Off-line signature verification and recognition by Support Vector Machine

OFF-LINE SIGNATURE VERIFICATION AND RECOGNITION BY SUPPORT VECTOR MACHINE

Emre Özgündüz,Tülin Şentürk and M. Elif Karslıgil

Computer Engineering Department, Yıldız Technical University
Yıldız , Istanbul, Turkey, phone: + (90) 212 3273673, fax: + (90) 212 3273673,
email: emre_ozgunduz@yahoo.com, tulinsenturk@hotmail.com, elif@ce.yildiz.edu.tr

ABSTRACT

In this paper we present an off-line signature verification and recognition system using the global, directional and grid features of signatures. Support Vector Machine (SVM) was used to verify and classify the signatures and a classification ratio of 0.95 was obtained. As the recognition of signatures represents a multiclass problem SVM's one-against-all method was used. We also compare our methods performance with Artifical Neural Network's (ANN) backpropagation method.

1. INTRODUCTION

Signatures are composed of special characters and flourishes and therefore most of the time they can be unreadable. Also intrapersonal variations and interpersonal differences make it necessary to analyse them as complete images and not as letters and words put together [1]. As signatures are the primary mechanism both for authentication and authorization in legal transactions, the need for research in efficient automated solutions for signature recognition and verification has increased in recent years.

Recognition is finding the identification of the signature owner. Verification is the decision about whether the signature is genuine or forgery. In this decision phase the forgery images can be classified in three groups: (i) random, (ii) simple, (iii) skilled [2]. Random forgeries are formed without any knowledge of the signer's name and signature's shape. Simple forgeries are produced knowing the name of the signer but without having an example of signer's signature. Skilled forgeries are produced by people looking at an original instance of the signature, attempting to imitate as closely as possible.

SRVS (Signature Recognition and Verification System) is often categorized in two major classes: on-line SRVS and off-line SRVS. The difference of on-line and off-line lies in how data are obtained. In the on-line SRVS data are obtained using an electronic tablet and other devices. In the off-line SRVS images of the signatures written on a paper are obtained using a scanner or a camera [1].

There are several implementations for signature recognition and verification. Justino, Bortolozzi and Sabourin proposed an off-line signature verification system using Hidden Markov Model [3]. Zhang, Fu and Yan (1998) proposed handwritten signature verification system based on Neural 'Gas' based Vector Quantization [4]. Vélez, Sánchez and Moreno proposed robust off-line signature verification system using compression networks and positional cuttings [1]. Arif and Vincent (2003) concerned data fusion and its methods for an off-line signature verification problem which are Dempster-Shafer evidence theory, Possibility theory and Borda count method [5]. Chalechale and Mertins used line segment distribution of sketches for Persian signature recognition [6]. Sansone and Vento (2000) increased performance of signature verification system by a serial three stage multi-expert system [2].

In this paper an off-line SRVS using SVM is proposed. SVM is a new learning method introduced by V. Vapnik et al. [7, 8]. SVMs are very universal learners. With a set of examples from two classes, a SVM finds the hyperplane, which maximizes the distance from either class to the hyperplane and separates the largest possible number of points belonging to the same class on the same side. Therefore the misclassification error of data both in the training set and test set is minimized.

Although in their basic form, SVMs learn linear threshold functions, in nonlinear case, they can be used to learn polynomial classifiers, radial basis function (RBF) nets, multi layer perceptron and the like by applying appropriate kernel functions. The dimensionality of the feature space does not have a direct relation to their ability to learn. In other words, SVMs measure the complexity of hypotheses according to the margin, which separates the data. Thus, even with many features present, we can apply SVMs if input data is separable with a wide margin using functions from the hypothesis space [9].

The system we introduced is divided into two major parts: (i) Training signatures, (ii) Verification or recognition of given signature. The block diagram of the system is given in Figure 1.

