offline signature verification,
signature identification, shape context method

Marcin ADAMSKI*, Khalid SAEED**

OFFLINE SIGNATURE IDENTIFICATION AND VERIFICATION USING NONITERATIVE SHAPE CONTEXT ALGORITHM

The paper presents experimental results on offline signature identification and verification. At the first stage of the presented system, the binary image of the signature undergoes skeletonization process using KMM algorithm to have a thinned, one pixel-wide line, to which a further reduction is applied. For each thinned signature image a fixed number of points comprising the skeleton line are selected. The recognition process is based on comparing the reference signatures with the questioned samples using distance measure computed by means of Shape Context algorithm. The experiments were carried out using a database containing signatures of 20 individuals. For the verification process random forgeries were used to asses the system error. The main advantage of the presented approach lies in utilizing only one reference signature for both identification and verification tasks, whereas the achieved results are comparable with respect to the systems that use several training samples per subject.

1. INTRODUCTION

The problem of automatic identity verification is crucial for security of data and restricting access to protected resources. Traditional methods such as passwords, PIN numbers and identification cards do not provide 100% safety and are cumbersome in everyday usage. In order to improve the security together with the comfort of their application the biometric methods start to gain popularity and become the first choice solutions in many environments, where there is a need for personal authentication. One of the most popular behavioural biometrics that is used on everyday basis is a handwritten signature. Despite of its drawbacks such as relatively weak permanence [5] and ease of producing a forgery, the handwritten signature received a lot of attention in both research and in commercial institutions. The ongoing development of new algorithms for feature extraction and signature recognition resulted in systems with error rates comparable with the results received for other biometrics. This work presents an offline system – recognition process is based on the analysis of static signature images. An example of other approaches, where dynamics of signing activity is being considered can be found in [3].

The main aim of this work is an attempt to create method for recognition of handwritten signatures using only one reference sample per individual. This may be an important asset in some practical applications where there is only one genuine signature available for comparison. The experiments that were carried out showed that described system gives good results that are competitive with approaches using several reference samples for each subject.

2. THE OUTLINE OF THE SYSTEM

Most of the information in this section was published in previous work [1] and given here for reader convenience. In order to prepare data to the algorithm of recognition, the images of signatures are first stored as Portable Network Graphics files. Images can be obtained by means of scanning devices from original documents. The segmentation of signatures from acquired scans is not considered in this work, but can be easily implemented by applying certain constraints on the position of the signature inside the analyzed document. Another problem is noise and defects caused by poor quality of documents and the scanning process. In our experiments we used the threshold technique to eliminate minor distortions and convert images from grayscale into black-and-white binary map. The threshold value is selected experimentally and applied to all signature images. If the value of a pixel in the original image exceeds the threshold it is then converted into a black pixel, otherwise it is treated as a part of the background (white).

To reduce the needed amount of information for processing during the recognition task the skeleton of a signature image is obtained. There are many algorithms designed for this purpose. In this work we use KMM algorithm [10] (named after the first name initials of its authors – **K**halid, **M**arek, **M**ariusz). The KMM algorithm preserves connectivity, produces one-pixel width line and was proved to give very good results for both handwritten script and picture images. Fig. 1 presents examples of signature images and their skeletonized versions obtained by means of KMM algorithm. Tab. 1 shows examples of binary images of signatures and their thinned versions

* The Computer Science Faculty of Białystok Technical University, 15-351 Bialystok, Wiejska 45A
** AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracow, Poland, e-mail: saeed@novell.ftj.agh.edu.pl

