ICDAR2007 Handwriting Segmentation Contest

B. Gatos[1], A. Antonacopoulos[2] and N. Stamatopoulos[1]

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

[2]*Pattern Recognition and Image Analysis (PRImA) Research Lab*
School of Computing, Science and Engineering, University of Salford, Manchester, M5 4WT, United Kingdom
http://www.primaresearch.org

Abstract

This paper presents the results of the Handwriting Segmentation Contest that was organized in the context of ICDAR2007. The aim of this contest was to use well established evaluation practices and procedures in order to record recent advances in off-line handwriting segmentation. Two benchmarking datasets (one for text line and one for word segmentation) were used in a common evaluation platform in order to test and compare all submitted algorithms for handwritten document segmentation in realistic circumstances. The results of the evaluation of five algorithms submitted by participants as well as of two state-of-the-art algorithms are presented. The performance evaluation method is based on counting the number of matches between the text lines or words detected by the algorithms and the text line or words of the ground truth.

1. Introduction

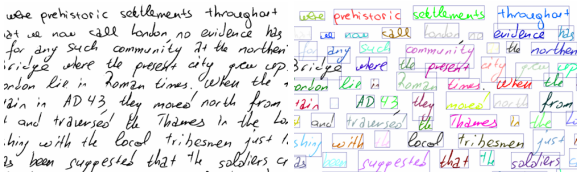
Document image segmentation is a critical stage towards unconstrained handwritten document recognition. Several problems that appear in this stage include the difference in the skew angle between text lines or along the same text line, the existence of overlapping words, adjacent lines or words touching, etc. Although successful methods have been reported in the literature, it is an open issue to have an objective evaluation of the performance of different handwriting segmentation methods in realistic circumstances. Such an evaluation requires both the creation of suitable ground truth based on manual annotation as well as the definition of a set of objective evaluation criteria. To this end, the handwriting segmentation contest was organized in the context of ICDAR2007 conference aiming to use well established evaluation practices and procedures in order to record recent advances in off-line handwriting segmentation. It is based on two benchmarking datasets (one for text line and one for word segmentation) and on evaluation strategies of previous document segmentation contests also held by the authors [1–3].

The contest details and an overview of the datasets are described in the next section. 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.

2. The contest

The ICDAR2007 Handwriting Segmentation Contest focused on the evaluation of text line and word segmentation methods using a variety of scanned handwritten documents. Based on these documents, the benchmarking datasets were created by manually annotating the ground truth for text line and word segmentation. The authors of candidate methods registered their interest in the competition and downloaded the training dataset (20 document images and associated ground truth) as well as the corresponding evaluation software. At a next step, all registered participants were required to submit two executables (one for text line detection and one for word detection) in the form of Win32 console application. 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 by the organizers of the contest, the test dataset (80 images and associated ground truth) was also given to all participants.

The documents used in order to build the training and test datasets came from (i) several writers that were asked to copy a given text; (ii) historical handwritten archives, and (iii) scanned handwritten document samples selected from the web. None of the documents included any non-text elements (lines, drawings, etc.) and were all written in several languages including English, French, German and Greek. A sample of a handwritten document image as well as the word segmentation ground truth given as part of the training dataset can be seen in Fig. 1.



(a) (b)

Figure 1. (a) A sample of a handwritten document image given as part of the training dataset and (b) the corresponding word segmentation ground truth.

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 [4]. A MatchScore table was employed whose values were calculated according to the intersection of the ON pixel sets of the result and the ground truth (a similar technique was used in [5]).

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. Based on a pixel based approach of [5], we define that:

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

We consider a match only if the matching score is equal to or above a specified acceptance threshold T_a. If N is the count of ground-truth elements, M is the count of result elements, and w_1,w_2,w_3,w_4,w_5,w_6 are pre-determined weights, we calculate the detection rate (DR) and recognition accuracy (RA) as follows:

$$DR = w_1 \frac{o2o}{N} + w_2 \frac{g\_o2m}{N} + w_3 \frac{g\_m2o}{N} \qquad (2)$$

$$RA = w_4 \frac{o2o}{M} + w_5 \frac{d\_o2m}{M} + w_6 \frac{d\_m2o}{M} \qquad (3)$$

where *o2o* (one to one match), *g_o2m* (one ground truth to many detected), *g_m2o* (many ground truth to one detected), *d_o2m* (one detected to many ground truth) and *d_m2o* (many detected to one ground truth) are calculated from MatchScore table (1) using acceptance threshold T_a and following the steps of [4].

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

$$FM = \frac{2DR \times RA}{DR + RA} \qquad (4)$$

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. Participating methods

Brief descriptions of the methods whose results were submitted to the competition are given in this section.

