Off-line handwritten signature contours analysis for identification

J. FRANCISCO VARGAS(1)(2), CARLOS M. TRAVIESO(1), MIGUEL A. FERRER(1), JESUS B. ALONSO(1)
(1)Department of Signals and Communications
Universidad de Las Palmas de Gran Canaria
Campus de Tafira, SPAIN
{ctravieso,mferrer,jalonso}@dsc.ulpgc.es
http://www.gpds.ulpgc.es
(2)Department of Electronic Engineering.
Universidad de Antioquia, COLOMBIA
jfvargas@udea.edu.co

Abstract: - An offline handwritten signature classification approach is presented. The signature is parameterized using the radius values for contour points in the polar coordinates space. In order to determine the optimum outline for the classification task, nearest and faraway points were took into account. HMM and SVMLight model are analyzed in this work. In order to determine the effect of the feature vector size on the performance system, experiments were carried out. Changes in feature vector size corresponding to both 20% increasing and 20% reducing with respect to initial size were analyzed. HMM and SVM model trained for early experiments were take into account for this last stage too. Results show that working with feature vector containing 360 elements offers better performance. Selecting faraway points offers better results, and using SVMLight model with polynomial kernel presents the best accuracy values obtained (99,02%) outperforming HMM model (88.52%).

Key-Words: - Automatic Offline Signature Classification, Hidden Markov Model, Support Vector Machine.

1 Introduction

Signature classification is one of the most important research areas in the field of biometric authentication. It has several applications such as personal verification for access control, banking applications, electronic commerce, etc [1,2,15].

Approaches for verification systems can be divided in two groups: on-line systems [3,4] for which the signature signal is captured during the writing process, so dynamic information (pressure, velocity, etc) is available, and off-line systems for which the signature is written on a sheet and then is scanned. Subsequently, from the scanned image, the usual step is parameterize their geometric structure as previous stage to their recognition [5, 6, 7].

In this paper, we present a simple and fast handwritten signature classification approach based on estimating the optimum outline. The vector containing signature contour, initially in Cartesian coordinates, is transformed to polar coordinates, then we analyzed the problem which one angle value may contains several angles values. Initially, we extract 360 values for both angles and radius feature vectors. We compare a Hidden Markov Model (HMM) [8, 9] and a Support Vector Machine SVMLight [10] model based approaches for off-line signature Classification.

Hidden Markov Models have attracted the attention of many researchers in the pattern recognition area. HMM provides a good probabilistic representation of sequences having large variations, in this work, a signature is modeled by a left-to-right HMM, The topology only authorizes transitions between each state to itself and to its immediate right-size neighbors.

SVMs have shown great capabilities in solving various classification problems. In many applications, SVMs have outperformed many other machine learning methods and have established themselves as a powerful tool for classification problems. When performing a pattern recognition task, the SVM first maps the input data into a high-dimensional feature space and then finds an optimum separating hyperplane to maximize the margin between two classes in this high dimensional space.

In the classification process, a given signature is assigned as belonging to an author depending on whether the matching score by the associated model is above a preset threshold. To examine the proposed approach, a signature database containing signature samples from 40 individuals, 24 repetitions from each person.

In order to determine the effect of the feature vector size on the performance system, experiments are carried out. We change the feature vector size by increasing and reducing 20% initial vector length. HMM and SVM model trained for early experiments were take into account for this last stage too.

In the rest part of this paper, Section 2 presents the feature extraction from a signature image and describes the signature database used, Section 3 discusses the classification task, and describes models take into account. Section 4 presents our experimental results. Finally, section 5 gives conclusions of the present work.

2. Feature Extraction

The geometrical features proposed by this paper are based on two vectors which represent stroke distribution in polar coordinates.

2.1 Outline Detection and Representation

The outline is calculated by means of morphological operations as is shown in Fig. 1: First, we apply a dilatation in order to reduce the signature variability and, afterward, a filling operation is applied to simplify the outline extraction process. When several objects are detected after filling, a horizontal dilatation is performed until all the objects are connected.

The outline is represented as a sequence of its Cartesian coordinates $(X_t, Y_t)_{t=1}^{T}$, T being its length. This sequence follows the contour counterclockwise and starts in the point $(X_1, Y_1) = C_x$, $\max\left(Y_t \big|_{X_t = C_x}\right)$, (C_x, C_Y) being the geometric center of the outline.

We transform the outline vector to Polar Coordinates, so radius and angle values were calculated. Figure 2 shows that some angles values could contain more than one radius values, that is, if we draw a line from (C_x, C_Y) point to some point outside, several points of the outline will be crossed.

In order to determine the optimum outline for the classification task, in this work, we take into account only the nearest and faraway points.

