Krzysztof WRÓBEL*, Rafał DOROZ*

NEW SIGNATURE SIMILARITY MEASURE BASED ON AVERAGE DIFFERENCES

This paper presents a new method of recognizing handwritten signatures. Signature was treated as a collection of features of specific values. As features the values of x, y coordinates of signature points and the pressure p in its consecutive points have been used. Additionally, before comparing them, the signatures were properly prepared. The method discussed in the paper is a modification of the method based on average differences. This modification consists in dividing signatures into windows of the preset size and measuring the value of similarity between the windows according to their position in the signature. The paper shows the construction of a new similarity measure taking into consideration the modifications introduced and the results of the research obtained by means of this similarity measure.

1. INTRODUCTION

The methods of automatic personal identification based on biometric techniques are used in many spheres of life. Signature biometrics is one of the oldest and longest-used protection techniques. A handwritten signature has become an official form of reliability confirmation (e.g. in banking) [5].

At present there are a lot of methods of determining the similarity between signatures, which include on-line and off-line methods. Off-line methods treat a signature as a scanned image [3,7,9,10,11]. However, such an approach, as opposed to on-line methods, does not allow for the analysis of the dynamics of the writing process. The device which is able to register the signature together with the dynamics of the signing process, and which then sends the data to the computer is called a tablet. A signature is treated as a collection of movements which include unique biometric data such as for example: accelerating and changing the pressure of a pen on the tablet, the number of breaks in writing, the position of the pen, and the angles at which the person who is writing a signature holds the pen [1]. Using the given parameters, the signature analysis can be much more complex than in the case of a signature made on the paper, thus the probability of the correct signature recognition is higher.

In the proposed method several important stages: signature rotation, normalization of its features and comparison can be distinguish. In order to compare signatures, the similarity measure based on average differences has been used. Additionally, the modification of this rate, where signature is divided into pre-defined windows size has been introduced.

2. SIGNATURE DATA PREPARATION

All compared signatures come from SVC database (Signature Verification Competition) available on http://www.cse.ust.hk/svc2004/index.html. A single signature is represented by a text file which contains the values of its particular features. Fig.1 shows two sample of signatures and the values of selected features: coordinates x, y of the signature points and the pressure p.

	x	y	p
q_1	2873	6154	130
q_2	2768	6133	187
.	.	.	.
.	.	.	.
.	.	.	.
q_c	3783	6802	291

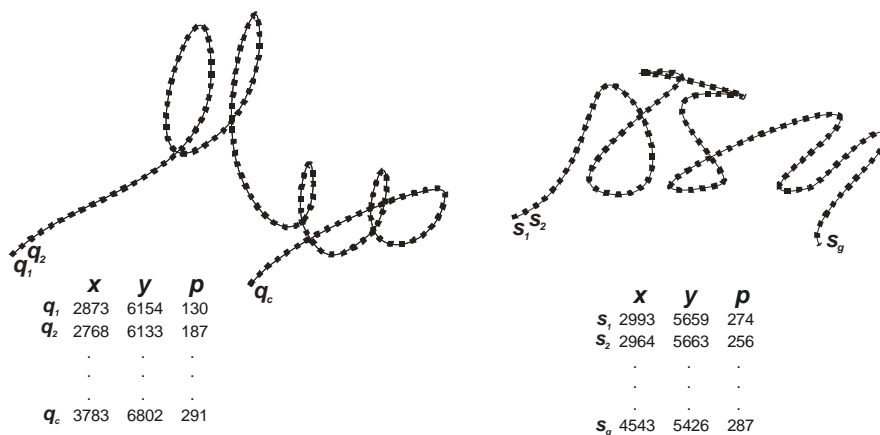|	x	y	p
s_1	2993	5659	274
s_2	2964	5663	256
.	.	.	.
.	.	.	.
.	.	.	.
s_g	4543	5426	287

Fig.1. Sample of signatures and their selected features.

