International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

A Writer Identification System of Greek Historical Documents using MATLAB

M. K. Konstantakis[1], E. J. Yannakoudakis[2]

[1]School of Social Sciences and Psychology, Department of Sociology, Panteion University, L.Syggrou 136, GR-17671 Athens, Greece

[2]Department of Informatics, Athens University of Economics & Business, Patission 76, GR-10434 Athens, Greece

***Abstract*--In this paper we present a system for writer identification from historical lines of text, where features are extracted and used to recognize individuals. The main goal is to analyze documents of different writing styles in order to identify the writers. We consider a complete 2D probability distribution that takes into account all possible combinations of angle pairs, outperforming original code. The images included are taken from the Greek digital library Hellinomnimon available at http://www.lib.uoa.gr/hellinomnimon/. The experiments are based on 7 pages of text written by 43 different writers.**

***Keywords*-- Writer identification, Handwriting analysis, Feature extraction, Edge-hinge distribution, Historical documents.**

I. INTRODUCTION

The growth of artificial intelligence and pattern recognition fields owes greatly to one of the highly challenged problem of handwriting identification. Identifying the handwriting of a writer is highly essential today due to the immense growth in technology and its applications in wide areas. The application of writer identification is in wide areas, such as, digital rights management in the financial sphere, to solve the expert problems in criminology by forensic expert decision-making systems, where a narrowed-down list of identified writers provided by the writer identification system. By combining with the writer verification as an authentication system this can be used to monitor and regulate the access to certain confidential sites or data where large amounts of documents, forms, notes and meeting minutes are constantly being processed and managed, knowing the identity of the writer would provide an additional value. It can also be used for historical document analysis [1], handwriting recognition system enhancement [2] and hand held and mobile devices [3].

To a certain extent its recent development and performance consider as a strong physiologic modalities of identification, such as DNA and fingerprints [4]. It is evident that the importance of writer identification has become more significant in these days. Obviously, the number of researchers involved in this challenging problem is going high as a result of these opportunities.

There are numerous languages throughout the world. Each language poses a different threat to the writer identification problem depending on the characteristics of the language. So it is very clear that the identification problem varies across multiple languages. The handwriting-based writer identification is an active research arena.

As it is one of the most difficult problems encountered in the field of computer vision and pattern recognition, the handwriting-based writer identification problem faces with a number of sub problems like :

- designing algorithms to identify handwritings of different individuals,
- identifying relevant features of the handwriting,
- basic methods for representing the features,
- identifying complex features from the basic features developed and,
- evaluating the performance of automatic methods.

The main goal of this paper is to analyze documents of different writing styles in order to identify the writers. In the end of the experimental procedure we insert to the database model some texts of unknown writers in order to recognize the writer and confirm the functionality of the system.

The rest of the paper is organized as follows. Section 2 describes the theoretical part of the historical documents. Section 3 describes the research about the writer identification procedure and the previous works in that field. In the following, the proposed method and statistical approach (feature extraction stage and classification stage) is detailed in section 4, while in section 5 the experimental results aredemonstrated. Finally, conclusions are drawn in section 6 and references in section 7.

II. HISTORICAL DOCUMENTS

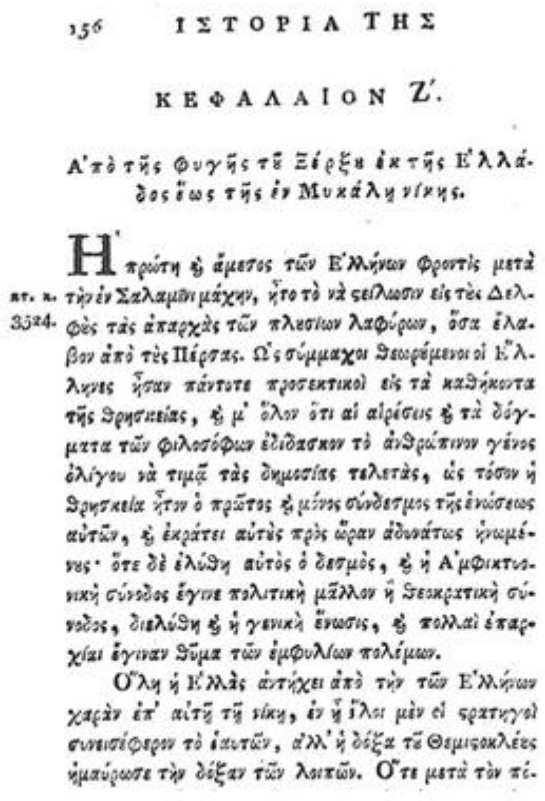
Historical document analysis is an emerging research topic that has gained increasing attention during the last decade [5]. Problems such as word spotting [8, 17], document layout analysis [10], and handwriting recogni¬tion [7, 9] have been investigated by the research commu¬nity.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

Especially the latter task, handwriting recognition, is challenging for a number of reasons, including training sets of small size, unusual writing styles, crossed out or overwritten words, and other artifacts.

