An Off-Line Signature Verification Method based on the Questioned Document Expert's Approach and a Neural Network Classifier

CESAR SANTOS[1] EDSON J. R. JUSTINO[1] FLÁVIO BORTOLOZZI[1] ROBERT SABOURIN[2]

[1] PUCPR - Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição, 1155, Curitiba, PR, Brazil
cesar.roberto@pucpr.br
{justino, fborto}@ppgia.pucpr.br
[2] ÉTS - École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal, Québec, Canada
Robert.Sabourin@etsmtl.ca

Abstract

In an off-line signature verification method based on personal models, an important issue is the number of genuine samples required to train the writer's model. In a real application, we are usually quite limited in the number of samples we can use for training (4 to 6). Classifiers like the Neural Network [5], the Hidden Markov Model [2] and the Support Vector Machine [1] need a substantial number of samples to produce a robust model in the training phase. This paper reports on a global method based on only two classes of models, the genuine signature and the forgery. The main objective of this method is to reduce the number of signature samples required by each writer in the training phase. For this purpose, a set of graphometric features and a neural network (NN) classifier are used.

Keywords: Signature verification, Expert's classifier, Neural network.

1. Introduction

Usually, two different pattern classes make up an off-line signature verification method in training and verification tasks *(W_1* and *W_2)* based on personal models. W_1 represents a genuine signature set for a specific writer and W_2 represents a forged signature set. In the latter case, the set of forgeries is divided into three different types (random, simple and simulated) [2,8]. The random forgery is usually a genuine signature sample belonging to a different writer, one who is not necessarily enroled in the signature verification system. The simple forgery is a signature sample with the same shape as the genuine writer's signature, while the simulated forgery is a reasonable imitation of the genuine signature model.

The training phase uses a set of genuine signature samples *(W_1)* to produce a robust personal model. Usually, a meaningful number of samples capable of representing personal variability make up this set (see Fig. 1). The verification phase uses a personal model to discriminate among writers and among all types of forgeries *(W_2)*.

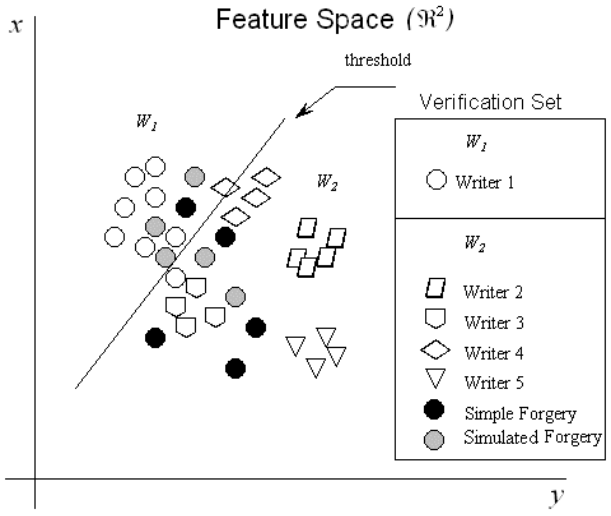

Figure 1. The signature models area for different authors in an off-line signature verification method based on personal models

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

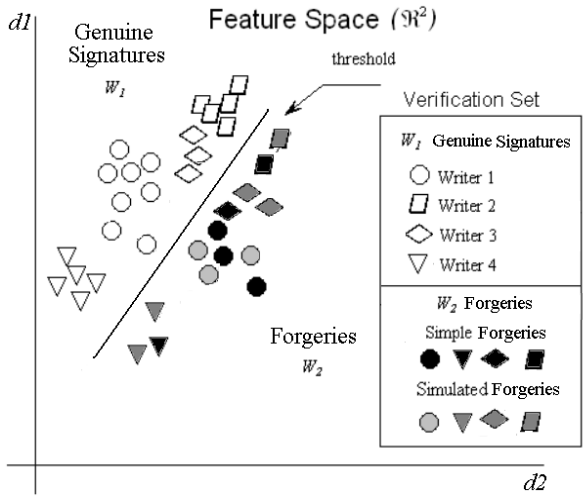

Figure 2. The general models in an off-line signature verification method based on the questioned document expert's approach

2. Questioned Document Expert's Approach

An expert on questioned documents classifies a signature sample in terms of authenticity, as either genuine or non genuine [4,11]. A non genuine or forged signature can be represented as a random, simple or simulated forgery.

