On-line Signature Verification: Discrimination Emphasised.

Ronny Martens, Luc Claesen
IMEC vzw
Kapeldreef 75, 3001 Heverlee, Belgium
Tel.: 32(0)16-28 12 11
Fax.: 32(0)16-28 15 01
E-mail: martens@imec.be

Abstract.

Narrowly seen, this paper presents an on-line signature verification system, based on 3D force patterns and pen inclination angles, as recorded during signing. The feature extraction mechanism is based on the well-known elastic matching technique. In contradiction to previous work in the same area however, we emphasise the importance of the final step in the process: the discrimination based on the extracted features. We show that by choosing the right discrimination approach we are able to improve the quality of the entire verification process drastically. The techniques we compare for discrimination, however, are not specific to signature verification, but should be considered carefully in every process, where a classification decision is made out of a set of parameters.

1. Introduction.

The problem of proving a subject's identity is as old as mankind itself. Just think about the use of a simple key, a password etc. Within the area of identity verification we denote two basic approaches: biometric and non-biometric verification.

Until now, non-biometric methods are most popular, mainly because of their simplicity. The disadvantages of this approach are obvious. The person whose identity is to be verified has to do a considerable effort. He has to remember a PIN-code or password, or to carry a characteristic object with him. Furthermore, the "secrets" that are used for identification can be easily stolen or lost. Biometric identity verification on the contrary is based on the use of physiological or behavioural characteristics. As a consequence, the disadvantages mentioned higher are no longer present, but the systems are much more difficult to implement. An overview of several biometric identity verification methods is presented in [1].

In this paper we deal with on-line signature verification. This means that we use signals like pen-tip position, velocity or acceleration, forces on the pen-tip, pen inclination angles etc. to verify whether a person is who he claims to be. A general overview of the field of on-line signature verification is given in [2] and [3].

The organisation of this paper is based on the different steps that should be taken when a signature is classified. In section 2 we describe very briefly the process of collecting and pre-processing the data used for verification. Section 3 describes the feature extraction process. Section 4 compares different discrimination approaches. 4.1 describes the result of Sato's discrimination approach [4]. 4.2 describes the statistical approach, while 4.3 focuses on Artificial Neural Networks. General conclusions are drawn in section 5.

2. Data collection and pre-processing.

Our input-instrument, the SmartPen™ (Figure 1 and [5]), looks like an ordinary pen. However, when a person is writing, 5 different signals are registered. These are the forces on the pen-tip in 3 directions, and the pen inclination angles. As we are not interested in rotations of the pen around its own axis, 2 angle signals instead of 3 are sufficient. All 5 signals are low-pass filtered with cut-off-frequency 40 Hz. It is generally accepted (e.g. [2]) that the bandwidth of handwriting is about 20 Hz, so the margin we use is safe. The resulting signals are sampled at 100 Hz. As already mentioned, the orientation of the pen in the signer's hand is not characteristic, but it does influence the other signals, as those are measured relative to the pen. We eliminate the effect of these pen-rotations by redefining the co-ordinate system that is used. The new reference axis's are chosen as the ones with extreme energy contents.

0-8186-7898-4/97 $10.00

Figure 1: The SmartPen.

3. Feature extraction.

In every approach to signature verification we can recognise a feature extraction and a classification process. Sometimes this separation is obvious, for instance when using a pure statistical approach [7]. In other cases it can be very difficult to clearly separate the two stages, e.g. when using neural nets to do verification [8]. In literature two classes of verification approaches are described, the parameter-based and the function-based approach. In the parameter-based approach we don't need a complete reference signature in order to classify a new signature, so typical parameters are energies, durations etc. In the function-based approach on the contrary, we do need a reference. Characteristic parameters used here are for instance correlation-coefficients. Since the main aim of this paper is to emphasise the importance of the classification process, we do not care very much about the actual feature extraction. Instead, we use a common function-based approach called dynamic time warping (DTW). This technique was introduced in the signature verification field in [4]. The aim of the algorithm is to find a to some criterion optimal time-alignment between the test (T) and reference (R) patterns to be compared. A time-alignment p is denoted as:

$$p = c(0), c(1), c(2), \dots, c(K) \tag{1}$$

$$p_k = c(k) = (i(k), j(k)) \tag{2}$$

I and j denote the i-th/j-th sample in respectively R and T. As described in [9] several obliged and optional constraints exist for the alignment, but these will not be explained further here. Defining

$$d(p_k) = d((i(k), j(k))) = \|R_i - T_j\| \tag{3}$$

$$D(p) = \frac{\sum_{k=1}^{K} w(k).d(p_k)}{\sum_{k=1}^{K} w(k)} \tag{4}$$

we are looking for the alignment P with D(P) minimal. The weight vector w can in theory be used to stress the importance of certain parts of a signature, but normally its elements are determined only by the local shape of the warping path [9]. The 2 features we use to classify are:

$$x_{Form} = D(P) \tag{5}$$

$$x_{Motion} = \sum_{(i,j)\in P} \|i - j\| \tag{6}$$

4. The discrimination process.

The purpose of the discrimination process is to extract a binary (genuine/forgery) decision out of a previously computed feature vector.

In the following sections we compare different discrimination approaches. Sato's approach is included to illustrate the importance of the classification scheme. In the statistical approach we try to construct an optimal (Bayes) classifier by estimating probability density functions (PDF's) from the populations involved. Using neural nets, we assume the net will be able to learn how to combine the available features in an optimal way.

As a performance indicator we will use the Equal Error Rate (EER). This is the error-percentage that occurs when False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. The databases we use have been constructed as follows. 18 persons provided us with 20 original signatures each. These signatures have been registered in 3 sessions spread over 3 months. For the construction of the discriminant functions we use the 10 originals resulting from the first session. The remaining 10 originals are used when testing the system performance. As forgeries for a certain person we use the genuine signatures produced by the other 17 participants. An other group of 41 persons provided us with a set of 15 random forgeries each. This set is, when appropriate, described as the "population". Since we are basically comparing classification-methods, we do not care about the rather poor quality of our forgeries.

4.1 Sato's approach[1].

The discriminant-function used by Sato (D_{Sato}) looks as follows:

$$D_{Sato}(\xi) = \xi'.\Pi^{-1}.\xi \tag{7}$$

$$\xi = [x_{Form}\ x_{Motion}]' \tag{8}$$

$$\Pi = E_{originals}(\xi.\xi') \tag{9}$$

Π, and thus the classification performance, is dependent on the reference selected. Since ideally $\xi = 0$

---

[1]Reading [4] one will notice that the author did not use exactly the same parameters as we do. However, we chose features with a similar statistical distribution.

we choose the reference that minimises det(Π). The EER using this approach is 11.1 %.

4.2 Statistical approach.

From a statistical point of view the design of an optimal classifier is easy, since we can obtain a mathematical expression for the expected cost (C) associated with a certain decision function.

$$C = K_g.ap_g.\int_{R_f} p_g(\xi)d\xi + K_f.ap_f\int_{R_g} p_f(\xi).d\xi \qquad (10)$$

The symbols have the following meaning:

$K_g/_f$: The cost for misclassifying a genuine/forged signature.

$ap_g/_f$: The a priori chance for a certain observation to be an original/forgery.

$p_g/_f(\xi)$: The probability density function for the genuine/forged signatures.

$R_g/_f$: The region where a certain feature vector is considered as coming from a genuine/forged signature.

Substituting $\int_{R_g} p_f(\xi).d\xi$ by $1-\int_{R_f} p_f(\xi).d\xi$ one can easily derive that the frontier between R_g and R_f is defined by:

$$\frac{p_g(\xi)}{p_f(\xi)} = \frac{K_f.ap_f}{K_g.ap_g} \qquad (11)$$

As we can not determine exact costs or a priori chances, the only relevance of (11) to us is that it is important to look at the value of:

$$\frac{p_g(\xi)}{p_f(\xi)} \qquad (12)$$

to make a decision.

