Automated Off-Line Writer Verification Using Short Sentences and Grid Features

Konstantinos Tselios, Elias N. Zois, *Member IEEE,* Athanasios Nassiopoulos, Sotirios Karabetsos, *Member IEEE* and George Economou, *Member IEEE*

***Abstract*—This work presents a feature extraction method for writer verification based on their handwriting. Motivation for this work comes from the need of enchancing modern eras security applications, mainly focused towards real or near to real time processing, by implementing methods similar to those used in signature verification. In this context, we have employed a full sentence written in two languages with stable and predefined content. The novelty of this paper focuses to the feature extraction algorithm which models the connected pixel distribution along predetermined curvature and line paths of a handwritten image. The efficiency of the proposed method is evaluated with a combination of a first stage similarity score and a continuous SVM output distribution. The experimental benchmarking of the new method along with others, state of the art techniques found in the literature, relies on the ROC curves and the Equal Error Rate estimation. The produced results support a first hand proof of concept that our proposed feature extraction method has a powerful discriminative nature.**

***Index Terms*—Writer Verification, Handwritten Sentences, Grid Features, ROC, EER**

I. Introduction

BIOMETRICS recognition is an appealing method for keeping numerous situations, including defense and economic transactions secured. Thus, access to important resources is granted by reducing potential vulnerability. Among other biometric features, online and offline handwriting, which is a subset of behavioral biometrics, has been frequently used for resolving the problem of recognizing writers either for security or forensic applications [1], [2]. In recent years, writer identification and verification tasks have received considerable attention among the scientific community. A special case of writer verification uses context based handwriting. So, the answer to the question: is this person who he claims to be? shall be provided by examining a predetermined text of known transcription. As stated by Siddiqi and Vincent [3] this kind of writer verification problem is similar to signature verification.

Although content dependent approaches using well defined semantics have been used at the early years of writer recognition there are at least three important reasons that justify the continuous study of handwriting patterns other than signatures. Firstly, biometric verification schemes based on handwritten words or small sentences can be potentialy used to real world security applications which are quickly emerging in a modern and continuous evolving mobile and Internet based environment. Secondly, content based retrieval systems could also benefit since their users could query handwriting images from various corpuses with similar handwriting styles [4]. Finally, an important reason emerges from the field of continuous verification [5]. By this, we mean that we could use the handwritten patterns, to grant access to resources not only to a person's initial entrance, but also within a cyclic and continuously verification loop, throughout the entire use of the application. In order to explore writer verification tasks, we can test a number of algorithms in a number of well established databases in the literature like IAM [6], Firemaker [7], CEDAR [8] and Brazilian Forensic letter database [9]. These databases carry rich handwriting information since they have a large sample size like 156 words and/or paragraphs. The use of these databases might bring around awkward circumstances if issues like those described in the continuous verification schemes need to be raised. This can be easily seen using the following example: Imagine the case that a person has to verify him/her by writing a entire letter in a relative small amount of time. In order to cope with this situation, an alternative idea would be either to use a portion of the afore-mentioned databases or to employ one small sentence content like the one provided by database like the HIFCD1 [10].

In this work, we are presenting a novel feature extraction method for writer verification based on the structured exploitation of the statistical pixel directionality of handwriting. This is achieved by counting, in a probabilistic way, the occurrence of specific pixel transitions along predefined paths within two pre-confined chessboard distances. Then, the handwritten elements described by their strokes, angles and arcs are modelled by fusing, in the feature level, two and three step transitional probabilities. This is an extension of the work proposed in [11] for signature verification.

E. N. Zois is with the Electronics Engineering Department, Technological and Educational Institute of Athens, Agiou Spiridonos Str., 12210, Aegaleo, Greece (phone: +302105387204 fax: +302105385304; e-mail: ezois@teiath.gr).

S. Karabetsos and A. Nassiopoulos are with the Electronics Engineering Department, Technological and Educational Institute of Athens Greece.

K Tselios is with the CMRI, University of Bolton, United Kingdom, (e-mail: kt1cmr@bolton.ac.uk).

