ICFHR2012 Competition on Writer Identification Challenge 1: Latin/Greek Documents

G. Louloudis, B.Gatos and N. Stamatopoulos
Computational Intelligence Laboratory
Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos"
GR-15310 Athens, Greece
{louloud, bgat, nstam}@iit.demokritos.gr

Abstract

Writer identification is important for forensic analysis, helping experts to deliberate on the authenticity of documents. The general objective of the ICFHR 2012 Writer Identification Contest is to record recent advances in the field of writer identification using established evaluation performance measures. Challenge 1 of the contest deals specifically with Latin scripts. The benchmarking dataset of challenge 1 of the contest was created with the help of 100 writers that were asked to copy four parts of text in two languages (English and Greek). This paper describes the contest details for this challenge including the evaluation measures used as well as the performance of the seven submitted methods along with a short description of each method.

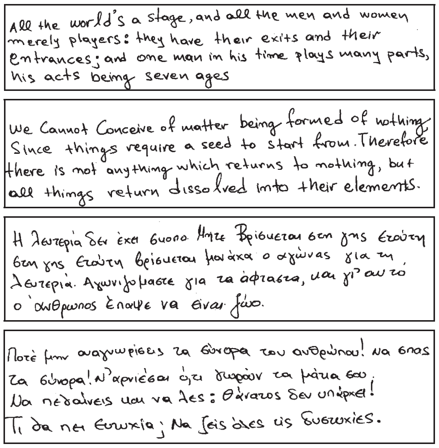
1. Introduction

Writer identification concerns the process of defining the identity of the writer of a document when a database of documents with known writer information is available. From the document image analysis scope, writer identification can be defined as the retrieval of handwritten samples of the same writer from a database using a handwritten sample as a graphical query. The large number of recent publications ([1]-[7]) as well as the organization of several competitions ([8]-[10]) prove that writer identification is a very active and promising area of research.

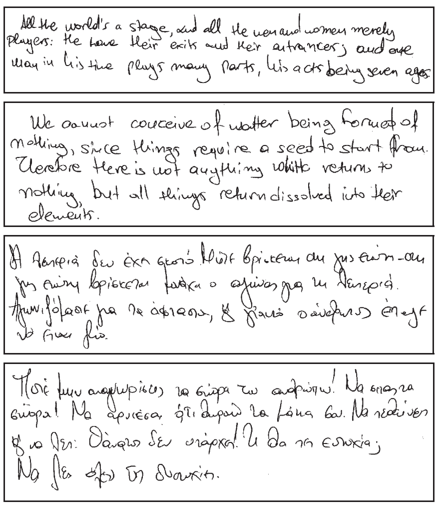
Following the successful organization of the ICDAR 2011 Writer Identification Contest [8], we organized the “ICFHR 2012 Writer Identification Contest, Challenge1: Latin Documents” providing a benchmarking dataset along with an objective and established evaluation methodology in order to record recent advances in the field of writer identification for Latin scripts. A new benchmarking dataset was created with the help of 100 writers that were asked to copy four parts of text in two languages (two in English and two in Greek) (see Fig.1) in order to test and compare recent algorithms for writer identification in realistic circumstances. These parts of text were the same for all users. Among all documents, only the Greek documents were written in the native language of the writer. We should also note that the number of text lines that were produced by the writers ranged between three and six.

The contest procedure was based on the following milestones. The authors of candidate methods registered their interest in the competition [11] and downloaded the experimental dataset (part of the ICDAR 2011 Writer Identification Contest dataset containing 104 cropped images written by 26 individual writers in English and Greek languages) [12]. At a next step, all registered participants were required to submit one executable in the form of a Win32 console application. This executable takes as input two binary images and produces a similarity score taking into account the corresponding writing style. After the evaluation of all candidate methods, the benchmarking dataset (400 images along with the corresponding writer id information) became publicly available [13].

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.219



(a)



(b)

Figure 1. Image samples from two different writers (a), (b) in English and Greek language.