Previous research on the recognition of handwriting in historical documents has been described in [9], where a hidden Markov model recognizer for holistic handwritten words has been applied to manuscripts of George Washington, and in [7] where HMMs as well as conditional ran-dom field models have been used for handwriting recog¬nition on the same manuscript. In [6] the attention has been on speeding up a the recognition task for indexing historical documents, and in [15] it has been focused on character recognition in historical Greek documents.

The system described in this paper is being developed for the recognition of historic manuscripts from Greek authors in the context of research in literature, physics, mathematics and geography (see Figure 1 for example image).


156 ΙΣΤΟΡΙΑ ΤΗΣ

ΚΕΦΑΛΑΙΟΝ Ζ'.

Α'πὸ τῆς φυγῆς τῦ Ξέρξυ ἐκ τῆς Ἑλλάδος ἕως τῆς ἐν Μυκάλῃ νίκης.

Η πρώτη ϕ ἄμεσος τῶν Ἑλλήνων ϕροντὶς μετὰ
πτ. π. τὴν ἐν Σαλαμῖνι μάχην, ἦτο τὸ νὰ ϕείλωσιν εἰς τὲς Δελ-
3524. ϕὸς τὰς ἀπαρχὰς τῶν πλουσίων λαϕύρων, ὅσα ἔλαβον ἀπὸ τὲς Πέρσας. Ὡς σύμμαχοι θεωρέμενοι οἱ Ἕλληνες ἦσαν πάντοτε προσεκτικοὶ εἰς τὰ καθήκοντα τῆς θρησκείας, ϕ μ' ὅλον ὅτι αἱ αἱρέσεις ϕ τὰ δόγματα τῶν ϕιλοσόϕων ἐδίδασκον τὸ ἀνθρώπινον γένος ὀλίγου νὰ τιμᾷ τὰς δημοσίας τελετὰς, ὡς τόσον ἡ θρησκεία ἦτον ὁ πρῶτος ϕ μόνος σύνδεσμος τῆς ἑνώσεως αὐτῶν, ϕ ἐκράτει αὐτὲς πρὸς ὥραν ἀδυνάτως ἡνωμένους· ὅτε δὲ ἐλύθη αὐτὸς ὁ δεσμὸς, ϕ ἡ Ἀμϕικτυονικὴ σύνοδος ἔγινε πολιτικὴ μᾶλλον ἤ θεοκρατικὴ σύνοδος, διελύθη ϕ ἡ γενικὴ ἕνωσις, ϕ πολλαὶ ἐπαρχίαι ἔγιναν θῦμα τῶν ἐμϕυλίων πολέμων.

Ὅλη ἡ Ἑλλὰς αὐτήχει ἀπὸ τὴν τῶν Ἑλλήνων χαρὰν ἐπ' αὐτῇ τῇ νίκῃ, ἐν ᾗ ὅλοι μὲν οἱ ϕρατηγοὶ συνεισέϕερον τὸ ἑαυτῶν, ἀλλ' ἡ δόξα τῦ Θεμιστοκλέες ἡμαύρωσε τὴν δόξαν τῶν λοιπῶν. Ὅτε μετὰ τὸν πέ-

Fig.1 Example of Greek historical document

III. Writer Identification

Research in writer identification has received renewed interest in the recent years. A wide variety of features, local or global and structural or statistical, have been proposed that serve to distinguish the writing of an individual from another.

Consequently the approaches proposed in the last several years renewed the interests in this scientific community for the research topic. Figure 2 presents the standard framework of writer identification [11].

The necessary features from the handwritten documents are extracted as the first step. Later the features extracted are used to classify to which writer the document belongs using similarity score method. The document is classified as belonging to a writer with high similarity score:

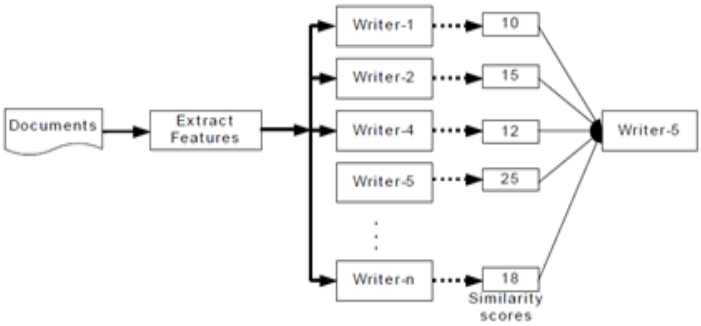


Fig.2 Example of historical Greek handwritten document

Based on the input method of writing, automated writer identification has classifieds into on-line and off-line. The on¬line writer identification task is considered to be less difficult than the offline one as it contains more information about the writing style of a person, such as speed, angle or pressure, which is not available in the off-line one [12, 13]. Based on the different features associated with the writing, such as character, word, line, paragraph and the document, this has classified.

