

ICDAR2009 Handwriting Segmentation Contest

B. Gatos¹, N. Stamatopoulos¹ and G. Louloudis^{1,2}

¹*Computational Intelligence Laboratory, Institute of Informatics and Telecommunications,
National Center for Scientific Research "Demokritos", GR-153 10 Agia Paraskevi, Athens, Greece
{bgat, nstam}@iit.demokritos.gr*

²*Department of Informatics and Telecommunications
University of Athens, Greece
louloud@mm.di.uoa.gr*

Abstract

The Handwriting Segmentation Contest was organized in the context of ICDAR2009 conference in order to record recent advances in off-line handwriting segmentation. This paper describes the contest details including the dataset, the ground truth and the evaluation criteria and presents the results of the 12 participating methods. The contest includes handwritten document images produced by many writers in several languages (English, French, German and Greek). These images are manually annotated in order to produce the ground truth which corresponds to the correct text line and word segmentation result. For the evaluation, a well established approach is used based on counting the number of matches between the entities detected by the segmentation algorithm and the entities in the ground truth.

1. Introduction

Segmentation of handwritten document images into text lines and words is one of the most important and challenging tasks in a handwritten recognition system. Several problems inherent in 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, seriously affect the segmentation and, consequently, the recognition accuracy. To this end, it is imperative to have a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current and new practices in handwritten document segmentation.

Following the successful organization of the ICDAR 2007 Handwriting Segmentation Contest [1], we organized the ICDAR 2009 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. For the evaluation, a well established approach that is also employed by other document segmentation contests ([1], [2], [3]) is used.

In the next Section, the contest details and an overview of the datasets are described. In Section 3, the performance evaluation method and metrics are described, while each of the participating methods is summarized in Section 4. Finally, the results of the competition are presented in Section 5 and the conclusions are drawn in Section 6.

2. The contest

In this 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 (100 document images and associated ground truth from the ICDAR 2007 Handwriting Segmentation Contest [1]) 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 (200 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 handwritten document image which is part of the test set and a sample of word segmentation ground truth annotation can be seen in Fig. 1.

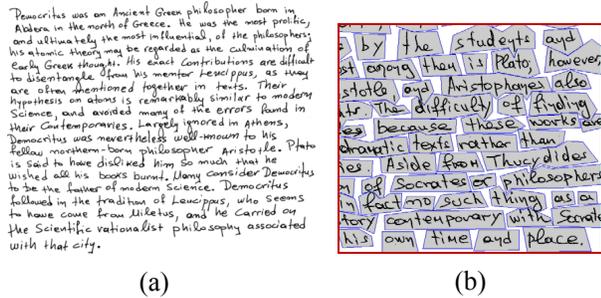


Figure 1. (a) A sample of a handwritten document image part of the test dataset and (b) a sample of word segmentation ground truth annotation.

3. 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:

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

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:

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

A global performance metric SM for handwriting segmentation is extracted by calculating the average values for FM metric for text line and word segmentation.

4. Methods and participants

We had 12 submissions to the competition while 4 of them included only a text line segmentation methodology. Brief descriptions of the methods are given in this section.

CASIA-MSTSeg method: Submitted by F. Yin, X.D. Zhou, Q.F. Wang and C.L. Liu of the Institute of Automation of Chinese Academy of Sciences (CASIA) in Beijing, China and based on [6]. Connected components are first split based on several geometric constraints and then grouped into a tree structure by the minimal spanning tree (MST) algorithm with the distance metric designed by supervised learning. Text lines are extracted from the tree by dynamically cutting selected edges. Concerning word segmentation, for each gap between adjacent connected components in a line, 11 geometric features are extracted and fed to an SVM classifier for classifying gaps into between-word and within-word ones.

CMM method: Submitted by A. Hassaïne and B. Marcotegui of the Center of Mathematical Morphology in Paris School of Mines, France. A first labeling of the image is applied using the minimum number of horizontal intersections with the text. Cases of components with several labels or with labels that have to be merged are then handled based on several rules. For word detection, the average distance between the bounding boxes of the connected components of each line is computed. A distance is considered to be an inter-word distance if it is larger than a threshold.

CUBS method: Submitted by Z. Shi, S. Setlur and V. Govindaraju of the Center for Unified Biometrics and Sensors (CUBS), University at Buffalo, SUNY, in New York, USA. The line separation algorithm is based on an improved directional run-length analysis [7], [8]. Concerning word segmentation, at each background pixel location, a horizontal background

run including the location is traced and the run-length is saved in a new image buffer for each pixel location. A simple thresholding of the new buffer reveals word primitives. Then, the distances between the consecutive word primitives are computed using convex hull distance. A threshold for grouping of the word primitives is calculated based on the mean and variance of the distances.

