Rafal DOROZ[1], Malgorzata PALYS[1], Tomasz ORCZYK[1], Hossein SAFAVERDI[1]

# METHOD OF SIGNATURE RECOGNITION WITH THE USE OF THE COMPLEX FEATURES

In this paper a new method of handwritten signatures verification has been proposed. This method, for each signature, creates complex features which are describing this signature. These features are based on dependencies analysis between dynamic features registered by tablets. These complex features are then used to create vectors describing the signature. Elements of these vectors are calculated using measures proposed in this work. The similarity between signatures is assessed by determining the similarity of vectors in the compared signatures. Research, whose results will be presented in the further part of this work, have shown a high efficiency of verification using proposed method.

## 1. INTRODUCTION

The gap in security issues, which occurred in the computerization and automation era, is filled by biometric methods. A natural, reliable and very effective solution to these problems can be found in biometrics. It is possible to accurately and clearly identify a person by using partially or fully automated human recognition schemes on the basis of personal biological features. Biometrics can be defined as the use of physiological or behavioral characteristics for identification purposes. Physiological biometrics includes data derived directly from the measurement of any part of the human body e.g. fingerprints, iris, retina or the shape of the face. Among various behavioral characteristics that can be measured, signature is recognized as one of the most reliable, unique, undeniable, and unchanging characteristic for identifying persons. Behavioral biometrics, in turn, analyzes the data, which records the mode of a person's behavior, such as manner of speaking or signature dynamics [9], [14].

Hand-written signature constitutes one of the most popular biometric methods. It is commonly used, because of the ease of obtaining signatures as well as its legal and social acceptance [5], [6]. This method has been used for many years, inter alia in forensics, document authentication, and bank transactions authorization.

Data collection process within a signature recognition process can be divided into two categories: static and dynamic. The static system collects data using *off-line* devices [13]. A signature is put on the paper, and then is converted into a digital form with the use of a scanner or a digital camera. In this case, the shape of the signature is the only data source, without the possibility of using dynamic data. On the other hand, dynamic systems use *on-line* devices, which register, apart from the image of the signature, also dynamic data connected with it.

[1] University of Silesia, Institute of Computer Science, Bedzinska 39, 41-200 Sosnowiec, Poland
e-mail:{rafal.doroz, malgorzata.palys, tomasz.orczyk}@us.edu.pl

A tablet is an example of such device. Having the dynamic features values, it is possible to model the dynamics of pen movements when writing. The dynamics of writing is difficult to be forged, as it is a feature that is individual for each person. The analysis of acquired signature features allows either for its verification or for its identification. At present, there are many signature recognition techniques, including methods based on neural networks [16], Hidden Markov Models (HMM) [18], fuzzy sets [10], statistic computations [3], [8], [11], etc. In this work a new method has been proposed, which is based on using a completely new features, called "complex features", for signature verification.

## 2. METHOD DESCRIPTION

The method proposed in this work uses a new kind of dynamic features describing these signatures. Features, called "complex features", are created based on dependencies analysis between basic features, recorded by a tablet. Additionally, the proposed method eliminates the need for normalizing the length of analyzed signatures, which is required for some of methods known in the literature [17].
The algorithm proposed in this paper may be described by the following steps:
- feature extraction - extraction of the basic signature features using a tablet,
- complex feature creation - establishing new, complex features on the basis of simple features,
- creation of the vector $\mathbf{W}$ - establishing a vector describing the signature using complex features,
- signature verification - signature verification using previously established $\mathbf{W}$ vectors.

## 2.1. FEATURE EXTRACTION

By using tablet, a signature $S$ can be represented by the set of $n$ points:

$$S = \{s_1, s_2, ..., s_n\}, \tag{1}$$

where $s_j$ is the $j$-th point of the signature $S$, $n$ is the number of signature points.
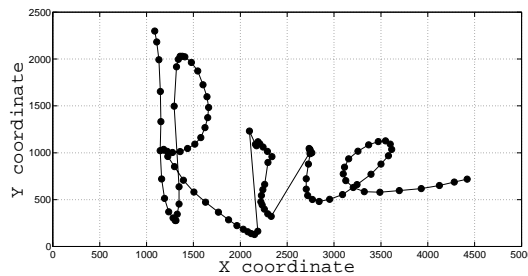Figure 1 presents the set of discrete points of signature captured by tablet.



Fig. 1. An example of signature $S$.