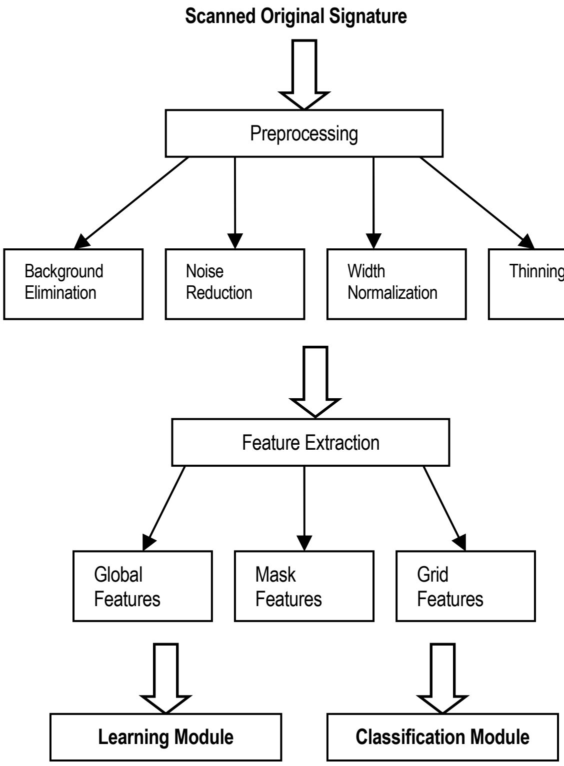


Fig. 1 Block diagram of proposed system

2. PREPROCESSING

The preprocessing step is applied both in training and testing phases. Signatures are scanned in gray. The purpose in this phase is to make signatures standard and ready for feature extraction. The preprocessing stage includes four steps: Background elimination, noise reduction, width normalization and skeletonization. The preprocessing steps of an example signature are shown in Fig. 2.

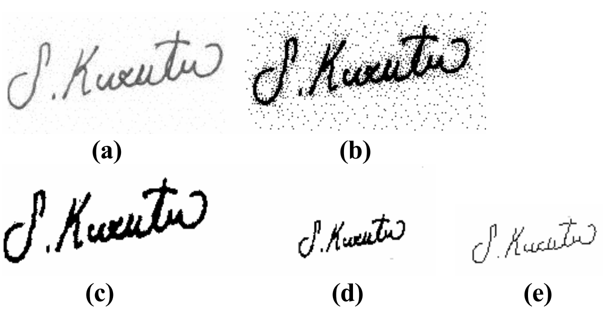


Fig 2. Preprocessing steps: (a) scanning, (b) background elimination, (c) noise reduction, (d) width normalization, (e) thinning applied signatures.

2.1 Background Elimination

Data area cropping must be done for extracting features. P-tile thresholding was chosen to capture signature from the background. After the thresholding the pixels of the signature would be "1" and the other pixels which belong to the background would be "0".

2.2 Noise Reduction

A noise reduction filter is applied to the binary image for eliminating single black pixels on white background. 8-neighbors of a chosen pixel are examined. If the number of black pixels is greater than number of white pixels, the chosen pixel will be black otherwise it will be white.

2.3 Width Normalization

Signature dimensions may have intrapersonal and interpersonal differences. So the image width is adjusted to a default value and the height will change without any change on height-to-width ratio. At the end of width normalization width dimension is adjusted to 100.

2.4 Thinning

The goal of thinning is to eliminate the thickness differences of pen by making the image one pixel thick. In this system Hilditch's Algorithm is used.

3. FEATURE EXTRACTION

Extracted features in this phase are the inputs of training phase. The features in this system are global features, mask features and grid features. Global features provide information about specific cases of the signature shape. Mask features provide information about directions of the lines of the signatures. Grid features provide overall signature appearance information. The feature extraction steps of an example signature are shown in Fig. 3.

3.1 Global Features

Signature area is the number of pixels which belong to the signature. This feature provides information about the signature density.

Signature height-to-width ratio is obtained by dividing signature height to signature width. Signature height and width can change. Height-to-width ratios of one person's signatures are approximately equal.

Maximum horizontal histogram and maximum vertical histogram: The horizontal histograms are calculated for each row and the row which has the highest value is taken as maximum horizontal histogram. The vertical histograms are calculated for each column and the column which has the highest value is taken as maximum vertical histogram.