Consider the problem where there are two categories of classes, genuine signatures and forged signatures [10]. The expert's approach places a questioned sample in one of these two classes to establish its authenticity. The expert uses a set of n genuine signature samples $S_{ki\ (i=1,2,3,\ldots,n)}$ against which to compare the questioned sample S_Q, and observes, based on F distinct graphometric features $f_{VKi,j(j=1,2,3,\ldots,n)(j=1,2,3,\ldots,F)}$ and $f_{VQj(j=1,2,3,\ldots,F)}$, differences in measurement between the genuine and the questioned samples $D_{i\ (i=1,2,3,\ldots,n)}$. The expert then takes "partial" decisions $R_{i\ (i=1,2,3,\ldots,n)}$ based on these comparisons, his or her final decision report D depending on the sum of the partial decisions obtained (see Fig. 3).

3. Signature Database

A set of signatures from 240 writers makes up the signature database (40 samples per writer), which was subdivided into two parts. The first part, composed of 180 writers, was used in the training and validation procedures. In the training procedure, 180 writers were considered, with 4 samples per writer (720 genuine samples). In the validation procedure, a subset of 40 writers was used. In this case, another 4 samples per writer (160 genuine samples) were used. The second part, composed of 60 writers, was used in the testing phase and as a reference database. A set of 5 samples per writer was used in testing (300 genuine samples), with a subset of 5 genuine samples as a reference (300 genuine samples). The testing database was expanded by the addition of forgeries; specifically, a set of 5 simple forgeries and 5 simulated forgeries for each writer (600 forgery samples).

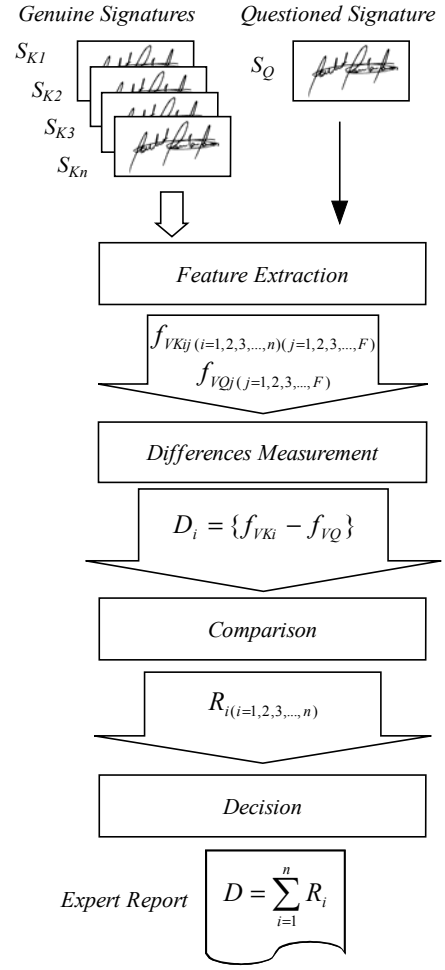

Figure 3. The off-line signature verification scheme based on the expert's approach

4. System Outline

The proposed system is based on the expert's method shown in Fig. 3. The signature images were subjected to a set of phases described in the sequence.

Grid-segmentation procedures have been used extensively in the off-line signature verification approach [1,2,6,7,9], these references also demonstrating how a

grid approach can be adapted to compute graphometric features.

The signature image is composed of a rectangle of 400x1000 pixels, with 300 dpi and 256 gray levels. Before grid segmentation, the image is moved? to the left in order to absorb the horizontal variability [2]. Then, a grid is put over the image area. A set of different grid resolutions was used in the experiments, but a grid with square cells of medium resolution (50x50 pixels) showed better results (see Fig. 4).

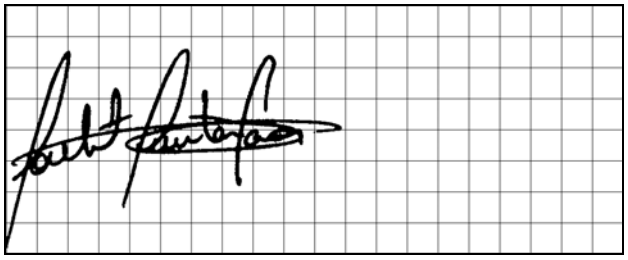

Figure 4. The grid-segmentation example, using 50x50 pixels (8x20 cells)

5. Feature Extraction

Graphometric studies have demonstrated the set of specific features used by questioned document experts [1,2]. Based on this set, a subset of static and pseudodynamic computational features was used [1,2] (see Table 1).

