Individuality of Handwriting: A Validation Study.

Conference Paper · January 2001

DOI: 10.1109/ICDAR.2001.953764 · Source: DBLP

4 authors, including:

Sargur N. Srihari
University at Buffalo, The State University of N…

Sung-Hyuk Cha
Pace University

Individuality of Handwriting: A Validation Study

Sargur N. Srihari Sung-Hyuk Cha Hina Arora Sangjik Lee
Center of Excellence for Document Analysis and Recognition (CEDAR)
University at Buffalo, State University of New York

Abstract

Motivated by several rulings in United States courts concerning expert testimony in general and handwriting testimony in particular, we undertook a study to objectively validate the hypothesis that handwriting is individualistic. Handwriting samples of 1,500 individuals, representative of the US population with respect to gender, age, ethnic groups, etc., were obtained. Analyzing differences in handwriting was done by using computer algorithms for extracting features from scanned images of handwriting. Attributes characteristic of the handwriting were obtained, e.g., line separation, slant, character shapes, etc. These attributes, which are a subset of attributes used by expert document examiners, were used to quantitatively establish individuality by using machine learning approaches. Using global attributes of handwriting and very few characters in the writing, the ability to determine the writer with a high degree of confidence was established. The work is a step towards providing scientific support for admitting handwriting evidence in court. The mathematical approach and the resulting software also have the promise of aiding the expert document examiner.

1 Introduction

The analysis of handwritten documents from the viewpoint of determining the writer has great bearing on the criminal justice system. Numerous cases over the years have dealt with evidence provided by handwritten documents such as wills and ransom notes. Handwriting has long been considered individualistic, as evidenced by the importance of signatures in documents. However, the individuality of writing in handwritten notes and documents has not been established with scientific rigor, and therefore its admissability as forensic evidence can be questioned.

Writer individuality rests on the hypothesis that each individual has consistent handwriting which is distinct from the handwriting of another individual. However, this hypothesis has not been subjected to rigorous scrutiny with the accompanying experimentation, testing, and peer review. Our objective was to make a contribution towards this scientific validation.

The task involved setting up a methodology for validating the hypothesis that everybody writes differently. The study is built upon recent advances in developing machine learning algorithms for recognizing handwriting from scanned paper documents; software for recognizing handwritten documents has many applications, such as sorting mail with handwritten addresses. The task of handwriting recognition focuses on interpreting the message conveyed—such as determining the town in a postal address—which is done by averaging out the variation in the handwriting of different individuals. On the other hand, the task of establishing individuality focuses on determining those very differences. What the two tasks have in common is that they both involve processing images of handwriting and extracting features.

There are two variabilities of concern while comparing handwriting: the variability of the handwriting of the same individual and the variability of the handwriting from one idividual to another. These two variabilities are seen when several individuals are asked to write the same word many times (Fig. 1). Intuitively, the *within-writer variation* (the variation within a person's handwriting samples) is less than the *between-writer variation* (the variation between the handwriting samples of two different people). The goal of this study was to establish this intuitive observation in an objective manner.

Figure 1. Variability in handwriting: Samples provided by eight writers (boxed), each of whom wrote the same word thrice.

The study consisted of three phases: *data collection, feature extraction*, and *individuality validation*. In the data collection phase, representative samples of handwriting were collected. The feature extraction phase was to obtain handwriting attributes that would enable the writing style of one writer to be discriminated from the writing style of another writer. The validation phase was to associate a statistical confidence level with a measure of individuality.

The study pertains to natural handwriting and not to forgery or disguised handwriting. Examination of handwritten documents for forensic analysis is different from recognition of content, e.g., reading a postal address, or in attempting to assess personality (also known as graphology).

2 Data Collection

A database of handwriting samples (written in English) of over one thousand individuals was created [1]. The sample population was made representative of the US population by stratification along different genders, age groups and ethnicities, and proportional allocation based on the US census data. This has been achieved by basing our sample distribution on the US census data (1996 Projections) using a sampling technique [2]. There are 510 female and 490 male population distributions and 36% of white ethnicity group, and so forth. The database comprises handwriting samples of $1,000$ distinct writers.