ETS method: Submitted by D. Rivest-Henault and M. Cheriet of the Ecole de technologie superieure (ETS) of the University of Quebec in Montreal, Canada. Both text line and word segmentation methods are based on text smearing and morphological operations. Most of the involved operations take into account the local text line orientation. This has the benefit of greatly reducing the frequency of accidental line merging. The text is smeared using a modified version of Weickert's coherence-enhancing diffusion filter while the smeared image is binarized using Otsu's algorithm.

ILSP-LWSeg-09 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 [9]. Text line detection makes use of the Viterbi algorithm. Candidate line separators are obtained and combined by minimizing a function which exploits the distance between the separators and the local foreground density. For word segmentation, as a metric of separability between the two sets (inter and intra-word gap), the negative logarithm of the objective function of a soft-margin linear SVM is used.

JadavpurUniv method: Submitted by R. Sarkar, A. Khandelwal, P. Choudhury, N. Das, S. Basu, M. Kundu, M. Nasipury, D. K. Basu and A. F. Mollah of the CSE Dept., Jadavpur University in Kolkata, India. Text line detection is based on Connected Component Labeling and on comparison of components in a neighborhood. Then, the dimensional features of the components are analyzed to determine the style of handwriting and threshold values are set for inter-word spacing in case of both isolated and cursive handwriting. Words are then identified on the basis of difference in intra-word and inter-word spacing.

LRDE method: Submitted by T. Geraud of the EPITA Research and Development Laboratory (LRDE) in Le Kremlin-Bicetre, France. The input image is sub-sampled in both dimensions while turning it into a gray-level image. Then, an anisotropic Gaussian filtering is applied (mainly horizontal). The morphological watershed transform

is computed, leading into a partition of the image into regions. To obtain the segmentation into lines, a simple merging procedure is run on the region adjacency graph. Word segmentation is based on attribute morphological closing as well as on morphological watershed transform. The source code can be downloaded from [10].

PAIS method: Submitted by S. Lu, S. Fan, Y. Wen and Y. Lu of the ECNU-SRI Joint Lab for Pattern Analysis and Intelligence System, Shanghai, China. The image is vertically divided into several strips. Potential text lines are detected based on the horizontal projection values of each strip in order to estimate the average distance between adjacent text lines. The text lines are then finalized by applying the knowledge of estimated line-distance and reasonable black-to-white traversal numbers. For word segmentation of each text line, the number of possible words is estimated by the black-to-white traversal numbers. A gap is considered as inter-word gap if it is larger than a threshold calculated by estimated number of possible words.

AegeanUniv method (text line segmentation only): Submitted by E. Kavallieratou of the University of Aegean in Samos, Greece and based on [11], [12]. The page is vertically separated into three areas and for each area a horizontal projection profile is employed. The valleys with minima less than a certain threshold are considered to be likely beginnings of line segments. Sequentially, the area is examined pixel by pixel until an entire white path is outlined.

PortoUniv method (text line segmentation only): Submitted by J. Cardoso of Faculdade de Engenharia, University of Porto in Porto, Portugal and based on [13]. The image is handled as a graph and the text lines as connected paths between the two lateral margins of the image. The paths to look for are the shortest paths between the two lateral margins while paths through black pixels are favoured. An efficient dynamic programming approach is used to find the minimum paths.

PPSL method (text line segmentation only): Submitted by A. Alaei, P. Nagabhushan and U. Pal of the University of Mysore in Mysore, India. The text page is vertically crumbled into few strip-like structures. In order to get Potential Piece-wise Separation Line (PPSL) between two consecutive lines, the white/black spaces in each strip are analyzed. Next, such PPSLs are concatenated or extended in both directions to produce the complete segmentation lines based on distance analysis of each PPSL with left and right neighboring PPSLs.

REGIM method (text line segmentation only): Submitted by M. Mezghani, W. Boussellaa and A. Alimi of the Research Group on Intelligent Machines (REGIM) of the University of Sfax in Tunisia and A. Zahour of the Equipe Gestion Electronique de Document (GED), University of Le Havre, France. The methodology is based on (i) document decomposition into columns and blocks covering all textual elements, (ii) classification of the generated blocks using several statistical parameters and (iii) text line detection based on a fuzzy base line determination using a fuzzy C-means algorithm.