4.1 The BESUS method

This method was submitted by S. Mandal, S. Sen, S. Ray and A. Das of the CST Department of the Bengal Engineering and Science University in Shibpur, India in association with B. Chanda of the ECS Unit of the Indian Statistical Institute in Kolkata, India and is based on previous work reported in [6].

For text line segmentation the method is based on the following steps:

Step 1: Using morphological operations adjacently placed words in a line are coalesced.

Step 2: Mean line width (l_w) and mean gap (g_w) width are then computed.

Step 3: From this data, a set of line separators are proposed.

Step 4: If a line width or line gap as proposed by the line separators is substantially greater than the value as determined by l_w and g_w, the dataset is divided into two equal halves and all computations are repeated from step 3. This continues until the width of the strip thus produced is very small.

In this way, a number of line separators are computed and used to denote each text line with separate markers. From the text line information of a given image, word segmentation is accomplished as follows:

Step 1: For each text line in the data set calclulate the vertical projections.

Step 2: Calculate mean (*M*) and standard deviation (*S*) of inter word gap width from the vertical projection data.

Step 3: For each line, we denote a gap as inter word gap if it is greater than *M-S*.

Step 4: Words in a line are marked with separate markers.

4.2 The DUTH-ARLSA method

This method was submitted by M. Makridis and N. Nikolaou of the Democritus University of Thrace in Xanthi, Greece. It is based on an adaptive run length smoothing algorithm (ARLSA) and it can be summarized in three steps.

Step 1: Connected component analysis is performed, some connected components with great height are being split and, finally, two types of obstacles are detected: column and text obstacles. Column obstacles are vertical white run lengths (background pixels) whose length exceeds a proper threshold. Basically, column obstacles are detected on the left and right margins of most documents and between columns in cases of multi-column documents. Text obstacles are detected between text lines. They are small white run lengths whose length does not exceed a threshold.

Step 2: An adaptive run length smoothing algorithm (ARLSA) is performed. According to it, additionally smoothing constraints are set in regard to the geometrical properties of neighboring connected components. The replacement of background pixels with foreground pixels is performed when these constraints are satisfied and a column or a text obstacle is not encountered.

Step 3: For each text line the distances between the connected components are calculated. The mean value of these distances is estimated and used as a threshold for detecting word gaps.

4.3 The ILSP-LWSeg method

This method was submitted by V. Papavassiliou, T. Stafylakis, V. Katsouros and G. Carayannis of the Institute for Language and Speech Processing (ILSP) in Athens, Greece. Its working principles are as follows:

For the line segmentation task, the document image is first splitted into a sequence of consecutive non-overlapping vertical zones and the projection for each zone is then calculated. Each projection profile is then smoothed by employing a non-causal filtering on neighbouring projections. By differentiating both the original and the smoothed profiles and considering the extreme points (upper and lower boundary of a text line) the first order statistics of the height and foreground density of the “within” and “between” line classes are estimated. A Viterbi algorithm is applied using the foreground density as the generative probability for the two classes and the height to model the transitions between the two classes. Combining the results of Viterbi on the original and smoothed profiles, the candidate line separators for each zone are obtained. Starting from left to right, each line separator of a zone is merged with the closest line separator of the following zone and the remaining separators either define new lines or extend existing ones. Finally, strokes which significantly lie to two consecutive lines are split using the 4-n cross point of the skeleton that is nearest to the separator.

The word segmentation starts with ordering the strokes of each line using the horizontal coordinate of their centroid. The initial candidate word boundaries are estimated by the local minima of the horizontal projection. Assuming that each stroke cannot be a part of two adjacent words, for each candidate boundary point, the strokes that are closest from the left and the right of the boundary point are grouped into two separate sets. As a metric of separability between the two sets the negative logarithm of the objective function of a soft-margin SVM is adopted. Since a writer and style independent threshold cannot be defined, a global threshold for the whole page is estimated using unsupervised learning. The two classes (“between” and “within” words boundary point) are modeled by the EM algorithm (expectation-maximization algorithm), and the threshold is defined as the equiprobable point.

4.4 The PARC method

This method was submitted by J. Chen, D. Kletter, J. Lin, P. Sarkar, E. Saund and Y. Wang of the Perceptual Document Analysis Area of the Palo Alto Research Center in Palo Alto, USA. The method consists of the following distinct steps:

Step 1: Using image morphology, form line masks whose connected components represent individual text lines.

Step 2: Assign connected components of the original image to text lines according to simple proximity rules. In this process, multiline components are broken up while ambiguous assignments are left out.

Step 3: Measure features of horizontal distance among connected components in a line.

Step 4: Using a trained classifier, classify nearby connected components as "same-word" versus "different word".

