A perceptive method for handwritten text segmentation

Aurélie Lemaitre, Jean Camillerapp, Bertrand Coüasnon

A perceptive method for handwritten text segmentation

Aurélie Lemaitre[a], Jean Camillerapp[b] and Bertrand Coüasnon[b]

[a]Université de Rennes 2 - Irisa - UEB, Campus de Beaulieu, 35042 Rennes Cedex, France;
[b]INSA - Irisa - UEB, Campus de Beaulieu, 35043 Rennes Cedex, France

ABSTRACT

This paper presents a new method to address the problem of handwritten text segmentation into text lines and words. Thus, we propose a method based on the cooperation among points of view that enables the localization of the text lines in a low resolution image, and then to associate the pixels at a higher level of resolution. Thanks to the combination of levels of vision, we can detect overlapping characters and re-segment the connected components during the analysis. Then, we propose a segmentation of lines into words based on the cooperation among digital data and symbolic knowledge. The digital data are obtained from distances inside a Delaunay graph, which gives a precise distance between connected components, at the pixel level. We introduce structural rules in order to take into account some generic knowledge about the organization of a text page. This cooperation among information gives a bigger power of expression and ensures the global coherence of the recognition. We validate this work using the metrics and the database proposed for the segmentation contest of ICDAR 2009. Thus, we show that our method obtains very interesting results, compared to the other methods of the literature. More precisely, we are able to deal with slope and curvature, overlapping text lines and varied kinds of writings, which are the main difficulties met by the other methods.

Keywords: Structure analysis, handwritten text, text line segmentation, word segmentation

1. INTRODUCTION

In this paper, we address the problem of the segmentation of handwritten pages into text lines and words. It is still a difficult problem as the handwriting is not constrained (slope, varying curvature inside a document) and the writing varies a lot depending on the writer (thick or thin, with large or small space between words). And yet, it is an important step for the recognition of content: the segmentation into words can significantly help the process of recognition. In this domain, the objective is to propose generic methods, which are able to deal with various kinds of unconstrained writings and various alphabets.

In this paper, we propose a new way to address this problem, thanks to the principles of perceptive vision. The perceptive vision consists of combining several levels of resolution of the images and use the saliency of structural elements. We have previously proposed to localize text lines thanks to a line segment extractor.[1] In this paper, we propose now to assign each pixel to the text lines they belong to. Indeed, the use of perceptive vision is an efficient way to deal with slope and curvature, overlapping text lines and varying styles of writers.

Concerning the segmentation into words, we propose to combine two aspects: some digital data and some structural rules. Thus, we propose to combine inter/intra word distances with a structural description of the structure of a page of text.

The paper is organized as follows. In section 2, we present the related work and show the originality of our approach. Then, we present our method for both line segmentation (section 3) and word segmentation (section 4). In section 5, we detail our implementation. At last, we show the efficiency of our method thanks to an evaluation on the database of ICDAR'2009 segmentation contest and the comparison with other methods of the literature. The results show that our method is particularly suitable to detect overlapping text lines and to deal with various kinds of writings.

Further author information: (Send correspondence to A. Lemaitre)
E-mail: aurelie.lemaitre@irisa.fr

2. RELATED WORK

The topic of handwritten text segmentation has been widely studied during the last years and various kinds of methods have been proposed.

In,[2] Louloudis *et al.* propose a complete state of the art on this topic. First, we can notice that the methods that are used for printed text analysis, such as projections, are not suitable for handwritten documents, due to the irregularity of the writings, the slope and the curvature of the text lines, and the overlapping between text lines. Thus, various works have been proposed that separate the two steps of text line segmentation and word segmentation.

For example, Li *et al.* present in[3] an interesting method for script independent text line segmentation but do not address in this paper the problem of word segmentation.

Most of the works on text and word segmentation have been compared during a contest at the last ICDAR 2009.[4] We propose to synthesize in table 1 the different techniques that are used by the participants of the contest.

