Document Title:	**Computational Methods for Handwritten Questioned Document Examination**
Author:	**Sargur N. Srihari, Ph.D.**
Document No.:	**232745**
Date Received:	**December 2010**
Award Number:	**2004-IJ-CX-K050**

Computational Methods for Handwritten Questioned Document Examination [1]

FINAL REPORT

Award Number: 2004-IJ-CX-K050

SUBMITTED TO:
U.S. Department of Justice
Office of Justice Programs
National Institute of Justice
810 Seventh Street N.W.
Washington, DC 20531

AWARDEE:
Research Foundation of the State University of New York

Author:
Sargur N. Srihari
520 Lee Entrance, Suite 202
University at Buffalo, State University of New York
Buffalo, New York 14260
Tel: 716-645-6164 ext 113
Email: srihari@cedar.buffalo.edu

August 12, 2010

[1] Previous Title: QUANTITATIVE ASSESSMENT OF THE DISCRIMINATIVE POWER OF HANDWRITING

Abstract

Computational approaches to the handwriting facet of questioned document (QD) examination were developed with a view towards providing a scientific basis for handwriting evidence, formalizing human expert-based approaches and validating existing methodology. Extended writing samples such as a paragraph of writing as well as signatures were considered. The task of verification, which is to determine whether two writing samples compared side-by-side, originate from the same person, was the principal problem addressed. A statistical model for writer verification was developed; it allows computing the likelihood ratio based on a variety of different feature types. Several new algorithms were developed for extended handwriting comparison including capturing uniqueness of the writing and discretizing results into a nine-point opinion scale (ranging from identified as same to identified as different). A signature verification approach was developed to deal with a small set of learning samples. The tools developed were shown to have value in: (i) establishing that the handwriting of twins can be discriminated, (ii) providing an upper bound on the error rate of expert human examiners, since system error rate is lower than that of lay persons but higher than that of experts, (iii) automatically extracting frequencies of various letter formations so that they can be used in forensic testimony, e.g., to state a probability of random correspondence, and (iv) determining authorship of a historical document. The research led to a United States patent, two journal papers, twelve conference papers and two book chapters. Several presentations were made to the QD community including user workshops. The work was cited in a Daubert and a Frye hearing. The software developed was made available to the QD community by distribution media (CDs and downloads from the internet).

Contents

Chapter 1

Executive Summary

Computational approaches to the handwriting facet of questioned document (QD) examination were developed with a view towards providing a scientific basis for using handwriting as evidence, formalizing human expert-based approaches and validating existing methodology. Methods were developed for both extended writing samples, e.g., a paragraph of writing, and signatures. The algorithms and software were incorporated into a previously developed platform. The task of verification, which is to determine whether two writing sample sets originate from the same person, was the principal problem addressed. Several new algorithms and tools were developed for the comparison of extended handwriting comparison. The signature verification method developed allows dealing with a small set of learning samples.

The main technical accomplishments were in the following areas:

1. Developed a statistical model for writer verification. It involves computing the likelihood ratio of prosecution and defense hypotheses using similarity measures over a variety of features.

2. Extracting document examiner features from handwriting. This goal was to extract the type of handwriting features used by human QD examiners.

3. Capturing uniqueness of writing. Since the task of writer verification is assisted by detecting unexpected writing styles, methods and interfaces were developed to perform this function.

4. Expressing the result of comparison on a discrete opinion scale. Several guidelines have been developed for expressing the results of handwriting comparison including the use of a nine-point scale by SWGDOC. A method for quantizing the log-likelihood ratio into the scale were developed.

5. Signature verification using Bayesian methods. The presence of only a handful of genuine signatures was overcome by using a fully Bayesian approach wherein the prior parameter distributions were learnt from a larger training set which are then modified by the available samples.

The tools developed were shown to have value in validating handwriting examination procedures. They included: (i) establishing that the handwriting of twins can be discriminated, (ii) providing an upper bound on the error rate of expert human examiners, since system error rate is lower than that of lay persons but higher than that of experts, (iii) automatically extract frequencies of various letter formations so that they can be used in forensic testimony, e.g., to state a probability of random correspondence, and (iv) determining authorship of a historical document.

Research results were disseminated in three ways: (i) *software* known as CEDAR-FOX distributed to the QD community via media (CDs and internet downloads), (ii) *scholarly works* including a U.S. patent, two journal papers, twelve full-length refereed conference papers and two book chapters, and (iii) *presentations* made at forensic, QD and pattern recognition conferences including software user workshops. The work was also cited in a Daubert and a Frye hearing both of which resulted in handwriting testimony being admitted.

Project Dates

This project (*Computational Methods for Handwritten Questioned Document Examination*) was awarded for the two-year period 7/1/2007 to 6/30/2009 with a no-cost extensions until June 30, 2010. The funds were added to a previously existing project with dates 7/1/2004 to 6/30/2007(whose title was *Quantitative Assessment of the Discriminative Power of Handwriting*). This report gives a synopsis of the effort during 2004-2010 and gives a detailed report for the project period: 2007-2010.

Chapter 2

Final Report Narrative

2.1 Introduction

Questioned document (QD) examination involves the comparison and analysis of documents, printing and writing instruments in order to identify or eliminate persons as the source. Questions about documents arise in research, business and finance, as well as civil and criminal trials. A significant aspect of QD examination is concerned with handwriting.

Handwriting is the product of an individuals neuromuscular function applied to an instrument or implement to create ordered markings, usually in a particular language, on a receptive medium or surface, commonly paper. Accordingly, the handwriting reflects the genetic, biological, and structural aspects of the writer, as well as the writers environment, experience and training.

The examination of handwritten items typically involves the comparison of a questioned item submitted for examination along with a known item of established origin associated with the matter under investigation. Comparisons are based on the high likelihood that no two persons write the same way, while considering the fact that the writing of each person has its own variabilities. "Thus, an analysis of handwriting must compare interpersonal variability – some characterization of how handwriting features vary across a population of possible writers – with intrapersonal variability – how much an individuals handwriting can vary from sample to sample. Determining that two samples were written by the same person depends on showing that their degree of variability, by some measure, is more consistent with intra-personal variability than with inter-personal variability" [1].

While handwriting has long been accepted by the courts as evidence [2][1], the need to establish a scientific basis came about under rulings such as Daubert [3] and its successors

[1]The modern development of the law of evidence relating to testimony of scientific or technical experts was guided by a decision of the Court of Appeals for the D. C. Circuit in a criminal case in 1923, *Frye v. United States* (Frye). The court in Frye set forth the standard for admissibility of expert testimony: "Just when a scientific principle or discovery crosses the line between the experimental and demonstrable states is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs".

[4] [2]. Judicial concerns over the reliability of handwriting analysis can be summarized as follows [5]: (i) fundamental premises there has been little empirical testing and few published studies relating to the basic theories on which handwriting analysis is based; (ii) reliability of methodology used by QD examiners the data on handwriting analysis is "sparse, inconclusive, and highly disputed, particularly as regards the relative abilities of [questioned] document examiners and lay examiners" [6] and [prior studies] did not conclusively establish that [QD examiners] can reliably do what they say they can do [7]; (iii) error rates there is little known about the error rates of QD examiners, and the error rates for real world conditions are higher [8]; (iv) peer review and publications peer review has been limited to the QD examiner community and those within the field of QD examination have failed to engage in critical study of the basic principles and methods of handwriting analysis; published articles on handwriting analysis are significantly different from scholarly articles in such fields as medicine or physics in their lack of critical scholarship [6]. There has been no peer review by a competitive, unbiased community of practitioners and academics [9]; and (v) general acceptance– the acceptance of the methodology should come from disinterested experts outside of the QD examiner community.

In view of the judicial concerns there are several issues in handwriting examination that need study and research. Principal among these are: whether the handwriting of every individual is distinct (the *individuality* problem), whether the procedures used by QD examiners are repeatable (the *validation* problem) and whether the the inevitable uncertainty in individualization/exclusion can be quantified (the *uncertainty* problem).

Computational methods of QD examination offer the promise of helping resolve each of these problems. They can be used to perform statistical tests on representative populations to address the individuality problem. By formalizing human procedures they can be used to validate existing methods as being repeatable. They can also be used to compute various probabilities of interest in quantifying uncertainty in the opinion. The development of such methods also could not only help resolve the legal issues of handwriting evidence, but could also be eventually useful as tools for the QD examiner.

[2] In 1993, the United States Supreme Court decided in *Daubert v. Merrell Dow Pharm, Inc.* (Daubert), a civil case, that the provisions of Federal Rule of Evidence 702, rather than the common law rule developed in Frye, governed admissibility of expert testimony in federal courts. In reviewing the admissibility of scientific testimony, the Court observed that evidentiary reliability will be based upon scientific validity and directed trial judges to ensure the reliability of all scientific testimony and evidence, by focusing on the experts principles and methodology. The Court listed five (non-exclusive) factors to be given consideration by the trial court in making a ruling on the admissibility of expert testimony: (1) whether a theory or technique can be (and has been) tested; (2) whether the theory or technique has been subjected to peer review and publications; (3) the known or potential rate of error of a particular scientific technique; (4) the existence and maintenance of standards controlling the techniques operation; and (5) a scientific techniques degree of acceptance within a relevant scientific community. The subsequent decision of the United States Supreme Court in *Kumho Tire Co., v. Carmichael* brought technical, as well as scientific, expert testimony within the ambit of Rule 702.

2.1.1 Statement of the problem

The goal of this project was to develop computational approaches for handwriting examination. Algorithms, statistical models and software were to be developed for two forms of handwriting: extended writing (e.g., a paragraph or page of writing), and signatures on paper documents. The principal problem to be addressed was the task of verification, which is to determine whether two writing sample sets compared side-by-side, originate from the same person. New algorithms were to be developed for various tasks in the process including: text segmentation, feature extraction, capturing uniqueness of writing and expressing the opinion of examination, e.g., on a scale ranging from identified as same to identified as different. The signature verification problem was to be explored for realistic situations, e.g., small learning set, absence of forgery samples. The resulting methods would be useful to resolve some of the outstanding issues, viz., individuality, validation and uncertainty quantification. The software resulting from the research was to be in a form that could be used independently by the QD examiner community.

2.1.2 Literature citation and review

Extended Handwriting

A survey of various computer methods of handwriting examination including systems such as FISH is given in [10]. QD examiners employ characteristics that are termed the seven S's, viz., size, slant, spacing, shading, system, speed, strokes [11]. Although statistical characterization of such characteristics was mentioned over a hundred years ago [12] and systematic study of frequencies of characteristics was begun nearly thirty years ago, e.g., [13] and colleagues who studied 200 individuals writing the letter pair *th*, such efforts do not appear to have been continued due to the enormity of such an undertaking.

Starting with our research on individuality of handwriting using computational tools [14, 15] we developed the CEDAR-FOX system for writer verification and identification. A statistical model for handwriting was incorporated into the system. Many features are used including features commonly used by QD examiners and computational ones developed for tasks such as OCR [16, 17]. The system used 13 macro features (which include pictorial features such as slant and spacing) and a variable number of micro features (which are at the letter and word level) for the comparison. It is first of all necessary to extract relevant characters and character combinations from a large set of samples. This is made possible by having a transcript mapping algorithm [18] whose output of image snippets can populate a database of images needing to be analyzed. Features analogous to those used by QD examiners was suggested in [19] which formed a starting point for including such features in our work.

