Pairwise Scores for Designing Handwritten Document Comparisons

Danica Ommen, Cami Fuglsby, Christopher Saunders, Michael Caligiuri, Linton Mohammed, JoAnn Buscaglia

Introduction

PCAST and OSAC have identified handwriting as one of the areas that would benefit from research into the scientific validity of the comparisons performed by forensic document examiners.

Objectives

- Design a study to validate conclusions expressed by forensic document examiners.
- Quickly and effectively pair handwritten documents with various levels of visual correspondence.
- Develop an automated *score* indicating the level of dissimilarity between two documents.
- Explore open-source and proprietary methods of producing scores.
- Compare the resulting document pairs between different methods.

Acknowledgements

- This research was supported by the National Institute of Justice, U.S. Department of Justice under Award No. 2017-DN-BX-0148. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the National Institute of Justice or the U.S. Government.
- This work was partially funded by CSAFE through Cooperative Agreement # 70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.
- Thank you to Susan Vanderplas for help with OpenCV installation, and to Sciometrics, LLC for the FLASH ID system.

Contact Information

Danica Ommen Assist. Professor
CSAFE, Iowa State University
Email dmommen@iastate.edu

Data

The *NIJ dataset* was collected on lined white paper using a ballpoint ink pen and then scanned.

- Collected handwriting samples from 33 different individuals
- 2 different styles for each individual: cursive and print
- 6 different phrases from the *London Letter* written in each style
- 5 replicates of each phrase

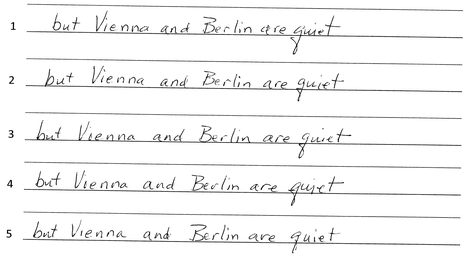

Figure 1: Print handwriting sample

SIFT Methods

SIFT (Scale-Invariant Feature Transform) is an algorithm designed to identify key points in an image regardless of the orientation and size [1]. For this project, SIFT was implemented using OpenCV in the C++ language. Overview of using SIFT to produce a score between two samples of handwriting is provided below:

Create the codebook:

- Start with a *training set* of images
- Detect the important *features* in each image
- Use the SIFT algorithm to compute the matrix of *descriptors* for all the features
- Use the k-Means algorithm to *cluster* the descriptors
- Use the BOW algorithm to create the *codebook* from the clusters

Compute the scores:

- Start with a *test set* of images
- Detect the important *features* in each image
- Use the SIFT algorithm to compute the matrix of *descriptors* for all the features
- Use the codebook and the k-NN algorithm to find the *encoded vector* for each image
- Compute the *Euclidean distance* between all of the encoded vectors

FLASH ID Methods

FLASH ID is a software developed by Sciometrics, LLC to quantify features of handwriting [2]. For this project, FLASH ID was used to produce a score vector for each writing sample, which were used to compute a single score for a pair of writing samples. This nonstandard score was developed for this project.

Using FLASH ID:

- Start with a *base set* of images
- Assign some writers from the FBI 100 to the *reference set*
- Use FLASH ID to get a *score vector* [2] (one score for each reference set writer) for each writing sample in the NIJ dataset

Using R:

- Compute the *Euclidean distance* between all pairs of score vectors
- Separate the distance scores into *within* or *between* writer
- Separate distance scores by phrase
- *Rank* the scores (higher score = two writing samples are more dissimilar)

Results

Scores are only comparable within system, writing style, and phrase. Example pairs from SIFT (Figures 2-5) and FLASH ID (Figures 6-9) scores are given below:

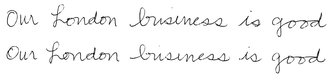

Figure 2: Same writer, low score=0.056

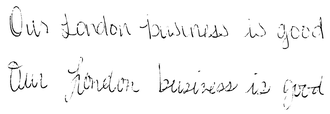

Figure 3: Different writer, low score=0.049

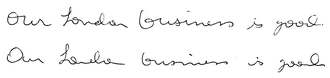

Figure 4: Same writer, high score=0.124

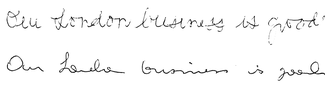

Figure 5: Different writer, high score=0.178

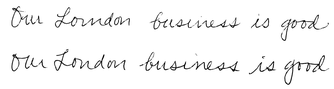

Figure 6: Same writer, low score=1.535

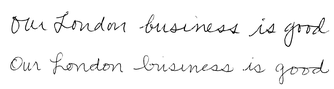

Figure 7: Different writer, low score=1.619

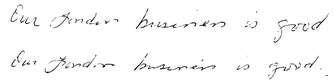

Figure 8: Same writer, high score=8.57

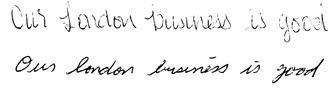

Figure 9: Different writer, high score=12.67

An example of a pair from different writers and a comparison of the SIFT and FLASH ID scores (star position on number line) and ranks (number on star) for that pair are given below:

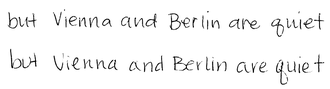

Figure 10: Pair with high visual similarity

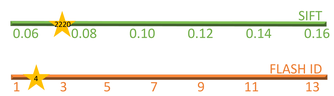

Figure 11: Scores & rank for the pair

Conclusion

- These methods can be extended to perform forensic identification of source
- FLASH ID scores seem to give better rankings in this particular case
- But, we know the SIFT methods have room to improve:
- Considering replacing SIFT with SURF or using a combination of SIFT & SURF
- Looking at different scoring metrics to replace the Euclidean distance

References

[1] D. G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2.

[2] Miller, et al. (2017), A Set of Handwriting Features for Use in Automated Writer Identification, *J Forensic Sci*, 62: 722-734. doi:10.1111/1556-4029.13345