Authors	Text line segmentation	Word segmentation
Yin, Liu[5]	Selection of edges in a Minimum Spanning Tree	Extraction of 11 characteristics about the distances between connected components and classification with a SVM
Shi, Setlur, Govindaraju[6]	Adaptive local connectivity map	Convex hull distance and threshold based on mean and variance
Rivest-Hénault, Cheriet[7]	Text smearing and morphological operators	Text smearing and morphological operators
Papavassiliou, Stafylakis, Katsouros, Caryannis[8]	Use of Viterbi algorithm to detect horizontal separators	Use o the negative logarithm of the objective function of a SVM
Khandelwal, Choudhury, Sarkar, Basu, Nasipuri, Das[9]	Analysis of features of the connected components in a given neighborhood	Difference in intra-word and inter-word distances
Louloudis, Gatos, Pratikakis, Halatsis[2]	Hough transform and adapted post processing	Combination of 2 distance metrics, and threshold obtained thanks to Gaussian mixtures

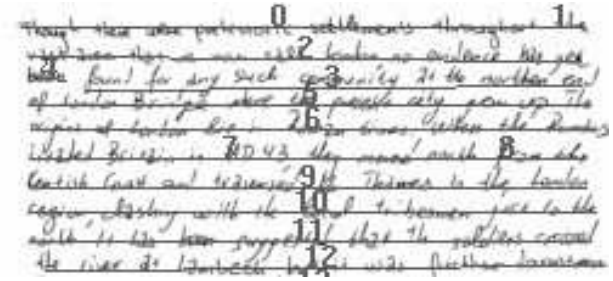
Table 1. Synthesis of the latest main methods of handwritten text segmentation

Among all these methods, we can pick out the works of Papavassiliou *et al.*[8] and the method of Shi *et al.*[6] that have obtained the best results for the segmentation contest of ICDAR 2009.[4] We can also mention that the work of Papavassiliou *et al.*[8] is one of the only works that is particularly efficient for both text line and word segmentation. These authors explain that their remaining difficulties occur when the writings are very segmented.

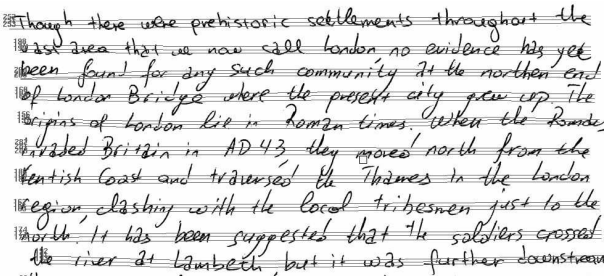
In order to be able to deal with any kind of writing, we have proposed in a previous work[1] to combine several levels of perception of the image. Thus, we consider that having a global vision of the document (at low resolution) enables the detection of text line position, as if they were line segments. Then, the vision at high resolution enables the confirmation of the presence of text lines. However, in this previous work, we were just giving the localization of the text lines without assigning each pixel to each text line. The novelty of our work in this paper is to assign each pixel of the image to a given text line and a given word.

3. OUR METHOD FOR TEXT LINE SEGMENTATION

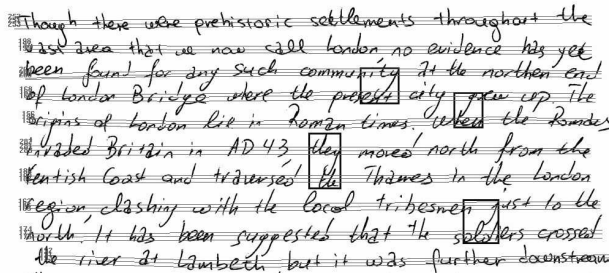
The segmentation into text lines is realized in four steps of analysis, detailed below: global localization of text lines as line segment at low resolution (figure 1(a)), adjustment of the position of the baselines at high resolution (figure 1(b)), detection of overlapping connected components and re-segmentation (figure 1(c)), assignment of each connected component to the final text lines (figure 1(d)).