* University of Silesia. Institute of Informatics, 41-200 Sosnowiec, Będzińska 39, Poland,
kwrobel@zsk.tech.us.edu.pl, doroz@zsk.tech.us.edu.pl

An algorithm of the comparison of the signatures should be independent of their direction resulting from the different position of the tablet on which the signatures were written. Therefore, every signature before comparing it was additionally turned by angle α The value of α is estimated on the basis of inclination of simple linear regression l of coordinate signature points to the axis X (Fig. 2). The detailed description of signature rotations is presented in the work [3].
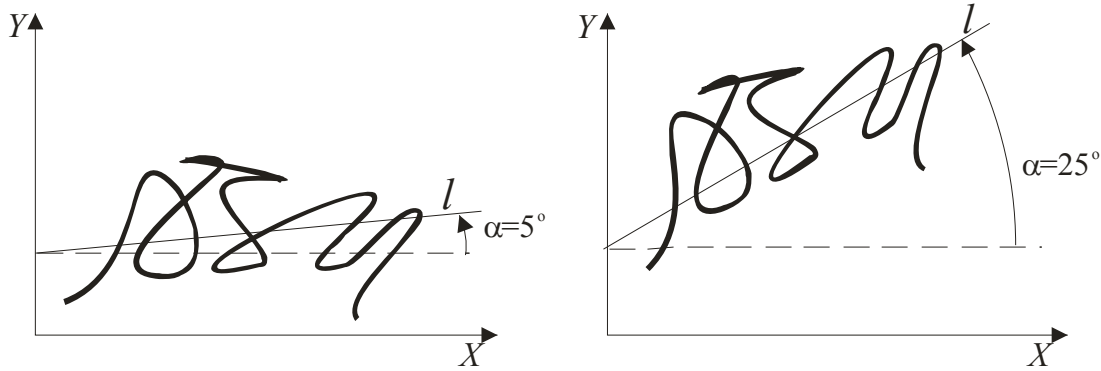

Fig. 2. Different angles of inclination of simple linear regression line l to the axis X for the same signature.

3. SIMILARITY MEASURE WHICH USES AVERAGE DIFFERENCES

The presented method of marking the similarity is based on the comparison of a set collection of signature features. There was assumed that for a single signature, the features are the values of coordinates x, y of signature points and pressure p in its consecutive points.

In order to determine the similarity between signatures, the similarity measure based on average differences has been used [2]. The average difference between elements M_1 and M_2 of the analyzed data collection is represented by the following equation:

$$DD(M_1, M_2) = \frac{1}{n}\sum_{i=1}^{n}|M_{1i} - M_{2i}| \qquad (1)$$

where:
n – umber of features being compared,
M_1 – normalized to the range [0,1] value of the i'th feature of the first element,
M_2 – normalized to the range [0,1] value of the i'th feature of the second element.

The average difference takes the values from the range [0,1]. Value 0 means that the elements are identical. In this paper equation (1) to the form (2), in which value 1 means that the elements are identical, has been modified. Such a modification allows for the comparison of the presented similarity measure with other normalized similarity measures to the range [0,1].

$$DD(M_1, M_2) = 1 - \sum_{i=1}^{n}|M_{1i} - M_{2i}|w_i \qquad (2)$$

In order to specify the influence of a given feature on the result of the comparison, the weight of the ith feature w was introduced, which fulfills the following conditions:

$$w_i \in [0,1], \quad \sum_{i=1}^{n}w_i = 1.$$

To compute the similarity between two signatures Q and S, the following equation was modified:

$$sim_{DD}(Q,S) = \frac{\sum_{i=1}^{c}\max\big(DD(q_i, s_1), DD(q_i, s_2), ..., DD(q_i, s_g)\big)}{c} \qquad (3)$$

where:
$Q = \{q_1, q_2, ..., q_c\}$ – collection of c data representing the first signature,
$S = \{s_1, s_2, ..., s_g\}$ – collection of g data representing the second signature.

A single data collection is described by means of three features, e.g. $q_1 = \{x_1^q, y_1^q, p_1^q\}$, $s_1 = \{x_1^s, y_1^s, p_1^s\}$.