G. Economou is with the Physics Department, University of Patras, 26500, Greece. (e-mail: economou@upatras.gr).

A two stage classification scheme based on similarity measures and an SVM has been enabled in the HIFCD1 corpus. The verification efficiency is evaluated by measuring the Equal Error Rate on the ROC curves, which is the point were the probability of misclassifying genuine samples is equal to the probability of misclassifying forgery samples. The EER is evaluated as a function of the word population. This is achieved by plotting the ROC curves each time we append a word for verification.

Finally, in order to benchmark our proposed method, comparisons are provided against recently described, state of the art methodologies for, off-line signature verification pre-processing and feature extraction, as well as writer verification and feature extraction approaches. Within this context, we are providing a feasibility study of the discriminative power of our method. This "feature benchmarking" concept can be justified by the fact that an ideal feature extraction method would make the classifier's job trivial whereas an ideal classifier would not need a feature extractor [12]. Thus, by keeping the classifier stage fixed, feature benchmarking could be rated in a comparative way.

The rest of this work is organized as follows: Section 2 provides the database details and the description of the feature extraction algorithm. Section 3 presents the experimental verification protocol which has been applied. Section 4 presents the comparative evaluation results while section 5 draws the conclusions.

II. Database and Feature Extraction Procedure

A. Database Description and Pre-Processing

In order to provide a confirmation of the proposed method and evaluate our approach, we have employed the HIFCD1 handwritten corpus which has been used formerly in the literature [10]. This corpus is under re-enlistment and enrichment since its initial appearance in 2000. The developed database consists of two different small sentences, one written in Greek and the other one in English. Additionally to the first twenty persons who have been enrolled in the past, another twenty persons have been enrolled later on creating a total temporary set of forty persons. This database is under restructuring in order to increase its size and diversity (e.g. include iris, fingerprints, gait, signatures, face, large scale handwritten text etc.) of biometric samples equivalent to these provided by modern databases like IAM [6] and BioSecure [13]. Each sentence was written by each writer 120 times. Consequently, 9600 sentences were recorded in our database containing a total of 48000 words. Both linguistic forms of the sentences are presented in Fig.1. The Greek language, being our native language, was used in order to maintain constant handwriting characteristics. The Greek sentence is made up of two small words of three letters, two medium length words of seven letters and a lengthy word of eleven letters. Each word has been created in its own cell thus making segmentation procedures trivial. For every word image of the corpus, pre-processing steps are applied in order to provide an enhanced image version with maximized amount of utilized information. The pre-processing stage includes thresholding of the original handwritten image using Otsu's method [14] and thinning in order to provide a one pixel wide handwritten trace, which is considered to be insensitive to pen parameters changes like size, colour and style. Finally, the bounding rectangle of the image is produced. It must be pointed out that we treat the handwritten image as a whole and we do not perform any character segmentation. Next, an alignment is carried out for every bounded image.

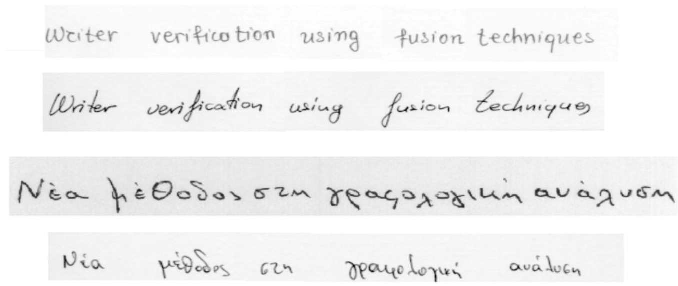

Fig. 1. HIFCD samples

This stage gathers the intrapersonal useful information from all the samples of a writer inside a region that is considered to be the one that contains the most useful handwriting information [9], [11]. In this work, we have used the estimated coordinates of the centre of mass $\bar{x}$ and $\bar{y}$ for each image. Fig. 2 presents in a graphical way the above discussion. In this work the term 'most informative window' (MIW) of the handwritten pattern is presented by considering the processed handwritten word sub-region, inside the bounded image, centred at $\bar{x}$ and $\bar{y}$ parameters while its length and width are determined empirical with trial an error method.

