A New On-Line Signature Verification Algorithm Using Variable Length Segmentation and Hidden Markov Models

Mohammad M. Shafiei and Hamid R. Rabiee
Digital Media Lab (http://www.aictc.com/dml)
Sharif University of Technology
mshafiei@mehr.sharif.edu , rabiee@sharif.edu

Abstract

In this paper, a new on-line handwritten signature verification system using Hidden Markov Model (HMM) is presented. The proposed system segments each signature based on its perceptually important points and then computes for each segment a number of features that are scale and displacement invariant. The resulted sequence is then used for training an HMM to achieve signature verification. Our database includes 622 genuine signatures and 1010 forgery signatures that were collected from a population of 69 human subjects. Our verification system has achieved a false acceptance rate (FAR) of 4% and a false rejection rate (FRR) of 12%.

1. Introduction

Biometric authentication methods including voice and fingerprint identification, face recognition, retina scan, and signature verification are becoming increasingly popular for applications ranging from access control to restricted areas to fraud prevention in financial transactions. Signature verification is of particular importance as it is the only widely accepted method for endorsing financial transactions.

An important advantage of the signature over other biometric is its long standing tradition in many commonly encountered verification tasks. It has been used for decades in civilian applications while other methods (e.g., fingerprints) still have the stigma of being associated with criminal investigation. In other words, signature verification is already accepted by the general public.

Handwritten signature verification can be divided into on-line (or dynamic) and off-line (or static) verification. On-line verification refers to a process where the signer uses a special pen called a stylus to create his/her signature that produces measurements such as pen location, speed, and pressure. Off-line verification is concerned with the verification of a signature made by a normal pen. Various different approaches to both classes have been proposed. For literature surveys, see [4], [9].

On-line signature verification schemes extract signature features that characterize spatial and temporal characteristics of a signature. The feature statistics of a training set of a genuine signature are used to build a model or template for validating further test signatures. Selecting a good model is the most important step in designing a signature verification system. Hidden Markov Model is one of the most widely used models for sequence analysis in signature verification. Handwritten signature is a sequence of vectors of values related to each point of signature in its trajectory. Therefore, a well chosen set of feature vectors for HMM could lead to the design of an efficient signature verification system.

In all verification systems, the signature to verify is compared to the prototypes of the genuine signature at disposal by means of a similarity measure, often based on Dynamic Time Warping (DTW), or a distance between the signer's model and the signature at hand [6, 13]. In the probabilistic frameworks such as HMM, the distance is actually the likelihood of the observation (the signature to verify) given a statistical model of the signer [3], [7], [11], [12].

In this paper, we have presented a new technique based on variable length segmentation of signatures in a HMM model for on-line signature verification. To achieve our goal, we have modified the algorithm in [1] to segment each signature based on its perceptually important points. Then after some preprocessing, we have associated to each segment a scale and displacement invariant feature vector. Finally, the sequence composed of such vectors is used as the observation sequence of the HMM.

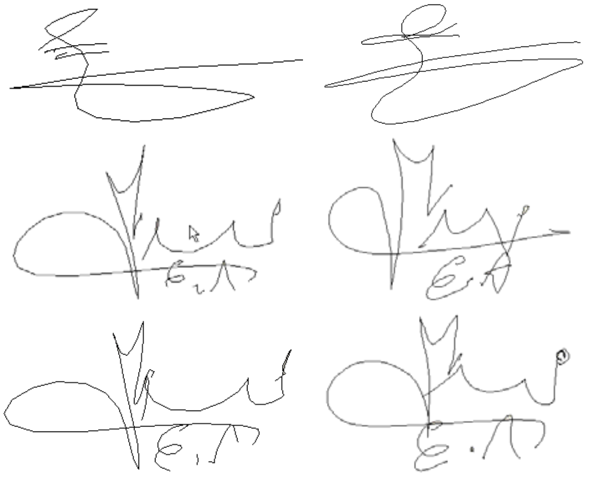
The organization of this paper is as follows. Section 2 presents the description of our data collection process. Section 3 is devoted to the detailed explanation of our feature extraction and selection algorithm. Section 4

0-7695-1960-1/03 $17.00 © 2003 IEEE

presents the HMM used in our system. Finally, the experimental results and concluding remarks are presented in Section 5.