This paper describes the contest details including the datasets, the evaluation criteria as well as the performance of the seven submitted methods along with a short description of each method.

2. Methods and participants

Four research groups submitted their methodologies to the contest. One of these research groups submitted three different methodologies while another one submitted two making the total number of participating methodologies equal to seven. A brief description of these methodologies is provided in this section.

TEBESSA-a method: Submitted by (a) Chawki Djeddi from the LAMIS Laboratory, Mathematics and Computer Science Department, University of Tebessa, Tebessa, Algeria, (b) Labiba Souici-Meslati from the LRI Laboratory, Computer Science Department, Badji Mokhtar University of Annaba, Annaba, Algeria and (c) Abdellatif Ennaji from the LITIS Laboratory, Rouen University, Saint Etienne du Rouvray, Rouen, France.

The submitted method is based on the edge-hinge features which estimate the joint distribution of edge angles in a writer's handwriting. They are constructed by performing an edge detection using a Sobel kernel on the input images, and subsequently, measuring the angles of both edge segments that emanate from each edge pixel. To compare two documents, the Manhattan Distance Metric is used.

TEBESSA-b method: Submitted by the same group as the previous method. It is based on multi-scale run length features [14] which are determined on the binary image taking into consideration both the black pixels corresponding to the ink trace and the white pixels corresponding to the background. The probability distribution of black and white run-lengths has been used. There are four scanning methods: horizontal, vertical, left-diagonal and right-diagonal. We calculate the runs lengths features using the grey level run length matrices and the histogram of run lengths is normalized and interpreted as a probability distribution. The method considers horizontal, vertical, left-diagonal and right-diagonal white run-lengths as well as horizontal, vertical, left-diagonal and right-diagonal black run-lengths extracted from the original image. To compare two documents, the Manhattan Distance Metric is used.

TEBESSA-c method: Submitted by the same group as the previous method. It is based on the combination of both types of features used by the previous two methods: multi-scale edge-hinge features and multi-scale run-length features [14]. Again for this method, the Manhattan Distance Metric is used to compare two documents.

QATAR-a method: Submitted by Abdelâali Hassaïne and Somaya Al-Ma'adeed from Qatar University. The submitted method combines hundreds of geometrical features that were made available in [9] using a logistic regression classifier. Those features are based on number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures and chain codes.

QATAR-b method: Submitted by the same group as the previous method. It uses the most discriminant of the features cited above.

TSINGHUA method: Submitted by Lu Xu, Xiaoqing Ding, Liangrui Peng and Xin Li from State Key Laboratory of Intelligent Technology and Systems, Department of Electronic Engineering, Tsinghua University, Beijing, P.R.China.

The methodology adopts a grid microstructure feature approach (GMSF) which processes handwritten texts in multi-line [7]. A set of microstructures are calculated using a moving grid window. The probability distribution of the microstructures forms the GMSF which describes the writing style. A method using variance weighted Chi-square distance is implemented for writer similarity measurement.

We should mention that this method was the winner of the ICDAR 2011 Writer Identification Contest [8].

HANNOVER method: Submitted by Karl-Heinz Steinke from the Hochschule Hannover, University of Applied Sciences and Arts, Germany.

The submitted method is a statistical approach. The handwriting is seen as a texture with a steady structure of line elements all over the image. For the description of such a texture a suitable set of primitive elements has to be found whose frequency of occurrence is suited to distinguishing different writers to the greatest possible extent. The line segments of which the writing is composed can be taken as primitive elements of a handwriting specimen. Straight line segments may be obtained by the run lengths of pixel chains. The number and length of pixel chains is determined in eight different directions and for each direction a frequency distribution is made. The features obtained by this shift-invariant transformation are nearly text independent as long as there is enough text at hand (about three to five handwriting lines). The feature vector furnishes information about the sloping position, size, regularity and roundness of the handwriting. The submitted method can be imagined as a shredder which is fed with 8 rotated documents. The feature vectors obtained by the described method have a very high dimension. As neighbored components of the vector are strongly correlated they are added to a certain degree so that only 8 features in each direction remain. The final feature vector used has 64 components.

