On-line Handwritten Signature Verification using Hidden Markov Model Features

R. S. Kashi
J. Hu
W. L. Nelson

Bell Labs, Lucent Technologies,
600 Mountain Ave
Murray Hill, NJ 07974

W. Turin

AT&T Research Labs
180 Park Ave
Florham Park, NJ 07932

Abstract

A method for the automatic verification of on-line handwritten signatures using both global and local features is described. The global and local features capture various aspects of signature shape and dynamics of signature production.

We demonstrate that with the addition to the global features of a local feature based on the signature likelihood obtained from Hidden Markov Models (HMM), the performance of signature verification improves significantly.

The current version of the program has 2.5% equal error rate. At the 1% false rejection (FR) point, the addition of the local information to the algorithm with only global features reduced the false acceptance (FA) rate from 13% to 5%.

1 Introduction

Signature verification is a common behavioral biometric to identify human beings for purposes of establishing their authority to complete an automated transaction, gaining control of a computer, or gaining physical entry to a protected area. Signatures are particularly useful for identification because each person's signature is highly unique, especially if the dynamic properties of the signature are considered in addition to the static shape of the signature. Even if skilled forgers can accurately reproduce the shape of signatures, it is unlikely that they can simultaneously reproduce the dynamic properties as well.

On-line signature verification schemes extract signature features that characterize spatial and temporal characteristics of a signature [1]. The feature statistics of a training set of genuine signatures are used to build a model or template for validating further test signatures [2]. Selecting a good model is the most important step in designing a signature verification system. The other important component of the system is a distance measure between a signature and its model. Signature models are usually described by the set of parameters (features) which can be roughly divided into local and global feature subsets. Global features are calculated for sufficiently large segments of signatures (such as pen-down segments or the whole signature) while local features are calculated for smaller segments (such as equally spaced sub-segments or even every signature sample). For example, total signature time and length of a signature are global features while slope tangent at each point is a local feature. Local features are more sensitive to handwriting variations than global features, but require more computer resources for processing and storage. The feature set selected should be a judicious choice of global and local features having maximum discriminative power while keeping the set cardinality small.

There are several methods of using local features in signature verification [3]. The most popular method is based on elastic matching by Dynamic Warping (DW) [4]. An alternative strategy (commonly used in speech recognition) is using Hidden Markov Modeling (HMM) [5, 6]. Dynamic warping performs flexible matching of local features of a model and a signature sample. The closeness of the match is used as discriminator. An HMM performs stochastic matching of a model and a signature using a sequence of probability distributions of the features along the signature. In this paper we model the signing process with several states that constitute a Markov chain, each of them corresponding to a signature segment. The states are not directly observable (hidden); we can only observe the signature local features (such as tangent angles). The observations are bound statistically with the model states and conditionally independent in each state. During training, the model parameters are estimated from a set of valid signatures. During a signature verification, the probability that the signature is genuine is calculated. If this probability (which is called the likelihood function) is high, the signature is accepted otherwise it is rejected. This approach can be viewed as a statistical matching of the test signature with the signature HMM. In our signature modeling, we use the handwriting tangent and its derivative as an observation vector in equal length segmentation. For this observation vector, the HMM likelihood method of the signature verification performed comparable to the Euclidean distance rule presented in an earlier paper [2]. However, when we combined the global and local features, we obtained significantly better performance.

The remainder of the paper is organized as follows: Section 2 presents signature normalization by Fourier transforms, Section 3 describes the global feature selection, Section 4 describes the log-likelihood computation using discrete HMM's and Section 5 describes the results of our signature verification algorithms on the Murray Hill signature database [2].

2 Signature Pre-processing

2.1 Definitions

It is convenient to represent planar curves (x_0, y_0), (x_1, y_1), ... (x_{N-1}, y_{N-1}) using complex vector notations.

Using vector notations, the combined affine transformation of a curve can be expressed as

$$\mathbf{z}^{'} = A\mathbf{z} + B \tag{1}$$

where $A = Kexp\,[j\theta]$, $K = |A|$ is related to scale, $\theta = arg\,(A)$ is related to angle of rotation, and B is a complex quantity related to the displacement.