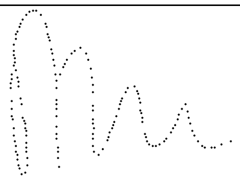
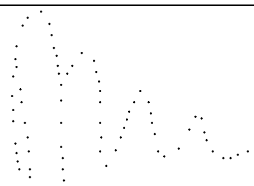
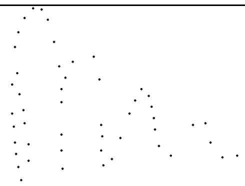
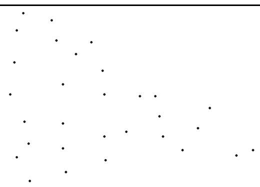
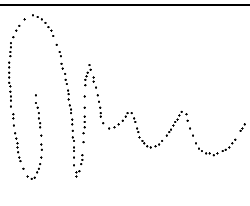
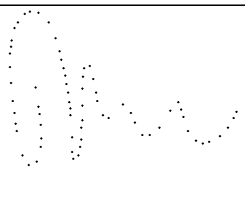
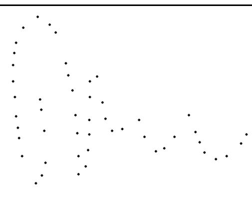
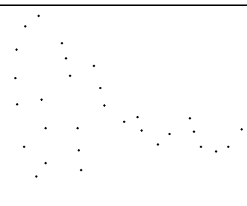
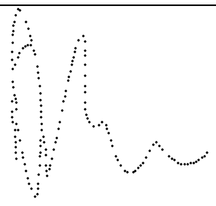
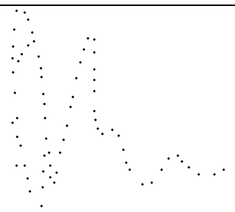
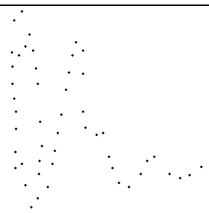
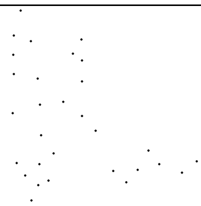
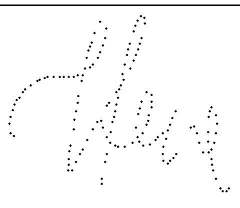
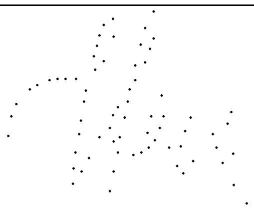
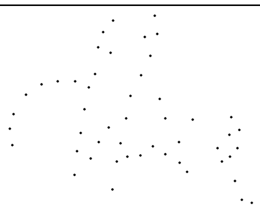
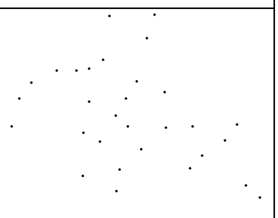
Table 1. Examples of signatures given by two individuals and their thinned versions

signature image	signature skeleton
	
	
	
	

For further reduction of data we use a sampling technique. During the procedure of sampling, we leave N equally spaced points from the thinned signature line, where N is chosen arbitrarily. The sampling algorithm iteratively deletes the signature pixels until only the required number of pixels (N) remains in the image. The signature pixels selected for deletion in each iteration have the smallest distance from their neighbouring pixel.

The number of selected points N was constant across all the signatures in the database. However, several experiments were carried out for different values of N, ranging from 10 up to 150. To illustrate points selected from the signature skeleton, tab. 2 shows points obtained for different examples of signatures using various values of N.

Table 2. Examples of the points selected form thinned signature lines for different values of N.
Signatures (a), (b) and (c) were given by one author, the signature (d) was produced by another author

	the number of selected points N			
	150	*70*	*50*	*30*
a				
b				
c				
d				

The Shape Context algorithm (SC) [2] allows for measuring shape similarity by solving the so called *correspondence problem* between two objects (A, B) and finding an aligning transform. Each of the two objects, whose shapes are compared, is represented by a set of points (1).

$$A = \{a_1, a_2, ..., a_n\},\\ B = \{b_1, b_2, ..., b_m\} \tag{1}$$

For each point on the first object (A) a corresponding point on the second object (B) is found. In order to find corresponding pairs every point is described by a shape context descriptor (2). This descriptor contains information about the configuration of the entire shape relative to the point being described. The shape context descriptor is formed by computing a coarse histogram representing distribution of points comprising the object relative to the reference point. Bins used in the histogram calculation are uniform in log-polar space. The number of bins (K) used during the experiments was 50 – this value was selected in [2] by Belogie, Malik and Puzicha.

$$\begin{matrix} h_i(k) = \#\ \{a \neq a_i : (a - a_i) \in bin(k)\} \\ h_j(k) = \#\ \{b \neq b_j : (b - b_j) \in bin(k)\} \end{matrix},\ k = 1...K \tag{2}$$

The matching cost of two points is calculated, where a pair of points – one point from object A and the other from B, are taken. It is based on statistic test χ^2 and given by (3):

$$C_{ij} = C(a_i, b_j) = \frac{1}{2} \sum_{k=1}^{K} \frac{\left[h_i(k) - h_j(k)\right]^2}{h_i(k) + h_j(k)} \tag{3}$$

The following step is to find pairs that minimize the expression (4). The result is a permutation P such that the expression (4) is minimized and represents the corresponding pairs of points.

$$H(P) = \sum_i C(a_i, b_{P(i)}) \tag{4}$$

The cost of matching two objects A and B can be computed using equation (5). In this work it is used as the *distance measure* for signature recognition.

