2013 12th International Conference on Document Analysis and Recognition

ICDAR2013 Handwriting Segmentation Contest

Nikolaos Stamatopoulos[1], Basilis Gatos[1], Georgios Louloudis[1], Umapada Pal[2] and Alireza Alaei[3]
[1]Computational Intelligence Laboratory, Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos", GR-153 10 Agia Paraskevi, Athens, Greece
{nstam, bgat, louloud}@iit.demokritos.gr
[2]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute
203 B. T. Road, Kolkata-700108, India
umapada@isical.ac.in
[3]Computer Science Laboratory, Universite Francois Rabelais
64 avenue Jean Portalis, 37200-Tours, France
alireza.alaei@univ-tours.fr

***Abstract* — This paper presents the results of the Handwriting Segmentation Contest that was organized in the context of the ICDAR2013. The general objective of the contest was to use well established evaluation practices and procedures to record recent advances in off-line handwriting segmentation. Two benchmarking datasets, one for text line and one for word segmentation, were created in order to test and compare all submitted algorithms as well as some state-of-the-art methods for handwritten document image segmentation in realistic circumstances. Handwritten document images were produced by many writers in two Latin based languages (English and Greek) and in one Indian language (Bangla, the second most popular language in India). These images were manually annotated in order to produce the ground truth which corresponds to the correct text line and word segmentation results. The datasets of previously organized contests (ICDAR2007, ICDAR2009 and ICFHR2010 Handwriting Segmentation Contests) along with a dataset of Bangla document images were used as training dataset. Eleven methods are submitted in this competition. A brief description of the submitted algorithms, the evaluation criteria and the segmentation results obtained from the submitted methods are also provided in this manuscript.**

Keywords- Handwritten Text Line Segmentation; Handwritten Word Segmentation; Performance Evaluation.

I. INTRODUCTION

Segmentation of a document image into its basic entities, namely, text lines and words, is considered as a non-trivial problem to solve in the field of handwritten document recognition. 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, etc. (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 image segmentation.

Following the successful organization of the ICDAR2007, ICDAR2009 and ICFHR2010 Handwriting Segmentation Contests [1-3], we organized the ICDAR2013 Handwriting Segmentation Contest to record recent advances in off-line handwriting segmentation. A major difference from the previous contests is that we extended the languages involved by including an Indian language (apart from Latin languages). This contest may provide a clear guideline for future research in this particular field of document image analysis.

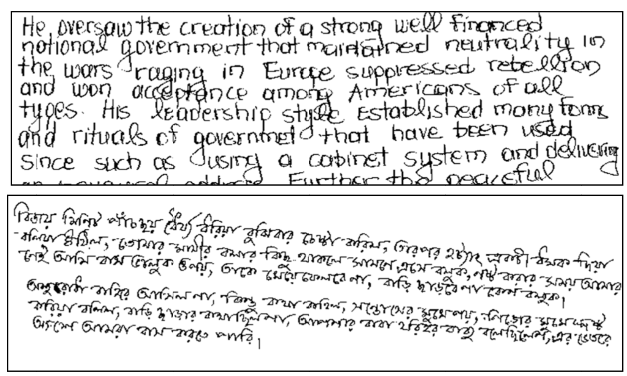
He oversaw the creation of a strong well financed
national government that maintained neutrality in
the wars raging in Europe suppressed rebellion
and won acceptance among Americans of all
types. His leadership style established many forms
and rituals of government that have been used
since such as using a cabinet system and delivering

Figure 1. Indicative portions of samples of the benchmarking dataset (English and Bangla).

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 image segmentation in realistic circumstances. Handwritten document images were produced with the help of several writers in English and Greek (Latin languages) and in Bangla (Indian language). The benchmarking datasets used in the previously organized contests (only the English and Greek parts) together with 50 document images from [4] were used for training. Concerning the evaluation stage, a well-established approach that was also employed by other document image segmentation contests was used.

The remainder of the paper is organized as follows. In Section II, the contest details and an overview of the datasets are described. In Section III, the performance evaluation method and metrics are detailed. A brief description of each participating method is provided in Section IV while the results of the competition are presented in Section V. Finally, some conclusions are drawn in Section VI.