2.2 Fourier Normalization

Fourier transform of a planar curve $\mathbf{z}$ is defined as a mapping

$$\mathbf{Z} = \mathbf{F}\mathbf{z} \tag{2}$$

where $\mathbf{F}$ is the Fourier matrix. The entry in row j and column k of F is defined as $F_{jk} = w^{jk}$ where w is the primitive N^{th} root of unity and is defined as $w = exp\ [2\pi i/N]$. Since this mapping is invertible:

$$\mathbf{z} = \mathbf{F}^{-1}\mathbf{Z} \tag{3}$$

and since the Fourier matrix is both symmetric and orthogonal [9]

$$\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^{*} \tag{4}$$

The sequence $\mathbf{Z}$ can be used as an alternative representation of a curve in the "spatial frequency" domain. This and similar transforms are very popular in image processing, because often many coefficients of $\mathbf{Z}$ are small and can be neglected thus giving a more compact and smoothed curve representation. In our implementation, we obtained the Fourier transform of the first difference of the coordinates, instead of the actual coordinates as shown in equation 2. This was done to reduce the end-point distortions in the reconstructed signature [7].

The objective of signature normalization is to transform a signature to some canonical form which is then processed by the signature verification algorithm. Our approach to signature normalization is based on the normalization of its Fourier coefficients. First we set $Z_0 = 0$ which according to equation (1) is equivalent to translating the coordinate system origin to the curve centroid

$$Z_0 = 1/N \sum_{k=0}^{N-1} z_k \tag{5}$$

Next, we divide the rest of the Fourier coefficients by Z_1. Since the Fourier transform is linear, each coordinate of the translated signature is divided by Z_1 which according to equation (1) is equivalent to scaling by $K = 1/|Z_1|$ and rotating by $\theta = -argZ_1$.

The normalized signature is obtained by the inverse transform of the normalized Fourier coefficients and is used for the computation of the global and local features.

3 Global feature-based verification procedures

Global features by definition are properties of the signature as a whole, or of a substantial part of the signature, rather than a property of the signature that depends on an individual shape or detail within the signature. The two main global aspects of the signature data, namely, *shape* and *dynamics* play somewhat complementary roles in distinguishing genuine signature from forgeries, i.e. the more forgers try to match every detail of a signature's shape, the less likely they are to match its dynamics, and vice versa. The advantage of using dynamic features is that they are "hidden", since they are not apparent from an examination of a paper copy of the signature. Therefore verification procedure should include a mixture of both shape and dynamic features.

In our current algorithm, we use 23 global features. The names and analytical expressions for computing these features are explained in the Appendix. There are two time-related features: The first is the total signature time T. The second is the time down ratio T_{dr}, which is the ratio of pen-down time to total time.

Six other dynamic features depend on the writing velocity and acceleration. The x,y components of velocity and acceleration were computed from the derivatives of the pen coordinates with respect to time

$$\mathbf{v} = (v_x, v_y) = (\dot{x}, \dot{y}),\ \mathbf{a} = (a_x, a_y) = (\dot{v_x}, \dot{v_y}).$$

The derivatives were calculated using cubic smoothing B-splines [10]. The path velocity magnitude (speed) and path-tangent angle of the pen motion are given in terms of the v_x, v_y components by

$$v = (v_x^2 + v_y^2)^{1/2},\ \theta = tan^{-1}(v_y/v_x).$$

The features derived from these speed components were root-mean-square (rms) speed V, average horizontal speed V_x, and integrated centripetal acceleration IA_c.

While these features are principally related to signature dynamics, they are all correlated to some degree with the signature shape (for example, the regions of high velocity are usually related to the regions of low curvature). Therefore, the shape-related features *should be carefully chosen to be not* strongly correlated with the selected dynamical features. We chose the following features: length-to-width ratio L_w, horizontal span ratio X_{wr}, horizontal centroid X_{cn}, and vertical centroid Y_{cn}. Eight features characterize the distribution density of the path-tangent angles θ_k, over the k = 1,2 ..., K data points for a given signature. We estimate it by the histogram of the angles that lie in eight sectors between zero and 2π by

$$S_l = \text{card}\{\theta_k : (l-1)\pi/4 < \theta_k \leq l\pi/4\}/K$$

where $k = 1, \ldots, K, l = 1, \ldots, 8$ and card$\{A\}$ denotes the cardinality (i.e. the number of elements) of the set A. The next four features are angle-sector densities of the angular changes $\delta_k = \theta_k - \theta_{k-1}$, that lie in four sectors between zero and 2π

$$C_m = \text{card}\{\delta_k : (m-1)\pi/2 \leq \delta_k \leq m\pi/2\}/(K-1)$$

where $k = 2, \ldots, K, m = 1, \ldots, 4$. Two more features, V_{xy}, relate to the correlation between the x and y component of pen speed, and the first global moment make up the global feature set.