3. Performance evaluation

In order to measure the accuracy of the submitted methodologies we use the soft *TOP-N* and the hard *TOP-N* criterion. For every document image of the benchmarking dataset we calculate the distance to all other document images of the dataset using the participants' submitted executables. Then, we sort the results from the most similar to the less similar document image.

For the soft *TOP-N* criterion, we consider a correct hit when **at least one** document image of the same writer is included in the *N* most similar document images. Concerning the hard *TOP-N* criterion, we consider a correct hit when **all** *N* most similar document images are written by the same writer. For all document images of the benchmarking dataset we count the correct hits. The quotient of the total number of correct hits to the total number of the document images in the benchmarking dataset corresponds to the *TOP-N* accuracy. The values of *N* used for the soft criterion are 1, 2, 5 and 10 while for the hard criterion are 2 and 3. Since we have 4 document images per writer, 3 is the maximum value of *N* for the hard criterion.

For each criterion (soft or hard), we calculate the ranking of every submitted methodology. The final ranking is calculated after sorting the accumulated ranking value for all criteria (as in [8]). Specifically, let $R(j)$ be the rank of the submitted method for the j^{th} criterion, where $j=1\ldots m$, m denotes the total number of criteria. As denoted in (1), for each writer identification method, the final ranking S is achieved by the m rankings summation. The smaller the value of S the better performance is achieved by the corresponding method.

$$S = \sum_{j=1}^{m} R(j) \tag{1}$$

4. Evaluation results

We evaluated the performance of all participating algorithms using the soft *TOP-N* and the hard *TOP-N* criterion presented in the previous section. The evaluation results of all participating methods using the entire dataset are presented in Tables 1 and 2 while the evaluation results for each language independently are presented in Tables 3 (English) and 4 (Greek). In all tables, the results that correspond to the highest accuracy are marked in bold. Also, the ranking position of each methodology is presented in parentheses. Concerning language dependent experiments only soft *TOP-N* criterion is feasible since only two documents are available per writer and the one is used as query. As it is mentioned in Section 1, the benchmarking set was created with the help of 100 writers that were asked to copy four parts of text in two languages (English and Greek).

Table 1. Soft evaluation using entire dataset (%)

Method	*TOP-1*	*TOP-2*	*TOP-5*	*TOP-10*
TEBESSA-a	92,3 (3)	96,5 (2)	98,8 (2)	99,0 (2)
TEBESSA-b	89,8 (4)	94,3 (4)	97,8 (3)	98,8 (3)
TEBESSA-c	**94,5 (1)**	**97,3 (1)**	**99,3 (1)**	**99,3 (1)**
QATAR-a	70,3 (7)	80,8 (7)	91,8 (6)	95,3 (7)
QATAR-b	80,0 (6)	87,3 (6)	95,0 (5)	98,0 (5)
TSINGHUA	92,8 (2)	95,8 (3)	97,8 (3)	98,3 (4)
HANNOVER	85,5 (5)	90,3 (5)	95,3 (4)	97,3 (6)

Table 2. Hard evaluation using entire dataset (%)

Method	*TOP-2*	*TOP-3*
TEBESSA-a	57,5 (2)	**38,0 (1)**
TEBESSA-b	57,5 (2)	29,3 (3)
TEBESSA-c	**65,0 (1)**	37,8 (2)
QATAR-a	32,3 (6)	11,3 (7)
QATAR-b	34,0 (5)	15,3 (6)
TSINGHUA	51,5 (3)	27,3 (4)
HANNOVER	41,5 (4)	22,8 (5)

Table 3. Soft evaluation using only the English documents of the dataset (%)

