Basic on-line handwritten signature features for personal biometric authentication

Bača Miroslav[*], Koruga Petra[*] Fotak Tomislav[*]

[*] Faculty for Organization and Informatics, Centre for Biometrics
Pavlinska 2, 42000 Varaždin, Croatia
{miroslav.baca; petra.koruga; tomislav.fotak}@foi.hr

Abstract – On-line handwritten signature-based personal authentication is still a challenging research topic. Although great efforts have been achieved in developing and defining a framework that systems for on-line authentication based on the handwritten signature of a person should adhere, these frameworks are still not enough because they do not include all the features that handwritten signature as a biometric feature has. In addition, there is a range of features that current system for capturing a signature posses by which it is possible to further and better define the characteristics of signatures required to be in the process of authentication of entities, but can be used not only for authentication but also for identification. The paper provides an overview of the differences between off-line and on-line mode of authentication system based on the handwritten signature as well as surveys of some properties that these features can have, supported with the latest systems to "capture" handwritten signatures.

I. Introduction

Application of a handwritten signature on a daily basis is a normal action that people do several times a day. Most of the time, handwritten signature is given as consent for an action or set of actions that have or may have an impact on a person. Handwritten signature as a method of authentication has become a part of everyday life, and there probably is not a single area in which it is not used. Handwritten signature usage dates from ancient times and is held until today as a mean of giving consent to something that needs to be done.

The problem arises when someone decides to imitate the signature of the person with the purpose of fraud or false representation. Therefore, there is a need for adequate protection of personal signatures. One good method of protection is the use of biometric systems based on handwritten signature. Every person has a different way of handwritten signing that only he/she knows how to repeat an indefinite number of times (though not entirely identical). For the purpose of this work, biometrics and biometric systems will be considered in the context of authentication/identification of person through his/her behavioral and/or physical characteristics [1]. Handwritten signature is defined as the first and last name written in your own handwriting [2]. It is often the case that the signature does not contain the full name but only one part of it, or sometimes a set of connected lines that do not resemble in the name of the signer. This type of signature is called paraph, and is defined as an abbreviated signature, sometimes only the initial letter of the name, and placed on administrative act within the routine procedure, which means that whoever put the initials agrees to endorse the content [2]. Signature is often equated with a person's handwriting. In the context of this work that is possible because all a person needs to leave in the system is the signature that will contain the basic elements of the handwriting, and some extracted characteristics will be based solely on the characteristics of handwriting. In other cases, it is necessary to distinguish between these two terms because the signature is only one factor that contains the features of the handwriting.

A. Forgery

To an ordinary person forging a signature can seem easy. It is a set of lines that are easy to imitate. But even in off-line systems that is not easy. This is part of criminology, i.e. forensic handwriting analysis because "the question of the origin of handwriting dates from ancient times when the writings and signatures appeared in the legal life as a means of communication and legalization of Legal Affairs" [3]. There are three basic methods for handwriting and signature expertise [3]:

- Overhead method - compares the properties of individual letters
- Graphometric method - determines the quantitative characteristics of individual signatures and the measurement of specific characteristics of the signature
- Contemporary graphics method - a combination of overhead method from which it takes a comprehensive and detailed examination of the letter forms, and other characteristics of the signature.

Handwritten signature can be forged in three ways [3]:

- Forging a signature by copying - can be done in several ways, which include placing a sheet of writing paper on a document with a true signature and drawing heavy pencil strokes across an authentic signature on paper to leave imprinted grooves which draw the signature,

- Forging signatures by mimicking - in this way a forger writes a forged signature based on a sample of a true signature in front of him. This process can be slow, but also fast and trained,
- Free forgery of signatures - forger doesn't write a disputed signature based on the authentic signatures, but he remembers it or writes quite freely because he does not know what the original signature looks like.

II. Static and dynamic signature

Analysis of 140 different sources in a signature domain, which is described in [4] separates the basic lines of data collection and preprocessing followed by signature characteristics extraction, verification and system performance measurement. In this context it is presented that for the extraction and preprocessing of signatures DTW[1] algorithms are commonly used, while for signature characteristic extraction neural networks and genetic algorithms are used. It was noted that the in phase of verification Euclidean and Mahalanobis distance as well as HMM[2] are used. In the system performance analysis (FAR[3], FRR[4], EER[5]) in the static analysis of signatures the best results are achieved by systems which segment the signature using the network, then those who use graphometric methods, neural networks and HMM. In the analysis of dynamic signatures dominate systems using discrete wavelet transforms, Fourier transforms and other functional properties.

