole region of interest including flowback, however, C_{llr} values were lower for Zernike moment magnitude and phase features than for Zernike moment magnitude-only feature, albeit only by 2%.

We ran an additional set of validations without rotation, using Zernike moment magnitude-only features and Zernike moment magnitude and phase features extracted from the whole region of interest including flowback. For all combinations of model and for both magnitude-only and magnitude-and-phase features, the C_{llr} values for rotated versus non-rotated image data were less than 1% different.[24] Thus, even if not theoretically rotation invariant, the magnitude and phase features in practice gave equally good results irrespective of whether rotation was applied to the data matrices or not.

For future work, including ultimate application to casework, we therefore consider the cost of performing rotation to be not justified.

6.4. Benefit of using more known-source cartridge cases

In §3.5.3 we described using different numbers of known-source cartridge cases for training, resulting in Model 1 v 1, Model 1 v 3, and Model 1 v 9 having successively smaller-valued within-source covariance matrices. The expected result of this is that models with smaller

[22] In addition, in order to assess the stability of the system using Zernike moment magnitude and phase features extracted from the whole of the region of interest including flowback, we performed randomization tests in which in each iteration we randomly selected a different 198 firearm training dataset versus 99 firearm calibration/validation dataset split. Based on the results, we were satisfied that the system is sufficiently stable with respect to the selection of data for such splits. For brevity, we do not include the results here.

[23] Compare, for example, the score distributions in Fig. 10a and b, and 16 of Morrison & Poh [82], and the corresponding Tippett plots in Figs. 11, 12 and 17 of Morrison & Poh [82].

[24] For Model 1 v 1, Model 1 v 3, and Model 1 v 9 respectively, for magnitude-only features the C_{llr} values were 0.529, 0.387, and 0.357 without rotation, compared to 0.531, 0.390, and 0.359 with rotation, and for magnitude and phase features the C_{llr} values were 0.520, 0.384, and 0.348 without rotation, compared to 0.519, 0.384, and 0.351 with rotation.

Table 1
C_{llr} values for each combination of feature set and segmented region for Model 1 v 1.

Feature Set	Segmented Region					
	whole region of interest		breech face	firing pin	breech face + firing pin	
	including flowback	excluding flowback			feature concat.	score-level fusion
central moments	0.616	0.671	0.710	0.923	0.682	0.697
circle-moment invariants	0.597	0.673	0.695	0.962	0.677	0.693
Legendre moments	0.577	0.679	0.719	0.923	0.709	0.707
concentric-circle features (mag.)	0.586	–	–	–	–	–
concentric-circle features (mag. & phase)	0.526	–	–	–	–	–
Zernike moments (mag.)	0.531	0.652	0.684	0.852	0.615	0.632
Zernike moments (mag. & phase)	0.519	0.645	0.689	0.841	0.605	0.635

Table 2
C_{llr} values for each combination of feature set and segmented region for Model 1 v 3.

Feature Set	Segmented Region					
	whole region of interest		breech face	firing pin	breech face + firing pin	
	including flowback	excluding flowback			feature concat.	score-level fusion
central moments	0.491	0.527	0.574	0.858	0.537	0.553
circle-moment invariants	0.467	0.532	0.557	0.901	0.537	0.551
Legendre moments	0.448	0.538	0.583	0.845	0.571	0.563
concentric-circle features (mag.)	0.435	–	–	–	–	–
concentric-circle features (mag. & phase)	0.390	–	–	–	–	–
Zernike moments (mag.)	0.390	0.502	0.547	0.752	0.459	0.476
Zernike moments (mag. & phase)	0.384	0.498	0.550	0.730	0.449	0.478

Table 3
C_{llr} values for each combination of feature set and segmented region for Model 1 v 9.