We asked each participant to copy out the *CEDAR letter* [1] three times in his or her most natural handwriting. In this data collection, provided uniform writing materials are used: plain unruled sheets and a medium black *Bic* round stic pen. Each of the collected handwritten documents are digitally scanned (300 dpi resolution, 0% contrast and 13% brightness) and stored as 8 bit (256 grey level) images. After scanning and digitization, line images are acquired from each document image by segmentation. Word images are segmented from the line image and each character is segmented from the word images.

3 Handwriting Attributes (Features)

Features are quantitative measurements that can be obtained from a handwriting sample in order to obtain a meaningful characterization of the writing style.

These measurements can be obtained from the entire document or from each paragraph, word, or even a single character. In pattern classification terminology, *measurements*, or attributes, are called *features*. In order to quantify the process of matching documents, each sample is mapped onto a set of features that correspond to it, called a *feature vector*. For example, if measurements, f_1, f_2, ..., f_d, are obtained from a sample, then these measurements form a column vector $[f_1, f_2, ..., f_d]^t$, which is a *data point* in d-dimensional space [3]; note that superscript t indicates vector transposition.

Computational features are those that can be determined algorithmically, e.g., by software operating on a scanned image of the handwriting. Computational features remove subjectivity from the process of feature extraction. While it could be argued that all document examiner features could eventually be computational features—when the correct algorithms have been defined—the fact remains that most of the document examiner features are not yet computable.

While some document examiner features like legibility and writing quality may be too subjective to be implemented, several of the other features are computable based on existing techniques for handwriting recognition [4, 5].

The micro-features consist of 512 binary (0 or 1 value) features corresponding to gradient (192 bits), structural (192 bits), and concavity (128 bits) features. The first gradient feature generator computes the gradient of the image by convolving it with a 3×3 Sobel operator [6, 7]. The direction of the gradient at every edge is quantized to 12 directions. The structural feature generator takes the gradient map and looks in a neighborhood for certain combinations of gradient values. These combinations are used to compute 8 distinct features which represent lines (strokes) and corners in the image. The concavity feature generator uses an eight point star operator to find coarse concavities in four directions, holes, and large scale strokes. The image feature maps are normalized with a 4×4 grid and a feature vector is generated.

4 Validation

We can approach the task of validating individuality in handwriting by taking one of two approaches: *verification* or *identification* model. In the former model, given two samples of handwriting of the letter from the sample population, we would like to tell whether they were written by the same person or by two different people. In the later one, the task is to assign a document of an unknown writer to one of known writers.

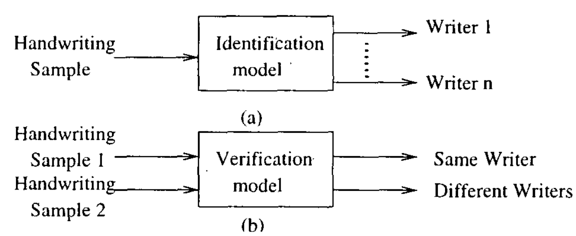

Figure 2. Two models for establishing the individuality of handwriting: (a) the identification model, and (b) the verification model.

Verification Model: The computer program is always presented with two handwritten samples and the question posed is: Are the two documents compatible or not? Given two documents, a multivariate statistical analysis of these computational features allows us to determine whether they were written by the same or different authors. Tests based on the 11 document level features resulted in accuracy levels of up to 96%. The detailed description and experimental results can be found in [8, 9]. Classification was done using an artificial neural network trained using backward error propagation.

Identification Model: The computer program is presented with the fundamental question of author identification which is defined as one of assigning an author to a test handwritten document of unknown authorship based on labeled prototype documents. There are m writing exemplars of each of n people. Given a writing exemplar, x, of an unknown writer, the task is to determine whether x was written by any of the n writers and if so, identify the writer. A *writer identification* system is designed by measuring spatial visual features from digitally scanned handwritten document images, quantifying their presence, and conducting statistical and geometrical analysis of features.