2. Data Collection

A total of 622 genuine signatures were collected from a population of 69 human subjects including 12 women and six left handed writers. Some subjects contributed as few as 4 genuine signatures and one subject wrote his signature more than 34 times. Additionally, 1010 forgery signatures were used for testing rejection capabilities of the system. All of these forgery signatures were “skilled forgeries”. Most of forgers have tried to forge both shape and dynamics of genuine signatures. We first asked from our forgers to try to forge only shape of a signature and then in second step, we provided them the animation of each signing process to be forged and asked them to forge the signature accurately by considering both the shape and the process of signing. They could see this animation many times to learn dynamic features of signatures, and then they tried to forge these signatures. An example of our database including genuine and forgery samples is shown in Figure 1.



(a) **(b)**

Figure 1. Sample signatures from our database. Column (a) contains the genuine signatures and column (b) forgery signatures of two writers.

3. Feature Extraction and Selection

We assume that a signature can be described by a left-to-right HMM with loop, forward and skip transitions whose probabilities are re-estimated during training.

Besides the choice of the HMM-topology, the probability density function modeling of the HMM is the most important part in order to design the most appropriate models for the verification task. We have chosen continuous HMM based on a Gaussian mixture model [3].

3.1. Signature segmentation

We have used a modified version of the method originally presented by Brault et al. [1] for segmentation. This method separates curve lines of an image in areas of nonzero values spaced by areas of zero values. A nonzero domain characterizes an important area of signature and is represented by the point of maximum significance. The significance (S) of a writing point depends on the neighboring writing angle change between the selected writing point and the neighbor as shown in Figure 2 and is computed by following expression:

$$S(i) = \sum_{k=n}^{\alpha_{\max}} \left(\cos(\alpha_+(k)) \cdot \cos(\alpha_-(k)) \right)$$

i *:selected writing point*

$\alpha_+(\alpha_-)$ *:angle with the left(right) neighborhood*

k *:selected neighboring writing point*

n *:first observed neighboring writing point*

$\alpha_{\max}$ *:limiting angle*

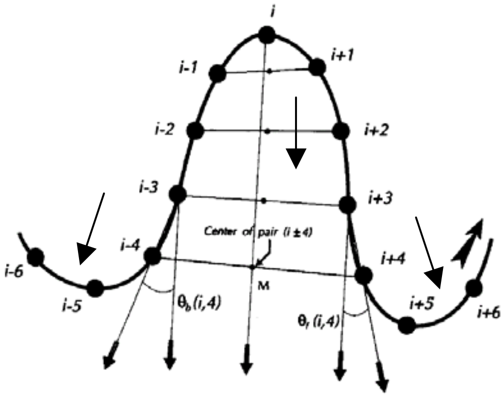


Figure 2. Significance calculation of writing point P_i

We compute $\cos(\alpha_+(k))$ by its relation to inner product of two vectors $\overrightarrow{P_iM}$ and $\overrightarrow{P_{i+k-1}P_{i+k}}$. Its value is computed as:

$$\cos(\alpha_+(k)) = \frac{\left\langle \overrightarrow{P_{i+k-1}P_{i+k}}, \overrightarrow{P_iM} \right\rangle}{\left\| \overrightarrow{P_{i+k-1}P_{i+k}} \right\| \times \left\| \overrightarrow{P_iM} \right\|}$$

Because this algorithm can’t identify endpoints of pen-down strokes as significant points, we add them to

0-7695-1960-1/03 $17.00 © 2003 IEEE

the list of significant points of a signature after applying the above algorithm to signatures.

Figure 3, shows a sample result of this segmentation algorithm on a signature from our database.

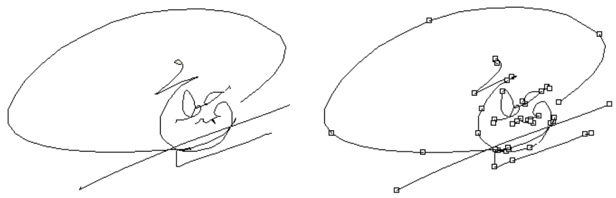


Figure 3. A sample signature and the result of our segmentation algorithm

3.2. Feature Extraction

The result of segmentation is a number of segments for each signature. Each segment is characterized by location of its most significant point in the signature, average velocity, average acceleration, average pressure, pressure variance and two angles of tangent lines to curve of segment in two segment end points (see Figure 4).