Horizontal and vertical center of the signature are calculated using the formulas in Eq. 1 [10].

$$\text{Center}_x = \frac{\sum_{x=1}^{Xmax} x \sum_{y=1}^{Ymax} b[x][y]}{\sum_{x=1}^{Xmax} \sum_{y=1}^{Ymax} b[x][y]} \quad \text{Center}_y = \frac{\sum_{y=1}^{Ymax} y \sum_{x=1}^{Xmax} b[x][y]}{\sum_{x=1}^{Xmax} \sum_{y=1}^{Ymax} b[x][y]} \qquad (1)$$

Local maxima numbers of the signature: The number of local maxima of the vertical and horizontal histogram is calculated.

Edge point numbers of the signature: Edge point is the pixel which has only one neighbour, which belongs to the signature, in 8-neighbor.

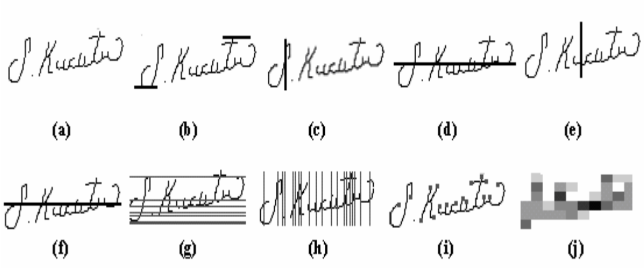


Fig. 3 Feature extraction steps: (a) preprocessed signature and (b) height, (c) maximum vertical histogram, (d) maximum horizontal histogram, (e) horizontal center, (f) vertical center, (g) horizontal local maxima numbers, (h) vertical local maxima numbers, (i) edge points, (h) grid features of the signature.

3.2 Mask Features

Mask features provide information about directions of the lines of the signatures. The angles of the signatures have interpersonal differences. In this system 8 different 3x3 mask features are used. Each mask is taken all around the signatures and the number of 3x3 parts of the signature, which are same with the mask, is calculated.

3.3 Grid Features

Grid features are used for finding densities of signature parts [10]. In this system 60 grid features are used. Signature is divided into 60 equal parts and the image area in each divided part is calculated.

4. SIGNATURE DATABASE

For training and testing of the signature recognition and verification system 1320 signatures are used. The signatures were taken from 70 persons.

For training the system 40 persons' signatures are used. Each of these persons signed 8 original signatures; other 30 persons imitated the signatures. For each person 4 forgery signatures are signed. In the training set the total number of signatures is 480 (12 x 40).

In order to make the system robust, signers were asked to use as much as variation in their signature size and shape and the signatures are collected at different times without seeing other signatures they signed before.

For testing the system, another 320 genuine signatures and 320 forgery signatures are taken from the same 40 persons in the training set.

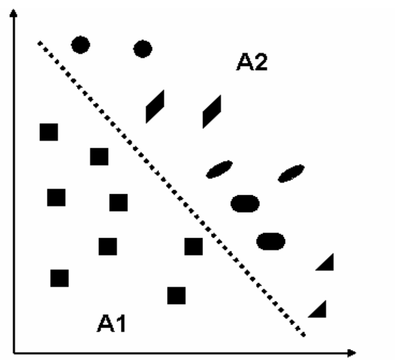
5. TRAINING AND TESTING

The recognition phase consists of two parts, training and testing respectively which is accomplished by SVM.

5.1 Training Phase

Signature recognition is a multi-class problem. Since SVM supports only two-class recognition, a multi class system can be constructed by combining two class SVMs. We used one-against-all technique to classify between each class and all the remaining classes [11] (Fig 4).

In the training phase, for each person we chose 8 positive (genuine) and 82 (39 x 2 + 4) negative (forgery) examples. 78 signatures of 82 forgery signatures are the random forgeries which are taken from other persons in the training set. Other 4 forgery signatures are skilled forgeries. Each example includes 77 (9 + 8 + 60) extracted global, mask and grid features which are normalized into [0, 1] space. Possible kernel options are linear, polynomial, radial basis function and sigmoid. In this system radial basis function is used which gave the best results.