We used the *nearest neighbor rule* [3] to classify the unknown input vector by finding the most similar template (prototype document) in the prototype set. Using character level features, the performance was 98% for 2 writers when all pairs of writers was considered and eight characters of only one word of the document was considered. The performance deteriorates to 72% when $1,000$ writers are simultaneously considered. What this means is that when there are $1,000$ possible writers the author of the test document is correctly assigned with a 72% probability.

We designed our classifier to see if we can classify a test document based on a reference set containing documents from n writers. In the experiment shown in the graph in Fig. 4, the reference set consists of all the documents written by each of n writers, except for a test document that is left out from the set. So the reference set has (3*n)-1 documents in it. The test document is assigned the class of the document nearest to it in the reference set. The distance between any pairs of documents, with feature vectors $(a_1, a_2, ..., a_d)$ and $(b_1, b_2, ..., b_d)$, is calculated as $\sqrt(\sum_{i=1}^{d}(a_i - b_i)^2)$, where d is the number of attributes. This process is repeated using each of the 3*n documents belonging to the n writers. The estimated accuracy for a given n writers is then (number correctly classified)/(3n). To generalize to all writers, a random representative set of 1000 writers was first chosen. Then for each n, 1000 random subsets of the random representative set were selected. The average of the 1000 accuracies determined from the leave-one-out method was plotted on the graph for each value of n. The graph shows results pertaining to the document (CEDAR letter), paragraph (destination address block in CEDAR letter), word ("referred" in CEDAR letter), and character (segmented letters of word "referred" in CEDAR letter) level features.(Results are shown for the two sets of character features discussed earlier - geometric and GSC. The GSC features result in better performance.) As can be seen, performance decreases with increase in number of writers, and decrease in document content.

4.1 Comparison of the Two Models

Validation of individuality was done using two different approaches, both based on classificatory models: (i) the approach of identifying the writer from a set of possible writers, and (ii) the approach of determining whether two documents were written by the same writer. Writer identification accuracy was close to 98% for two writers. In the verification approach, the features were mapped onto the feature distance domain, and the individuality problem was tackled as a 2-class problem of classifying within- and between-author distances. Verification accuracy was about 95%.

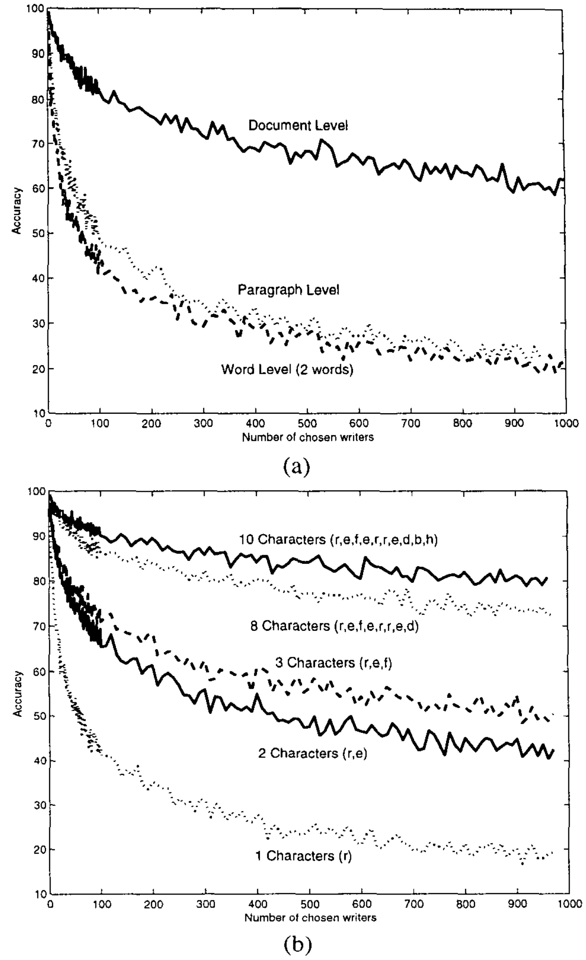

Figure 3. Variation of degree of confidence in writer identification with number of writers on (a) different level features (b) various number of characters.