During a signing process the tablet is capable to measure in each point of signature many dynamics features. This implies that there is a feature vector $\mathbf{s}_j$ associated with each $j$-th point $s_j$ of signature $S$:

$$s_j \rightarrow \mathbf{s}_j, \quad \text{where} \quad \mathbf{s}_j = [f_{1,j}, f_{2,j}, ..., f_{m,j}]^T, \; j = 1, .., n, \tag{2}$$

where $f_{i,j}$ is the value of $i$-th feature registered in the $j$-th point $s_j$ of the signature $S$, $m$ is the number of all features recorded in each point of the signature.

Taking into account the all features, the signature $S$ can be described by the elements of the matrix $\mathbf{S}$:

$$\mathbf{S} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,n} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,n} \end{bmatrix}. \tag{3}$$

## 2.2. COMPLEX FEATURE CREATION

In the next stage of the method new complex features of the signature are established. These features are established by determining dependencies between dynamic features. Method of establishing complex features has two stages. In the first stage the individual columns of matrix $\mathbf{S}$ are assigned to one of two matrices, denoted as $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$. Assignation is based on analysis of the feature value, registered in $i$-th row of analyzed column of matrix $\mathbf{S}$. All columns have to be assigned using values from the same row. If value of the analyzed feature is lower than or equal to the threshold value "$Tr^i$" then the column is assigned to the matrix $\mathbf{S}^{(1)}$, in the other case it is assigned to the matrix $\mathbf{S}^{(2)}$.

$$\mathbf{S}^{(1)} = \{\mathbf{s}_j \in \mathbf{S} : f_{i,j} \leqslant Tr^i\}, \quad \mathbf{S}^{(2)} = \{\mathbf{s}_j \in \mathbf{S} : f_{i,j} > Tr^i\}, \tag{4}$$

where $i \in \{1, ..., m\}$, $j = 1, ..., n$ and $\mathbf{s}_i = [f_{1,i}, f_{2,i}, \cdots, f_{m,i}]^T$.
The threshold value $Tr^i$ is assumed to be an average value of all elements of the $i$-th row:

$$Tr^i = \frac{1}{n} \sum_{j=1}^{n} f_{i,j}, \quad f_{i,j} \in \mathbf{S}, \quad i \in \{1, ..., m\}. \tag{5}$$

The methodology of assigning points of signature to matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ has been illustrated in figures 2a-b. Figure 2a) shows an example of signature containing $n = 104$ points. Figure 2b) shows values of a dynamic feature (in the presented case it is the pressure of the pen on a tablet surface) registered in following $n = 104$ points of the signature. Additionally, an average value of $Tr^i$, obtained from values of all points, has been marked on the graph.
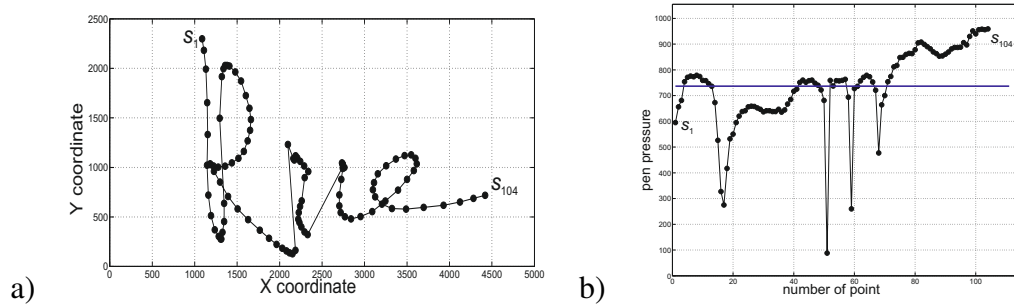


Fig. 2. The methodology of assigning points of signature to matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$.
a) a signature consisting of $n$=104 points, b) a graph of pen pressure feature value with the line representing an average value of the feature.

