# A Reduced Feature Representation based Online Signature Authentication

Youssef Elmir[1], Nasreddine Karour[2], Souleymane Ouled Jaafri[2]
[1]Department of Mathematics and Computer Science
University of Tahri Mohammed of Bechar
Algeria
elmir.youssef@yahoo.fr

[2]Department of Mathematics and Computer Science
University of Adrar
Algeria

***ABSTRACT:*** *In order to achieve perfection in security field, the biometric recognition has been used as a tool, because it is based on physiological and behavioral characteristics of a person which unlike the traditional prevailing methods of human identification, cannot be forgotten, stolen, eliminated or changed with time. In this paper, a theoretical and general presentation of an online signature based biometric system using a reduced set of features. The authenticity of a signer is determined by comparing an input signature to a stored reference set (template) based on a limited number of simple extracted features like the total length of the lines, pressure, the distance between the first and the last point, the average of horizontal displacement...etc. The proposed system was compared to QU-PRIP system which used more extracted features, this system which is the winner system in ICDAR2011 Signature Verification Competition. After several experiments, the obtained results show that even though the overall system performance are not very good enough but it is approximately similar to the one of the QU-PRIP system that uses more extracted features, the fact that affect directly on the total space of representation and the total time necessary for authentication.*

**Keywords:** Handwritten Signature, Biometric Authentication, Dynamic Features, Form Features, Signature Verification Methods, Online Signatures

**Received:** 19 January 2019, Revised 1 March 2019, Accepted 9 March 2019

**DOI:** 10.6025/jisr/2019/10/2/39-47



## 1. Introduction

Human beings have qualities that distinguish them from the rest of creatures, although these qualities distinguish people from each other, including skin color, eye color, physical composition, sound, full facial appearance and so on; and by using the greatest gift of the creator namely mind, human were able to use these qualities that distinguish them, and that brought a scientific approach to use in the field of individual identity [1].

Starting from fingerprints which began to be used in the research sector and criminal to help identify the criminals and murderers, and still this trait is used today. But with the technological development, the computer revolution, the proliferation of large-scale computer networks, the increasing number of applications making use of such networks, the need for a quick systems to identify persons, the large number of criminal acts, thefts, and other fraud, make humans looking on their characteristics and try to use them as a way to identify persons, and then biometrics system resulted a system has the ability of authentication accurately, rapidly, reliably, and without touching the existing infrastructures.

This work is released to study the possibility of reducing the size of representation space of online signature templates without decreasing the performance of the recognition system by making experiments on QU-PRIP [2] signature database, to reach good recognition rates of individuals on a subset of this database. The most important part is to extract features from each signature in the database according to the defined vectors [3].

This paper is structured as follows; the second section includes a detailed working principle of the proposed approach to extract online signature features. In section three, the process of classification is given in details. Section four represents the experiment process and what it gives as obtained results. Finally, a conclusion of the complete work study is given.

## 2. Features Extraction

This is an important step in the recognizing of the online hand-written signature. According to the coordinates, curvatures and the recorded time information, a set of features of the signatures are obtained like time, length of strokes and speed, then obtain a resultant function using the Gauss function (named after Carl Friedrich Gauss, a function of the form in (1):

$$f(x) = ae^{\frac{-(x-b)^2}{2c^2}} \tag{1}$$

To calculate the probability density through other density functions also, then obtain the corresponding averages, variances and standard deviations which will be the unique features for the signatures (Figure 1). Mathematically, that can be translated to (2):

$$D = \sum_{i=0}^{N-1} \{(x_i - x_i+1)^2 + (y_i - y_{i+1})^2\}^{1/2} \tag{2}$$

With $D$ represents the total distance of the pen travelled on the hand written signature the Euclidian distance of all the points. With $xi$ is the coordinate in direction $x$ and $yi$ is the coordinate in direction $y$.

Speed $Vx$ and $Vy$ express the functions of time, which can be calculated with the following formulae in (3) and (4):

$$V_{x_m} = \frac{(x_{m+1} - x_m)}{(t_{m+1} - t_m)} \tag{3}$$

$$V_{y_m} = \frac{(y_{m+1} - y_m)}{(t_{m+1} - t_m)} \tag{4}$$

The speed between two consecutive critical points $V^*$, and the speed between any two points $V\bullet$ as shown in Figure 2, are calculated as the distance between those points, since the points are equidistant in time. Where $m = 0, 1... N - 1$.