Feature Set	Segmented Region					
	whole region of interest		breech face	firing pin	breech face + firing pin	
	including flowback	excluding flowback			feature concat.	score-level fusion
central moments	0.485	0.497	0.542	0.843	0.506	0.534
circle-moment invariants	0.465	0.494	0.527	0.913	0.524	0.529
Legendre moments	0.420	0.501	0.549	0.822	0.546	0.534
concentric-circle features (mag.)	0.416	–	–	–	–	–
concentric-circle features (mag. & phase)	0.363	–	–	–	–	–
Zernike moments (mag.)	0.359	0.441	0.493	0.699	0.406	0.421
Zernike moments (mag. & phase)	0.351	0.450	0.508	0.678	0.401	0.430

ratios of within-source versus between-source covariance matrix magnitudes will produce a larger range of log-likelihood-ratio values, extending from higher-magnitude negative log-likelihood-ratio values for different-source comparisons to higher-magnitude positive log-likelihood-ratio values for same-source comparisons.

For Zernike moment magnitude and phase features extracted from the whole region of interest including flowback, the results were as expected for different-source comparisons, but not so for same-source comparisons (see Fig. 9). Increasing the number of known-source cartridge cases clearly improved results for different-source comparisons but did not clearly do so for same-source comparisons: in Fig. 9c the largest same-source log-likelihood-ratio value was actually less than in Fig. 9a and Fig. 9b.

Although the 9% reduction in C_{llr} values for Model 1 v 9 compared to Model 1 v 3 (0.351 compared to 0.384) appears to be substantial, if it is primarily due to large-magnitude negative log likelihood ratios from different-source comparisons getting even more negative, the increase in performance indicated by the C_{llr} values may not be particularly pertinent in casework.

In the context of casework, in which firing 3 cartridge cases from a seized firearm is currently the norm, the cost of firing 9 cartridge cases instead may not be justified. This is an issue to revisit once a larger database is collected and potentially better performing systems are developed.

For training and validation purposes, we recommend firing 10 cartridge cases from each firearm. These can be used to make multiple sets of data for training and validating Model 1 v 3.

6.5. *Best segmentation*

As mentioned in §1.3, in current casework practice, practitioners tend to visually compare the firing-pin impressions and the breech-face regions of pairs of fired cartridge cases. As mentioned in §1.4.3, in previous research using data, quantitative measurements, and statistical models, flowback has usually been excluded from analysis. In §3.3, however, we hypothesized that the flowback region would contain information related to the firearm that fired the cartridge.

Combining information from the firing-pin impression and the breech-face region was expected to result in better performance than using one of these alone, and that result was obtained: in Table 1, Table 2, and Table 3 it can be observed that any of the means tested for combining firing-pin-impression and breech-face-region information almost always resulted in lower C_{llr} values than using either of these alone.[25]

As we hypothesized, however, the best performance was obtained by

[25] There were a couple of exceptions for circle-moment invariants.

extracting features from the whole region of interest, including not only the breech-face region and the firing-pin impression, but also the flow-back region. It therefore appears that, contrary to received wisdom, the flowback region does contain useful information about the firearm that fired the cartridge case.

Practically, only having to segment the region of interest from the headstamp region, and not having to additionally segment the firing-pin impression and the breech-face region will result in a simpler and faster system for comparing fired cartridge cases.

For future work, including ultimate application to casework, we therefore consider the whole region of interest including flowback to be the best segmented region to use.

7. Conclusion

The present paper described and validated a feature-based system for calculation of likelihood ratios from 3D digital images of fired cartridge cases. The system includes a database of 3D digital images of the bases of approximately 3,000 fired cartridge cases, consisting of 10 cartridges fired per firearm from approximately 300 firearms of the same class (semi-automatic pistols that fire 9 mm diameter centre-fire Luger-type ammunition, and that have hemispherical firing pins and parallel breech-face marks). The images were captured using Evofinder®, an imaging system that is commonly used by operational forensic laboratories. Although in terms of the combination of number of firearms of the same class and number of fires per firearm, this may be one of the largest databases in existence, we consider it relatively small for training statistical models that take account of both within-source and between-source variability. Given this relatively small database, we were encouraged by the relatively good validation results.

