IPAN 99, DTW,
lineał regression

Rafał DOROZ*, Stanisław WIDUCH*

SIGNATURE CHARACTERISTIC POINTS DETERMINATION BY MEANS OF THE IPAN99 ALGORITHM

The handwritten signature is often used for the identity confirmation. From the specialized graphic tablet, we receive the information in form of time strings. In this paper, the research results are presented which refer to the effect of applying the method of point detection of the highest curvature. Signatures are normalized by means of a DTW method, where two time strings representing the signature features are matched.

1. INTRODUCTION

The methods for the automatic identification of a signature can be applied in many areas of human beings' life. Signature is one of the longest known security techniques. Pattern of signature has been for many years an accepted form of credibility determination (e.g. in case of bank transactions). We can generally distinguish between two different categories of verification systems: online, for which the signature signal is captured during the writing process, thus making the dynamic information available, and offline for which the signature is captured when the writing processing is over and, thus, only a static image is available. Tablet is able to execute the measurement of several parameters during performance of signature, e.g. it measures the pressure of the pen on the tablet surface, determines the location of the pen, etc. For this reason, the analysis of the presented signature can be more complex and based on a greater number of extracted signature features. There are a lot of methods for determination of signature similarity [1,3,10]. In the methods based on signature features analysis, isolated signature points are investigated (usually the edge points of signature are considered). Hence, signature analysis problem can be reduced to appropriate signature point adjustment.

Described in this paper the linear regression method requires the usage of the strings with the same data quantity. In case of the signatures, theirs differences can be considered as changes local sizes (width, height of the characters) as well as with the redundancy or shortage of the elements forming the signature. One of the methods of adapting two time series is the DTW(Dynamic Time Warping) algorithm [1,2,8,9].

Analysing the entire received data string may be too time-consuming taking into the consideration the base of e.g. 10000 signatures. Therefore, it is advisable to find the methods for the reduction of the string length, which describe the given signature.

*University of Silesia. Institute of Informatics, 41-200 Sosnowiec, Będzińska 39, Poland

The goal of the research is not to determine the optimal method for signature classification, but only to analyse the effect of applying the algorithm IPAN99 [4,5,6], where can be detected points with the highest curvature. Taking into account detected signature points, some results of string adaptation by means of DTW method will be also presented.

2. DETECTION OF POINTS WITH THE HIGHEST CURVATURE

Points with the highest curvature play an important role in perceiving the shapes by a human being. In the present thesis, a very fast and efficient algorithm IPAN99 has been applied, which was published in year 1999 [6]. The algorithm IPAN99 defines the corner as a point of a given curve, if we are able to inscribe in that curve a triangle with a given size and angle of the vertical gap.

In the first stage, the algorithm reviews the string of points and selects the candidates for corners. In the second stage, the redundant candidates are eliminated:

$$\begin{aligned} d_{\min}^2 &\le \left|p - p^+\right|^2 \le d_{\max}^2 \\ d_{\min}^2 &\le \left|p - p^-\right|^2 \le d_{\max}^2 \\ \alpha &\le \alpha_{\max}, \end{aligned} \qquad (1)$$

where:

$\left|p - p^+\right| = |a| = a$ the distance between points P and $P+$,

$\left|p - p^-\right| = |b| = b$ the distance between points P and P^-,

$\alpha \in [-\pi, \pi]$ the triangle angle of vertical crack defined as follows:

$$\alpha = \arccos\frac{a^2 + b^2 - c^2}{2ab}$$

Values d_{min} and d_{max} are entered into the algorithm as its parameters.

