Handwriting Analysis for Writer Verification

Ameur BENSEFIA, Thierry PAQUET, Laurent HEUTTE
PSI Laboratory, FRE CNRS-2645
UFR des Sciences, University of Rouen
76821, Mont-Saint-Aignan, FRANCE
{ameur.bensefia, thierry.paquet, laurent.heutte}@univ-rouen.fr

Abstract

This communication deals with the writer verification task. This task consists in deciding whether two handwritten samples have been written by the same writer or not. Handwritings are first characterized by the graphemes that have been segmented by a segmentation procedure. Handwritten samples are then analysed according to two different procedures. Text samples are described in a feature space common to the two writers. The statistic of a mutual information criteria allows to build a robust hypothesis test. In the case of small samples of handwritings such as single words, the Levenstein distance is used to build a second hypothesis test. The two approaches are evaluated on a PSI database as well as the IAM database.

1. Introduction

As for any biometric-based identification applications (fingerprints, faces, voices, signatures…), forensic analysis of handwriting requires to query large databases of handwritten samples of known writers due to the large number of individuals to be considered. Therefore in order to come to a conclusion about the identity of an unknown writer, two tasks must be considered:

- the writer identification task concerns the retrieval of handwritten samples from a database using the handwritten sample under study as a graphical query [2]. It provides a subset of relevant candidate documents [9][6][11].
- the writer verification task, on its own, must come to a conclusion about two samples of handwriting and determines whether they are written by the same writer or not [15][12].

When dealing with large databases, the writer identification task can be viewed as a filtering step prior to the verification task. In this case, the verification consists in matching the unknown writer with each of those in the selected subset. Therefore the verification task can sometimes be adapted to each known reference writer based on the individual description of their handwriting. On the contrary, when the number of potential writers is too large even unknown or infinite, an individual description of each handwriting cannot be used. In this case one can for instance derive a specific set of feature differences to model the overall within and between writer distance (intra and inter writer variability) on a set of examples [12].

Most of the time the writer authentication task is carried out by a Forensic Handwriting Examiner (FDE) and is prone to an important subjectivity [7]. In any case, the confidence that can be associated to this kind of decision is not scientifically proven. Recent works have however proposed a scientific methodology of handwriting analysis for the writer verification task [1]. It should be noted that this task has been less studied than the identification task. This is undoubtedly due to the fact that verification implies a local process of decision-making which generally depends on the textual contents. Indeed, one generally has to compare the possible shapes of a character or a specific word that occur on the document under study. This is why the complete automation of this task does not seem to be very realistic because it would depend on automatic recognition.

The first stage in the design of a writer verification system requires to define a similarity measure between two handwritings. The second stage is a decision between two classes: within-writer and between-writer. To this end an hypothesis test appears to be useful by providing means too quantify the false acceptance and false rejection errors.

In this communication, we propose to deal with the writer verification task in 2 different ways - independently of the textual contents for samples of big size (bloc of texts) - according to the textual contents for small samples (handwritten words). The first part of this communication is devoted to the writer verification from text blocs. The approach rests on a mutual information criterion and a hypothesis test for the decision-making. Two databases are used to evaluate the method: *PSI_DataBase*, made up of 88 writers, was built at PSI and was used as a training base, and *IAM_DataBase* made up of 150 writers was used as a test base and part of it has been provided by the authors of [14]. In the second part of this paper we deal with the writer verification task based on the analysis of handwritten words. For this purpose we use a Levenstein distance between two handwritten chains [13]. Then we

Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004)
0-7695-2187-8/04 $20.00 © 2004 **IEEE**

built a hypothesis test to make the decision. The method is evaluated on a database of 100 words written by 20 different writers.

2. Writer verification from text blocs

For the purpose it is needed to define a set of relevant graphical features able to describe any handwriting. To circumvent this difficulty, the method rests on the adaptation of the feature set to the examined handwritings. We briefly point out the feature detection principle. For more precise details one can see [8]. The document is segmented into lines of text and each line is segmented into graphemes based on the analysis of upper contours of the connected components. This stage produces a set of elementary patterns which are either characters, parts of characters, or connected characters (figure 1). Following this segmentation step, handwritten texts can be described at the grapheme level, at the bi-gram and tri-gram level or even more.

The set of graphemes obtained on the two documents is submitted to a unsupervised classification phase. This stage makes it possible to define groups of similar patterns that are more or less frequent on the two documents and which will constitute the characteristics of the writer verification process. This method therefore doesn't depend on any set of pre-defined features and it is precisely for this reason that it is adaptable to any unknown handwritings. We now describe the main stages which lead to the decision.