Table 1. Graphometric and computational feature relations

Graphometric Features	Computational Feature
Caliber (Static)	Space occupation
Proportionality (Static)	Space occupation
White spaces (Static)	Space occupation
Base behavior (Static)	Space occupation
Apparent pressure (Pseudodynamic)	Pressure area
Curvature (Pseudodynamic)	Stroke curvature
Progression (Pseudodynamic)	Stroke regularity

Apart from space occupation analysis, grid segmentation makes it possible to absorb a set of static features like caliber, proportionality, white spaces around the signature traces and base behavior [2]. The pseudodynamic features were computed using the signature traces. The pressure areas are represented by trace width variability and material deposition in a given area of the trace. This feature is dependent on the writing instrument and the type of paper used (see Fig. 5b). The pressure area is computed by the gray-level average inside the cell. The stroke curvature describes the hand velocity (see Fig. 5c). This feature is computed by finding the most significant signature segment inside the cell using a signature skeleton image. Then, the angle variability is computed using a chain code [3]. Stroke regularity corresponds to the writer's ability: the signature traces of a capable writer will show characteristics of firmness and speed [2], while the traces of a writer lacking ability will show distortion and irregularity [7] (see Fig. 5d). This feature is computed by finding the biggest stroke inside the cell, computing the variability of the pixel coordinates using a chain code, and then normalizing by stroke length [3].

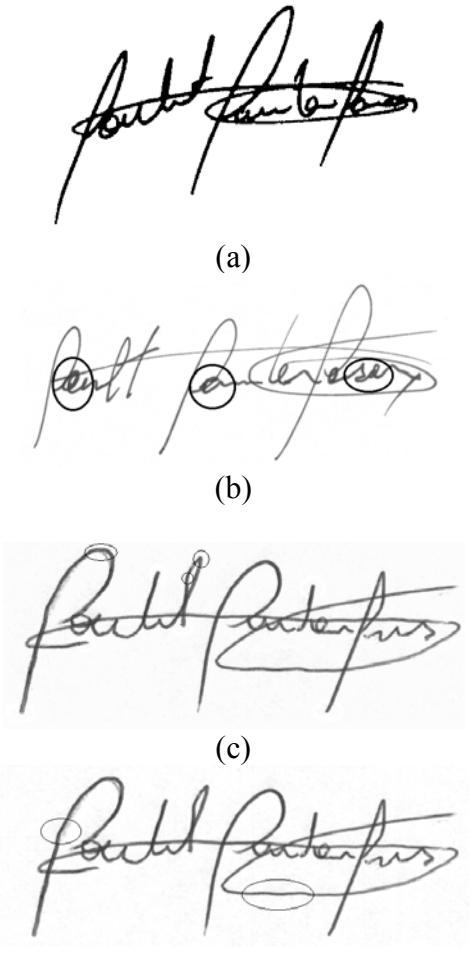

Figure 5. (a) Genuine signature; and forgeries with (b) Pressure areas; (c) Stroke curvature; (d) Stroke regularity

6. Feature Distance Measurement

Cha [4] proposed a method to establish individuality in automatic handwriting verification using distance measurement in a dichotomization process. We have used this method to produce a feature vector to discriminate genuine signatures from forgeries in the questioned document expert's method.

In the signature verification case, all databases were converted into a set of feature vectors. These vectors are obtained from both the S_{Ki} and S_Q signatures (Eqs. 1 and 2).

$$fv^F_{Ki_{(i=1,2,\ldots,n)}} = (f_1^F, f_2^F, \cdots f_m^F) \quad (1)$$

$$fv^F_Q = (f_1^F, f_2^F, \cdots f_m^F) \quad (2)$$

where $F(1 \leq F \leq 4)$ represent the feature subsets and m represents the maximum number of cells (160).

The Euclidian distance vectors $D_{fi\ (i=1,2,3,\ldots,n)}$ between genuine feature vectors and between questioned feature vectors are computed to obtain the neural network input in the training, validation and verification phases (Eq. 3) [4].

$$D_{f\,i_{(i=1,2,\ldots,n)}} = \bigcup_{F=1}^{4} \sqrt{(fv^F_{Ki} - fv^F_Q)^2} \quad (3)$$

7. Comparison

There are two stages in the comparison phase, training and verification. In training, the feature distances $D_{fi\ (i=1,2,3,\ldots,n)}$ are computed using a pair of signature samples. If the signature samples have been written by the same person, the feature vector is set to 1 (authorship). If the writers were different (random forgery), the feature vector is set to 0 (no authorship). Using the hypothesis that all distances between two signature samples produced by one writer are small, then the NN is trained to pick up small feature distances (genuine signature) and large feature distances (forgery).

