

## Forged Handwriting Detection

Hung-Chun Chen

### Abstract

Many important documents require signatures to verify the identity of the writer, but handwriting experts are often required to differentiate between authentic and forged signatures. For this reason it is important to build an objective system to identify forged handwriting, or at least to identify those handwritings that are likely to be forged. It seems reasonable that, with care, people can successfully forge handwriting in terms of shape and size by tracing authentic handwriting. Also, carefully tracing an authentic handwriting sample takes time, and the writing would tend to be written rather slowly. Therefore because we believe that most, or at least a large portion of, forged handwritings tend to be written slower than authentic handwriting, we hypothesize that it is also likely to be wrinklier (less smooth) than authentic handwriting. Although, with practice a forger might be able to increase his speed so that it might not be easy to detect the difference in wrinkliness, it is generally believed that it is difficult to forge the details of handwriting speed and acceleration, and that these properties tend to be unique to the individual. Therefore, the wrinkliness, speed, and acceleration of forged and authentic handwriting will be studied with the aim of developing an objective system to identify forged handwriting. These studies employ the IBM Transnote, pen-enabled notebook computer and associated SDK that provides the x and y coordinates as a function of time. By using this information the speed and acceleration of any portion of the handwriting sample can be calculated from the online handwriting. Also, by digitally scanning the collected handwriting samples in two different resolutions, a fractal number estimate of the wrinkliness of the writing can be calculated, and comparisons made to differentiate between the authentic and forged handwriting samples from scanned documents.

### Introduction

It is very important to verify the writership of a document, but it can hardly be done by human. So it is necessary to build a system that objectively identifies forged handwriting. Various automatic writer identifications on computer techniques have been studied. However, those models were based on one assumption that subjects provide their handwriting samples in their natural handwriting style, and the study did not cover forgery and disguised writing. So I start a study to find out the capability of machines to detect forgery. There are three stages in this study. *i)* handwriting samples collection, *ii)* feature extraction (wrinkles and speed), *iii)* statistical analysis. Subjects are asked to

write the test samples in their natural handwriting style in a pad three times and to forge other subjects' handwriting samples. These samples are scanned and stored digitally.

Although the shape might be copied, it is difficult to copy others' speed and acceleration. Forged handwriting tends to be written slowly in order to copy the shape perfectly, it might be wigglier than the authentic ones. I'll use the Fractal dimension measure on the wrinkliness feature. Then I'll analyze the speed to see if the forged samples were really done by slower speed. If it is not, then I'll compare the speed to check if it is possible to achieve the same speed and acceleration as the authentic samples'.

