

A Proposal for a Text-Indicated Writer Verification Method

Yasushi YAMAZAKI

Naohisa KOMATSU

Dept. of Electronics, Information and Communication Eng., WASEDA Univ.
Tokyo 169, Japan

Abstract

We propose an on-line writer verification method to improve the reliability of verifying a specific system user. In the proposed method, a different text including ordinary characters is used on every verification process. This text can be selected automatically by the verification system so as to reflect the specific writer's features. A specific writer is accepted only when the same text, which is indicated by the verification system, is written, and the system can verify the writer's personal features from the written text. The proposed method makes it more difficult to disguise writer himself with forged handwriting data than the previous methods using only signatures.

1 Introduction

We discuss an identity verification scheme based on handwriting information. Many study results on writer verification schemes are reported[1]. Two approaches are applied in the writer verification. One is based on static information (the result of handwriting) and the other is based on dynamic information (the action of writing). In this paper, we utilize dynamic information, considering the application of access control systems to allow specific users to enter important rooms or use computer resources.

Most of the recent research focus on signature verification especially in the field of on-line writer verification. In signature verification, the parameters which reflect the personal features are extracted from characters which are familiar to the writer. This characteristic of signature verification makes it possible to realize stable writer verification compared to the verification method without signatures.

However, signature verification has the serious problem of forged handwriting, because the same signature is used in both the enrolment process and the verification process. To solve the problem of forged handwriting, dynamic information such as velocity, acceleration, and force exerted on the pen are utilized[1].

Using ordinary characters is another measure to counter forged handwriting. It is hard to imitate the handwriting, because the verification text can be changed at every verification. Moreover, it is expected that the reliability of writer verification is improved by selecting characters which reflect personal features. Yoshimura, et al. described in their paper[2] that an on-line writer verification method using ordinary characters is effective to eliminate forgeries. However, the

definite on-line method has not been proposed yet.

Considering the above suggestions, we proposed an on-line writer verification method using ordinary characters, and evaluated its reliability[3]. However, in the proposed writer verification method, there is a problem that the reliability varies according to the kind or number of the characters. Namely, if the writer selects characters which do not contain personal features or the number of written characters is small, the extraction of personal features is not performed sufficiently. Furthermore, if the same text is used repeatedly in every writer verification, it is easy for a forger to attack with forged handwriting just as in the case of signature verification. Therefore, to carry out writer verification using ordinary characters, it is very important to know how to select a suitable text.

2 Text-indicated writer verification method

We propose a writer verification method which is based on a challenge and response type of authentication process to solve the problem described in the previous section. We call it a text-indicated writer verification method in this paper (see Figure 1).

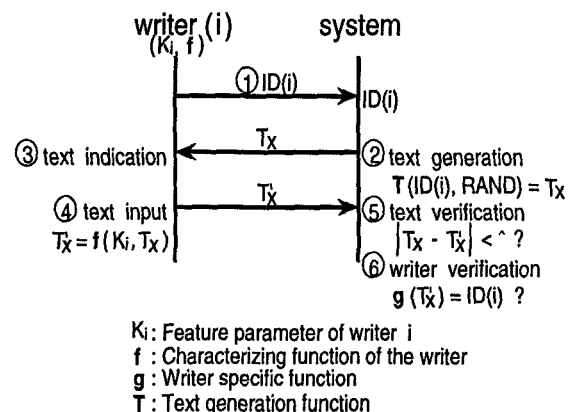


Figure 1: Text-indicated writer verification method

The verification process is as follows:

- 1 A writer sends his $ID(i)$ to the system.
- 2 The system generates a random number $RAND$ and a text T_x by using a text generation function

