Utilising Baum-Welch for On-line Signature Verification.

Ronny Martens Luc Claesen

IMEC vzw

Kapeldreef 75, B-3001, Heverlee, Belgium

Tel.: +32(0)16281211

Fax.: +32(0)16281501

E-mail: martens@imec.be

Abstract.

This paper describes a fully operational on-line signature verification system. From a hardware point of view, its heart is the SmartPen™, a special input-device allowing the perception of force and angle signals. The most important software aspect that we focus on here is the exploitation of the Baum-Welch procedure in the feature extraction process. This algorithm provides a mathematical basis for classifying a signature taking into account the relative importance of both the different signals under observation, and the distinct phenomena that are present in these. The usefulness of the approach is illustrated by presenting the results of a full-scale field test.

Key words.

On-line signature verification, Baum-Welch.

1 Introduction.

The most typical characteristic of our current society is undoubtedly its high degree of automation. At the moment a very large amount of jobs, previously done exclusively by humans, can be executed without quality loss by machines. The weakness of this automation process is however in many cases situated in the problem of autonomously identifying the subject the work has to be done for.

Until very recently most of the identity verification systems have been using in general very easy to implement non-biometric identity verification techniques. These techniques do not base the actual classification on a unique characteristic of a certain person, but on some kind of secret - a key, a pass-word - shared between the subject being verified and the object that is performing the verification task. In return for the simplicity of the concept, a rather high price has to be paid. First of all, it is for a rigid criminal relatively easy to gain insight in the secrets that are used for verification, or to get possession of the objects employed for this purpose. This means that the degree of security that is achieved by a non-biometric identity verification system is in practice rather low. Secondly, the person who is verifying his identity has to do a non-trivial effort, like remembering a PIN-code or carrying a key.

As the number of occasions where a certain person has to prove his identity is rapidly growing, the disadvantages mentioned above are getting more and more coercive. For this reason, and thanks to the technological evolution that makes more complex verification techniques feasible, there is in the last few years a steadily growing interest for biometric identity verification. These alternative identity verification systems use some kind of physiological (fingerprint, iris-pattern) or behavioural characteristic (handwriting, typing rhythm) to identify a certain person.

An excellent overview of the different biometric identity verification techniques is given in (Miller 1994). In this paper we concentrate on on-line signature verification. This means that we classify a certain person as a genuine or forger by observing the characteristics of his/her way of signing. An overview of the work that has recently been done in this area can be found in (Plamondon and Lorette 1989, Leclerc and Plamondon 1994).

A description of the different steps that have to be taken when building a complete on-line signature verification system is given in for instance (Plamondon 1994). In general, we distinguish between three major stages: data-acquisition and pre-processing, feature-extraction, and classification.

This separation is the backbone of our paper. Section 2 focuses on data-acquisition. Sections 3 and 4 deal with parameter-extraction respectively classification. We evaluate the performance of the global system in section 5. Conclusions are drawn in section 6.

2 Data-acquisition and pre-processing.

Two different trends can be observed in data-acquisition for on-line signature verification. A majority of the researchers (Bromley et al. 1993; Wirtz 1995; Huang and Yan 1995) opt for a classic tablet for the registration of the positions of the pen-tip over time. A second group hopes to end up in a better cost-benefit situation by building a special input-device that allows the observation of signals like forces, accelerations etc. (Baron and Plamondon 1989; Herbst and Liu 1977). In this work the second approach is chosen for. Our input-device, the SmartPen™ (Fig. 1 and Claesen et al. 1996), captures 5 different signals:

- Forces on the pen-tip in 3 directions.
- Angles of the pen-shaft relative to the writing surface in 2 dimensions.

Figure 1: The SmartPen™.

We don't have any information about the rotation of the pen around its own axis. In practice, this is not a problem since this rotation is not characteristic for the signing process. The actual pen-rotation does however influence the other signals that are registered, as these are measured relative to the pen. Using a new reference co-ordinate system eliminates this effect. The main axes of this system are defined as the ones in which the force signals have extreme energy contents. The resulting signals are low-pass filtered with cut-off frequency ±30 Hz. It is generally accepted (Leclerc and Plamondon 1994) that higher frequency regions contain no signer-specific information. After A/D conversion and encryption the signals are transmitted wirelessly to the computer doing the actual feature extraction and classification.

