Australian Journal of Basic and Applied Sciences, 6(12): 423-428, 2012
ISSN 1991-8178

Off line signatures Recognition System using Discrete Cosine Transform and VQ/HMM

[1]Behrouz Vaseghi, [2]Somayeh Hashemi,

[1,2]Department of Electrical Engineering of Islamic Azad University, Abhar Branch Abhar, Iran.

Abstract: This paper presents a new recognition system for offline signatures using Discrete Cosine Transform (DCT) and pseudo 2 Dimension Discrete Hidden Markov Model (P2D-DHMM). Each signature image is scanned in two ways. One way from right to left and one way from top to bottom by a sliding window and two sets of features are extracted. 2D-DCT coefficients as features are extracted. K-means clustering is used for generation two codebooks and then by the vector quantization (VQ) two code words for each signature image are generated. These code words are used as observation vectors in training and recognition phase. Two separate Discrete HMM (each HMM for each way) is trained by Baum Welch algorithm for each set of containing image of the same signature (λ_c^v, λ_c^h). A test signature image is recognized by finding the best match (likelihood) between the image and all of the HMMs ($\lambda_c^v + \lambda_c^h$) signature models using forward algorithm. Experimental results show the advantages of using P2D-DHMM recognizer engine.

Key words: Discrete Cosine Transform; Hidden Markov Model; K-means algorithm.

INTRODUCTION

Biometric recognition has been described as automatic identification of an individual based on physiological and behavioral characteristics. Biometrics traits (signature, voice, iris, fingerprint etc.) are preferred to traditional methods (namely passwords, PIN numbers, smartcards etc.) because biometric characteristics of individual are not easily transferable; they are unique and cannot be stolen. Within biometric methods, automatics signature recognition is an important research area because of the social, legal and wider acceptance of handwritten signature as means of identification. Signature recognition systems can generally divided into two classes called online and offline. In an offline technique, signature is signed on a piece of paper and scanned to a computer system. In an on-line technique, signature is signed on a digitizer and dynamic information like speed, pressure is captured in addition to image of the signature. Recognition decision is usually based on local or global features extracted from signature under processing. Excellent recognition results can be achieved by comparing the robust model of the query signature with all the user models using appropriate classifier. Signature recognition system can be described as a two-class classification system, the classes are genuine and forgery. The aim of any signature recognition/verification system is to detect one or more category of signature forgeries namely random, simple, and skilled. Random or zero-effect forgery is any scribbled written signature of genuine signature of another person. Simple or casual signature forgery is forged by a forger who is familiar with the name of the genuine user but has no access to his/her signature samples. Skilled signature forgery is forged by a forger who has unrestricted access to one or more signature samples of the genuine user. The performance of a signature verification or recognition system is generally evaluated according to the error representation of a two-class pattern recognition problem, the error representations are False Rejected Ratio (FRR) and False Acceptance Ratio (FAR). (F. Leclerc and R. Plamondon. 1994; R. Sabourin. 1997; J. P. Edward. 2002). An effective signature recognition system must have high recognition rate. Recognition accuracy depends on the ability of the system to reduce intra variation within the signatures of the same person while increase the inter variation between signatures of different people. And this depends on techniques adopted in training and classification of signatures. It also depends on the extracted features.

Many researchers have used combination of different features and classifiers to developed signature recognition systems. Among various stochastic approaches, HMMs have proven very effective in modeling both dynamic and static signals (J. Coetzer, 2004; V.V Kohir and U.B. Desai. 1998; V. Kiani, 2010; E. Yacoubi, 2000). Previous HMM based signature recognition systems used continuous HMM topology, different number of states for users and weak features for training and classification of signature images (J. Coetzer, 2004; E. Yacoubi, 2000; E. Justino, 2001; E. Justino, 2005; E. Justino, 2000). In this paper, combination of 2D-DCT signature features and discrete HMM are incorporated to develop a robust model framework and signature classification algorithm.

Section 2 provides the description of the system, the preprocessing, frame generation, and feature extraction technique, Clustering and quantization. Hidden Markov Model and Recognition are given in section 3. Finally Experimental results and Conclusion presented in section4.

Corresponding Author: Behrouz Vaseghi, Department of Electrical Engineering of Islamic Azad University, Abhar Branch Abhar, Iran.
E-mail: Behrouz.vaseghi@gmail.com

2. The Signature Recognition System:

Off-line signature recognition system proposed in this paper is basically divided into six stages namely, data acquisition, preprocessing, frame generation, feature extraction, clustering and vector quantization, training and recognition stages as shown in Figure 1.