Step 5: Apply transitive closure to group connected components into words.

Step 6: Assign ambiguous fragments remaining from Step 2 to nearby words, using simple proximity rules.

4.5 The UoA-HT method

This method was submitted by G. Louloudis of the University of Athens in Athens, Greece and is based on an efficient Hough Transform (HT) mapping.

The first step of the UoA-HT method calculates the connected components and the average character height of the document image. The connected component space is partitioned into three sub-domains depending on the connected components height and width. The connected components belonging to the first sub-domain (which contains all components that correspond to the majority of the characters) are partitioned to equally sized blocks. The gravity center of each block is calculated and fed to the Hough transform procedure so as to extract the first candidate lines. A final step is used to correct possible splitting, to detect text lines that the previous step did not reveal and, finally, to separate vertically connected characters and assign them to text lines.

The word segmentation stage takes as input the result of the text line detection stage. For each text line detected in the previous stage, the connected components are extracted by looking for non zero regions in the vertical projection profile. The Euclidean distance of adjacent connected components is calculated. A distance among two adjacent connected components is considered as inter-word if it is larger than a threshold otherwise it is considered as being an intra-word. The threshold is calculated by the median of horizontal white run-lengths in the row with the maximum number of black to white transitions multiplied by a predefined factor *f*.

5. Results

We evaluated the performance of the 5 algorithms for text line and word segmentation using equations (1)–(4) for all 80 test images with parameters $w_1 = 1$, $w_2 = 0.25$, $w_3 = 0.25$, $w_4 = 1$, $w_5 = 0.25$ and $w_6 = 0.25$ (one to one matches have higher significance). We used acceptance threshold T_a=95% for text line segmentation and T_a =90% for word segmentation. For the sake of clarity, we have also implemented two state-of-the-art techniques: RLSA and projection profiles. All evaluation results are shown in Table 1 while a graphical representation of the evaluation results is given in Fig. 2. As it can be observed, the ILSP-LWSeg method has an overall advantage with SM=94.2% while the UoA-HT method is second with SM=92.5%. The complete ranking list is:

1. ILSP-LWSeg method (SM=94.2%)
2. UoA-HT method (SM=92.5%)
3. PARC method (SM=85.4%)
4. BESUS method (SM=73.1%)
5. DUTH-ARLSA method (SM=70.7%)
6. PROJECTIONS (SM=61.6%)
7. RLSA (SM=60.1%)

More specifically, concerning text line segmentation, the ILSP-LWSeg method achieved the highest results with FM=97.1% while the UoA-HT method is second with FM=95.4%. The complete ranking list for text line segmentation is:

1. ILSP-LWSeg method (FM=97.1%)
2. UoA-HT method (FM=95.4%)
3. PARC method (FM=92.6%)
4. BESUS method (FM=83%)
5. DUTH-ARLSA method (FM=72%)
6. PROJECTIONS (FM=65.9%)
7. RLSA (FM=44.8%)

Table 1. Detailed evaluation results.

		M	o2o	g_o2m	g_m2o	d_o2m	d_m2o	DR	RA	FM	SM
BESUS	Text lines	1904	1494	9	151	72	21	**86,6%**	**79,7%**	**83,0%**	**73,1%**
	Words	19091	9114	327	6172	2449	823	**80,7%**	**52%**	**63,3%**	
DUTH-ARLSA	Text lines	1894	1214	149	227	107	354	**73,9%**	**70,2%**	**72,0%**	**70,7%**
	Words	16220	9100	394	5896	2440	954	**80,2%**	**61,3%**	**69,5%**	
ILSP-LWSeg	Text lines	1773	1713	5	34	17	10	**97,3%**	**97,0%**	**97,1%**	**94,2%**
	Words	13027	11732	303	834	378	819	**90,3%**	**92,4%**	**91,3%**	
PARC	Text lines	1756	1604	40	76	34	85	**92,2%**	**93,0%**	**92,6%**	**85,4%**
	Words	14965	10246	422	3482	1524	1088	**84,3%**	**72,8%**	**78,1%**	
UoA-HT	Text lines	1770	1674	14	54	27	28	**95,5%**	**95,4%**	**95,4%**	**92,5%**
	Words	13824	11794	263	1418	668	602	**91,7%**	**87,6%**	**89,6%**	
RLSA	Text lines	1877	632	264	346	122	757	**44,3%**	**45,4%**	**44,8%**	**60,1%**
	Words	13792	9566	639	2064	920	1633	**76,9%**	**74,0%**	**75,4%**	
PROJECTIONS	Text lines	1892	1109	91	344	155	192	**68,8%**	**63,2%**	**65,9%**	**61,6%**
	Words	17820	8048	391	4265	1694	963	**69,2%**	**48,9%**	**57,3%**	

