Response to reviewers

Eric Hare, Heike Hofmann, Alicia Carriquiry

June 19, 2017

1. Page 2, line 46-47: I do not understand the phrase, "the transcription of the individual characteristics into a 2D plotting framework." Please write the idea more clearly.

   Thank you. We have clarified this statement.

2. Figures. 1, 3, 4, 5, 6, 7, 8, 9, 10, 12b, 12c, 14: These figures all have labels that are too small to be readable or they lack units or both. The authors must fix all of them in one way or the other. Without the units, the reader cannot be sure whether the scales along the x direction represent pixels (data points) or um.

   We have updated the figures with more readable labels and with units. Thanks for the suggestion.

3. Page 8, lines 207-209: I do not understand the sentence, "Here, d is defined as the distance between xi and the location of the fit xo and the maximum distance of the range of the x-values for span alpha in xo." It appears that the quantity d has two definitions, "distance between xi and the location of the fit xo" and "maximum distance of the range of the x-values". This cannot be correct. Please re-write.

   Apologies for the confusion. This was intended to convey the fact that in the traditional definition, the value d is in an interval between 0 and 1, and hence to map our distances to this scale, we divide by the range of the x values. We have rewritten this statement to hopefully make it more clear.

4. Page 11, line 240: Something is missing in the description of Step 3. How can the calculation of a rolling average by itself yield a criterion for identifying peaks and valleys in the profile? Please rewrite the description of the procedure more completely.

   We have added some clarification to this point. The rolling average is then used to determine peaks and valleys by identifying points at which the derivative is equal to zero. Given sufficient smoothing, this should correspond to identification of a striation marking.

5. Page 13, figure 7: Even though the figures are tiny and the scales are too small, I should be able to discern a shift of (1.5625 x 17 um) = 27 um in one of the profiles between graphs 7a and 7b. Instead the two graphs look identical. I think the wrong graph was plotted somewhere.

   You are correct below that the actual optimal lag was -2. We have corrected this error.

6. Page 13, figure 8: The graph does not agree with its caption. The caption states that correlation is maximized at a lag of minus 17 whereas the graph clearly shows the maximum at a lag of about minus 2. There must be an error to rectify here. If the lag is only 2, that would answer the issue I raised about figure 7.

   As stated above, you are correct. The caption has been corrected.

7. Page 20, figure 16: Why are the AUC values the same (0.91) for all graphs although the graphs are all different. This seems to result from clerical errors.

   This clerical error has been corrected and the labels are now accurate.

8. Page 22, figure 19: The perfect separation observed between matching and non-matching data points is not consistent with the modest predicted probabilities shown on the abscissa. For example, at predicted probabilities of 0.6 to 0.75, I would have expected some dark blue non-matching data points to appear

on the right side of the graph. It would be good to explain what "predicted probability" means precisely for this graph.

The primary difference in the numbers comes from the fact that one is an out-of-bag error rate, i.e., an error rate computed on the land comparisons that were held out of the trees during the computation of the random forest. Meanwhile, figure 19 is the predicted probability on all land-to-land comparisons provided by our final forest.

9. Page 3, lines 68-70: The authors' term for the transition region between land impressions and groove impressions on a bullet surface is unconventional. The authors have chosen to use the word "groove" instead of "shoulder" to describe the transition region between the land impression and the groove impression areas. This could be confusing to a firearms examiner.

   We have carefully gone through the paper and adjusted the language so that it will be more clear to a firearms examiner.

10. Page 5, line 142: For better clarity, the authors could refer to Fig. 1 when describing the coordinate "fixed height x along the bullet land".

    We have included this reference.

11. Page 6, line 168: I think it would avoid confusion later if the authors make it clear that the smoothing factor s is specified in terms of data points and not in terms of distance. Adding "(pixels)" or (data points)" to the number "35" on this line would make the difference more explicit at the start.

    This does make it much more clear, and we have made the change.

12. Page 11, lines 240-254: There appear to be logical gaps in the explanations for Steps 3 and 4. It is difficult to grasp the difference between "common" intervals and "joint" intervals. From Step 3 it seems clear that common intervals must be specified for a pair of profiles, whereas joint intervals seem to be specified for single profiles. This leads to confusion about why the region labelled "X" in Fig. 6c is not considered to contain matching striations when the intervals on both profiles overlap.

    We hope we have clarified portions of these steps sufficiently. First, we've made it more clear about how the peak and valley detection operates. Secondly, you are correct that the region labeled X in Figure 6c was an error. This has since been correct, and hence your intuition about common intervals was in fact correct. If there are further steps that can be taken to improve this portion, please let us know.

13. Page 14, line 334: The phrase "an equidistant resolution" is confusing here. I think that "equal sampling intervals" would be more conventional.

    We agree, and have accordingly made this change.

14. Page 15, line 372: The authors should supply a reference for the term "tank rash."

    This reference has now been included (Hamby 2009).

15. Page 18, figure caption 13a: Change "bullet C-3" to "bullet land C-3".

    Done.

16. Page 19, figure 15: The histograms seem unrealistically smooth, particularly the histograms for matches. How were they calculated?

    These are actually density plots rather than histograms. Essentially, they're curve fits to the actual histograms. This is why they appear smoother than the raw results would appear.

17. Supplement, page 6, figure caption 4: The caption claims a max CMS (consecutive matching striae) of 22, but I was only able to count a max CMS of 11. How is CMS defined here?

The location of the Os at the bottom of the figure identify the locations that our algorithm detects a striation marking. The primary reason for the discrepancy is that we treat valleys as a striation marking for the purpose of the CMS computation as well as peaks. This is described further on page 12 in the main paper.