Signatures

The most common task in QD analysis is that of authenticating signatures [12, 20, 21, 22, 23]. The problem most frequently brought to a QD examiner is the question relating to the authenticity of a signature: *Does this questioned signature (Q) match the known, true* *signatures (K) of this subject?* [24] A QD examiner uses years of training in examining signatures in making a decision in case work. Another interesting and related question that is associated with the same problem is whether the questioned signature is disguised or forged.

There is scattered literature on automatic methods of signature verification [25, 26, 27, 28, 29]. Automatic methods for independently identifying disguised and forged signatures, as well as handling multiple types of signature have not been researched thus far. Since both of the above are very common in QD analysis, these along with automatic extraction of signatures are a focus of this research.

2.1.3 Rationale for the research

The goal of this research is to develop automated methods for the comparison of handwriting specimens. The methods to be developed have three purposes:

1. Formalizing the intuitive approach of human experts by means of precisely stated algorithms
2. Use the developed methods to extract characteristics from large data sets to provide scientific justification to methods used by examiners
3. Make the tools available to the QD community for performing comparisons more efficiently

2.2 Methods

2.2.1 Data Sets

The handwriting data sets used in this research were as follows:

1. **Extended Handwriting.** Two data sets were used:
   (a) The *CEDAR* data set consists of writing samples of over 1,500 individuals representing the United States population. Each individual copied a document known as the CEDAR-letter three times each. The CEDAR letter contains all possible letters in the English language, including many common letter pairs such as "th", "an", "he" and "nd". Each document is scanned at a resolution of 300 pixels per inch. Individuals who wrote the samples are stratified by gender and age and selected to be representative of the United States population[3].
   (b) *Twins' handwriting* collected by the U.S. Secret Service. The same text as in the CEDAR letter was used by the US Secret Service to have a set of 206 pairs of twins copy them[4].

[3] A detailed description of the CEDAR data set is given by us in [14] which can be downloaded from `http://www.cedar.buffalo.edu/~srihari/papers/JFS-2002.pdf`

[4] The demographics of the USSS data set is described further by us in [15] which can be downloaded from `http://www.cedar.buffalo.edu/~srihari/papers/JFS2008-color.pdf`.

2. **Signatures.** The following five data sets were used:
   (a) The *Georgia* data set for two individuals. Person1: 140 genuine, 35 disguised and 90 spurious and 459 forgeries, Person2: 170 genuine, 21 disguised and 650 forgeries,
   (b) The *NISDCC* data set of 12 persons, for each there are 5 genuine and 155 forgeries of each person
   (c) The *NISDCC* data set of 100 persons, for each person 20 genuine and 20 forgeries (some missing data)
   (d) The *CEDAR* data set of 55 persons each person has 24 genuines and 24 forgeries
   (e) The *NFI* data set of 2 persons with each person having 9 genuine and 11 forgeries

The following areas were explored:

1. document properties and user interfaces
2. writer verification model
3. writer adaptation
4. QD examiner features
5. detection of writer uniqueness
6. use of automated tools developed
   (a) twin's study
   (b) scripts other than contemporary English
7. signature verification

Each of these are described below.

2.2.2 Document Properties and User Interfaces

A number of user interfaces for QD examiner interaction as well for understanding the system's decision are provided. One of these is a display showing the system's understanding of the line structure in the image (see Figure 4.1). It displays the estimated number of lines, words and characters. The degree to which the writing differs from the Palmer style of handwriting is displayed. An indication is made as to whether the writing is cursive or printed shown on a continuous scale.

The letters and letter pairs of interest in the side-by-side comparison are extracted largely automatically but with user correction and input. This is made possible by the transcript mapping function of CEDAR-FOX [18]. The transcript mapping function allows the user to associate characters with correct recognition decisions before a comparison is made. Given the typed transcript of the handwritten document, this function maps words to images, as shown in Figure 4.2. This process provides word image snippets from which the letter combination snippets can be extracted.

2.2.3 Writer Verification Model

In the writer verification problem, the input is (i) the evidence: a scanned questioned document and (ii) the known: a set of documents (scanned) created by a single writer. The output is an opinion whether the writer of the questioned document is the same as the writer of the known set. This can be expressed in probabilistic terms. Let $\mathbf{x}$ be a feature representation, e.g., a vector or matrix, characterizing the evidence and $\mathbf{y}$ be the corresponding feature representation characterizing the known. We define two hypotheses, H and $\overline{H}$ referred to the prosecution and defense hypotheses. According to the prosecution hypothesis $\mathbf{x} = \mathbf{y}$ and according to the defense hypothesis $\mathbf{x} \neq \mathbf{y}$.

Define two probabilities $p(\mathbf{x}, \mathbf{y}|H)$ and $p(\mathbf{x}, \mathbf{y}|\overline{H})$ which are the joint probabilities of the the evidence and known given the prosecution and defense hypotheses respectively. Then the relative strength of the prosecution hypothesis over the defense hypothesis is given by the likelihood ratio [30]

$$\frac{p(\mathbf{x}, \mathbf{y}|H)}{p(\mathbf{x}, \mathbf{y}|\overline{H})}. \tag{2.1}$$

If we were able to pre-determine the probabilities of all possible values of $\mathbf{x}$ and $\mathbf{y}$ under each hypothesis, then we can provide the likelihood ratio. However, in general, the exact joint probabilities are impossible to compute. For instance if $\mathbf{x}$ and $\mathbf{y}$ are each extracted from a paragraph of 100 handwritten characters and each character has 26 possible values, then there are 26^{100} possible character strings. If each character is represented by a four-bit feature vector that can take 16 possible values, then the number of possible values of $\mathbf{x}$ is $16^{100} \times 26^{100}$ and the number of possible values of the pair $\mathbf{x}, \mathbf{y}$ is its square. Even if the text in the evidence and known are limited to a single known sequence of 100 characters (say the London letter), the number of possible values of the pair is still an impossibly large 16^{200}. The number of samples needed to estimate probabilities would be even larger.

One way to make the probability evaluations tractable is to introduce a kernel function $d(\mathbf{x}, \mathbf{y})$, which is a similarity or distance between evidence and known. The relationship between the evidence, known and similarity is shown as a probabilistic graphical model in Figure 4.5. It shows that while the evidence and known are independent, their similarity is dependent on both. The joint probability of interest $p(\mathbf{x}, \mathbf{y})$ in Eq. 2.1 for each of the prosecution and defense hypotheses is replaced, with appropriate conditioning by H and $\overline{H}$ by

$$p(\mathbf{x}, \mathbf{y}, d(\mathbf{x}, \mathbf{y})) = p(\mathbf{x})p(\mathbf{y})p(d(\mathbf{x}, \mathbf{y})|\mathbf{x}, \mathbf{y}) \tag{2.2}$$

Note that $p(\mathbf{x})$ and $p(\mathbf{y})$ represent the rarity of the evidence and known respectively, while $p(d(\mathbf{x}, \mathbf{y})|\mathbf{x}, \mathbf{y})$ represents the probability of the similarity.

Since the the term $p(\mathbf{x})p(\mathbf{y})$ will be the same under both the prosecution and defense hypotheses we can rewrite the likelihood ratio in Eq. 2.1 as

$$\frac{p(d(\mathbf{x}, \mathbf{y})|H)}{p(d(\mathbf{x}, \mathbf{y})|\overline{H})}. \tag{2.3}$$

Since d is a scalar value then we need to evaluate a univariate distribution instead of a multivariate joint distribution. Although $p(d(\mathbf{x}, \mathbf{y}))$ still depends on all possible pairs of values, the number of samples needed would only be determined by the resolution needed to capture the variations of $d(\mathbf{x}, \mathbf{y})$. In some forensic domains methods to compute a distance between two samples exist, e.g., the score between two fingerprints or the difference between two speech patterns. Such an approach is used in computing likelihood ratios for fingerprints [31].

For a pair of handwritten documents it is not simple to calculate $d(\mathbf{x}, \mathbf{y})$, e.g., when the features do not lend themselves to a vector representation. We can regard $d(\mathbf{x}, \mathbf{y}) = f(d_1, d_2, ..)$ where f is a function of the differences $d_i = |x_i - y_i|$ among the features. If we choose d to be the Euclidean distance (L_2 norm) we have $f(d_1, d_2, ..) = (\sum_{i=1}^{n} d_i^2)^{1/2}$. If d is the Manhattan distance (L_1 norm) we have $f(d_1, d_2, ..) = \sum_{i=1}^{n} d_i$.

Rather than accumulate the differences into a single value, we can regard them as elements of a vector $\mathbf{d}(\mathbf{x}, \mathbf{y}) = [d_1, d_2, ..]$. This is useful because the differences provided by different features may not be commensurate, e.g., for characters we use a correlation similarity metric and for macro features we use simple differences. We note that this approach is not equivalent to one using a sum of differences, since knowing the distributions of d_1 and d_2 does not specify the distribution of $d = d_1 + d_2$. However since we are only interested in introducing an easily computed relationship between $\mathbf{x}$ and $\mathbf{y}$, as suggested in Eq. 2.2, introducing a set of differences should be more informative than a single distance.

Finally, we make the naiive Bayes assumption that differences along features are independent, we replace $p(d(\mathbf{x}, \mathbf{y}))$ in Eq. 2.3 by $\prod_{i=1}^{n} p_i(d_i)$ where p_i is the probability distribution associated with the ith difference. Since differences are positive-valued we can expect the distributions to have a Gamma distribution, which is a generalization of the Gaussian distribution. The distance between features of ensemble of pairs of same and different writer documents are modeled parametrically using Gamma/Gaussian density functions [15]. If p_i^s denotes the density function modeling the distance between *same* writer document pairs for the i^{th} feature and p_i^d denotes the density function modeling the distance between *different* writer pairs for the i^{th} feature, then the likelihood ratio between two documents with distances d_i between the features, is given by $LR = \prod_i \frac{p_i^s(d_i)}{p_i^d(d_i)}$ and the log-likelihood ratio (LLR) is given by

$$LLR = \sum_i log\ p_i^s(d_i) - log\ p_i^d(d_i) \tag{2.4}$$

The number of micro features used when comparing a pair of documents is variable since it depends on the content and recognition results on the documents. For example, if there were N occurrances of character "a" in the first document and M occurrences of character "a" in the second document, then that results in $N \times M$ comparisons resulting in $N \times M$ features for the character "a" alone. Similar, logic applies to other characters and bigrams such as "th", "he" etc.

A screen display showing the LLR value of a side-by-side document comparison is shown in Figure 4.3. The overall opinion based on comparison is shown in Figure 4.4– it includes a mapping from the raw LLR to a nine-point opinion scale.

2.2.4 Writer Adaptation

The design of the CEDAR-FOX system was based on handwriting samples from a general population as described in Section 2.2.1. This includes both the tools for automatic recognition, such as of characters and words, and statistical models of prosecution and defense hypotheses. There are several clear advantages to changing this design by adapting to the specific known writer. For one, automatic recognition will become more efficient thereby allowing fewer manual corrections to be made by the QD examiner. Another is that the within-writer distance distribution will capture only the particular writer's variations rather than the whole population thereby leading to higher verification accuracy.