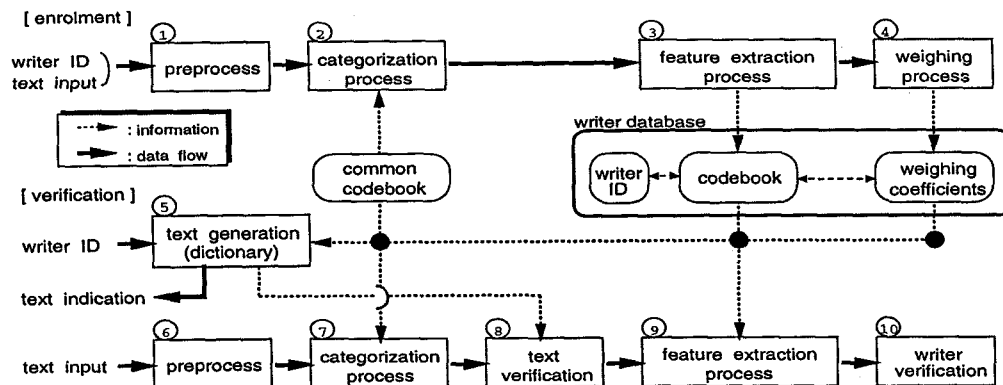


Figure 2: Enrolment and verification process of text-indicated writer verification

T with parameters RAND and $ID(i)$. The generated text is selected so as to reflect the feature of writer i .

- ③ The system indicates the generated text T_x to the writer.
- ④ The writer produces his handwritten version of the indicated text T_x . We define K_i as the feature parameter which represents the unique feature of writer i , and f as the function which gives the personal feature of handwriting. T'_x , the text written by writer i , is given as follows: $T'_x = f(K_i, T_x)$.
- ⑤ The system recognizes characters of the text T'_x and judges whether T_x and T'_x are the same. If they are the same, the writer verification process(⑥) is executed. Otherwise, the verification process terminates and rejects the writer as not being the specified writer.
- ⑥ The function g discriminates the writer by referring to the handwriting information of written text. The system judges whether the values of $g(T'_x)$ and $ID(i)$ are the same. If so, the claimed writer is accepted. Otherwise, the claimed writer is rejected.

The characteristics of the proposed method are as follows:

- The verification system selects the text automatically so as to reflect the specific writer's features. So the verification system can prepare enough characters which contain the specific writer's features at every verification instance.
- A writer is accepted only if he produces the handwritten version of the text indicated by the verification system.
- The indicated text is varied for every verification. This makes it difficult to attack with forged handwriting.
- In the verification process, the verification system can indicate characters which were not used in the enrolment process.

3 Enrolment and verification process of text-indicated writer verification

In this section, we describe the enrolment and verification processes of the text-indicated writer verification method. In the proposed processes, the feature extraction method based on categorized handwriting information, which was proposed in our previous work[3] is applied. The categorized handwriting information is used in the character recognition as well as in the writer verification. This is one of the unique points in our proposal. We also describe the reliability of the proposed method with simulation results. In the following subsections, we explain the enrolment and verification process (see Figure 2).

3.1 Enrolment process

3.1.1 Preprocess

Handwriting information is taken from the verification terminal and features are extracted (see Figure 2-①). These features should satisfy the following requirements: ①The variation of features of the same writer is small enough, while the differences between features of different writers are large. ②Text-independent; the identity verification system can verify whatever a writer writes. Referring to the above requirements, we use an internal angle between two lines as a feature (see Figure 3(a)). We define a set of one stroke (i.e. from pen down to pen up) and one inter-stroke (i.e. from pen up to pen down) as a feature extraction unit, where a set of features is extracted (see Figure 3(b)). The extracted features are transformed into a set of Fourier descriptors[4], which express the shape of the strokes in the frequency domain. We selected the P-type Fourier descriptor[5] which is applicable to open curves, and convenient for the description of handwriting. These P-type Fourier descriptors are then transformed into feature vectors, which are used in the next categorization process. As a feature vector, we use the lower frequency part of the P-type Fourier descriptor including n coefficients.