A. *Off-line (Static) signature*

Systems that deal with off-line signatures are based on statistical indicators, primarily graphs and HMM. Application of HMM in the analysis of signatures is relatively new. In [5] is described the application of graphometric properties such as pixel density and axial slant of certain parts of signatures. As part of that, signature is segmented by the grid and is divided into four zones. Each zone contains a column containing the cell with horizontal and vertical projections, which are later converted into vector with assigned numerical value that describes the signature alone. Using vector quantization vectors are converted into series of symbols depending on codebook. Using HMM the learning process and verification of signatures is implemented.

Approach based on the global characteristics of the signature alone is described in [6]. On the binary image of signature discrete Radon transform is applied where each column represents a projection or shadow of a signature at a certain angle. In this way the function of total pixels in the image and the intensity of a given pixel are calculated. It uses a defined number of non-overlapping beams at an angle for each given angle. Every such sequence (for each angle) is represented with a vector and later modeled using HMM. Flexible methods for generating additional samples in order to better learn signatures from a limited number of samples is described in [7]. The classification and verification, which is described, is based on Mahalanobis distances that can be calculated after the central vector and full covariance matrix are calculated and trained. Statistical model is also used in [8]. Using GSC[6] approach for the feature extraction on local, medium and large scale, paper presents global, statistical, geometrical and topological features. A unique approach to verification using global features such as signature pixels gradient, statistical characteristics derived from the distribution of pixels, geometrical and topographical features are described in [9]. In contrast, a method that uses the geometric mean for the extraction of features does the verification through vertical and horizontal splitting of signature image [10]. Some approaches [11] involve several different models, so a system for verifying off-line signatures uses grayscale images from which information about High Pressure Points are extracted. Two images, one containing the High Pressure Points extracted and other with a binary version of the original signature, are transformed to polar coordinates where a pixel density ratio between them is calculated. Approach that includes graphs is also interesting, so in the method of signature verification [12] bipartite graphs are used. Verification is the problem of graph matching. Application of bipartite graphs is also used in [13] where the similarity of two signatures is measured by bipartite graph structure.

B. *On-line (Dynamic) signature*

Analysis of signature images leaves great opportunities for signature forgery and is not suitable to be used alone in on-line system. As a result of this thinking we have systems based on dynamic properties of signatures extracted on the spot, which means that the signer must be physically present during the signature extraction. For this purpose, a set of 42 features (29 dynamic and 13 static) was created, which subsequently, through normalization, came to 49 features [14]. Since this is a large number of features an algorithm was developed, that for every person decides the best subset based on registered signature. In [15] the problem of verification is approached by applying the DTW algorithm. In the first iteration the global features of signatures are extracted (average pressure, average speed, tilt means for writing, etc.) and signature vectors are compared by Euclidean distances. In the second iteration local characteristics are extracted through segmentation of the signature into series of strokes and the best match among them is found.

Application of HMM and neural network in analysis of the dynamic signature is visible in [16]. System extracts five signatures per person through a digitizing tablet, and each of them is rotated and scaled in order to be consistent with the first signature. Looking at the pen position, time, velocity and pressure parameters, during the signature verification process two neural networks with feed-forward architecture are used.

The first network deals with the computation of the pressure distribution data whilst the second estimates the

[1] DTW Dynamic Time Warping

[2] HMM Hidden Markov Model

[3] FAR False Accept Rate

[4] FRR False Reject Rate

[5] EER Equal Error Rate

[6] Gradient, Structural and Concavity

velocity magnitude. Both networks apply Lavenberg-Marquardt training algorithm. HMM is also used in [17]. From the initial set of 19 features, they reduced five that were not of influence to the system. The training process includes Baum-Welch and Segmential K-Means algorithms.

Fully dynamic system for extraction of signature is described in [18] in which the signature is extracted by Synaptics Touchpad specialized to send spatial information. The signatories had to sign with finger and they did not use their personal signature. They had the ability to create personal paraphs or use existing ones. From every signature ten characteristics were extracted. Data were shown in a matrix, and Hausdorff distance was used for the measure of similarity.

C. *Hybrid Solutions*

In addition to separating the dynamic and static signatures it is possible to use the so-called hybrid systems that work with both on-line and off-line signatures. One is presented in [19] and consists of on-line/off-line hybrid module that takes care of the collection and management of dynamic and static data of a signature using a digitizing tablet. The first module is the on-line module and is responsible for creating a "window" around the lines of a signature, which is later used in the learning process, in which global characteristics of a signature are viewed as well as local characteristics of each "window". The second module is an off-line module that processes the images of test signatures and compares them with reference data collected during the extraction.