We have made some significant progress on the work with adaptation in recognition [32]. This leads to increased performance wherever word recognition is used, including tasks such as writer identification/verification, transcript mapping, and plain recognition. The adaptation engine is quite flexible, allowing trained, adapted writer profiles to be saved and loaded from disk in addition to the built-in writer prototypes. While the adaptation code is written and implemented, still need to make an interface for saving/restoring the writer profiles (i.e., building a way to allow the user to access the code that already exists).

We extensively evaluated performance increases that can be expected with writer verification [33]. One variation on this theme was the alternate model of writer verification which dynamically learns/builds a same-different writer model based on a specific case. In the past, the same-same and same-different writer models were built over large populations and kept static. We instead created the same-different writer model by comparing directly the case in question to the large population of different writers and building the model at runtime. We demonstrated how this leads to significantly better performance in many cases. Both models have merit; CEDAR-FOX still defaults to the original model since this original approach is well documented and accepted. We still need to allow more flexibility for the operator to select the new method as well.

2.2.5 QD Examiner Features

Since the micro features used in the CEDAR-FOX system originate from algorithms developed for character recognition rather than QD examination it was felt necessary to develop newer algorithms for extracting such features. Starting with the work of [19] we developed and implemented such features [34]. They are based on an ability skeletonize such lexemes, using an algorithm such as [35]. The features for the lexeme “th” are shown in Figure 4.6. The use of the “th” feature within CEDAR-FOX is shown in Figure 4.7.

Based on such features the statistical distributions of QD examiner features can be computed. There are several frequency tables that need to be determined. These are the marginal frequencies of the individual characteristics and the conditional frequencies. The conditioning would be on each of the characteristics and their combinations. Conditional frequency tables of each of the four characteristics, conditioned on the other three characteristics (taken one at a time) from 531 handwriting samples are given in Figures 4.8 and 4.9. The probability of random correspondence for some examples of “th” are shown in Figure 4.10.

2.2.6 Uniqueness of Writing

An important step in writer verification is a character-by-character comparison of the documents to determine whether the characters present in all the documents are consistent. Since it is extremely time consuming to manually identify the characters individually, an automated approach is important. In prior work, we have focused on identifying well written prototype characters and used these for comparisons. However, such an approach falls short in one key respect-uniquely written characters are often excluded from comparison since they do not appear as well generated prototypes. This is especially unfortunate because such uniquely written characters are often the strongest evidence of writership. A motivating example is available in Figure 4.11.

There are several key ways to integrate such knowledge: i) relaxing the prototype restrictions, we can simply incorporate the less "well written" characters into the character-by-character score, ii) we can seek to identify uniquely written characters themselves, and then attempt to incorporate the additional value that these characters have as evidence of writership and incorporate them into the entire verification model as a separate score. We have taken two approaches to identifying characters as unique. One is an addition to the CEDAR-FOX interface of a way for a document examiner to identify characters as unique in a supervised fashion. The second is an automated approach through semi-supervised learning to identify uniquely written characters. Automatic identification the challenge of the presence of poorly segmented characters. In this setting, a character was incorrectly isolated from a word image and subsequently assigned an incorrect label as a character. This problem is, in fact, the underlying motivation to exclude characters straying too far from prototypical appearances (in effect tagging them as "poorly written"). Our approach to overcome this difficulty is to check for internal consistency among the unique characters. That is, if the unique characters are not similar to the prototypes, but have a sufficient similarity to one another, it implies there was no segmentation error. We have also developed a semi-supervised method that learns from characters labeled with sufficient certainty, modifying the prototype model, and re-evaluates the document in additional passes to identify additional characters missed in the first pass. The next challenge is the method by which we incorporate the unique characters into the writer verification framework. Our initial approach, with good results, was simply to include them in the character-by-character comparison. Another idea is to give some weight to the degree of uniqueness and give some weight to the degree of consistency of writing of the unique characters-both implying different ways of estimating the inherent value each character class has as being specific for the questioned or known writer. An example is present in Figure 4.12 where two different samples of the same writer contain similarly written 'I' productions that are somewhat unique. A selection of prototype 'I' characters is shown in Figure 4.13.

The two unique samples are incorrectly identified as prototype 'g' characters by the automatic identifier, and thus the previous method flagged them as inappropriate for comparison. However, their discriminality power is quite high and losing this information significantly compromises potential writer verification. The new method helps avoid this problem.

There is an interface for learning uniqueness (See Figure 4.14). For example, we have areas where the user can flag unique characters. We are still working on the best way to integrate this information into scores, but as it is they can at least force the non-prototypical characters to be used for writer verification scoring even now. One direction we have begun to explore with the uniqueness is the feature-space analysis of the writing, starting with expanding on the previous th work. In addition, we have significantly improved the transcript mapping algorithm. The performance goes up significantly with the fixes, although there are still some obvious improvements that could be made in the future.

2.2.7 Use of Developed Tools

The tools developed were used to further advance the scientific basis of QD examination.

Individuality of Handwriting

In our previous work on handwriting individuality [14] we used a handwriting sample representative of the U. S. population. Since that can be criticized as being too diverse a population, and thereby providing over-optimistic results, it was felt necessary to evaluate the methods with a cohort population such as twins. In a study of the handwriting of twins, error rates of an automatic verification system was determined for twin and non-twin populations [15]. Considering an input of a half-page of extended handwriting, it was determined that error rates with the handwriting of twins was higher (13%) compared to that of non-twins (4%). Within the twin population error rates with identical (mono-zygotic) twins was higher than that of fraternal twins. Since system error rate was lower than that of lay persons but higher than that of experts, the reported error rates are upper bounds on human forensic expert rates.

Scripts Other than Contemporary English

Another area of use of the methodologies developed is the analysis of handwriting in other scripts, such as historical documents and non-Latin script, with regards to writership. Since the training of our methods was done largely on contemporary writing in the United States, such a use establishes the robustness of the developed methods.

CEDAR-FOX was used to determine whether a historical document of unknown authorship was written by a well-known writer [36]. The questioned document was compared against two known handwriting samples of Herman Melville, a 19th century American author who has been hypothesized to be the writer of this document. The comparison led to a high confidence result that the questioned document was written by the same writer as the known documents. The same conclusion had also been reached by studying the linguistic content of the documents. Such methodology can be applied to many such questioned documents in historical writing, both in literary and legal fields. This case was published as an interesting application of writer verification and was additionally interesting in that CEDAR-FOX's utility was extended to a new application of historical document comparison.

The methods were also modified to see how well handwriting in Arabic could be verified. In a first ever such effort, and using primarily macro features promising results were obtained [37].

2.2.8 Signature verification

Signature verification is the most common task in forensic document analysis. It is also an ideal domain to anchor this study since the degree of uncertainty with signatures is likely to be higher than in say latent prints. This is because of two reasons: (i) greater variability within genuine signatures, and (ii) a typically small set of signatures to learn from. The goal of signature verification is to make a decision whether a questioned signature belongs to a set of known signatures of an individual or not. The principal uncertainty in this task is comparison uncertainty.

In the signature problem the modeling necessarily has to pertain to a given individual. There are a set of genuine signatures and a set of skilled forgeries. There is considerable variation in the known (genuine). This can be modeled either as a generative model in feature space or as a discriminative model in similarity space. Modeling signatures in feature space becomes infeasible when the genuine set is small to capture all variations for the given individual. Thus the discriminative approach is more promising.

A complicating factor is that in a typical forgery case not only a very limited number of known signatures may be available, with as few as four or five knowns [24], and there may be no forgery samples available for that individual at all. See example in Figure 4.15 for a typical case. Thus the problem seems to be not very promising for a statistical approach.

When samples are few, the Bayesian approach comes to the rescue. A simple example of the power of the Bayesian approach over the frequentist approach is that of estimating the probability of *heads* in coin tossing. If only three tosses are available and all three come up heads, then the frequentist approach assigns a probability of one to heads– which is unacceptable since in subsequent tosses it might turn out that we do have a fair coin. On the other hand the Bayesian approach assigns a prior probability distribution over the probability of heads, with a high probability for 1/2 and decreasing probabilities for values approaching 0 and 1. When there are tosses made this distribution is appropriately modified into a posterior distribution. Thus if three heads are observed in succession then the probability of heads lying between 1/2 and 1 are given higher probabilities than those less than 1/2.

In the Bayesian approach for signature verification one can assume reasonable prior distributions for "same" and "different" class similarity distributions. These distributions would be modified by the samples at hand. The hyper-parameters for the prior distributions would be determined from samples from other signature verification problems contained in a database. In a sense the prior distributions represent prior knowledge of signature verification by the questioned document examiner and the posterior distributions represent knowledge as modified by the case at hand.

Thus the discriminative model based on the Bayesian approach is a natural one for signature verification. Our approach is a fully Bayesian one whose goal is to overcome the limitation of having too few genuine samples. The algorithm has three steps: (i) assume reasonable forms for the prior distributions,e.g., Gaussian or gamma, for the two classes, (ii) prior distributions of parameters are learnt from a population of signatures, (ii) determine posterior distributions of parameters using genuine samples of a particular person, and (iii) Determine probabilities of the query from both genuine and forgery classes and the Log Likelihood Ratio (LLR) of the query. Rather than give a hard decision, this method provides a probabilistic measure (LLR) of the decision. The distributions involved are illustrated in Figure 4.16. The results are better than a previous one where distributions were directly modeled without a Bayesian prior [38].

2.3 Conclusions

Computational methods were developed for the handwriting facet of QD examination. These have covered the areas of statistical models for writer verification, identification of the uniqueness of a writer, and extraction of QD features.

Statistical models based on the general population were shown to be value. Obtaining models for specific individuals was shown to further improve performance. The feasibility of comparing handwriting to a standard style such as "Palmer" was demonstrated.

Capturing unexpected writing styles is an important aspect of QD examination that was explored. It involves interactively selecting unusual writing styles. The feasibility of extracting features similar to those used by QD examiners was demonstrated for the most commonly encountered letter formation of "th".

The methods developed have been shown to be of value in developing a scientific basis for handwriting examination, e.g., establishing individuality using cohort studies and non-contemporary English.

An approach to signature verification based on a fully Bayesian formulation has been proposed. In this task there are usually few samples to determine probabilities based on frequencies. In the Bayesian formulation parameters of prior distributions of between-class and same-class distributions are learnt from a database. The distributions are modified using any available samples.

2.3.1 Discussion of Findings

The methods have demonstrated the value of computational approaches to an area of forensic analysis largely based on the recall of human experts. The work was considered to be of sufficient value that it was cited in two Daubert hearings (U.S. v Prime [39] and U.S. v Lindsay) and a Frye hearing (DC Courts 2008) in all of which handwriting testimony was allowed to be presented.

2.3.2 Implications for Policy and Practice

The summary assessment of the NAS report [1] relative to handwriting analysis is concisely stated: "The scientific basis for handwriting comparisons needs to be strengthened. Recent studies have increased our understanding of the individuality and consistency of handwriting and computer studies and suggest that there may be a scientific basis for handwriting comparison, at least in the absence of intentional obfuscation or forgery. Although there has been only limited research to quantify the reliability and replicability of the practices used by trained document examiners, the Committee agrees that there may be some value in handwriting analysis".

