

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221389525>

ICFHR 2010 Handwriting Segmentation Contest

Conference Paper · November 2010

DOI: 10.1109/ICFHR.2010.120 · Source: DBLP

CITATIONS

41

READS

130

3 authors, including:



Nikolaos Stamatopoulos

National Center for Scientific Research Demokritos

46 PUBLICATIONS **745** CITATIONS

[SEE PROFILE](#)



Georgios Louloudis

National Center for Scientific Research Demokritos

45 PUBLICATIONS **695** CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



tranScriptorium [View project](#)



OldDocPro [View project](#)

ICFHR 2010 Handwriting Segmentation Contest

Basilis Gatos, Nikolaos Stamatopoulos and Georgios Louloudis

Computational Intelligence Laboratory, Institute of Informatics and Telecommunications

National Center for Scientific Research "Demokritos"

GR-153 10 Agia Paraskevi, Athens, Greece

{bgat, nstam, louloud}@iit.demokritos.gr

Abstract— The general objective of the ICFHR 2010 Handwriting Segmentation Contest organized in the context of ICFHR 2010 conference was to use well established evaluation practices and procedures in order to record recent advances in off-line handwriting segmentation. Two new benchmarking datasets, one for text line and one for word segmentation, were created in order to test and compare recent algorithms for handwritten document segmentation in realistic circumstances. Handwritten document images were produced by many writers in several languages (English, French, German and Greek). The dataset of previously organized contest (ICDAR ICDAR 2009 Handwriting Segmentation Contest) was used as training dataset. This paper describes the contest details including the datasets, the ground truth, the evaluation criteria as well as the performance of the 7 submitted methods along with a short description of each method.

Keywords— Handwritten Document Segmentation; Performance Evaluation

I. INTRODUCTION

In handwritten document recognition pipeline one of the most important and challenging tasks is the segmentation of handwritten document images into text lines and words. This task becomes really challenging due to the characteristics of unconstrained handwritten documents such as the difference in the skew angle between text lines or along the same text line, the existence of adjacent text lines or words touching, the existence of characters with different sizes and variable intra-word gaps (see Fig.1). All these problems seriously affect the segmentation and, consequently, the recognition accuracy. Therefore, it is imperative to have a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current practices in handwritten document segmentation.

Following the successful organization of the ICDAR 2007 & ICDAR 2009 Handwriting Segmentation Contests ([1], [2]), we organized the ICFHR 2010 Handwriting Segmentation Contest in order to record recent advances in off-line handwriting segmentation. Two new benchmarking datasets, one for text line and one for word segmentation, were created in order to test and compare recent algorithms for handwritten document segmentation in realistic circumstances. Handwritten document images were produced by many writers in several languages (English, French, German and Greek). The dataset of previously

organized contest was used as training dataset. For the evaluation, a well established approach that is also employed by other document segmentation contests ([1], [2], [3]) is used. This paper describes the contest details including the datasets, the ground truth, the evaluation criteria as well as the performance of the 7 submitted methods along with a short description of each method.

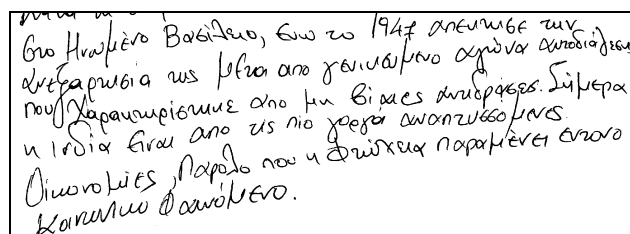
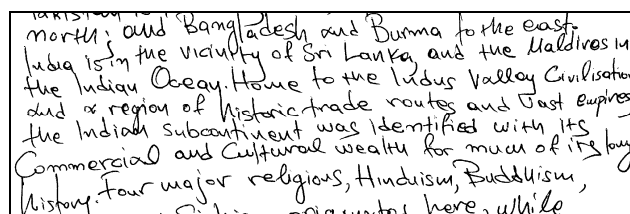


Figure 1. Samples of unconstrained handwritten documents.

