IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 2, Ver. IV (Mar - Apr.2015), PP 48-53
www.iosrjournals.org

Online Signature Verification Using Vector Quantization and Hidden Markov Model

Behrouz Vaseghi[1], Somayeh Hashemi[2]
[1,2] *(Department of Electrical Engineering ,Abhar Branch, Islamic Azad University, Abhar, Iran.)*

Abstract: *In this paper an on-line signature verification system, using vector quantization and Hidden Markov Model (VQ-HMM) is presented. After the signature acquisition, a Chebichef filter is used for noise reduction, and size and phase normalization is performed using Fourier transform. Each signature is segmented and mean velocity, acceleration and pressure are used as extracted features. K-means clustering is used for generation a codebook and VQ generates a code word for each signature. These code words are used as observation vectors in training and recognition phase. HMM models are trained using Baum Welch algorithm. In the verification phase, the forward algorithm is used. The Threshold used in the verification phase is a function of the minimum probability in training phase. Equal Error Rate obtained from this system is 14%.*
Keywords: *online signature verification, HMM, Vector Quantization, Baum Welch, K-means*

I. Introduction

NOWADAYS, the society demands secure means for person authentication. traditional authentication methods are based on the knowledge (password, Personal Identification Number numbers) or on the possession of a token (Identificator card, keys), which can be forgotten or stolen. This fact places a lot of attention in biometrics as an alternative method for person authentication and identification. Biometrics is defined in [1] as the use of physiological or behavioral characteristics for person recognition, and hence, they are not affected by the disadvantages of the traditional authentication methods since they cannot be forgotten or stolen. Biometrics can be coarsely categorized into behavioral and physiological biometrics. Physiological biometrics are based on measurable physiological traits, such as fingerprints or the iris pattern. Behavioral biometrics are based on measurements extracted from an activity performed by the user, in a conscious or unconscious way, and they are inherent to his/her own personality or learned behavior, although influenced also by the physical characteristics of the person. in this sense, behavioral biometrics have some interesting advantages, like user acceptance and cancelability, but they still lack the same level of uniqueness as physiological biometrics. Among all the biometric traits that can be categorized as pure behavioral, the signature, and the way that we sign, is the one that has the widest social acceptance for identity authentication. On the basis of the signature acquisition method, signature recognition methods can be categorized into static (or offline) and dynamic (or online) methods [2]. Offline signature verification uses the shape of the signature to authenticate the signer. Online signature verification uses dynamic characteristics of the signature (time-dependent signals) to authenticate the signer. Learning the dynamics of the real signer is a difficult task for an impostor, compared to replicate only the shape of a signature. Moreover, the use of devices with built-in pen input such as smartphones, Personal Digital Assistants, or tablet PCs has been spread in the last years. These facts have motivated great research efforts in the last decade on dynamic or online signature verification, as reviewed in [3].

II. Signatures Database And Preprocessing

2.1 data collection

By using a graphic tablet, a total of 500 genuine signatures were collected from a population of 50 human subjects including 25 men and 25 women that each of the men and women contributed 10 genuine signature samples. Additionally, 250 forgery signatures (5 forgery signatures for each person) were used for testing rejection capabilities of the system. All of these forgery signatures were “skilled forgeries”. In fact, the data acquisition procedure involves a signer and a forger. While the signer is seated and writes signatures, the forger is looking over the shoulder of the signer to capture both the image and the dynamic information of the written signature.

2.2 preprocessing

After the signatures collected, a four-order Chebychev filter with cutoff frequency 20 hertz is used for noise reduction and size and phase normalization is performed using Fourier transform [4],[5]. Then each signature is segmented based on the points where $v_y = 0$, $v_x = 0$ [6],[7] . After initial segmentation, in order to divide each person’s signatures into the same number of segments, dominant number of segments for each signature is calculated and signatures with segments different from this number are redivided based on this number (figure 1).

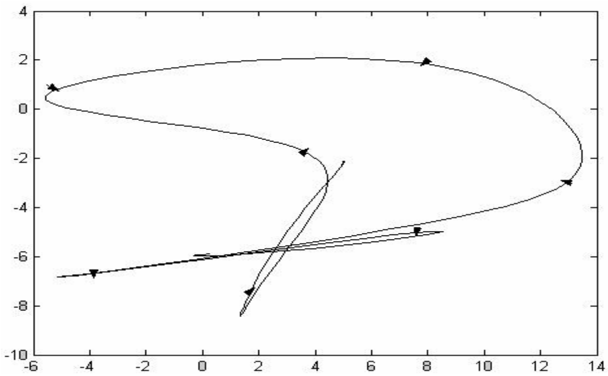


Fig.1 Result of signature segmentation

III. Feature Extraction and Vector quantization

3.1 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 2).

