Automatic Detection of Handwriting Forgery Using a Fractal Number Estimate of Wrinkliness

Sung-Hyuk Cha¹, Charles C. Tappert¹, Michael Gibbons¹, and Yi-Min Chee²

¹Computer Science Department, Pace University, Pleasantville, NY, USA
²IBM T.J. Watson Research Center, Yorktown Heights, NY, USA
scha@pace.edu, ymchee@us.ibm.com, ctappert@pace.edu

Abstract

We investigate the detection of handwriting forged by novices. To facilitate document examination it is important to develop an automated system to identify forgeries, or at least to identify those handwritings that are likely to be forged. Because forgers often carefully copy or trace genuine handwriting, we hypothesize that good forgeries – those that retain the shape and size of genuine writing – are usually written more slowly and are therefore wrinklier (less smooth) than genuine writing. From online handwriting samples we find that the writing speed of the good forgeries is significantly slower than that of the genuine writings. From corresponding offline samples we find that the wrinkliness of the good forgeries is significantly greater than that of the genuine writings, showing that this feature can help identify candidate forgeries from scanned documents. Using a total of eight handwriting distance features, including the wrinkliness feature, we train a neural network to achieved 89% accuracy on detecting forged handwriting on test samples from ten writers.

Key Words: Forgery detection, handwriting analysis, wrinkliness, fractal

1. Introduction

Since questioned document examinations play an important investigative and forensic role in many types of crime [1, 2], it is necessary to build a system that objectively identifies forged handwriting from offline images. Various automatic writer identification computer techniques,
feature extraction, comparison, and performance evaluation methods have been studied (see [3, 4] for an extensive survey). In a study to establish the individuality in handwriting, Srihari, et al., successfully designed two models to establish the individuality with high confidence: a writer identification system and a writer verification system [5, 6]. However, these models were based on the assumption that subjects provide their handwriting samples in their natural handwriting style, and the study did not cover forgery and disguised writing.

It is unknown whether writership can be verified from offline images if some of the writing samples are forgeries. For this reason, we conducted a preliminary study to measure the capability of humans and machines to detect forgery [7]. There were three stages in this study: i) collecting handwriting samples, ii) developing a classification system, and iii) running experiments. Subjects were asked to write test samples in their natural handwriting style and to forge other subjects’ handwriting samples. The resulting handwriting samples were scanned and stored digitally. Word-level features were then computed from the writing samples, and a neural network was trained using the dichotomy model [5, 6, 8] to distinguish between genuine and forged handwriting.

We recently extended the study of the individuality of handwriting by developing a pilot system to automatically detect handwriting forgery [7, 9]. One of the interesting features of this study was a measure of the variability of the handwriting on a small scale. Although one can copy the shape of another’s handwriting, it is difficult to mimic the dynamic aspects, such as speed and acceleration. Because forged handwriting tends to be drawn slowly, it should be wigglier than genuine handwriting. This wrinkliness feature can be measured using a fractal dimension measure. For example, in the paper, “How long is the coastline of Britain?” a fractal measure of
the wrinkliness of the coastline was suggested [10]. Because the measured length of the coastline depends upon the size of the measuring stick, this problem can be answered in terms of fractal dimension. Hence, we applied this measure as one of the features to detect forgery. While fractal dimensions have been used for handwriting analysis and style recognition [11, 12], our study is the first to use them to detect forgeries.

The key idea of this study is that successful forgers, especially novice ones, often forge handwriting shape and size by carefully copying or tracing the genuine handwriting. Our first hypothesis, therefore, is that good forgeries – that is, those that retain the shape and size of genuine writing – tend to be written more slowly than genuine writing. Based on the first hypothesis, our second hypothesis is that good forgeries are also likely to be wrinklier (less smooth) than genuine handwriting. We examine these hypotheses by utilizing both online and offline data from the same handwriting samples. Our objective is to develop algorithms that work with offline images. We use online data only to verify that novices write forgeries slower than usual to support the correctness of the wrinkliness hypothesis.

This paper is organized as follows. In section 2, we explain the method of collecting genuine and forged handwriting samples. Section 3 discusses the features extracted from the handwriting. Section 4 describes the experiments that verify the speed and wrinkliness hypotheses and that test whether forgery can be detected automatically. Finally, section 5 draws conclusions of this work.