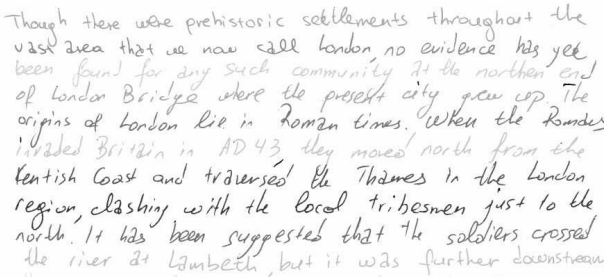
(a) Extraction of line segments at low resolution (reported for an easy look in a higher resolution)



(b) Adjustment of the position of baselines



(c) Detection of overlapping connected components



(d) Final segmentation (various grey levels)

Figure 1. Steps of our method of segmentation into text lines (image 014 of ICDAR base)

3.1 Extraction of line segments

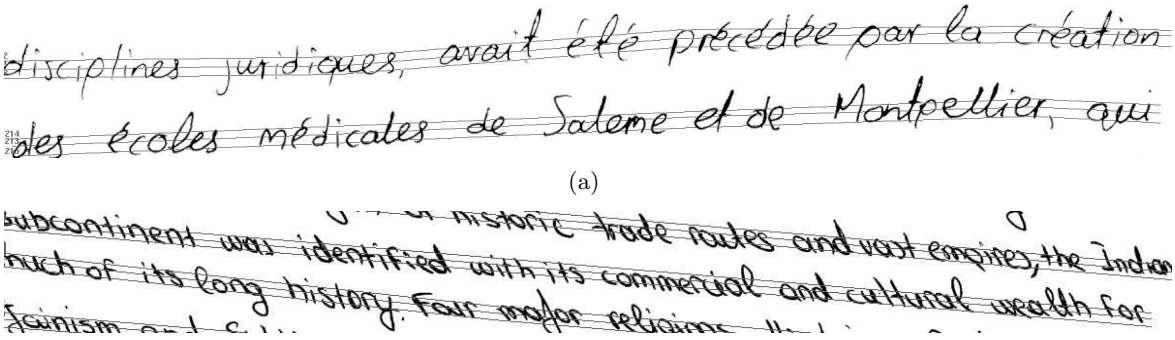
At low resolution, the text lines appear as line segments. Indeed, at low resolution, we apply a line segment extractor based on Kalman filtering, as we described in.[1] This robust operator is able to follow line segments, even if they are dotted (as text lines), curved or irregular. At the end of this step, the system emits some hypotheses on the presence of text lines that will be confirmed using high resolution analysis. This method is particularly suitable as it enables the prediction of text line position, whatever their kind of writing, thickness, slope or curvature.

3.2 Positioning of baselines

In this step, we use the global vision of the text lines in order to find the precise position of the body of the words. Thus, we have presented in[10] our operator to position the baselines of the writing, using the global context of the text line. This operator is based on the global detection of a zone of interest in the image. Then, it locally adjusts the position of the baseline, taking into account the presence of upper and lower black pixels. In this work, we find the position of three baselines: at the top, at the bottom and at the middle of the text body (figure 2). The interest of this method is to precisely obtain the position of the body of the words, even in the case of slope and curvature. This is very interesting for a process of word recognition.

3.3 Detection of overlapping connected components

The third part of the work consists of allotting each connected component to the associated text line. For this purpose, we study the position of each connected component, relative to the baselines that have been obtained at the previous step. This analysis makes it possible to detect conflicting connected components when they cross the baselines of several text lines. In that case, our method launches a re-segmentation of the connected components. We use the global perception of baselines to define the point where the connected component should be cut. In this first version of our work, our method is very simple: we segment the connected component at the average position of the upper and the lower middle baselines (figure 3). This method is not very precise, but it is enough for the proposed application. We are planning to improve this method in future work.



(a)

(b)

Figure 2. Positioning of three baselines that delimit the text body, following the global slope and curvature

(a) One connected component that overlaps two text lines: computation of the average position of the upper and the lower middle baselines

(b) Result of the segmentation: two connected components

Figure 3. Re-segmentation of a connected component

3.4 Assignment of each connected component to the final text lines

Once every connected component has been re-segmented if necessary, we associate each one with the nearest text line, taking into account the local position of the baselines.