The verification model has a slightly lower accuracy, as can be expected due to its mapping into a space of distances before performing classification. It was seen that performance deteriorated with a decrease in document content for both models. The verification model cannot be parameterized corresponding to the number of writers considered, unlike the identification model. However, repeated application of the verification model, considering one writer at a time, will yield a method of identification. Such a use of the verification model will have a reject option built in.

The principal advantage of the verification model over the identification model is its statistical generality. The identification model is easy to set up for establishing individuality as long as a substantial number of instances for every class is observable. When the number of classes is too large, e.g., the US population, most parametric or non-parametric multiple classification techniques are of no use to validate the individuality of classes, and the problem is seemingly insurmountable.

In the verification model, one need not observe all classes, yet it allows for inferential classification of patterns. It is a method for measuring the reliability classification about the entire set of classes based on samples obtained from a small sample of classes.

5 Summary and Conclusion

A study was conducted for the purpose of establishing the individuality of handwriting. The work was motivated by US high court rulings that require expert testimony be backed by scientific methodology. Since handwriting had not been subjected to such a study, we decided to undertake this endeavor.

A database was built representing the handwriting of 1500 individuals from the general US population. The sample population was made representative of the US population by stratification and proportional allocation. The population was stratified across different genders, age groups and ethnicities. Each individual provided three handwritten samples, produced by copying-out a source document which was designed to capture many attributes of the English language: document structure; positional variations of alphabets, numerals, and punctuation; and interesting alphabet and numeral combinations. Computer software was used to extract features from digitally scanned images of handwriting. Features were extracted at a global level of the document, from the entire document, from a paragraph of the document, and from a word of the document. Finer features were extracted at the character level from each sample.

Validation of individuality was done using a machine-learning approach where some samples are used to learn writer characteristics, and other samples are used to test the learnt models. Based on a few macro-features that capture global attributes from a handwritten document and micro-features at the character level from a few characters, we were able to establish with a 98% confidence that the writer can be identified. Taking an approach that the results are statistically inferable over the entire population of the US, we were able to validate the individuality hypothesis with a 95% confidence. By considering finer features, we should be able to make this conclusion with a near-100% confidence.

Our work has employed handwriting features similar to, but not exactly the same as, those used by document analysts in the field. However, the objective analysis that was done should provide the basis for the conclusion of individuality when the human analyst is measuring the finer features by hand.

There are many important extensions of the work that could be done. Some of these are to study the handwriting of similarly trained individuals, to study temporal variations of handwriting over periods of time, etc.

Acknowledgments

This work was funded by National Institute of Justice (NIJ) grant 1999-IJ-CX-K010. We are grateful to Dr. Richard Rau who introduced us to the domain of forensic science and provided guidance. We would like to thank Eugenia Smith for assistance in collecting handwriting samples and Kristen Pfaff for editorial assistance.

References

[1] S.-H. Cha and S. N. Srihari, "Handwritten document image database construction and retrieval system," in *Proceedings of SPIE, Document Recognition and Retrieval*, vol. 4307, pp. 13–21, January 2001.

[2] S. L. Lohr, *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA, 1999.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, Inc., 2nd ed., 2000.

[4] S. N. Srihari, "Feature extraction for locating address blocks on mail pieces," in *From Pixels to Features* (J. C. Simon, ed.), pp. 22–27, August 1988.

[5] S. N. Srihari, "Recognition of handwritten and machine-printed text for postal address interpretation," *Pattern Recognition Letters*, vol. 14, no. 4, pp. 291–303, 1993.

[6] J. T. Favata, G. Srikantan, and S. N. Srihari, "Hand-printed character/digit recognition using a multiple feature/resolution philosophy," in *IWFHR-IV*, pp. 57–66, December 1994.

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, 3rd ed., 1992.

[8] S.-H. Cha and S. N. Srihari, "Writer identification: Statistical analysis and dichotomizer," in *Proceedings of SS&SPR 2000 LNCS - Advances in Pattern Recognition*, vol. 1876, pp. 123–132, Springer-Verlag, September 2000.

[9] S.-H. Cha and S. N. Srihari, "Multiple feature integration for writer verification," in *Proceedings of 7th IWFHR 2000*, pp. 333–342, September 2000.