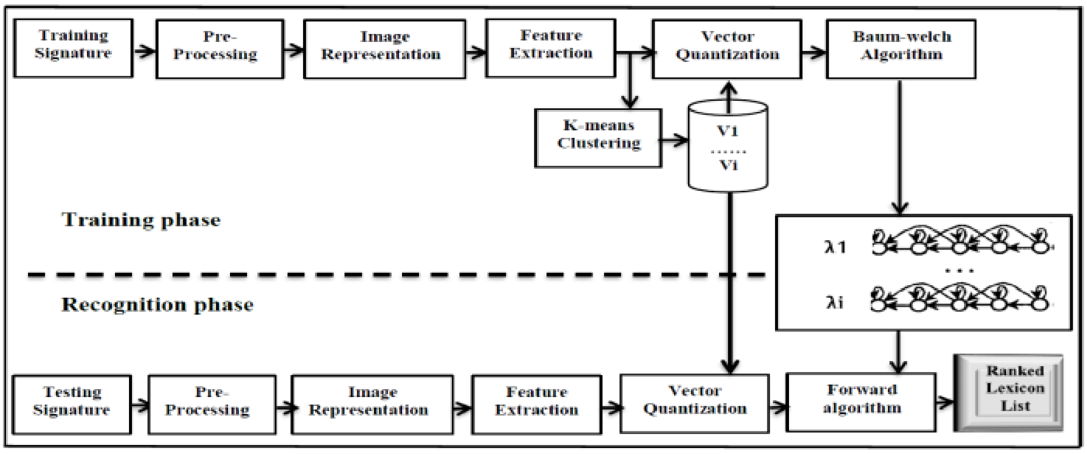

Fig. 1: An overview of proposed signature recognition system.

2.1.Input Signature Data:

The input data to the proposed system are genuine signatures of the registered users. Genuine signatures are collected from 50 men and 50 women that each of the men and women contributed 15 genuine signature samples. The signature samples are scanned by a scanner with 300-dpi resolution and in 256 gray levels. Therefore we have 1500 signature images in our database.

2.2.Preprocessing:

The preprocessing consists of the following steps:

- Binarization: The gray level image of a signature is binarized at a threshold determined by modified version of maximum entropy sum and entropic correlation methods. (Figure 2-2).
- Noise removal: The binarized image often has spurious segments which are removed by a morphological closing operation followed by a morphological opening operation both with a 3x3 window as the structure element. (Figure 2-3).
- To Surround: For decrease of the memory volume and increase the speed of the processing the binarized image is surrounded in a circumferential rectangular (Figure 2-4).

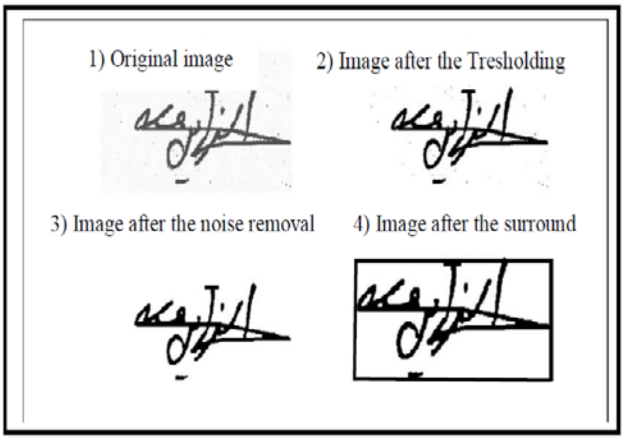

Fig. 2: An example image during preprocessing.

2.3.Frame Generation and Feature Extraction:

In this phase, the signature image is converted to an appropriate sequential form suitable for HMM recognition engine.The area of the image is divided into a set of vertical and horizontal fixed-width frames (strips) from right to left and top to bottom. The width of a frame is set to approximately twice of the average stroke width of the signature image and there is a 50% overlap between two consecutive frames. From each frame 2D-DCT coefficients are calculated and 9 higher coefficients (3*3 arrays) for generation the observation vector are selected.In these way two sets of observation vectors (one set from vertical scanning and one set from horizontal scanning) are generation.

2.4.Clustering and Quantization:

A set of feature vectors extracted ($V = \{v_1, v_2, ..., v_n\}$) from more than 100,000 signature strips (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 (A.K. Jain, 1999). After generating the codebook, a given feature vector (v_i) is mapped into a membership vector $U_i = [u_{i1}, u_{i2}, ..., u_{ic}]$. Thus, each signature image, 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.

3. Hidden Markov Model:

In this method, each signature class is modeled by two single right-left HMM. A HMM (λ_c) , is defined by the following parameter: (Mikael Nilsson).

- 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:

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

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:

$$A = \{a_{ij}\} \tag{3}$$

That

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

And

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

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 :

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

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:

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

That

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

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

In this way each signature image in the test data set is represented as two sequences of T_h, T_v observations and for each signature image set, two separate right-left DHMM is trained by these observations and Baum-Welch algorithm. The distributions Π and Γ are not re-estimated since they are predefined in right-left HMM as show in equations (5-10).