4. OUR METHOD FOR WORD SEGMENTATION

When the segmentation into text lines has been done, we can realize a segmentation into words. We propose to combine two kinds of information: digital data and structural knowledge.

4.1 Digital data

The digital data is a distance threshold between connected components.

We assume that the writings are regular inside of a same page. Thus, we compute the distance threshold for each page. For this purpose, we calculate all the distances between neighbor connected components, inside of the whole page. We choose the distance based on the Delaunay graph (figure 4). Thus, we compute a precise distance between the pixels of the connected components,[11] instead of using a distance between bounding boxes.

Figure 4. Delaunay graph for distance computation: the vertexes give the distance between pixels of the neighbor connected components

We obtain a list of distance, from which we remove the `nbl` larger distances, where `nbl` is the number of text lines. Indeed, we consider that the too big distances will disturb the computation of the threshold. With the remaining distances, we compute a k-mean (k=2) to separate the inter and intra word distances. We obtain a threshold that will be used to group together the connected components of the same words for all the page. An example of result is presented in figure 5.

Figure 5. Words that are found by the application of the inter-word/intra-word distance threshold between connected components

4.2 Structural knowledge

The originality of our approach is that we enrich the digital threshold between connected components by using some structural knowledge. For example, in our ground truth, the punctuation signs are always associated with the previous words. However, this may cause troubles when the writer makes a too large space between a word and the following sign (dot, coma ...): two separate words are computed. Consequently, we have proposed to introduce some knowledge in a post-process that groups together, when necessary, a word and the following punctuation. This is illustrated in figure 6: the words in black have been associated with the following dots and comas.

Figure 6. Introduction of structural knowledge: the punctuation is associated with the previous word (black boxes), instead of making two separate words

This introduction of structural knowledge has been realized really simply as we work in the context of a grammatical generic method, DMOS-P.

5. IMPLEMENTATION WITH DMOS-P METHOD

5.1 Presentation of the method

As an example of implementation of our work, we present how we have implemented our process of segmentation using an existing method for document structure recognition, DMOS-P (Description and Modification of the Segmentation with Perceptive vision).[12]

This method is made of a bidimensionnal grammatical language, EPF (Enhanced Position Formalism), which enables a physical and a logical description of the structure of documents. For each kind of document to study, the user builds a grammatical description that consist in explaining the relative position of each structural element in the image. The terminals of the grammatical analysis are the connected components and the line segments extracted from the image. The terminals are organized into perceptive layers, that enable the cooperation between various resolution levels of the images. When a description has been realized in EPF, the associate parser is automatically produced by a compilation step.

Thanks to this EPF formalism, the knowledge is separated to the system, and the digital level is entirely guided by the symbolical description. Thus, this method is generic and can be applied on any kind of document. It has been validated on various kinds of documents: musical scores, mathematical formulas, military forms and at a large scale on more than 500,000 document pages.[13]

5.2 Our implementation

For our application, we have realized a generic grammatical description of the organization of a page of text into text lines and words. We use two levels of perception of the image, that is to say two perceptive layers that are made of:

- the line segments that are extracted in a low resolution image (dimensions divided by 16) thanks to our Kalman based extractor (called `LowResolution` layer),
- the connected components that are extracted in the initial image, after a binarization step (called `HighResolution` layer).

These two layers are the base for the grammatical description of the text lines and words.