An important component of the research reported in the present paper was the comparison of different methods for feature extraction. Key conclusions were:

- Of the feature sets tested, the best performance was achieved using Zernike moment magnitude and phase features.
- Performance of Zernike moment magnitude and phase features was equally good irrespective of whether the data matrices were rotated prior to feature extraction or not. Use of costly rotation procedures is therefore not necessary.
- The best performance was achieved by directly extracting features from the whole of the region of interest (the firing-pin impression plus the flowback region plus the breech-face region), rather than by any process that involved separately segmenting the firing-pin impression and the breech-face region.
- In the context of casework involving comparison of a fired cartridge case recovered from a crime scene with cartridges fired from a seized firearm, using 3 cartridges fired from the seized firearm would appear to be sufficient to achieve good results. Use of a larger number of fires per firearm would, however, be advisable for system training and validation.

In future work aimed at developing better performing systems, we will therefore use Zernike moment magnitude and phase features extracted from the whole of the region of interest without rotation of data matrices prior to feature extraction.

Planned future work includes expanding the size of the database to the point where it will be sufficient for training a DNN-embedding based system, which is currently the state-of-the-art approach in forensic voice comparison, and which is expected to lead to substantial improvements in system performance. Planned future work will also ultimately include field testing by practitioners of a later version of the system.

Disclaimer

All opinions expressed in the present paper are those of the authors, and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the authors are associated.

Author contributions

Nabanita Basu: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Rachel S Bolton-King:** Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing - Review & Editing. **Geoffrey Stewart Morrison:** Conceptualization, Funding acquisition, Methodology, Supervision, Visualization, Writing - Original Draft, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by Research England's Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2023.

Thanks to Dr Michael Derenovskiy and his colleagues at ScannBI Technology Europe GmbH for the loan of the Evofinder® imaging system.

Thanks to the organizations and individuals who donated the fired cartridge cases. To maintain their anonymity, we do not thank them by name.

References

[1] G.S. Morrison, Advancing a paradigm shift in evaluation of forensic evidence: the rise of forensic data science, Forensic Sci. Int.: Synergy 4 (2022), 100270, https://doi.org/10.1016/j.fsisyn.2022.100270.

[2] R.S. Bolton-King, Preventing miscarriages of justice: a review of forensic firearm identification, Sci. Justice 56 (2016) 129–142, https://doi.org/10.1016/j.scijus.2015.11.002 [Corrigendum: (2018) 58, 83.].

[3] R. Nichols, Firearm and Tool Mark Identification: the Scientific Reliability of the Forensic Science Discipline, Academic Press, London, UK, 2018, https://doi.org/10.1016/C2016-0-04649-1.

[4] W.A. Tobin, P. Blau, Hypothesis testing of the critical underlying premise of discernible uniqueness in firearms-toolmarks forensic practice, Jurimetrics 53 (2013) 121–142. https://ssrn.com/abstract=2185742.

[5] X.H. Tai, W.F. Eddy, Automatically Matching Topographical Measurements of Cartridge Cases Using a Record Linkage Framework, 2020. http://arxiv.org/abs/2003.00060.

[6] T.P. Smith, G.A. Smith, J.B. Snipes, A validation study of bullet and cartridge case comparisons using samples representative of actual casework, J. Forensic Sci. 61 (2016) 939–946, https://doi.org/10.1111/1556-4029.13093.

[7] E.J.A.T. Mattijssen, C.L.M. Witteman, C.E.H. Berger, N.W. Brand, R.D. Stoel, Validity and reliability of forensic firearm examiners, Forensic Sci. Int. 307 (2020), 110112, https://doi.org/10.1016/j.forsciint.2019.110112.

[8] E.J.A.T. Mattijssen, C.L.M. Witteman, C.E.H. Berger, X.A. Zheng, J.A. Soons, R. D. Stoel, Firearm examination: examiner judgments and computer-based comparisons, J. Forensic Sci. 66 (2021) 96–111, https://doi.org/10.1111/1556-4029.14557.

[9] N. Scurich, B.L. Garrett, R.M. Thompson, Surveying practicing firearm examiners, Forensic Sci. Int.: Synergy 4 (2022), 100228, https://doi.org/10.1016/j.fsisyn.2022.100228.

[10] P. Thumwarin, C. Prasit, P. Boonbumroong, T. Matsuura, Firearm identification based on FIR system characterizing rotation invariant feature of cartridge case image, in: Proceedings of the 2008 23rd International Conference Image and Vision Computing New Zealand, 2008, https://doi.org/10.1109/IVCNZ.2008.4762085.