2. Handwriting Data Collection
In order to study the general problem of forged handwriting, we selected ordinary words to forge rather than signatures, and for this preliminary study we used only the word *April*. The use of ordinary words rather than signatures provides the broadest application and generality of the study, and the use of a single word should not overly bias the results of this preliminary study. Because signatures are generated by a highly learned automatic process we would expect the differences, especially for the speed and wrinkliness measures, to be even more pronounced for signatures than for ordinary words.

Each of ten subjects was asked to write the word *April* three times on provided ruled paper mounted on an IBM ThinkPad TransNote, pen-enabled notebook computer in their most natural handwriting style. As expected, the handwritings consisted of both cursive and handprinted words depending on writers’ natural handwriting styles (Fig. 2). Subjects were also asked to forge the word *April* three times from the handwriting samples of each of the nine other writers. Figure 1 shows three genuine writing samples and six forgeries. All of the collected handwritten samples were scanned, digitized, and stored in an image database together with demographic information on the subjects. While forging practice was not formalized, the writers tended to practice the forging of each writing sample two or three times. Thus, we obtained 30 genuine samples of the word *April* (three from each of the ten writers) and 270 forgeries (three word forgeries of the nine other writers from each of the ten writers) for a total of 300 words.

---

1 We originally planned to use a set of words that contains all word-initial alphabet characters for both upper and lowercase, and word middle and terminal positions for the lowercase alphabet [6]: {April, Bob, California, December, English, February, Greg, Halloween, Iraq, June, Kentucky, Los Angeles, Markov, November, October, Pennsylvania, Queen, Raj, States, Texas, United, What, Xray, York, Zorro, alumni, boy, come, date, enjoy, false, great, have, interest, jazz, keep, leave, millennium, now, of, picnic, question, run, six, time, unique, video, where, xenophobia, you, zero}.
Using the IBM Ink Manager SDK, we saved the online data from the IBM TransNote digitizer. In addition, using a Canon CanoScan LiDE 30 Flatbed Scanner, the paper on which the handwriting samples were written was digitally scanned at both 300 and 600 dpi. This yielded a total of 900 database records – 300 online and 600 offline records. The record format consisted of: ID of the subject, online/offline, ID of the copied subject, word written, first/second/third try, and sampling rate (online) or resolution (offline).

3. Feature Extraction

We developed eight word-level features: some come from the handwriting recognition and identification literature [13], and one comes from Fractal theory. Since the dichotomy model, which transforms the features into a distance space, is used to detect the forgeries, the features need not be homogeneous, and can be in any form as long as good distance measures are
associated with them [6, 9]. In addition, the speed of online handwriting is also recorded for the correlation study between wrinkliness and speed of handwriting.

3.1. Computing Handwriting Features

The first feature is the centroid ratio. After counting all black pixels on each row and column, the centroid can be found by averaging them. After a bounding box is computed from a word, the centroid ratio is found by dividing the x-centroid by its height and y-centroid by its width. Hence, the centroid is in the form of a two-dimensional vector. To compute the distance between the centroids from two different handwritings, the simple Euclidean vector distance is used.

We borrowed the next three features from those established in many document recognition systems: slant, stroke width, and ascender/descender. The slant feature is a simple numeric value. Average stroke width was computed separately from three parts of the word, since the average stroke width does not vary greatly when computed in pixel form, and thus it forms a three-dimensional vector. The ascender and descender information was combined into a two-dimensional vector.

Our next two features were the popular projected histograms. The side-projected histogram and bottom-projected histograms are stored separately and the histogram distance [14] is used to measure the distance. Our next feature is the gradient histogram, and two gradient histograms are compared using the angular histogram distance.

Finally, the wrinkliness, which is a simple numeric value, is computed, and a detailed description of this is given in the next section. In summary, we have eight features: centroid ratio,
slant, stroke width, ascender/descender, side-projected histogram, bottom-projected histogram, gradient histogram, and wrinkliness.

3.2. Computing Handwriting Wrinkliness

One of the interesting features is the wrinkliness of the handwriting. One can copy the shape of another’s handwriting, although it was found to be difficult for many subjects. However, the speed and acceleration are more difficult to mimic. Hence, forged handwriting tends to be written slowly and, when scanned, is wrinklier than genuine handwriting. This feature can be measured using the fractal dimension measure [10], and we applied this feature as one of the features to detect forgery.