First we take a look at Mahalanobis decision making, which assumes a Gaussian probability distribution of the parameters used. Afterwards we describe the more general kernel approach.

4.2.1 Mahalanobis distances.

A very well known discriminant function is based on the use of the Mahalanobis distance function. This vector to cluster distance ($D_{Mahalanobis}$) is defined as follows:

$$D_{Mahalanobis}(\xi) = (\xi-\xi_\mu)'.\Sigma^{-1}.(\xi-\xi_\mu) \qquad (13)$$

$$\xi_\mu = E_{originals}(\xi) \qquad (14)$$

$$\Sigma = E_{originals}((\xi-\xi_\mu).(\xi-\xi_\mu)') \qquad (15)$$

Mahalanobis decision making is, when the right offset is chosen, optimal on condition that $p_f(\xi)$ is constant, and $p_g(\xi)$ is multivariate normal. The EER using $D_{Mahalanobis}(\xi)$ equals 1.6%. The evolution of FAR as a function of FRR can be seen in Figure 2.

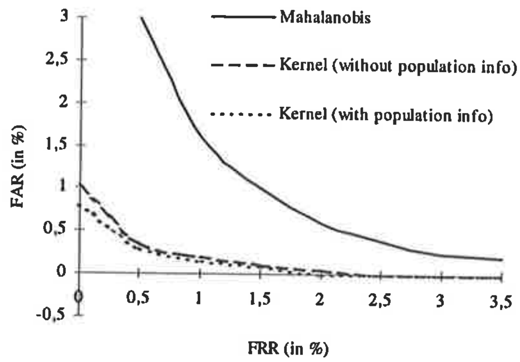

Figure 2: FAR and FRR for different classification approaches.

4.2.2 Kernel-approach.

Figure 3 contains a plot of a typical one-dimensional Gaussian PDF that has been constructed out of the original x_{Motion}-data for a given person. Clearly, the Gaussian approximation is not very accurate.

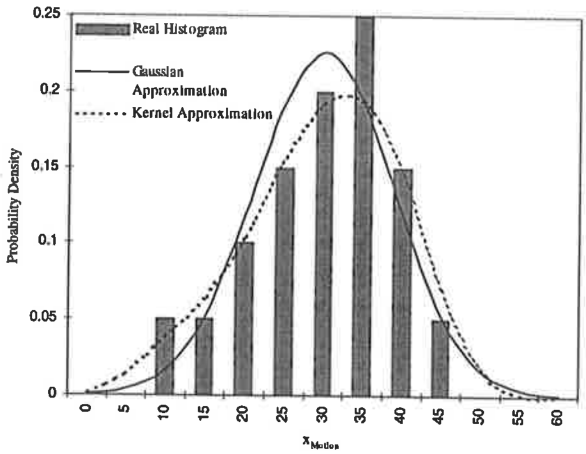

Figure 3: PDF-estimations.

Using the kernel approach we are able to perform a more accurate estimation of one or both the PDFs used by the Bayes-classifier. A complete description of the kernel method can for example be found in [10]. The key idea behind the approach is to describe a PDF as a sum of (kernel) functions, one for each observation point x_i. In theory, any type of function can be used as a kernel function (k(x)). Our results are best with:

$$k(x,x_i) = \frac{1}{h}.e^{-\frac{(x-x_i)^2}{h^2}} \qquad (16)$$

h can be chosen for optimal performance. Our results are best with h equal to $3.std_{originals}(x)$. Since we assume x_{Motion} and x_{Form} are independent, we can easily construct the global distribution function $p_g(\xi)$ out of both one-dimensional ones $p_{g,Form}(x_{Form})$ and $p_{g,Motion}(x_{Motion})$ as follows:

$$p_g(\xi) = p_{g,Form}(x_{Form}) \cdot p_{g,Motion}(x_{Motion}) \quad (17)$$

This method produces an EER of 0.4%. Figure 2 shows FAR as a function of FRR. The reason for the good performance can be seen in Figure 3. The PDF constructed by using kernel functions looks like a much more natural approximation of the real PDF than the Gaussian one. For instance, the right-skewness of the original distribution is clearly present in the kernel PDF-approximation.