In verification, the NN has two outputs, one to indicate that two signature samples were produced by the same writer W_1 and the other to indicate that two signature samples were produced by different writers W_2. In this case, it is possible to have genuine, random, simple and simulated forgeries like the questioned signature sample, and one of the other five genuine samples like the reference. When samples by the same writer are identified as samples by different writers, a type I error has occurred. When two different writers write two signature samples and they are identified as having been produced by the same writer, a type II error has occurred.

The MLP topology used is composed of an input layer with 640 neurons, changing from 4 to 16 neurons in the hidden layer, and an output layer composed of 2 neurons.

8. Decision

Usually a set of comparisons is performed in the expert's procedure. In this case, each known genuine signature sample (reference) is compared to the questioned signature sample. For this purpose, a small set of genuine signatures is used (4 to 10 samples) [2].

In our experiment, a set of 5 genuine signature samples per writer (second database) was used as a reference database. To produce a final decision, the proposed system combines all classifier outputs in a majority vote. This last stage represents the expert's procedure described previously (see Fig. 3).

9. Experimental Results

Table 2 shows the results obtained using the second database. The experiments have shown promising results in terms of general error rate. The high rejection level (a type I error rate of around 10%) was produced by the general model's inability to absorb intrapersonal variability. This inability was caused by too small a number of writers being used in the training phase. The simulated forgery acceptance rate was high because the model was not prepared, during the training phase, to identify this type of forgery.

Table 2. Experimental results obtained using the second signature database

Majority Vote	Type I Error (%)	Type II Error (%)			Total Error (%)
		Random	Simple	Simulated	
Comp. Features Set	10.33	4.41	1.67	15.67	8.02

10. Conclusion

The main purpose of this work is to report on a robust off-line signature verification method based on the questioned document expert's methodology. Two important advantages emerge. The first is the potential of the general method to reduce the number of signature

samples required for training and validation. The second is the model's ability to absorb new writers without generating new personal models. In terms of error rate, the results shown in Table 2 are promising, especially in the case of simple and random forgeries. It is possible to note the NN's ability to classify different types of forgeries (random, simple and simulated) without previous knowledge of the simple and simulated forgeries.

11. References

[1] Justino, E. J. R., Bortolozzi, F., Sabourin R., "An Off-line Signature Verification Method Based on SVM Classifier and Graphometric Features," The 5th International Conference on Advances in Pattern Recognition, 2003, Calcutta, 2003.

[2] Justino, E. J. R., Bortolozzi, F., Sabourin R., "Off-Line Signature Verification Using HMM for Random, Simple and Skilled Forgeries," ICDAR 2001, International Conference on Document Analysis and Recognition, Seattle, USA, v.1, pp. 105-110, 2001.

[3] Justino, E. J. R., "O Grafismo e os Modelos Escondidos de Markov na Verificação Automática de Assinaturas", Doctor Theses, Pontifícia Universidade Católica do Paraná, Brazil, p. 127, 2001.

[4] Cha, S., "Use of Distance Measures in Handwriting Analysis," Doctoral thesis. State University of New York at Buffalo, p. 231, USA, 2001.

[5] Baltzakis, H., Papamarkos N., A New Signature Verification Technique Based on a Two-Stage Neural Network Classifier, Engineering Applications of Artificial Intelligence, No.14, 2001.

[6] Yingyong Q., Hunt B. R., Signature Verification Using Global and Grid Features, Pattern Recognition – vol. 22, no. 12, Great Britain, pp. 95-103, 1994.

[7] Huang, K., Yan, H., Off-line Signature Verification Based on Geometric Feature Extraction and Neural Network Classification, Pattern Recognition, Vol.30, No.1, pp.9-17,1997.

[8] Chuang P. C., "Machine Verification of Handwritten Signature Image," Proc. Int. Conf. On Crime Countermeasures-Sci, J.S. Jackson and R. W. De Vore, University of Kentucky, Lexington, pp.105-109,1977.

[9] Sabourin R., Genest G., "An Extended-Shadow-Code Based Approach for Off-Line SignatureVerification," 12th IAPR International Conference on Pattern Recognition, Israel, pp.450- 460,1994.

[10] Van Erp, Merijn, Vuurpijl, Louis G., Franke, Katrin, Schomaker, Lambert R. B., " The WANDA Measurement Tool for Forensic Document Examination", Proceedings of the 11th Conference of the International Graphonomics Society, p. 282-285, Scottsdale, Arizona USA, 2003.

[11] Sita, Jodi C., Rogers, D., Found, B., " Spatial Comparison of Questioned to Specimen Signatures using Matrix Analysis Software", Proceedings of the 11th Conference of the International Graphonomics Society, p. 299-2303, Scottsdale, Arizona USA, 2003.