In responding to the the assessment that the scientific basis for handwriting comparison needs to be strengthened, it seems clear that further research will need to be conducted. Computer-assisted handwriting analysis has been used in research, business and financial sectors for some time. A number of systems have been available to assist QD examiners in performing their work in forensic applications, either by making some tasks simpler, or by assisting in narrowing the number of items for them to review. As experience with systems like CEDAR-FOX grows, benchmarks of performance may be established, leading to the development of more testing, establishment of standards for the systems, and computation of recognized error rates.

Advances in the development of computer-assisted handwriting analysis have led to the consideration of a computational system by courts in the United States. Computer-assisted handwriting analysis has been introduced in the context of Frye or Daubert hearings conducted to determine the admissibility of handwriting testimony by QD examiners, as expert witnesses, in civil and criminal proceedings. This research provides a basis for scientific methods, and mitigates concerns over reliability of handwriting analysis expressed in judicial decisions.

Research has not disclosed the level of acceptance of computer-assisted handwriting analysis within the questioned document examiner community. Similarly, research has not yielded an indication of whether there is general acceptance in the field in which computer-assisted handwriting analysis belongs (under the Frye standard); nor does it appear that any study has been conducted or published to determine the degree of acceptance of computer-assisted handwriting analysis methodology within the relevant community (under the Daubert standard). While the issue of general acceptance of the methodology used by QD examiners appears to have been settled by appellate courts in the judicial system, there had been vigorous debate over whether the methodology belonged to the community comprised of QD examiners, or to a broader group of scholars and scientists. This debate was reflected in the judicial decisions , as well as peer review of the handwriting studies . In determining whether and when computer-assisted handwriting analysis achieves general acceptance, the relevant community will need to be ascertained by the courts, in order to determine the admissibility of the computer-assisted methodology and results.

Because courts, in making rulings in the context of Frye or Daubert hearings, are not required to follow rules of evidence, the admissibility of the computer-assisted methodology, computations, and results themselves have not been the subject of judicial evidentiary rulings. To date, the role of computer-assisted handwriting analysis in the judicial system has been limited to supporting, from a scientific viewpoint, the fundamental premise of the individuality of handwriting, and the ability to test and ascribe potential error rates to handwriting analysis. As computational systems and algorithms become more robust, and the accuracy and reliability of those systems improve, their performance may approach levels identified with the performance of questioned document examiners. As computational systems develop over time, parties in judicial proceedings may decide to offer direct evidence of the newer computational methods, as well as the results of the methods, either in Frye or Daubert hearings, or at trial. The offer of proof may come either as a computation performed by a QD examiner, or by a witness properly trained as a user of the computer-assisted handwriting analysis system. It does not appear, to date, that such an effort has been undertaken in a published case. In the event that both traditional QD analysis and computer-assisted handwriting analysis is conducted and offered as evidence in the same proceeding, and the Frye and Daubert standards are satisfied, the performance of the questioned document examiner may be measured against the output of the computational analysis. In this manner, the scores produced by the human examiner and the computational system may be compared and contrasted, permitting opportunity either for cross-validation, or for cross-examination.

Daubert teaches that a reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community. There also is guidance from a Frye jurisdiction, (New York): in defining the relevant scientific field, the Court must seek to comply with the Frye objective of obtaining a consensus of the scientific community. If the field is too narrowly defined, the judgment of the scientific community will devolve into the opinion of a few experts. The field must still include scientists who would be expected to be familiar with the particular use of the evidence at issue, however, whether through actual or theoretical research. . The National Research Council, in reviewing the scientific basis for handwriting comparisons, observed that questions about documents arise in any matter affected by the integrity of written communications and records. This observation may provide a cornerstone for analysis of the relevant community under either the Frye or Daubert standard. It is expected that, as computer-assisted handwriting analysis becomes more widely used and recognized, and handwriting analysis research continues to advance, the debate over the relevant community to which it belongs will resume.

2.3.3 Implications for Further Research

Some of the areas for further research are as follows.

1. A future potential role for computer-assisted handwriting analysis in the courts has been identified. Methods have to be further improved for general acceptance by the QD community.

2. One potentially important research is in determining the frequencies of different handwriting formations so that the probability of random correspondence can be stated. This involves developing generative models of handwriting. The potential of doing this in the case of the "th" structure was shown in this research. A larger effort covering other letter combinations needs to be pursued. It involves developing methods to extract features for a large number of letter combinations.

3. The methods for evaluating comparison uncertainty based on likelihood ratios need further evaluation, e.g., other similarity measures and statistical formulations.

4. Methods for comparing handwriting to standard style such as Palmer can be expanded to common styles used in schools today such as the Denelian

5. Identifying characteristics of the writing that make it unique needs further exploration.

6. The fully Bayesian approach proposed for signature verification can also be of value in the verification of extended handwriting as well as in other impression evidence domains such as latent prints, shoe-prints, etc. Further studies need to be made with forgeries, tracings and disguised signatures.

Chapter 3

Dissemination

3.1 Software Distribution

CEDAR-FOX can be downloaded freely for evaluation purposes. CEDAR-FOX can be downloaded from `http://www.cedartech.com/products/cedarfox.html` using username=cedarfox and password=buffalo.ny.

3.2 Scholarly Publications

3.2.1 United States Patent

Srihari, S. N., Y. C. Shin, et.al., Method and Apparatus for Analyzing and/or Comparing Handwritten and/or Biometric Samples, *United States Patent No. 7,580,551*, Aug 29, 2009. Assigned to the Research Foundation of the State University of New York.

3.2.2 Journal Papers

1. S. N. Srihari, H. Srinivasan and K. Desai, "Questioned Document Examination Using CEDAR-FOX," *Journal of Forensic Document Examination*, vol 18, 2007, pp. 1-20. Issue published in Fall 2008..

2. S. N. Srihari, C. Huang, H. Srinivasan and V. Shah, "On the Discriminability of the Handwriting of Twins," *Journal of Forensic Sciences*, 53(2), March 2008, 430-446.

3.2.3 Papers in Refereed Conference Proceedings

1. H. Srinivasan, S. Kabra, C. Huang and S. N. Srihari, "On Computing the Strength of Evidence for Writer Verification", *Proc. International Conference on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 844-848.

2. A. Bharadwaj, A. Singh, H. Srinivasan, S. N. Srihari, "On the Use of Lexeme Features for Writer Verification," *Proc. International Conference on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 1088-1092.

3. C. Huang and S. N. Srihari "Automatic Writer Verification with Handwriting of Twins and Non-Twins," *Proc. International Graphonomics Society Conference*, Melbourne, Australia, 2007, pp. 184-187.

4. K. Kuzhinjedathu, H. Srinivasan and S. N. Srihari, "Segmentation of Overlapping Handwritten Lines," *Proc. International Graphonomics Society Conference*, Melbourne, Australia, 2007, pp. 28-31.

5. K. Kuzhinjedathu, H. Srinivasan and S. N. Srihari, "Robust Line Segmentation for Handwritten Documents," in *Document Recognition and Retrieval XV*, B. A. Yanikoglu and K. Berkner (eds.), SPIE & IS&T, Bellingham, WA, 2008, pp. 68150C 1-9.

6. C. Huang and S. N. Srihari, "Word Segmentation of Off-line Handwritten Documents," in *Document Recognition and Retrieval XV*, B. A. Yanikoglu and K. Berkner (eds.), SPIE & IS&T, Bellingham, WA, 2008, pp. 68150E 1-6.

7. S. N. Srihari, K. Kuzhinjedathu, H. Srinivasan, Chen Huang, and D. Pu,"Off-line Signature Verification using Bayesian Approach," *Computational Forensics: Proceedings of International Workshop, IWCF 2008*, Washington DC, 2008, Springer LNCS 5158, pp. 192-203.

8. G. R. Ball and S. N. Srihari,"Prototype Integration in Off-line Handwriting Recognition Adaptation," *Proceedings International Conference on Frontiers in Handwriting Recognition*, Montreal, Canada, 2008, pp. 529-534.

9. S. N. Srihari and G. R. Ball,"Writer Verification for Arabic Handwriting," *Proceedings Document Analysis Systems (DAS 2008)*, Nara, Japan, September 17-19, 2008, IEEE Computer Society Press, pp. 28-34.

10. G. R. Ball and S. N. Srihari, "Comparison of statistical models for writer verification," *Proceedings of Document Recognition and Retrieval XV*, San Jose, CA, January 2009, Bellingham, WA: SPIE.

11. G. R. Ball, V. Ramakrishnan and S. N. Srihari, "Identification of forgeries in handwritten ballot propositions," *Proceedings of Document Recognition and Retrieval XV*, San Jose, CA , January 2009, Bellingham, WA: SPIE.

12. D. Pu, G. R. Ball and S. N. Srihari, "A Machine Learning Approach to Off-Line Signature Verification using Bayesian Inference," *Computational Forensics: Proc. Third International Workshop, IWCF 2009*, The Hague, The Netherlands, Springer 2009, pp. 125-136.

13. K. Manning and S. N. Srihari, "Computer-Assisted Handwriting Analysis: Interaction with Legal Issues in U. S. Courts," *Computational Forensics: Proc. Third International Workshop, IWCF 2009*, The Hague, The Netherlands, Springer 2009, pp. 137-149.

3.2.4 Book Chapters

1. S. N. Srihari, C. Huang, H. Srinivasan and V. A. Shah, "Machine Learning for Signature Verification," in *Machine Learning in Document Analysis and Recognition,* S. Marinai and S. Fujisawa (eds.), Springer, 2008, pp. 387-408.

2. H. Srinivasan and S. N. Srihari, "Use of Conditional Random Fields for Signature-based Retrieval of Scanned Documents," in *Computational Methods for Counterterrorism,* S. Argamon and N. Howard (eds). Springer, 2009, pp. 17-32.

3.3 Presentations

3.3.1 Conferences

Portions of this research were presented at the following conferences:

1. At the *International Conference on Document Analysis and Recognition*, Curitiba, Brazil on September 25-26, 2007

2. At the *Northeast Association of Forensic Scientists (NEAFS)*, Lake George, NY on October 1, 2007

3. At the *International Graphonomics Society Conference*, Melbourne, Australia on November 12 and 14, 2007

4. "Robust Line Segmentation for Handwritten Documents," *Document Recognition and Retrieval XV Conference*, San Jose, CA, January 30, 2008.

5. "Word Segmentation of Off-line Handwritten Documents," *Document Recognition and Retrieval XV Conference*, San Jose, CA, January 30, 2008.

6. "Quantitative Assessment of Handwriting," *NIJ Grantees meeting in conjunction with American Academy of Forensic Sciences (AAFS) Annual Conference*, Washington, DC, February 19, 2008.

7. "The CEDAR-FOX system and Recent research results," *Southwest Association of Forensic Document Examiners (SWAFDE)* . This was a one-day plenary workshop at the SWAFDE annual meeting in San Diego on April 12, 2008.

8. "Off-line Signature Verification using Bayesian Approach," *International Workshop on Computational Forensics*, Washington DC, August 7, 2008.

9. "Prototype Integration in Off-line Handwriting Recognition Adaptation," *International Conference on Frontiers in Handwriting Recognition*, Montreal, Canada, August 19, 2008.

10. "Writer Verification for Arabic Handwriting," *International Workshop on Document Analysis Systems*, Nara, Japan, September 17, 2008.