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.283

II. THE CONTEST

The authors of candidate methods registered their interest in the contest and downloaded the training dataset (150 document images written in English and Greek as well as 50 images written in Bangla along with the associated ground truth) and the corresponding evaluation software [5]. At a next step, all the participants registered for the contest were asked 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 positive integer values each corresponding to a segmentation region. After the evaluation of all candidate methods, the benchmarking dataset (50 images written in English, 50 images written in Greek and 50 images written in Bangla) (see Fig.1) along with the evaluation software became publicly available [6].

The training and benchmarking datasets contain black & white handwritten document images produced by many writers. The corresponding document images do not include any non-text elements (lines, drawings, etc.). During the creation phase of the Latin part of the benchmarking dataset, 50 writers were asked to copy two samples of text in English and Greek language. For the Indian part, 50 document images with different content and sizes were considered.

III. PERFORMANCE EVALUATION

The method used to evaluate the performance of the submitted algorithms is based on counting the number of matches between the entities detected by the algorithm and the entities in the ground truth [7]. For the detection of matches, 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 points of set s. Table *MatchScore(i,j)* represents the matching results of the j ground truth region and the i result region:

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

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

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

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

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

A global performance metric SM for handwriting segmentation is extracted by calculating the average values of the FM metric for text line and word segmentation. The performance evaluation method is robust and well established since it has been used in other contests [1-3] and it depends only on the selection of the acceptance threshold T_a.

IV. METHODS AND PARTICIPANTS

Nine research groups participated in the competition with eleven different algorithms (two participants submitted two algorithms each). Nine submissions included both text line and word segmentation algorithms while two submissions included only a text line segmentation method. Brief descriptions of the methods are given in this section.

CUBS method: Submitted by Z. Shi, S. Setlur and V. Govindaraju from 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 [8, 9]. 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. At a next step, the distances between consecutive word primitives are computed using the convex hull distance. A bi-modal fitting is applied to find the threshold in determining the minimal word gap in the document image.

GOLESTAN method (two methods): Submitted by M. Ziaratban from the Electrical Engineering Department, Golestan University in Iran.

a. In the text line extraction algorithm, a handwritten text image is first filtered by a 2D Gaussian filter. The size and the standard deviation of the Gaussian filter as well as the block size are calculated for each text image, separately. The filtered image is then divided into a number of overlapped blocks. For each block, a local skew angle is estimated. The filtered block is binarized using an adaptive threshold and with respect to the estimated local skew angle. Binarized blocks are concatenated to get the overall path of text lines. Finally, the text lines are extracted by thinning the background of the path image. A similar approach is used to extract words from each text line. A detected text line is first filtered by a 2D Gaussian filter. At a next step, ascenders and descenders are then eliminated and an adaptive thresholding is used to determine the words.

b. Line segmentation method remains the same while for the word segmentation a 2D Gaussian filter is used in the same way without eliminating the ascenders and descenders.

INMC method: Submitted by J. Ryu and N.I. Cho from the INMC, Department of Electrical Engineering and Computer Science, Seoul National University, Korea and H.I. Koo from the Ajou University, Suwon, Korea. The line segmentation algorithm is based on an energy minimization framework considering the fitting errors of text lines and the distances between detected text lines [10]. However, the state-estimation was improved by performing over-segmentation at the initial stage. Therefore, unlike [10], the algorithm is able to handle cursive and Indian scripts where many graphemes are connected. The energy minimization algorithm is also improved by developing additional steps based on dynamic programming. Concerning the word segmentation, method [11] is modified in order to deal with the irregularity in handwriting documents. A text line is segmented into words using the statistical information of spacing in each text-line and then, based on the local statistical information of word segments, a refining is applied.