Hybrid solutions could become universal solutions in the signature domain because they combine the best of static and dynamic signature system.

III. ARCHITECTURE OF AN AUTHENTICATION SYSTEM

Our system architecture includes two main modules. The first module is the user registration and the second is authentication. During the first encounter with the system user has to register in order to be able to use the system. Using digitizing tablet, user signature characteristics are extracted and signature features of interest are also extracted. After deducting all necessary parameters, statistical indicators such as the arithmetic mean, median, standard deviation, minimum values and maximum values are calculated. Calculated statistics are represented by vector and stored in a database.

In the process of authentication user is introduced using a handwritten signature and extracted characteristics are compared with the stored vector. If the user is authenticated, preferred activities are allowed, otherwise they are denied. The basic principle of the system is given in Figure 1.

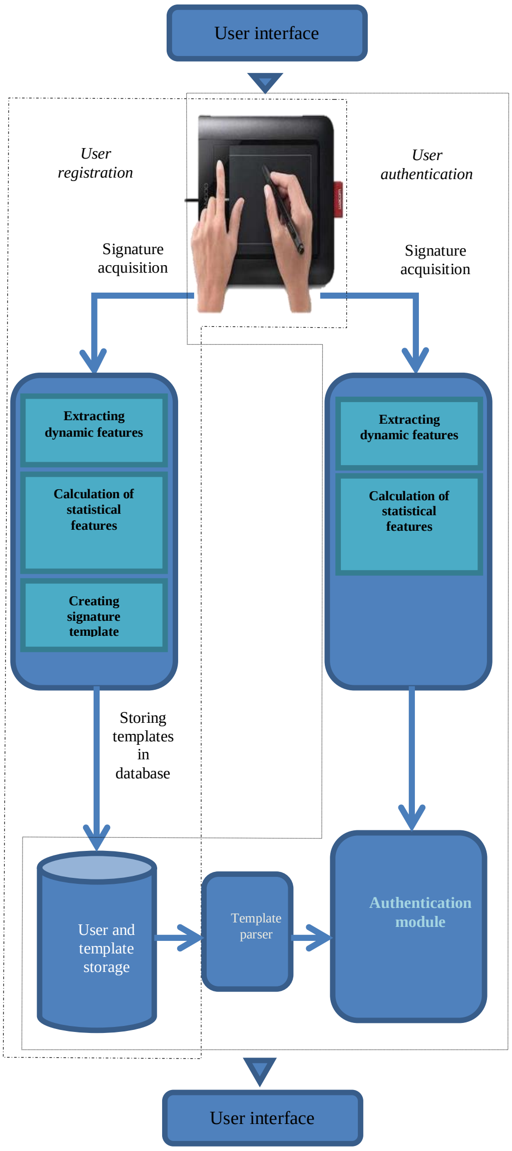

User interface
User registration
User authentication
Signature acquisition
Signature acquisition
Extracting dynamic features
Calculation of statistical features
Creating signature template
Extracting dynamic features
Calculation of statistical features
Storing templates in database
User and template storage
Template parser
Authentication module
User interface

Figure 1. The architecture of the handwritten signature-based system for registration and user authentication

A. *Characteristics of extracted dynamic signature*

In dynamic signature there is always the question of representative characteristics. Although so far 42 or 49 properties where used, for normal application of the system with satisfactory results following features are sufficient:

- Number of strokes - this feature is the total number of lines contained in the signature. One line is from the time since the user put down the pen to the contact surface until it is filed, i.e. until pen-up occur.
- Number of pen-ups - this feature shows how many times signer picked up a pen during signing. It should be noted that the last lifting of pen is not counted because it marks the end of the signing.
- Signature aspect ratio - feature takes into consideration the width of the signature (signature size on the x-axis) expressed in pixels of a tablet

and normalized on pixels of the screen and the height of a signature (the size of signatures on the y-axis) expressed in the same way that puts them in proportion. The assumption is that the user will sign each time the same in terms of creating a signature in one, two or more lines and that the size of signatures each time will be approximately the same.

- Signature length - feature represents length of all lines drawn in the signature first expressed in pixels of a digitizing tablet and carried out in the pixels of a screen. Tablet delivers the data about interaction and coordinates of writing resource in a form package. By calculating the Euclidean distance between two consecutive packages and adding it up to the total distance the total length of the signature is obtained.