## **Database Construction**

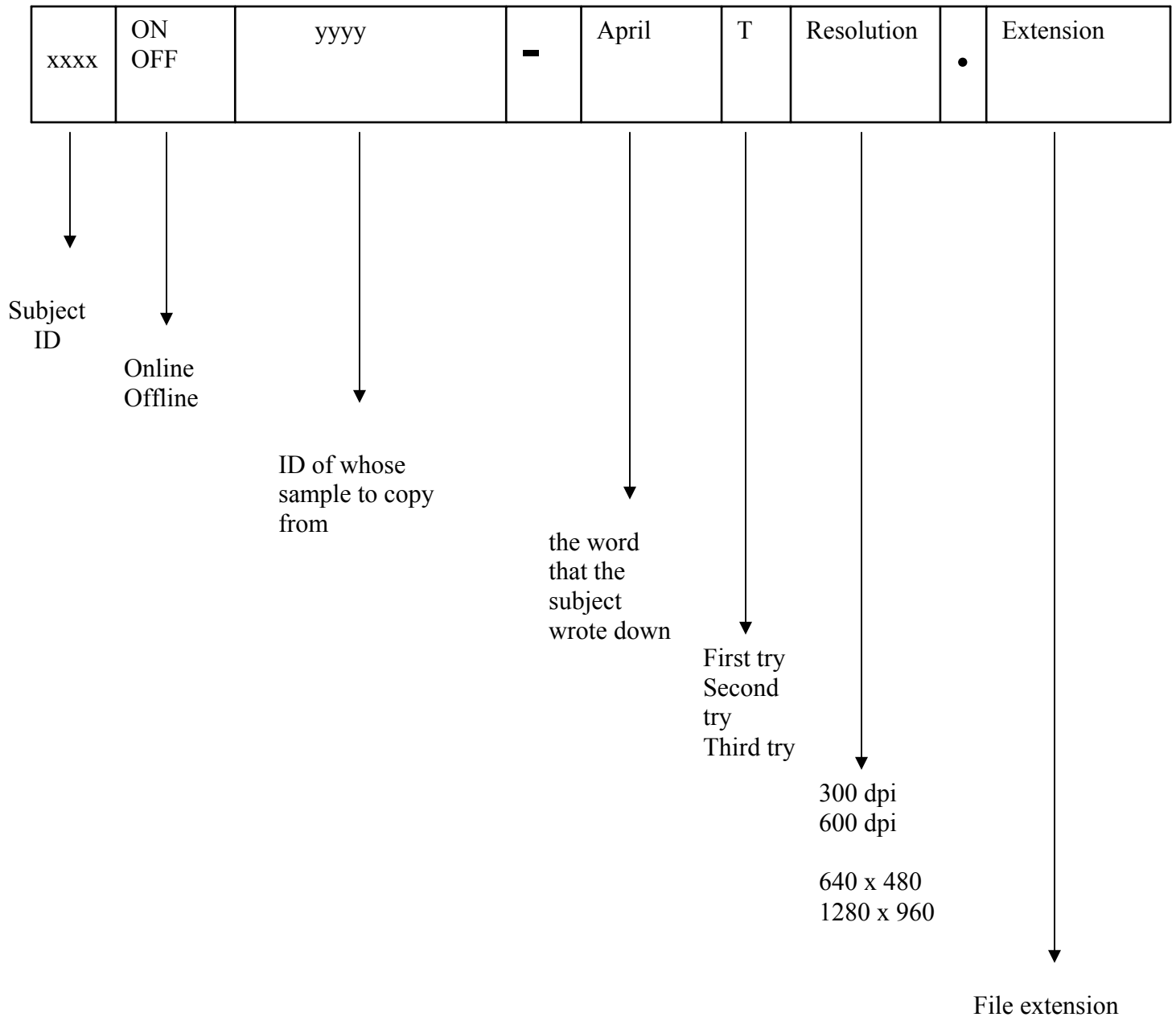
I invite ten people to be my subjects this experiment. Each subject is asked to write an authentic sample and to forge the other nine subjects' handwriting samples. Subjects actually wrote on a paper which is on top of the device called "Thinkscribe Digital NotePad". By the software installed inside the laptop, we can get the information about the speed. Thus, I separate them into two parts. The online parts are the data collected by the software and the off-line part is the images which are digitally scanned into the computer as a picture file or taken pictures by using the digital camera. Since the data will contain many different types, I find a way to name it in order to tell the differences between each data. I use seven parts in a filename to represent the subject ID, online or off line, ID of whose samples to copy from, the word that the subject wrote down, first, second or third try, the resolution, the extension.

The first indicator is the "Subject ID", that is, who writes this sample. The second indicator "ON" or "OFF" follow the subject ID are to separate the online and off-line data. "ON" means it is the data saved in the computer and we can use them to get the speed information. "OFF" means that we get the data by either scanning or taking pictures on the papers those subjects wrote down. The indicator right after the "ON" or "OFF" is the "ID of whose sample to copy from". If the subject ID is 0001 and he's trying to forge subject 0005's handwriting, then the first indicator would be 0001 and the third indicator would be 0005. If the subject is writing his authentic samples, then the first part and the third part would be the same. By looking at the first and the third indicator, we can tell that if this sample an authentic one or forged one. If the first indicator is identical to the third indicator, then it is an authentic handwriting belongs to that subject. Otherwise, it is a forged handwriting sample. The fourth part is the word that the subject actually writes down. I use just one word "April" this time, so this part will always be "April" in every data. The following part is that every subject must write each word three times, so the indicator shows the first try, second try and the third try by using "t1", "t2", and "t3". The sixth part is about the resolution. If the data is a scanned image, this part would be either 300 or 600. 300 means the resolution setting on the scanner is 300 dpi, 600 means 600 dpi. If the data is a picture taken by the digital camera, it would be 640x480 or 1280x960. The last part of the file name is the extension. It depends on which type of data it is. If it is

an image, then it would be jpg. If it is the online data directly from the computer, then it would be nbk.

For example, if subject number 0001 copy out “April” from subject number 0007’s sample on his first try, the image from the sample sheet is going to be scanned using 600 dpi resolution setting. Then the file name would be “0001OFFF0002-AprilT1600.jpg”.

<File>



**The process of collecting samples**

In order to get the information of some unique characteristics of one's handwriting, I invite ten people to be my subjects (*Figure 1*). I ask each of them to write down a chosen word "April" in their normal handwriting style. After collecting ten authentic handwriting samples, I ask them to forge other nine subjects' handwriting samples. For each subject, I have one authentic handwriting of his own and other nine forged samples.