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

The forward variable for a given signature image sample K is calculated as:

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

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

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

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

$$P_{K,v} = P(O^{(K)} \mid \lambda_v) = \sum \alpha_{T_K}^{(K)}(i).\gamma_i \tag{17}$$

Finally the observation probability is calculated as:

$$P_K = P_{K,h} + P_{K,v} \tag{18}$$

And a sorted list of candidate classes is obtained.

4. Exprimental Results and Conclusion:

1500 signature images of 100 men and women that were collected in a database and used for developing pattern recognition. After applying preprocessing steps including binarization, noise removal and besieged in a circumferential rectangular each signature image is scanned from right to left and top to bottom by a sliding window and from each window 2D-DCT features is extracted. A codebook consisting 10 codeword is constructed using K-means clustering method from a pool of about 100000 feature vectors extracted from signature images. By using this codebook each signature is represented as a sequence of membership vectors. For each men/women signatures two separate right-left HMM are trained by Baum-Welch algorithm. In recognition phase the probability of generating the test image by each HMM is computed by forward algorithm. (by equ. 16,17 and 18) and a sorted list of candidate class is obtained.

For computing the recognition rate three distinct sets (A, B, C) that each set contain 5 signature images are predefined in the database usable for training and testing system. We use two of them for training and one set for testing and the mean of recognition rate in each condition is considered as recognition rate of the system in that condition. Performance of the proposed signature recognition system is shown in table 1.

Table 1: Recognition rate of proposed system.

Test	Training sets	Test sets	Recognized	Not Recognized (FAR)
1	A,B	C	93.14%	6.86%
2	B,C	A	91.1%	8.9%
3	A,C	B	92.75%	7.25%
Mean of recognition rate			92.33%	7.67%

In table 2 the performance of the propose signature recognition system is compared with the other recognition systems.

Table 2: comparisons to other signature recognition system.

Method	Recognized	Not Recognized (FAR)
[14]	92%	8%
Proposed system	92.33%	7.67%
[15]	92.5%	7.5%
[16]	99.1%	0.9%

Examinations our results and comparison with the continuous HMM show the performance of propose system in signature image recognition. In this work is used from a basic concept of HMM and avoided from the complex methods and algorithms. These are advantage for this work.

REFERENCES

Plamondon, R. and S.N. Srihari, 2000. "On-line and off-line Handwriting Recognition: A comprehensive Survey", IEE tran. on Pattern Analysis and Machine Intelligence, 22(1): 63-84.

Leclerc, F. and R. Plamondon, 1994. "Automatic Verification and Writer Identification: The State of the Art 1989-1993", International Journal of Pattern Recognition and Artificial Intelligence, 8: 643 -660.

Sabourin, R., 1997. "Off-line signature verification: Recent advances and perspectives" BSDIA', 97: 84-98.

Edward, J.P., 2002. "Customer Authentication-The Evolution of Signature Verification in Financial Institutions", Journal of Economic Crime Management, 1(1).

Coetzer, J., B.M. Herbst and J.A. Du Preez, 2004. "Off-line Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model", Eurasip Journal on Applied Signal Processing - Special Issue on Biometric Signal Processing, 100(4): 559-571.

Kohir V.V. and U.B. Desai, 1998. "Face Recognition Using A DCT-HMM Approach", Fourth IEEE workshop on Applications of Computer vision (WACV'98).

Kiani, V., R. Pourre za and H.R. Pourreza, 2010. "Offline Signature Verification Using Local Radon Transform and Support Vector Machines", International journal of Image Processing (IJIP), (3).

Yacoubi, E., E.J.R. Justino, R. Sabourin and F. Bortolozzi, 2000. "Off-line signature verification using HMMs and cross-validation", IEEE International Workshop in Neural Networks for Signal Processing, 21 International Journal of Computer Applications (0975 - 8887) 10(2): 859-868. November 2010.

Justino, E., F. Bortolozzi and R. Sabourin, 2001. "Off-line signature verification using HMM for random, simple and skilled forgeries", Proceedings of Sixth International Conference on Document Analysis and Recognition, 1: 1031-1034.

Justino, E., F. Bortolozzi and R. Sabourin, 2005. "Comparison of SVM and HMM classifiers in the off-line signature verification", Pattern Recognition Letters, pp: 1377-1385.

Justino, E., 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.

Jain, A.K., M.N. Murty and P.J. Flynn, 1999. "Data Clustering: A Review", ACM Computing Surveys, 31(3).

Mikael Nilsson, "First order Hidden Markov Model Theory and implementation issues"Research Report Department of Signal Processing Blekinge institute of Tecnology Sweden. February, 20

Camino, J., C. Travieso, C. Morales and M. Ferrer, 1999. "Signature Classification by Hidden Markov Model"; 5-7: 481-484.

Fard, M.M., N. Mozayani, 2008. "A new on-line signature verification by Spatio-Temporal neural network"; 17-20 June 2008 Page(s): 233-235.

Zois, E.N., A.A. Nassiopoulos, V. Anastassopoulos, 2005. "Signature verification based on line directionality"; 2-4 Page(s):34.