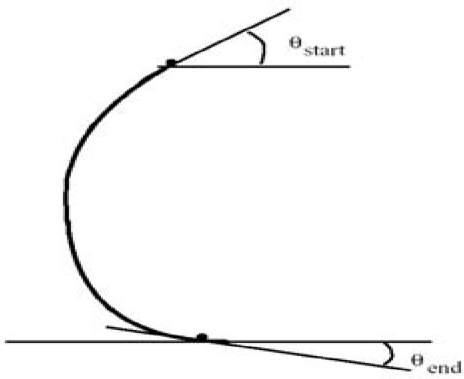


Fig.2 Angle of tangent lines at two end points of a segment

As described, signatures are captured using a pen tablet model Graphire2 Universal Serial Bus, branded by [6]. The tablet spatial resolution is 100 lines per millimeter. The precision is ±0.25 mm. The maximum detection height is 1 cm, and the sampling rate is 100 Hz. Time signals for each segment provided by the digitizer are as follows [8],[9],[10]:

1) Coordinate information: x and y.
2) Pressure: pr .
3) Angles: θ_{start} , θ_{end}

Additional dynamic features are derived from these features.

1) Velocity $v = \sqrt{(\delta_x/\delta_t)^2 + (\delta_y/\delta_t)^2} = \sqrt{v_x^2 + v_y^2}$ and logVelocity $\log V = \log(1.0 + v)$
2) Acceleration $a = \delta_v/\delta_t$.
3) Tangential angle $\varphi = \tan^{-1}(v_y/v_x)$.
4) Curvature radius $cr = (\delta_\varphi/\delta_t)/v$.
5) logPressure $\log pr = \log(1.0 + pr)$.

Therefore, eleven parameters are extracted on each segment of the signature. In this way, each signature is represented by a sequence of frames with aforementioned features.

$$p_i = [x, y, pr.\theta_{start}, \theta_{end}, v, a, \varphi, cr, \log v, \log pr]$$

3.2 Vector Quantization

A set of feature vectors extracted ($P = \{p_1, p_2, \ldots, p_n\}$) from signature segments (frames) were gathered to generate a codebook. The K-means algorithm is used to segment this vector space into C-partitions represented by a set of cluster centers [11],[12]. After generating the codebook, a given feature vector (v_i) is mapped into a membership vector $U_i = [u_{i1}, u_{i2}, \ldots, u_{ic}]$. Thus, each signature, represented by a sequence of feature vectors extracted from signature frame, is now mapped into an observation sequence of membership vectors (instead of an observation sequence of single values in the case of conventional VQ/HMM). Therefore, the Baum Welch Re-estimation algorithm (in training phase) and Forward algorithm (in recognition phase) must be used to take into account this observation sequences.

IV. Hidden Markov Model

In this method, each signature class is modeled by a single HMM. A HMM, is defined by the following parameter: [12],[13](see figure 3)

- The number of states (N) which is set for each class is proportional to the average numbers of frames in training samples in that class. The individual states are denoted as:

  $$S = \{S_1, S_2, \ldots, S_N\} \tag{1}$$

  and the state at time t as q_t .

- The number of distinct observation symbols per state (M), which is set equal to 10 in this case. we denote the individual symbols as

  $$V = \{v_1, v_2, \ldots, v_M\} \tag{2}$$

  Which are c cluster centers obtained by K- means clustering algorithm.

- The state transition probability distribution:

  $$A = \{a_{ij}\}$$

  (3)

  That $a_{ij} = p[q_{t+1} / q_t = s_i]\ , 1 \le i, j \le N$ (4)

  And $a_{ij} = 0 \quad if\ (j \prec i) or (j \succ i + \Delta)$ (5)

  The maximum number of forward jumps in each state (Δ) is chosen experimentally to be between 2 and 4 for each class during training.