II. THE CONTEST

We focused on the evaluation of text line and word segmentation methods using a variety of scanned handwritten documents. Based on these documents, we manually annotated the ground truth for text line and word segmentation and created the benchmarking datasets. The authors of candidate methods registered their interest in the competition and downloaded the training dataset (200 document images and associated ground truth from the ICDAR 2009 Handwriting Segmentation Contest) as well as the corresponding evaluation software. At a next step, all registered participants were required to submit two executables (one for text line segmentation and one for word segmentation). Both the ground truth and the result information were raw data image files with zeros corresponding to the background and all other values defining different segmentation regions. After the evaluation of all candidate methods, the testing dataset (100 images and associated ground truth) along with the evaluation software became publicly available [4].

The documents used in order to build the training and test datasets came from several writers that were asked to copy a given text. All documents did not include any non-text elements (lines, drawings, etc.) and were written in several languages (English, French, German and Greek). A sample of a text line and word segmentation ground truth annotation can be seen in Fig. 2(a),(c). Based on these annotations we build the corresponding raw image files in which all pixels that have the same value (greater than zero) belong to the same segmentation regions (see Fig. 2(b),(d)).

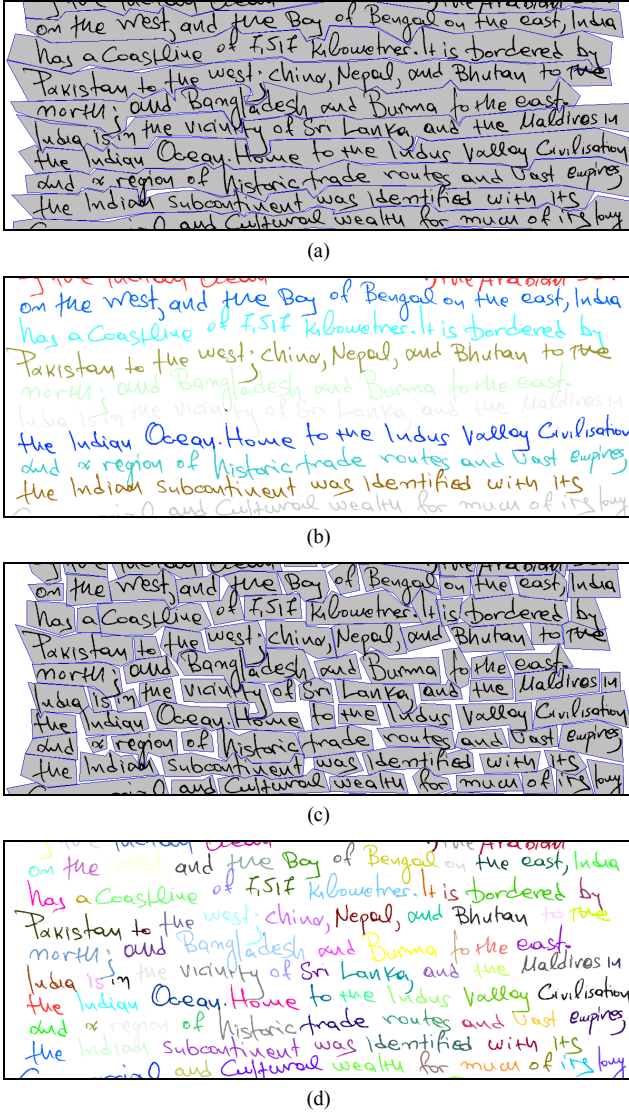


Figure 2. (a), (c) Samples of text line and word segmentation ground truth annotation and (b), (d) the corresponding raw image files.

III. PERFORMANCE EVALUATION

The performance evaluation method used was based on counting the number of matches between the entities detected by the algorithm and the entities in the ground truth

[5]. We used a MatchScore table whose values are calculated according to the intersection of the ON pixel sets of the result and the ground truth.