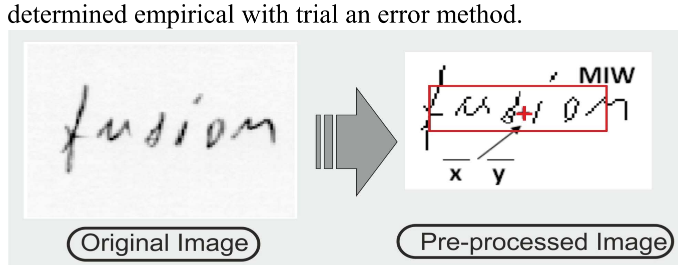

Fig. 2. Original and pre-processed handwritten image with MIW

B. Feature Extraction

The feature extraction method maps the handwriting information, represented by the sequence of MIW words, to a feature vector which models handwriting by estimating the distribution of local features like orientation and curvature. The idea behind this originates from the simplest form of chain code. Analytically, chain code describes an eight set of sequences of two pixels and codes the succession of different orientations on the image grid. When sequences of three successive pixels are examined, line, convex and concave curvature features are generated. Since we do not utilize the features' order of appearance, the corresponding features which can be defined uniquely, beginning from a central pixel to another one, inside a chess-board distance equal to 2 are twenty-two (22). The enforcement of the symmetry condition limits the number of independent convex and concave features to 11. This subset is enriched with the use of four line-features describing the fundamental line segments of slope 0, 45, 90, 135. This 15-dimensional feature space defines the new embedding space. Furthermore we have partitioned the MIW image to a 2×2 sub-window grid, and the respective outputs have been fused in feature level by simple appending.

Following the above idea, we explore an additional feature set by measuring the pixels paths which are obeying the following statement. Find the four pixel connected paths, while restraining the chess-board distance among the first and the fourth pixel equal to three and co instantaneously restraining the chess-board distance among the first and the third pixel equal to two, by ignoring the prior path selection that has taken place in the inner two-step transition. This provides a feature with dimensionality of 28 since we do not partition the image. The final feature vector is generated by appending, in a feature fusion way, the aforementioned two and three step features. Its dimensionality equals to 88 (*four sub-images x 15 features + one image x 28 features*) and it is depicted graphically in Fig. 3. Algorithmically, a rectangular grid of 4 × 7 dimension scans every input of MIW words sequence. This mask aligns each aforementioned pixel with the {5, 3} coordinate, thus enabling 15 potential 2-step paths and 28 3-step paths from the central pixel according to the previous discussion. Then, the paths which are included in the feature set are marked and a counter updates the corresponding features found. Finally, the feature components are normalized by their total sum in order to provide a probabilistic expression.

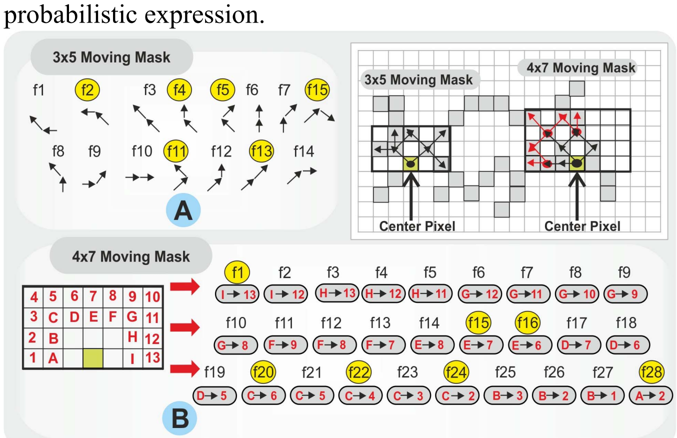

Fig. 3. Feature extraction methodology. Example with activated feature components (represented in yellow circles). a) Basic feature generating mask within chessboard distance of two. b) The feature mask within chessboard distance of three, irrespective of the inner, two-step path.