3 Feature extraction.

The ability to deal with non-linearities in the time-domain is an important property of a good on-line signature verification system. This is illustrated by the fact that many recent publications describe verification techniques that allow handling them (Wirtz 1995, Martens and Claesen 1997, Kashi et al. 1997). The traditional approach for removing timing-differences in the signature verification area is to use Dynamic Time Warping (DTW). This technique originates from the speech recognition field. In the last decade, Hidden Markov Models (HMMs) have been applied very successfully in this area for the same purpose. Recently, the HMM-approach migrated to the signature verification field, with good results as well (Yang et al. 1995; Kashi et al. 1997). However, while the concept of a state is very clear in the speech recognition area (Markel and Gray 1976), it is not in the signature verification field. A first step towards a solution for this problem is the use of the Bakis type model[1], where the concept of a state is clearly weakened. We make one more step in this softening process, as will become clear in the next sections. We start by discussing the general characteristics of our approach in section 3.1, and focus on its peculiarities in 3.2.

3.1 General characteristics.

In previous work (Martens and Claesen 1996) we have illustrated the importance of taking into account the relative weight of phenomena situated in different frequency regions during the classification process. As a consequence, an observation has to contain information about several succeeding time-samples. In this context, a useful observation sequence for a certain signal might look like in Figure 2.

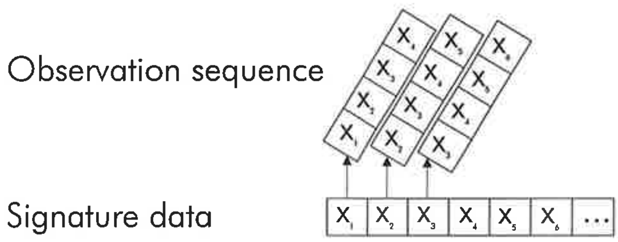

Figure 2: Construction of observation sequence out of signature data.

---

[1] In a Bakis type model the number of states is proportional to the average duration of the observation sequences.

The use of a classical HMM for signal classification requires the ability to compute the chance of occurrence for every state at each time instant. As we have only a very limited number of genuine signatures at our disposal, we have to make some good assumptions about the characteristics of the observed – by nature continuous – features used for classification. If we don't, it will become impossible to compute the PDFs used for the computation of these chances with a reasonable accuracy. As illustrated in (Martens and Claesen 96), we can approach the PDF for an observation - like the one in Figure 2 - to occur at a certain time-instant, by a product of univariate Gaussian distributions for each of the different features included in the observation. In order to do this, it is however essential that these features are first decorrelated by e.g. the Gabor-transform or the discrete wavelet transform (DWT).

Next to the construction of the PDFs as discussed in the previous paragraph, we have to compute transition probabilities from one state to another. Our approach is based on the concept of a left-to-right HMM. This means that when a certain state is left, only one new state can be reached. In many situations the transition probabilities are modelled by using a simple transition matrix. This means that the probability to stay in a certain state decays exponentially as a function of time. This duration-modelling procedure is obsolete for our application. As a result we use explicit time duration modelling. We assume the duration to stay in a certain state is a Gaussian distributed variable. The fact that all the parameters in our model have a Gaussian distribution will have an important effect on the actual classification process, as discussed in section 4.

3.2 Special features.

When building a HMM with the characteristics mentioned higher, a state is associated with every observation. More important, every observation contributes to the final probability for a certain observation sequence to be generated by the model under study. Because we don't have the same strong state-concept as in for instance speech, we just want to perform a good time-alignment between a model constructed out of a set of reference patterns and the observation sequence to be classified.

We solve this difficulty by associating just one observation to a state, using the majority of the observations only for the computation of the duration of a state. For instance, in the case of Figure 2, on average only 1 observation out of 4 should be matched to a state. Of course, one can increase the number of states in the model – and thus the number of observations that is really matched to a state -, but this only introduces overhead into the classification process. As is generally known, overhead should be eliminated as it effects the performance of a pattern classification system in a negative way.