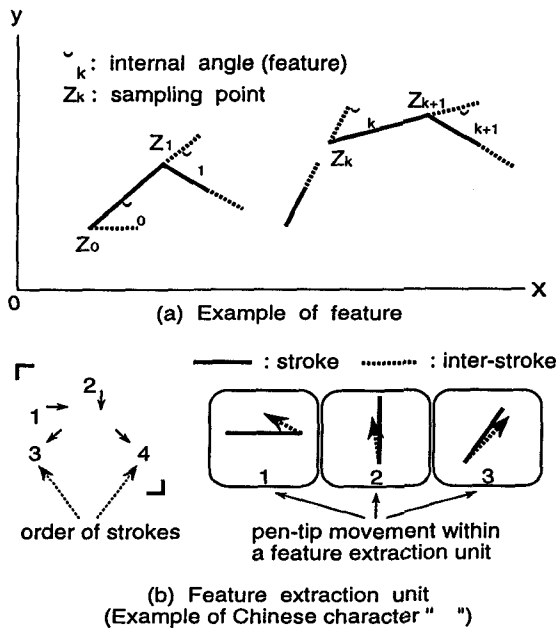


Figure 3: Feature and its extraction unit

3.1.2 Categorization process

The feature vectors are classified into several categories by referring to templates which are prepared in advance (see Figure 2-②). These templates contain personal features obtained from many writers. In the proposed method, such templates are obtained from feature vectors using an improved clustering algorithm[6] based on the LBG algorithm[7] which is a well known method in the design of vector quantizers. We define each reproduced vector in a codebook as a 'template'. The codebook is produced by the clustering algorithm and a set of feature vectors is used as a training sequence. In this categorization process, feature vectors are classified into categories by referring to templates using the vector quantization method. The codebook produced in this process is referred to as a 'common codebook' in this paper.

3.1.3 Feature extraction process

Personal features are extracted from the categorized feature vectors (see Figure 2-③). A learning algorithm is applied to obtain a relationship between the writers' IDs and their categorized feature vectors. In this paper, we call a set of feature vectors within a writer a 'class'. The learning algorithm has two functions; registration and discrimination. In registration, the algorithm repeats the renewal of reference vectors so as to match each feature vector with the correct writer of the correct class. We define a set of reference vectors in a category produced by the learning algorithm as a 'codebook' and regard it as the personal features of

writers. As a learning algorithm, we use the Learning Vector Quantization (LVQ)[8] algorithm.

3.1.4 Weighing process

To emphasize the personal features in each category, weighing coefficients are calculated according to the uniqueness of personal features (see Figure 2-④). Weighing coefficients are produced as follows:

$$W_{ij} = \frac{R_{ij}^{(i)} - \max_{1 \leq k \leq M, k \neq i} (R_{ij}^{(k)})}{\max_{1 \leq k \leq M} (R_{ij}^{(k)})} \quad (1)$$

M : number of writers.

$R_{ij}^{(k)}$: number of times i 's test data is accepted as k 's data in category j . The distance measure used here is the Euclidean distance.

W_{ij} shows writer i 's weighing coefficient in category j and it decides the discriminating power of the learning algorithm.

3.2 Verification process

3.2.1 Generation of text

To make a verification text, N kinds of characters are selected in the dictionary by referring to the weighing coefficients (see Figure 2-⑤). Each of the selected characters includes several categories which have sufficient personal features of the claimed writer. In this paper, the selected characters consist of several categories that satisfy following requirements;

1. The category that satisfies $W_{ij} \geq W_{th}$. W_{ij} is a weighing coefficient and W_{th} is a threshold of W_{ij} .
2. The category that satisfies $P_{ij} \geq P_{th}$. P_{ij} is the occurrence probability of category j in character i and P_{th} is a threshold of P_{ij} .

3.2.2 Preprocess and categorization process

Feature vectors are produced and classified into categories by referring to a common codebook (see Figure 2-⑥⑦).

3.2.3 Verification of text

A decision is made whether the writer has written the same characters as the system requested by performing character recognition of the written text (see Figure 2-⑧). It should be emphasized that the same feature vectors are used in this character recognition process as in the writer verification process. When the written text is different from the indicated text, the writer is rejected. Written text is compared with indicated text by a character using DP (Dynamic Programming) matching technique[9] based on the categories. When the following condition is satisfied, it is decided that the indicated text was correctly written by the writer; $D < D_{th}$, where D is the Euclidean

distance between two characters, one of which is in the written text and the other in the indicated text, and D_{th} is a threshold of D and is not dependent on characters.