LRDE method: Submitted by E. Carlinet and T. Géraud from the EPITA Research and Development Laboratory (LRDE) in Le Kremlin-Bicetre, France. For text line segmentation, the inter-line spacing is first detected using a correlation measure of the projected histogram of the image on the y-axis. 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) whose kernel support depends on the inter-line spacing detected above. The morphological watershed transform is computed, leading into partitioning the image into regions. To obtain line segmentation, a simple merging procedure is applied on the region adjacency graph. Word segmentation relies on the text lines detected above to compute the inter-word spacing. The horizontal distances between each pair of adjacent connected component of a text line give the intra-word and inter-word spaces. A 2-means clustering allows setting a decision boundary between the two classes. At a next step, dilation is performed with a horizontal structuring element whose width depends on inter-word spacing detected above. Finally, an attribute morphological closing followed by a morphological watershed transform produces the final word segmentation result.

MSHK method: Submitted by L. Mengyang from the Department of Management Sciences, City University of Hong Kong. The text line segmentation algorithm is based on connected component analysis. The average width and height of connected components (CCs) are first estimated using statistical metrics methods. The CCs of normal size that are close to each other and almost at the same latitude are grouped into short text lines. At a next step, the previously detected text lines are merged into long text lines according to their direction, latitude and the intersections between them. Finally, the CCs with abnormal size are merged with the existing text lines by checking the neighborhood. Once the text lines are detected, the horizontal density of each text line is estimated and a closing operation is applied according to it. Finally, the average distance between adjacent words is calculated and is used to merge adjacent words whose distances are smaller than this value.

NUS method: Submitted by X. Zhang and C. L. Tan from the School of Computing at the National University of Singapore. For text line extraction, all small strokes and large connected components (CCs) are first removed and a skew correction method is applied. The possible locations of the text lines are detected using a seam carving algorithm. When constructing the energy accumulation matrix, the accumulative energies are normalized by their distance to the current position using only the newest W/2 energies, where W is the width of the image. Seams with an energy value smaller than a threshold are removed and for each remaining seam the CCs which are intersected with the seam are labeled with the same number. Finally, each unlabeled stroke is merged with the nearest CC and the image is rotated back to its original skew angle. Concerning the word segmentation, the small strokes and other floating strokes which are located above or below the main body of the text line are removed. The gap between every pair of consecutive CCs is calculated using soft margin SVM and the second most dominant of these gap metrics value is used as a threshold for word segmentation.

QATAR method (two methods): Submitted by A. Hassaine and S. Al Maadeed from the Qatar University.

a. First, the script of the handwritten document image is automatically detected using the features presented in [12]. Text line segmentation is then 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. Some lines might be split into several connected components which are subsequently merged using standard proximity rules trained separately for each script category. The word segmentation is performed by thresholding a smoothed version of a generalized chamfer distance in which the horizontal distance is slightly favored.

b. The second method is similar to the first one with the exception that it is trained on both the provided training dataset as well as the QUWI dataset [13].

CVC method (text line segmentation only): Submitted by D. Fernandez, F. Cruz, J. Llados, O.R. Terrades and A. Fornes from the Computer Vision Center, Universitat Autonoma de Barcelona in Spain. In this algorithm, the text line segmentation problem is formulated as finding the central path in the area between two consecutive text lines. This is solved as a graph traversal problem. A graph is constructed using the skeleton of the image. At a next step, a path-finding algorithm is used to find the best path to segment the text lines of the document.

IRISA method (text line segmentation only): Submitted by A. Lemaitre from the IRISA Laboratory, University of Rennes 2, France. The text line segmentation algorithm combines two levels of information: a blurred image and the extracted connected components. This method aims at imitating the human perceptive vision that combines two different points of view of a single image: i) a blurred global point of view and ii) a local precise point of view. On the one hand, the blurred image provides the position of text body in the parts of the image that contain a high density of writings. On the other hand, the analysis of connected components gives the position of text lines in large spaced handwriting or for large characters (like titles or uppercase). The blurred image is obtained by a recursive low-pass filter on columns, followed by a low-pass filter on rows. In this blurred image, we detect the significant holes of luminosity, which are grouped among the columns, depending on size and position criteria. This first step of analysis provides parts of segments of text lines. In the second step of analysis, the presence of connected components is used to locally extend, if necessary, the pieces of text lines that have were found previously. Thus, a local analysis of the alignments of connected components is used, taking into account the global organization of the page. Consequently, the body for each text line (position and thickness) is obtained. At a final step, each connected component is associated to the nearest text line, after having re-segmented the connected components that belong to several text lines.