$$\text{cost}(A, B) = \frac{1}{n} \sum_{a \in A} \min_{b \in B} C(a, b) + \frac{1}{m} \sum_{b \in B} \min_{a \in A} C(a, b) \tag{5}$$

The initial estimate of correspondences may contain some errors. In order to improve this estimation iterative process is applied where in each step a transformation function that aligns entire images is constructed [2]. In every iteration, before finding corresponding pairs, one of the objects is transformed according to this function and search for pairs is performed using the transformed object's coordinates. This transformation function is computed using the corresponding pairs obtained in previous iteration step by means of regularized thin-plate splines.

3. RESULTS

In order to assess the effectiveness of the method presented in this work, several experiments were carried out. The database of signatures that was used was created by 20 different people, with each signature repeated three times. The experiments were divided into to main parts related to signature identification and verification tasks. In the identification task, the system was evaluated by computation of the percentage of properly identified signatures. In the verification the *FAR* and *FRR* curves were determined for selected system configurations. Based on these results, the *EER* was computed and used for comparisons with other approaches.

3.1. SIGNATURE IDENTIFICATION

The identification task was conducted using the following scheme: during the experiments each instance of the genuine signature was used as a reference pattern to identify the remaining two. The experiments were repeated three times, taking consecutive instances of genuine signatures as reference patterns. According to this scheme every test comprised of 2400 comparisons (20 users x 40 test signatures x 3 repetitions). The main purpose of the identification process was to recognize the owner of the questioned signature without considering its authenticity. The applications of this process can be found in Document Retrieval Systems. Based on an example of a handwritten word or, as in this case, a signature, a database of documents is searched for pages that contain graphic elements most similar to the given example. Such a system, specifically designed to use signatures (called Signature Based Document Retrieval), can be found in [12].

Tab. 3 presents the achieved results in a form of percentages of properly identified signature owners. Each row groups the results obtained in one series of the experiments – using a particular value of *N* given by in the first column. The columns: r_1, r_2, r_3, contain the percentages of properly identified signatures using different genuine samples of each registered user taken as his reference signature. The last column shows the average values computed for the columns: r_1, r_2, r_3. As can be seen the results depend on the number of points and the selected reference pattern. The worst average result – 40% was obtained for *N*=10 and this clearly shows that this number of points is insufficient. The best average value of 99.2% of properly identified subjects was achieved by using 70 points per signature. In other words, it means that the system made one mistake during 120 identification attempts.

Table 3. Percent of properly identified signatures

	percent of properly identified signatures			
N	***r_1***	***r_2***	***r_3***	***average***
10	40.0%	47.5%	32.5%	40.0%
20	77.5%	77.5%	82.5%	79.2%
30	92.5%	92.5%	92.5%	92.5%
40	97.5%	97.5%	95.0%	96.7%
50	95.0%	**100.0%**	92.5%	95.8%
60	97.5%	97.5%	97.5%	97.5%
65	**100.0%**	97.5%	97.5%	98.3%
70	**100.0%**	**100.0%**	97.5%	**99.2%**
75	97.5%	**100.0%**	97.5%	98.3%
80	**100.0%**	97.5%	97.5%	98.3%
90	**100.0%**	97.5%	97.5%	98.3%
100	**100.0%**	97.5%	95.0%	97.5%
110	97.5%	97.5%	92.5%	95.8%
120	**100.0%**	97.5%	97.5%	98.3%
130	**100.0%**	97.5%	97.5%	98.3%
140	**100.0%**	**100.0%**	97.5%	**99.2%**
150	**100.0%**	**100.0%**	97.5%	**99.2%**

Further analysis of Tab. 3 shows, that for some of the reference samples all of the remaining signatures were properly recognized. It is also interesting that the selection of the reference sample that is not most representative version of the authentic signature doesn't lead to significant incensement of the system error. For example, considering the series for *N*=70, the worst result was obtained for *r3*, but still, the value of 97.5% properly identified signatures is quite good.

To achieve the above results the non-iterative version of the SC algorithm was used. The experiments have shown that, using the described method of selection of characteristic points, the iterations of the original SC method increased identification error. Tab. 4 illustrates average percent of properly identified signatures using SC algorithm with 3 iterations, presented in paper [1]. These were obtained on the same database and by means of identical procedure for choosing reference samples and averaging the final results. As can be seen the correctness of the system has decreased. The other drawback of the iterative version was its increased time of computation.