Figure 1. Samples of graphemes of level 1 on the handwriting word “man”

2.1. Distance between text blocs

Assume that two handwritten documents D_1 and D_2 have been written by writers S_1 and S_2 respectively. Let us denote S the set of these two writers :

$$S = \{S_1, S_2\}$$

Define G the set of clusters common to the two analyzed documents and obtained by an unsupervised classification stage:

$$G = \{g_1, g_2, g_3 \ldots\ldots g_N\}$$

Some of these features can be present on the two documents, while others can appear specifically on one single document. Mutual information then allows to measure the dependence between the set of writers S and the set of features G. Low values of the mutual information indicate a strong independence between the two random variables while high values denote a strong dependence between them. Independence between S and G should indicate that the set of features G is distributed in the same way on the two documents and should reflect the same identity for the two writers S_1 and S_2. On the contrary, the mutual information criterion should allow to detect cases that exhibit a strong dependence between S and G and to reveal different identities of the two writers. We recall the expression of the mutual information between G and S:

$$I_M(G,S) = H(G) - H(G/S)$$

Where $H(G)$ denotes the Shannon entropy [10]:

$$H(G) = -\sum_{i=1}^{card(G)} P(g_i)\, H(G=g_i)$$

and $H(G/S)$ denotes the conditional entropy defined by:

$$H(G|S) = -\sum_{i=1}^{card(G)} \sum_{j=1}^{2} P(S_j)\, P(g_i|S_j)\ \log_2 \left[P(g_i|S_j) \right]$$

2.2. Evaluation of the mutual information criterion

To attest the relevance of this criterion we carried out two distinct series of tests on two different handwritten databases. The first one has been constituted at the PSI laboratory (*PSI_DataBase*) and contains 88 writers who have been asked to copy a letter (in French) that contains 107 words. The scanned images have been cut into two parts, in order to have two samples of each writer. The second database is part of the IAM database (*IAM_DataBase*) [14]. The fraction of this database that we have used contains texts written in English by 150 writers. Textual content varies from one writer to another. We chose to use the *PSI_DataBase* as the training database, and the *IAM_DataBase* as the test database.

Figure 2 gives the distribution of the mutual information criterion in the two following cases : figure 2.a gives the distribution of the mutual information criterion in the case where the two writers are identical, while figure 2.b gives the distribution of the criterion in the case where the two writers are different. From the observation of these two distributions it seems clear that mutual information should provide a quantitative criterion for the writer verification task. Furthermore, this figure shows that these two distributions can be approximated using a normal distribution.

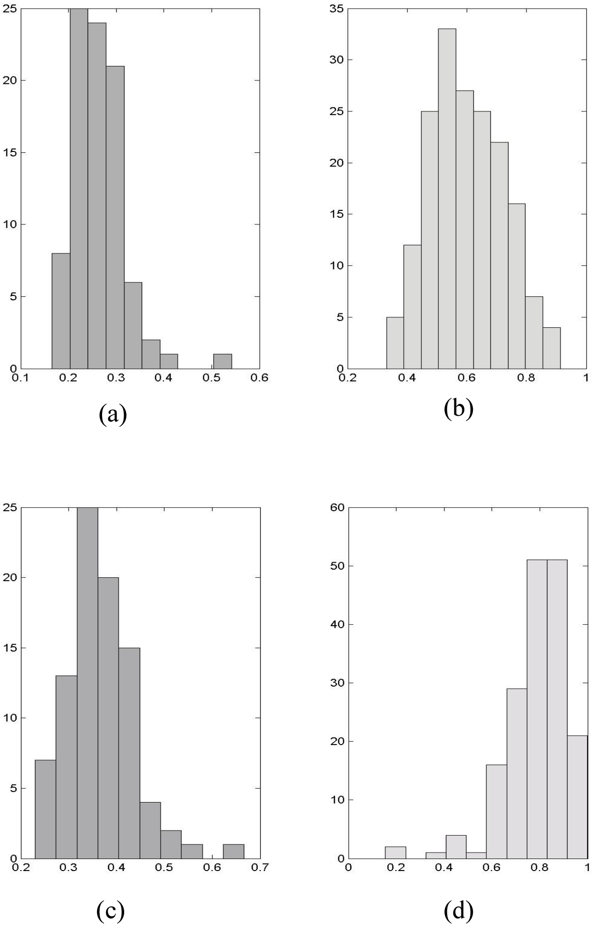

Figure 2. Mutual information criterion distributions on the PSI database, with the graphemes of level 1 (a, b), with the graphemes of level 2 (c, d), in the intra (a, c) and in the inter case (b, d)