[11] C.Y. Liong, N.A.M. Ghani, S.B.A. Kamaruddin, A.A. Jemain, Firearm classification based on numerical features of the firing pin impression, Procedia Comput. Sci. 13 (2012) 144–151, https://doi.org/10.1016/j.procs.2012.09.123.

[12] D. Ott, R. Thompson, J. Song, Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests, Forensic Sci. Int. 271 (2017) 98–106, https://doi.org/10.1016/j.forsciint.2016.12.014.

[13] K. Addinall, W. Zeng, P. Bills, P.T. Wilcock, L. Blunt, The effect of primer cap material on ballistic toolmark evidence, Forensic Sci. Int. 298 (2019) 149–156, https://doi.org/10.1016/j.forsciint.2019.02.054.

[14] L.P. Xin, J. Zhou, G. Rong, A cartridge identification system for firearm authentication, in: Proceedings of the 5th International Conference on Signal Processing/Proceedings of the 16th World Computer Congress, 2000, https://doi.org/10.1109/icosp.2000.891807.

[15] A. Legrá, E. Marañón, H. Pérez, L. de la Torre, A. Quintana, R. Quirós, Automatic identification of weapons from images of the cartridge case head, in: *Proceedings of the 7th WSEAS International Conference on Applied Computer and Applied Computational Science* (ACACOS'08), 2008, pp. 236–241. https://www.researchgate.net/publication/236109928.

[16] T.G. Fadul Jr., G.A. Hernández, S. Stoiloff, S. Gulati, An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides, 2012. Report for National Institute of Justice Award Number 2009-DN-BX-K230, https://www.ojp.gov/pdffiles1/nij/grants/237960.pdf.

[17] K. Zhang, Y. Luo, Slight variations of breech face marks and firing pin impressions over 3070 consecutive firings evaluated by Evofinder®, Forensic Sci. Int. 283 (2018) 85–93, https://doi.org/10.1016/j.forsciint.2017.11.035.

[18] E.F. Law, K.B. Morris, C.M. Jelsema, Determining the number of test fires needed to represent the variability present within 9mm Luger firearms, Forensic Sci. Int. 276 (2017) 126–133, https://doi.org/10.1016/j.forsciint.2017.04.019.

[19] L. Lightstone, The potential for and persistence of subclass characteristics on the breech faces of SW40VE Smith and Wesson Sigma pistols, Assoc. Firearm Tool mark Exam. J. 42 (4) (2010) 308–322.

[20] D. LaPorte, An empirical and validation study of breechface marks on .380 ACP caliber cartridge cases fired from ten consecutively finished Hi-Point Model C9 pistols, Assoc. Firearm Tool mark Exam. J. 43 (4) (2011) 303–309.

[21] J. Song, T.V. Vorburger, W. Chu, J. Yen, J.A. Soons, D.B. Ott, N.F. Zhang, Estimating error rates for firearm evidence identifications in forensic science, Forensic Sci. Int. 284 (2018) 15–32, https://doi.org/10.1016/j.forsciint.2017.12.013.

[22] J. Zhou, L.P. Xin, D.S. Gao, C.S. Zhang, D. Zhang, Automated cartridge identification for firearm authentication, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. I–749, https://doi.org/10.1109/cvpr.2001.990551. –I-754.

[23] D.G. Li, Image processing for the positive identification of forensic ballistics specimens, in: Proceedings of the Sixth International Conference of Information Fusion, 2003, pp. 1494–1498, https://doi.org/10.1109/ICIF.2003.177417.

[24] C. Gambino, P. McLaughlin, L. Kuo, F. Kammerman, P. Shenkin, P. Diaczuk, N. Petraco, J. Hamby, N.D.K. Petraco, Forensic surface metrology: tool mark evidence, Scanning 33 (2011) 272–278, https://doi.org/10.1002/sca.20251.