- The observation symbol probability distribution :

  $$B = \{b_j(m)\} \tag{6}$$

  That

  $$b_j(m) = p[v_m at / q_t = s_j]\ , 1 \le j \le N, 1 \le m \le M \tag{7}$$

- The initial state distribution :

  $$\Pi = \{\pi_i\} \quad , 1 \le i \le N \tag{8}$$

  That

  $$\pi_i = p[q_i = s_i] = \begin{cases} 0, i \ne 1 \\ 1, i = 1 \end{cases} \tag{9}$$

- The last state distribution :

$$\Gamma = \{\gamma_i\} \quad ,1 \le i \le N \tag{10}$$

That

$$\gamma_i = p[q_T = s_i] = \begin{cases} 0, i \ne N \\ 1, i = N \end{cases} \tag{11}$$

- The set of K observation sequences (training samples) for each signature class:

$$O = \{O^{(1)}, O^{(2)}, ... O^{(K)}\} \tag{12}$$

That

$$O^{(K)} = \{O_1^{(1)}, O_2^{(2)}, ... O_{T_K}^{(K)}\} \tag{13}$$

And $O_t^{(K)}$ is the observation vector at frame t in the Kth training sample.

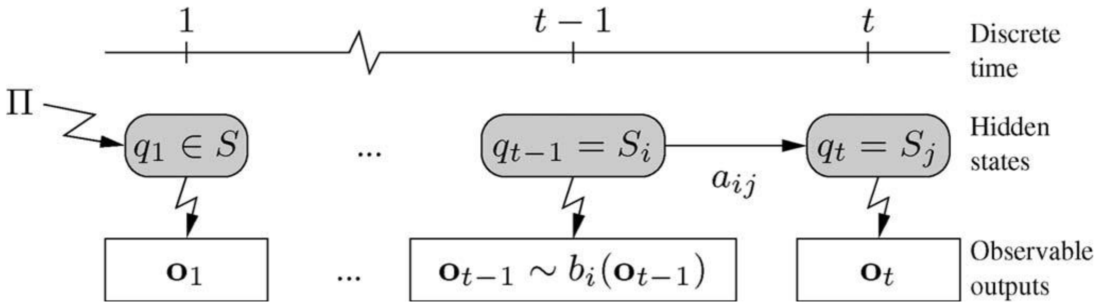


Fig. 3. HMM structure and parameters

In this way each signature in the test data set is represented as sequences of T observation and for each signature set, a separate left-right HMM is trained by these observations and Baum-Welch algorithm. The distributions Π and Γ are not re-estimated since they are predefined in left-right HMM as show in equations (8-11). For each signer i, an HMM is trained using 7 genuine signatures of i. The number of states of each HMM model equals 0.9 times the number of segments that in segmentation step is computed for each signature in the training set.

After training all of the HMMs by Baum-Welch algorithm the probability that o has been generated by signature model, ($P(o \mid \lambda_h), 1 < h < 50$) ,was computed by forward algorithm as follows:

$$\alpha_t^{(K)}(j) = \begin{cases} \pi_j . b_j .(o_t^k), t = 1 \\ \left[\sum_{i=1}^{N} \alpha_{t-1}^{(k)}(i). a_{ij} \right] b_j(O_t^{(k)}) \quad ,2 \le t \le T_k \end{cases} ,1 \le j \le N \tag{14}$$

Similarly, the backward variable for given signature sample K is calculated as:

$$\beta_t^{(k)}(j) = \begin{cases} \gamma_j , t = T_K \\ \left[\sum_{i=1}^{N} a_{ji} b_j (O_t^{(k)}) \beta_{t+1}^{(k)}(i) \right], t = T_K - 1,...,1 \end{cases} ,1 \le j \le N \tag{15}$$

Finally the observation probability is calculated as:

$$P_{K,h} = P(O^{(K)} \mid \lambda_h) = \sum \alpha_{T_K}^{(K)}(i). \gamma_i \tag{16}$$

The Threshold used in the verification phase is a function of the minimum probability in training phase

$$p_{thr} = p_{train} * 10^{-x} \tag{17}$$

That p_{train} was obtained from the train phase and x is a constant value for all signatures.

V. Experimental results

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 7 signatures for each person. All other signatures of this person and forgeries of his/her signature were used for testing purposes.