Figure 3 shows a possible way to implement this approach. The figure reveals a part of the model that should be constructed when the observation sequences look like the ones from Figure 2. The complete model should consist of a sequence of these basic blocks.

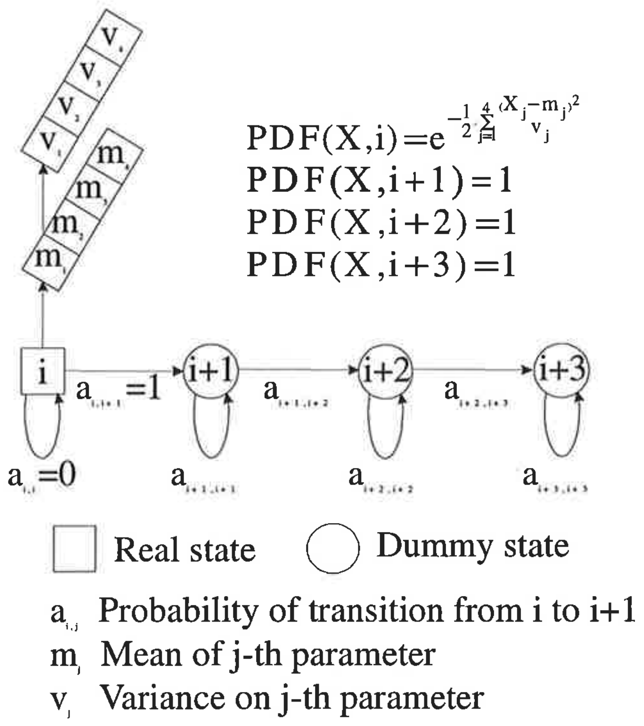

Figure 3: Possible implementation of alternative HMM-approach.

As is illustrated, the observation in a dummy state is of no importance. The a-values of all dummy states between succeeding real states are chosen in order to give the duration between these real states a Gaussian distribution.

By increasing the number of real states, relative to the number of dummy states, it becomes clear that this approach is just an evolution of the Bakis type model.

At this moment we can define the features that are used as an input to the actual classification process that is described in section 4. After computing the most probable sequence of state transitions for a certain signature, we have 2 types of parameters at our disposal:

- Motion information parameters.

  These describe the time-interval between 2 succeeding real states. The vector that contains all motion information parameters is denoted as M.

- Form information parameters.

  They form a vector that containing all the observations that are matched to a real state. This vector is indicated by F.

We define the resulting feature vector T as $\begin{bmatrix} M \\ F \end{bmatrix}$. The number of components in T is denoted by N.

4 Classification.

As has already been mentioned in section 3, the distribution of the components of both M and F can be reasonably well approximated in a Gaussian way. It is also intuitively clear that the degree of correlation between M and F is small. This is also true for the correlation coefficients between different components in M. Thanks to the fact that the components of F have been computed using the DWT, these are also decorrelated (Martens and Claesen 1996). Both conditions – Gaussian distribution and absence of relevant correlations - are enough to conclude that X^2 (1) has a chi-square distribution (Papoulis 1990).

(1) $$X^2 = \sum_{i=1}^{N} \frac{(t_i - \bar{t}_i)^2}{\sigma(t_i)}$$

As the number of degrees of freedom (DoF, in this case equal to N) is large, we can construct a variable Z (2) out of X that has a standard normal distribution (Papoulis 1990).

(2) $$Z = \frac{\left(\frac{X^2}{\mathrm{DoF}}\right)^{\frac{1}{3}} - \left(1 - \frac{2}{9.\mathrm{DoF}}\right)}{\sqrt{\frac{2}{9.\mathrm{DoF}}}}$$

To decide about the originality of a signature, we have to compare $|Z|$ to a threshold. Due to the solid mathematical fundaments, this threshold is person-independent, and – if the class of reference signatures is really representative for the class of signatures that a certain person can produce – we can even predict the FRR[2] for a certain threshold value.

