HAL
open science

Use of Perceptive Vision for Ruling Recognition in Ancient Documents

Aurélie Lemaitre, Bertrand Coüasnon, Jean Camillerapp

To cite this version:

Aurélie Lemaitre, Bertrand Coüasnon, Jean Camillerapp. Use of Perceptive Vision for Ruling Recognition in Ancient Documents. Recent advances in graphic recognition, Springer Berlin/heildeberg, pp.1-11, 2010, LNCS 6020. hal-00545050

HAL Id: hal-00545050

https://hal.science/hal-00545050v1

Submitted on 9 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Use of Perceptive Vision for Ruling Recognition in Ancient Documents

Aurélie Lemaitre[1], Bertrand Coüasnon[2] and Jean Camillerapp[2]

[1] Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes
[2] INSA, Avenue des Buttes de Coësmes, F-35043 Rennes
UMR IRISA, Campus de Beaulieu, F-35042 Rennes
Université Européenne de Bretagne, France

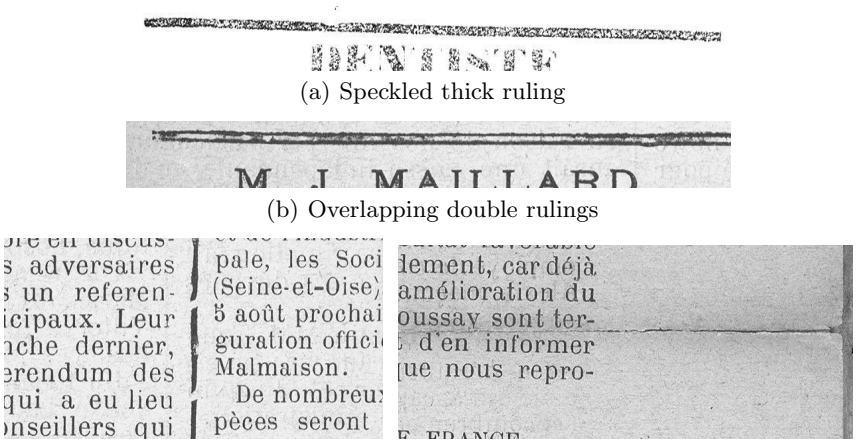
Abstract. Rulings are graphical primitives that are essential for document structure recognition. However in the case of ancient documents, bad printing techniques or bad conditions of conservation induce problems for their efficient recognition. Consequently, usual line segment extractors are not powerful enough to properly extract all the rulings of a heterogeneous document. In this paper, we propose a new method for ruling recognition, based on perceptive vision: we show that combining several levels of vision improves ruling recognition. Thus, it is possible to put forward hypothesis on the nature of the rulings at a given resolution, and to confirm or infirm their presence and find their exact position at higher resolutions.
We propose an original strategy of cooperation between resolutions and present tools to set up a correspondence between the elements extracted at each resolution. We validate this approach on images of ancient newspaper pages (dated between 1848 and 1944). We also propose to use the extracted rulings for the structure analysis of newspaper pages. We show that using more reliable extracted rulings simplifies and improves document structure recognition.

1 Introduction

Rulings are a base for the analysis of strongly structured documents like forms, tables or newspapers [3]. However, in the case of ancient or damaged documents, the recognition of rulings is complex: old printing techniques may produce irregular, speckled or dashed lines; bad method of conservation causes smearing ink and folding of the paper; the digitalization step can cause skew and curvature. Some examples of rulings that are difficult to detect in ancient documents are presented on figure 1.

The classical methods of the literature, based on projection or Hough transform are not always convenient: for example they do not easily deal with curvature and skew, and they use global parameters for a whole page. Consequently, other methods have been proposed in the literature. Gatos *et al.* present in [2] a specific process for the extraction of lines, based on gray scale transformation of the binary image. However, they require an a priori known length and width

in order to determine the pixels that belong to the final ruling. Hadjar *et al.* in [4] try to detect discontinuous rulings, with an approach based on connected components. Nevertheless, they need to know a maximum distance to group components into a single ruling. A threshold is also required in the work of Liu *et al.* presented in [8]. In all these methods dedicated to damaged documents, a strong a priori knowledge is used to answer the problem of the variation of the thickness of the rulings in the same document, and to deal with over segmented rulings. Xi *et al.* [9] propose a method based on curvelets, using different resolutions of an image. However, their method is not able to deal with curvature or with slope bigger than 2 degrees, that frequently occurs in digitized documents.



(a) Speckled thick ruling

(b) Overlapping double rulings

(c) Dashed ruling

(d) Line due to a tear

Fig. 1. Examples of rulings that are difficult to detect in ancient documents

Hori and Doerman present in [5] a multi-level analysis for form recognition. We follow this idea in order to solve more generally the problem of ruling recognition. We propose a new approach based on a mechanism used by the human eye: the perceptive vision. Indeed, in order to understand a document, our brain combines the visions at various levels of perception. In the case of rulings, we show that combining the analysis at various levels of resolution makes more accurate their extraction. Moreover, this method is able to extract rulings whatever their thickness, and even if they are damaged, without dedicated knowledge on their nature.