As it shows in the figure 2b) individual points of the signature have been registered with a various pen pressure. The analysis of the pen pressure value of the first three points of the signature indicates that it is lower than the average value $Tr^i$, and thus these points have been assigned to the matrix $\mathbf{S}^{(1)}$. The result of the classification of all three points has been shown in Figure 3. Points of signature assigned to the matrix $\mathbf{S}^{(1)}$ have been marked with white color, while points assigned to the matrix $\mathbf{S}^{(2)}$ have been marked with black color.
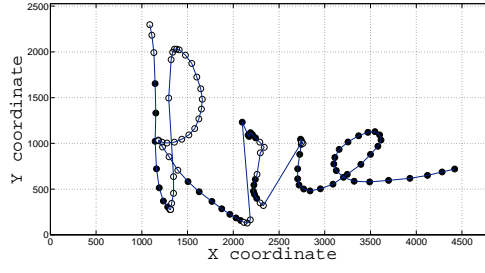
Fig. 3. Points of the signature assigned to the matrix $\mathbf{S}^{(1)}$ (white) and to the matrix $\mathbf{S}^{(2)}$ (black).

The exemplary establishing matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ from matrix $\mathbf{S}$ have been shown in the Figure 4.
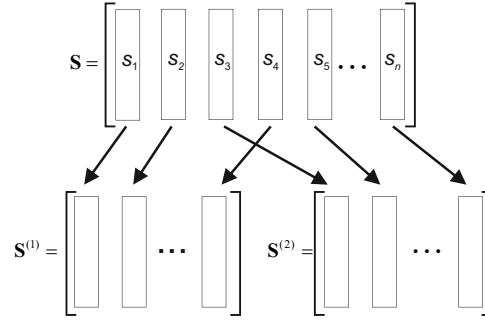


Fig. 4. The rules of establishing matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$.

As a result of assigning all the points of signature $\mathbf{S}$ we obtain the following matrices:

$$
\mathbf{S}^{(1)} = \begin{bmatrix} f_{1,1}^{(1)} & f_{1,2}^{(1)} & \cdots & f_{1,k}^{(1)} \\ f_{2,1}^{(1)} & f_{1,2}^{(1)} & \cdots & f_{2,k}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1}^{(1)} & f_{m,2}^{(1)} & \cdots & f_{m,k}^{(1)} \end{bmatrix}, \quad \mathbf{S}^{(2)} = \begin{bmatrix} f_{1,1}^{(2)} & f_{1,2}^{(2)} & \cdots & f_{1,l}^{(2)} \\ f_{2,1}^{(2)} & f_{2,2}^{(2)} & \cdots & f_{2,l}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1}^{(2)} & f_{m,2}^{(2)} & \cdots & f_{m,l}^{(2)} \end{bmatrix}, \tag{6}
$$

where $f_{i,j}^{(1)}, f_{i,j}^{(2)}$ is the $i$-th feature in the $j$-th column of the corresponding matrices $\mathbf{S}^{(1)}$ or $\mathbf{S}^{(2)}$, $k$ is the column count of the matrix $\mathbf{S}^{(1)}$, $l$ is the column count of the matrix $\mathbf{S}^{(2)}$.

Matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ have the same row count. Column count in these matrices may be different, but their sum have to be equal to column count of matrix $\mathbf{S}$.

Establishing $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ ends the first stage of obtaining the complex features of a signature. In the following stage a single row is selected from each of the matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$. The number of selected row in both matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ has to be the same. Selected $i$-th row of the matrix $\mathbf{S}^{(1)}$ creates a vector of a complex feature $\mathbf{FS}_i^{(1)}$, and by analogue a row of the matrix $\mathbf{S}^{(2)}$ is used for creating a vector of a complex feature $\mathbf{FS}_i^{(2)}$.

$$
\mathbf{FS}_i^{(1)} = \left[ f_{i,1}^{(1)}, f_{i,2}^{(1)}, ..., f_{i,k}^{(1)} \right], \quad \mathbf{FS}_i^{(2)} = \left[ f_{i,1}^{(2)}, f_{i,2}^{(2)}, ..., f_{i,l}^{(2)} \right], \tag{7}
$$

$$
f_{i,j}^{(1)} \in \mathbf{S}^{(1)}, \ j = 1, ..k, \quad f_{i,j}^{(2)} \in \mathbf{S}^{(2)}, \ j = 1, .., l, \ i = 1, ..., m.
$$