Scanned digital handwriting images are typically binary images and they are represented by a rectangular array called pixel whose value is either 0 or 1. Whereas the wrinkliness of the coastline was computed using two different measuring sticks [10], the wrinkliness of handwriting in binary digital images can be computed using two different resolution pixels. One can simply count the number of high and low resolution pixels on the boundary of handwriting and the formula for the wrinkliness is:

\[
\text{Wrinkliness} = \log\left(\frac{\text{boundary in high res.}}{\text{boundary in low res.}}\right) / \log(2)
\]

If the character is a smooth straight line in either horizontal or vertical directions, the wrinkliness value is 1.
Figure 2. Computing the Fractal Wrinkliness of handwriting in binary digital image format.

Figure 2 illustrates how to compute the fractal wrinkliness. First, a high resolution count is obtained by counting the number of pixels on the boundary of the character. Then a low resolution count is obtained by considering the adjacent four pixels as one large pixel and recounting the number of larger pixels on the boundary of the character. In the example of Figure 2, there are 69 small pixels and 32 large boundary pixels. The following measure is the wrinkliness of the above character: \[\text{Wrinkliness} = \frac{\log(69/32)}{\log(2)} = 1.1085.\]

3.3. Computing Handwriting Speed

The IBM TransNote digitizer recorded the x-y coordinates of the pen movement at a resolution of 100 points/centimeter (254 points/inch) and a sampling rate of 133Hz. From these online data the average speed (velocity) of the pen tip during the writing of each word was computed by dividing the length of the trajectory of the word by the time to write the word.
4. Experimental Results

In this section, we first show the result of our experiment to distinguish between genuine handwriting and forgery. Using the dichotomy model, we trained an artificial neural network. The dichotomy model transforms the feature space into a feature distance space and the forgery detection problem becomes a two-class classification problem. Finally, we show the correlation between wrinkliness and speed.

4.1. Automatic Forgery Detection System

As depicted in Figure 3, we generate two different class sets: one is the within-genuine handwriting distance set and the other is the between-genuine-and-forgery handwriting distance set.

**Figure 3.** Generating the Training and Testing sets.
set. The within-genuine handwriting distance set is collected by taking two natural handwriting samples from the same person. Features mentioned in the previous section are extracted and their corresponding distance measures are computed. In this manner we construct the d-dimensional within-genuine handwriting distance set.

Similarly, the between-genuine-and-forgery handwriting distance set is collected by taking a natural handwriting sample of subject x and a forgery sample of x attempted by a different subject y. These two sets are divided into training, validation, and testing sets to train and test an artificial neural network. Figure 4 illustrates our neural network dichotomy model.

Figure 4. Neural network using the dichotomy model to detect forgery.

We used a fully connected, feed forward, back-propagation neural network with the dichotomy model for training. There are eight input units (features), five hidden-layer units, and one output unit. After training, forged handwriting samples can be detected in an application system as
follows. Consider the case of a person presented with a document claimed to be in his/her handwriting but the person claims that it is a forgery. To verify whether the sample is genuine or a forgery, the person is asked to provide his/her handwriting sample. Then, the known sample and the sample in question are fed into the forgery detection system (Figure 4) and the system provides the result of “genuine” or “forgery.” On the test data of this experiment we obtained 9.6% and 11.2% for type I and type II errors, respectively. Type I errors occur when the system classifies genuine handwriting as forgery and vice versa for type II errors.

4.2. **Correlation between wrinkliness and speed**

A sample of the calculated measurements is shown in Table 1, listing the shorthand filename (e.g., 0101T1 means subject 1, copying subject 1, try 1), the number of pixels at 300dpi, the number of pixels at 600dpi, the wrinkliness, and the average pen tip speed. For each of the two hypotheses, we set the level of significance to 5% and applied the t-Test to the data to obtain the probability, p, of the null hypothesis – that the genuine and forged samples came from the same probability distribution.