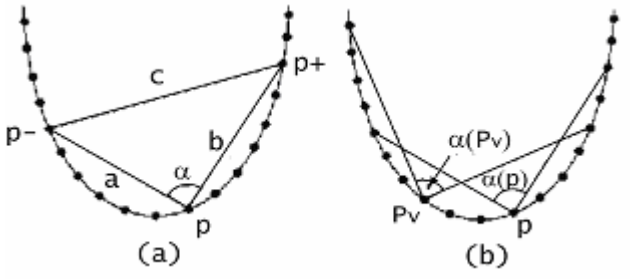


Fig.1. Detection of points with the highest curvature by means of the IPAN99

Triangle, which meets the conditions (1) is so called an acceptable one. The searching for an acceptable triangle is stared from the point P on the outside that is from the shortest length of the triangle sides and stops, if any part of the conditions (1) is fulfilled (therefore

the restricted number of neighbouring points is taken into consideration). Among all acceptable triangles, the one with the smallest angle of vertical crack is selected $\alpha(P)$.

P_i belongs to the neighbourhood of P point, if $|P - P_i|^2 \leq d^2_{min}$.

Algorithm parameters:

d_{min} – the parameter restricting the length of sides from the "bottom". Small values cause that algorithm reacts to small corners.

d_{max} – the parameter restricting the length of sides from the "top". It is necessary to avoid false acute angles created by distant points of the curve

α_{max} – boundary angle specifying the minimal acuteness, which has to have a point in order to classify it as the candidate for the corner.

The candidate P point is rejected, if it has a shaper neighbour that is P_i point, which is also a candidate, and which was assigned a greater strength of the corner:

$$\alpha(P) > \alpha(P_i) \tag{2}$$

3. SIGNATURE SIMILARITY DETERMINATION BY USING LINEAR REGRESSION METHOD

One of the methods, which is most often used for determination of the functional dependency between two data strings is the method of the smallest squares. This method enables forming of a mathematical formula describing the relation between the studied elements.

In the described method as the measure for the quality of point adjustments of different strings, the coefficient R^2 was adopted, where the variables X and Y can be correlated: [7].

$$R^2 = \frac{\left[\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})\right]^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2 \sum_{i=1}^{n}(y_i - \overline{Y})^2} \tag{3}$$

Where R^2 has the following:

1. Reflexivity, i.e. $R^2(X,X) = 1$
2. Symmetry, i.e. $R^2(X,Y) = R^2(Y,X)$
3. $R^2 \in [0,1]$ where $R^2=1$ means the full linear conformity (high similarity), and $R^2=0$ means the lack of linear conformity (the lack of similarity).

4. DATA ADAPTION BY MEANS OF THE DTW METHOD

Some inconvenience of the above described method, where factor R^2 is used, is the possibility of similarity description for only two sequences with the same number of n points. As it has been described in the Section 1, the biometric data used in the signature analysis have different lengths. In order to standardise two time series, the non-linear transformation based on the DTW methods [1,2,8,9],was used in the research.

The basic DTW algorithm defines the distance between two time series $X=(x_1,x_x,...,x_n)$, $Y=(y_1,y_2,...,y_m)$ The calculation are started with the formation of the matrix $n\times m$ (Fig. 2), which enables the setting of the "best adjustment path " according to the dependencies(3):

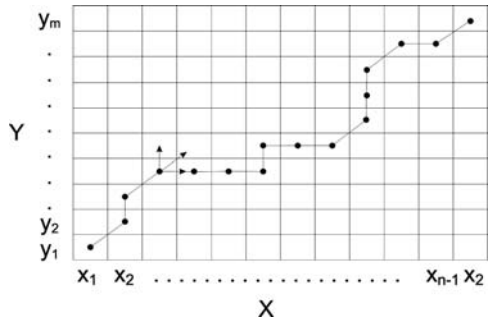


Fig.2. The matrix adjustment plan of an exemplar path.

$$D(i,j)=\min\begin{Bmatrix} D(i-1,j-1)+d(x_i,y_i) \\ D(i-1,j)+d(x_i,y_i) \\ D(i,j-1)+d(x_i,y_i) \end{Bmatrix} \qquad (4)$$

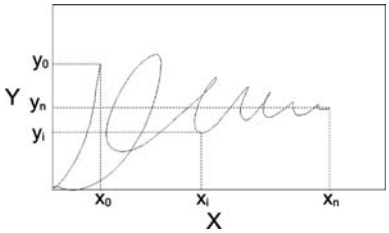
Where: $d(x_i,y_i)=\sqrt{(x_i-y)^2}$ is the cost of the series matching.