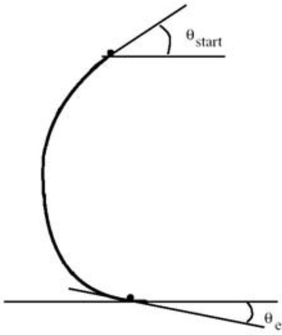


Figure4. Angle of tangent lines at two end points of a segment.

Similar to [2], in order to estimate the speed at time t, we consider the point visited at time t, $P(t) = (x(t), y(t))$ and its two neighbors in the sequence (see Figure 5), that is $P(t-1)$ and $P(t+1)$. We estimate the speed in the x and y directions respectively by

$$v_x(t) = \delta x(t) = x(t+1) - x(t-1)$$

and

$$v_y(t) = \delta y(t) = y(t+1) - y(t-1)$$

and an estimate of speed magnitude at time t is given by

$$\|v(t)\| = \sqrt{\delta x(t)^2 + \delta y(t)^2}$$

Therefore, seven parameters are extracted on each segment of the signature: four dynamic and three static.

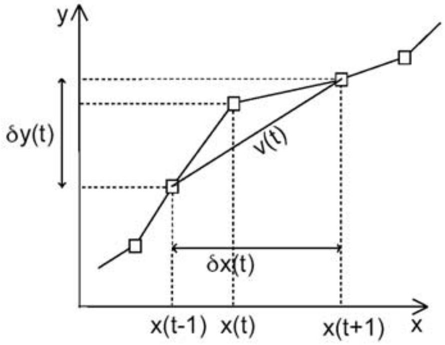


Figure 5. Computation of velocity at each point

In this way, each signature is represented by a sequence of frames with aforementioned features. This sequence is used as observation sequence for training our HMMs.

4. HMMs for Signature Verification

For each signer *i*, an HMM is trained using 5 genuine signatures of *i*. We assume mixture of ten Gaussians for emission probabilities of this HMM. The number of states of each HMM model equals 0.5 times the average number of segments that in segmentation step is computed for each signature in the training set. The EM algorithm was used during training and the Viterbi algorithm during the verification phase to approximate the likelihood of the signature [10].

Similar to previous works in [3], [6] and [11], we have used the Viterbi log-likelihood score divided by the number of segments in the signature in order to perform the verification. We calculate for each person, mean (μ) and variance (σ) of log-likelihood obtained from its HMM model divided by average segments number for his/her signatures in the training set and these values is saved beside of HMM model as a template for his/her signing process. These two values show acceptable range of log-likelihood value for genuine signatures. Then, in verification phase for each test signature, we compute log-likelihood score and this signature is accepted if the following condition for the resulting value of log-likelihood (p) is satisfied:

$$\mu - C \times \sigma \ \pi \ p \ \pi \ \mu + C \times \sigma$$

where C is a constant factor that scales the log-likelihood deviation.

5. Experimental Results and Conclusions

Two quantities can characterize the performance of a signature verification algorithm; False Rejection (FR), that is when a true signature is rejected, and False Acceptance (FA), that is when a forgery is accepted as true signature. Traditionally, the value of threshold is chosen such as to realize the Equal Error Rate(ERR).

As mentioned earlier, an HMM is trained using 5 signatures for each person. All other signatures of this person and forgeries of his/her signature and all other signatures in database were used for testing purposes. After analysis of all signatures in database, the developed system achieved a false acceptance rate (FAR) of 4% and a false rejection rate (FRR) of 12% for both random and skilled forgeries. The Equal Error Rate (ERR) of the system was 11.5%.