$x_m$ is the coordinate in direction $x$, same about $y_m$ in direction $y$.

$t_m$ is the time of movement, $v_x$ is the speed at point tm in direction $x$, and $v_y$ is the speed at point $tm$ in direction $y$.

Acceleration $a_x$ and $a_y$ at point $t_m$ in directions $x$ and $y$ can be expressed in (5) and (6):

Speed plays an important role in the process of signature verification, which usually signatures take between 2 to 10 seconds

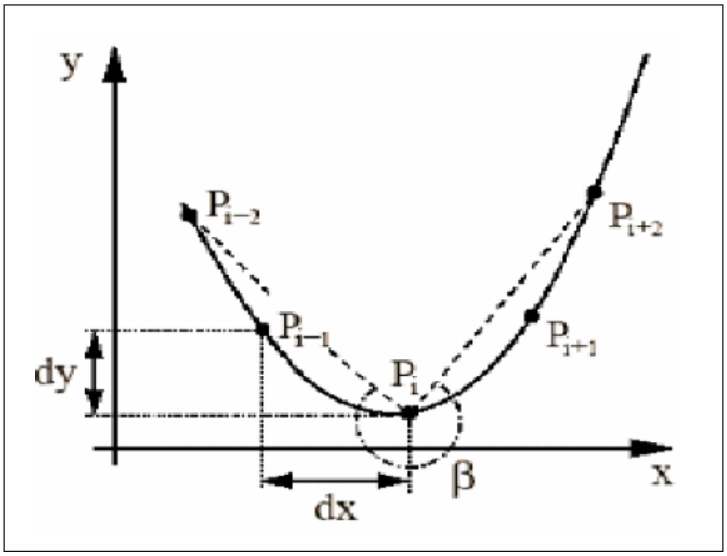



Figure 1. The features are computed at point $p_i$, using the preceding point $p_{i-1}$ and $p_{i-2}$ and the two succeeding points are $p_{i+1}$ and $p_{i+2}$

$$a_{x_m} = \frac{(v_{x_{m+1}} - v_{x_m})}{(t_{m+1} - t_m)} \tag{5}$$

$$a_{y_m} = \frac{(v_{y_{m+1}} - v_{y_m})}{(t_{m+1} - t_m)} \tag{6}$$

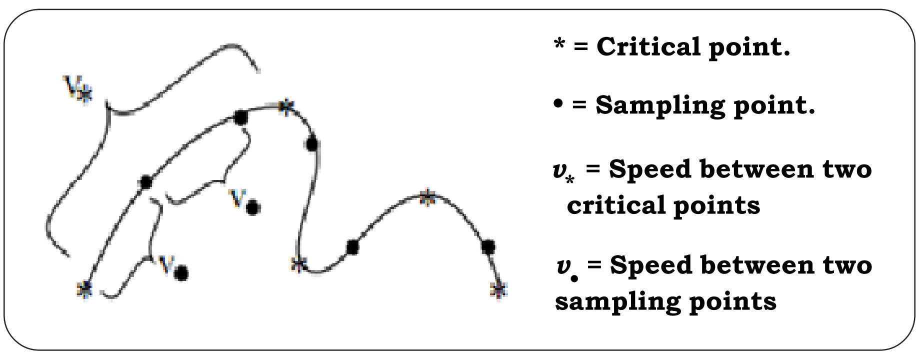



Figure 2. Computations of Speed Features

with an average time of 5 seconds, and the continuous practice play an important role in that. Based on the previous studies, the average amount of time for a genuine signature is usually 3 to 6 seconds, and for a forged was 10 to 11 seconds, due to slow writing in order to ensure the quality and try to reach the ideal. There are problems that arise when using the time dependent features is non-linear timing differences exist in signature made by the same person (Figure 3) due to physical or emotional state of the signer [4][5][6].

Not only speed which plays an important role, also the total time, the length of the strokes, the time for lifting the pen, which can symbolize one person's biometric features. Back to the stroke, all strokes are combined into one long stroke during pre-processing.