Text-dependent & text-independent are the other classification of automated writer identification. Dependent on the text content, text-dependent methods only matches the same characters and requires the writer to write the same text consequently. The text-independent methods are able to identify writers independent of the text content and it does not require comparison of same characters. Thus it is very similar to signature verification techniques and uses the comparison between individual characters or words of known semantic (ASCII) content. This method considers as the global style of hand writing text as the metric for comparison, and also got better identification results. As it requires the same writing content this method is not apt for many practical situations. Even though it got a wider applicability, text-independent methods do not obtain the same high accuracy as text-dependent methods do.

Previous Works

Features used for the writer identification task are mainly global features which are based on statistical measurements, extracted from the whole block of text to be identified.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

These features can be broadly classified in two families:

- features extracting from texture: the document image is simply seen in this case as an image and not as a handwriting. For example, application of Gabor filters and co-occurrence matrices were considered in [14].
- structural features: in this case the extracted features attempt to describe some structural properties of the handwriting. One can quote for example the average height, the average width, the average slope and the average legibility of characters [16].

Note that it is also possible to combine the two families of features [18]. The nature of these statistical features, extracted from a block of text, has allowed to reach interesting performance, which are however always difficult to compare due to a lack of common references.

One can also categorize the previous works according to the number of writers and the nature of training samples used by the system (see table 1).

On the one hand the system is required to deal with as much writers as possible while on the other hand, training samples of each handwriting may represent several lines of text or on the contrary a few words.

Table 1.
Comparison of performance and test conditions for writer identification in most recent studies.

System	# Writers	Sample size / Features	Classification Methodology dependency	Performance	Language
Said 2000 [19]	40	25 samples per writer (Few lines of handwritten text) – Gabor filtering, gray-scale matrices	Nearest centroid classification using weighted Euclidean distance	96%	English
Marti 2001 [21], Hertel and Bunke [22]	20	5 samples of the same text –Height of the three main writing zones	A k-nearest neighbor and a feed forward neural network classifier	90%	English
Bulacu 2003 [23, 24, 25, 26]	650	Edge based directional PDF's as features	A k-nearest neighbor and a feed forward neural network	92%	English
Pitak 2004 [27]	81	Velocities of the barycenter of the pen movements	Fourier transformation approach	98.5%	English
Schlapbach 2004 [20]	100	X-Y coordinates	Hidden Markov Model	96%	English
Bangy 2007 [32]	242	Hierarchical Structure in Shape Primitives + Fusion Dynamic and Static	Nearest neighbor classifier	Chinese 90% English 93%	English, Chinese
Chan 2008 [28, 36]	82	X-Y coordinates, direction, curvature of x-coordinates and the status of pen up or pen down	Discrete character prototype distribution approach (Euclidean distance)	95%	French
Neils 2008 [29]	43	Allograph prototype matching approach using the dynamic time warping distance function	Af-iwf (allograph frequency – inverse writer frequency) measure	60%	English
Helli 2008 [30, 31]	100	Point-based , stroke-based	Tey proposed an LCS (Longest Common Subsequence) based classifier	95%	Persian
YuChen Yan 2009 [33]	200	Spectral feature based on Fast Fourier Transformation	Euclidean and WED classifiers	98% -top 10 64%-top1	Chinese

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

IV. The Proposed Method – Statistical Approach

Research in automatic writer identification has mainly focused on the statistical approach. This has led to the specification and extraction of statistical features such as run-length distributions, slant distribution, entropy, and edge-hinge distribution [23]. The edge-hinge distribution feature outperforms all other statistical features.

We have improved and used Luigi Rosa and Laurens van der Maaten's excellent package for writer identification [34, 37]. Luigi Rosa consider a complete 2D probability distribution that takes into account all possible combinations of angles pairs, outperforming original code.

The statistical approach consist of two stages: a feature-extraction stage and a classification stage. In the feature-extraction stage, features are extracted from handwriting and are stored in feature vectors. In the classification stage, the feature vectors are mapped onto classes representing the writers.

IV.I Feature Extraction Stage

Edge Distribution

Feature extraction starts with conventional edge detection (we used Sobel, followed by thresholding) that generates a binary image in which only the edge pixels are "on".[23]

We then consider each edge pixel in the middle of a square neighborhood and we check in all directions emerging from the central pixel and ending on the periphery of the neighborhood for the presence of an entire edge fragment:

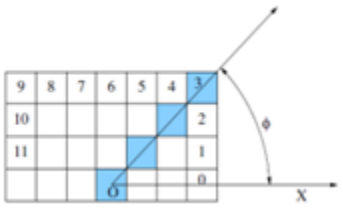


Fig.3 Extraction of edge direction distribution

All the verified instances are counted into a histogram that is finally normalized to a probability distribution which gives the probability of finding in the image an edge fragment oriented at the angle measured from the horizontal.