11. "Comparison of statistical models for writer verification," Conference on *Document Recognition and Retrieval*, under the auspices of the Society of Photo Instrumentation Engineers (SPIE), San Jose, CA, on January 22, 2009.

12. "Individuality of Handwriting: A Twins' Study," Invited talk at QD section of the *American Academy of Forensic Sciences (AAFS)*, annual meeting in Denver on February 19, 2009.

13. "CedarFox," Invited talk at annual meeting of Southern Association of Forensic Document Examiners (SAFDE) in Atlanta, GA on April 16, 2009.

14. G. R. Ball and S. N. Srihari, "Semi-supervised learning in handwriting recognition," *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, Barcelona, Spain, July 2009, IEEE Computer Society Press. Presented as a full paper with an oral presentation, which had an acceptance rate of 20%.

15. D. Pu, G. R. Ball and S. N. Srihari, "Off-line Signature Verification using Bayesian Approach," *Proceedings of International Workshop on Computational Forensics (IWCF)*, The Hague, Netherlands, August 2009, Springer.

16. K. Manning, Esq and S. N. Srihari, "Computer-Assisted Handwriting Analysis: Interaction with Legal Issues in U. S. Courts," *Proceedings of International Workshop on Computational Forensics (IWCF)*, The Hague, Netherlands, August 2009, Springer.

Chapter 4

Figures

Figures referred to in the narrative section of the report (Chapter 2) follow.

List of Figures

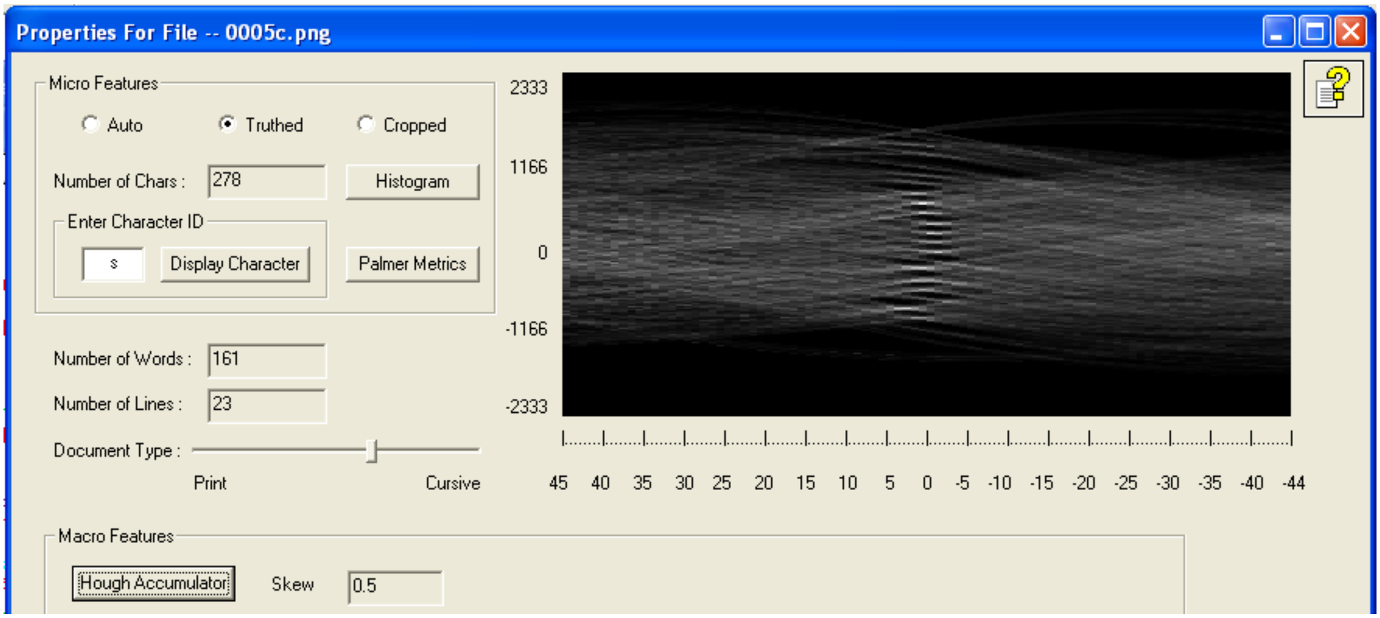
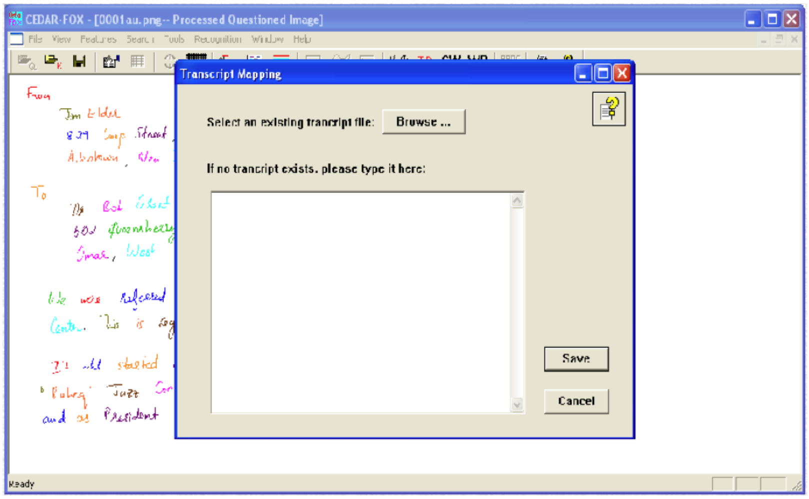
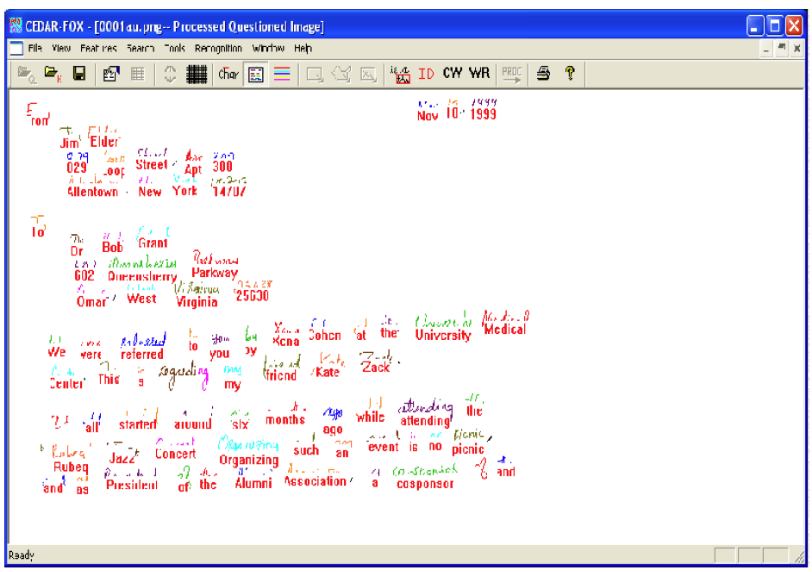


Figure 4.1: Document properties are displayed after pre-processing. It includes: an assessment of whether the writing is cursive or printed, the estimated number of lines, words and characters in the writing, and the line structure (wavy forms on top right are the output of a Hough transform accumulator array that shows the relative orientation and distance to the origin). Clicking on the buttons yield further information such as the degree of similarity to Palmer style of handwriting, and display any selected character in the writing

.



(a) Interface to type transcript.



(b) Display of words associated with images.

Figure 4.2: Tool for extracting characters from image. (a) Transcript mapping allows automatic matching of word image snippets to associated text from a typed transcript of the handwriting. (b) Truth super-imposed on handwriting image. Results can be filtered to get desired letter combination snippets.

.

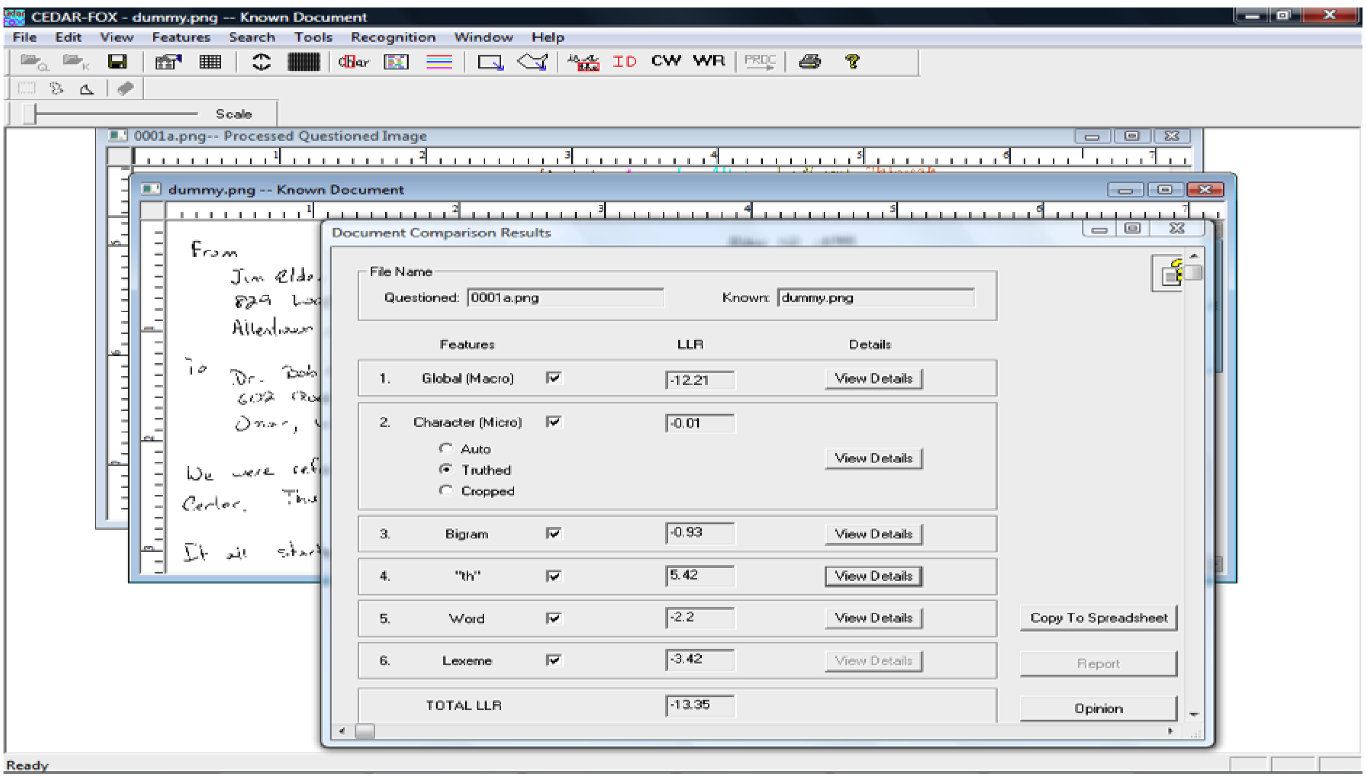


Figure 4.3: Display of document comparison results. Relative contributions of different types of features, including macro, micro and "th", are shown. In this case the LLR is negative indicating that the samples were not written by the same person.

.