<table>
<thead>
<tr>
<th>Filename</th>
<th>300 dpi Pixels</th>
<th>600 dpi Pixels</th>
<th>Wrinkliness</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0101T1 (genuine)</td>
<td>14894</td>
<td>30583</td>
<td>1.0380</td>
<td>0.1140</td>
</tr>
<tr>
<td>0201T1 (forged)</td>
<td>12182</td>
<td>25719</td>
<td>1.0781</td>
<td>0.0600</td>
</tr>
<tr>
<td>0301T1 (forged)</td>
<td>15127</td>
<td>32416</td>
<td>1.0996</td>
<td>0.0453</td>
</tr>
<tr>
<td>0401T1 (forged)</td>
<td>14839</td>
<td>31470</td>
<td>1.0846</td>
<td>0.0903</td>
</tr>
<tr>
<td>0501T1 (forged)</td>
<td>13114</td>
<td>27350</td>
<td>1.0604</td>
<td>0.0660</td>
</tr>
<tr>
<td>0601T1 (forged)</td>
<td>13026</td>
<td>27138</td>
<td>1.0589</td>
<td>0.0411</td>
</tr>
<tr>
<td>0701T1 (forged)</td>
<td>11747</td>
<td>24409</td>
<td>1.0551</td>
<td>0.0525</td>
</tr>
</tbody>
</table>
For the speed hypothesis we obtained the value $p = 0.0000000059$. This extremely low $p$ value clearly indicates that the writing speed of the forgeries was significantly slower than that of the genuine writings, strongly verifying the first hypothesis.

For the wrinkliness hypothesis we realized that many of the forged writings were obviously poor ones in terms of shape and size, indicating that several of the subjects either did not or could not reasonably mimic the genuine writing. It was usually easy to identify a “poor” forgery compared to a “good” one (Fig. 4)

![Figure 4](image1.png)

**Figure 4.** Genuine writing (left), “poor” forgery (center), “good” forgery (right).

Also, there were a few cases where one of a writer’s three genuine writing samples differed in writing style – for example, one written cursively and two printed (Fig. 5).

![Figure 5](image2.png)

**Figure 5.** Different genuine writing styles.

We determined by eye and manually eliminated the poor forgeries and the genuine samples written in a different style (about 25 percent of the data). For the remaining samples we found that the wrinkliness of the good forgeries was significantly greater than that of the genuine writings, $p = 0.0205$. Although not nearly as strong a verification as for the first hypothesis, the second
hypothesis was clearly verified, indicating that the wrinkliness of forged writing is significantly greater than that of genuine writing.

To further verify that the wrinkliness measure helps to identify forgeries we reran the forgery detection experiment without the wrinkliness measure. However, the improvement in forgery detection with the wrinkliness measure was not significant. The reason for this, we believe, is that the novice forgeries were easy to detect even without the wrinkliness measure. Also, the first six features, several of which are complex, contain a large amount of information and may overwhelm the wrinkliness feature.

5. Conclusion

In this paper, we developed and tested an automatic forgery detection system. From the experiments, we observed the ease of forging handwriting, and found that many subjects can successfully forge the handwriting of others in terms of shape and size by tracing the genuine handwriting. We also observed that even though the handwriting shape can easily be copied or traced by forgers, the exact dynamics of speed and acceleration are difficult to mimic. To capture a measure of writing speed, we proposed a measure of wrinkliness, and showed that forged handwriting tends to be wrinklier than natural handwriting.

We obtained experimental handwriting data from subjects writing samples in their natural style and writing forgeries of other subjects’ handwriting. These handwritings were digitally scanned and stored in an image database. Eight distance features were computed using various types of features including wrinkliness. A neural network was trained using a dichotomy model to
distinguish between genuine and forged handwriting, and we achieved 89% accuracy in detecting the forgeries.

In summary, using a fractal measure we found that the wrinkliness of good forgeries is significantly greater than that of genuine writings, corresponding to the writing speed of good forgeries being significantly slower than that of genuine writings. Also, … (SUNG, COMPLETE THIS SENTENCE DESCRIBING YOUR NEW RESULT). However, in the dichotomy system the forgery detection accuracy with the added wrinkliness measure (total of eight features) was about the same as that without the wrinkliness measure (seven features), probably because novice forgeries are easy to detect and the other features combined were stronger. Nevertheless, the overwhelming evidence clearly shows that the wrinkliness measure can be valuable in detecting forgeries in scanned documents.

Reference