The similarity measure presented by equation (3) has the following characteristics:

1. $sim_{DD}(Q,S) \in [0,1]$. Value 1 means that the signatures are identical.
2. The similarity measure is not symmetrical, which means that $sim_{DD}(S,Q) \neq sim_{DD}(Q,S)$.

The similarity measure presented by equation (3) compares only the features of single signature points. In practice, two theoretically different signatures may have points of the same coordinates. For this reason, similarity measure can be wrongly determined. That is why, while comparing the signatures, it is better to analyze the string of consecutive signature points. In the next part of the paper the string h of the signature points will be called 'window' and marked as *win*. The division of the signatures Q and S into windows is shown in Fig. 3
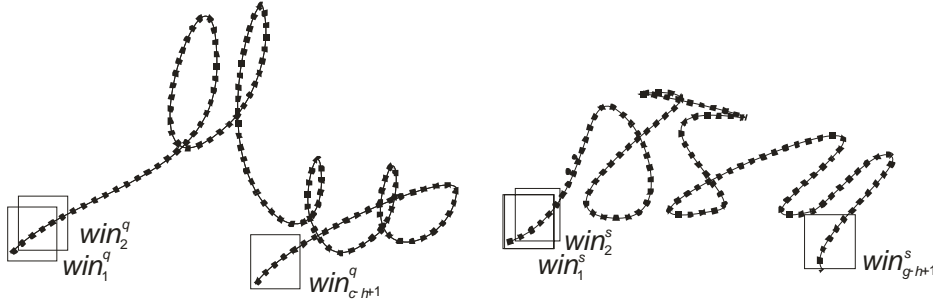

Fig.3. Division of signatures Q and S into windows, where h=5.

After dividing the signatures into windows, average differences are calculated for the windows of the signatures being compared. In order to calculate the average difference for two windows, it is necessary to calculate average differences between the corresponding lines in the windows (Fig.4).

Fig.4. Rule for calculating the average difference for two windows.

For every pair of windows being compared, the average difference, has been calculated:

$$DD_{i,j}^{h} = \frac{\sum_{k=j}^{h+j-1} DD(q_{i+k-1}, s_k)}{h} \qquad (4)$$

where:
i – number of the window in signature Q, $1 \leq i \leq c-h+1$,
j – number of the window in signature S, $1 \leq j \leq g-h+1$,
h – number of signature points in the window.
In addition, equation (4) was modified by means of two factors which considered:
  o standard deviation between the values of average differences in the compared windows,
  o distance between the windows being compared.
The average value does not allow us to estimate the discrepancy between individual values DD in the window. Therefore, the result of the comparison is additionally dependent on the value of the calculated standard deviation *SD*. It was assumed that the higher the value of standard deviation, the smaller the similarity between the windows. Moreover, distance d between the windows of signatures being compared was taken into consideration. This distance is defined by equations (6a) and (6b).
After considering the above modifications, equation (4) looks as follows:

$$DD_{i,j}^{h} = DD_{i,j}^{h} \cdot \left(1 - SD_{i,j}^{h}\right) \cdot d\left(win_i^q, win_j^s\right) \qquad (5)$$

where:

$SD_{i,j}^h$ – standard deviation between the values of average differences in windows i and j,

$d\left(win_i^q, win_j^s\right)$ – distance between the compared windows win_i^q and win_j^s, according to the equations:

$$d\left(win_i^q, win_j^s\right) = 1 - \left| \frac{i}{c-h+1} - \frac{j}{g-h+1} \right| \qquad (6a)$$

$$d\left(win_i^q, win_j^s\right) = \begin{cases} 0, & d\left(win_i^q, win_j^s\right) < 0.95 \\ d\left(win_i^q, win_j^s\right), & d\left(win_i^q, win_j^s\right) \geq 0.95 \end{cases} \qquad (6b)$$

Value 0.95 in equation (6b) was experimentally estimated.