Let I be the set of all image points, G_j the set of all points inside the j ground truth region, R_i the set of all points inside the i result region, $T(s)$ a function that counts the elements of set s . Table $MatchScore(i,j)$ represents the matching results of the j ground truth region and the i result region:

$$MatchScore(i,j) = \frac{T(G_j \cap R_i \cap I)}{T((G_j \cup R_i) \cap I)} \quad (1)$$

An example of how to calculate the $MatchScore(i,j)$ table is given in Fig. 3.

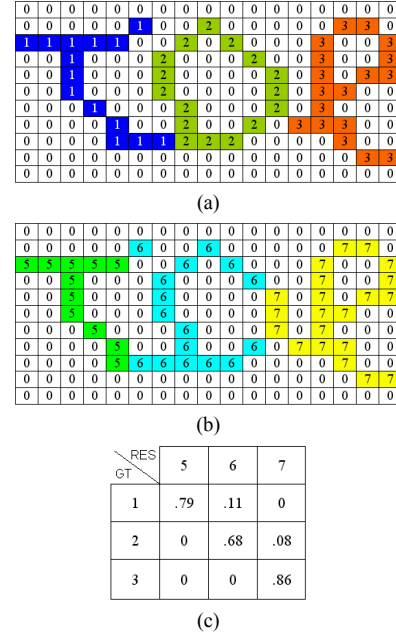


Figure 3. (a) Segmentation ground truth image, (b) segmentation result image and (c) the corresponding $MatchScore(i,j)$ table.

We consider a region pair as a one-to-one match only if the matching score is equal to or above the evaluator's acceptance threshold T_a . If N is the count of ground-truth elements, M is the count of result elements, and $o2o$ is the number of one-to-one matches, we calculate the detection rate (DR) and recognition accuracy (RA) as follows:

$$DR = \frac{o2o}{N}, \quad RA = \frac{o2o}{M} \quad (2)$$

A performance metric FM can be extracted if we combine the values of detection rate and recognition accuracy:

$$FM = \frac{2DR \cdot RA}{DR + RA} \quad (3)$$

A global performance metric SM for handwriting segmentation is extracted by calculating the average values for FM metric for text line and word segmentation. The evaluation software [4] that calculates FM metric is shown in Fig. 4.

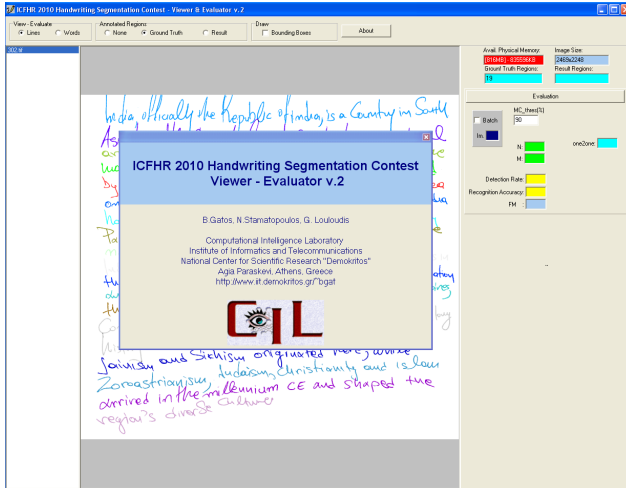


Figure 4. Contest evaluation software.

IV. METHODS AND PARTICIPANTS

Five research groups have participated in the competition with seven different algorithms (two participants submitted two algorithms each). Six submissions included both text line and word segmentation algorithms while one submission included only a text line segmentation methodology. Brief descriptions of the methods are given in this section.