Complex features $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$ can be represented in a form of graphs. Two exemplary complex features have been presented in Figure 5.
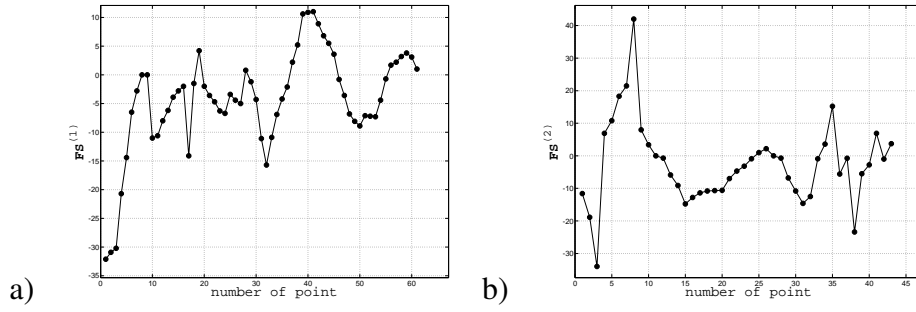
Fig. 5.   Exemplary complex features $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$.

## 2.3. CREATION OF THE VECTOR W

Many methods known from literature require the same length of feature vectors between which a similarity value is to be calculated [17]. This condition is hard to fulfill because even signatures coming from the same person often have different lengths. This implies the need for using methods for length equalization of compared data. Examples of such methods are DTW (Dynamic Time Warping) and FNP (Fixed Number of Points) [12], [17]. Usage of these methods may lead to decrease of verification accuracy [12], [17], [20]. Because of that, in the presented paper, a new method has been proposed. It is achieved by creating for each signature a vector $\mathbf{W} = [w_1, w_2, ..., w_6]$ describing this signature. Using elements of vectors $\mathbf{W}$ processed signatures are classified as genuine or forged. The advantage of using $\mathbf{W}$ vectors is the fact that they always have the same length, which does not depend on the length of the signature they describe. This eliminates a need for equalizing lengths of compared signatures.

The vector $\mathbf{W}$ consists of six elements. Value of the element $w_1$ from $\mathbf{W}$ vector describes dependency between $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$ complex features count. If their count is equal, then $w_1 = 1$. Along of a rise of the disproportion of the features count, the value of $w_1$ is decreased. Value of the $w_1$ is calculated in the following way:

$$w_1 = \frac{\min\{k, l\}}{\max\{k, l\}}, \quad w_1 \in [0, 1]. \tag{8}$$

Value of $w_2$ describes a dependency between the greatest values from $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$.

$$w_2 = \min \left\{ \frac{\max\{\mathbf{FS}_i^{(1)}\}}{\max\{\mathbf{FS}_i^{(2)}\}}, \frac{\max\{\mathbf{FS}_i^{(2)}\}}{\max\{\mathbf{FS}_i^{(1)}\}} \right\}, \quad i \in \{1, ..., m\}, \; w_2 \in [0, 1]. \tag{9}$$

For calculating a value of $w_3$ it is required to determine the index of the greatest value in the vector $\mathbf{FS}_i^{(1)}$. The $w_4$ element is calculated in the same way, but using $\mathbf{FS}_i^{(2)}$ vector.

$$w_3 = \frac{\arg\max\limits_{j=1,...,k} \left\{ f_{i,j}^{(1)} : f_{i,j}^{(1)} \in \mathbf{FS}_i^{(1)} \right\}}{k}, \; w_4 = \frac{\arg\max\limits_{j=1,...,l} \left\{ f_{i,j}^{(2)} : f_{i,j}^{(2)} \in \mathbf{FS}_i^{(2)} \right\}}{l}, \; i \in \{1, ..., m\}. \tag{10}$$

The $w_5$ element of the $\mathbf{W}$ vector describes a number of the extremes $NE^{(1)}$ determined on the graph of a complex feature $\mathbf{FS}_i^{(1)}$. The value of $w_6$ is the number of extremes $NE^{(2)}$ on the graph of a complex feature $\mathbf{FS}_i^{(2)}$. For calculating a number of the extremes an algorithm [19] has been used. The number of extremes allows to determine a frequency of value changes in the analyzed feature. Calculated number of extremes is normalized to the range $[0, 1]$.

$$w_5 = \frac{NE^{(1)}}{k}, \quad w_6 = \frac{NE^{(2)}}{l} \tag{11}$$

## 2.4. SIGNATURE VERIFICATION

The last element of the proposed method is the signature verification. Verification is done by means of classification of **W** vectors calculated for each signature. Proposed method allows usage of almost any available classifier. In the presented work four different classifiers have been tested in their so called "out of box" configuration. All those classifiers were suited to work with such data, but each belongs to a different family of classification algorithms. Also all these classifiers can return so called "support value", which describes a probability that analyzed sample belongs to a given class.