The path determination has some restrictions:

- initial and final points of the *X* and *Y* signal should be equal,
- the path cannot have any interruptions,
- the path points have to be arranged in time, it cannot come back to already assigned points.

5. CONCLUSIONS AND RESEARCH RESULTS

The signature data were produced by means of the SigLite LCD 4x3 tablet, where the static and dynamic signature features can be captured. It allows the measure of the signature features such as: signature time intervals, pen-pressure and pen location on the tablet surface. The sizes of the active area, where the signature can be putted are as follows: 112 x 64mm, the speed of the data conversion: 377 points per second, resolution: 410 ppi. The signatures signed by 30 different people were the set of the input data for the research. Each of them made 5 signatures every several days. Finally, the signature database contains 150 signatures.



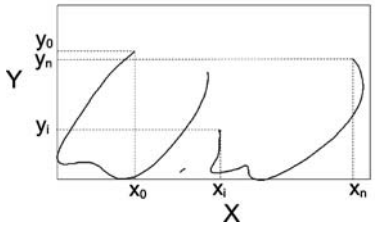


Fig.3. Coordinates x_i or y_i of the analysed signature.

Each signature was treated as a series of points $X = (x_1, x_x, ..., x_n)$. The values of the coordinates x_i or y_i of the analysed signature can be considered as the string values. (Fig.3)

In the next step, the characteristic function of coordinates y_i points were designated by using method of IPAN99. The sets of the characteristic points of two signatures were normalised by means of DTW algorithm. In proposed approach the influence of the basic parameters d_{min},d_{max},α_{max} of the IPAN99 algorithm on the DTW working has been investigated. The Figs.4a and 5a present the signature characteristics of the same person. Additionally, in these figures characteristic points of signature were marked. The Fig. 5a presents all characteristic points both signatures, and Fig. 5b presents the same signatures, normalised by means the DTW algorithm, where signatures have the same length. The number of characteristic points depending on IPAN99 parameters. From Fig.5 follows, that the new number of characteristic points is less than number points observed in Fig.4.

These points by means of the DTW procedure were normalized. The results, obtained from mentioned algorithm, were showed by Figs.4b and 5b, respectively. The Figs.5 and 6 present the analogous approach for signatures of two different persons.

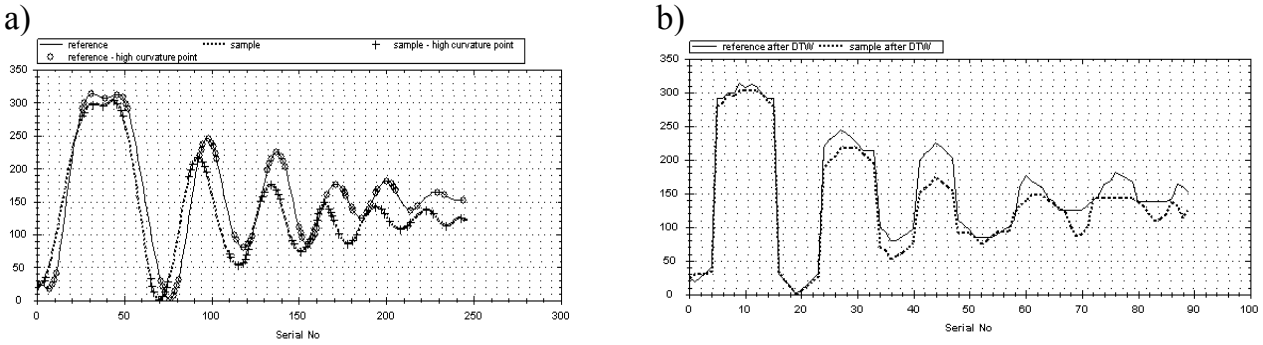


Fig.4. The two signatures of the same person, analysed with aid of the of the IPAN99 parameters: d_{min}=5 d_{max}=10 α_{max}=160. The number of the characteristic points in the reference signature=62, in the analysed signature=71. The linear regression coefficient =0,9717.