A text line is described as a line segment at low resolution and a succession of connected components, in the same place, at high resolution. This rule is expressed as follows in the EPF language *:

```
textLine ::=
    USE_LAYER(LowResolution) FOR(oneLineSegment) &&
    AT(overLineSegment)&&
    USE_lAYER(HighResolution) FOR(setOfConnectedComponents).
```

Once every text line has been built, we can segment them into words, using the fact that a text line is a succession of words. This is expressed by arecursive rule:

```
textLineInWords ::=
    word &&
    AT(rightWord)&&
    setOfWords.
```

where a word is described as a succession of connected components, that are close (with a lower distance than the inter-word threshold `T` that is computed for each page) :

```
word ::=
    connectedComponent &&
    AT(closeEnough T)&&
    setOfConnectedComponents.
```

The main interest of this method is that it can include easily structural knowledge such as the rule presented in section 4.2. Indeed, thanks to EPF language, we can easily add another description of a word that take into account the punctuation (dots, coma). This rule is:

```
word ::=
    word &&
    AT(rightWord)&&
    punctuationSign.
```

where a `punctuationSign` is a small isolated connected component.

As a conclusion, the use of DMOS-P method is very interesting as it easily enables the introduction of structural knowledge, such as the rule concerning the punctuation presented above. Moreover, as this method is generic, it enables the use of the result of text line segmentation for the analysis of more complex documents. For example, we have applied the results of the segmentation for the analysis of incoming mail and on text pages as we will show in the results.

*AT is the position operator and && is the concatenation operator

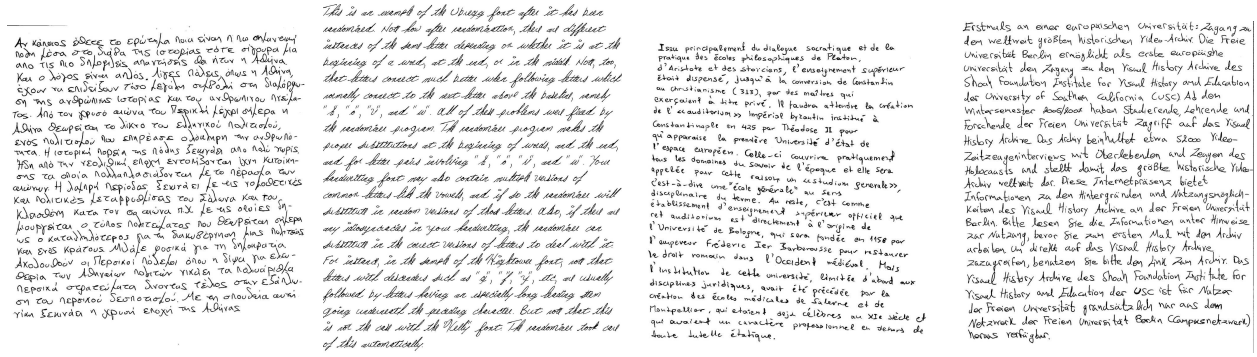
6. APPLICATION

In order to validate the interest of our method, we present our results on the international database proposed for the segmentation contest of ICDAR'2009.[4]

6.1 Data base

This database is made of handwritten pages, written by different writers who had to copy a given text. These pages are written in four languages: English, French, German and Greek.

We consider two sets: a training set made of 100 images, and a test set made of 200 images. For each set, we were given a ground truth that associates each pixel to its line and to its word. Some example of images are given on figure 7.



(a) (b) (c) (d)

Figure 7. Example of studied images

6.2 Metrics from ICDAR'2009 segmentation contest

The evaluation of the performance is realized by counting the number of matches between the entities (lines, words) detected by the algorithm and the entities in the ground-truth.[4] Let N, the number of entities in the ground-truth. M is the number of entities that are retrieved by the analysis. Then, we calculate the number $o2o$ of one to one matching entities between N and M. Two entities (lines, words) are considered as matching if they have more than a threshold S of common pixels. In the segmentation contest, this threshold S is 95% for lines and 90% for words. The detection rate DR and the recognition accuracy RA are defined as follows.