The verification procedure used in this study are based on the one discussed in detail in a previous paper [2]. Briefly, the procedure builds a model based on the genuine signature data since the statistical properties of genuine signature data obtained from a known signer should be reasonably predictable. A signature model for entrant i is a set of means, μ, and standard deviations, σ, obtained during training from six instances of signatures. Therefore, using the above model for classifying signatures from genuine signatures only, one can define an error measure E_i for a given signature which is claimed to be of entrant i by

$$E_i = \left(\sum_{k=1}^{N} \left((M_{ik} - \mu_{ik})/\sigma_{ik} \right)^2 \right)^{1/2} \qquad (6)$$

Here, N is the total number of global features to be used for verifying the signatures of entrant i. M_{ik} is the value of the k-th feature as evaluated on a signature claimed to be that of entrant i, and μ_{ik} and σ_{ik} are, respectively the mean and standard deviation of that feature over the reference set of entrant i.

Because global features do not relate directly to the shape of the components of a signature, a signature model's discriminative power can be improved by adding a complementary set of local features which is described in the next section.

4 Hidden Markov Model For Signatures

4.1 Model Description

We assume that a signature can be described by a left-to-right HMM whose state-transition diagram is shown in Fig. 1. According to this model, the hidden chain transfers from state i to state $i+1$ with probability $a_{i,i+1}$ or stays in state i with probability $a_{ii} = 1 - a_{i,i+1}$. Let q_t denote the state the chain is in at time t, the probability that an observation vector $\mathbf{O}_t$ is inside some quantization region R_j while the chain is in state i is defined by the state conditional probability $b_i(j) = P\{\mathbf{O}_t \in R_j | q_t = i\}$.

It is easy to verify that state duration (state-holding time) distribution is geometrical for any Markov chain.

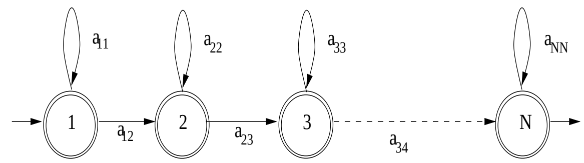

Figure 1: Left-to-Right HMM state-transition diagram.

The probability of having d consecutive observations in state i, is implicitly defined as:

$$p_i(d) = (a_{ii})^{d-1}(1 - a_{ii}) \qquad (7)$$

However, for the relatively small number of states that correspond to segments of a signature this distribution is inappropriate. To model more accurately handwriting process we chose to use explicit duration modeling, resulting in a variable duration HMM [11] [12]. This model can also be called a Hidden Semi-Markov Model (HSMM), because the underlying process is a semi-Markov process [13]. In this case, state-duration distribution can be a general discrete probability distribution. It is easy to prove that any HSMM model is equivalent to the so-called canonical HSMM in which $a_{ii} = 0$ [13]. Thus, for the left-to-right variable-length HMM there is no need of estimating state transitional probabilities.

In general, the variable duration HMM is not equivalent to an HMM with the finite number of states, but in all practical cases it can be approximated by an HMM by increasing the number of states. A brute-force approach to this modeling consists of decomposing an HSMM state i into unit-duration sub-states $i_1, i_2, \ldots$ with the transitional probabilities $a_{i-1,i_k} = p_{i-1}(k)$, $a_{i_j,i_{j+1}} = 1$ as shown in Fig. 2.

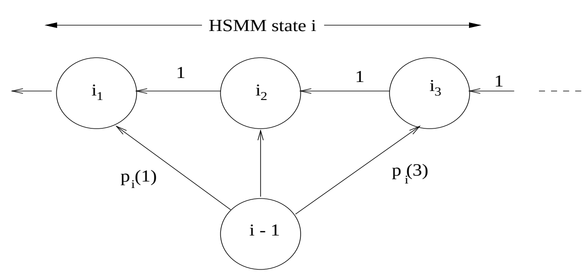

Figure 2: Equivalent HMM.