In order to avoid redundancy, the algorithm only checks the upper two quadrants in the neighborhood because, without on-line information, we do not know which way the writer "traveled" along the found oriented edge fragment.

Edge-hinge Distribution

A method of feature extraction that is similar to the one previously described, but it has added complexity. Edge-hinge distribution is mainly focused in the neighborhood, not one, but two edges fragments emerging from the central pixel and, subsequently, compute the joint probability distribution of the orientations of the two fragments.

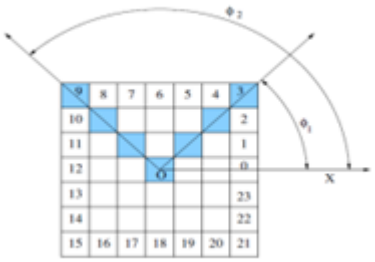


Fig.4 Extraction of edge-hinge distribution

Edge-hinge distribution is a feature that characterizes the changes in direction of a writing stroke in handwritten text. The edge-hinge distribution is extracted by means of a window that is slid over an edge-detected binary handwriting image. Whenever the central pixel of the window is on, the two edge fragments (i.e. connected sequences of pixels) emerging from this central pixel are considered. Their directions are measured and stored as pairs. A joint probability distribution is obtained from a large sample of such pairs. (p(x,y) = P(X = x and Y = y)).

Run-Length Distribution

Run lengths, first proposed for writer identification by Arazi [35], are determined on the binarized image taking into con-sideration either the black pixels corresponding to the ink trace or, more beneficially, the white pixels corresponding to the back¬ground. Whereas the statistical properties of the black runs mainly pertain to the ink width and some limited trace shape characteris¬tics, the properties of the white runs are indicative of character placement statistics. There are two basic scanning methods: hori¬zontal along the rows of the image and vertical along the columns of the image. Similarly to the edge-based directional features pre¬sented above, the histogram of run lengths is normalized and inter¬preted as a probability distribution.

IV.II Classification Stage

Classification is performed with pdist function, which calculates the distance between vectors using various distance types (Manhattan distance, Euclidian distance or Chi-square distance).

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

D= pdist(X) computes the Euclidean distance (set as default) between pairs of objects in m-by-n data matrix X. Rows of X correspond to observations, and columns correspond to variables. D is a row vector of length m(m–1)/2, corresponding to pairs of observations in X. The distances are arranged in the order (2,1), (3,1), ..., (m,1), (3,2), ..., (m,2), ..., (m,m–1)). D is commonly used as a dissimilarity matrix in clustering or multidimensional scaling.

To save space and computation time, D is formatted as a vector. However, we can convert this vector into a square matrix using the squareform function so that element i, j in the matrix, where i < j, corresponds to the distance between objects i and j in the original data set.

Given an m-by-n data matrix X, which is treated as m (1-by-n) row vectors x1, x2, ..., xm, the various distances between the vector xs and xt are defined as follows:

- *Euclidean distance*

$$d_{st}^2 = (x_s - x_t)(x_s - x_t)'$$

Notice that the Euclidean distance is a special case of the Minkowski metric, where p = 2.

- *Standardized Euclidean distance*

$$d_{st}^2 = (x_s - x_t)V^{-1}(x_s - x_t)'$$

Where V is the n-by-n diagonal matrix whose jth diagonal element is S(j)2, where S is the vector of standard deviations.

- *Minkowski metric*

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} | x_{sj} - x_{tj} |^p}$$

Notice that for the special case of p = 1, the Minkowski metric gives the city block metric, for the special case of p= 2, the Minkowski metric gives the Euclidean distance, and for the special case of p = ∞, the Minkowski metric gives the Chebychev distance.

V. Experiments And Results

We consider a complete 2D probability distribution that takes into account all possible combinations of angles pairs, outperforming original code.

We used a Matlab implementation and some of the main functions of the platform are:

- Select image: read the input image
- Add selected image to database: the input image is added to database and will be used for training
- Database Info: show informations about the images present in database.
- Writer Recognition: writer identification. The selected input image is processed.
- Delete Database: remove database from the current directory.
- Remove a specified image: we can remove a specified image from the database.
- Exit: quit program.