5 Performance evaluation.

The use of the Baum-Welch algorithm to model a class of genuine signatures, as has been described higher, is closely related to the use of the DTW-algorithm for feature extraction (Martens and Claesen 1996). This last approach also allows dealing with the relative importance of phenomena situated in different frequency or time-intervals for a certain class of signatures. The only additional feature of the approach that is described here is that we can take into account the

---

[2] FRR stands for False Rejection Rate. This is the percentage of genuine signatures that is misclassified by the system. This number has no meaning on itself, but it should be looked at in combination with the FAR (False Acceptance Rate). This is the number of forgeries that is considered as genuine signatures.

relative importance of the different signals that are observed. This importance is even allowed to change dynamically in the time and frequency domain. Figure 4 reveals the evolution of the FRR and FAR as a function of the classification offset. In order to clarify the effect of the added functionality of the approach, we have linearly scaled the classification-offsets, making the points of EER of both approaches appear at the same offset-value. It is clear that our alternative performs relevantly better than the DTW-based approach. The EER[3] is reduced from about 5.5% to about 3%, and the classes of genuines and forgeries are separated much clearer.

---

[3] EER stands for Equal Error Rate. This is the percentage of misclassifications that occurs when the classification-offset is chosen in order to make FAR and FRR equal. This number is very dependent on the database that has been used. Here we use a database with 45 people involved. Each person provides us with 15 real signatures. 10 are used for training, and the remaining 5 for testing. Every person provides us with a forgery for each of the other persons involved in the database.

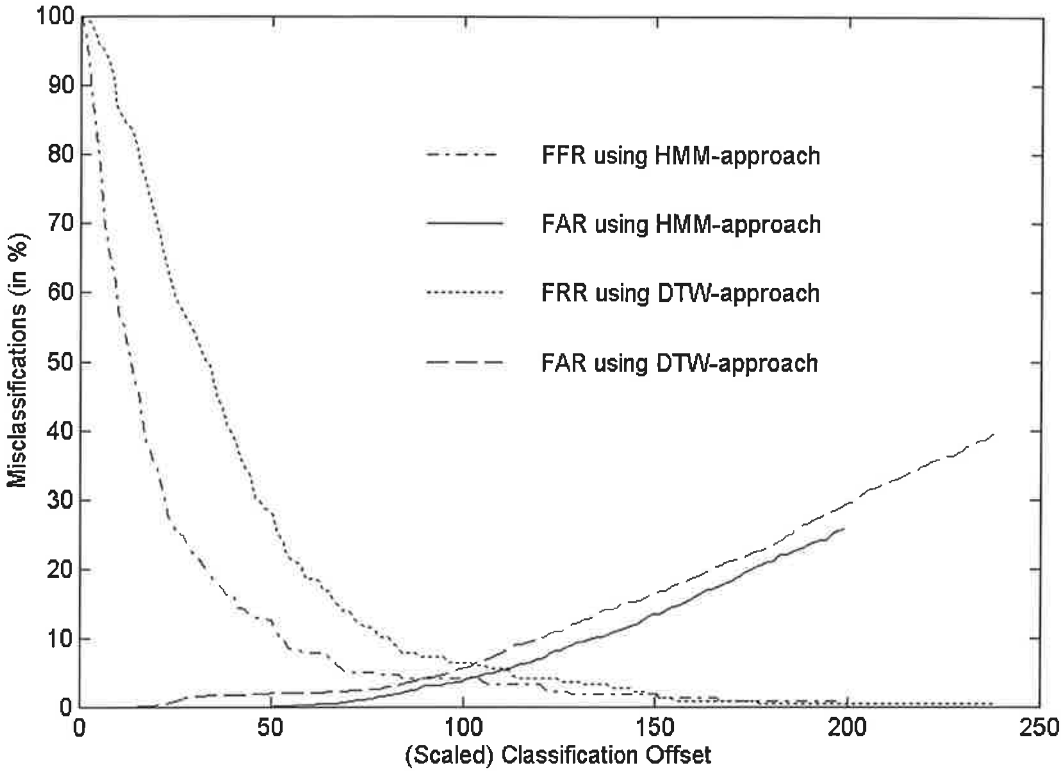

Figure 4: The benefits of the HMM-based verification approach.