In order to know which point the classification stage works better, we carried out some probes and the results are presented in the Section 4

2.2 Signature Database

The database we used contains data from 40 individuals: 24 genuine signatures for each individual. The repetitions of each genuine signature were collected using black or blue ink on white A4 sheets of paper featuring two different box sizes: the first box is 5 cm wide and 1.8 cm high and the second box is 4.5 cm wide and 2.5 cm high.

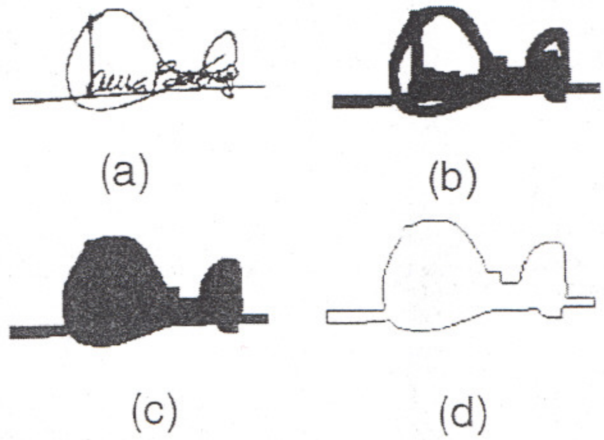

Fig. 1. **(a) Original. (b) Dilated. (c) Filled. (d) Outline of the signature.**

Half of the specimens were written in each type of boxes. The 24 genuine specimens of each signer were collected in single day writing sessions. All the signature images in black and white and noise cleaned were saved in PNG format. A freely distributed version of the described database is available from: http://www.gpds.ulpgc.es/download/index.htm.

3 Classifiers

In the present paper, two classification models were analyzed, HMM and SVMLight model are described as follow:

3.1 The HMM signature model

In this case, a signature is modeled by a left-to-right HMM. The number of states in each signer´s HMMs signature is 35. The topology only authorizes transitions between each state to itself and to its immediate right-size neighbors. Forward-Backward algorithm, the Viterbi algorithm, and Baum-Welch algorithm were used to solve training and classification problems. The initialization was carried out using equal-occupancy method [11].

The K-means algorithm is used during train to create the multilabeling VQ, which make a soft decision about which code words are closes to the input vector. Therefore, a vector containing the relative closeness of the 10 closest code words to input is generated. The components of the features vector p_t are considered independent, so the three sets of observations symbols contains 32 symbols. The number of symbols was judged a good trade-off between the number of centers describing the space of the signers and the number of observation vectors collected in the training database. 32 symbols were experimentally established.

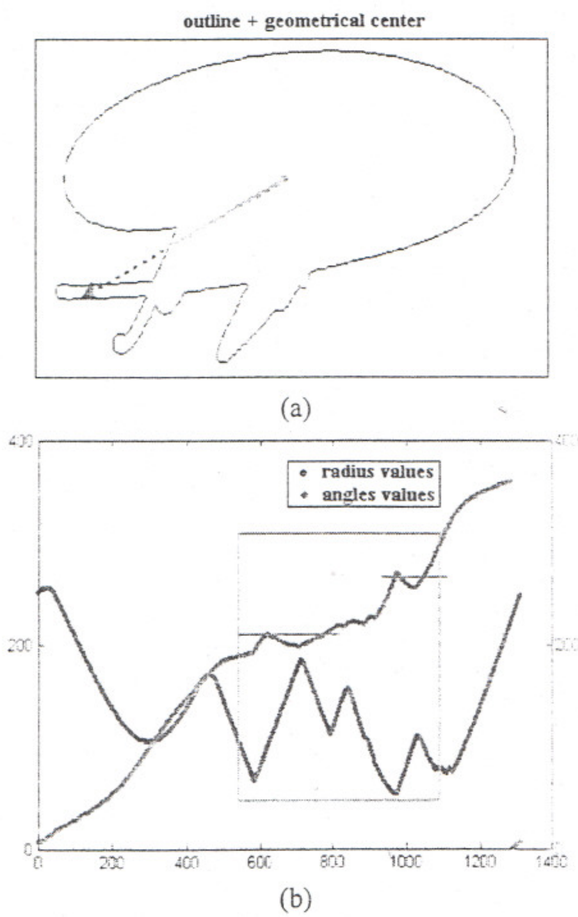

Fig. 2. (a) Signature outline. (b) Radius and angle vectors, box shows the angles values repeated problem.

To verify a signature, the log likelihood oh the two HMM that model the signature is obtained [12]. The usual normalization technique necessary to consolidate the score are not necessary. If the final score is greater than a threshold, the signature is accepted. The HHM model used is the GPDShmm toolbox which is freely available from http://www.gpds.ulpgc.es/download/index.htm [9].