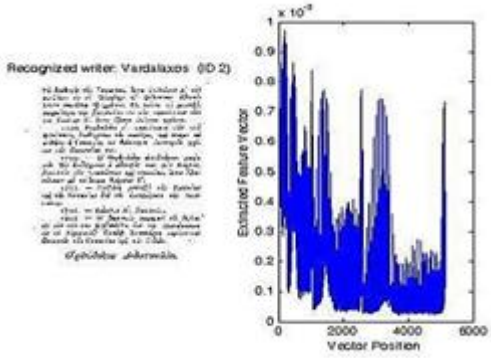
In the experiments, 7 pages of text written by 43 different writers are used, Greek authors in the context of research in literature, physics, mathematics and geography.

From the Matlab menu of our program we inserted the images (7x43=301 images) to the database and categorized them by ID and author name. (43 ID's of 43 different writers).

For testing the program we used test images of known writers (in example Vardalaxos_test and Velestinlis_test) and test images of unknown writers taken from the Greek digital library Hellinomnimon (at the bottom of the Hellinomnimon page):

- Διδασκαλία Εντελής συστηματική απάσης της Εμπορικής Επιστήμης (Ιάσιο, 1817) – ID: agnostos1.
- Ζυγόμετρον ήτοι τα ζύγια και μέτρα διαφόρων εμπορίων (Βενετία, 1803) – ID: agnostos2.
- Πινακίδιον Αριθμητικόν (Βιέννη, 1791) – ID: agnostos3.

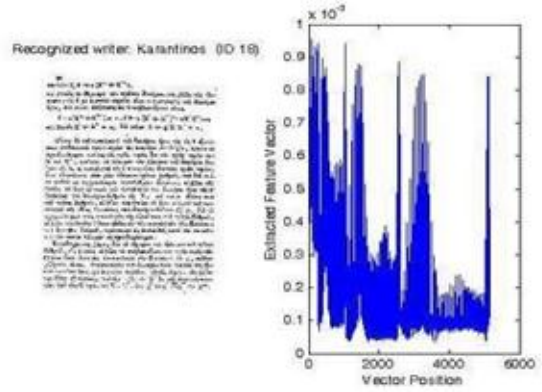
By using Vardalaxos_test we received the above results after the feature extraction and the writer recognition process:



The program analyzed that the Vardalaxos_test image matches exactly with the Vardalaxos as a writer.

By using agnostos1_2 (we used 4 images from agnostos1 writer) image we received the above results after the feature extraction and the writer recognition process:

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)



The program analyzed that the agnostos1_2 image matches exactly with the Karantinos (ID) as a writer.

In the table below we can see the authors, the Id's of each author, the context or research and the Identification name. We can see the authors and the context of research in the table below:

Author	ID	Context of research	Id name
Αλεξανδρίδης Δημήτριος (1784 - 1851;)	1	Ιστορία της Ελλάδος : Από της πρώτης καταβολής των Ελληνικών πραγμάτων άχρι της αλώσεως της Κωνσταντινούπολης ύπο των Οθωμανών	Alexandridis
Βαρδαλάχος Κωνσταντίνος (1775 - 1830)	2	Μαθήματα δια τους Παίδας. Τόμος Τρίτος (Οδησσός 1830)	Vardalaxos
Βάρκοσης Νικόλαος (; - περ. 1782)	3	Λογική Βαϋμαϊστέρου (Βιέννη, 1795)	Varkosis
Βενιαμίν Λέσβιος (1759/1762 - 1824)	4	Γεωμετρίας Ευκλείδου Στοιχεία [Τόμος Β'] (Βιέννη, 1820) - Στοιχεία Αριθμητικής [Τόμος Α'] (Βιέννη, 1818)	Veniamin
Βλεμμίδης Νικηφόρος (1197-1272)	5	Επιτομή Λογικής (Λειψία, 1784)	Vlemmidis
Βούλγαρις Ευγένιος (1716 - 1806)	6	Α. Τακουετίου, Στοιχεία Γεωμετρίας (Βιέννη, 1805) - Στοιχεία της Μεταφυσικής,(Βενετία, 1805)	Voulgaris
Γλυτζούνης Μανουήλ (1530/1540 - 1596)	7	Βιβλίον πρόχειρον τοις πάσι περιέχον την τε πρακτικήν αριθμητικήν... (Βενετία, 1568)	Glytzounis
Γοβδελάς Δημήτριος (1780 - 1831)	8	Στοιχεία Αλγέβρας (Χάλλη, 1806) - Στοιχεία Αριθμητικής (Ιάσιο, 1818)	GovdelasD
Γοβδελάς Μιχαήλ (τέλη 18ου αι. - μέσα 19ου αι.)	9	Αι της Φύσεως Εποχαί (Βενετία, 1816)	GovdelasM
Δαμασκηνός ο Στουδίτης (αρχές 16ου αι.- 1577)	10	Μερική διάγνωσις εκ των παλαιών φιλοσόφων περί Φύσεως (Βενετία, 1682)	Damaskinos
Δημητρίου Θωμάς (μέσα 18ου αι. - αρχές 19ου αι.)	11	Χειραγωγός Έμπειρος της Πραγματείας (Βιέννη, 1809)	Dimitriou
Ζαφαράνας Σπυρίδων (περ. 1770 - περ. 1830)	12	Του ναυκλήρου Εφημερινή εις το πέλαγος πράξις (Κωνσταντινούπολη, 1803)	Zafaranas
Θεοτόκης Νικηφόρος (1731 - 1800)	13	Στοιχείων Μαθηματικών εκ παλαιών και νεωτέρων συνερανισθέντων, (Μόσχα, 1798)	Theotokis
Ιωαννίδης Παναγιώτης (τέλη 18ου αι. - μέσα 19ου αι.)	14	Βεζούτ Λέων, Η σύντομος έκθεσις του πλανητικού συστήματος (Παρίσι, 1825)	Ioannidis
Κάβρας Ζήσης (περ. 1760 - 1844;)	15	Στοιχεία της Αριθμητικής και Αλγέβρης (Ιένα, 1800)	Kavras
Κανέλος Στέφανος (1792 - 1823)	16	Φυσική δημώδης εις παύσιν της δεισιδαιμονίας (Βενετία, 1810)	Kanelos
Καρακιουλάφης Αναστάσιος (περ. 1780 - περ. 1840)	17	Αριστοτέλους φυσιογνωμονικά (Κωνσταντινούπολη, 1819)	Karakioulafis
Καραντηνός Ιωάννης (1784 - 1834)	18	Διασάφησις εις την επίτομον θεωρίαν... του κυρίου Λακρουά (Κέρκυρα, 1826) - Σειρά στοιχειώδους Μαθηματικής, τόμος Α': Στοιχεία Αριθμητικής (Βιέννη, 1828)	Karantinos
Κοδρικάς Παναγιώτης (1750/1755 - 1827)	19	Ομιλίαι περί Πληθύος Κόσμων (Βιέννη, 1794)	Kodrikas