6 Conclusion.

We have built a fully operational biometric identity verification system based on on-line signature analysis. The hardware-heart of the system is the SmartPen™. This instrument allows observing forces on the pen-tip and angles of the pen-shaft while a person is writing. The main topic of this paper is to answer the question on how to extract features that satisfy 3 conditions:

- They should be local in both time and frequency-domain.
- They reflect the relative importance of the different types of signals in a mathematically correct manner.
- They allow dealing with non-linear phenomena.

A first step in solving this problem is the use of the discrete wavelet transform to construct observation sequences where the individual features are decorrelated. Afterwards, we can build a model for the class of genuine signatures using the Baum-Welch algorithm. The approach that is presented is in fact an evolution of the Bakis type HMM.

Using this model, we extract form and motion parameters for a certain signature that should be verified by simply computing the optimal sequence of state transitions.

The last step in the verification process is the classification. The construction of a good classifier is in general a difficult problem, but in this case it is simplified a lot because the form and motion parameters we use have a Gaussian distribution and are decorrelated. Because the number of features that we use is rather big, we can easily combine them into a single parameter that has a standard normal distribution for the class of genuine signatures under study. If the class of reference signals we have at our disposal is really representative for the class of genuine signatures this even allows to predict the FRR, by examining the classification offset.

We have illustrated that the technique offers better classification results than a technique based on the combined use of the discrete wavelet transform and dynamic time warping. The price that has to be paid for this is a drastic increase of the complexity of the training process. The actual verification procedure does not suffer very much from this drawback, as the major complexity of the process really has to be situated in this training stage.

7 Acknowledgement.

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

8 References.

Baron R, Plamondon R (1989) Acceleration Measurement with an Instrumented Pen for Signature Verification and Handwriting Analysis. IEEE Trans. on Instrumentation and Measurement 38:1132-1138

Bromley J, Bentz JW, Bottou L, Guyon I, Lecun Y, Moore C, Säckinger E, Shah R (1993) Signature Verification Using a "Siamese" Time Delay Neural Network. Int. Journal of Pattern Rec. and Art. Intelligence 7:669-688

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

Herbst NM, Liu CN (1977) Automatic Signature Verification Based on Accelerometry. IBM Journal of Research and Development 21:245-253

Huang K, Yan H (1995) On-Line Signature Verification Based on Dynamic Segmentation and Global and Local Matching. Optical Engineering 34:3480-3487

Leclerc F, Plamondon R (1994) Automatic Signature Verification: The State of the Art -- 1989-1993. Int. Journal of Pattern Rec. and Art. Intelligence 8:643-660

Martens R, Claesen L (1997) An Evaluation of Different Handwriting Observation Techniques from a Signature Verification Point of View. In: Nabeel A. Murshed, Flávio Bortolozzi (eds) Advances in Document Image Analysis. Springer-Verlag Berlin Heidelberg New York pp273-282

Miller B (1994) Vital Signs of Identity. IEEE Spectrum 31: 22-30

Plamondon R (1994) The Design of an On-Line Signature Verification System: From Theory to Practice. Int. Journal of Pattern Rec. and Art. Intelligence 8:795-811

Plamondon R, Lorette G (1989) Automatic Signature Verification and Writer Identification - The State of the Art. Pattern Recognition 22:2:107-131

Wirtz B (1995) Stroke-Based Time Warping for Signature Verification, In: Proc. 3rd Int. Conf. on Doc. Analysis and Recognition pp 179-182

Martens R, Claesen L (1996) On-line Signature Verification by Dynamic Time Warping, In: Proc. 13th Int. Conf. on Pattern Recognition, vol. 3, pp. 38-42

Markel JD, Gray AH (1976) Linear Prediction of Speech. New York, NY: Springer Verlag

Papoulis A (1990) Probability and Statistics. Prentence-Hall International Editions

Yang L, Widjaja BK, Prasad R (1995) Application of Hidden Markov Models for Signature Verification. Pattern Recognition 28:161-170

Kashi RS, Hu J, Nelson WL (1997) On-line Handwritten Signature Verification using Hidden Markov Model Features, In: Proc. 4th Int. Conf. on Doc. Analysis and Recognition, pp. 253-257