5. Evaluation results

We evaluated the performance of all participating algorithms for text line and word segmentation using equations (1)–(3), the test dataset (200 images) and the corresponding ground truth. The acceptance threshold we used was $T_a=95\%$ for text line segmentation and $T_w=90\%$ for word segmentation. The number of text lines and words for all 200 document images was 4034 and 29717, respectively. All evaluation results are shown in Table 1 while a graphical representation of the evaluation results is given in Fig. 2-4. In order to get an overall ranking for both text line and word segmentation, we used the global performance metric SM (see Section 3) in order to compare the 8 algorithms that provide both text line and word segmentation results (CASIA-MSTSeg, CMM, CUBS, ETS, ILSP-LWSeg-09, JadavpurUniv, LRDE and PAIS). As it can be observed (Fig.2), the ILSP-LWSeg-09 method outperforms all other methodologies in the overall ranking, achieving $SM=96,91\%$. The ranking list for the first five methodologies is:

1. ILSP-LWSeg-09 ($SM=96,91\%$)
2. PAIS ($SM=94,53\%$)
3. CMM ($SM=93,66\%$)
4. CUBS ($SM=93,24\%$)
5. CASIA-MSTSeg ($SM=90,27\%$)

Concerning text line segmentation, the CUBS method achieved the highest results with $FM=99,53\%$ (Fig. 3). The ranking list for the first five methodologies for text line segmentation is:

1. CUBS ($FM=99,53\%$)
2. ILSP-LWSeg-09 ($FM=99,05\%$)
3. PAIS ($FM=98,52\%$)
4. CMM ($FM=98,42\%$)
5. CASIA-MSTSeg ($FM=95,68\%$)

Table 1. Detailed evaluation results.

		M	$o2o$	DR (%)	RA (%)	FM (%)	SM (%)
CASIA-MSTSeg	Lines	4049	3867	95,86	95,51	95,68	90,27
	Words	31421	25938	87,28	82,55	84,85	
CMM	Lines	4044	3975	98,54	98,29	98,42	93,66
	Words	31197	27078	91,12	86,80	88,91	
CUBS	Lines	4036	4016	99,55	99,50	99,53	93,24
	Words	31533	26631	89,62	84,45	86,96	
ETS	Lines	4033	3496	86,66	86,68	86,67	85,80
	Words	30848	25720	86,55	83,38	84,93	
ILSP-LWSeg-09	Lines	4043	4000	99,16	98,94	99,05	96,91
	Words	29962	28279	95,16	94,38	94,77	
Jadavpur Univ	Lines	4075	3541	87,78	86,90	87,34	85,04
	Words	27596	23710	79,79	85,92	82,74	
LRDE	Lines	4423	3901	96,70	88,20	92,25	88,09
	Words	33006	26318	88,56	79,74	83,92	
PAIS	Lines	4031	3973	98,49	98,56	98,52	94,53
	Words	30560	27288	91,83	89,29	90,54	
AegeanUniv	Lines	4054	3130	77,59	77,21	77,40	-
PortoUniv	Lines	4028	3811	94,47	94,61	94,54	-
PPSL	Lines	4084	3792	94,00	92,85	93,42	-
REGIM	Lines	4563	1629	40,38	35,70	37,90	-

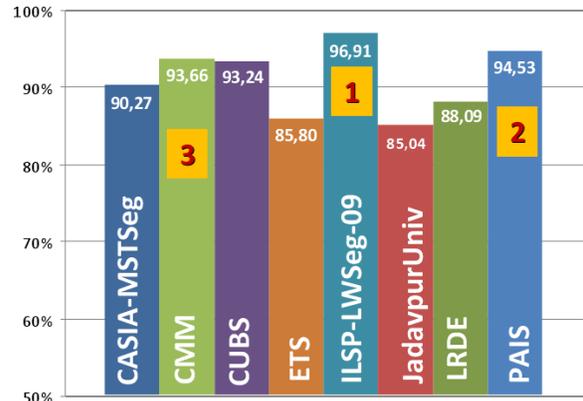


Figure 2. Overall evaluation performance for both text line and word segmentation.

For the word segmentation stage, the ILSP-LWSeg-09 method obtained the highest results with $FM=94,77\%$ (Fig. 4). The ranking list for the first five methodologies for word segmentation is:

1. ILSP-LWSeg-09 ($FM=94,77\%$)
2. PAIS ($FM=90,54\%$)
3. CMM ($FM=88,91\%$)
4. CUBS ($FM=86,96\%$)
5. ETS ($FM=84,93\%$)

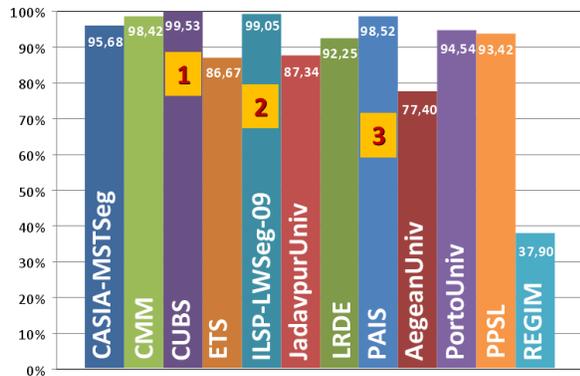


Figure 3. Evaluation performance for text line segmentation.