III. Classification Protocol

As described in section II, the input to the classification system are the training and testing feature vectors denoted hereafter as $\{v_{Tw}, v_{TSw}\}$. The training set v_{Tw} is composed of the genuine and forgery vectors $\{G_{TW}, F_{TW}\}$ of each writer W_i, $i = 1,2,...,40$. The G_{TW} vectors are modeling the genuine class population by means of their average value $\bar{\mu}_{v_{GTW}}$ and standard deviation $\hat{\sigma}_{v_{GTW}}$. Next, the similarity scores of the genuine training vectors are evaluated by using the weighted distance as eq. (1) provides [12] and their pdf $S\left(v_{G_{TW}} \mid W_i\right)$ is stored. A similar procedure, described by eq. (2), has been applied in order to derive the distribution of the similarity scores $S\left(v_{F_{TW}} \mid W_i\right)$ for the case of the false train samples $\{F_W\}$.

$$S\left(v_{G_{TW}} \mid W_i\right) = \left(\sum_{j=1}^{88} \hat{\sigma}(j)_{v_{GTW}}^{-2} \left(G(j)_{TW} - \bar{\mu}(j)_{v_{GTW}}\right)^2\right)^{-0.5} \quad (1)$$

$$S\left(v_{F_{TW}} \mid W_i\right) = \left(\sum_{j=1}^{88} \hat{\sigma}(j)_{v_{FTW}}^{-2} \left(F(j)_{TW} - \bar{\mu}(j)_{v_{FTW}}\right)^2\right)^{-0.5} \quad (2)$$

Following the first stage, a two-class support vector machine is employed in order to provide a mapping of the training similarity scores to another distance space, induced by the SVM. Accordingly, inputs to the second stage are the genuine and impostor distribution scores $S\left(v_{G_{TW}} \mid W_i\right)$, $S\left(v_{F_{TW}} \mid W_i\right)$. The output of the SVM is a continuous-valued distance of the optimal separating hyper-plane from the unknown test input sample vector [24]. The mapping function has been represented by a Gaussian Radial Base kernel function after a number of trials.

The testing phase uses the remaining samples of the genuine and forgery sets $\{v_{TSw}\} = \{G_{TSW}, F_{TSW}\}$. Thus, for each writer, the similarity scores, evaluated from the samples of the testing set, are presented as an input to the second stage SVM mapping function. A negative value from the SVM output indicates that the unknown feature vector is below the optimal separating hyper plane and near the hyper-plane which corresponds to the genuine class. On the other, a positive value denotes that the unknown input vector tends to fall towards the impostor hyper-plane class [15]. Finally, the continuous SVM output models both the overall distribution of the genuine writers along with the impostor ones. The selection of the training samples for the genuine class is accomplished using random samples with the hold-out validation method.

Evaluation of the verification efficiency of the system is accomplished with the use of a global threshold on the overall SVM output distribution. This is achieved by providing the system's False Acceptance Rate (FAR: samples not belonging to genuine writers, yet assigned to them) and the False Rejection Rate (FRR: samples belonging to genuine writers, yet not classified) functions. With these two rates, the receiver operator characteristics (ROC) are drawn by means of their FAR / FRR plot. Then, classification performance is measured with the utilization of the system Equal Error Rate (EER: the point which FAR equals FRR).

IV. Results

A. Benchmarking With Relative Feature Algorithms

We have benchmarked the proposed methodology against three other feature extraction methods for signature verification and writer identification, which can be found in the literature. The first is a signature verification texture based approach, which is provided by Vargas, Ferrer, Travieso and Alonso [16]. Secondly, we are examining the performance of a shape descriptor proposed by Aguilar, Hermira, Marquez and Garcia, which is based on the use of predetermined shape masks [17]. In all cases, the pre-processing as well as the feature extraction steps have been realized according to the description described by the authors. The third method uses the *f1* contour direction pdf features and the *f2* contour hinge features which are a part of the work proposed by Bulaku and Schomaker [18]. It is of great interest that the *f2* feature is one of the most powerful descriptors for modelling the handwriting. It must be noted that, an appropriate pre-processing step has been carried out in order to provide the contours of the handwritten images.