NifiSoft method: Submitted by Abdelâali Hassaïne of the NifiSoft, Saint-Etienne, France.

a. Line segmentation is performed by adaptively thresholding a double-smoothed version of the original image. The size of the thresholding window is chosen in such a way that it maximizes the number of vertical lines that intersect with each connected component at exactly two transition pixels: the aim of this step is to ensure that each connected component belongs to only one line. However, some lines might be split into several connected components which are subsequently merged using standard proximity rules. These rules are combined using a logistic regression classifier. Finally, foreground pixels are assigned to the closest connected component. Word segmentation is performed by thresholding a smoothed version of a generalized chamfer distance in which the horizontal distance is slightly favored. The global threshold is determined using a logistic regression according to distance, size and proportion features of each line.

b. Line segmentation methodology remains the same while concerning word segmentation the distance between each pair of neighboring connected components is estimated from the Voronoi diagram of all the connected components. The global threshold is also determined in the same way.

IRISA method: Submitted by Aurélie Lemaître of the IRISA Laboratory, IMADOC team, Université de Rennes I, Rennes, France. The method is based on the principles of the perceptive vision, that is to say combine several levels of resolution of the images and use the saliency of structural

elements. An implementation based on a grammatical method, DMOS-P (Description and Modification of the Segmentation with Perceptive vision) [6] is used. Thus, a generic grammatical description of the organization of a page of text into text lines and words, using two levels of resolution has been realized. The associate parser is automatically produced by a compilation step. The localization of the text lines is realized using a low resolution image. Indeed, at low resolution, the text lines appear as line segments. Then, an analysis in the resolution of the initial image enables to associate each connected component to a text line. Thanks to the use of the global vision, conflicting connected components can be detected when two text lines overlap. In that case, the grammatical level asks a re-segmentation of the connected components. When each connected components has been associated to one text line, the distances between connected components are computed using a Voronoi graph. Then, a k-mean enables to separate the inter and intra word distances.

CUBS method: Submitted by Zhixin Shi, Srirangaraj Setlur and Venu Govindaraju of the Center for Unified Biometrics and Sensors (CUBS), University at Buffalo, SUNY, New York, USA. Both text line and word segmentation methods are based on a connectivity mapping using directional run-length analysis ([7], [8]). A handwritten document image is firstly mapped into a connectivity map which reveals the text line patterns, from which the text lines are extracted. For word segmentation, a different parameter is used to show word-like primitives in the map. Then, the distances between the consecutive word primitives are computed using convex hull distance. A bi-modal fitting is applied to find the threshold in determining the minimal word gap in the document image.

TEI method: Submitted by A. Nicolaou of the Technological Educational Institution of Athens, Greece. Line segmentation is done with an improved shredding [9] technique. The image is separated in horizontal strips along the white most paths (local minima tracers) of a pyramid blur of the original binary image. Each connected component of the original image is assigned to a line strip. The main innovation in this method is the complex shape of the blurring filter. On the training set this method achieved an arbitrary score of 99.53% while the previous implementation achieved 98.9% by the same standards. Concerning word segmentation, for each detected line in a page we fill all bounding boxes of each component, we then smear vertically and so produce a sequence of shapes which we call “syllables”. A “syllable” almost always, is a sequence of letters never extending a word. We extracted all “syllable” sequences from the training set and extracted features for each gap along two consecutive “syllables”. For each gap in between “syllables”, we extracted 7 features which were normalised linearly in [0,1] according to all the patterns (“syllable”gaps) extracted from the training set. The features were extracted by taking various geometrical aspects of the gaps and the page (in pixels) and the histogram of the gap's size within a particular page of the

training set. We trained a 10 feed-forward Neural Network with identical parameters and training sets to distinguish “syllable” gaps that separate words from those that don't. For each pattern (“syllable” gap) we round the average of the NN outputs and merge the two “syllables” into a larger one accordingly. We used 10 NN to maximize the reliability of our classifiers generalization ability.