One problem of estimating duration probabilities $p_i(k)$ directly is that it requires many more training samples than those necessary for the training of observation probabilities. Since we do not usually use more than 6 training samples for each signature, an alternative duration model is used. In this alternative method, only the duration boundaries τ_i and T_i are estimated for each state using the segmentations obtained in the last iteration of training. Then the duration probabilities outside the range $[\tau_i, T_i]$ are assigned a very small value and those inside this range are assumed to be evenly distributed. Our experiments show that with limited amount of training data, this method with a uniform distribution provides much better results than using an exponential distribution.

Our current method of explicit duration modeling is based on statistical pattern matching and feature extraction. We would like to emphasize that in signature verification it is important to find a good discriminator – a function which separates valid signatures from forgeries. Signature likelihood and state-durations obtained in this way allow us to do consistent pattern matching.

4.2 Encoding Signature Samples

Each signature sample is represented as a sequence of observation vectors. In the current implementation, each observation vector is composed of the inclination angle α and the difference between adjacent inclination angles $\Delta\alpha$, sampled at points evenly distributed in terms of arc length (Fig. 3). The distance between adjacent sample points is proportional to the total arc length of the script so that each sample has roughly the same number of sample points. We assume that α and $\Delta\alpha$ are statistically independent.

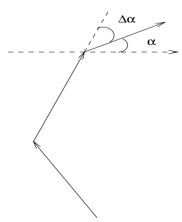

Figure 3: Components of an observation vector

Even though the observation vectors are continuous in nature, we chose to use discrete HMMs instead of continuous density HMMs to avoid making assumptions on the form of the underlying distribution. Since the two features are assumed to be independent, they are each quantized separately, and a separate probability distribution is estimated for each feature at each state. The joint conditional probability is simply the product of the probabilities of the two independent features.

A group of simple preprocessing procedures are applied to the original signature sample so that the observation vectors can be extracted more reliably. Following a cleaning operation during which irregular points (isolated points generated by electro-magnetic interference during data collection) and redundant points are removed, cusps (points with sharp direction change) are detected and marked. Then, a cubic B-spline smoothing filter is applied and, the signature is re-sampled at intervals of equal arc length. Cusps are preserved during smoothing and re-sampling to avoid loosing sharp angles. Finally the smoothed and re-sampled signatures are normalized using their Fourier transforms as explained in section 2.2.

4.3 The Likelihood Score

The Viterbi algorithm [5] is used to obtain the likelihood that the signature can be modeled by the HMM of the particular subject. The Viterbi algorithm searches for the most likely state sequence corresponding to the given observation sequence and gives the accumulated likelihood score along this best path. Using explicit state duration modeling, the increment of the log-likelihood score for the transitions to the $i+1$-th state are given by

$$\Delta L_{i,i+1} = \log b_{i+1}(k) + p_i(d) \qquad (8)$$

$$\Delta L_{i+1,i+1} = \log b_{i+1}(k) \qquad (9)$$

Although the length of the sequence of vectors is normalized by the total arc length of the signature, the exact number of observation vectors is not constant due to the slight length change introduced by the re-sampling process preserving cusps. To reduce the effect of length variation, the accumulated log-likelihood score given by the Viterbi algorithm is divided by the number of sample points to obtain the final normalized score to be used in verification.

4.4 Training

The HMM of each subject is trained by applying a segmental k-means iterative procedure [5] across all the training samples. The procedure is composed of iterations of the following 2 steps:

1. Segmentation of each training sample by Viterbi algorithm, using the current model parameters.

2. Parameter re-estimation using their means along the path.

The initial model parameters are obtained through equal-length segmentation of all the training samples. The re-estimation stops when the difference between the likelihood scores of the current iteration and those of the previous one is smaller then a threshold (usually after 5-7 iterations).

The observation probabilities are estimated in the usual way [5]. The mean of the accumulated log-likelihood obtained from the Viterbi algorithm is computed for the training signatures.

4.5 Testing

Given a test signature, the Viterbi algorithm is used to obtain the likelihood that the signature can be modeled by the HMM of the particular subject. The difference of this score and the mean log-likelihood obtained during training is then used as an error measure to classify a test signature as a valid or forgery. Further, this error measure was used individually and in combination with the global error obtained by equation 6. The global and local errors were combined using an Euclidean method by computing their root mean-square weighted combination.