- Signing time – feature expresses the total time needed to get a person to sign, usually in milliseconds, since the beginning of the signing. It is assumed that the time for a trained signature will always be nearly equal.

- Time-down ratio - describes how much of total signing time the pen was in contact with the singing surface. The rule is expressed in promille and its values range from 0 to 1. To a person who has trained signature, this characteristic will be fairly constant because it is directly related to the signing time.

- Time-up ratio - feature opposite to Time-down ratio, and indicates how long of total signing time a pen was separate from the signature area. This feature is used for authentication algorithm, which may favor one of two opposing feature in relation to represent a more stable signature characteristic of the person.

- Signature speed – feature is derived from the total length of the signature and the time in which the pen was in direct contact with the signing area. It tells us the speed of signing expressed in pixels per millisecond (px/ms). Trained signature should not have significant differences over the time. However, this feature strongly depends on the physical and mental condition of the person.

- Velocity along the x-axis - feature represents speed expressed in number of pixels per millisecond, which indicates how quickly people sign if we consider only the x coordinate system. It calculates the total length which pen passed along the x-axis and is divided by the total time in which the pen was lowered to the signing area. This feature depends on the physical and mental condition of the person and is often used less than other characteristics.

- Velocity along the y-axis - feature represents speed expressed in number of pixels per millisecond, which indicates how quickly people sign if we consider only the y coordinate system. It calculates the total length which pen passed along the y-axis and is divided by the total time in which the means for writing was lowered to the signing area. This feature depends on the physical and mental condition of the person and is often used less than other characteristics.

- Average pressure - feature is obtained by monitoring the level of pressure which pen leaves on the signing area. In order to obtain the average pressure it is necessary to add up all levels of the received pressure to one variable and divide by the total number of packages. The biggest influence on this characteristic has signers body.

- Strongest pressure moment - feature can be characterized as the only local characteristics of signatures which can be global as it is unique in the entire signature. In order to extract this feature, monitoring the level of pressure which pen leaves on the signing area is needed. The highest level is observed and the time of its creation is recorded. We assume that the signature always has nearly the same moment of the strongest pressure.

Twelve described features are only the beginning in the process of determining the ideal feature subset to be used in personal authentication. One could notice that these features are mainly global handwritten signature features. It does not mean that we disregard local features; we rather give the basic set of features that can be used to compute some others and can also be used on local level, e.g. we can determine all these features for each stroke.

B. *Registration*

This step is necessary due to the application of the system.

For the authentication needs it is necessary to use the arithmetic mean, standard deviation and median. These indicators are calculated for each feature separately.

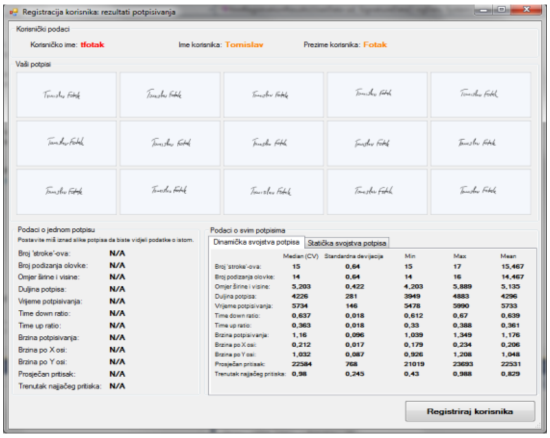
The arithmetic mean is the most common statistical indicator which needs to be calculated if an average that will continue to be used needs to be calculated. It takes into consideration all the elements together, adds to their value and divides the sum by the total number of elements in the set. Mathematically we can represent it with the formula $\bar{x} = \frac{1}{n} \cdot \sum_{i=0}^{n} x_i$. In the system it is used for the purpose of calculating the average value of each signature feature.

The average deviation of elements in the sets from mean value is expressed as the standard deviation. It also represents a measure of dispersion in the cluster. Within one standard deviation from the mean are about 68% of elements in set for the normal distribution, while within three standard deviations are approximately 99.7% of elements. Mathematically it can be described with the formula: $S_N = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2}$. Since the signature depends on many things and changes all the time, it is not possible to determine whether further signatures of the same user will be within given limit. Therefore, the system will use a slightly altered formula for standard deviation, which is usually used in calculating the population when the number of elements in the set is constantly growing and it is not possible to take into consideration the entire set. It gets a little more tolerance, and the formula is: $s_N = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2}$.