$$DR = \frac{o2o}{N}, RA = \frac{020}{M}$$

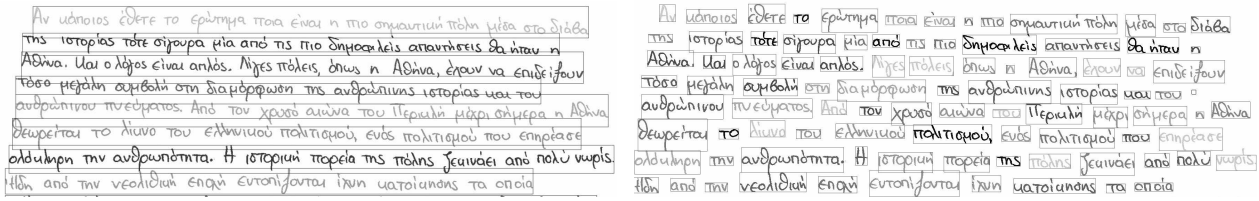
The performance metric FM is computed by combination of the detection rate and the recognition accuracy.

$$FM = \frac{2 * DR * RA}{DR + RA}$$

6.3 Results

We apply our method for line and word segmentation on both the training set and the test set. An example of segmentation is presented in figure 8. The obtained statistical results are presented in table 2. We obtain a very good result for line segmentation, with a performance FM of 99.25% on the test set, which means that 99.25% of text lines have more than 95% correct pixels. Concerning words, our algorithm obtains a performance of 94.20%, which means that 94.20% of words have more than 90% correct pixels. We can notice that the detection rate, DR, and the accuracy, RA, are similar for each of our processes.

The results that we have obtained on the test set enable the comparison of our approach with the methods that have been proposed in ICDAR'2009 contest.[4] We present this comparison on table 3. We can notice that our method obtains the second place for both words and lines. Our results are very close to the first ones (about



(a) Text line segmentation (b) Word segmentation

Figure 8. Result of the segmentation, represented with various grey levels and bounding boxes (image 03 of ICDAR base)

	Training set		Test set of ICDAR'2009 contest	
	Lines	Words	Lines	Words
Nb images	100	10	200	200
N	2242	1957	4034	29717
M	2244	1930	4032	29663
o2o	2199	1757	4003	27969
DR	98.08%	89.78%	99.23%	94.11%
RA	97.99%	91.03%	99.28%	94.28%
FM	98.03%	90.40%	**99.25**%	**94.20**%

Table 2. Detailed results obtained by our method

0.5%). As we have seen in table 2, our results vary depending on the studied database (from 90.4% for words on the training database to 94.20% on test database). Consequently, we can consider that our results are similar to the best methods of ICDAR'2009 contest, and that it is necessary to use a bigger database in order to evaluate more precisely the difference of performance between methods. We can also notice that our method obtains good results for both line and word segmentation, which is only the case for the method of ILSP-LWseg-09.[8] The interest of our work is that it properly deals with the varied kinds of writings, even with very segmented characters, contrary to Papavassiliou *et al.*[8] that meets difficulties with segmented characters.

Author	Method	FM lines	FM words
Yin *et al.*	CASIA-MSTSeg[5]	95.69%	84.85%
Hassane *et al.*	CMM	98.42%	88.91%
Shi *et al.*	CUBS[6]	**99.53%**	86.96%
Rivest-Hénault *et al.*	ETS[7]	86.67%	84.93%
Papavassiliou *et al.*	ILSP-LWseg-09[8]	99.05%	**94.74**%
Sarkar *et al.*	Jadavpur Univ[9]	87.34%	82.74%
Geraud *et al.*	LRDE[14]	92.25%	83.92%
Lu *et al.*	PAIS	98.52%	90.54%
	Proposed method	**99.25%**	**94.20%**

Table 3. Comparison of the performance (FM) with results obtained on the set of ICDAR'2009 contest[4]

We will soon obtain other results as we have taken part to the ICFHR'2010 segmentation contest, which results will be published on November 2010. This new competition will make possible another comparison with the other methods.

7. CONCLUSION

In this paper, we have proposed a new method for text line and word segmentation.

For the detection of text lines, we use a method based on perceptive vision. Thus, the global perception of text lines as line segment is a way to deal with varying writings, overlapping characters, slope and curvature of text lines. Then, the use of a high resolution analysis enables a re-segmentation of connected components and a precise assignment of each pixel to each text line.