Method	*TOP-1*	*TOP-2*	*TOP-5*	*TOP-10*
TEBESSA-a	89,5 (3)	**96,0 (1)**	97,0 (2)	**98,5 (1)**
TEBESSA-b	83,0 (4)	90,0 (4)	96,0 (3)	97,0 (3)
TEBESSA-c	91,5 (2)	95,5 (2)	**97,5 (1)**	98,0 (2)
QATAR-a	53,5 (7)	66,5 (7)	85,0 (7)	90,0 (6)
QATAR-b	72,5 (6)	82,5 (6)	92,5 (5)	96,5 (4)
TSINGHUA	**94,0 (1)**	94,5 (3)	95,5 (4)	98,0 (2)
HANNOVER	82,0 (5)	88,0 (5)	91,5 (6)	95,0 (5)

Table 4. Soft evaluation using only the Greek documents of the dataset (%)

Method	*TOP-1*	*TOP-2*	*TOP-5*	*TOP-10*
TEBESSA-a	92,0 (2)	95,0 (2)	98,5 (2)	99,0 (2)
TEBESSA-b	85,5 (5)	93,5 (4)	95,5 (4)	99,0 (2)
TEBESSA-c	**93,5 (1)**	**97,0 (1)**	**99,5 (1)**	**99,5 (1)**
QATAR-a	76,0 (6)	86,0 (7)	94,5 (5)	96,5 (4)
QATAR-b	85,5 (5)	90,0 (6)	96,0 (3)	98,5 (3)
TSINGHUA	90,0 (3)	94,0 (3)	98,5 (2)	99,0 (2)
HANNOVER	87,5 (4)	93,0 (5)	98,5 (2)	**99,5 (1)**

Table 5 presents the ranking of all participating algorithms for each experiment independently as well as the final ranking. The best overall performance is achieved by TEBESSA-c method which has been submitted by Chawki Djeddi, Labiba Souici-Meslati and Abdellatif Ennaji (LAMIS Laboratory, Mathematics and Computer Science Department, University of Tebessa, Tebessa, Algeria - LRI Laboratory, Computer Science Department, Badji Mokhtar University of Annaba, Annaba, Algeria - LITIS Laboratory, Rouen University, Saint Etienne du Rouvray, Rouen, France). Table 6 demonstrates a query image together with the four most similar document images returned by the winning methodology (correct case). It also includes information about the similarity (distance d) as well as the writer information (W) and the id per writer (Id) of each document image. Notice that for this query we measure a correct hit both for the soft *TOP-1* and for the hard *TOP-3* criterion.

Table 5. Overall ranking *S* for all experiments *T1* to *T4* presented in tables 1 to 4 respectively.

Method	*T1*	*T2*	*T3*	*T4*	*S*	*OVERALL RANK*
TEBESSA-a	9	3	8	7	**27**	**2**
TEBESSA-b	14	5	15	14	**48**	**4**
TEBESSA-c	4	3	4	7	**18**	**1**
QATAR-a	27	13	22	27	**89**	**7**
QATAR-b	22	11	17	21	**71**	**6**
TSINGHUA	12	7	10	10	**39**	**3**
HANNOVER	20	9	12	21	**62**	**5**

Table 6. Example of a correct writer identification result from the winning method

Q/R	Image	W	Id	d
Q	All the world's a stage, and all the men and women merely players: they have their exits and their entrances; and one man in his time plays many parts, his acts being seven ages.	27	1	
R1	We cannot conceive of matter being formed of nothing, since things require a seed to start from. Therefore there is not anything which returns to nothing, but all things return dissolved into their elements.	27	2	3,21
R2	Η λευτεριά δεν έχει σκοπό. Μήτε βρίσκεται στης γης στοίχει- στη γης στοίχη βρίσκεται μονάχα ο αγώνας για τη λευτεριά. Αγωνιζόμαστε για τα άστρα, και γι'αυτό ο άνθρωπος έπαψε να είναι ζώο.	27	3	3,74
R3	Ποτέ μην αναγνωρίσεις τα σύνορα του ανθρώπου! Να σπας τα σύνορα! Ν'αρνιέσαι όλα θωρούν τα μάτια σου. Να πεθαίνεις και να λες: Θάνατος δεν υπάρχει! Τι θα σε ευτυχία; Να ζεις όλες τις δυστυχίες.	27	4	4,07
R4	Η λευτεριά δεν έχει σκοπό. Μήτε βρίσκεται στη γης στοίχη - στη γης στοίχη βρίσκεται μονάχα ο αγώνας για τη λευτεριά. Αγωνιζόμαστε για τα άστρα, και γι'αυτό ο άνθρωπος έπαψε να είναι ζώο.	43	3	4,19