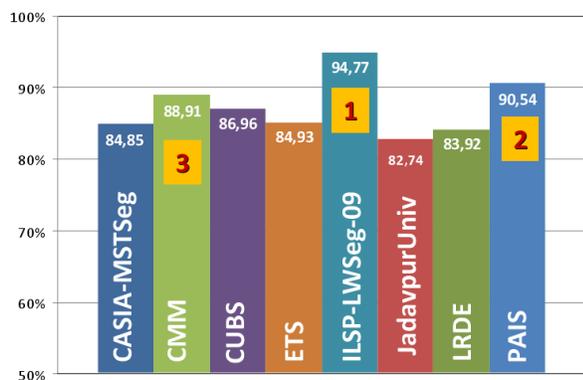


Figure 4. Evaluation performance for word segmentation.

6. Conclusions

ICDAR 2009 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 as well as a ranking only for word segmentation, was achieved by the ILSP-LWSeg-09 method of the Institute for Language and Speech Processing (ILSP) with overall global performance metric $SM = 96,91\%$ and word segmentation performance metric $FM = 94,77\%$. Considering only text line segmentation, the best performance was achieved by the CUBS method of the Center for Unified Biometrics and Sensors (CUBS) with performance metric FM equal to $99,53\%$.

References

[1] B. Gatos, A. Antonacopoulos, N. Stamatopoulos, "ICDAR2007 Handwriting Segmentation Contest", *9th International Conference on Document Analysis and Recognition (ICDAR'07)*, Curitiba, Brazil, September 2007, pp. 1284-1288.

[2] A. Antonacopoulos, B. Gatos, D. Bridson, "ICDAR2007 Page Segmentation Competition", *9th International Conference on Document Analysis and Recognition (ICDAR'07)*, Curitiba, Brazil, September 2007, pp. 1279-1283.

[3] A. Antonacopoulos, B. Gatos, D. Bridson, "ICDAR2005 Page Segmentation Competition", *8th International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, Korea, August 2005, pp. 75-79.

[4] www.iit.demokritos.gr/~bgat/HandSegmCont2009/Benchmark

[5] I. Phillips, A. Chhabra, "Empirical Performance Evaluation of Graphics Recognition Systems", *IEEE Trans. of Patt. Analysis and Machine Intell.*, Vol. 21, No. 9, September 1999, pp. 849-870.

[6] F. Yin, C.-L. Liu, "Handwritten text line segmentation by clustering with distance metric learning", *Proc. 11th Int. Conf. on Frontiers in Handwriting Recognition*, Montreal, Canada, August 2008, pp. 229-234.

[7] Z. Shi, S. Setlur, V. Govindaraju, "Text Extraction from Gray Scale Historical Document Images Using Adaptive Local Connectivity Map". *8th International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, Korea, August 2005, pp. 794- 798.

[8] Z. Shi, S. Setlur, V. Govindaraju, "A Steerable Directional Local Profile Technique for Extraction of Handwritten Arabic Text Lines", to appear in *10th International Conference on Document Analysis and Recognition (ICDAR'09)*, Spain, July 2009.

[9] T. Stafylakis, V. Papavassiliou, V. Katsouros, G. Carayannis, "Robust Text-line and Word Segmentation for Handwritten Documents Images", in *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, 2008, pp. 3393-3396.

[10] <http://olena.lrde.epita.fr/ModuleIcdar>

[11] E. Kavallieratou, N. Dromazou, N. Fakotakis, G. Kokkinakis, "An Integrated System for Handwritten Document Image Processing", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 17, No. 4, 2003, pp. 101-120.

[12] E. Kavallieratou, N. Fakotakis, G. Kokkinakis, "An Off-line Unconstrained Handwriting Recognition System", *International Journal of Document Analysis and Recognition*, No 4, 2002, pp. 226-242.

[13] J. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, "A Connected Path Approach for Staff Detection on a Music Score", *Proceedings of the International Conference on Image Processing (ICIP 2008)*, San Diego, USA, October 2008, pp. 1005 - 1008.