Next, a single window in the first signature is compared with every window in the second signature (Fig.5). From among all comparisons the most similar window is selected. This step is repeated for every single window of the first signature.
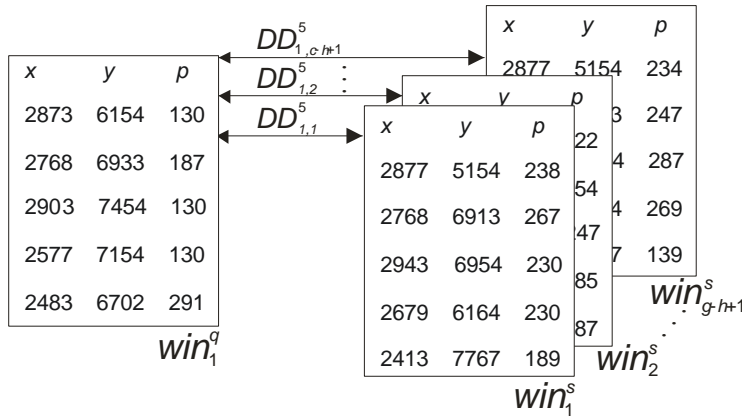

Fig.5. Example of comparing windows of signatures Q and S, respectively.

Finally, the similarity measure which considers all the above modifications is calculated by means of the following equation:

$$sim_{DD}^h(Q,S) = \frac{\sum_{i=1}^{c-h+1} \max\left(DD_{i,1}^h, DD_{i,2}^h, ..., DD_{i,g-h+1}^h\right)}{c-h+1}, \text{ for } h < c, h < g \qquad (7)$$

For the parameter value $h = 1$ equation (7) is equal to equation (3).

4. RESEARCH RESULTS

The research was conducted on 50 signatures written by 10 people. Each person wrote 5 signatures. Forgeries were not taken into consideration during research. All in all, 19 measurements were made, taking in each measurement different values of weights of particular features. During the research, the signature window size was arbitrarily chosen as h = 10 points. Since the selected similarity measure is not symmetrical, the comparison of signatures was made according to the rule 'everyone with everyone else'. Altogether, it resulted in 2,500 comparisons for one measurement.

For each measurement, Equal Error Rate (EER) was calculated [3]. The smaller the EER, the smaller the error of the method. Table 1 shows the results obtained, while Fig.6 shows the dependence diagram FAR-FRR-ERR for the measurement with weights $w_x = 0.4$, $w_y = 0.4$, $w_p = 0.2$.

Table 1. Measurements results for $h = 10$ and different values of weights of particular signature features.

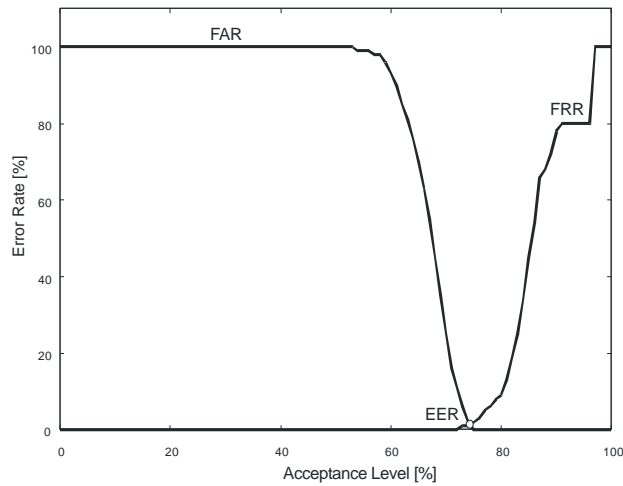| w_x	w_y	w_p	EER [%]
0.4	0.4	0.2	1.06
0.33	0.33	0.33	1.15
0.5	0.25	0.25	1.46
0.4	0.2	0.4	1.82
0.25	0.5	0.25	1.86
0.5	0.5	0.0	2.22
0.2	0.6	0.2	2.26
0.2	0.4	0.4	2.26
0.6	0.2	0.2	3.02
0.25	0.25	0.5	3.06
0.2	0.2	0.6	3.95
0.5	0.0	0.5	4.17
0.0	0.5	0.5	4.75
0.8	0.1	0.1	6.26
0.1	0.8	0.1	6.48
0.1	0.1	0.8	9.06
1.0	0.0	0.0	10.62
0.0	1.0	0.0	11.11
0.0	0.0	1.0	15.46

Fig.6. Diagram FAR-FRR-ERR for the measurement with weights $w_x = 0.4$, $w_y = 0.4$, $w_p = 0.2$.