[25] N.D.K. Petraco, H. Chan, P.R. De Forest, P. Diaczuk, C. Gambino, J. Hamby, F. L. Kammerman, B.W. Kamrath, T.A. Kubic, L. Kuo, P. McLaughlin, G. Petillo, N. Petraco, E.W. Phelps, P.A. Pizzola, D.K. Purcell, P. Shenkin, Application of Machine Learning to Toolmarks: Statistically Based Methods for Impression Pattern Comparisons, Report for National Institute of Justice Award, 2011. Number 2009-DN-BX-K041, https://www.ncjrs.gov/pdffiles1/nij/grants/239048.pdf.

[26] Y. Pan, Z. Chen, M. Tong, X. Zhao, Extraction of individual characteristics of breech face impressions in ballistic identification using optimal Gaussian filter parameters, in: *Proceedings of the 11th International Conference on Computer Science & Education* (ICCSE), 2016, pp. 519–523, https://doi.org/10.1109/ICCSE.2016.7581634.

[27] N.A.M. Ghani, C.Y. Liong, A.A. Jemain, Analysis of geometric moments as features for firearm identification, Forensic Sci. Int. 198 (2010) 143–149, https://doi.org/10.1016/j.forsciint.2010.02.011.

[28] Z.L. Chuan, A.A. Jemain, C.Y. Liong, N.A.M. Ghani, L.K. Tan, A robust firearm identification algorithm of forensic ballistics specimens, J. Phys. Conf. 890 (2017), 012126, https://doi.org/10.1088/1742-6596/890/1/012126.

[29] J. Leng, Z. Huang, On analysis of circle moments and texture features for cartridge images recognition, Expert Syst. Appl. 39 (2012) 2092–2101, https://doi.org/10.1016/j.eswa.2011.08.003.

[30] R. Fischer, C. Vielhauer, Digital crime scene analysis: automatic matching of firing pin impressions on cartridge bottoms using 2D and 3D spatial features, in: Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec '14), 2014, pp. 77–82, https://doi.org/10.1145/2600918.2600930.

[31] R. Fischer, C. Vielhauer, Automated firearm identification: on using a novel multiple-slice-shape (MSS) approach for comparison and matching of firing pin impression topography, in: Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec '15), 2015, pp. 161–171, https://doi.org/10.1145/2756601.2756619.

[32] K.B. Morris, E.F. Law, R.L. Jefferys, E.C. Dearth, Interpretation of cartridge case evidence using IBIS and Bayesian networks, Report on research conducted under Cooperative Agreement Number W911NF-12-2-0056, https://www.ncjrs.gov/App/AbstractDB/AbstractDBDetails.aspx?id=272547, 2016.

[33] D.G. Li, A new approach for firearm identification with hierarchical neural networks based on cartridge case images, in: Proceedings of the 2006 5th IEEE International Conference on Cognitive Informatics, 2006, pp. 923–928, https://doi.org/10.1109/COGINF.2006.365616.

[34] N.A.M. Ghani, C.Y. Liong, A.A. Jemain, Neurocomputing approach for firearm identification, Pertanika J. Sci. Technol. 26 (2018) 341–352. http://www.pertanika.upm.edu.my/pjst/browse/regular-issue?article=JST-S0297-2017.

[35] O. Giudice, L. Guarnera, A.B. Paratore, G.M. Farinella, S. Battiato, Siamese ballistics neural network, in: *Proceedings of the 2019 IEEE International Conference on Image Processing* (ICIP), 2019, pp. 4045–4049, https://doi.org/10.1109/ICIP.2019.8803619.

[36] N.A. Razak, C.-Y. Liong, A.A. Jemain, N.A.M. Ghani, S. Zakaria, Automatic firing pin impression identification based on feature fusion of fractal dimension and geometric moment, J. Telecommun. Electron. Comput. Eng. 12 (2) (2020) 7–10. https://jtec.utem.edu.my/jtec/article/view/5823.

[37] J. Roth, A. Carriveau, X. Liu, A.K. Jain, Learning-based ballistic breech face impression image matching, in: Proceedings of the 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015, https://doi.org/10.1109/BTAS.2015.7358774.

[38] J. Song, Proposed "congruent matching cells (CMC)" method for ballistic identification and basic concepts valid and invalid correlation region, Assoc. Firearm Tool mark Exam. J. 47 (3) (2015) 177–185. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=911193.