In my hypothesis, I want to find out the relationship between "speed" and "wrinkliness", so I need a device that can record the position of the points in a certain sampling rate when the pen moves. The data is called the "online" information. I also need the word that the subjects actually wrote down on the paper to calculate the wrinkliness. I choose the IBM Transnote and the SDK they provided to collect the samples.

The machine (Transnote) has two parts. One part is the digital notepad and the other part is the computer. When you put a piece of paper on top of the digital notepad and write on it with the provided pen, it will transmit the image from the digital pad to the computer, so you can get the shape of the words. Meanwhile, you have the paper that has the handwriting samples on it in order to scan or take photo for further use. It also records the information of all positions (x, y coordinates) of the pen movements at the sampling rate of 100MHz.

I start from collecting each subject's authentic handwriting sample. Each subject is asked to write down the word "April" on the same paper three times. After finish this part, I give the others' handwriting samples to the subject and ask him or her to forge that sample. Because each authentic sample contains three words, I ask the subject to forge all three words. So the subject would forge the top most word appears in the authentic samples and write down on the top most position, and then the middle word, and the bottom word.

However, I notice that there are some problems collecting the samples. The first problem is that even the same person, his handwriting might show a significant difference between the three tries. It might be a problem later on when I want to find out the unique of everybody's handwriting. The second problem is that some subjects may stop at some point when writing the sample, it will cause a problem when we want to calculate the speed, and it might not represent the real speed information. The third problem is that some subjects might not be used to the provided pen because it is board. So the words he or she wrote down might be different from using other pens. We'll discuss this more in the next section.

The image displays three handwritten samples of the word "April" from a single writer. The top sample is written in a cursive style with a vertical orientation. The middle sample is written in a similar cursive style but is oriented horizontally. The bottom sample is written in a more fluid, cursive style, also oriented horizontally.

(a) Authentic handwriting samples from one writer

After this, I have the samples stored in the computer and also have the papers with their handwriting samples. To deal with the papers with the handwriting samples on it, I scanned them through the scanner in two different resolutions. I choose 300 dpi and 600dpi to be the two resolutions.

### **Computing Handwriting Wrinkliness**

The wrinkliness of the handwriting is an important feature to tell the difference between the authentic and forged handwriting. The wrinkliness can be measured using the *fractal* dimension measure. For example, in the paper “How long is the coastline of Great Britain?”, there is a suggestion of the measure of the wrinkliness of the coastline [1, 2]. Scanned handwriting images are binary images and are presented by the pixel of value 0 or 1. One can count the number of pixels on the boundary of handwriting in both low and high resolutions. The formula for the wrinkliness is:

$$\text{Wrinkliness} = \log(\text{high\_resolution} / \text{low\_resolution}) / \log(2)$$

high\_resolution is the number of pixels on the boundary of the higher resolution handwriting samples

low\_resolution is the number of pixels on the boundary of the lower resolution handwriting samples

I calculate the wrinkliness of each subject's handwriting samples and compare them with each other. After that I run a hypothesis test to check if the wrinkliness and speed is different from the authentic handwriting to the forged ones. The following is the result of the hypothesis test:

### Wrinkliness

t-Test: Two-Sample Assuming Equal Variances  
(Wrinkliness)

	<i>Forged</i>	<i>Authentic</i>
Mean	1.093926505	1.083367476
Variance	0.001337773	0.001038271
Observations	270	30
Pooled Variance	0.001308627	
Hypothesized Mean Difference	0	
df	298	
t Stat	1.516693798	
P(T<=t) one-tail	0.065202033	
t Critical one-tail	1.649982551	
P(T<=t) two-tail	0.130404067	
t Critical two-tail	1.967955541	

We can say that it's 93.5% probability that the wrinkliness of the forged handwriting is higher than the wrinkliness of the authentic ones. So we can tell that there are some differences between authentic and forged handwritings, and the wrinkliness value of the forged ones are higher than the authentic ones.

### Reference:

[1] B. B. Mandelbort, "How Long is the Coast of Great Britain, Statistical Self Similarity and Fractional Dimension," *Science*, 155, 1967, pp 636-638

[2] Sung-Hyuk Cha and Charles C. Tappert, "Automatic Detection of Handwriting Forgery," *Proc. 8th Int. Workshop Frontiers Handwriting Recognition (IWFHR-8)*, Niagara, Canada, August 2002, pp. 264-267.