ILSP method: Submitted by V. Papavassiliou, T. Stafylakis, V. Katsouros and G. Carayannis of the Institute for Language and Speech Processing (ILSP) in Athens, Greece and based on [10], [11].

a. Firstly, we divide the image document page into vertical zones and obtain initial sets of text and gap areas in each zone by exploiting the piece-wise projections. Then, we find the optimal succession of text and gap stripes with the application of the Viterbi algorithm on an HMM with parameters drawn from statistics of each type of area from the whole document image. The line separators are obtained by combining the boundaries of the individual areas along the width of the page. Finally text lines are located with the application of simple geometrical constraints that conclude if a connected component (CC) can be directly assigned or it should be split because it lies across successive text lines. Word segmentation requires that the document is already segmented into text lines. We assume that successive words do not touch each other and as a result word separators would lie at the gap between two successive CCs. Therefore, word segmentation can be seen as a problem which requires the formulation of a gap metric and the clustering of the gaps in “inter” or “intra” word classes. To measure the gap metric of successive CCs, we use the negative logarithm of the objective function of a soft-margin linear SVM. We employ a nonparametric approach to estimate the probability density function of the gap metrics and have observed that the “inter” words gaps are accumulated to the most right lobe of the pdf while the “intra” word gaps are gathered to the left lobe. The classification threshold is chosen to be equal to the minimum between the two main lobes.

b. A text-line segmentation method is based on [12] (based on binary morphology). The basic steps of our approach are: a) apply dilation and sub-sampling to produce a low resolution image, in which the underlying texture of text lines is apparent while preventing aliasing, b) use binary rank order filtering to enhance the text-line structures and c) apply dilations and (p,q)-th generalized foreground rank openings successively to join close and horizontally overlapping regions while preventing a merge in the vertical direction. These operations evolve the candidate text lines and distinguish special patterns, which imply that text lines have come very close or have been merged. Then, the image is over-sampled to its original resolution and the connected components (CCs) of the resulting image correspond to the text lines of the initial document image. Finally, each CC of the initial document image is assigned to the text line that intersects, whereas if it intersects more than one text lines, i.e. a touching component, we cut it using the local ridges produced with the application of the watershed algorithm.

V. EVALUATION RESULTS

We evaluated the performance of all participating algorithms for text line and word segmentation using equations (1)–(3), the test dataset (100 images) and the corresponding ground truth. The acceptance threshold we used was $T_a=95\%$ for text line segmentation and $T_a=90\%$ for word segmentation. The number of text lines and words for all 100 document images was 1629 and 15130, respectively. All evaluation results are shown in Table I while a graphical representation of the evaluation results is given in Fig. 5, 7, 9. In order to get an overall ranking for both text line and word segmentation, we used the global performance metric SM (see Section III) in order to compare the 6 algorithms that provide both text line and word segmentation results (NifiSoft-a, NifiSoft-b, IRISA, CUBS, TEI and ILSP-a). From Table I we observe no significant deviation in the performance among participating methods since all submitted algorithms achieved a global score from 92.18% to 94.20%. Submitted text line segmentation methods found to perform better than the submitted word segmentation methods since they achieve a score from 94.86% to 97.63% compared to a score from 87.7% to 91.17% for word segmentation. The NifiSoft-a method outperforms all other methodologies in the overall ranking, achieving $SM=94,20\%$. Representative examples of text line and word segmentation results of the NifiSoft-a method are shown in Fig. 6. The ranking list for all six methodologies is:

1. **NifiSoft-a** ($SM=94,20\%$)
2. **NifiSoft-b** ($SM=93,97\%$)
3. **CUBS** ($SM=93,45\%$)
4. **ILSP-a** ($SM=93,29\%$)
5. **TEI** ($SM=92,42\%$)
6. **IRISA** ($SM=92,18\%$)

TABLE I. DETAILED EVALUATION RESULTS.