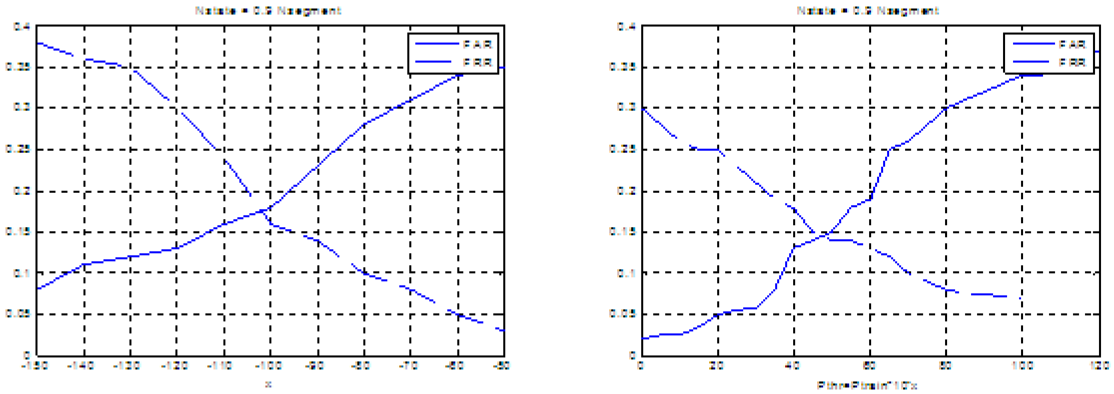
In this work effects of Threshold that used in the verification phase, number of signatures in training phase, optimum selection of signatures for training the model and number of state in the model relative to the number of segments of the signature have been studied.

In Figure 4, the FR/FA Error rate diagrams are shown for the system that uses 7 signatures of each person for training and all of remaining signatures for testing purpose. As shown in figure 4, when The Threshold is the function of the minimum probability in training phase ERR is reduced.

Figure 5 shows the FR/FA Error rate diagrams for the system that uses 4 and 7 signatures of each person for training .In this case, we observed that when we use 7 signatures for training the model, reduced the verification error.

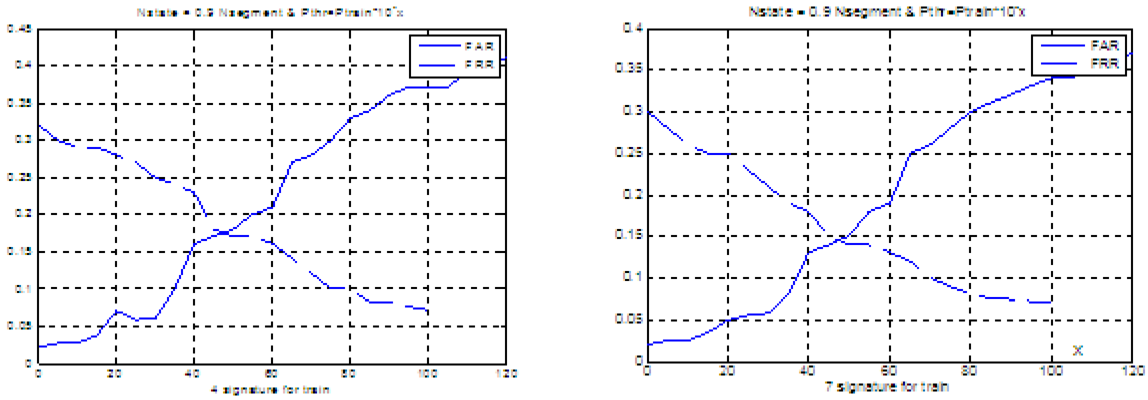
In a third experiment, we tested the verification reliability based on the signature selection for training the model.in this case we observed that the optimum selection of signatures for training the model will reduce the verification error.

Finally as shown in table 1, when we use Nstates= 0.9 Nsegments, the Equal Error Rate (ERR) of the system was 14%.



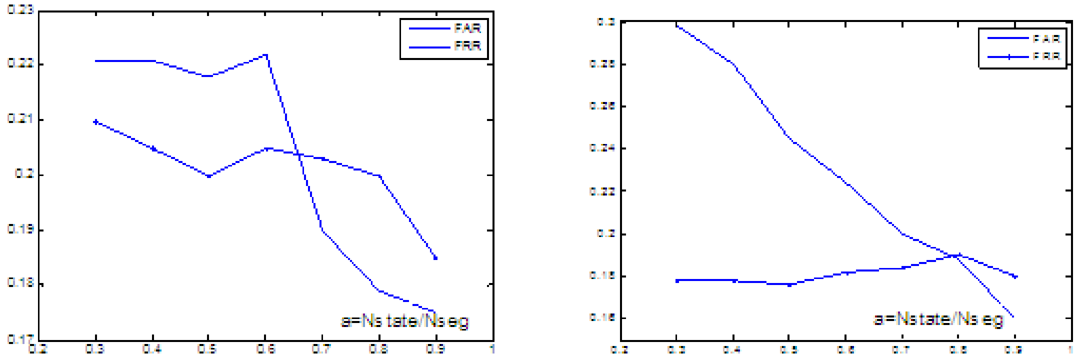
(a) (b)