V. Evaluation results

We evaluated the performance of all participating algorithms for text line and word segmentation using equations (1)–(3), the benchmarking dataset (150 images) [6] and the corresponding ground truth. The acceptance threshold 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 150 document images was 2649 and 23525, respectively. We have also applied three state of the art techniques: NCSR method [14], ILSP method [15] and TEI method [16]. NCSR method is based on Hough transform for text line segmentation and on the combination of the Euclidean and convex hull-based distance metrics for word segmentation. ILSP method makes use of the Viterbi algorithm and the objective function of a soft-margin linear SVM. Finally, TEI method is based on an improved shredding technique for text line segmentation. Concerning word segmentation, it is based on a Neural Network that combines various geometrical features extracted from the whole image as well as the gaps between connected components.

The evaluation results obtained from all the algorithms submitted to the contest as well as from the state of the art methods descripted above are presented in Table I, while graphical representations of them are also shown in Figs. 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 III). The GOLESTAN method outperforms all other methods in the overall ranking achieving *SM* = 94.17% (Fig. 2). The ranking list for the first four methods is as follows:

1. **GOLESTAN-a** (*SM*=94.17%)
2. **GOLESTAN-b** (*SM*=94.06%)
3. **INMC** (*SM*=93.96%)
4. **NUS** (*SM*=93.77%)

Considering only text line segmentation results, the INMC method achieved the best results with *FM* = 98.66% (Fig. 3). The ranking list for the first four text line segmentation methods is as follows:

1. **INMC** (*FM*=98.66%)
2. **NUS** (*FM*=98.41%)
3. **GOLESTAN-a** (*FM*=98.28%)
4. **CUBS** (*FM*=97.45%)

Based on the word segmentation results, the GOLESTAN method obtained the highest results with *FM* = 90.05% (Fig. 4). The first four word segmentation methodos obtained the highest results are listed in the following:

1. **GOLESTAN-a** (*FM*=90.05%)
2. **GOLESTAN-b** (*FM*=89.83%)
3. **NCSR (SoA)** (*FM*=89.62%)
4. **INMC** (*FM*=89.26%)

TABLE I. Detailed Evaluation Results

		M	*o2o*	*DR* (%)	*RA* (%)	*FM* (%)	*SM* (%)
CUBS	Lines	2677	2595	**97.96**	**96.94**	**97.45**	**92.41**
	Words	23782	20668	**87.86**	**86.91**	**87.38**	
GOLESTAN-a	Lines	2646	2602	**98.23**	**98.34**	**98.28**	**94.17**
	Words	23322	21093	**89.66**	**90.44**	**90.05**	
GOLESTAN-b	Lines	2646	2602	**98.23**	**98.34**	**98.28**	**94.06**
	Words	23400	21077	**89.59**	**90.07**	**89.83**	
INMC	Lines	2650	2614	**98.68**	**98.64**	**98.66**	**93.96**
	Words	22957	20745	**88.18**	**90.36**	**89.26**	
LRDE	Lines	2632	2568	**96.94**	**97.57**	**97.25**	**92.05**
	Words	23473	20408	**86.75**	**86.94**	**86.85**	
MSHK	Lines	2696	2428	**91.66**	**90.06**	**90.85**	**85.29**
	Words	21281	17863	**75.93**	**83.94**	**79.73**	
NUS	Lines	2645	2605	**98.34**	**98.49**	**98.41**	**93.77**
	Words	22547	20533	**87.28**	**91.07**	**89.13**	
QATAR-a	Lines	2626	2404	**90.75**	**91.55**	**91.15**	**88.36**
	Words	24966	20746	**88.19**	**83.10**	**85.57**	
QATAR-b	Lines	2609	2430	**91.73**	**93.14**	**92.43**	**88.25**
	Words	25693	20688	**87.94**	**80.52**	**84.07**	
CVC	Lines	2715	2418	**91.28**	**89.06**	**90.16**	**-**
IRISA	Lines	2674	2592	**97.85**	**96.93**	**97.39**	**-**
NCSR (SoA)	Lines	2646	2447	**92.37**	**92.48**	**92.43**	**91.02**
	Words	22834	20774	**88.31**	**90.98**	**89.62**	
ILSP (SoA)	Lines	2685	2546	**96.11**	**94.82**	**95.46**	**91.81**
	Words	23409	20686	**87.93**	**88.37**	**88.15**	
TEI (SoA)	Lines	2675	2590	**97.77**	**96.82**	**97.30**	**92.47**
	Words	23259	20503	**87.15**	**88.15**	**87.65**	