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 10, October 2014)

Κοκκινάκης Κωνσταντίνος (1781 - 1831)	20	Ιστορίας του Εμπορίου επιτομή (Βιέννη, 1809)	Kokkinakis
Κομμητάς Στέφανος (περ. 1770 - 1830;)	21	Παιδαγωγικά Μαθήματα Γεωγραφίας (Πέστη, 1828)	Kommitas
Κορυδαλεύς Θεόφιλος (1563/1574 - 1646)	22	Είσοδος Φυσικής Ακροάσεως κατ' Αριστοτέλην (Βενετία, 1779) - Εις άπασαν την Λογικήν του Αριστοτέλους (Βενετία, 1729)	Koridalevs
Κούμας Κωνσταντίνος (1777 - 1836)	23	Σύνοψις Φυσικής (Βιέννη, 1812)	Koumas
Κωνσταντίνου Γεώργιος (1710/1730 - περ. 1790)	24	Παγκόσμιος Ιστορία της Οικουμένης (Βενετία, 1759)	Konstantinou
Κώπας Γεράσιμος (1778 - 1832)	25	Πίναξ της κλίσεως του Ηλίου (Κωνσταντινούπολη, 1816)	Kopas
Λασσάνης Γεώργιος (1793 - 1870)	26	Η διά φωνής Αριθμητική (Μόσχα, 1820)	Lassanis
Μακραίος Σέργιος (1734/1740 - 1819)	27	Επιτομή Φυσικής ακροάσεως (Βενετία, 1816)	Makraios
Μαρμαροτούρης Ιωάννης (μέσα 18ου αι. - αρχές 19ου αι.)	28	Ο ανταποκριτής Τεργεστίου (Τεργέστη, 1800)	Marmarotouris
Μελέτιος, Μητροπολίτης Αθηνών (Μιχαήλ) (1661 - 1714)	29	Επιτομή Γεωγραφίας (Βενετία, 1728)	Meletios
Μοισιόδακας Ιώσηπος (1725/1730 - 1800;)	30	Θεωρία της Γεωγραφίας (Βιέννη, 1781)	Misiodakas
Μπαλάνος Βασιλόπουλος (1694 - περ. 1760)	31	Έκθεσις ακριβεστάτη της Αριθμητικής (Βενετία, 1803)- Μέθοδος Γεωμετρικώς χωρούσα περί ευρέσεως των δύο μέσων... (Βενετία, 1756)	MpalanosB
Μπαλάνος Κοσμάς (1731 - 1807/1808)	32	Αντιπελάργησις (Βιέννη, 1816)	MpalanosK
Παλαιολόγος Γρηγόριος (τέλη 18ου αι. - πρώτο μισό 19ου αι.)	33	Ερμηνεία της καλλιέργειας του γεωμήλου (Παρίσι, 1828)	Palaiologos
Παμπλέκης Χριστόδουλος (1733 - 1793)	34	Περί Φιλοσόφου, Φιλοσοφίας... (Βιέννη, 1786)	Pamplekis
Σπανόπουλλος Παναγιώτης (μέσα 18ου αι. - αρχές 19ου αι.)	38	Αριθμητική Εμπορική (Τεργέστη, 1803)	Spanopoulos
Σπαρμιώτης Ιωνάς (περ. 1770 - περ. 1830)	39	Σύνοψις των κωνικών τομών (Βιέννη, 1802)	Sparmiotis
Σταγειρίτης Αθανάσιος (περ. 1780 - περ. 1840)	40	Επιτομή Αριθμητικής (Βιέννη, 1810) - Ηπειρωτικά, ήτοι Ιστορία και Γεωγραφία της Ηπείρου (Βιέννη, 1819)	Stagiritis
Φατζέας Γεώργιος (περ. 1722 - 1768)	41	Γραμματική Γεωγραφική (Βενετία, 1760)	Fatzeas
Φιλιππίδης Δανιήλ (1750/1755 - 1832)	42	Γεωγραφία νεωτερική (Βιέννη, 1791) - Η Λογική παρά του Κονδιλλιάκ (Βιέννη, 1801)	Filippidis
Χρησταρής Μιχαήλ (1773 - 1831)	43	Στοιχεία της Αριθμητικής και Αλγέβρας (Πάντοβα, 1804)	Xristaris