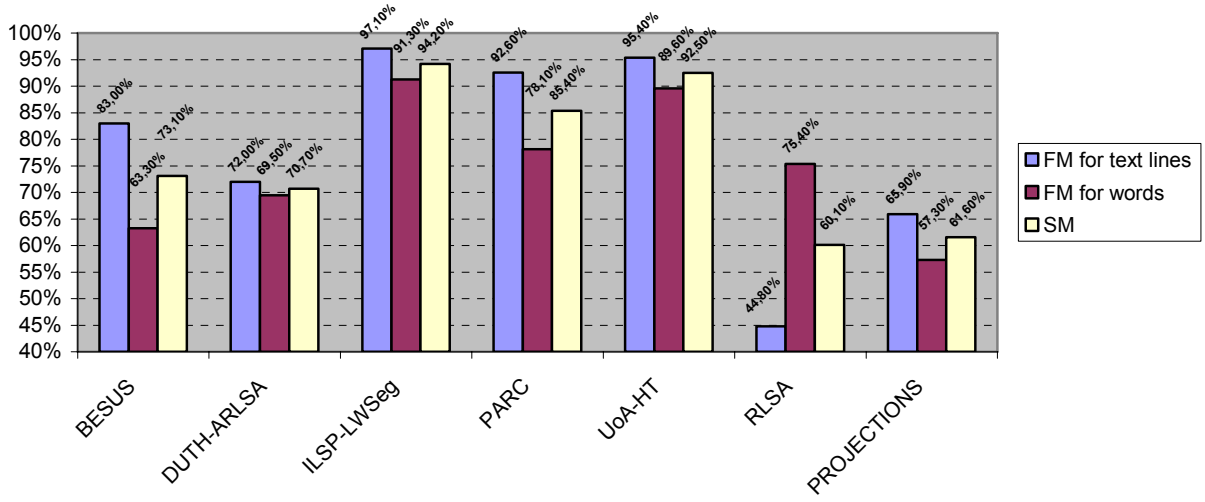


Figure 2. Graphical representation of the evaluation results

Concerning word segmentation, the ILSP-LWSeg method achieved the highest results with FM=91.3% while the UoA-HT method is second with FM=88.1%. The complete ranking list for word segmentation is:

1. ILSP-LWSeg method (FM=91.3%)
2. UoA-HT method (FM=89.6%)
3. PARC method (FM=78.1%)
4. RLSA (FM=75,4%)
5. DUTH-ARLSA method (FM=69,5%)
6. BESUS method (FM=63.3%)
7. PROJECTIONS (FM=57,3%)

6. Conclusions

The motivation of the ICDAR2007 handwriting segmentation contest was to use well established evaluation practices and procedures in order to record recent advances in off-line handwriting segmentation. In this paper, the results of the evaluation of five algorithms submitted by participants as well as of two state-of-the-art algorithms (RLSA and Projections) are presented. As it is shown in the evaluation results, the best performance for text line and word segmentation (global performance metric *SM*=94.2%) was achieved by the ILSP-LWSeg method of the Institute for Language and Speech Processing (ILSP) while in the second place (with *SM*=92.5%) is the UoA-HT method of the University of Athens and in the third place is the PARC method of the of the Perceptual Document Analysis Area of the Palo Alto Research Center. The state-of-the-art RLSA and Projections algorithms received scores of *SM*=60.1% and *SM*=61.6% respectively.

References

[1] A. Antonacopoulos, B. Gatos and D. Bridson, "ICDAR2005 Page Segmentation Competition", 8th Int. Conf. on Document Analysis and Recognition (ICDAR'05), 2005, pp. 75-79.

[2] A. Antonacopoulos, B. Gatos and D. Karatzas, "ICDAR 2003 Page Segmentation Competition", Proc. of the 7th Int. Conf. on Document Analysis and Recognition (ICDAR'03), 2003, pp. 688-692.

[3] B. Gatos, S. L. Mantzaris and A. Antonacopoulos, "First International Newspaper Segmentation Contest", Proc. of the 6th Int. Conf. on Document Analysis and Recognition (ICDAR'01), 2001, pp. 1190-1194.

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

[5] B.A. Yanikoglu, and L Vincent, "Pink Panther: a complete environment for ground-truthing and benchmarking document page segmentation", Pattern Recognition, volume 31, number 9, 1994, pp. 1191-1204.

[6] A.K. Das, A. Gupta and B. Chanda, "A fast algorithm for text line & word extraction from handwritten documents", Image Processing & Communications, 3, No. 1-2, 1997, pp. 85-94.