Table 4. Percent of properly identified signatures (iterative version of SC method)

N	30	50	100
average percent of properly identified signatures	87%	92%	95%

3.2. SIGNATURE VERIFICATION

The second part of the experiments was related to the signature verification. For the verification task the database that was used was the same as in previously described identification procedure. The main difference was introduction the threshold value *T*. If the distance between the reference and tested signature was below selected threshold, the test signature was identified as authentic example of the reference signature, otherwise it was rejected. The set of pairs (reference sample, tested sample) where the reference signature was compared to the tested example could be divided into two subsets (6 and 7).

$$G = \left\{ D(C_W(i), C_Q(j)) : i = j \right\}, \qquad (6)$$

$$H = \left\{ D(C_W(i), C_Q(j)) : i \neq j \right\}, \qquad (7)$$

where:

$C_W(i)$ – representation of the reference signature of the i-th person,

$C_Q(j)$ – representation of the questioned signature of the j-th person,

$D(C_W(i), C_Q(j))$ – the comparison of $C_W(i)$ i $C_Q(j)$ using SC-based distance measure.

The subset G represents these comparisons where the authentic signature of a particular user is tested against his reference sample. In this kind of comparisons the value of distance measure D should be below selected threshold. The comparisons given by the subset H represent so called random forgeries. In these cases, the original signature of one person was verified against reference example of another. In this case, the question signature should be rejected as not belonging to the subject who produced reference sample.

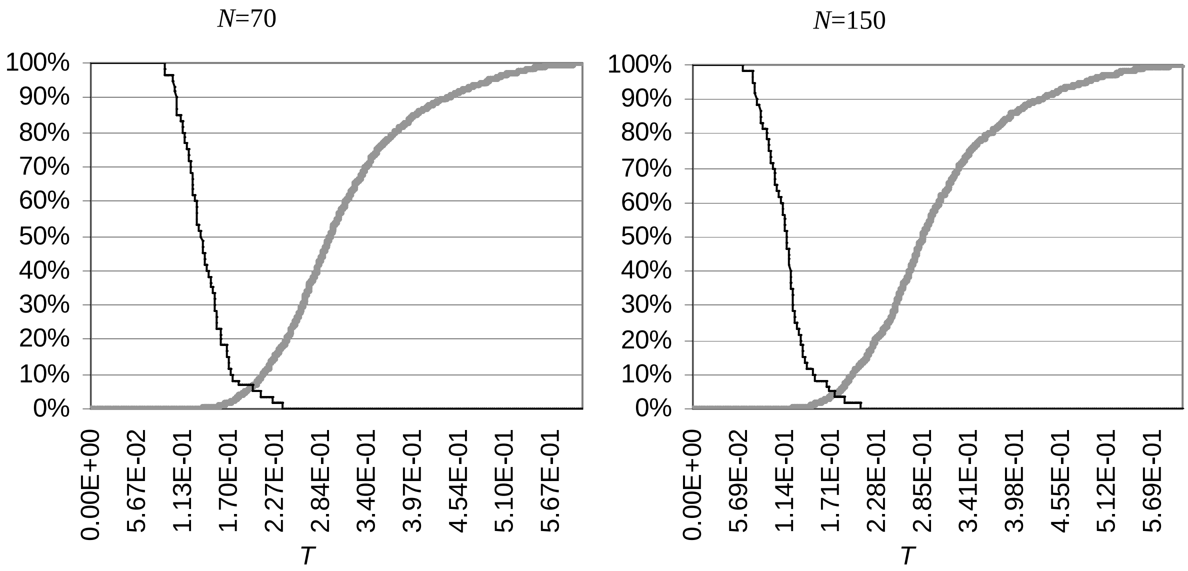


Fig. 2. *FAR* (thick line) and *FRR* (thin line) as functions of the threshold T for N=70 and N=150

The verification experiment was carried out for two values of N: 70 and 150. The achieved *EER* was 6.67% in the first, and 5% in latter case. Fig. 2 show the values of *FRR* and *FAR* as functions of T parameter. The characteristics were obtained for about 6000 different values of T.

3.3. COMPARISON TO OTHER WORKS

Tab. 5 compares the results of the system described in this paper with other approaches in the task of identification of offline signature. As can be seen the results are similar to those approaches that used greater number of learning samples. However, it should be mentioned that directed comparison of the results is difficult due to different databases that were used for the assessment of showed systems.