		M	$o2o$	DR (%)	RA (%)	FM (%)	SM (%)
NifiSoft-a	Lines	1634	1589	97,54	97,25	97,40	94,20
	Words	15192	13796	91,18	90,81	91,00	
NifiSoft-b	Lines	1634	1589	97,54	97,25	97,40	93,97
	Words	15145	13707	90,59	90,51	90,55	
IRISA	Lines	1636	1578	96,87	96,45	96,66	92,18
	Words	14314	12911	85,33	90,20	87,70	
CUBS	Lines	1626	1589	97,54	97,72	97,63	93,45
	Words	15012	13454	88,92	89,62	89,27	
TEI	Lines	1637	1549	95,09	94,62	94,86	92,42
	Words	14667	13406	88,61	91,40	89,98	
ILSP-a	Lines	1656	1567	96,19	94,63	95,40	93,29
	Words	14796	13642	90,17	92,20	91,17	
ILSP-b	Lines	1655	1559	95,70	94,20	94,95	
	Words	-	-	-	-	-	

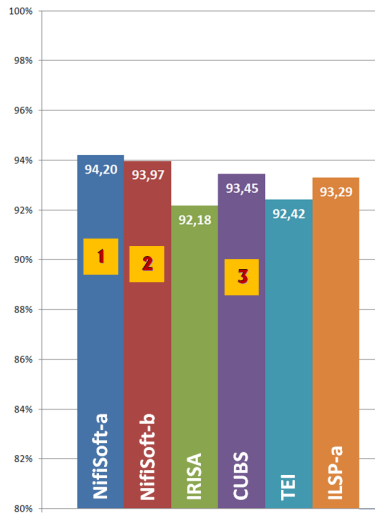
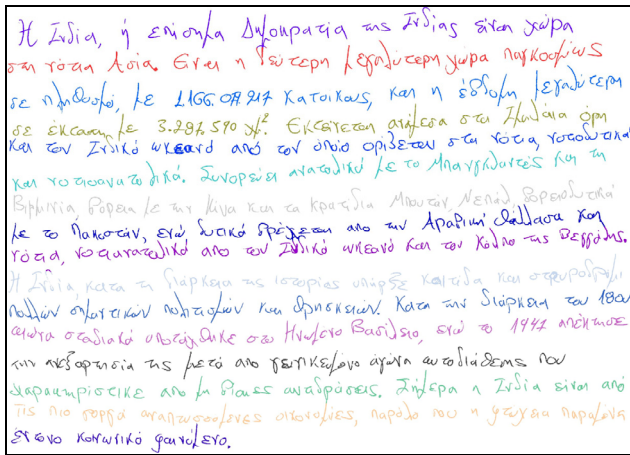
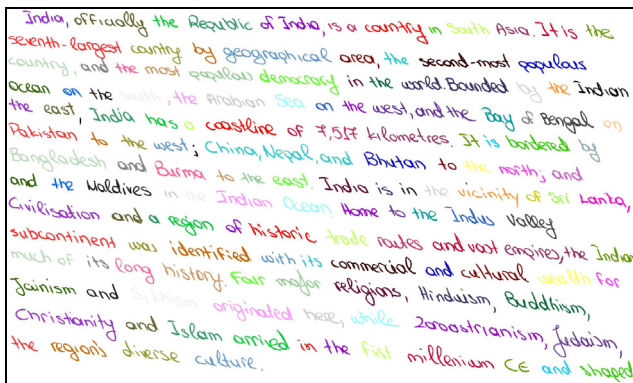


Figure 5. Overall evaluation performance for both text line and word segmentation.



(a)



(b)

Figure 6. Representative (a) text line ($FM=100\%$) and (b) word ($FM=89.61\%$) segmentation results of the NifiSoft-a method.

Concerning text line segmentation, the CUBS method achieved the highest results with $FM=97.63\%$ (Fig. 7). A representative example of text line segmentation result of

the CUBS method is shown in Fig. 8. The ranking list for text line segmentation methodologies is:

1. **CUBS** ($FM=97.63\%$)
2. **NifiSoft-a** ($FM=97.40\%$)
3. **NifiSoft-b** ($FM=97.40\%$)
4. **IRISA** ($FM=96.66\%$)
5. **ILSP-a** ($FM=95.40\%$)
6. **ILSP-b** ($FM=94.95\%$)
7. **TEI** ($FM=94.86\%$)