So far, we have been approximating only $p_g(\xi)$, assuming $p_f(\xi)$ constant. (12) stresses that information about the population of forgeries can be useful when trying to improve classification. Using the kernel approach to approximate both $p_g(\xi)$ and $p_f(\xi)$[2], and (17) to perform the actual classification, we obtain an EER of 0.3%. Classification results are once more visualised in Figure 2.

4.3 Artificial neural networks (ANN).

A general overview of ANN's can be found in [11]. We use the most common ANN-classification approach: a feedforward net using logsig activation functions ([11]) and trained by backpropagation. Since we have only 2 features, we can easily use an augmented input space [12], consisting of our 2 features and the 3 quadratic terms that can be constructed out of them. Results are best when we use a single hidden layer with only 2 hidden neurons. In this case we have 1.3% FRR, and 1.5% FAR.

5. Conclusion.

It is shown that the actual performance of a signature verification system is highly dependent on the discrimination process used. Several possible discrimination techniques are evaluated. Best results are obtained when the kernel approach is utilised to estimate the real PDF for the class of original signatures. This is because using the kernel approach we can approximate this PDF much more accurately then when using for instance a Gaussian approximation. The resulting PDF simply has to be compared to a threshold. This reflects the fact that using random forgeries to construct $p_f(\xi)$ is not relevantly better than assuming $p_f(\xi)$ constant. In other words: a set of random signatures doesn't tell us very much useful about the characteristics of the class of forged signatures for a certain signer. This last remark is confirmed by the observation that ANN's, which normally need information about both classes of original signatures and of forgeries do not result in a good system performance at all.

[2] The kernel approximation of $p_f(\xi)$ is computed using the set of signatures previously described as the "population".

6. Acknowledgement.

This research was supported by a scholarship from the Flemish Institute for the promotion of the scientific-technological research in industry (IWT).

7. References.

[1] B. Miller, "Vital signs of identity", IEEE spectrum, February 1994, pp. 22-30, 1994.

[2] R. Plamondon, G. Lorette, "Automatic Signature Verification and Writer Identification - State Of The Art", Pattern Recogn., vol. 22, no. 2, pp. 107-131, 1989.

[3] F. Leclerc, R. Plamondon, "Automatic Signature Verification: The State Of The Art -- 1989 - 1993", Int. J. Pattern Recogn. Artif. Intell., vol. 8, pp. 643-660, 1994.

[4] Y. Sato, K. Kogure, "Online signature verification based on shape, motion and writing pressure", Proc. 6th Int. Conf. on Patt. Rec., pp. 823-826, 1982.

[5] L. Claesen, D. Beullens, R. Martens, R. Mertens, S. De Schrijver, W. de Jong, "SmartPen: An Application of Integrated Microsystem and Embedded Hardware/Software CoDesign", ED&TC'96 User Forum, pp. 201-205, 1996.

[6] G. A. Bills, M. J. Varady, K. P. Zimmermann, "Spectral Analysis of Script and Signatures", Proc. 3rd Int. Symp. on Handwriting and Comput. Applic., pp. 178-180, 1987.

[7] H. D. Crane, J. S. Ostrem, "Automatic signature verification using a three-axis force-sensitive pen", IEEE Trans. Syst. Man Cybern., vol. 13, pp. 329-337, 1983.

[8] J. Bromley , J. W. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Säckinger, R. Shah, "Signature verification using a "siamese" time delay neural network", Int. J. Pattern Recogn. Artif. Intell. , vol. 7, pp. 669-688, 1993.

[9] H. Sakoe, S. Chiba, "Dynamic Programming Algorithm Optimisation for Spoken Word Recognition", IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-26, pp. 43-49, 1978.

[10] M. Nadler, E. P. Smith, Pattern Recognition Engineering, John Wiley and Sons, 1993.

[11] J. M. Zurada, Introduction to Artificial Neural Systems, West Publishing Company, 1992.

[12] Y. H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley Publishing Co., 1989.