CLASS A1	CLASS A2
Signer 1 (8 genuine)	**Signer 2 (4 skilled forgeries)** **Signer 3 (2 random forgeries)** **Signer 4 (2 random forgeries)** . . . **Signer 41 (2 random forgeries)**

Fig. 4 The structure of SVM used in this system

5.2 Testing Phase and Results

Testing can be done in two different ways which are verification and recognition. Verification is the decision about whether the signature is genuine or forgery. Recognition is the process of finding the identification of the signature owner.

5.2.1 Verification

In this system for each person 8 original and 8 forgery signatures are tested. The possible cases in verification are true acceptation (TA), false rejection (FR), true rejection (TR), false acceptation (FA). In verification phase the same 77 features are used.

The verification results of SVM and ANN's Backpropagation method are given in Table 1.

Table 1 Comparison of the verification results

	TAR	FRR	TRR	FAR
SVM	0.98	0.02	0.89	0.11
ANN	0.78	0.22	0.84	0.16

5.2.2 Recognition

In this step only the original signatures are used for recognition. Table 2 shows the recognition performances of SVM and ANN (Artificial Neural Network).

Table 2 Comparison of the recognition results

	True Classification Ratio	False Classification Ratio
SVM	0.95	0.05
ANN	0.75	0.25

6. CONCLUSION

We proposed and presented a new off-line signature verification and recognition technique which is based on global, mask, grid features of signatures and SVM. The training set was prepared using one-against-all approach. The results show that SVM outperforms ANN in both verification and recognition processes. Carefully chosen discriminating features of signatures combined with the use of SVM made our system more powerful compared to other existing systems both in terms of success ratio and ease of implementation and optimized run time.

REFERENCES

[1] J. F. Vélez, Á. Sánchez , and A. B. Moreno, "Robust Off-Line Signature Verification Using Compression Networks And Positional Cuttings", Proc. 2003 IEEE Workshop on Neural Networks for Signal Processing, vol. 1, pp. 627-636, 2003.

[2] Sansone and Vento, "Signature Verification: Increasing Performance by a Multi-Stage System", Pattern Analysis & Applications, vol. 3, pp. 169–181, 2000.

[3] E. J. R. Justino, F. Bortolozzi and R. Sabourin, "Off-line Signature Verification Using HMM for Random, Simple and Skilled Forgeries", ICDAR 2001, International Conference on Document Analysis and Recognition, vol. 1, pp. 105--110. 2001

[4] B. Zhang, M. Fu and H. Yan, "Handwritten Signature Verification based on Neural 'Gas' Based Vector Quantization", IEEE International Joint Conference on Neural Networks, pp. 1862-1864, May 1998.

[5] M. Arif and N. Vincent, "Comparison of Three Data Fusion Methods For An Off-Line Signature Verification Problem", Laboratoire d'Informatique, Université de François Rabelais, 2003

[6] A. Chalechale and A. Mertins, "Line Segment Distribution of Sketches for Persian Signature Recognition", IEEE Proc. TENCON, vol. 1, pp. 11–15, Oct. 2003

[7] C.Cortes and V.Vapnik, "Support-vector networks. Machine Learning", vol. 20, pp. :273-297, Nov. 1995.

[8] V.N.Vapnik, "The Nature of Statistical Learning Theory", Springer, 1995.

[9] Joachims, T, "Text categorization with support vector machines: Learning with many relevant features". Proceedings of the Tenth European Conference on Machine Learning . Berlin:Springer, pp. 137 -142, 1998.

[10] H.Baltzakis and N. Papamarkos, "A new signature verification technique based on a two-stage neural network classifier", Pergamon, pp 95-103, 2001.

[11] H. Gómez-moreno, P. Gil-jiménez, S. Lafuente-arroyo, R. Vicen-bueno and R. Sánchez-montero, "Color images segmentation using the Support Vector Machines", 2003.