[39] X.H. Tai, W.F. Eddy, A fully automatic method for comparing cartridge case images, J. Forensic Sci. 63 (2018) 440–448, https://doi.org/10.1111/1556-4029.13577.

[40] N.F. Zhang, The use of correlated binomial distribution in estimating error rates for firearm evidence identification, J. Res. Natl. Inst. Stand. Technol. 124 (2019), 124026, https://doi.org/10.6028/jres.124.026.

[41] J. Song, T.V. Vorburger, L. Ma, J.M. Libert, S.M. Ballou, A metric for the comparison of surface topographies of standard reference material (SRM) bullets and casings, in: Proceedings of the American Society for Precision Engineering, 2005 (ASPE), https://www.nist.gov/publications/metric-comparison-surface-topographies-standard-reference-material-srm-bullets-and.

[42] H. Zhang, J. Song, M. Tong, W. Chu, Correlation of firing pin impressions based on congruent matching cross-sections (CMX) method, Forensic Sci. Int. 263 (2016) 186–193, https://doi.org/10.1016/j.forsciint.2016.04.015.

[43] H. Zhang, J. Zhu, R. Hong, H. Wang, F. Sun, A. Malik, Convergence-improved congruent matching cells (CMC) method for firing pin impression comparison, J. Forensic Sci. 66 (2021) 571–582, https://doi.org/10.1111/1556-4029.14634.

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

[45] M. Tong, Y. Pan, Z. Li, W. Lin, Valid data based normalized cross-correlation (VDNCC) for topography identification, Neurocomputing 308 (2018) 184–193, https://doi.org/10.1016/j.neucom.2018.04.059.

[46] M. Tong, X. Yu, S. Huang, Automatic identification of firing pin impressions based on the Congruent Matching Cell (CMC) method, Neurocomputing 367 (2019) 246–258, https://doi.org/10.1016/j.neucom.2019.08.033.

[47] F. Riva, C. Champod, Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases, J. Forensic Sci. 59 (2014) 637–647, https://doi.org/10.1111/1556-4029.12382.

[48] F. Dong, Y. Zhao, Y. Luo, W. Zhang, K. Zhang, Specificity of characteristic marks on cartridge cases from 3070 consecutive firings of a Chinese Norinco QSZ-92 9 mm Pistol, J. Forensic Sci. Med. 5 (2) (2019) 87–94, https://doi.org/10.4103/jfsm.jfsm_6_19.

[49] F. Riva, E.J.A.T. Mattijssen, R. Hermsen, P. Pieper, W. Kerkhoff, C. Champod, Comparison and interpretation of impressed marks left by a firearm on cartridge cases – towards an operational implementation of a likelihood ratio based technique, Forensic Sci. Int. 313 (2020), 110363, https://doi.org/10.1016/j.forsciint.2020.110363.

[50] J. Song, Z. Chen, T.V. Vorburger, J.A. Soons, Evaluating likelihood ratio (LR) for firearm evidence identifications in forensic science based on the Congruent Matching Cells (CMC) method, Forensic Sci. Int. 317 (2020), 110502, https://doi.org/10.1016/j.forsciint.2020.110502.

[51] G.S. Morrison, E. Enzinger, Score based procedures for the calculation of forensic likelihood ratios – scores should take account of both similarity and typicality, Sci. Justice 58 (2018) 47–58, https://doi.org/10.1016/j.scijus.2017.06.005.

[52] C. Neumann, M. Ausdemore, Defence against the modern arts: the curse of statistics – part II: 'Score-based likelihood ratios, Law Probab. Risk 19 (2020) 21–42, https://doi.org/10.1093/lpr/mgaa006.

[53] C. Neumann, J. Hendricks, M. Ausdemore, Statistical support for conclusions in fingerprint examinations, in: D. Banks, K. Kafadar, D.H. Kaye, M. Tackett (Eds.), Handbook of Forensic Statistics, CRC, Boca Raton, FL, 2020, pp. 277–324, https://doi.org/10.1201/9780367527709.