VI. CONCLUSIONS

The recognition of handwritten manuscripts has gained increasing attention in recent years. In this paper we have presented an effective method for writer identification for Greek historical documents using Matlab. Our particular focus of attention is on the Greek digital library Hellinomnimon available at http://www.lib.uoa.gr/hellinomnimon/.

The experiments are based on 7 pages of text written by 43 different writers.

We have improved and used Luigi Rosa and Laurens van der Maaten's excellent package for writer identification [34]. Luigi Rosa consider a complete 2D probability distribution that takes into account all possible combinations of angles pairs, outperforming original code.

The statistical approach consist of two stages: a feature-extraction stage and a classification stage.

The achieved identification rates are very promising:

- For the Vardalaxos_test image the nearest ID is number 2 and the distance from Database is equal to 0.0600 and

- For the agnostos1_2 image the nearest ID is number 18 and the distance from Database is equal to 0.1064.

It is well known fact that the more training data is available for a writer the higher it is expected recognition performance.

Our future research will address the exploration of handwritten historical documents and the upgrade of the system to include more writers.

REFERENCES

[1] Fornes, A., Llados, J., Sanchez, G., Bunke, H. "Writer Identification in Old Handwritten Music Scores". 8th IAPR Workshop on Document Analysis Systems, 347—353, 2008.

[2] Sas, J. "Handwriting Recognition Accuracy Improvement by Author Identification". L. Rutkowski et al. (eds.), ICAISC, LNAI 4029, 682--691.Springer, Heidelberg, 2006.

[3] Chaudhry, R., Pant, S. K. "Identification of authorship using lateral palm print - a new concept". Journal ForensicScience International, volume (141), 49-57, 2004.

[4] Schomaker, L. "Advances in Writer Identification and Verification". 9th International Conference on Document Analysis and Recognition(ICDAR), volume 2, 1268—1273, 2007.

[5] Antonacopoulos, A., Downton, A. "Special issue on the Analysis of Historical Documents". International Journal Document Analysis Recognition, 9(2):75–77, 2007.

[6] Govindaraju, V., Xue H. "Fast handwriting recognition for indexing historical documents". 1st International Workshop on Document Image Analysis for Libraries, pages 314–320, 2004.

[7] Feng, S., Manmatha R., McCallum, A. "Exploring the use of conditional random field models and HMMs for historical handwritten document recognition". 2nd International Conference on Document Image Analysis for Libraries, pages 30–37, 2006.

[8] Kornfield, E., Manmatha, R., Allan, J. "Text alignment with handwritten documents". 1st International Workshop on Document Image Analysis for Libraries, pages 195–205, 2004.

[9] Lavrenko, V., Rath, T., Manmatha, R. "Holistic word recognition for handwritten historical documents". International Workshop on Document Image Analysis for Libraries, pages 278–287, 2004.

[10] Feldbach, M., Tonnies, K. "Line detection and segmentation in historical church registers". 6th International Conference on Document Analysis and Recognition, pages 743–747, 2001.

[11] Gupta, S. "Automatic Person Identification and Verification using Online Handwriting". Master Thesis, International Institute of Information Technology Hyderabad, India, 2008.

[12] Schlapbach, A., Marcus, L. Bunke, H. "A writer identification system for on-line whiteboard data". Pattern Recognition Journal 41 23821–23897, 2008.

[13] Schomaker, L. "Advances in Writer identification and verification". 9th International Conference on Document Analysis and Recognition (ICDAR), 2007.