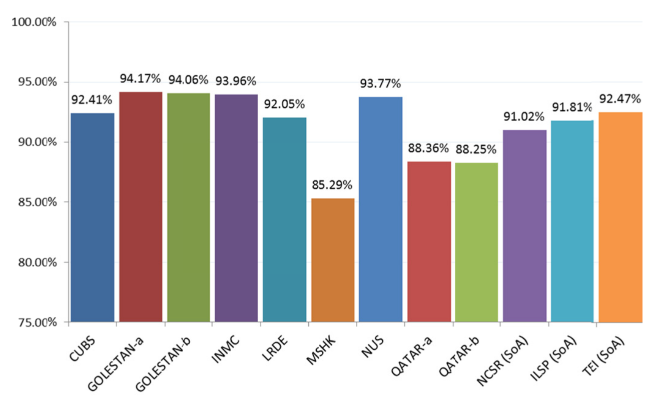

Figure 2. Overall evaluation performance for both text line and word segmentation.

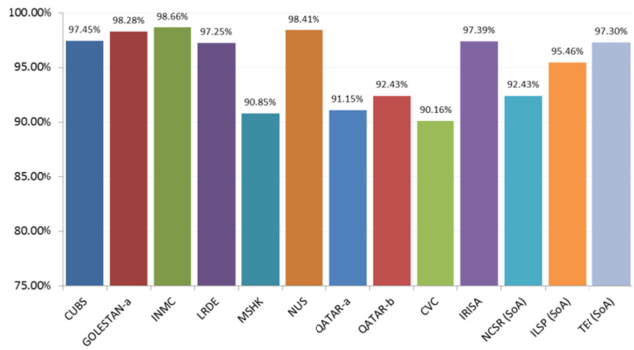

Figure 3. Evaluation performance for text line segmentation.

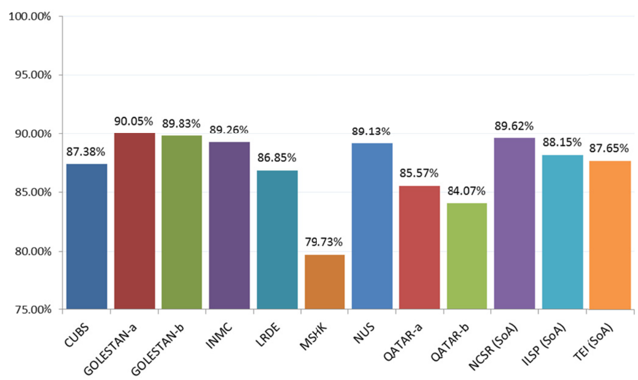

Figure 4. Evaluation performance for word segmentation.

After a careful analysis of the data presented in Table I we can stress that:

a. There is no significant deviation in the performance of the first four submitted methods since a global score between 93.77% to 94.17% is achieved.

b. The winning method (GOLESTAN) outperforms all other methods in the overall ranking as well as in the word segmentation stage. Moreover, it achieves the third best result at the text line segmentation stage.

c. The second method in the overall ranking (INMC) outpefrorms all other methods in the text line segmentation stage.

d. More than half of the submitted text line segmentation methods perform very well achieving a score above 97%. However, concerning word segmentation, the highest accuracy performed is 90.05% which implies that there exists a good potential for improvement.

e. TEI method achieved the best results in the overall ranking among the state of the art methods with $SM = 92.47\%$ and it was ranked fifth.