[54] D.M. Ommen, C.P. Saunders, A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio, Stat. Sci. 36 (2021) 344–359, https://doi.org/10.1214/20-STS805.

[55] Y. Wang, Class characteristic classification of test fired cartridge cases: a digital image decision tree approach to Kensington's matrix for initial stages of criminal investigation, J. Forensic Sci. Crim. Invest. 6 (2017), 555693, https://doi.org/10.19080/JFSCI.2017.06.555693.

[56] G.S. Morrison, E. Enzinger, D. Ramos, J. González-Rodríguez, A. Lozano-Díez, Statistical models in forensic voice comparison, in: D. Banks, K. Kafadar, D. H. Kaye, M. Tackett (Eds.), Handbook of Forensic Statistics, CRC, Boca Raton, FL, 2020, pp. 451–497, https://doi.org/10.1201/9780367527709.

[57] G.S. Morrison, P. Weber, E. Enzinger, B. Labrador, A. Lozano-Díez, D. Ramos, J. González-Rodríguez, Forensic voice comparison – human-supervised-automatic approach, in: M. Houck, L. Wilson, S. Lewis, H. Eldridge, P. Reedy, K. Lothridge (Eds.), Encyclopedia of Forensic Sciences, third ed., Elsevier, 2022. https://www.elsevier.com/books/encyclopedia-of-forensic-sciences/houck/978-0-12-823677-2. http://forensic-voice-comparison.net/encyclopedia/. available at, In press, A preprint is available at:.

[58] P. Weber, E. Enzinger, B. Labrador, A. Lozano-Díez, D. Ramos, J. González-Rodríguez, G.S. Morrison, Validation of the alpha version of the E^3 Forensic Speech Science System (E^3FS^3) core software tools, Forensic Sci. Int.: Synergy 4 (2022), 100223, https://doi.org/10.1016/j.fsisyn.2022.100223.