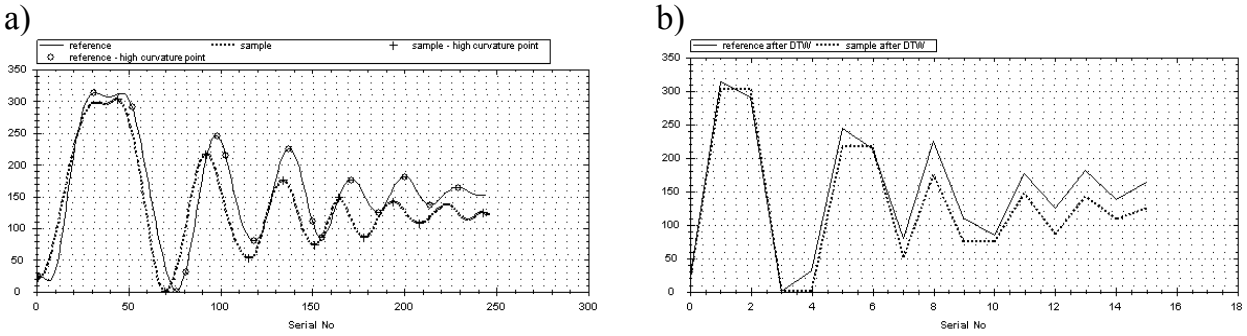


Fig.5. The two signatures of the same person, analysed with aid of the of the IPAN99 parameters: d_{min}=5 d_{max}=10 α_{max} =130. The number of characteristic points in the reference signature= 16, in the analysed siganture = 12. The linear regression coefficient =0,9637.

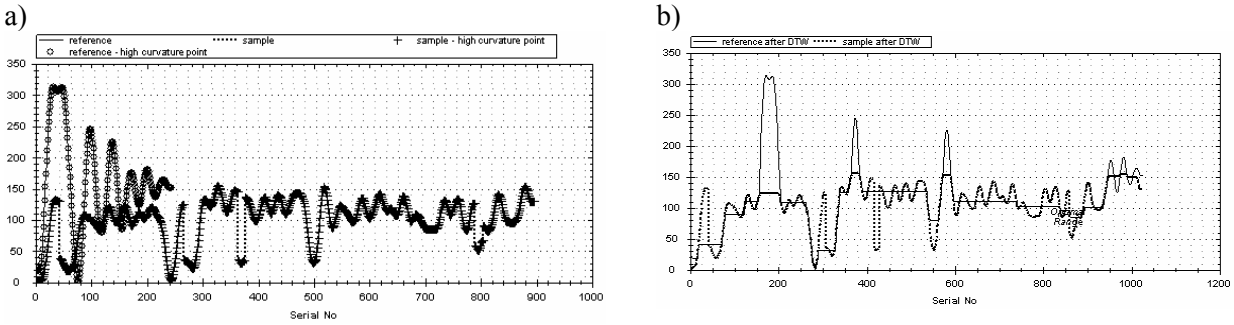


Fig.6. The two signatures of the different persons, for the IPAN99 parameters: d_{min}=5 d_{max}=50 α_{max}=160. The number of characteristic points in the reference signature= 245, in the analysed signature= 886. The linear regression coefficient = 0,40