The original number of strokes is recorded and used as a global feature, and based on the coordinates $x$ and $y$, a set of local

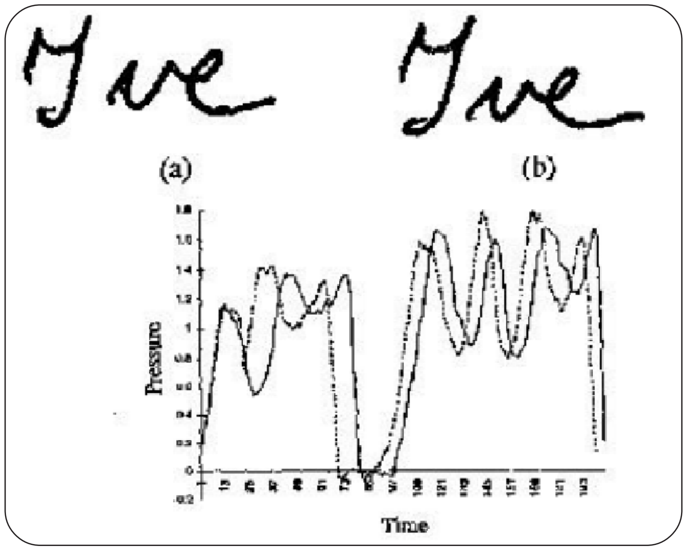



Figure 3. Depending on other properties. The two signatures are identical, but the difference appears depending on the time

features are extracted and divided into two categories: spatial features which is static features that are extracted from the shape of the signature (Figure 4), and temporal features. Local spatial features that are extracted and studied for their saliency for signature verification are:

- The $x$ and $y$ coordinate differences between two consecutive points and.
- The absolute $y$-coordinate with reference to the center of the signature, $y$.
- The sine and cosine of the angle with the x-axis, sin and cos.
- The curvature.

There are two types of features to be extracted, form features and dynamic features.

**2.1 Form Features**

According to [7]; the use of only dynamic features doesn’t allow us to identify the signer, while using the form allows the user to know why the signature is rejected. As examples of this type of features we have, the total length of the signature, the Euclidean distance of the signature and movement ratios. The most used form feature and the most features is the total length of the signature.

**2.2 Dynamic Features**

These features can be extracted only when the signing time is included with the signature vectors. As examples of this type of features we have, the total signing time, the average velocity and the average acceleration. The most used dynamics features are the total signing time and the average velocity by $X$ coordinate and by $Y$ coordinate and the most stable feature is the total signing time.

All these features are collected and structures in one column vector as feature vector for each individual.

**3. Classification Process**

To evaluate the performance of the extraction module, a toolbox is used, this toolbox is PhD_toolbox [8][9], the overall process works as follows:

After the feature extraction stage, for both systems, the signature feature vectors partitioned to samples for training, samples for evaluation and samples for testing, this step generates an $N \times M$ matrix, where $N$ is the number of individuals multiplied by the

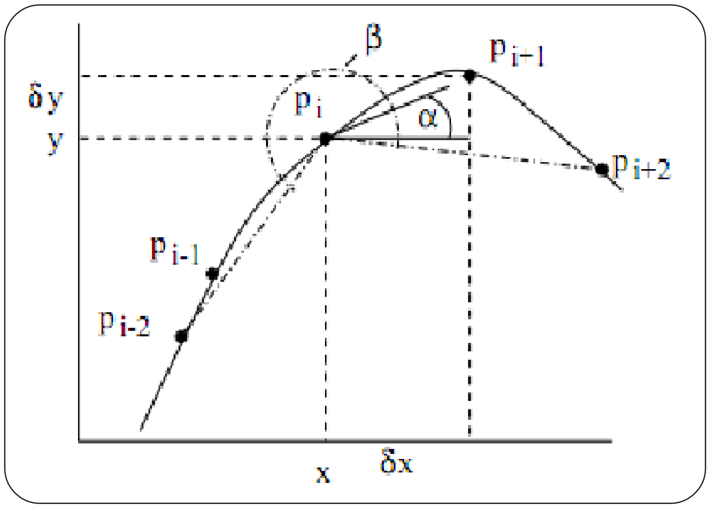



Figure 4. Spatial features extracted based on the coordinates $x$ and $y$

number of samples for each individual, and M is the number of the extracted features. By the end of this operation the process starts constructing principal component analysis (PCA) subspace by computing training, evaluation and test feature vectors using "PCA" for dimensionality reduction in case of big number of extracted features. In the following step, it computes the matching scores between gallery/training/target feature vectors and evaluation feature vectors using the Mahalanobis cosine distance similarity, the choice of this similarity measure is motivated by the results obtained in [10].