[59] F. Zernike, Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode, Physica 1 (1934) 689–704, https://doi.org/10.1016/S0031-8914(34)80259-5.
[60] M.R. Teague, Image analysis via the general theory of moments, J. Opt. Soc. Am. 70 (1980) 920–930, https://doi.org/10.1364/JOSA.70.000920.
[61] A. Khotanzad, Y.H. Hong, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 489–497, https://doi.org/10.1109/34.55109.
[62] D.R. Iskander, M.J. Collins, B. Davis, Optimal modeling of corneal surfaces with Zernike polynomials, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 48 (2001) 87–95, https://doi.org/10.1109/10.900255.
[63] M. Sun, J. Birkenfeld, A. de Castro, S. Ortiz, S. Marcos, OCT 3-D surface topography of isolated human crystalline lenses, Biomed. Opt Express 5 (2014) 3547–3561, https://doi.org/10.1364/BOE.5.003547.
[64] S.V. Pinhasi, R. Alimi, S. Eliezer, L. Perelmutter, Fast optical computerized topography, Phys. Lett. 374 (2010) 2798–2800, https://doi.org/10.1016/j.physleta.2010.04.05.
[65] N. Vretos, N. Nikolaidis, I. Pitas, 3D facial expression recognition using Zernike moments on depth images, in: Proceedings of the 18th IEEE International Conference on Image Processing, 2011, pp. 773–776, https://doi.org/10.1109/ICIP.2011.6116669.
[66] C. Teh, R.T. Chin, On image analysis by the methods of moments, IEEE Trans. Pattern Anal. Mach. Intell. 10 (1988) 496–513, https://doi.org/10.1109/34.3913.
[67] S.O. Belkasim, M. Shridhar, M. Ahmadi, Pattern recognition with moment invariants: a comparative study and new results, Pattern Recogn. 24 (1991) 1117–1138, https://doi.org/10.1016/0031-3203(91)90140-Z.
[68] M.-K. Hu, Visual pattern recognition by moment invariants, IEEE Trans. Inf. Theor. 8 (2) (1962) 179–187, https://doi.org/10.1109/TIT.1962.1057692.
[69] J. Flusser, On the independence of rotation moment invariants, Pattern Recogn. 33 (2000) 1405–1410, https://doi.org/10.1016/S0031-3203(99)00127-2.
[70] J. Flusser, T. Suk, Rotation moment invariants for recognition of symmetric objects, IEEE Trans. Image Process. 15 (2006) 3784–3790, https://doi.org/10.1109/TIP.2006.884913.
[71] K. Pearson, On lines and planes of closest fit to systems of points in space, Lond. Edinb.Dublin Phil. Mag. J. Sci. 2 (1901) 559–572, https://doi.org/10.1080/14786440109462720.
[72] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (6) (1933) 417–441, https://doi.org/10.1037/h0071325.
[73] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eug. 7 (1936) 179–188, https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.
[74] C.R. Rao, The utilization of multiple measurements in problems of biological classification, J. Roy. Stat. Soc. B 10 (1948) 159–203. http://www.jstor.org/stable/2983775.
[75] S.J.D. Prince, J.H. Elder, Probabilistic linear discriminant analysis for inferences about identity, in: Proceedings of the IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8, https://doi.org/10.1109/ICCV.2007.4409052.
[76] P. Kenny, Bayesian speaker verification with heavy tailed priors, in: Proceedings of Odyssey 2010: the Speaker and Language Recognition Workshop, 2010 paper 014, https://www.isca-speech.org/archive_open/odyssey_2010/od10_014.html.
[77] N. Brümmer, E. de Villiers, The speaker partitioning problem, in: Proceedings of Odyssey 2010: the Speaker and Language Recognition Workshop, 2010, pp. 194–201. https://www.isca-speech.org/archive_open/odyssey_2010/od10_034.html.
[78] A. Sizov, K.A. Lee, T. Kinnunen, Unifying probabilistic linear discriminant analysis variants in biometric authentication, in: P. Fränti, G. Brown, M. Loog, F. Escolano, M. Pelillo (Eds.), Structural, Syntactic, and Statistical Pattern Recognition, Springer, Berlin, 2014, pp. 464–475, https://doi.org/10.1007/978-3-662-44415-3_47.
[79] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Appl. Stat. 53 (2004) 109–122, https://doi.org/10.1046/j.0035-9254.2003.05271.x. https://doi.org/10.1111/j.1467-9876.2004.02031.x [Corrigendum: (2004) 53, 665–666.].
[80] J. González-Rodríguez, P. Rose, D. Ramos, D.T. Toledano, J. Ortega-García, Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition, IEEE Trans. Speech Audio Process. 15 (2007) 2104–2115, https://doi.org/10.1109/TASL.2007.902747.
[81] G.S. Morrison, Tutorial on logistic-regression calibration and fusion: converting a score to a likelihood ratio, Aust. J. Forensic Sci. 45 (2013) 173–197, https://doi.org/10.1080/00450618.2012.733025.
[82] G.S. Morrison, N. Poh, Avoiding overstating the strength of forensic evidence: shrunk likelihood ratios/Bayes factors, Sci. Justice 58 (2018) 200–218, https://doi.org/10.1016/j.scijus.2017.12.005.
[83] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Natl. Bur. Stand. 49 (1952) 409–436, https://doi.org/10.6028/jres.049.044.
[84] T.P. Minka, A comparison of numerical optimizers for logistic regression, Technical report, https://tminka.github.io/papers/logreg/, 2003.
[85] G.S. Morrison, E. Enzinger, V. Hughes, M. Jessen, D. Meuwly, C. Neumann, S. Planting, W.C. Thompson, D. van der Vloed, R.J.F. Ypma, C. Zhang, A. Anonymous, B. Anonymous, Consensus on validation of forensic voice comparison, Sci. Justice 61 (2021) 229–309, https://doi.org/10.1016/j.scijus.2021.02.002.
[86] N. Brümmer, J. du Preez, Application independent evaluation of speaker detection, Comput. Speech Lang 20 (2006) 230–275, https://doi.org/10.1016/j.csl.2005.08.001.
[87] D. Meuwly, Reconnaissance de locuteurs en sciences forensiques: l'apport d'une approche automatique, Doctoral dissertation, University of Lausanne, 2001. https://www.unil.ch/files/live/sites/esc/files/shared/These.Meuwly.pdf.