In contrast with Table 6, in Table 7 it is observed that for the underlying query the four more similar document images retrieved from the winning methodology correspond to different writers (error case). Although we have a miss for the soft *TOP-1* criterion we can observe in this example the visual similarity among all these document images.

Table 7. Example of a erroneous writer identification result from the winning method

Q/R	Image	W	Id	d
Q	All the world's a stage, and all the men and women merely players: they have their exits and their entrances; and one man in his time plays many parts, his acts being seven ages.	20	1	
R1	We connot conceive of matter being formed of nothing, since things require a seed to start from. Therefore there is not anything which returns to nothing, but all things return to dissolved into their elements.	60	2	3,33
R2	Η λευτεριά δεν έχει σκοπό. Μήτε βρίσκεται στης γης στοίχη- στης γης στοίχη βρίσκεται μονάχα ο αγώνας για τη λευτεριά. Αγωνιζόμαστε για τα άστρα, και γι'αυτό ο άνθρωπος έπαψε να είναι ζώο.	60	3	3,42
R3	Ποτέ μην αναγνωρίσεις τα σύνορα του ανθρώπου! Να σπας τα σύνορα! Ν' αρνιέσαι ότι θωρούν τα μάτια σου. Να πεθαίνεις κι να λες: Θάνατος δεν υπάρχει. Τι θα σε ευτυχία; Να ζεις όλες τις δυστυχίες.	60	4	3,62
R4	All the world's a stage, and all the men and women merely players: they have their exits and their entrances; and one man in his time plays many parts, his acts being seven ages.	50	1	3,66

The ranking list for the first three methodologies is:

1. TEBESSA-c ($S = 18$)
2. TEBESSA-a ($S = 27$)
3. TSINGHUA ($S = 39$).

Figure 2 presents the ranking of all participating algorithms in terms of S concerning only the English documents (m=4, experiments presented in Table 3). Figure 3 shows the ranking of all participating algorithms in terms of S concerning only the Greek documents (m=4, experiments presented in Table 4). Finally, figure 4 presents the final ranking of all participating algorithms in terms of S with m=14 (experiments presented in Tables 1-4).

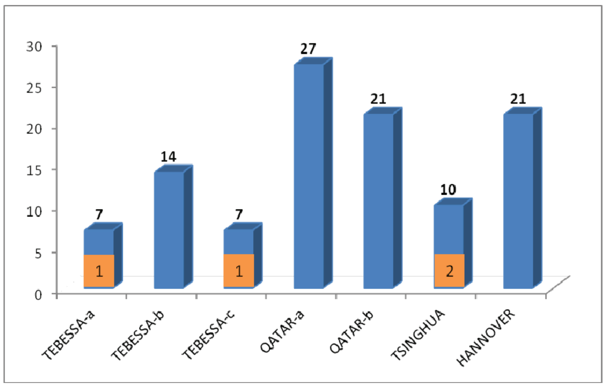


Figure 2. Ranking in terms of *S* only for English documents. The smaller the value of *S* the better performance is achieved by the corresponding method.

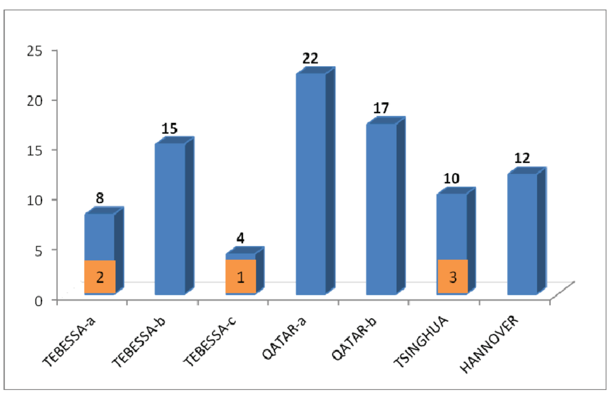


Figure 3. Ranking in terms of *S* only for Greek documents.