2.3. Decision

The writer verification task is a decision process between two classes : within and between-writer. The construction of the best decision leads us to model the statistical distribution of the decision criterion for the two considered classes. The modeling of the writer verification problem corresponds to the modeling of the two following hypothesis :

- H_0 : The two handwritings come from the same hand
- H_1 : The two handwritings come from different hands

We can also model the two hypothesis under a mathematical form. This can be accomplished using classical hypothesis test [10]. H_0 will serve as the null hypothesis or the default hypothesis :

$$H_0 : S_1 = S_2 \text{ and } H_1 : S_1 \neq S_2$$

Each of the two possible decisions is associated to a probability of correct and false decision. Probability of error on the null hypothesis is the Type I error and is denoted α, while probability of error on H_1 is the Type II error and is denoted β. Table 1 summarizes the possible situations.

Decision \ Truth	H_0 is true	H_1 is true
Accept H_0	$1-\alpha$	β
Accept H_1	α	$1-\beta$

Table 1. Associated probabilities to the different decisions

2.4. Statistical test

Several statistical tests of very different designs are available. The test which provides the smallest error β, for the same value of α, is by definition most powerful (which has the greatest power value $1-\beta$). Indeed, it can detect the smallest differences between the populations without increasing the Type I error. The choice of the more adapted statistical test to the hypothesis is a central aspect of statistics. Each statistic of a test has a probability distribution. If the standard deviation of the population is unknown, which is generally the case, then the most suitable statistics test are founded on a t-student distribution. This law is often used when the number of individuals in the population is lower than 30. Beyond this value the t-Student law converges towards a normal law. This is why we assessed our writer verification system using a normal law (figure 3).

Assuming normal distribution of the mutual information criterion for the two hypothesis, it is very simple to quantify the Type I and Type II errors. Using these distributions and choosing a value for the Type I error, we can define the rejection and acceptance regions of the null hypothesis and deduce the experimental value of β.

The area of rejection of H_0, noted W_0, is defined by the first order error. The limit of this area allows to define the rejection area of H_1, denoted W_1 and to deduce the second order error by the following relations:

$$P(W_0 \mid H_0) = \alpha \quad \text{and} \quad P(W_1 \mid H_1) = \beta$$

In the same way, one determines the acceptance regions of the two hypothesis, $\overline{W}_0$ for H_0 and $\overline{W}_1$ for H_1. We have:

$$P(\overline{W_0} \mid H_0) = 1\text{-}\alpha \quad \text{and} \quad P(\overline{W_1} \mid H_1) = 1 - \beta$$

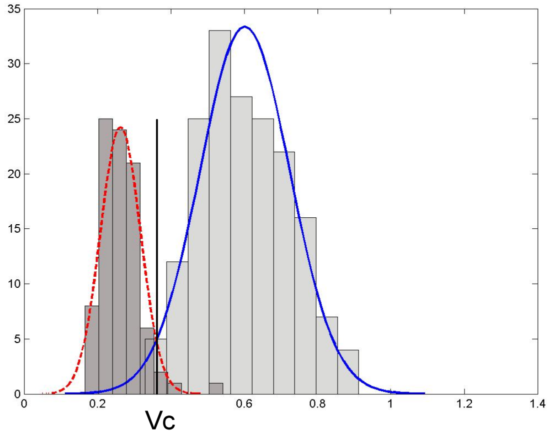

Figure 3. Adjustment of the within and between-writer distances distributions with a normal law on *PSI_DataBase*, using graphemes of level 1.

2.5. Results

Table 2 shows the various values of α, 1-β and the decision threshold Vc which will be associated to these 2 errors on the *PSI_DataBase*. Knowing the distribution of the H_1 hypothesis, we can choose the threshold α theoretically according to the theory of the Bayesian laws [4] : the minimization of the two areas of errors is the intersection point of the two distributions.

PSI_DataBase	α	1–β	*Vc*
Graphemes of level 1	3.5%	97.5%	0.3616
Graphemes of level 2	3.4%	98.5%	0.4984

Table 2. Summary of the various values of α, 1-β and Vc calculated on *PSI_DataBase*.

As we indicated previously, the *IAM_DataBase* was used for the evaluation of our writer verification system. The test consists in measuring the within-and between-writers similarities in order to validate the decision rule. The results obtained are summarized in table 3 below.

IAM_DataBase	Error I	Error II
Graphemes of level 1	4%	4%
Graphemes of level 2	10.66%	3.33%

Table 3. Evaluation of the first and second order error on the *IAM_DataBase*.