The next indicator is the central value or median. Certain features can have a large range of values and by calculation of its mean value it is possible to get a value that is actually not produced by any signature in the set, nor will it be caused by a subsequent signature. This would automatically mean that this characteristic will fail during the authentication. Therefore, the central value of set is introduced. The set is sorted and it exempts the value which is located right in the middle.

If number of elements in set is even, elements on N/2 and N/2+1 positions are used and their arithmetic mean is calculated. In ordered set S = {3.0, 3.5, 3.5, 3.6, 3.7, 3.9, 4.0} it is clear that the central value is 3.6. The same principle will be applied on a set of signature features.

The last two indicators that will be taken into account are the minimum value and maximum value in the set. By using the arithmetic mean and standard deviation the extreme elements usually cannot be reached. In order to allow a person to be about to repeat his/her signature which contains such values, these two indicators are taken and certain deviation is allowed.



Registracija korisnika: rezultati potpisivanja

Korisnički podaci

Korisničko ime: tfotak   Ime korisnika: Tomislav   Prezime korisnika: Fotak

Vaši potpisi

Podaci o jednom potpisu

Postavite miš iznad slike potpisa da biste vidjeli podatke o istom.

Broj 'stroke'-ova:	N/A
Broj podizanja olovke:	N/A
Omjer širine i visine:	N/A
Duljina potpisa:	N/A
Vrijeme potpisivanja:	N/A
Time down ratio:	N/A
Time up ratio:	N/A
Brzina potpisivanja:	N/A
Brzina po X osi:	N/A
Brzina po Y osi:	N/A
Prosječan pritisak:	N/A
Trenutak najjačeg pritiska:	N/A

Podaci o svim potpisima

Dinamička svojstva potpisa | Statička svojstva potpisa

	Median (CV)	Standardna devijacija	Min	Max	Mean
Broj 'stroke'-ova:	15	0,64	15	17	15,467
Broj podizanja olovke:	14	0,64	14	16	14,467
Omjer širine i visine:	5,203	0,422	4,203	5,889	5,135
Duljina potpisa:	4226	281	3949	4883	4296
Vrijeme potpisivanja:	5734	146	5478	5990	5733
Time down ratio:	0,637	0,018	0,612	0,67	0,639
Time up ratio:	0,363	0,018	0,33	0,388	0,361
Brzina potpisivanja:	1,16	0,096	1,039	1,349	1,176
Brzina po X osi:	0,212	0,017	0,179	0,234	0,206
Brzina po Y osi:	1,032	0,087	0,926	1,208	1,048
Prosječan pritisak:	22584	768	21019	23693	22531
Trenutak najjačeg pritiska:	0,98	0,245	0,43	0,988	0,829

Figure 2. The display of signature features with statistical indicators

Calculated statistical parameters are then written to the database in the form of a series of bytes (BLOB variable). The way all the data should be stored, if the classical method of notation through a series of characters is used, they would occupy much space and burden the system performance. Therefore, for this work purposes, and perhaps further development of the system, a special way of recording data in the database is proposed. It is based on the fact that the data are stored in binary format and are compressed by up to 10 times.

Binary representation of a feature vector:

```
000011110001101011001111000010100000000011110100010O
```

Hexadecimal:

```
0F-1A-CF-0A-00-F4-43-88-33-82-80-39-02-81-E5-14-C0-
D3-10-6B-2E-02-43-20-84-10-11-91-ED-A4-C4-CB-32-96-
66-01-24-55-98-BB-34-FF-3E-82-53-25-3C-B4-AD-60-92-
94-C2-09-30-24-40-0C-02-07-8A-8A-06-70-1A-80-08-81-
66-07-50-83-02-04-00-AE-1C-F0-97-16-00-EC-1C-03-00-
A4-37-72-37-3D-F5-0F-56-BB-DC
```

Generated signature vector will be used in the authentication process. Since it is well formed, we know exactly how to find certain value in it. With a little help of a vector (database string) parser we can get all the previously calculated statistical parameters. Considering these parameters there are numerous ways to verify the person's identity. Starting from the basic idea that allows signature feature to be consider as genuine if it satisfies the 'one deviation' rule (value must be within one deviation of arithmetic mean, maximum or minimum value) and setting threshold to some number (N ≤ number of features), all the way to the fuzzy approach, in our future work we will try to develop a system that is properly dealing with the problem of personal authentication using handwritten signature.