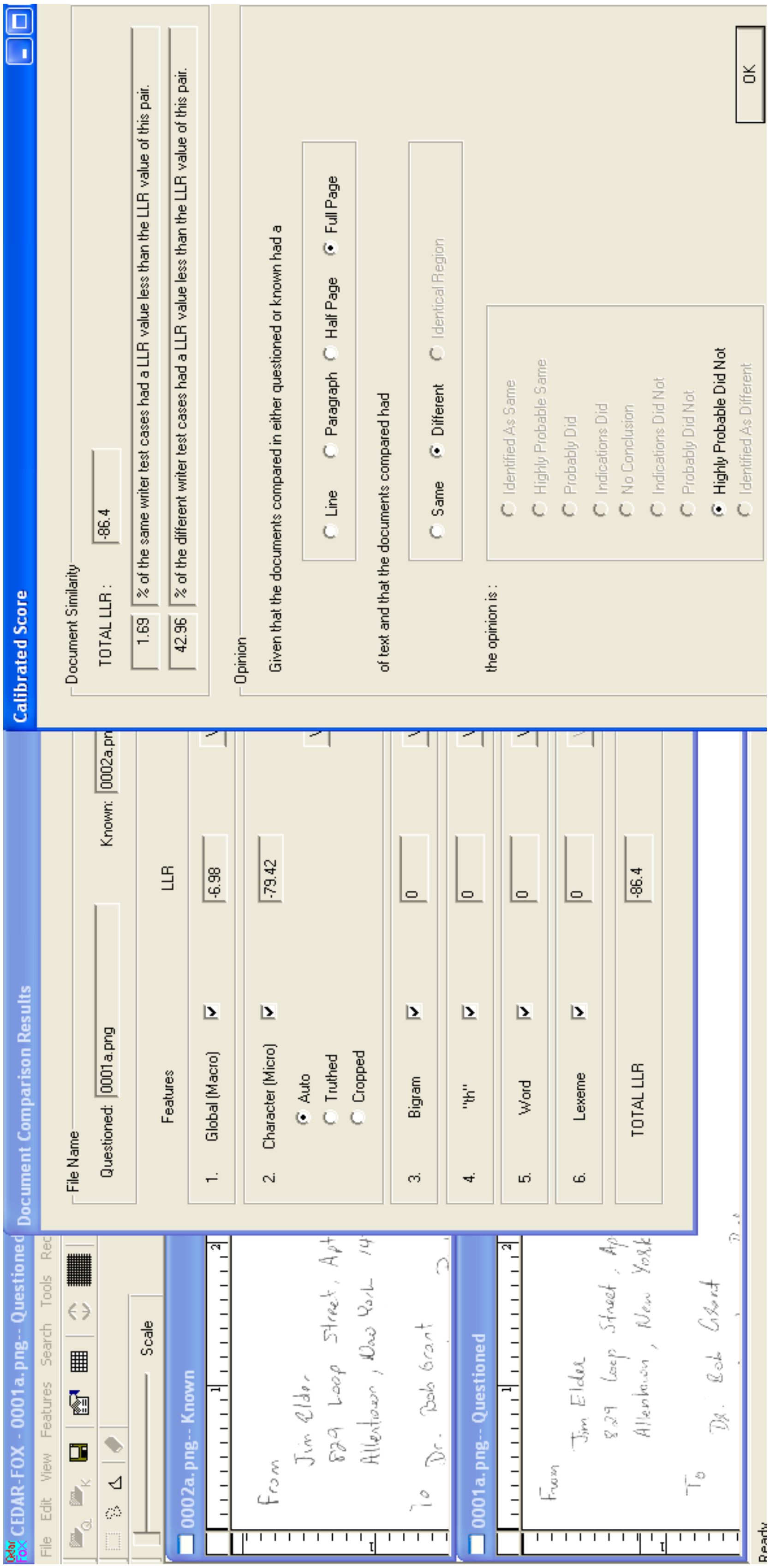


Figure 4.4: Opinion of comparison which consists an LLR value as well as a conclusion on a nine-point scale.

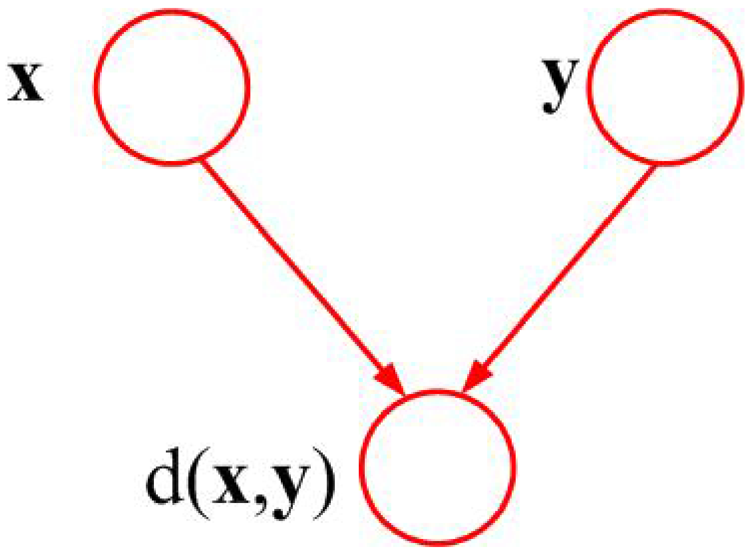
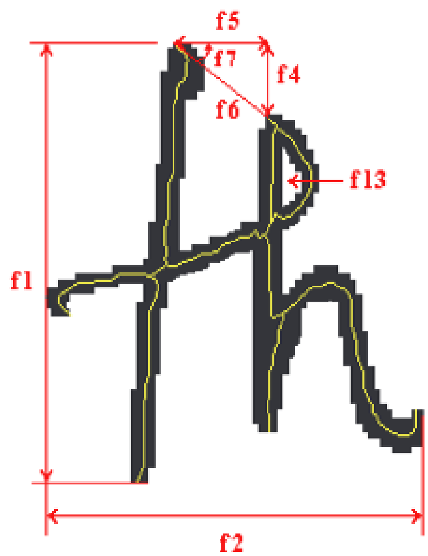


Figure 4.5: Probabilistic graphical model that relates evidence $\mathbf{x}$, known $\mathbf{y}$, and similarity $d(\mathbf{x}, \mathbf{y})$
.

 LIST OF FIGURES

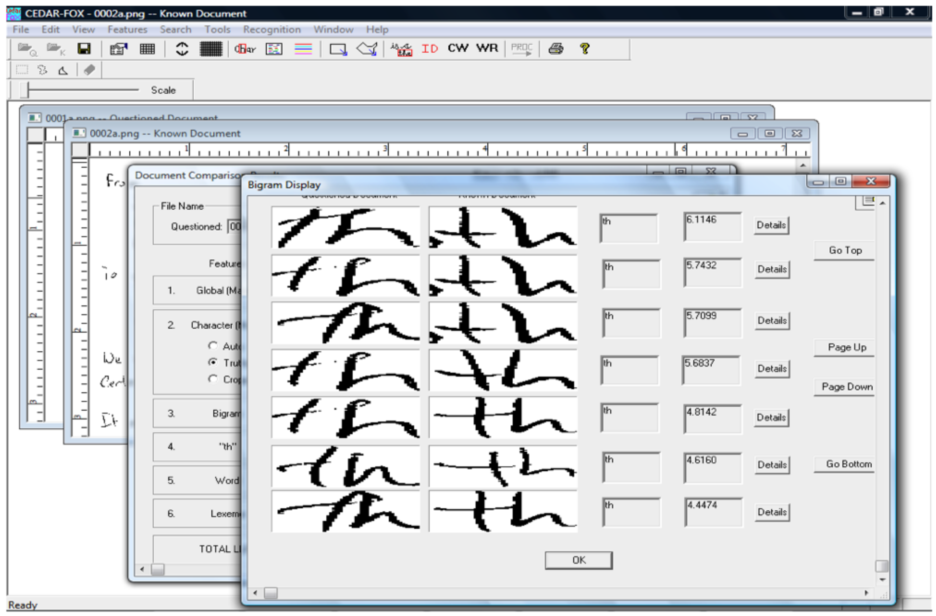


(a) Letter "th" combination.

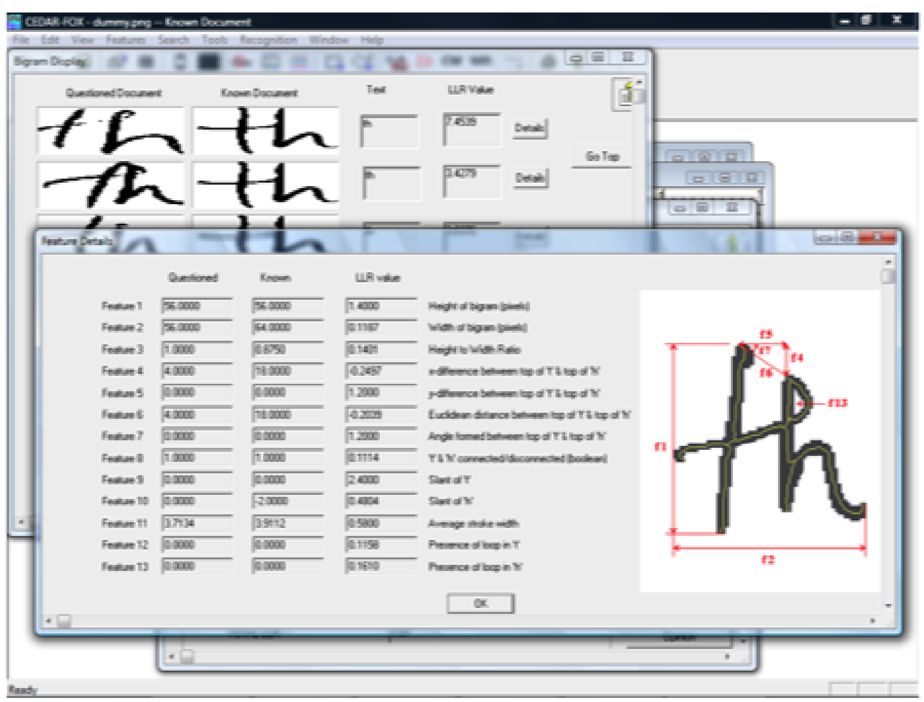
- Height of bigram (pixels)
- Width of bigram (pixels)
- Height to Width Ratio
- x-difference between top of 't' & top of 'h'
- y-difference between top of 't' & top of 'h'
- Euclidean distance between top of 't' & top of 'h'
- Angle formed between top of 't' & top of 'h'
- 't' & 'h' connected/disconnected (boolean)
- Slant of 't'
- Slant of 'h'
- Average stroke width
- Presence of loop in 't'
- Presence of loop in 'h'

(b) Features extracted.

Figure 4.6: Letter combination "th" and thirteen features as computed by the system.

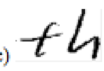


(a) Six pairs of *th* from two documents and their relative contributions to discrimination. All six have positive LLRs (fourth column) indicating different writership. Clicking the "Details" button displays the following screen.



(b) A single *th* and the LLR contributions of its 13 characteristics.

Figure 4.7: Contribution of "th" features in comparison.