3.2 Support Vector Machine signature model

Support Vector Machines (SVM) [13] provides generally better generalization performance when the amount of data is small and robustness to the quantization effects. [14].

Roughly, the principle of SVM relies on a linear separation in the feature space where the data have been previously mapped in order to take into account the eventual nonlinearities of the problem. In order to achieve a good level of generalization performance, the distance (margin) between the separator hyperplane and the data is maximized. In the Structural Risk Minimization principle, Vapnik has proven that maximizing the margin means, in fact, minimizing the VC-dimension of the linear separation model, which has been shown to be a good way to reduce the generalization risk.

To generalize the linear case, one can project the input space into a higher-dimensional space in the hope of a better training class separation. In the case of SVM, this is achieved by using the so-called kernel trick. In essence, it replaces the inner product used to calculate the distance between the input and the separator hyperplane with a kernel function K. Among the commonly used kernel functions are the polynomial and the RBF kernel.

The software used to train and test the model is the SVMLigth, which can be downloaded free of charge from http://www.kernel-machines.org/software.html. To verify a sequence, the SVM of a signature calculates the distance of the input sequence to the separator hyperplane. If the distance is greater than a threshold, the signature is accepted.

4 Results

With a view to demonstrating the performance of the proposed approach worked out using basic feature vectors from the signature contour, we provide results with two different classifiers: HMM and SVMLight models.

Figure 3 shows the classification accuracy obtained selecting nearest or faraway points as feature vectors and for different size of database. We used the HMM model trained and described in section 3.1, and as can be seen, the best results are obtained working with the faraway points. Respect to training and testing datasets construction, a half (chosen randomly) of the 24 samples available per each individual was used as train set, and the other half as test set. Table 1 presents numerical results for this probe.

Table 2 shows classification results for different values of the trade-off between training error and margin for the SVMLight model. As can be seen, the model with C=10 presents maximum accuracy with minimal standard deviation after cross validation procedure involving 5 iterations.

Table 1
Accuracy classification achieved for different number of individuals, using HMM.

	Type of point	
# signers	Nearest	Faraway
10	72.13	84.91
20	73.96	87.05
30	73.72	86.17
40	73.85	88.52

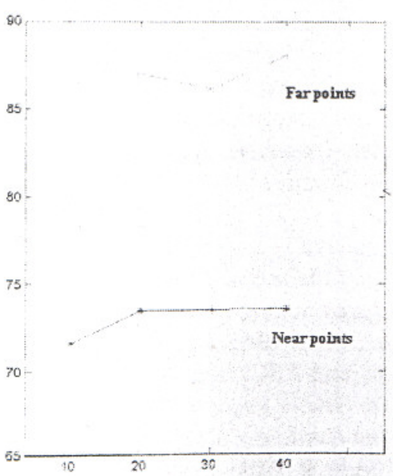

Fig. 3. Accuracy classification achieved for

Table 2
SVMLight model accuracy for different values of parameter C

	SVMLight	
C value	accuracy	std. Dev.
1	98.22	0.92
10	98.89	0.73
100	98.59	1.18

Table 3
SVMLight model accuracy for different type of kernel function

	SVMLight		
Kernel type	% Acc.	% Precis.	% Recall
linear	98.89	83.04	72.44
polynomial	99.02	83.43	79.78

Table 3 presents accuracy obtained for both linear and polynomial kernel for SVMLight model. The data in term of precision and recall values are reported too. RBF kernel was probed but results show that SVM model trained presents very low generalization ability.

In order to determine the effect of the feature vector size on the performance system, experiments were carried out. Changes in feature vector size corresponding to both 20% increasing and 20% reducing with respect to initial size (360 values), so two new feature vectors were analyzed, one with 288 elements and the other containing 432 elements from signature contour. HMM and SVM model trained for early experiments were take into account for this last stage too.

Table 4
SVMLight model accuracy for different feature vector size

	SVMLight		
Vector Size	% Acc.	% Precis.	% Recall
288	$98.92_{0.81}$	$80.02_{16.74}$	$83.04_{15.38}$
360	$99.02_{0.65}$	$83.43_{14.65}$	$79.78_{16.96}$
432	$98.94_{0.72}$	$82.89_{16.01}$	$75.94_{18.57}$

Table 4 shows accuracy obtained for SVMLight model trained with different feature vector size. Subscript numbers represent standard deviation for each case. Feature vector containing 360 elements presented the better performance with de minimum standard deviation. Highest precision score is reported for this vector size too.