2.6. Discussion

The proposed approach for the writer verification task, the results seem particularly promising for several reasons. First of all the choice of a local representation based on the segmented graphemes seems very relevant since it allows a level of description which is close to characters without however requiring a recognition stage. In addition, it is remarkable to obtain similar performances on the IAM database than on the PSI database on which the hypothesis test was learnt. We are able, thus, to bring relevant quantitative elements for the handwriting individuality assumption. We show moreover here that it is possible to build a robust statistical test on several databases of handwritings. It will naturally be necessary to validate the approach on more consequent databases of documents.

3. Verification from handwriting words

To tackle the writer verification task based on the analysis of handwritten words, we have decided to build the analysis based on the textual contents in order to overcome the lack of data in this case. Let us recall that the Forensic document Examiners (FDE), during their analysis of documents verification, compare features extracted on the same characters or on the same words from the two documents [7][5]. When using text blocs we succeeded in providing text independent writer verification test because in this situation a sufficent amount of data (graphemes) was available.
The approach that we propose now, for the writer verification from handwritten words, will be dependant of the analyzed word but will not require the knowledge of the characters within the analyzed word. Indeed, we only consider situations where we compare similar entries of the lexicon. In this case, the verification task is closely related to the problem of words comparison like signature verification [3], or to some approaches of cursive handwriting recognition. Inspired from this studies we

propose a new criterion based on the estimation of a distance between two handwritten words.

3.1 Distance between handwritten chains

Our writer verification problem can be considered as a problem of distance calculation between 2 handwritten chains, since we are able to segment words into graphemes by using segmentation process. One can then formulates the problem in the following way:

let us consider two segmented handwritten words X and :

$$X = x_0\, x_1 \,..\, x_n \qquad\qquad Y = y_0\, y_1 \,..\, y_m$$

where the x_i, y_j represent respectively, graphemes of the chain X and the chain Y. We define the transformation cost of a grapheme x_i into a grapheme y_j by a similarity measure between graphemes (in our case the correlation measure). Then the distance calculation algorithm between handwritten words is simply expressed by the following algorithm :

$D(0, 0) := sim(x_0, y_0)$
For i :=1 to n Do
$D(i, 0) := D(i-1, 0) + sim(x_i, y_0)$
For j :=1 to m Do
$D(0, j) := D(0, j-1) + sim(x_0, y_j)$
For i :=1 to n Do
For j := 1 to m Do
$M_1 := D(i-1, j-1) + sim(x_i, y_j)$
$M_1 := D(i-1, j) + sim(x_i, y_j)$
$M_1 := D(i, j-1) + sim(x_i, y_j)$
$D(i, j) := Max(M_1, M_2, M_3)$
$Distance(X, Y) = D(n, m)/ (n+m)$

where $sim(x_i, y_j)$ is the correlation measure between the grapheme x_i and the grapheme y_j

By making this choice, maximum values of the transformation cost of X into Y corresponds to a low variability between the two chains, and is a good crietrion for considering that the two analysed words come from the same hand. Conversely, low values of the final transformation cost will be significant of an important difference between the two writers of the two analyzed words.

3.2 Experimentation

For this new experiment a new handwritten word database was built. We asked 20 different writers to copy 5 occurrences of the French word "manuscrit". Some examples are presented on figure 4. We used 60 intra-writer distances and 60 inter-writer distances, to estimate the corresponding distributions that we also modeled with normal laws.

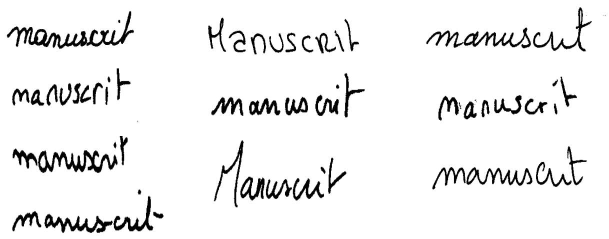

Figure 4. Some samples of our handwritten word database.

We adopted the same process as that adopted for the writer verification from text blocs and we formalized the problem according to two hypothesis H_0 and H_1 and determined the two corresponding errors. The adjustments carried out on the distributions have allowed us to determine the rejection and acceptation regions according to a threshold decision Vc (figure 5). We can notice overlapping between the two distributions, which leads naturally to a more important risk of error than in the preceding case. We can estimate the power of the test to 78% and the first order error to 15.1%. These results seem completely natural due to the small amount of information used to build this second test.