Concerning the segmentation into words, we propose to combine digital data and symbolical knowledge. Thus, we use distances obtained from Delaunay graph, in order to precisely compute an inter/intra word distance threshold, at pixel level. Then, the novelty consists of combining this digital information with structural knowledge about the organisation of the page of text, such as the localization of the punctuation.

We have evaluated our method in an international context, based on ICDAR'2009 contest. We have shown that our method obtains results that are similar to the best methods. More precisely, we propose one of the two methods that are able to properly segment both lines and words. Our method is able to deal with overlapping text lines and varied kinds of writings, which are the difficulties met by other methods. We now wait for the results of ICFHR'2010 contest, in order to improve our method. More precisely, in the future work, we are planning to use some more elaborate techniques for the re-segmentation of connected components and for the classification of intra and inter word distances.

REFERENCES

[1] Lemaitre, A. and Camillerapp, J., "Text line extraction in handwritten document with Kalman filter applied on low resolution image," in [*IEEE Workshop - Document Image Analysis for Libraries (DIAL'06)*], 38–45 (2006).

[2] Louloudis, G., Gatos, B., Pratikakis, I., and Halatsis, C., "Text line and word segmentation of handwritten documents," *Pattern Recognition* **42**, 3169–3183, (Dec. 2009).

[3] Li, Y., Zheng, Y. F., Doermann, D., and Jaeger, S., "Script-independent text line segmentation in freestyle handwritten documents," *IEEE Trans. Pattern Analysis and Machine Intelligence* **30**, 1313–1329 (Aug. 2008).

[4] Gatos, B., Stamatopoulos, N., and Louloudis, G., "ICDAR 2009 handwriting segmentation contest," in [*ICDAR*], 1393–1397 (2009).

[5] Yin, F. and Liu, C. L., "Handwritten chinese text line segmentation by clustering with distance metric learning," *Pattern Recognition* **42**, 3146–3157, (Dec. 2009).

[6] Shi, Z., Setlur, S., and Govindaraju, V., "A steerable directional local profile technique for extraction of handwritten arabic text lines," in [*ICDAR*], 176–180 (2009).

[7] Henault, D. R. and Cheriet, M., "Image segmentation using level set and local linear approximations," in [*ICIAR*], 234–245 (2007).

[8] Papavassiliou, V., Stafylakis, T., Katsouros, V., and Carayannis, G., "Handwritten document image segmentation into text lines and words," *Pattern Recognition* **43**, 369–377, (Jan. 2010).

[9] Khandelwal, A., Choudhury, P., Sarkar, R., Basu, S., Nasipuri, M., and Das, N., "Text line segmentation for unconstrained handwritten document images using neighborhood connected component analysis," in [*PReMI*], Chaudhury, S., Mitra, S., Murthy, C. A., Sastry, P. S., and Pal, S. K., eds., *Lecture Notes in Computer Science* **5909**, 369–374, Springer (2009).

[10] Lemaitre, A., Camillerapp, J., and Coüasnon, B., "Multi-script baseline detection using perceptive vision," in [*14th Biennial Conference of the International Graphonomics Society (IGS 2009)*], (2009).

[11] Lemaitre, A., Coüasnon, B., and Leplumey, I., "Using a neighbourhood graph based on Vorono tessellation with DMOS, a generic method for structured document recognition," in [*Graphics Recognition - Ten years review and future perspectives, Sixth IAPR International Workshop GREC 2005, Revised Selected Papers*], **LNCS 3926**, 267–278 (2006).

[12] Lemaitre, A., Camillerapp, J., and Coüasnon, B., "Interest of perceptive vision for document structure analysis," in [*Human Vision and Electronic Imaging XV*], (January 2010).

[13] Coüasnon, B., "DMOS: A generic document recognition method to application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems," in [*Proceedings of International Conference on Document Analysis and Recognition (ICDAR'01)*], 215–220 (2001).

[14] Geraud, T., "http://www.lrde.epita.fr/cgi-bin/twiki/view/olena/moduleicdar."