5 Performance results and discussion

The performance of signature verification methods tested on the Murray Hill database will be shown in terms of Type-1 false rejection(FR) errors vs. Type-2 false acceptance (FA) errors, as the decision threshold is increased in small increments (Figures 4). The test database consisted of 542 genuine signatures and 325 forgeries. Each reference set consisted of the first 6

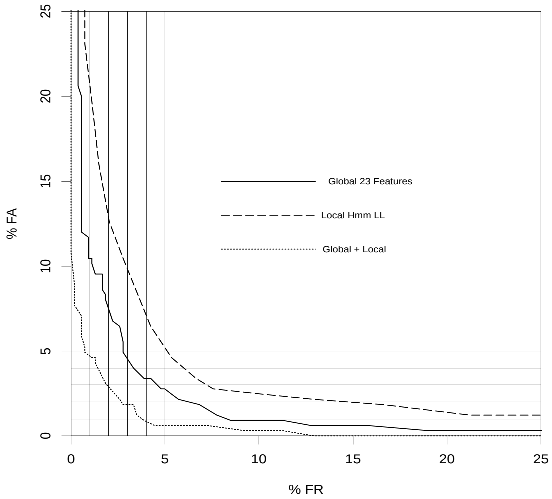

Figure 4: Error trade-off performance curves

signatures of every one of the 59 subjects. The digitizing tablet used for gathering the signatures had an 80 X 80 mm glass surface and a writing stylus that was electronically connected by a tether to the capacitive sensing system [14].

In Figure 4, the FR/FA error trade-off curves are shown for the algorithm with only 23 global features (solid line), with only the HMM log-likelihood score (dashed line) and for the same 23 features augmented with the HMM log-likelihood score (dotted line). As seen from Figure 4, the combination of the HMM log-likelihood and global feature information improves the performance in all regions of the plot when compared to either the local or global methods used independently. The equal error rate as seen from the figure decreases from about 4.5% to about 2.5% with the enhanced technique. At the 1% false rejection (FR) point, the addition of the local information reduced the false acceptance.

6 Summary

In this study, we have presented an on-line signature verification scheme using a hybrid technique of global and local features. The global features captures various spatial and temporal characteristics of the signature and the local features using the discrete HMM capture the dynamics of signature production. Together, they accurately model the signature variability of individuals and discriminate between valid and forged attempts. The algorithms described here can be used in both stand-alone point-of-sales and networked applications.

Acknowledgments

We would like to thank Mike Brown for the numerous helpful discussions.

References

[1] G. Lorette and R. Plamondon. Dynamic approaches to handwritten signature verification. In *Computer Processing of Handwriting*, pages 65–85, 1990.

[2] W. Nelson, W. Turin, and T. Hastie. Statistical methods for on-line signature verification. *International Journal of Pattern Recognition and Artificial Intellingence*, 8, 1994.

[3] M. Parizeau and G. Lorette. A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):710–717, 1990.

[4] Y. Sato and K. Kogure. Online signature verification based on shape, motion, and writing. In *Proc. 6th of ICPR*, pages 823–826, 1982.

[5] L. R. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[6] L. Yang, B.K. Widjaja, and R. Prasad. Application of hidden markov models for signature verification. *Pattern Recognition*, 28:161–170, 1995.

[7] R. S. Kashi, W. Turin, and W. Nelson. On-line signature verification using stroke direction coding. *Optical Engineering Journal*, 35(9), 1996.

[8] F. Sinden and G. T. Wilfong. On-line recognition of handwritten symbols. AT&T Bell Laboratories Internal Document, 1994.

[9] Gilbert Strang. *Linear Algebra and its Applications.* HBJ College Publishers, Orlando, Florida, 1988.

[10] C. de. Boor. *A practical guide to splines.* Springer Verlag, 1978.

[11] J. D. Ferguson. Variable duration models for speech. In *Proc. Symp. on the Application of HMM to Text and Speech*, pages 143–179, Priceton, NJ, 1980.

[12] S. E. Levinson. Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech & Language*, 1(1):29–45, March 1986.

[13] W. Turin. *Performance Analysis of Digital Transmission Systems.* Computer Science Press, New York, 1990.

[14] G. E. Blonder, R. A. Boie, L. W. Ruedisueli, E. R. Wagner, and Y. H. Wong. Capacitive moment sensing for electronic paper. AT&T Bell Laboratories Internal Document, 1991.