B. *Verification Results*

According to the material exposed in section III, representation of the genuine class has been realized with various schemes by utilizing 5, 10, 15, 20, 25, *30* samples for the $\{G_{TW}\}$ training and 115, 110, 105, 100, 95 and *90* samples for the $\{G_{TSW}\}$ testing. On the other, the $\{F_{TW}\}$ training set for the forgery class has been formed using one sample of all the remaining writers which results to a number of 39 samples. The $\{F_{TSW}\}$ samples are formed by employing the remaining $119\left(samples/writer\right)\times 39\ writers$, resulting to a total number of 4641. The ROC curves, which are drawn as a function of the number of words and presented to figs, 4-8, illustrate the classification efficiency of our method against to those mentioned to the previous section. These curves have been evaluated for the last training scheme, i.e 30 and 90 training samples for $\{G_{TW}\}$ and $\{G_{TW}\}$ population. Similar results regarding the evaluation taxonomy have been obtained.

Commenting on the results, it can be easily inferred that our method provides a challenging, first hand proof of concept of its enhanced writer verification capabilities. Another interesting issue is that the verification efficiency is enhanced when the number of the inserted words to the feature stage increases, which is intuitively correct. An Additional comment is that the English sentence provides a boosted EER when compared to the Greek sentence, even though Greek is our native language. This might be due to the fact that the text used in the English sentence incorporates lengthier words when compared to the Greek one. Another standpoint for the enhanced Latin EER measure could be that when Greeks or individuals which are not having English as their native language are forced to write in Latin, their response provides less spontaneous handwritten samples. This may have introduced less writer specificity in the data which in its turn provides higher verification rates. Although the results are quite encouraging however; they must be further tested in larger databases and under a number of different feature and classifications schemes. The best EER rates corresponding to figures 4-8 are presented in tabular form in table 1.

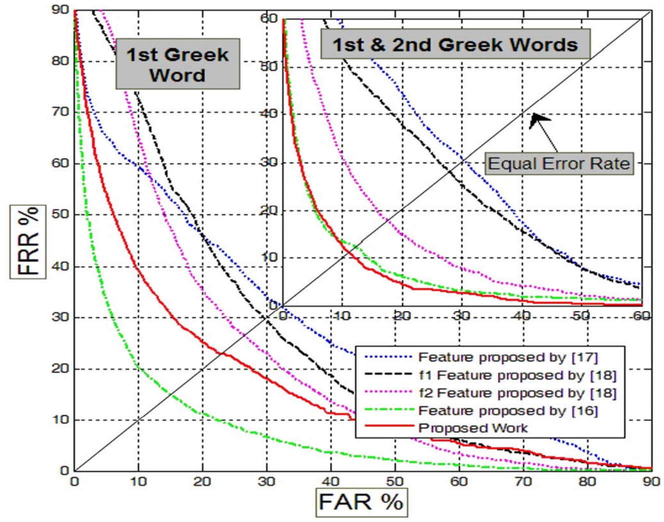

Fig. 4. ROC curves and EER of the proposed and the competitive methods. The lower left part presents the results from one Greek word while the upper right uses a sequence of the first and second words.

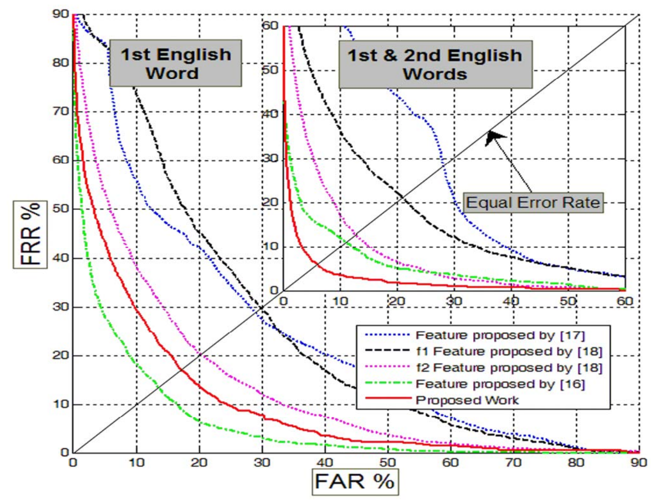

Fig. 5. ROC curves and EER of the proposed and the competitive methods. The lower left part presents the results from one English word while the upper right uses a sequence of the first and second Enlish words.