Fig.4. FAR and FRR diagram: (a) Threshold is constant for all signatures and (b) threshold is a function of the minimum probability in training phase



(a) (b)

Fig.5. FAR and FRR diagram: (a) 4 signatures for train (b) 7 signatures for training



(a) (b)

Fig.5. FAR and FRR diagram: (a) random selection of signatures (b) optimum selection of signatures

Table.1. Equal Error Rate (ERR) versos the a=Nstate/Nseg

a	0.3	0.4	0.5	0.6	0.7	0.8	0.9
EER	19%	19%	0.19%	0.17%	0.17%	0.16%	0.14%
x	27	28	29	29	30	31	32

VI. Conclusion

The purpose of this paper is to design an automatic on-line signature verification system, using Vector Quantization and Hidden Markov Model (VQ-HMM). We have used a graphic tablet for signature acquisition. Our database contains 500 authentic signatures and 250 forgeries, signed by 50 persons. A Chebichef filter is used for noise reduction, and size and phase normalization is performed using Fourier transform. Each signature is segmented based on the points where $v_y = 0$, $v_x = 0$. After initial segmentation, in order to divide each person's signatures into the same number of segments, dominant number of segments for each signature is calculated and signatures with segments different from this number are redivided based on this number. Mean velocity, acceleration and pressure are used as extracted features. K-means clustering is used for generation a codebook and VQ generates a code word for each signature. These code words are used as observation vectors in training and recognition phase. HMM models are trained using Baum Welch algorithm. In the verification phase, the forward algorithm is used. The Threshold used in the verification phase is a function of the minimum probability in training phase. Equal Error Rate obtained from this system is 14%. We observed that the optimum selection of signatures for training the model, calculating the verification threshold as a function of the training probability and increasing the number of states will reduce the verification error.

Acknowledgements

This work has been supported by the Islamic Azad University, Abhar Branch, Abhar-Iran

References

[1]. N. K. Ratha, A. W. Senior, and R. M. Bolle, "Automated biometrics," in Proc. 2nd Int. Conf. Adv. Pattern Recog., Rio de Janeiro, Brazil, Mar. 2001, pp. 445–474.
[2]. B.Vaseghi , S.Hashemi 'Offline signatures Recognition system using Discrete Cosine Transform' Australian Journal of Basic and Applied Sciences, 6(12): 423-428, 2012 ISSN 1991-8178
[3]. P.O. D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 38, no. 5,pp. 609–635, Sep. 2008.
[4]. F. D. Griess, K. Jain, "Online signature verification", Project Report, Michigan State University, department of computer science
[5]. D.S. R. Plamondon, G. Lorette, "Automatic signature verification and writer identification – the state of the art", Pattern Recognition Journal, vol. 22, no. 2, pp. 107-131, 1989.
[6]. R. S. Kashi, W. L. Nelson, "Online handwritten signature verification using hidden markov model features", Proceeding of 4thInter. conf. on document analysis and recognition, pp. 253-257, 1997.
[7]. L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition", Proceedings of IEEE, vol. 77, no.2, pp. 257-286, February 1989.
[8]. J. Fiérrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target dependent score normalization techniques and their application to signature verification," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 35, no. 3, pp. 418–425, Aug. 2005.
[9]. J. Ortega-Garcia, J. Fiérrez-Aguilar, J. Martin-Rello, and J. Gonzalez-Rodriguez, "Complete signal modeling and score normalization for function-based dynamic signature verification," in Proc. 4th Int. Conf. Audio Video-Based Biometrie Person Authentication, Guildford, U.K., Jun. 2003, pp. 658–667.
[10]. B. L. Van, S. Garcia-Salicetti, and B. Dorizzi, "Fusion of HMMs likelihood and Viterbi path for on-line signature verification," in Proc. Eur. Conf. Comput. Vis., Prague, Czech Republic, May 2004, pp. 318–331.
[11]. E. Justino, A. Yacoubi, R. Sabourin and F. Bortolozzi. 2000"An off-line signature verification system using HMM and graphometric features", Proc. of the 4th International Workshop on Document Analysis Systems, pp. 211-222.
[12]. A.K. Jain, M.N. Murty, andP.J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, September 1999.
[13]. Vaseghi.B., Alirezaee.Sh., "Off-line Farsi/Arabic Handwritten word recognition using vectorquantization and hidden markov model," Proceedings of the 12th IEEE International Multitopic Conference, 978-1-4244-2824-3/08/$25.00