The analysis of the research results shows:
- o the best results were obtained during the analysis of coordinates *X* and *Y* with weight 0.4 and pressure *P* with weight 0.2,
- o calculating the comparison by means of only one feature leads to a major error (>10%),
- o the further away the values of weights of particular features, the worse the final result,
- o in the 5 best results, the weights of the features *X* and *Y* play a significant part.

Table 2 shows comparison results provided by the most recently proposed techniques, where pen tables were used.

Table 2. Different online signature verification methods [6]

Methods (for genuine signatures)	ERR (%)
Proposed technique (the best result)	1.06
Data glove, Kamel et al. (2008)	2.37
Maramatsu et al. (2003)	2.60
Kholmatov et al. (2005)	2.80
Nakanishi et al. (2005)	3.30
Shinatro et al. (2006)	4.10
Fierez-Aguilar et al. (2005)	from 5 to 7
Lei et al. (2004)	7.2

5.   FINAL REMARKS

The presented results lead to the conclusion that the method based on average differences is very useful in the process of signature recognition. The error obtained is relatively low, although other dynamic features such as the speed of pen were not considered in the research. The results of the experiment are an incentive to do further research into the method proposed. In the next experiments we are going to determine the influence of additional parameters, such as the size and distance between the windows, on algorithm function. Addition to the database forgeries is also the aim of next investigations.

BIBLIOGRAPHY

[1]   AL-SHOSHAN A.I., Handwritten Signature Verification Using Image Invariants and Dynamic Features, Computer Graphics, Imaging and Visualisation, International Conference on Volume, pp. 173 – 176, 2006.

[2]   CZEKANOWSKI J., Zarys metod statystycznych w zastosowaniu do antropologii. Prace Towarzystwa Naukowego Warszawskiego, Warszawa 1917.

[3]   DOROZ R., PORWIK P., PARA T., WRÓBEL K., Dynamic signature recognition based on velocity changes of some features. International Journal of Biometrics, Vol. 1, No. 1, pp. 47-62.

[4]   FOLEY J.D., Introduction to Computer Graphics, Addison-Wesley, 1993.

[5]   IMPEDOVO S., PIRLO G., Verification of Handwritten Signatures: an Overview, 14th International Conference on Image Analysis and Processing ICIAP, 2007.

[6]   KAMEL N.S., ELLIS G.A., SAYEED S., Glove-based approach to online signature verification, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.

[7]   KHAN M.K., KHAN M.A., KHAN M.A.U., AHMAD I., On-Line Signature Verification by Exploiting Inter-Feature Dependencies, Pattern Recognition, 2006. ICPR 2006. 18th International Conference on Volume 2, pp. 796 – 799.

[8]   PARA T., DOROZ R., Using dynamic features and linear regression to signature verification. Techniki Komputerowe. Biometria 2007, pp. 129-138. Warszawa 2008.

[9]   PORWIK P., The Compact Three Stages Method of the Signature Recognition. Proceeding of 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07), Ełk. 2007, pp. 282-287.

[10]   PORWIK P., PARA T., Some Handwritten Signature Parameters in Biometric Recognition Process, Information Technology Interfaces, 2007. ITI 2007. 29th International Conference, pp. 185-190.

[11]   SAEED K., ADAMSKI M., Klasyfikacja podpisu offline z wykorzystaniem metody DTW – Krajowa Konferencja Naukowa - KBIB'05, tom I, Wydawnictwa Politechniki Częstochowskiej, Częstochowa 2005, pp. 455-460.