

I. Introduction

A. Motivation

1. part of the work a firearms and toolmark examiner has to do is to assess evidence for its source. Generally, a pair of cartridge cases or bullets (from the crime scene or retrieved in the lab from a suspect's firearm) is being checked whether or not the patterns on the items are similar enough to be coming from the same source. FTEs are testifying in court according to the AFTE rules of identification (include citation) XXX
 - a. Show bullet and explain different parts.
2. PCAST report and NRC have criticized a lot of forensic for its lack of scientific validity - FTE examinations are subjective and do not have established error rates. XXX cite reports

B. Background- Current Practices?

1. Paper by Hare et al introduces an algorithm for an automatic matching method of bullets based on 3d scans of land engraved areas (LEAs): objective and allows establishing error rates for different situations based on studies (firearm/ammunition combination) XXX overview picture?
2. Part of the matching process in the Hare et al paper is to locate the location of the groove engraved areas and separate from the land engraved areas. The rollapply method suggested in the paper is not working 100% and leads to erroneous conclusions down stream in the analysis.
3. Chu et al (2013?) uses Canny Edge detection to sharpen the image of striae in a bullet land and determine "valid striae". Previous work by these authors also highlight the use of CCF for finding the effective correlation area which is used to generate an average profile of the bullet signature.(Chu et al. 2010)

C. Background- Hough Transform and Image Prep

1. Data comes in form of scans, scans are a result of high-resolution 3d microscopy; each scan consists of height measurements collected over a regular x-y grid. This makes the data somewhat similar to a (grey-scale) image. More information needed about data sources
 - a. Orientation of the image differs from traditional views of the x-y axis due to the way images are stored in memory connect to CImage library
 - b. Show bullet scan and explain parts.
2. The Hough Transform is a computer vision algorithm that detects certain shapes or features in a 2-dimensional image
 - a. In our case we are interested in detecting lines. Traditionally we think of lines being defined on the x-y axis in terms of both slope and intercept. This means every point on a line can be defined as a line in the feature space.
 - b. Points on the same line are then intersections in the feature space. XXX tikz plot here XXX
 - c. Users can select the number of bins that divide the feature space. Each line in the feature space is associated with a set of two parameters. This

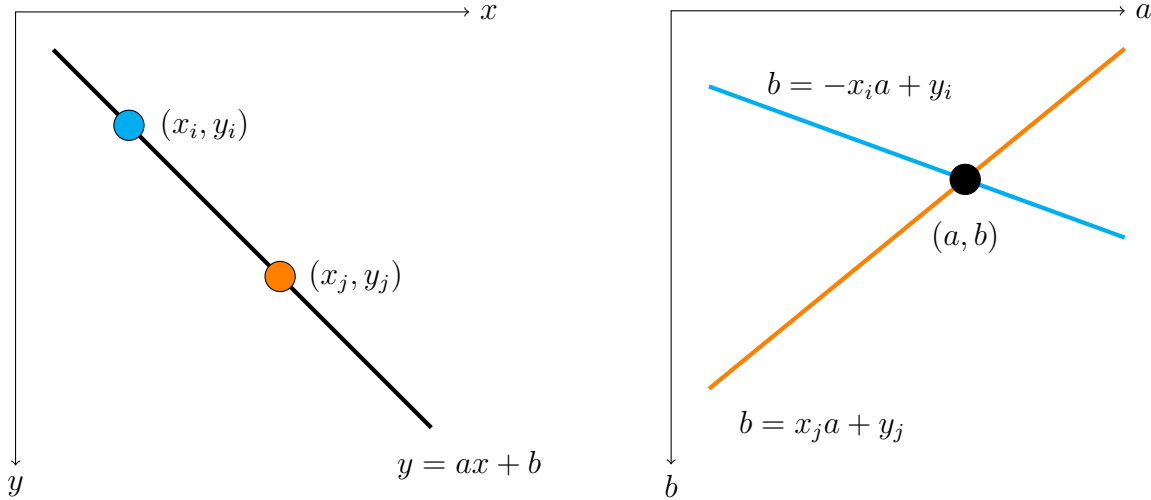


Figure 1: Diagram of feature space linea transformation oriented for image origin.

can be visually represented as a grid with the number of squares in the grid determined by the bins. Each time a line is detected that falls within one of the feature space bins, that bin is incremented. INSERT TIKZ PLOT HERE OF FEATURE SPACE GRID

- d. The resulting output of Hough transforms in our case is thus a series of linear equations with a corresponding count of the times this particular set of features was detected. Essentially the Hough score, as it is often called, counts the number of points detected in the image that all line on the same line.
- e. We note that Hough lines are parametrized in the Hessian Normal Form. Instead of the line being defined by the slope and intercept, it is defined by two parameters θ and ρ . XXX insert tikzpicture here of hessian normal form oriented over bullet land image XXX

D. Evaluation of Results

1. To evaluate the success of our results we will utilize a method developed by Kiegan Rice that calculates the "area of misidentification"
2. A series of consecutive crosscuts is taken from the bullet land based on the optimized crosscut from 'x3p_crosscut_optimize', five rows of data above and below the optimized crosscut. An average is then created from these crosscuts.
3. A robust loess fit is then created from the averaged crosscuts, a prediction is made based on the robust loess fit on our summarized crosscut. We then calculate the residual of the robust loess fit and the absolute residual
4. We then calculate the groove locations using the Hough method and input the manually identified groove locations. To calculate the area of misidentification we take the sum of the absolute value of the robust loess fit residuals between the identified grooves location and the predicted groove location and multiply it by the resolution of the x3p object. XXX need nice picture explaining differenceXXX

II. Methods

A. Application of Hough transform and Pre-processing

1. Scans can be inputted as an x3p or as a dataframe, but needs to be oriented in the proper fashion. i.e. width is larger than height. We then save some of the x3p information such as width, height, and resolution.
2. We have a significant amount of missing data in our scans, whether it is from break off or because the light of the high resolution scans cannot penetrate far enough into grooves. We have decided to impute these NAs by replacing them with z-values that are 5% higher than the maximum height found in the scan.
2. Scans are then converted to CImages where a gradient image in the x and y direction are created. We then find the gradient magnitude of each image.
2. Each gradient image is thresholded so only strong features are selected, those in the 99th percentile and the Hough transform is applied from the `image` packages ("`hough_line`"). The output of the '`hough_line`' function is a dataframe containing ρ , θ , and a score.

B. Line selection and Output

1. We eliminate all duplicate lines detected in the image and we transform the θ parameter to be between $-\pi$ and π rather than between 0 to 2π . This allows us to filter any lines that are not roughly vertical, from this information we can then find both the x intercept and the y intercept of the each line detected in the image.
2. We then calculate what is known as x-bottom and x-top which is where each line detected in the image intersects with the top and bottom of the bullet land. This can then be used to calculate the slope of each line with respect to the y-direction. Explain why the y-direction is more robust. XX insert picture of calculating xtop and xbottom from over bullet landXX
3. As mentioned before the score associated with the Hough transform is essentially the number of points detected on each line in the image so longer lines will necessarily have a higher possible maximum score. To account for this we create a normalized score. XX insert normalized score graphic hereXXX
4. Segments are then separated into right hand and left hand segments (take the left most fifty percent of the data etc.). We then select the line with the largest normalized score on each side of the image. If the line with the largest normalized score is further to the left than the left most fourth of the data, we arbitrarily set the groove at the lower 1/4th of the image. A similar process is undertaken for the right hand groove.
5. The final output is a function to describe the detected groove in microns.

IV. Results Some sort of presentation of scores from Phoenix, Hamby, and Houston showing similar patterns across each