J26 Pairwise Comparison Scores for Handwritten Questioned Documents

Danica Ommen, PhD, Iowa State University Department of Statistics, Ames, IA 50011; Cami Fuglsby, MS, Department of Mathematics and Statistics, Brookings, SD 57007; Christopher P. Saunders, PhD, Brookings, SD 57006; Michael Caligiuri, PhD, University of California, San Diego, CA 92093; Linton Mohammed, PhD, Forensic Science Consultants, Inc, Burlingame, CA 94010-2017; JoAnn Buscaglia, PhD, FBI Laboratory, CFSRU, Quantico, VA 22135*

Learning Overview: After attending this presentation, attendees will better understand the value of pairwise comparison methods for designing tests of scientific validity for forensic feature comparison methods.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating an alternative use for pairwise comparison methods beyond automated source identification applications.

Hypothesis Statement: Methods resulting in pairwise comparison scores can be used in designing tests of scientific validity for a variety of forensic feature comparison methods, specifically in the case of validating questioned document examiners' conclusions.

In 2016, the President's Council of Advisors on Science and Technology released a report that advocated for further research into the scientific validity of all forensic feature comparison methods, including methods for handwriting evidence.[1] One of the research initiatives identified by the Organization of Scientific Area Committees, overseen by the National Institute of Standards and Technology, in response to this report seeks to validate the conclusions expressed by forensic document examiners regarding the weight or the value of the evidence in a particular comparison. To design the experiments needed to test examiners' conclusions, it is often necessary to pair handwritten documents with various levels of correspondence for the test comparisons performed by the examiners. In this way, the document examiners' conclusions will not be skewed by receiving, for example, only very easy comparisons. Since this pairing needs to be done quickly and effectively, we have been developing an automated method of comparing two documents that would produce a "score" indicating the level of similarity between two documents. Scores that are high indicate writing that is visually similar, and scores that are low indicate writing that is visually dissimilar. This score can then be used to provide the examiners with a range of difficulty in the test comparisons. The idea is that the person designing the test comparisons can pick out test comparisons that are from different writers with a high score to provide examiners with difficult exclusions, and test comparisons that are from the same writer with a low score to provide examiners with difficult inclusions. The test designer may also want to select easier test comparisons for both inclusions and exclusions, and a variety of pairs for comparison with moderate levels of difficulty.

In this presentation, a pairwise comparison score for images of scanned handwritten questioned documents using both open-source and proprietary automated image feature-detection methods will be developed. The Scale-Invariant Feature Transform (SIFT) and the Speeded-Up Robust Features (SURF) algorithms have demonstrated potential to identify important features of images for the purpose of comparing two general images of any content.[2] Woodard et al. demonstrated the use of these open-source algorithms for automated writer identification systems in the case of Arabic handwriting.[3] Gannon Technologies Group developed the proprietary FLASH ID algorithm to perform a similar writer identification task, regardless of the written language.[4] However, the algorithms first need to be trained on a set of images to create a dictionary/codebook of similar features to look for in the test images. Once the algorithm is trained, a single image is provided as the input, then a numerical representation of the important features in that image is output to the user. A method of comparing these numerical outputs to produce a similarity score indicating the level of correspondence between two scanned images of handwriting (cursive or print) will be demonstrated. The use of the scores on a collection of handwritten and handprinted English-language documents collected as part of an ongoing research effort will also be presented.

Reference(s):

1. President's Council of Advisors on Science and Technology. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.* Washington, DC: Executive Office of The President's Council of Advisors on Science and Technology, https://www.whitehouse.gov/administration/eop/ostp/pcast/docs-reports (2016).
2. Lowe, D. (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, pp. 91-110.
3. Woodard, J., Lancaster, M., Kundu, A., Ruiz, D., and Ryan, J. (2010). Writer Recognition of Arabic Text by Generative Local Features. *IEEE*, https://ieeexplore.ieee.org/document/5634495/\.
4. Gantz, D., Miller, J., and Walch, M. (2005). Multi-Language Handwriting Derived Biometric Identification. *Proceedings 2005 Symposium on Document Image Understanding Technology*, pp. 197-210.

Scores, Handwriting, Pairwise Comparisons