VI. Conclusions

The ICDAR2013 Handwriting Segmentation Contest was organized in order to record recent advances in off-line handwriting segmentation. As shown in the evaluation results section, the best results were obtained by the GOLESTAN method submitted by M. Ziaratban from the Golestan University in Iran with an overall global performance of 94.17% (for both text line and word segmentation) and a word segmentation performance of 90.05%. Considering only text line segmentation, the best result was obtained by the INMC method submitted by J. Ryu and N.I. Cho from the Seoul National University and H.I. Koo from the Ajou University in Korea and the performance was of 98.66%.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600707 - tranScriptorium.

References

[1] B. Gatos, A. Antonacopoulos and N. Stamatopoulos, “ICDAR2007 Handwriting Segmentation Contest”, 9th International Conference on Document Analysis and Recognition, pp. 1284-1288, 2007.

[2] B. Gatos, N. Stamatopoulos and G. Louloudis, “ICDAR2009 Handwriting Segmentation Contest”, 10th International Conference on Document Analysis and Recognition, pp. 1393-1397, 2009.

[3] B. Gatos, N. Stamatopoulos and G. Louloudis, “ICFHR 2010 Handwriting Segmentation Contest”, 12th International Conference on Frontiers in Handwriting Recognition, pp. 737-742, 2010.

[4] A. Alaei, U. Pal and P. Nagabhushan, “Dataset and ground truth for handwritten text in four different scripts”, International Journal of Pattern Recognition and Artificial Intelligence, 26, 2012.

[5] http://users.iit.demokritos.gr/~nstam/ICDAR2013HandSegmCont/TrainingToolkit/

[6] http://users.iit.demokritos.gr/~nstam/ICDAR2013HandSegmCont/BenchmarkingDataset/

[7] I. Phillips and A. Chhabra, “Empirical Performance Evaluation of Graphics Recognition Systems”, IEEE Transactions on Pattern Analysis and Machine Intellgence, 21(9), pp. 849-870, 1999.

[8] Z. Shi, S. Setlur and V. Govindaraju, “Text Extraction from Gray Scale Historical Document Images Using Adaptive Local Connectivity Map”, 8th International Conference on Document Analysis and Recognition, pp. 794- 798, 2005.

[9] Z. Shi, S. Setlur and V. Govindaraju, “A Steerable Directional Local Profile Technique for Extraction of Handwritten Arabic Text Lines”, 10th International Conference on Document Analysis and Recognition, pp. 176-180, 2009.

[10] H.I. Koo and N.I. Cho, “Text-Line Extraction in Handwritten Chinese Documents Based on an Energy Minimization Framework”, IEEE Transactions on Image Processing, 21(3), pp.1169-1175, 2012.

[11] H.I. Koo and D.H. Kim, “Scene Text Detection via Connected Component Clustering and Non-text Filtering”, IEEE Transactions on Image Processing, to appear, doi: 10.1109/TIP.2013.2249082.

[12] A. Hassaine, S. Al-Maadeed and A. Bouridane, “A set of geometrical features for writer identification”, Neural Information Processing, Springer Berlin Heidelberg, pp. 584-591, 2012.

[13] S. Al-Maadeed, W. Ayouby, A. Hassaine and J. Aljaam, “QUWI: An Arabic and English Handwriting Dataset for Offline Writer Identification”, International Conference on Frontiers in Handwriting Recognition, pp. 746 – 751, 2012.

[14] G. Louloudis, B. Gatos, I. Pratikakis and C. Halatsis, “Text line and word segmentation of handwritten documents”, Pattern Recognition, 42(12), pp. 3169-3183, 2009.

[15] T. Stafylakis, V. Papavassiliou, V. Katsouros and G. Carayannis, “Handwritten document image segmentation into text lines and words”, Pattern Recognition, 43(1), pp. 369-377, 2010.

[16] A. Nicolaou and B. Gatos, “Handwritten Text Line Segmentation by Shredding Text into its Lines”, 10th International Conference on Document Analysis and Recognition, pp. 626-630, 2009.