Table 5. Comparison to other works in the identification task

authors	**approach**	**subjects**	**reference samples**	**correctness**
Adamski M.,Saeed K.	SC	20	1	99.2%
Szedel J. [13]	Hidden Markov Models (HMM)	26	7	99.74%
Pavlidis I. et al. [7]	Defermable Models	40	1	70.83%
Riba J. and et al. [9]	Neural Networks	6	50	99.89%
Han K., Sethi I.[4]	Structural Approach	120	1	97.5%

Tab. 6 compares the results of the system described in this paper in signature verification task for random forgeries. The proposed method, in this case, also gave results comparable to other systems that used several reference samples per individual.

Table 6. Comparison to other works in the verification task (all the results were obtained using random forgeries)

authors	approach	subjects	reference samples	error
Adamski M., Saeed K.	SC	20	1	*EER*=5.0%
Szedel J. [13]	HMM	12	7	*FAR*=4.45%, *FRR*=4.17%
Murshed N. et al. [6]	ARTMAP	5	3	*FAR*=6.25%, *FRR*=17.27%
Santos C. et al. [11]	Neural Networks	40	5	*FAR*=10.33%, *FRR*=4.41%
Ramachandra A. et al. [8]	Graph Matching	5	3	*EER*=7.92%

4. CONCLUSIONS

In this paper a signature identification and verification system was presented. The aim of the system is to recognize offline handwritten signatures using only one reference sample per person. The achieved results are competitive with other approaches that use larger training sets. This encourages for further work.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Science and Higher Education under research project nr N N516 377636 (G/WI/1/2009). The contribution of the second author to the conference was supported by grant Socionical no. 27.27.220.7024 (H71)

BIBLIOGRAPHY

[1] ADAMSKI M., SAEED K., Signature Image Recognition by Shape Context Image Matching. Journal of Medical Informatics and Technologies, Vol. 11, pp. 89-95, 2007.

[2] BELONGIE S., MALIK J., PUZICHA J., Shape Matching and Object Recognition Using Shape Contexts. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 4, 509-522, 2002.

[3] DOROZ R., PORWIK P., PARA T., WROBEL K., Dynamic Signature Recognition Based on Velocity Changes of some Features. International Journal of Biometrics, Vol. 1, No. 1, pp. 47-62, 2008.

[4] HAN K., SETHI I. K., Handwritten Signature Retrieval and Identification. Pattern Recognition Letters, Vol. 17, pp. 83-90, 1996.

[5] JAIN A. K., ROSS A., PRABHAKAR S., An Introduction to Biometric Recognition. IEEE Trans. on Circuits and Systems for Video Technology, Vol. 14, No. 1, 4-20, 2004.

[6] MURSHED N. A., BORTOLOZZI F., SABOURIN R., Off-Line Signature Verification Using Fuzzy ARTMAP Neural Network. IEEE Int. Conf. on Neural Networks, Vol. 4, pp. 2179-2184, 1995.

[7] PAVLIDIS I., MAVUDURU R., PAPANIKOLOPOULOS N., Off-Line Recognition of Signatures Using Revolving Active Deformable Models. IEEE Int. Conf. on System, Man and Cybernetics, 1994.

[8] RAMACHANDRA A. C., PAVITHRA K., YASHASVINI K., RAJA K. B., VENUGOPAL K. R., PATNAIK L. M., Cross-Validation for Graph Matching based Offline Signature Verification. Annual IEEE India Conf., INDICON, Vol. 1, pp. 17-22, 2008.

[9] RIBA J. R., CARNICER A., VALLMITJANA S., JUVELLS I., Methods for Invariant Signature Classification. 15th Int. Conf. on Pattern Recognition, Vol. 2, pp. 953-956, 2000.

[10] SAEED K., RYBNIK M., TABĘDZKI M., Implementation and Advanced Results on the Non-interrupted Skeletonization Algorithm. Lecture Notes in Computer Science - W. Skarbek (Ed.), Computer Analysis of Images and Patterns: Springer-Verlag Heidelberg, pp. 601-609, 2001.

[11] SANTOS C., JUSTINO E. J. R., BORTOLOZZI F., SABOURIN R., An Off-Line Signature Verification Method Based on the Questioned Document Expert's Approach and a Neural Network Classifier. 9th Int. Workshop on Frontiers in Handwriting Recognition, pp. 498-502, 2004.

[12] SRIHARI S. N., SHETTY S., SIYUAN C., SRINIVASAN H., HUANG C., AGAM G., FRIEDER O., Document Image Retrieval Using Signatures as Queries. 2nd Int. Conf. on Document Image Analysis for Libraries, pp. 198 - 203, 2006.

[13] SZEDEL J., Metoda automatycznego rozpoznawania podpisów. Phd thesis, Silesian Technical University, 2004.