### 3.1 Mahalanobis Cosine Distance

In general, for normal (Gaussian) random variable $X$ with variance $S = 1$ and mean $\mu = 0$, any other normal random variable $R$ can be defined in terms of $X$ by (7):

$$R = \mu_1 + \sqrt{S_1}\, X \tag{7}$$

Conversely, to recover a normalized random variable from any normal random variable, typically (8) should be solved:

$$X = \frac{(R - \mu_1)}{\sqrt{S_1}} \tag{8}$$

Then, both sides are squared, and the square-root is calculated, (9) for a metric that looks a lot like the Mahalanobis distance:

$$D = \sqrt{X^2} = \sqrt{\frac{(R - \mu_1)^2}{S_1}} = \sqrt{(R - \mu_1)\, S_1^{-1}\, (R - \mu_1)} \tag{9}$$

The resulting magnitude is always positive and varies with the distance of the data from the mean, attributes that are convenient when trying to define a model for the data. Formally, the Mahalanobis distance of a multivariate vector $x = (x_1, x_2, x_3, \dots, x_N)T$ from a group of values with mean $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)\ T$ and covariance matrix $S$ is defined as in (10) (11):

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \tag{10}$$

The Mahalanobis distance (or "generalized squared interpoint distance" for its squared value (12)) can also be defined as a dissimilarity measure between two random vectors $x$ and $y$ of the same distribution with the covariance matrix $S$:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} \tag{11}$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called a normalized Euclidean distance:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{s_i^2}} \tag{12}$$

where $s_i$ is the standard deviation of the $x_i$ and $y_i$ over the sample set.

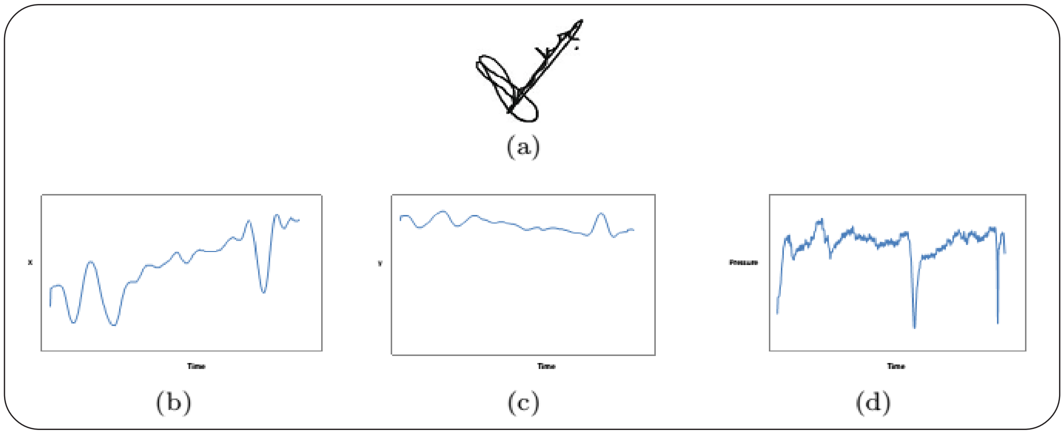



Figure 5. Example of a signature and the corresponding *X*, *Y* and Pressure signals [13]

## 4. Experiments and Results

### 4.1 Datasets

To perform this research QU-PRIP database is used. Online signatures contain a set of samples, each sample corresponds to the point coordinates on the digitizing tablet along with the corresponding pressure ($X_t$, $Y_t$, $P_t$) where $t$ correspond to time (see Figure 5). In this study, we used online signatures acquired using a WACOM Intuos4 digitizing tablet with a sampling rate of 200 Hz, a resolution of 2000 lines/cm and a precision of 0.25 mm. The pressure information is available in 1024 levels. 194 volunteers participated in the data collection process. They were instructed to provide occurrences of their natural signatures, and then to change their signatures in order to deny their identity at a later stage. Other volunteers were then asked to produce a simulation of the genuine signatures that they could see [13]. However, the first fifty individuals have only three samples as reference signatures and three samples as forgery signatures, and to unify the number of samples for each individual of the database, a subset is constructed using the individuals from 51 to 138, for each individual four samples are used for training, one sample for evaluation and the last one for testing.

### 4.2 Extracted Features

Since the signing time is not included in the database (QU_PRIP), only form features are extracted by the proposed system. Table 1 shows the extracted features. These 13 features are used to construct a feature vector of 13 elements for each sample of each individual where the system proposed by QU-PRIP generates from the same sample a feature vector of 2802 elements (Table. 1).