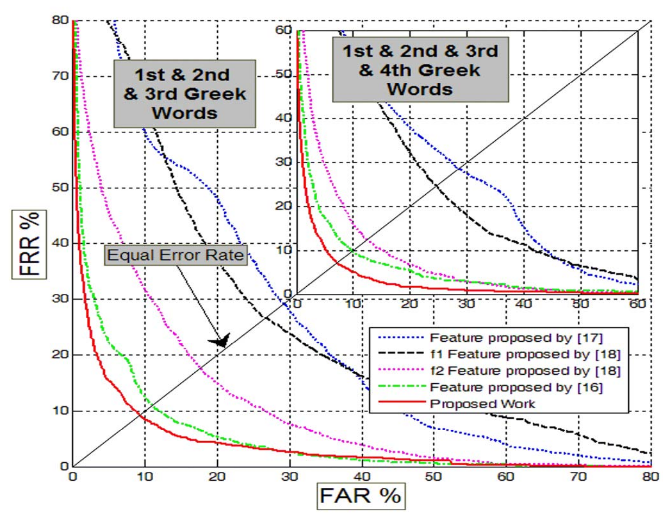

Fig. 6. ROC curves and EER of the proposed and the competitive methods. The lower left part presents the results by employing a sequence of the first three words of the Greek sentence while the upper right uses a sequence of the first four Greek words.

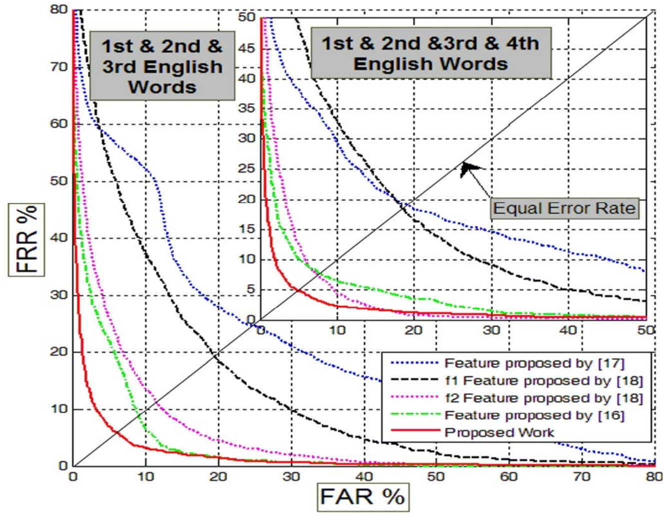

Fig. 7. ROC curves and EER of the proposed and the competitive methods. The lower left part presents the results by employing a sequence of the first three words of the English sentence while the upper right uses a sequence of the first four English words.

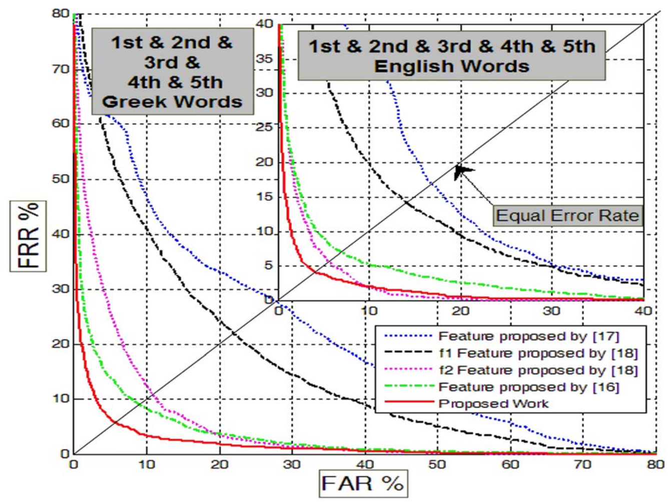

Fig. 8. ROC curves and EER of the proposed and the competitive methods. The lower left part presents the results by employing a sequence of the five words of the Greek sentence while the upper right uses a sequence of the five words of the English sentence.