3.2.4 Feature extraction process

Via the codebook, the corresponding writer is specified using a LVQ algorithm (see Figure 2-⑨). In each category, the reference vector in the codebook, which is the nearest to the produced feature vector, is selected. Then, the registered writer corresponding to the reference vector is selected. These processes are applied to each feature vector in a category. Therefore, each feature vector is finally assigned to a specific writer. We count the number of times each registered writer was selected as a candidate for the genuine writer in the above processes, and define it as an 'expected value'.

3.2.5 Verification of writer

The writer is verified based on the expected value defined above (see Figure 2-⑩). The largest value of an expected value is defined as w_1 , and the second largest value of an expected value is defined as w_2 . If the ID shown by the writer before the verification process agrees with the ID corresponding to w_1 , the verification process is continued, otherwise the writer is rejected and the process is terminated. Then, if w_1 and w_2 satisfy the following threshold condition, the writer is accepted, otherwise rejected; $w_2/w_1 \leq th$ ($0 \leq th \leq 1$).

4 Reliability test

In this section, the reliability of the proposed method is shown with simulation results.

In the proposed method, it is important that each category can represent many characteristics of handwriting efficiently without lack of personal features. We use a set of Chinese characters as the text in the reliability test, because Chinese characters seem to have many features of handwriting. In the reliability test, we use two kinds of text. One is the text for making a common codebook, and the other is the text for a reliability test. The former is called a 'codebook text' and the latter is called an 'experimental text' in this paper. The codebook text consists of 100 characters including seven types of fundamental elements of Chinese characters. 5 writers wrote 20 characters to make the codebook text. On the other hand, the experimental text consists of 100 characters (see Table 1). 13 writers who did not contribute to the codebook text wrote 20 characters, repeated 5 times to make this experimental text. We selected 45 characters from the former part of the experimental text and used them as training data. We used the remaining 33 characters of the experimental text as weighing data and 22 characters as test data. It is important that training data is different from test data.

The dictionary in the text generation process consists of 11 characters which are the same characters as the test data. These characters were written by another group of 10 writers. All writing data were

gathered at a spatial resolution of 10 points/mm and a sampling rate of 110 samples/s. The P-type Fourier descriptors are produced by the 128-point FFT, from which a 82-component feature vector is formed by grouping of the first 41 lowest order complex coefficients. Other parameters are decided as follows; $W_{th}=0$, $P_{th}=0$.

Table 1: Experimental text

Training data	早稲田大学理工学部 × (5 times)
Weighing data	電子通信学科小松研究室 × (3 times)
Test data	電子通信学科小松研究室 × (2 times)

4.1 Extracted categories

Figure 4 shows the parts of handwriting curves obtained from reproduction vectors in the common codebook. Each category reflects the part of strokes which contain personal features. And each label from A to M in the same figure shows the writers who are correctly recognized in the experiment, where the registered writer who has the nearest reference vector to the test data is identified as the writer of the test data. This result shows that personal features can be detected in the simplest elements of a character which consist of one stroke and one inter-stroke adjoining the next stroke. We can find that there are some categories which contain many personal features as compared with the others, such as categories no.1, no.8 and no.9. Of course, categories which reflect personal features are different for each person. This result suggests the effect of the weighing process.

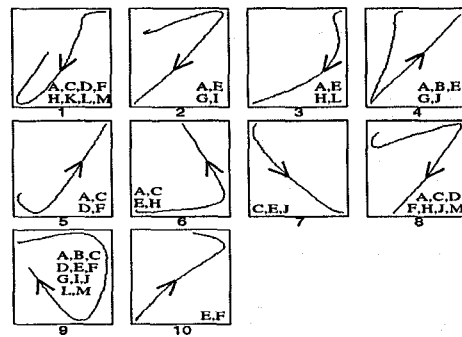


Figure 4: Handwriting curves extracted from common codebook

4.2 Results of writer verification

Figure 5 shows the FRR (False Rejection Rate or the rate of incorrectly rejecting the specific writer), and the FAR (False Acceptance Rate or the rate of accepting the wrong writer). In the reliability test, any writer writes twice the three different characters which are selected from the test data. The recognition system automatically selects the characters which contain the specific writer's features. The test result is shown as the curve named 'indicated' in the Figure 5.