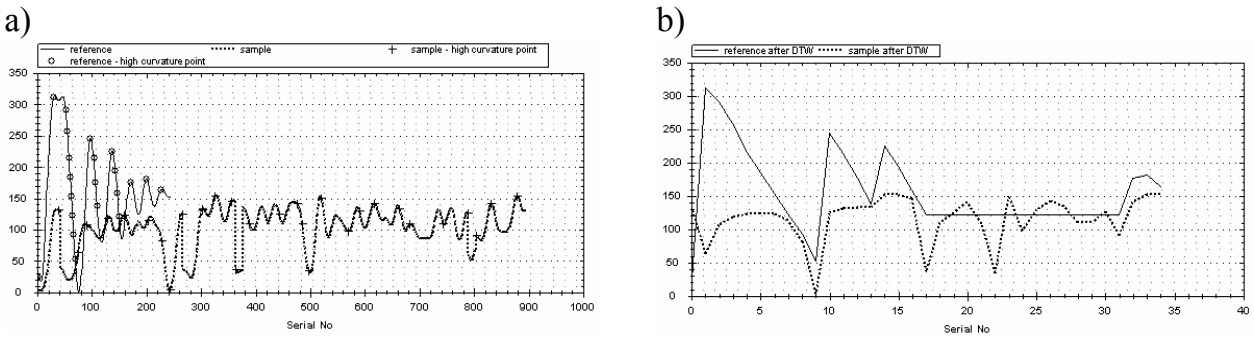


Fig.7. The two signatures (from Fig. 6), analysed with aid of the IPAN99 algorithm d_{min}=5 d_{max}=50 α_{max} =100. The number of characteristic points in the reference signature= 21, in the analysed signature= 30. The linear regression coefficient = 0,054.

In the last stage, the linear regression coefficient $\underline{R}^2$ is calculated by means of the smallest square method for the set of the characteristics points of the two signatures.

In investigation carry out, the comparisons of signatures have been performed, where particular person signature with different signature was analysed.. The entire research was performed twice. In the first case, the IPAN99 algorithm was applied in the method defining its parameters (d_{min},d_{max}, α_{max}), and in the second case, the mentioned algorithm was not used. In the Tables 1 and 2, the values of the linear regression coefficient are presented depending on the set of IPAN99 algorithm parameters.

Table 1. The exemplary percentage error coefficients depending on the algorithm IPAN99 parameter set

Parameters of IPAN99 algorithm d_{min} = 5, d_{max} = 40, α_{max} = 170, the number of characteristic points =25%												
	DTW with the usage of IPAN99						DTW without the usage of IPAN99					
	Signatures of the same person			Signatures of the another persons			Signatures of the same person			Signatures of the another persons		
	average	min	max	average	min	max	average	min	max	average	min	max
sign1	0,93	0,88	0,98	0,53	0,40	0,68	0,96	0,95	0,97	0,55	0,42	0,68
sign 2	0,90	0,89	0,91	0,40	0,22	0,78	0,91	0,89	0,90	0,52	0,22	0,77
sign 3	0,92	0,87	0,98	0,27	0,09	0,74	0,93	0,92	0,94	0,40	0,14	0,71
sign 4	0,86	0,79	0,90	0,34	0,11	0,59	0,84	0,80	0,88	0,41	0,11	0,57
sign 5	0,84	0,75	0,85	0,53	0,42	0,69	0,92	0,88	0,96	0,55	0,32	0,77
sign 6	0,85	0,76	0,87	0,45	0,17	0,72	0,91	0,90	0,96	0,48	0,24	0,71
sign 7	0,85	0,80	0,92	0,38	0,38	0,38	0,90	0,88	0,91	0,44	0,26	0,68
sign 8	0,95	0,93	0,97	0,38	0,10	0,64	0,96	0,95	0,98	0,45	0,21	0,69
sign 9	0,93	0,90	0,97	0,34	0,02	0,73	0,97	0,96	0,97	0,41	0,06	0,73
sign 10	0,94	0,89	0,98	0,53	0,40	0,68	0,96	0,95	0,97	0,55	0,42	0,68
average	0,90			0,42			0,93			0,48		