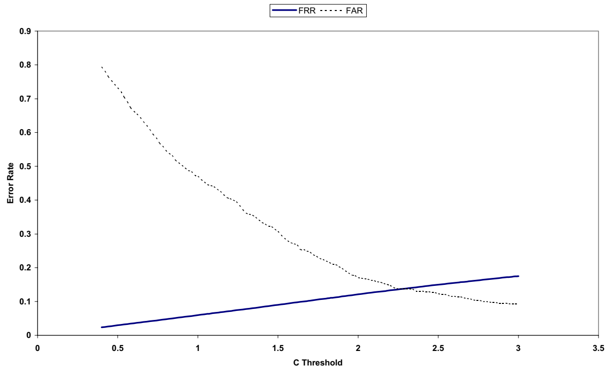


Figure 6. FAR and FRR diagram for our Database

In Figure 6, the FR/FA Error rate diagrams are shown for the system that uses 5 signatures of each person for training and all of remaining signatures for testing purpose.

Figure 7 shows the FR/FA Error rate diagrams for the system that uses 5 signatures of each person for training and all of his/her remaining signatures and simulated forgeries of his/her signature for testing purposes. In this case, the Equal Error Rate (ERR) of the system was 25.5%.

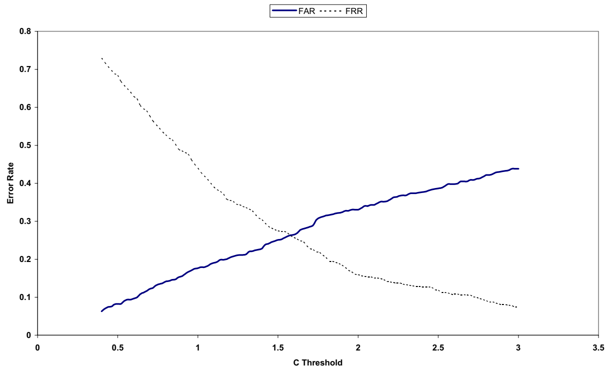


Figure 7. FAR and FRR diagram for skilled forgeries

The high FRR in this case, comparing to other works, were caused by the small number of signatures used in our training phase. In spite of using Gaussian mixtures for modeling interpersonal variability, the HMM doesn't learn adequately these variability when using small number of signatures in the training phase.

6. References

[1] J. J. Brault and R. Plamondon, “Segmenting handwritten signatures at their perceptually important points,” IEEE transaction on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 953-957, 1993.

[2] J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. L. Cun, C. Moore, E. Sackinger, and R. Shah, “Signature verification using a siamese time delay neural network,” International Journal of Pattern Recognition and Artificial Intelligence, 7(4):669–688, 1998.

[3] J. Dolfing, E. Aatrs, and J. Osterhout, “On-line signature verification with Hidden Markov Models,” In *ICDAR*, pages 1309–1312, 1998.

[4] G. Gupta, A. McCabe, "A Review of Dynamic Handwritten Signature Verification", tech. rep., James Cook University, Computer Science Dept., 1997.

[5] G. Gupta and R. C. Joyce, "A Study of Shape in Dynamic Handwritten Signature Verification," tech. rep., James Cook University of north Queensland, Computer Science Dept., 1997.

[6] R. S. Kashi, J. Hu, W. L. Nelson and W. Turin, "On-line Handwritten Signature Verification using Hidden Markov Model Features", Proceedings of ICDAR 97, pp. 253-257.

[7] R. Kashi, J. Hu, W. Nelson, W. Turin. “A Hidden Markov Model approach to online handwritten signature verification”, International Journal on Document Analysis and Recognition, 1:102-109, 1998.

[8] L. L. Lee, T. Berger, and E. Aviczer, "Reliable On-Line Human Signature Verification Systems," IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 18, no. 6, pp. 643-647, June 1996.

[9] R. Plamondon and S. N. Srihari, "On-line and Off-line Handwriting Recognition: A Comprehensive Survey," IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 22, no. 1, pp. 63-84, January 2000.

[10] L. Rabiner B. Juang, “Fundamentals of speech recognition,” Signal Processing Series. Prentice Hall, Englewood Cliffs, N. J., 1993.

[11] G. Rigoll, A. Kosmala, “A systematic comparison of online and off-line methods for signature verification with hidden markov models,” In *ICPR*, pages 1755–1757, 1998.

[12] L. Yang, B. Widjaja, R. Prasad, “Application of hidden markov models for signature verification,” *Pattern Recognition*, 28(2):161–170, 1995.