After a careful analysis of the data in Tables 1-4 we can stress that:

(a) The winning methodology (TEBESSA-c) outperforms all other methodologies in most cases. From all 14 experiments, it has the best performance in 10 of them while the second place in the rest 4.

(b) None of the participating methodologies manages to achieve 100% accuracy even to the *TOP-10* criterion. This is because we have cases with very similar writing styles (see Fig. 5).

(c) If we observe the language experiments independently it seems that TEBESSA-c is the winning methodology for the Greek documents whereas TEBESSA-a has the same ranking as TEBESSA-c for the English documents experiment.

(d) Concerning the Greek language, it seems that some methodologies present a drop in their accuracy compared to the case where all documents are used (e.g. TEBESSA methods, TSINGHUA) whereas some methodologies present an increase in the accuracy for the *TOP-1* and *TOP-2* cases (e.g. QATAR methods, HANNOVER).

(e) Regarding the English language, only the TSINGHUA method introduces an increase in the accuracy compared to the case of using the entire dataset whereas in all other cases we observe that there is a drop in the accuracy.

(f) The accuracy of the hard *TOP-3* criterion is very low (<40%). We can thus claim that participating methodologies are far from succeeding to cluster all similar documents at the top of the ranking list.

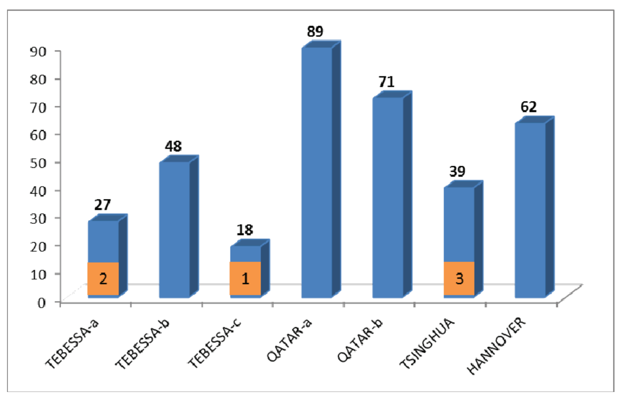


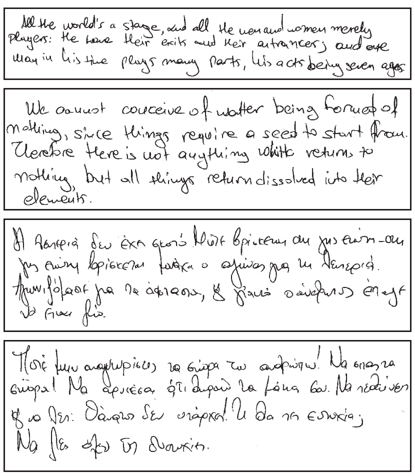
Figure 4. Final ranking in terms of *S* only for Greek documents.

5. Conclusions

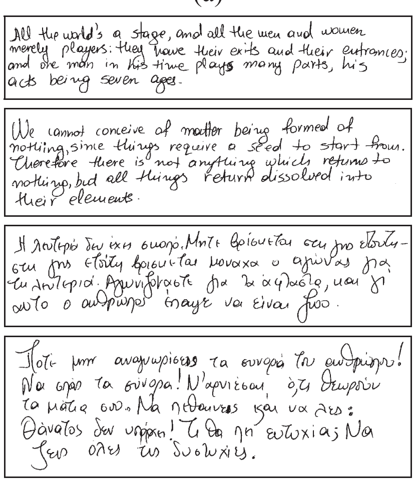
"ICFHR 2012 Writer Identification Contest, Challenge1: Latin Documents" is dedicated to record recent advances in the field of writer identification in Latin documents using established evaluation performance measures. The benchmarking dataset of the contest was created with the help of 100 writers that were asked to copy four parts of text in two languages (English and Greek). In order to measure the accuracy of the submitted methodologies we used the soft TOP-N and the hard TOP-N criterion. Four research groups with seven submitted methodologies participated in the contest. The best overall performance is achieved by TEBESSA-c method which has been submitted by Chawki Djeddi, Labiba Souici-Meslati and Abdellatif Ennaji. The winning method is based on the combination of multi-scale edge-hinge and multi-scale run-length features.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° 215064 (project IMPACT).