REFERENCES

[1] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 63-84, 2000.

[2] G. X. Tan, C. Viard-Gaudin, and A. C. Kot, "Automatic writer identification framework for online handwritten documents using character prototypes," *Pattern Recognition*, vol. 42, pp. 3313-3323, 2009.

[3] I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognition*, vol. 43, pp. 3853-3865, 2010.

[4] A. Bhardwaj, A. O. Thomas, Y. Fu, and V. Govindaraju, "Retrieving handwriting styles: A content based approach to handwritten document retrieval," in *Proc. International Conference on Handwriting Recognition*, Kolkata, India, 2010, pp. 265-270.

[5] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar, "Continuous verification using multimodal biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 687-700, 2007.

[6] U.-V. Marti and H. Bunke, "The IAM-database: An English sentence database for off-line handwriting recognition "*International Journal on Document Analysis and Recognition*, Vol. 5, pp. 39-46, 2002.

[7] M. Bulaku and L. Schomaker, "Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems", Technical Report, NICI 2000.

[8] S. N. Srihari, S.-H. Cha, H. Arora and S. Lee, "Individuality of handwriting", *Journal of Forensic Science*, Vol. 47, pp.1-17, 2002.

[9] R. K. Hanusiak, L. S. Oliveira, E. Justino and R. Sabourin, "Writer verification using texture-based features", *International Journal of Document Analysis and "Recognition*, [DOI:10.1007/s10032-011-0166-4] , 2011.

[10] E. N. Zois and V. Anastassopoulos, "Fusion of correlated decisions for writer verification," *Pattern Recognition*, vol. 34, pp. 47-61, 2001.

[11] E. N. Zois, K. Tselios, E. Siores, A. Nassiopoulos, and G. Economou, "Off-Line Signature Verification Using Two Step Transitional Features," in *Proc 12th IAPR Conference on Machine Vision Applications*, Nara, Japan, 2011.

[12] R. O. Duda and P. E. Hart, *Pattern classification*. New York: John Wiley and Sons, 2001.

[13] http://biosecure.it-sudparis.eu/AB/

[14] N. Otsu, "A threshold selection method from gray-level histogram", *IEEE Transactions on System, Man and Cybernetics*, Vol. 8, pp.62-66, 1978.

[15] Lutz Hamel: "*Kernel Knowledge discovery with support vector machines*", Wiley, New Jersey, 2009.

[16] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso, "Off-line signature verification based on grey level information using texture features", *Pattern Recognition*, Vol. 44, pp. 375-385, 2011.

[17] J. F. Aguilar, N. A. Hermira, G. M. Marquez and J. O. Garcia, "An off-line signature verification system based of local and global information", *LCNS 3087*, pp.295-306, 2004.

[18] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, pp. 701-717, 2007.

TABLE I
CLASSIFICATION EFFICIENCY (%) BASED ON THE EQUAL ERROR RATE DERIVED FROM FIGS. 4-8

Feature Extraction Method	Sequences of Words (1^{st} / {1^{st} & 2^{nd} } / {1^{st}& 2^{nd}&3^{rd}} / {1^{st}& 2^{nd}&3^{rd}&4^{th}} / {all}	
	English Sentence	Greek Sentence
Proposed work	15.53 / 6.05 / 5.92 / 4.90 / 4.08	22.78 / 11.13 / 9.21 / 7.14 / 5.71
Feature proposed by [16]	13.54 / 11.10 / 9.08 / 7.69 / 6.92	15.04 / 12.29 / 10.99 / 9.76 / 8.96
f1 Feature proposed by [18]	29.81 / 21.06 / 19.46 / 18.41 / 14.12	29.78 / 28.08 / 26.49 / 23.85 / 21.98
f2 Feature proposed by [18]	20.22 / 12.72 / 11.36 / 7.48 / 5.58	26.55 / 17.72 / 17.57 / 12.41 / 10.82
Feature proposed by [17]	28.95 / 28.19 / 24.64 / 19.07 / 16.90	32.30 / 30.44 / 29.18 / 28.47 / 27.63