Table 2. The exemplary percentage error coefficients depending on the algorithm IPAN99 parameter set

Parameters of IPAN99 algorithm d_{min} = 10 d_{max} = 40 α_{max} = 150, the number of characteristic points =16%												
	DTW with the usage of IPAN99						DTW without the usage of IPAN99					
	Signatures of the same person			Signatures of the another persons			Signatures of the same person			Signatures of the another persons		
	average	min	max	average	min	max	average	min	max	average	min	max
sign1	0,87	0,78	0,95	0,45	0,23	0,73	0,97	0,95	0,98	0,56	0,43	0,68
sign 2	0,60	0,02	0,91	0,33	0,10	0,76	0,90	0,90	0,90	0,52	0,22	0,77
sign 3	0,73	0,23	0,98	0,35	0,16	0,73	0,94	0,93	0,95	0,40	0,14	0,71
sign 4	0,80	0,72	0,86	0,28	0,15	0,43	0,84	0,80	0,88	0,42	0,12	0,58
sign 5	0,75	0,65	0,84	0,48	0,35	0,64	0,92	0,89	0,96	0,55	0,32	0,77
sign 6	0,33	0,22	0,42	0,38	0,15	0,65	0,96	0,95	0,96	0,49	0,25	0,72
sign 7	0,76	0,55	0,93	0,38	0,38	0,38	0,90	0,89	0,92	0,44	0,26	0,68
sign 8	0,93	0,88	0,98	0,31	0,06	0,61	0,96	0,95	0,98	0,46	0,21	0,70
sign 9	0,92	0,86	0,98	0,27	0,02	0,72	0,97	0,96	0,98	0,43	0,07	0,73
sign 10	0,88	0,78	0,95	0,45	0,23	0,73	0,97	0,95	0,98	0,56	0,43	0,68
average	0,75			0,37			0,93			0,48		

It results from the performed research that the appliance of the IPAN99 method decreased the average value of the linear regression coefficient. It happened both in case of the comparison the signatures of the same person (from 0,93 to 0,88 for IPAN99 parameters in the Table 1 and from 0,93 to 0,75 in the Table 2) and in case of comparing the signatures of different persons (accordingly from 0,48 to 0,42 and from 0,48 to 0,37). The exception is the signature 4, for which the mentioned situation did not occur in the Table1. In that case, the the linear regression coefficient after appliance of IPAN99 increased from 0,84 to 0,86. The differences in the linear regression coefficient are definitely smaller in the Table1. It is obviously joined with the IPAN99 parameters. In the Table2, the borderline angle $\alpha_{\max}$ specifying the minimal sharpness, which the point has to have to be qualified as the candidate for corner, is smaller. It causes that part of the points cannot be qualified as a characteristic point. The IPAN99 algorithm with the parameters specified in the Table1

qualified in average 25% of the points as the characteristic points, whereas in the Table2 the percentage of these points amounted only to 16%.

Also the analysis of the difference between the minimal and maximal values of the regression coefficient for particular measures provides interesting results. In case of a research without the usage of the IPAN99 the biggest difference between the minimal and maximal values of the regression coefficient for the signatures of the same person amounts to 0,08 (signature4). In case of comparing the signatures of different persons this value amounts to 0,67 (signature9). It testifies that the regression coefficient in case of comparing the signatures of the same persons shows only very small diversity. The usage of the IPAN99 algorithm influenced the increase of this difference. In the Table1, the maximal difference amounts to 0,11 (signature6) for the comparison of the signatures of the same persons and 0,65 (signature3) for the comparison of the signature of different persons. In the Table2, the maximal difference amounts to 0,88 (signature2) for the signatures of the same person. Such a high difference between the smallest and the biggest linear regression coefficients testifies that the IPAN99 algorithm parameters were selected wrongly.