(a)



(b)

Figure 5. Document images from two different writers (a), (b) with very similar writing styles. Notice the similarity of the first image in (a) with the first image in (b).

References

[1] A.A. Brink, J. Smit, M.L. Bulacu and L.R.B. Schomaker, "Writer identification using directional ink-trace width measurements", Pattern Recognition, Vol. 45, Issue 1, pp. 162–171, 2012.

[2] S. Fiel and R. Sablatnig, "Writer Retrieval and Writer Identification using Local Features", 10th IAPR International Workshop on Document Analysis Systems, pp. 145 – 149, Gold Coast, Australia, 2012.

[3] H. Cao, R. Prasad and P. Natarajan, "OCR-Driven Writer Identification and Adaptation in an HMM Handwriting Recognition System", 11th International Conference on Document Analysis and Recognition, pp. 739 - 743, Beijing, China, 2011.

[4] R. Jain and D. Doermann, "Offline Writer Identification using K-Adjacent Segments", 11th International Conference on Document Analysis and Recognition, pp. 769 - 773, Beijing, China, 2011.

[5] Q. Anh Bui, M. Visani, S. Prum and J.M. Ogier, "Writer Identification using TF-IDF for Cursive Handwritten Word Recognition", 11th International Conference on Document Analysis and Recognition, pp. 844 - 848, Beijing, China, 2011.

[6] A. Chaabouni, H. Boubaker, M. Kherallah, H. El Abed and A.M. Alimi, "Combining of Off-line and On-line Feature Extraction Approaches for Writer Identification", 11th International Conference on Document Analysis and Recognition, pp. 1299 - 1303, Beijing, China, 2011.

[7] L. Xu, X. Ding, L. Peng and X. Li, "An Improved Method Based on Weighted Grid Micro-structure Feature for Text-Independent Writer Recognition", 11th International Conference on Document Analysis and Recognition, pp. 638 - 642, Beijing, China, 2011.

[8] G. Louloudis, N. Stamatopoulos, and B. Gatos, "ICDAR 2011 Writer Identification Contest", 11th International Conference on Document Analysis and Recognition, pp. 1475 –1479, Beijing, China, 2011.

[9] A. Hassaïne, S. Al-Ma'adeed, J. Mohamad Alja'am, A.M. Jaoua and A. Bouridane, "The ICDAR2011 Arabic Writer Identification Contest", 11th International Conference on Document Analysis and Recognition, pp. 1470-1474, Beijing, China, 2011.

[10] A. Fornés, A. Dutta, A. Gordo and J. Lladós, "The ICDAR 2011 Music Scores Competition: Staff Removal and Writer Identification", 11th International Conference on Document Analysis and Recognition, pp. 1511 - 1515, Beijing, China, 2011.

[11] http://users.iit.demokritos.gr/~louloud/WritIdentCont2012/

[12] http://www.iit.demokritos.gr/~louloud/WritIdentCont2012/experimental_dataset_2012.rar

[13] http://www.iit.demokritos.gr/~louloud/WritIdentCont2012/benchmarking_dataset_2012.rar

[14] C. Djeddi, L. Souici-Meslati, ''A texture based approach for Arabic Writer Identification and Verification'', International Conference on Machine and Web Intelligence, pp 115–120, Algiers, Algeria, 2010.