For the word segmentation stage, the ILSP-a method obtained the highest results with $FM=91.17\%$ (Fig. 9). A representative example of word segmentation result of the ILSP-a method is shown in Fig. 10. The ranking list for the six word segmentation methodologies is:

1. **ILSP-a** ($FM=91.17\%$)
2. **NifiSoft-a** ($FM=91.00\%$)
3. **NifiSoft-b** ($FM=90.55\%$)
4. **TEI** ($FM=89.98\%$)
5. **CUBS** ($FM=89.27\%$)
6. **IRISA** ($FM=87.70\%$)

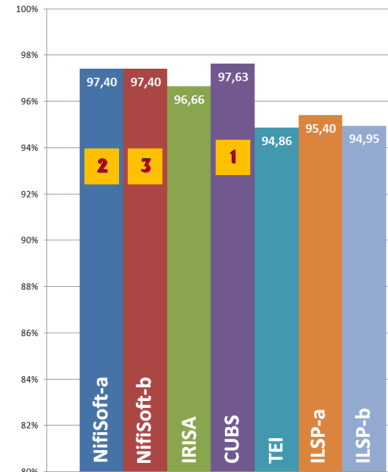


Figure 7. Evaluation performance for text line segmentation.

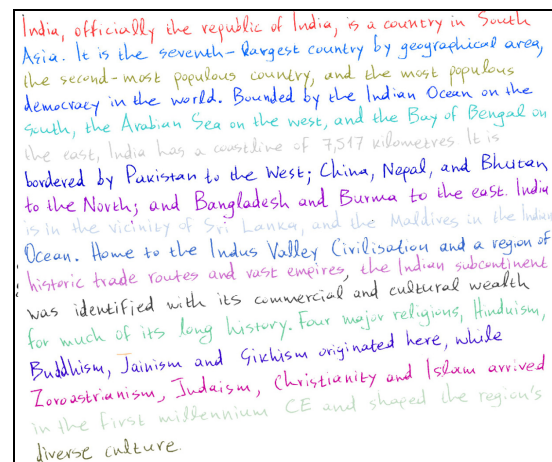


Figure 8. Representative text line segmentation result ($FM=97.14\%$) of the CUBS method.

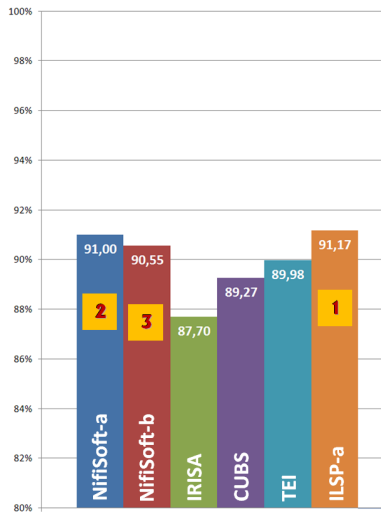


Figure 9. Evaluation performance for word segmentation.

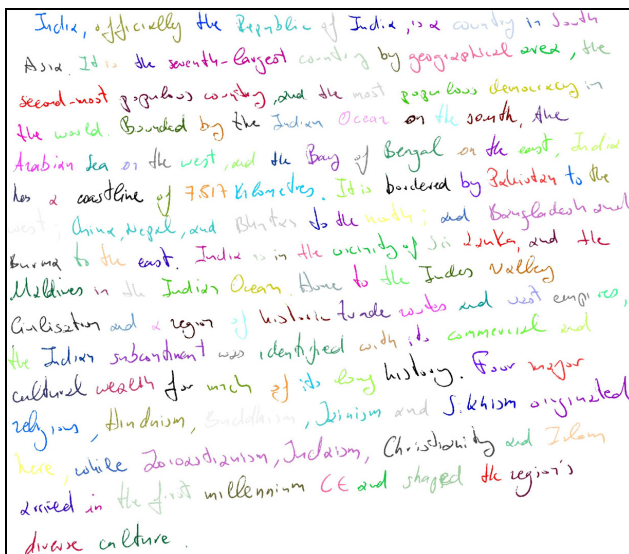


Figure 10. Representative word segmentation result ($FM=90.96\%$) of the ILSP-a method.