TABLE B - Height relation of the "t" to the "h": (a)  (b)  (c)  (d)

	samples	2a	2b	2c	2d	3a	3b	4a	4b	4c	4d
(a)"t" shorter than "h"	47.27% (251)	9.96% (25)	15.54% (39)	7.97% (20)	66.53% (167)	78.09% (196)	21.9% (55)	41.4% (104)	18.32% (46)	21.11% (53)	11.1% (28)
(b) * "t" even with "h"	22.79% (121)	9.09% (11)	12.40% (15)	3.31% (4)	75.21% (91)	80.17% (97)	19.83% (24)	12.40% (15)	0	87.60% (106)	0
(c)t" taller than "h"	29.94% (159)	11.32% (18)	5.03% (8)	5.66% (9)	77.99% (124)	76.1% (121)	23.90% (38)	28.93% (46)	38.36% (61)	24.53% (39)	8.18% (13)

* A sample reading of the table is: of the 121 people making "t" even with "h", 91 made "t" and "h" with no loops and 97 made "t" and "h" connected.

(a) Height Relationships.

TABLE C - Presence of loops:

	samples	1a	1b	1c	3a	3b
(a) "th" with only a loop in "t"	10.17% (54)	46.3% (25)	20.3% (11)	33.3% (18)	88.8% (48)	11.11% (6)
(b) "th" with only a loop in "h"	11.68% (62)	62.9% (39)	24.1% (15)	12.9% (8)	85.4% (53)	14.52% (9)
(c) made "th" with loop in both	6.21% (33)	60.6% (20)	12.1% (4)	27.2% (9)	81.8% (27)	18.18% (6)
(d) made "th" with no loops *	71.94% (382)	43.7% (167)	23.8% (91)	32.4% (124)	74.8% (286)	25.13% (96)

* A sample reading of the table is: of the 382 people making "t" and "h" with no loops, 91 made "t" even with "h" and 96 made "t" and "h" disconnected.

(b) Presence of Loops.

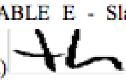
Figure 4.8: Conditional distributions of "th" conditioned on the first and second characteristics.

TABLE D - "t" and "h" connected/disconnected:

	samples	2a	2b	2c	2d	4a	4b	4c	4d
(a) both of them connected *	77.97% (414)	11.59% (48)	12.80% (53)	6.52% (27)	69.08% (286)	34.54% (143)	26.57% (110)	26.81% (111)	12.08% (50)
(b) both of them disconnected	22.03% (117)	5.13% (6)	7.69% (9)	5.13% (6)	82.05% (96)	40.17% (47)	28.21% (33)	24.79% (29)	6.84% (8)

* A sample reading of the table is: of the 414 people making "t" connected with "h", 48 made a loop only in "t" and 50 made "t" and "h" shaped outward.

(a) Connectedness of "t" and "h".

TABLE E - Slant of "t" and "h": (a)  (b)  (c) 
(d) 

	samples	1a	1b	1c	2a	2b	2c	2d	3a	3b
(a) both positive slants *	35.78% (190)	56.84% (108)	18.95% (36)	24.21% (46)	7.37%(14)	17.37% (33)	9.47(18)	65.79% (125)	75.26% (143)	24.74% (47)
(b) both negative slants	26.93% (143)	33.57% (48)	23.78% (34)	42.66% (61)	9.79% (14)	5.59% (8)	4.90% (7)	79.72% (114)	76.92% (110)	23.08% (33)
(c) inward	26.37% (140)	45.71% (64)	26.43% (37)	27.86% (39)	14.29% (20)	7.86% (11)	3.57% (5)	74.29% (104)	79.29% (111)	20.71% (29)
(d) outward	10.92% (58)	53.45% (31)	24.14% (14)	22.41% (13)	10.34% (6)	17.24% (10)	5.17% (3)	67.24% (39)	86.21% (50)	13.79% (8)

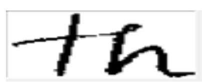
* A sample reading of the table is: of the 190 people making both "t" and "h" with positive slants, 46 made "t" taller than "h" and 125 made "t" and "h" with no loops.

(b) Slant of "t" and "h".

Figure 4.9: Conditional distributions for "th" conditioned on the third and fourth characteristics.

.

Case A:



From the feature values of CEDAR fox the following inferences for the above image can be made

a. The writer made't' taller than 'h'.
b. Neither of them contained a loop.
c. 't' and 'h' are connected.
d. Slant of both't' and 'h' is positive.

Thus the joint probability is:

p(x1, x2, x3, x4) = .2893.748*.7797*.3578 = .0604*

Case B:

From the feature values of CEDAR fox the following inferences for the above image can be made

a. The writer made't' shorter than 'h'.
b. Neither of them contained a loop.
c. 't' and 'h' are disconnected.
d. Slant of both't' and 'h' is negative.

Thus the joint probability is:

p(x1, x2, x3, x4) = .1832.748*.2203*.2693 = .0081*

(a)

Case C:

From the feature values of CEDAR fox the following inferences for the above image can be made

a. The writer made't' shorter than 'h'.
b. Neither of them contained a loop.
c. 't' and 'h' are disconnected.
d. Slant of both't' and 'h' is inward.

Thus the joint probability is:

p(x1, x2, x3, x4) = .2111.2513*.2203*.2637 = .0031*

Case D:

From the feature values of CEDAR fox the following inferences for the above image can be made

a. The writer made't' shorter than 'h'.
b. Neither of them contained a loop.
c. 't' and 'h' are disconnected.
d. Slant of both't' and 'h' is inward.

Thus the joint probability is:

p(x1, x2, x3, x4) = .2111.2513*.2203*.2637= .0031*

(b)

Figure 4.10: Four examples of *th* images and their computed probabilities. A: 0.06, B: 0.008, C: 0.003 and D: 0.003.

.

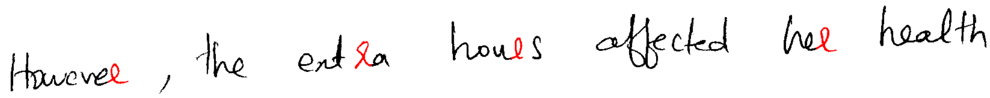


Figure 4.11: Uniquely written ‘r’ characters are consistently present in this writer’s sample

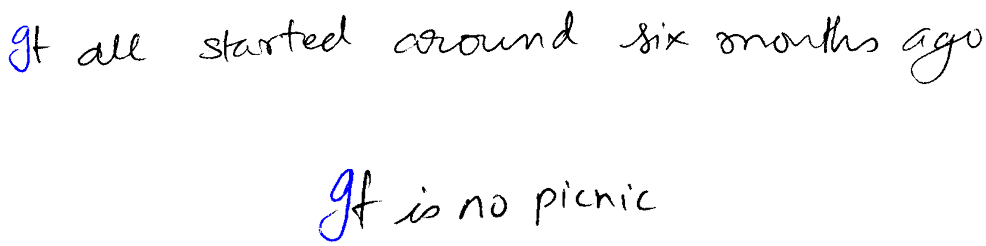
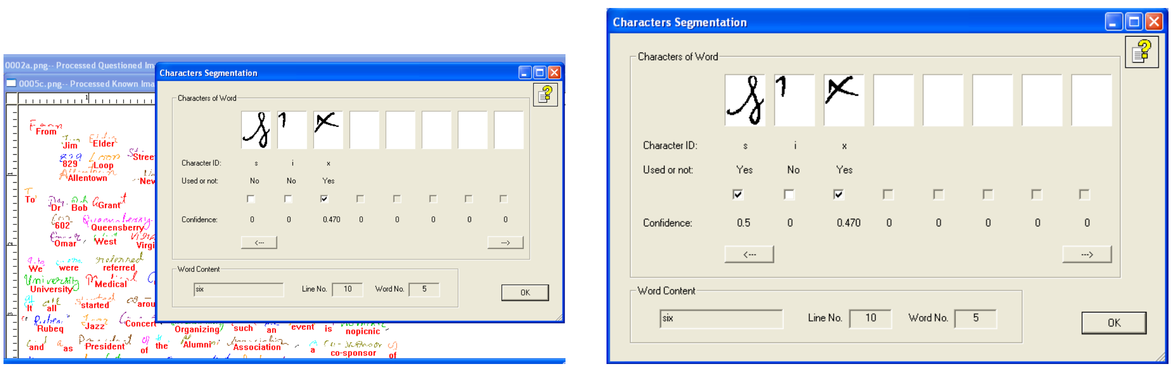


Figure 4.12: Uniqueness of writing: unusual ‘I’ characters present in two samples reflect their origin by the same writer.

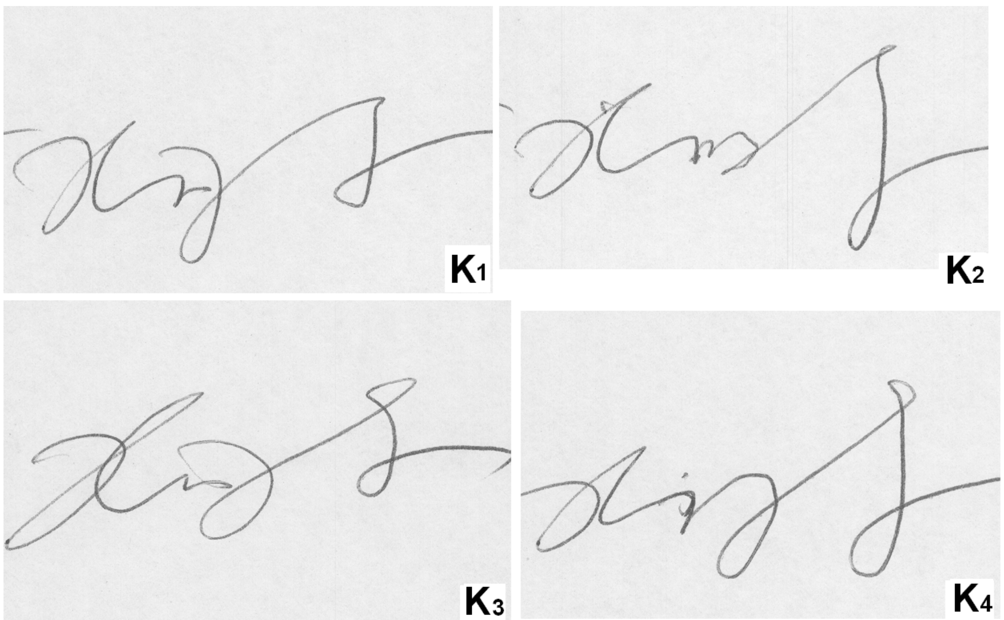
Figure 4.13: Uniqueness of writing: prototype ‘I’ characters written by different authors



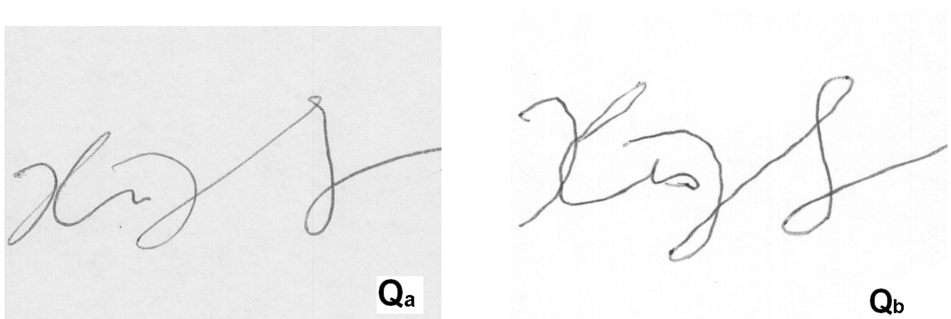
(a) (b)

Figure 4.14: Uniqueness of writing. Screenshots show selecting a character for inclusion even when it does not conform to the prototype. Originally we added a second box whether or not it is explicitly unique (we have both versions still), but if it was flagged unique, it would be something that should be included so we went back to a single box–either way, we can easily do either. After the checkbox is selected, the character is used for comparison regardless of its conformity to the prototypes.