IV. Conclusion

In this paper we presented basic differences between on-line and off-line handwritten signature verification systems. It is well known that on-line systems are more robust and it is more difficult to forge signature when dealing with it. There are many ways to implement that kind of system, so we showed some with a special attention on a feature extraction. The importance of capturing these characteristics is enormous because it makes a base for further steps in verification and identification process. The proposed system defines 12 signature characteristics that are being used for user registration. We calculate their statistical parameters and store them in the database as the series of bytes. These characteristics and statistical parameters will be used in our future work which will bring authentication and identification procedure into the daylight. We will also try to implement biometric authentication based on the described features using a smart card. For now, we have given a good base for our future work.

Acknowledgments

Shown results come out from the scientific project Methodology of biometrics characteristics evaluation (016-01611992-1721) and technological project Multiple biometric authentication using smart card (2008-043) financed by the Ministry of Science, Education and Sport, Republic of Croatia.

LITERATURE

[1] M. Bača, “Uvod u računalnu sigurnost“, Narodne novine, Zagreb, 2004

[2] V. Anić i dr, „Hrvatski enciklopedijski riječnik“, Zagreb, Novi Liber, 2002

[3] Ž. Sobol, “Identitet rukopisa, Ogled o vještačenju rukopisa”, Informator, Zagreb 1986

[4] S. Impedovo, G. Pirlo, „Verification of handwritten Signatures: an Overview“, 14th International Conference on Image Analysis and Processing, ICIAP, Mondena, Italy 2007

[5] E.J. Justino, A. Yacoubi, F. Bartolozzi and R. Sabourin, “An Off-line Signature Verification Using HMM and Graphometric Features“, Proceedings of 4th IAPR International Workshop on Document Analysis Systems, Rio de Janeiro, Brasil 2000

[6] J. Coetzer. B. Herbst and J. Du Preez, “Offline Signature Verification Using the Discrete Radon Transform and Hidden markov Model“, EURASIP Journal of Applied Signal Processing 2004

[7] B. Fang, C. Leung, Y. Tang, P. Kwok, K. Tse and Y. Wong, “Offline signature verification with generated training samples“, IEEE Proceedings Vision, Image and Signal Processing 149, 2002

[8] M.K. Kalera, S. Srihari and A. Xu, “Offline Signature Verification and Identification Using Distance Statistics“, International Journal of Pattern Recognition and Artificial Intelligence 2004

[9] H. Srinivasan, S. Srihari and M. J. Beal, “Sigranture Verification Using Kolmogrov Smirnov Statistic“, Proceedings of International Graphonomics Society Conference, Salermo, Italy 2005

[10] B. Majhi, Y.S. Reedy and D.P. Babu, “Novel Features fo Off-line Signature Verification“, International Journal of Computers, Communications & Control, 1, 2006

[11] J.F. Vargas, M.A. Ferrer, C.M. Travieso and J.B. Alonso, “Off-line Signature Verification Based on High Pressure Polar Distribution“, Proceedings on Int Conf on Frontiers in Handwriting Recognition, Montreal, Kanada, 2008.

[12] I.S. Abduhaiba, “Offline Signature Verification using Graph Matching“, Turk Elc Eng, 15, 2007

[13] A.C. Ramachandra, K. Pavithra, K. Yashasvini, K.B. Raja, K.R. Venugopal and L.M. Patnaik, “Off line signature authentication using cross-validated graph matching“, Proceedings of the 2nd Bangalore Annual Compute Confernce, Bangalore, India, 2009

[14] L.L. Lee, T. Berger and E. Aviczer, “Reliable On-Line Human Signature Verification Systems“, IEEE Transaction on Pattern Analysis and Machine Intelligence, 18, 1996

[15] C.E. Pippin, „Dynamic Signature Verification Using Local and Global Features“, Technical Report, Georgia Institute of Information Technology, Atlanta, USA, 2004

[16] M. Musa, B.H. Lim, “Biometric signature verification using pen position, time, velocity and pressure parameters“, Jurnal Teknologi, 2008

[17] A. McCabe, J. Trevathan, “Handwritten Signature Verification Using Complementary Statistical Models“, Journal of Computers 4, 2009

[18] T. Alpcan, S. Kesici, D. Bicher, M.K. Mihacak, C. Bauckhage and S.A. Camtepe, “A lightweighr biometric signature scheme for user authentication over networks“, Proceedings oft he 4th international conference on Security and privacy in communication networks, Istanbul, Turkey, 2008.

[19] A. Zimmer, L.L. Ling, „A Hybrid On/Off Line Handwritten Signature Verification System“, Proceedings of the Seventh International Conference on Document Analysis and Recognition, Edinbourgh, Scotland, 2003