VI. CONCLUSIONS

ICFHR 2010 Handwriting Segmentation Contest was organized in order to record recent advances in off-line handwriting segmentation. As it is shown in the evaluation results section, the best performance considering an overall ranking for text line and word segmentation was achieved by the NifiSoft-a method submitted by Abdelâali Hassaïne of the NifiSoft, Saint-Etienne, France with overall global performance metric $SM = 94,20\%$. Considering only text line segmentation, the best performance was achieved by the CUBS method submitted by Zhixin Shi, Srirangaraj Setlur and Venu Govindaraju of the Center for Unified Biometrics and Sensors (CUBS), University at Buffalo, SUNY, New York, USA. Considering word segmentation, the best

performance was achieved by the ILSP-a method submitted by V. Papavassiliou, T. Stafylakis, V. Katsouros and G. Carayannis of the Institute for Language and Speech Processing (ILSP) in Athens, Greece.

ACKNOWLEDGMENTS

This work has been partially funded by the European Community's Seventh Framework Programme under grant agreement n° 215064 (project IMPACT).

REFERENCES

- [1] B. Gatos, A. Antonacopoulos and N. Stamatopoulos, "ICDAR2007 Handwriting Segmentation Contest", Proc. 9th International Conference on Document Analysis and Recognition (ICDAR'07), Curitiba, Brazil, September 2007, pp. 1284-1288.
- [2] B. Gatos, N. Stamatopoulos and G. Louloudis, "ICDAR2009 Handwriting Segmentation Contest", Proc. 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, July 2009, pp. 1393-1397.
- [3] A. Antonacopoulos, B. Gatos and D. Bridson, "ICDAR2005 Page Segmentation Competition", Proc. 8th International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea, August 2005, pp. 75-79.
- [4] <http://www.iit.demokritos.gr/~bgat/HandSegmCont2010/Benchmark>
- [5] I. Phillips and A. Chhabra, "Empirical Performance Evaluation of Graphics Recognition Systems", in IEEE Trans. of Patt. Analysis and Machine Intell., Vol. 21, No. 9, September 1999, pp. 849-870.
- [6] A. Lemaitre, J. Camillerapp and B. Coüasnon, "Interest of perceptive vision for document structure analysis", Proc. Human Vision and Electronic Imaging XV, 2010, doi:10.1117/12.838453.
- [7] Z. Shi, S. Setlur and V. Govindaraju, "Text Extraction from Gray Scale Historical Document Images Using Adaptive Local Connectivity Map", Proc. 8th International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea, August 2005, pp. 794- 798.
- [8] Z. Shi, S. Setlur and V. Govindaraju, "A Steerable Directional Local Profile Technique for Extraction of Handwritten Arabic Text Lines", Proc. 10th International Conference on Document Analysis and Recognition (ICDAR'09), Spain, July 2009, pp. 176-180.
- [9] A. Nicolaou and B. Gatos, "Handwritten Text Line Segmentation by Shredding Text into its Lines", Proc. 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, July 2009, pp. 626-630.
- [10] T. Stafylakis, V. Papavassiliou, V. Katsouros and G. Carayannis, "Robust Text-line and Word Segmentation for Handwritten Documents Images", Proc. Int'l Conf. Acoustics, Speech and Signal Processing, 2008, pp. 3393-3396.
- [11] V. Papavassiliou, T. Stafylakis, V. Katsouros and G. Carayannis, "Handwritten Document Image Segmentation into Text Lines and Words", in Pattern Recognition, Vol. 43, Issue 1, January 2010, pp. 369-377.
- [12] V. Papavassiliou, V. Katsouros and G. Carayannis, "A Morphological Approach for Text-Line Segmentation in Handwritten Documents", Proc. 15th International Conference on Frontiers in Handwriting Recognition (ICFHR 2010), Kolkata, India, November 2010.