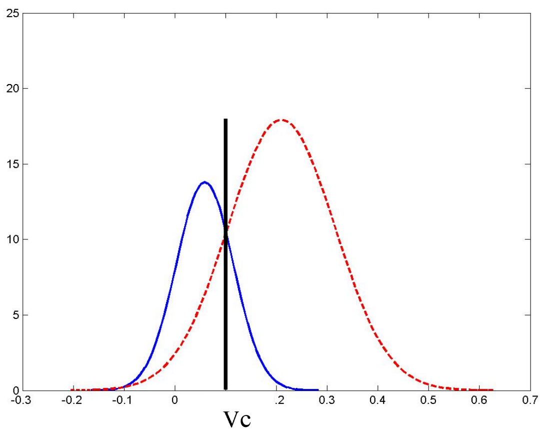

Figure 5. Inter and intra-writer distribution distances of the handwritten words.

4. Conclusion

In this communication we have investigated the writer verification problems. It constitutes a complementary and necessary approach for any identification approach, because it offers the only mean of individual rejection

(writer) when one writer is unknown by the system. We were interested in the writer verification from text blocs, independently of their textual contents, and in the writer verification from handwritten words, where we tried to avoid the use of characters location. In both cases, we built a discriminating criterion and a decision rule based on a hypothesis test. This allowed us to quantify the two kinds of errors (false acceptance and false rejection).
The verification performances obtained from text blocs analysis was definitely better than those obtained from the handwritten word analysis, what seems completely natural. We showed that it is possible to authenticate the handwritings of two documents without needing to know beforehand its properties. The tests were carried out on a basis of 150 writers unknown by the system and have allows us to obtain 4% error of false acceptance. Results obtained for the verification on handwritten words are about 15% of false acceptance what is completely honorable without using the characters knowledge in the words.

These results highlight the capacity to characterize the writings from the elementary patterns which compose them and which can be extracted thanks to a segmentation algorithm. These characteristics allows to automate the writer verification task without introducing additional knowledge, generally used by the Forensic Document Examiner (knowledge on the words or the characters). The performances obtained must still be validated on more constraining verification tasks. We think naturally to the detection of forgery. But in the current state, the approach makes it possible right now to supplement the results which we obtained for the writer identification task [2], by providing a rejection method of unknown writers. Finally this study highlights the capacity of the handwriting to constitute an interesting biometric characteristic for certain cases of use.

5. References

[1] S.H. Cha, S. Srihari, "Multiple Feature Integration for Writer Verification", IWFHR'7, pp 333-342, 2000.

[2] A. Bensefia, T. Paquet, L. Heutte, "Information Retrieval Based Writer Identification", ICDAR'03, pp. 946-950, 2003.

[3] L. Bovino, S. Impedovo, G. Pirlo, L. Sarcinella, Mutli-Expert Verification of Hand-Written Signatures, ICDAR'03, Vol II, pp. 932-936, 2003.

[4] Duda R., Stork D., Hart P., Pattern Classification and Scene Analysis, Wiley & Sons, 2nd Edition, 2000.

[5] K.H. Jarman, R. C. Hanlen, P. A. Manzolillo, Handwriting Examination : Moving From Art to Science, Pacific Northwest National Laboratory, 1999.

[6] U.V. Marti, R. Messerli, H. Bunke, "Writer Identification Using Text Line Based Features", Proc. ICDAR'01, pp 101-105, 2001.

[7] R.N. Morris, Forensic Handwritting Identification. Academic Press, 2000.

[8] A. Nosary, ''Automatic recognition of handwritten texts through writer adaptation'', PhD dissertation (in french), University of Rouen, France, 2002.

[9] H.E.S. Said, T.N Tan, K.D. Baker, "Personal Identification Based on Handwriting", Pattern Recognition, vol. 33; pp 149-160, 2000.

[10] G. Saporta, Probabilités analyse des données et statistiques. *Edition Technip*. pp 317-330. 1990.

[11] L. Schomaker, M. Bulacu, M. Van Erp, Sparse-Parametric Writer Identification Using Heterogeneous Feature Groups, *Proceeding of IEEE International Conference on Image Processing ICIP*, Vol. 1, pp. 545-548, 2003.

[12] S. Srihari, S.H.Cha, H. Arora, S. Lee, "Individuality of Handwritig : A Validity Study", ICDAR'01, pp 106-109, 2001.

[13] R. A. Wagner, M.J. Fisher, The String to String Correction Problem, JACM vol, 21, N° 1, pp 168-173. 1974.

[14] M. Zimmermann, H. Bunke, " Automatic Segmentation of the IAM Off-line Handwritten {English} Text Database". 16th International Conference on Pattern Recognition, Canada, Vol 4, pp. 35-39, 2002.

[15] E.N. Zois, V. Anastassopoulos, "Morphological Waveform Coding for Writer Identification", Pattern Recognition, vol. 33, n°3, pp 385-398, 2000.