The verification stage is preceded by a classifier training stage. Training data consists of **W** vectors calculated for genuine as well as forged signatures of a given person. In the result of classification a signature is assigned to one of two classes: genuine signatures or forged signatures.

## 3. RESEARCH

The effectiveness of the method discussed in this paper was assessed experimentally. The tests were performed using signatures from the MCYT signature database [21]. Proposed classifier has been used in verification mode. The test database included signatures from 100 people. From each individual 12 original signatures and 12 forged signatures were selected randomly from the database. Verification of a test signature has been done by comparing it against a set of 12 genuine and 12 forged signatures of a person being verified.

Classification of test signatures has been performed using many classifiers, proven in literature and implemented in the WEKA system [1], [7]:

- Random Forest - forest of random trees (RanF) [2],
- PART - PART decision list (PART) [4],
- $k$-Nearest Neighbours Classifier ($k$-NN) [15].

During the research an efficiency of verification has been tested. It is based on a complex features created using common and easy to calculate dynamic features. These features were: $\mathbf{Vx} = [vx_1, vx_2, ..., vx_n]$ - vector of the horizontal velocity of the pen in successive signature points, $\mathbf{Vy} = [vy_1, vy_2, ..., vy_n]$ - vector of the vertical velocity of the pen in successive signature points, $\mathbf{Vxy} = [vxy_1, vxy_2, ..., vxy_n]$ - the general pen velocity defined between successive points, $\mathbf{P} = [p_1, p_2, ..., p_n]$ - the vector of pressure in successive points of signature, where: $n$ - number of signature points.

Thus our **W** vector contains four dynamic features in each of the $i$-th point of the signature $S$.

$$\mathbf{s}_j = [vx_j, vy_j, vxy_j, p_j]^T.$$

Using four dynamic features $Vx$, $Vy$, $Vxy$, $P$ in the research allowed to create four pairs of matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ - each of the pairs has been created using a different feature. Next, from each pair of the matrices, four vectors of the complex features have been created - each vector for a different feature. So in total, during the research, 16 different complex features have been created. The usefulness of each of the features had been experimentally examined. For this purpose **W** vectors have been created for each complex feature and used for signature verification. The results obtained are shown in Table 1.

Results presented in Table 1 shows that the best verification accuracy of 96.67% has been achieved when using a complex feature created using two features: $Vxy$ speed and $Vy$ speed. Definitely the highest classification errors have been achieved when using $Vx$ feature for creation of complex features. In case of this feature there was no difference whether it was

Table 1. Accuracy [%] achieved using various features and various classifiers.

| creating matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ | creating vectors $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$ | Random Forest | PART | 5-NN | 1-NN |
|---|---|---|---|---|---|
| | | | | | |
| | Vx | 89.75% | 78.67% | 91.25% | 95.67% |
| | Vy | 94.17% | 86.58% | 76.08% | 87.33% |
| Vx | Vxy | 93.25% | 85.83% | 95.42% | 96.25% |
| | P | 89.25% | 79.08% | 92.00% | 94.25% |
| | Vx | 94.25% | 85.75% | 77.75% | 89.58% |
| | Vy | 91.75% | 87.33% | 94.33% | 96.00% |
| Vy | Vxy | 89.92% | 82.25% | 91.33% | 93.67% |
| | P | 93.42% | 85.33% | 87.67% | 93.08% |
| | Vx | 91.58% | 81.00% | 92.33% | 95.58% |
| | Vy | 96.67% | 88.50% | 90.75% | 95.50% |
| Vxy | Vxy | 94.33% | 83.92% | 87.67% | 95.17% |
| | P | 95.58% | 86.08% | 75.50% | 87.17% |
| | Vx | 93.67% | 86.92% | 90.25% | 95.00% |
| | Vy | 92.42% | 83.08% | 87.42% | 93.17% |
| P | Vxy | 94.17% | 89.83% | 82.25% | 89.08% |
| | P | 93.75% | 85.75% | 76.33% | 86.17% |

The header spans "Features used for:" over first two columns and "Classifiers" over last four.

used for creating the $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ matrices, or for selecting the $\mathbf{FS}_i^{(1)}$ and $\mathbf{FS}_i^{(2)}$ vectors from these matrices.

When evaluating an impact of used classifiers to the method efficiency it can be noticed that the best results were achieved when using multinomial variant of Random Forest classifier, where achieved accuracy was equal to 96.67%. Definitely the worst efficiency have been achieved for the PART classifier.