Table 5
HMM model accuracy for different feature vector size

	HMM	
Vector Size	% Acc.	Std. Dev.
288	88.11	2.63
360	88.52	1.05
432	87.17	2.51

Table 5 shows accuracy obtained for HMM model trained with different feature vector size. Again, feature vector containing 360 elements presents the better performance with de minimum standard deviation but lower than reported for SVM model.

5 Conclusions

This work presented a method to determine the optimum contour of handwritten signature for offline classification. The initial outline vector in Cartesian coordinates was transformed to polar coordinates, and looking for and optimum outline for the classification task, the nearest and faraway points was analyzed separately as unique feature vectors. Experimental results show that using only faraway points as optimum contour features outperforms nearest points.

The proposed features have been tested using HMM and SVM models. The results show that SVMLight model works better then HHM for this task. SVM model report an accuracy classification rate of 99% for polynomial kernel and 98% for linear kernel. Initial probes were carried out for a feature vector containing 360 elements from signature contour, very similar at all but in terms of precision and recall, polynomial kernel improve linear kernel performance.

For different feature vector size, the experiments showed that even reducing or increasing the number of elements, system performance remains better for a feature vector containing 360 elements. For all cases, SVM model with polynomial (cubic) kernel presents the better performance.

Acknowledgment

F. Vargas, is supported by the high level scholarships program, Programme Alßan **No. E05D049748CO**.

References:

[1] Plamondon R., Srihari SN., *Online and Off-Line Handwriting Verification: A Comprehensive Survey,* IEEE. Transactions on Pattern Analysis & Machine Intelligence vol.22, no.1, 2000.

[2] Leclerc F. and Plamondon R., *Automatics Signature Verification: The State of the Art,* 1989–1993. Pattern Verification and Artificial Intelligence, special issue signature verification, vol. 8, no. 3, 1994, pp. 643–660.

[3] Rhee Taik H., Cho Sung J. and Kim J. H., *On – Line Signature Verification Using Model–Guided Segmentation and Discriminative Feature Selection for Skilled Forgeries*, Proceedings Sixth International Conf. In Document Analysis and Verification, 2001 pp. 645–649.

[4] Lee L., Berger T. and Aviczer E, *Reliable On–Line Human Signature Verification Systems*, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 18, no. 6, 1996, pp. 643–647.

[5] Miguel A. Ferrer, Jesús B. Alonso, and Carlos M. Travieso, *Offline Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic*, IEEE Transaction on Pattern Analysis & Machine Intelligence, nol. 27, no. 6, 2005.

[6] J. Edson, R. Justino, F. Bortolozzi, and R. Sabourin, "*An Off-Line Signature Verification Using HMM for Random, Simple and Skilled Forgeries*," Proc. Sixth Intl Conf. Document Analysis and Recognition, pp. 1031-1034, 2001.

[7] J. Edson, R. Justino, A. El Yacoubi, F. Bortolozzi, and R. Sabourin, "*An Off-Line Signature Verification System Using HMM and Graphometric Features*," Proc. Fourth Int'l Workshop Document Analysis Systems, pp. 211-222, 2000.

[8] Rabiner L. R., *A tutorial on hidden Markov models and selected application in speech recognition*, Proc. IEEE 77, pp. 267-293. 1989.

[9] S. David, M.A. Ferrer, C.M. Travieso, J.B. Alonso, "*gpdsHMM: A Hidden Markov Model Toolbox in the Matlab Environment*," CSIMTA, Complex Systems Intelligence and Modern Technological Applications, pp. 476-479, 2004.

[10] Thorsten Joachims, *Making Large-Scale SVM Learning Practical.* LS8-Report, 24, Universität Dortmund, LS VIII-Report, 1998.

[11] M.A. Ferrer, I. Alonso, and C: Travieso, *Influence of the Initialization and Stop Criteria on HMM Based Recognizers*, Electronic Letters of IEE, vol. 36, pp. 1165-1166, 2000.

[12] A. Ross and A.K. Jain, *Multimodal Biometrics: An Overview*, Proc. XII European Signal Processing Conf., pp. 1221-1224, 2004.

[13] K. Müller, S. Mika, et al., *An introduction to Kernel-Based Learning Algorithms,* IEEE. Trans. Neural networks, vol. 12, no. 2 , pp. 181-201, 2001.

[14] D. Anguita, A. Boni, and S. Ridella, *A Digital Architecrture for Support Vector Machines: Theory, Algorithm, and FPGA Implementation*, IEEE Trans. Neural Networks, vol.14, no. 5, pp. 993-1009, 2003.

[15] R. Plamondon, "The Handwritten Signature as a Biometric Identifier: Psychophysical Model and System Design," Proc. European Convention Security and Detection, 1995, pp. 23-27, May 1995.