.

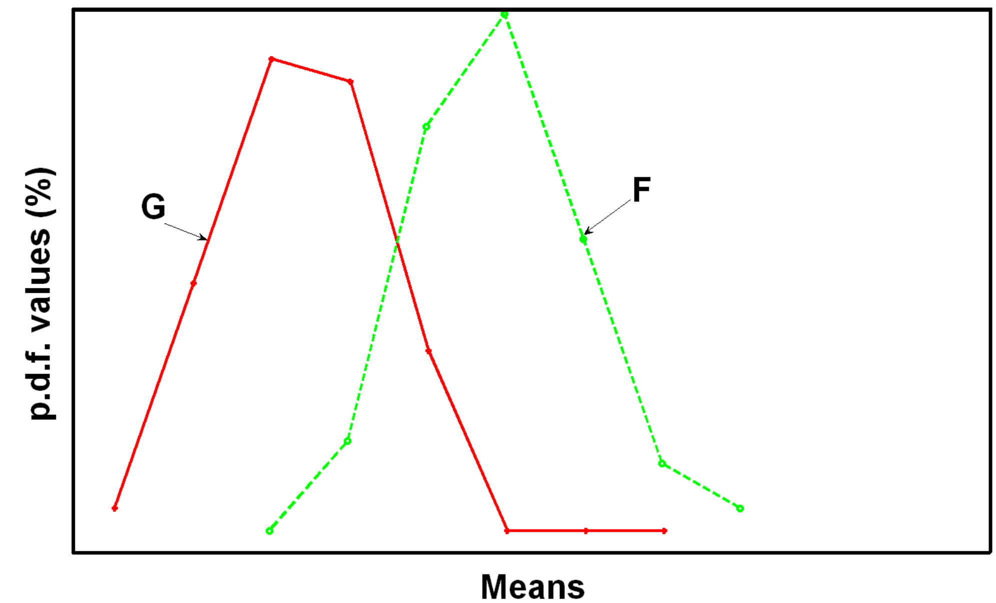


(a)

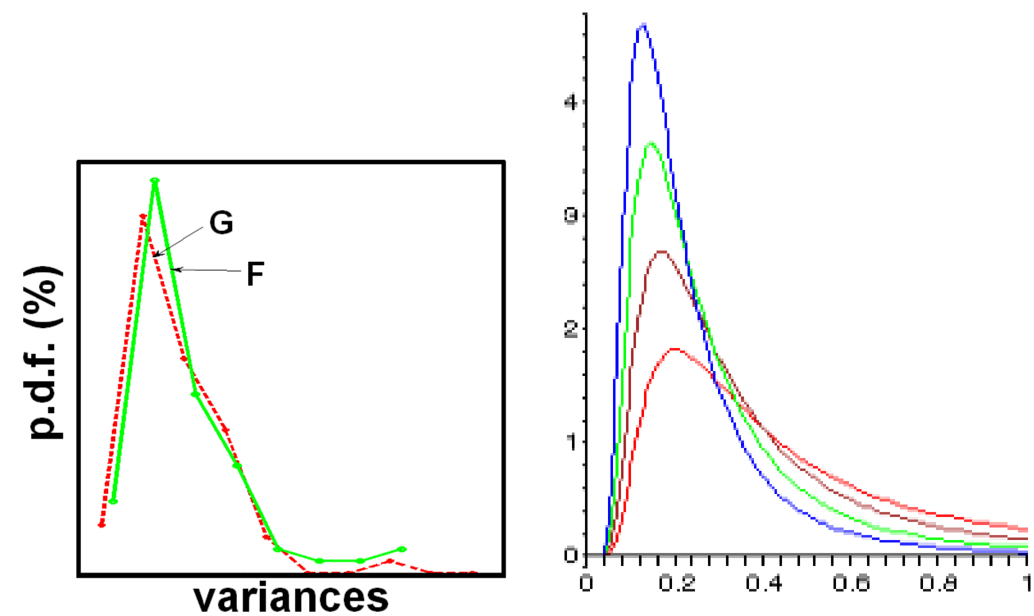


(b) (c)

Figure 4.15: Bayesian Signature Verification: (a) small learning set of four known signatures, (b) genuine signature (Q_a) with LLR=**9.19**, (c) skilled forgery (Q_b) with LLR= **-2.47**.



(a)



(b) (c)

Figure 4.16: Choice of parameter distributions. (a) Histogram of mean of genuine and forgery distances which are Gaussian-like; (b)Histogram of variance of genuine and forgery distances which are Inverse-chi-square-like; (c)Inverse-chi-square distributions.

Chapter 5

References

The references in the Program Narrative are given below. The references are in the order in which they are referred to in the text.

Bibliography

[1] NAS: Strengthening the Forensic Sciences: A Path Forward. National Academy of Sciences press (2009)

[2] App DC: Frye v United States. 54, 293 F **1013** (1923)

[3] United States Supreme Court: Daubert v Merrell Dow Pharmaceuticals. 509 U.S. **579** (1993)

[4] United States Supreme Court: Kumho v Carmichael et. al. 526 U.S. **137** (1999)

[5] Manning, K., Srihari, S.N.: Computer-assisted handwriting analysis: Interaction with legal issues in U. S. courts. In: Proc. Third International Workshop on Computational Forensics, The Hague, Netherlands, Springer (2009) 137–149

[6] S.D NY: U. S. v Starzecpyzel. 880 F. Supp **1027,1037** (1995)

[7] D. Alaska: U. S. v Saelee. 162 F. Supp 2d **1097,1102** (2001)

[8] 4th Cir: U. S. v Crisp. 324 F 3d **261,280** (2003)

[9] D. Mass: U. S. v Hines. 55 F Supp 2d **62,68** (1999)

[10] Srihari, S., Leedham, C.: A survey of computer methods in forensic document examination. In: Proceedings of the 11th Conference of the International Graphonomics Society, IGS 2003, Scottsdale, AZ. (2003)

[11] Harrison, D., Burkes, T.M., Sieger, D.P.: Handwriting examination: Meeting the challenges of science and the law. Forensic Science Communications **11** (2009)

[12] Osborn, A.: Questioned Documents. Nelson Hall Pub (1929)

[13] Muehlberger, R.J., Newman, K.W., Regent, J., Wichmann, J.G.: A statistical examination of selected handwriting characteristics. Journal of Forensic Sciences (1977) 206–210

[14] Srihari, S.N., Cha, S., Arora, H., Lee, S.: Individuality of handwriting. Journal of Forensic Sciences **44(4)** (2002) 856–872

[15] Srihari, S.N., Huang, C., Srinivasan, H.: Discriminability of the handwriting of twins. Journal of Forensic Sciences **53(2)** (2008) 430–446

[16] Srihari, S.N., Zhang, B., Tomai, C., Lee, S., Shi, Z., Shin, Y.C.: A system for handwriting matching and recognition. In: Proc. Symposium on Document Image Understanding Technology. (2003) 67–75

[17] Lee, S., Cha, S., Srihari, S.N.: Combining macro and micro features for writer identification. SPIE, Document Recognition and Retrieval IX, San Jose, CA (2002) 155–166

[18] Huang, C., Srihari, S.: Mapping transcripts to handwritten text. In: Proc. International Workshop on Frontiers in Handwriting Recognition (IWFHR-10), La Baule, France. (October 2006) pp. 15–20

[19] Pervouchine, V.: Ph.D. Thesis: Discriminative Power of Features Used by Forensic Document Examiners in the Analysis of Handwriting. Nanyang University (2006)

[20] Robertson, E.W.: Fundamentals of Document Examination. Nelson-Hall (1991)

[21] Bradford, R.R., Bradford, R.: Introduction to Handwriting Examination and Identification. Nelson-Hall (1992)

[22] Hilton, O.: Scientific Examination of Questioned Documents. CRC Press (1993)

[23] Huber, R.A., Headrick, A.M.: Handwriting identification:facts and fundamentals. CRC Press (1999)

[24] Slyter, S.A.: Forensic Signature Examination. Charles C. Thomas Pub (1995)

[25] Leclerc, F., Plamondon, R.: Automatic signature verification: the state of the art, 1989-1993. International Journal of Pattern Recognition and Artificial Intelligence **8** (1994) 643–660

[26] Guo, J.K., Doermann, D., Rosenfield, A.: Local correspondences for detecting random forgeries. In: Proceedings of the Third International Conference on Document Analysis and Recognition. (1997) 319–323

[27] Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 63–84

[28] Kalera, M.K., Zhang, B., Srihari, S.N.: Off-line signature verification and identification using distance statistics. International Journal of Pattern Recognition and Artificial Intelligence **18** (2004) 1339–1360

[29] Fang, B., Leung, C.H., Tang, Y.Y., Tse, K.W., Kwok, P.C.K., Wong, Y.K.: Off-line signature verification by the tracking of feature and stroke positions. Pattern recognition **36** (2003) 91–101

[30] Lindley, D.: A Problem in Forensic Science. Biometrika **64** (1977) 207–213

[31] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Bromage-Griffiths, A.: Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. Journal of Forensic Sciences **51** (2007) 1255–1266

[32] Ball, G.R., Srihari, S.N.: Prototype integration in off-line handwriting recognition adaptation. In: Proceedings International Conference on Frontiers in Handwriting Recognition, Montreal, Canada. (2008) 529–534

[33] Ball, G.R., Srihari, S.N.: Comparison of statistical models for writer verification. In: Proceedings Document Retrieval and Recognition XVI, San Jose, CA, SPIE (2009) 7247OE 1–8

[34] Bharadwaj, A., Singh, A., Srinivasan, H., Srihari, S.N.: On the use of lexeme features for writer verification. In: Proc of the International Conference on Document Analysis and Recognition. (2007)

[35] M. Melhi, S. S. Ipson, W.B.: A novel triangulation procedure for thinning handwritten text. In: Pattern Recognition Letters. Volume 22. (2001) 1059–1071

[36] Ball, G.R., Stittmeyer, R., Srihari, S.N.: Writer verification in historical documents. In: Proceedings Document Recognition and Retrieval XVII San Jose, CA, SPIE (2010)

[37] Srihari, S.N., Ball, G.R.: Writer verification of arabic handwriting. In: Proc. Eighth Int. Workshop on Document Analysis Systems, Nara Japan, IEEE Computer Society Press (2008) 28–34

[38] Pu, D., Srihari, S.: A probabilistic measure for signature verification based on Bayesian learning. In: Proc International Conference on Pattern Recognition, IEEE Computer Society Press (2010)

[39] 9th Cir: United States v Prime. 431F 3d **1147,1153** (2005)