## 4. CONCLUSIONS

Conducted research have proven a high efficiency of signature verification using the proposed method. Despite the preliminary stage of this research, achieved results can compete with contemporary methods known from the literature.

Current research has been conducted using only four features of a signature, but further research may include more signature features recorded by a tablet. Following research also include use of a different classifiers in the process of signature verification, as well as different signature databases for evaluating the method.

## BIBLIOGRAPHY

[1] BERTHOLD M. R., CEBRON N., DILL F., GABRIEL T. R., et al., KNIME: The Konstanz Information Miner. In: Studies in Classification, Data Analysis, and Knowledge Organization, Springer, 2007, pp. 319-326.

[2] BREIMAN L., Random forests, Machine Learning, 2001, Vol. 45, pp. 5-32.

[3] DEMSAR J., Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research, 2006, Vol. 7, pp. 1-30.

[4] FANK F., GENERATING I. H., Accurate Rule Sets Without Global Optimization, In: 15th Int. Conf. on Machine Learning, 1998, pp. 144-151.

[5] FAUNDEZ-ZANUY M., On-line signature recognition based onVQ-DTW, Pattern Recognition 40, 2007, pp. 981-992.

[6] FIERREZ-AGUILAR J., NANNI L., LOPEZ-PENALBA J., ORTEGA-GARCIA J., MALTONI D., An on-line signature verification system based on fusion of local and global information, The AVBPA05 Proceedings of the 5th international conference on audio and video-based biometric person authentication, 2005, pp. 523-532.

[7] HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P., WITTEN I. H., The WEKA Data Mining Software: An Update. SIGKDD Explorations, 2009, Vol. 11, No. 1, pp. 10-18.

[8] HARDLE W., SIMAR L., Applied multivariate statistical analysis, Springer-Verlag, Berlin-Heidelberg-New York, 2003.

[9]   IMPEDOVO S., PIRLO G., Verification Of Handwritten Signatures An Overview, 14th International Conference on Image Analysis and Processing - ICIAP, Modena, Italy, 2007, pp. 191-196.

[10]  KUDLACIK P., PORWIK P., A new approach to signature recognition using the fuzzy method, Pattern Analysis and Applications, 2014, Vol. 17, No. 3, pp. 451-463.

[11]  PALYS M., DOROZ R., PORWIK P., Statistical analysis in signature recognition system based on Levenshtein distance, The 8 Int. Conf. on Computer Recognition Systems, CORES, 2013, Poland, pp. 217-226.

[12]  PALYS M., DOROZ R., PORWIK P., On-line signature recognition based on an analysis of dynamic feature, In: IEEE Int. Conf. on Biometrics and Kansei Engineering (ICBAKE 2013), Tokyo Metropolitan University Akihabara, 2013, pp. 103-107.

[13]  PLAMONDON R., SRIHARI S. N., On-line and off-line handwriting recognition: A comprehensive survey, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, Vol. 22, No. 1, pp. 63-84.

[14]  RUUD M. B., CONNEL J. H., PANKANTI S., RATHA N. K., SENIOR A. W., Guide to biometrics, Springer, 2008.

[15]  SHAKHNAROVICH G., DARRELL T., INDYK P., Nearest-neighbor methods in learning and vision: theory andpractice, Neural Information, Processing, The MIT Press, 2006.

[16]  SPECHT D. F., Probabilistic Neural Networks, Neural Networks, 1990, Vol.3, pp. 109-118.

[17]  TAVENARD R., AMSALEG L., Improving the Efficiency of Traditional DTW Accelerators, Soumis a Knowledge and Information Systems (KAIS), Rennes, France, 2012.

[18]  VAN B. L., GARCIA-SALICETTI S., DORIZZI B., On using the Viterbi path along with HMM likelihood information for online signature verification, In: IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics, 2007, Vol. 37, No. 5, pp. 1237-1247.

[19]  VILLAVERDE K., KREINOVICH V., A linear-time algorithm that locates local extrema of a function of one variable from interval measurement results, Interval Computations, 1993, pp.176-194.

[20]  WROBEL K., DOROZ, R., Method of Signature Recognition with the Use of the Mean Differences, Proceedings of the 31st International IEEE Conference on Information Technology Interfaces (ITI 2009), Croatia, 2009, pp. 231-235.

[21]  http://atvs.ii.uam.es/databases.