### 4.3 Results

To make tests some samples of each database are chosen for training, and some other samples for evaluation and testing. Tests are made on the 88 chosen using the 6 samples, 4 of them are used for training, the 5th one is used for evaluation and the last one is used for testing. The following tables and graphs show the obtained results.

Figure 6 shows the obtained results as a cumulative matching characteristic curves for both systems, it is clear that in identification task, results are much similar and form rank 5 the proposed system gives better results than the QU-PRIP system, however, for verification purposes, the proposed system still need some enhancement especially in equal error rate because QU-PRIP still better in this task as it is shown in the receiver operating characteristic curves in Figure 6.

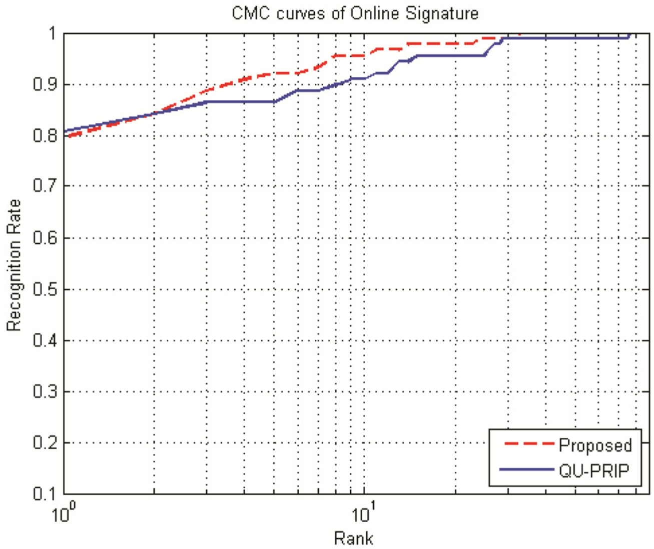
| Feature | Formula |
|---|---|
| Total lenth of the line | $\sum_{i=2}^{n-1} (dist_{Encl} (Pt_i, Pt_{i+1}))$ |
| Distance between the first and last point | $dist_{Encl} (Pt_1, Pt_n)$ |
| Ratio of movements to right and to left | $\frac{-\sum_{i=1}^{n-1} max (x_i - x_{i+1}, 0)}{\sum_{i=1}^{n-1} max (x_i - x_{i+1}, 0)}$ |
| Ratio of movements to up and to down | $\frac{-\sum_{i=1}^{n-1} max (y_{i+1} - y_i, 0)}{\sum_{i=1}^{n-1} max (y_{i+1} - y_i, 0)}$ |
| Ratio of movements following $X$ and following $Y$ | $\frac{-\sum_{i=1}^{n-1} max (x_i - x_{i+1}, 0)}{\sum_{i=1}^{n-1} max (y_i - y_{i+1}, 0)}$ |
| Angle formed by the horizontal line and that joins the first and last point measurements | $arctan \frac{y_n - y_1}{x_n - x_1}$ |
| Average Velocity | $\|\|AvgV\|\| = \sqrt{V_x^2 + V_y^2}$ |
| Average horizontal movement | $\frac{Max (X) - Min (X)}{n}$ |
| Maximum pressure | $Max (Pre)$ |
| Average pressure | $\frac{\sum_{i=1}^{n}(Pre (i))}{n}$ |
| Ratio of pens up and Number of stroke sequences or zero pressure | $\frac{\sum_{i=1}^{n}(Pre (i) \neq 0)}{\sum_{i=1}^{n}(Pre (i) = 0)}$ |
| Number of sroke sequences | $\sum_{i=1}^{n}(Pre (i) = 0)$ |
| Pen up sequences | $\sum_{i=1}^{n}(Pre (i) \neq 0)$ |

Table 1. Feature extracted from the QU-PRIP database [3]

The same observation is noticed in Table 2, for identification and verification purposes, performances of both systems are approximately similar where it is very clear that QU-PRIP system needs more time (~ 5/4 times) than the proposed system, the fact that make it faster with a small required space for representation.