The performed investigations showed that the smallest differences between the values of the linear regression coefficient achieved by means of the IPAN99 algorithm and the regression coefficient without usage of this algorithm were achieved for the following the IPAN99 algorithm parameters: $d_{min}=5$, $d_{max}=40$, $\alpha_{max}=170$. In case if using the IPAN99 algorithm, the average regression coefficient decreased by about 0,05. The comparable differences between the minimal and maximal regression coefficient value for the particular measurements were preserved. The value of the linear regression coefficient was subject to a slight change with the significant reduction of the number of characteristic points.

6. SUMMARY AND PLANNED RESEARCH DIRECTION

The goal of the research was not to determine the best method for signature classification, but only to analyse the effect of applying the algorithm of point detection of the highest curvature IPAN99 on the results of string adaptation by means of the DTW method.

The goal of the performed research was to determine the effect of applying the method of point detection of the highest curvature on the distance determination between two data strings, and not the determination of the best method for signature classification.

From the point of view of the research result analysis, it is worth paying attention to the difference between the values of the linear regression coefficients achieved by means of IPAN99 algorithm and the regression coefficients, which do not use this algorithm.

Summarising the research results, it can be stated that the usage of the IPAN99 algorithm enables the preliminary selection of characteristic signature points; however, it requires the appropriate selection of algorithm parameters.

The usage of the IPAN99 algorithm could be useful e.g. in case of preliminary selection of the signatures. It could allow the comparison of the studied signature only with a certain group of signatures, which shall be qualified as similar. The again comparison would be performed by means of the analysis of all signature points.

The achieved research results are encouraging and may be the basis for further research and analyses in that direction. In the next work stage, the appliance of all

information kinds is expected (coordinates *x*, y, pressure, signature time, etc.) in the classification process. Also the dynamic selection of the IPAN99 algorithm parameters is an interesting issue for further research.

BIBLIOGRAPHY

[1] RATH T. M., MANMATHA R., Word Image Matching using Dynamic Time Warping, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, Madison, Wisconsin. 2003, pp. 521-527.
[2] SAEED K., ADAMSKI M., Klasyfikacja podpisu offline z wykorzystaniem metody DTW, – Krajowa Konferencja Naukowa - KBIB'05, tom I, Wydawnictwa Politechniki Częstochowskiej, Częstochowa, Polska. 2005, pp 455-460.
[3] PARISSE C., Global Word Shape Processing in Off-Line Recognition of Handwritting. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1996. Vol. 18, no. 5, pp. 460-464.
[4] DOROZ R., WIDUCH S., GOŁUCH P., WRÓBEL K., A method of objects similarity searching using edge points with the highest curvature, TIBRD 04. 2004, pp. 126-135.
[5] DOROZ R., WIDUCH S., Estimation of signature similarity by means of IPAN99 algorithm. XXXVIII International Autumn Colloquium ASIS 2006 Advanced Simulation of System. Ostrawa. Czech Republic. 2006 pp. 85-90.
[6] CHETVERIKOV D. SZABO Z., Detection of high curvature points in planner curves. In 23rd Workshop of the Austrain Pattern Recognition Group. 1999, pp. 175–184.
[7] LEI H. PALLA S., GOVINDARAJU V., ER^2 : an Intuitive Similarity Measure for ON-line Signature Verification Frontiers in Handwriting Recognition. 2004. IWFHR-9 2004. Ninth International Workshop, pp. 191-195.
[8] FENG H., WAH CH.CH., Online signature verification using a new extreme points warping technique Source, Pattern Recognition Letters. 2003. Vol. 24 , pp. 2943 – 2951.
[9] WIROTIUS M, RAMEL J.Y., Vincent N., Distance and Matching for Authentication by On-Line Signature, Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05). pp. 230-235.
[10] 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.