[14] Said, H., Tan, T., Baker, K. "Personal Identification Based on Handwriting". Pattern Recognition Journal, vol. 33; pp 149-160, 2000.

[15] Ntzios, K., Gatos, B., Pratikakis, I., Konidaris, T., Perantonis, S. "An old Greek handwritten OCR system based on an efficient segmentation-free approach". International Journal Document Analysis Recognition, 9(2):179–192, 2007.

[16] Marti, U., Messerli, R., Bunke, H. "Writer Identification Using Text Line Based Features". 6th International Conference on Document Analysis and Recognition (ICDAR), Seattle (USA), pp 101-105, 2001.

[17] Rath, T., Manmatha, R. "Word spotting for historical documents". International Journal Document Analysis Recognition, 9(2):139–152, 2007.

[18] Srihari, S., Cha, S., Arora,H., Lee, S. "Individuality of Handwriting : A Validity Study", 6th International Conference on Document Analysis and Recognition (ICDAR), Seattle (USA), pp 106-109, 2001.

[19] Said, H., Peake, G., Tan, T., Baker, K. "Writer Identification from Non-Uniformly Skewed Handwriting Images". 9th British Machine Vision Conference, pp. 478-487, 1998. 172, 2004

[20] Schlapbach, A., Bunke, H. "Using HMM based recognizers for writer identification and verification". International Workshop on Frontiers in Handwriting Recognition, IWFHR, Tokyo, pp. 167–

[21] Marti, U.V., Messerli, R., Bunke, H. "Writer Identification Using Text Line Based Features". 6th International Conference Document Analysis and Recognition (ICDAR), pp. 101-105, 2001.

[22] Hertel, C., Bunke, H. "A Set of Novel Features for Writer Identification". Fourth International Conference Audio and Video-Based Biometric Person Authentication, pp. 679-687, 2003.

[23] Bulacu, M., Schomaker, L., Vuurpijl, L. "Writer identification using edge-based directional features". 7th International Conference on Document Analysis and Recognition (ICDAR), 2003.

[24] Bulacu, M. Schomaker, L. "Writer Style from Oriented Edge Fragments". 10th International Conference Computer Analysis of Images and Patterns, pp. 460-469, 2003.

[25] Schomaker, L., Bulacu, M., Franke, K. "Automatic Writer Identification Using Fragmented Connected - Component Contours". 9th International Workshop Frontiers in Handwriting Recognition (IWFHR), pp. 185-190, 2004.

[26] Bulacu, M., Schomaker, L. "A Comparison of Clustering Methods for Writer Identification and Verification". 8th International Conference Document Analysis and Recognition, vol. II, pp. 1275-1279, 2005.

[27] Pitak, T., Matsuura, T. "On-line writer recognition for Thai based on velocity of barycenter of pen-point movement". International Conference on Image Processing, ICIP `04, vol. 2, pp. 889–892, 2004.

[28] Chan, S. K., Viard-Gaudin, C., Tay, Y. H. "Online writer identification using character prototypes distributions". inProceedings of SPIE - The International Society for Optical Engineering, 2008.

[29] Niels, R., Gootjen, F. Vuurpijl, L. "Writer Identification through Information Retrieval: The Allograph Weight Vector". International Conference on Frontiers in Handwriting Recognition, pp. 481-486, 2008.

[30] Helli, B., Moghaddam, M.E. "A text-independent Persian writer identification system using LCS based classifier". IEEE International Symposium on Signal Processing and Information Technology, ISSPIT, 2008.

[31] Helli, B., Moghaddam, M.E. "A writer identification method based on XGabor and LCS". IEICE Electronics Express 6, 2009.

[32] Li, B., Sun, Z., Tan, T.N. "Hierarchical Shape Primitive Features for Online Text-independent Writer Identification". Proc. of 2th ICB Conference, pages 201–210, 2007.

[33] Yan, Y., Chen, Q., Deng, W., Yuan, F. "Chinese Handwriting Identification Based on Stable Spectral Feature of Texture Images". International Journal of Intelligent Engineering and Systems, Vol.2, No.1, 2009.

[34] Laurens van der Maaten, Eric Postma "Improving Automatic Writer Identification". Proceedings of the 17th BNAIC Conference, 2005.

[35] Arazi, B. "Handwriting identification by means of run-length measurements". IEEE Trans. Syst. Journal, Man and Cybernetics, SMC-7(12):878-881, 1977.

[36] Tan, X., Viard-Gaudin, C., Kot, A. "Automatic writer identification framework for online handwritten documents using character prototypes". Pattern Recognition Journal 42 pages 3313–3323, 2009.

[37] Kam, M., Fielding, G., Conn, R. "Writer identification by professional document examiners". Journal of Forensic Sciences, 42:778-785, 1997.