## 6. Conclusion

This work proposes a reduced feature representation of online signature for authentication purpose. Evaluation and experiments are conducted on QU-PRIP database. The obtained results and performances show the benefit gained by the proposed system against QU-PRIP system in the space representation by using only 13 features instead of 2802, the fact that reduce the space necessary for representation of signature references (templates) and by consequence the time necessary for computing similarity



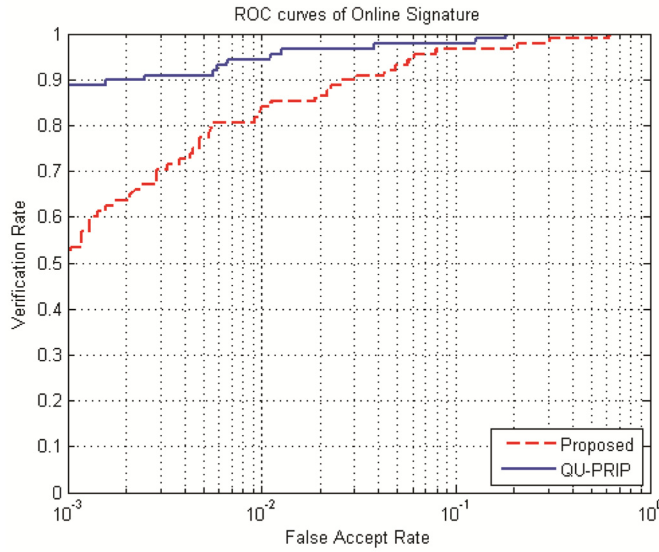




Figure 6. Cumulative matching characteristic and Receiver operating characteristic curves using 88 individuals from QU_PRIP

| Metric | Proposed | QU-PRIP |
|---|---|---|
| Recognition Rate (%) | 79.55 | 80.68 |
| Equal Error Rate (%) | 5.68 | 3.40 |
| Minimal Half Total Error Rate (%) | 5.33 | 2.34 |
| Verification Rate at 1% FAR (%) | 84.09 | 94.32 |
| Verification Rate at 0.1% FAR (%) | 52.27 | 88.64 |
| Verification Rate at 0.01% FAR (%) | 20.45 | 81.82 |
| Total Time for Evaluation computing (s) | 1.931 | 11.736 |
| Total Time for Test computing (s) | 2.533 | 11.487 |

Table 2. Performance metrics extracted using 88 individuals from QU-PRIP database

measure between references and target signatures. The lack of the dynamic features because of the absence of time data in the original captured data affect the overall performance of both systems, this factor should be considered in future works to enhance the performance of the proposed system.

**References**

[1] Jain, A. K., Ross, A., Prabhakar, S. (2004). An introduction to biometric recognition. Transactions on Circuits and Systems for Video Technology. IEEE. 14, 4-20. 10.1109/TCSVT.2003.818349, 30 January 2004.

[2] Al-maadeed, S, et al. (2012). Arabic signature verification dataset. *In*: Proceedings of the International Arab Conference on Information Technology.

[3] Karour, N., Ouled Jaafri, S. (2013). Online Signature Based Biometric System. University of Adrar, BSc thesis, 2013.

[4] Martens, R., Claesen, L. (1997). Dynamic programming optimisation for on-line signature verification. *In*: Proceedings of the

fourth International Conference on Document Analysis and Recognition. IEEE. (2) 653-656.

[5] Martens, R., Claesen, L. (1996). On-line signature verification by dynamic time-warping. *In*: Proceedings of the 13th International Conference on Pattern Recognition. IEEE. (3) 38-42.

[6] Phillips, P., et al. (1996). FERET (face recognition technology) recognition algorithm development and test results. Adelphi University, 1996.

[7] Nalwa, V., S. (1997). Automatic on-line signature verification. *In*: IEEE. 85 (2) 215-239.

[8] Struc, V., Pavešic, N. (2010). The complete gabor-fisher classifier for robust face recognition. *EURASIP Journal on Advances in Signal Processing*.

[9] Struc, V., Paveši, N. (2009). Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20, 115-138.

[10] Elmir, Y., Elberrichi, Z., Adjoudj, R. (2014). Multimodal Biometric Using a Hierarchical Fusion of a Person's Face, Voice, and Online Signature. *Journal of Information Processing Systems*, 10, 555-567.

[11] De Maesschalck, R., Jouan-Rimbaud, D., Massart, L., D. (2000). The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50, 1-18.

[12] Gnanadesikan, R., Kettenring, R., J. (1972). Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics, 81-124.

[13] Hassaïne, A., Al-Maadeed, S. (2012). An online signature verification system for forgery and disguise detection. *In*: Proceedings of the International Conference on Neural Information Processing. Springer. 552-559. Berlin, Heidelberg. 2012.