In this paper, we first present intuitively how using several levels of perception of an image can improve ruling recognition. It enables to deduce the perceptive mechanism that we implemented for ruling recognition. In section 4, we show the interest of the perceptive vision for ruling recognition with experiments on

old newspapers. At last, thanks to the good obtained results, we propose an application for newspaper structure recognition.

2 Intuitive approach

For the perceptive vision of rulings, we choose to combine three levels of perception. Indeed, we have experimentally noticed that using three levels of perception enables to obtain differences between visions that are significant but not too important. Thus, we build a multiresolution pyramid, by recursive low-pass filtering and sub-sampling. It recursively divides the dimensions of the images by 4.

On the three obtained resolutions, we apply an efficient line segment extractor, based on Kalman filtering [7]. Since our visual perception of the document varies, the detected line segments can differ depending on the resolution.

High resolution (initial image, ex: 300 dpi) When watching a document with a detailed local vision, that is to say at high resolution, we can see: pieces of double rulings (figure 1(b)), pieces of thin rulings (figure 1(c)) and elements due to noise. Thick rulings may not be perceptible because the white speckle takes too much importance.

Medium resolution (image divided by 4, ex: 75 dpi) When watching this document at medium resolution, we can see: pieces of double rulings (figure 1(b)), pieces of thin rulings (figure 1(c)), pieces of thick rulings (figure 1(a)), elements due to noise (figure 1(d)), and straight parts of capital letters.

Low resolution (image divided by 16, ex: 20 dpi) When watching this document with a global vision, that is to say at low resolution, we can see: thick rulings (figure 1(a)) that are dark enough, double rulings that appear as a single ruling (figure 1(b)), and some bold text lines that also appear as line segments. Thin rulings are too bright to appear at this level.

These different perceptions are gathered in table 1. We distinguish two kinds of rulings: the "true" rulings that are real structural elements (thick, thin, and double ruling) and the "false" ones that are text lines, pieces of characters and more globally noise.

Resolution	Low	Medium	High
Thin ruling	No	Yes	Yes
Thick ruling	Yes	Yes	No
Double ruling	Yes	Yes	Yes
Text lines	Yes	No	No
Characters	No	Yes	No
Noise	No	No	Yes

Table 1. Perception of line segments at each resolution

The table 1 shows that the perception of each element at the three resolutions can determine the kind of studied ruling. Thus, we deduce a strategy of analysis to recognize each kind of ruling. We now present this mechanism in details.

3 Principles of implementation of the perceptive vision of rulings

In this part, we present the elements that are required to implement a perceptive method of rulling. We present our strategy for combining the line segments that are detected at different resolutions. Then, we explain how we have treated two essential points that are the correspondance between resolutions, and the precise adjustment of the position of the rulings.

3.1 Strategy of perceptive cooperation

The perceptive strategy consists in putting forward hypothesis on the presence of a ruling, based on the perception at lower resolutions. This hypothesis may be confirmed or infirmed by the presence of elements on upper resolutions. This strategy is presented on figure 2.

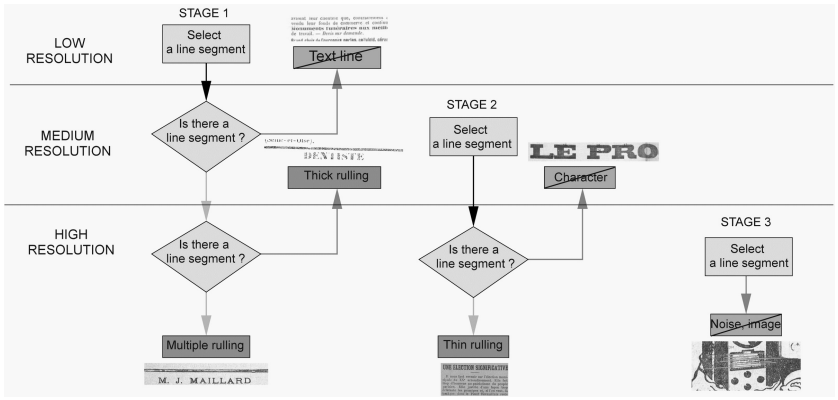


Fig. 2. Strategy for the combination of visions

The full analysis is realized in three stages: first the study of low resolution lines and their presence at highest resolutions (medium and high), then the study of the remaining medium resolution lines and their presence at high resolution, and at last the remaining line segments at high resolution.

First, we select a line segment at low resolution. If it has not any corresponding element at medium resolution, it may be a text line. Else, the presence of corresponding elements at high resolution makes it possible to confirm the hypothesis of a thick or a multiple line. We apply the same kind of mechanism to

recognize the thin lines that must have corresponding elements at both medium and high resolutions. This strategy enables to eliminate line segments that are detected only at medium or high resolution.

It is important to see that the position of the line segment at the lowest resolution determines the localization of the search zone at the other resolutions. Moreover, the vision at the lowest resolution gives knowledge on the studied line: curvature, skew, length, thickness, that are used to determine how to gather the pixels at the highest resolution and to constitute the final ruling.

3.2 Correspondence between resolutions

The strategy of perceptive cooperation requires to compare some line segments extracted at various resolutions. In order to easily realize this comparison, we use a single coordinate system for all the elements of each resolutions of the image. Thus, the figure 3 represents the three sets of line segments that are extracted at three resolutions, but that are stored in the same coordinate system.

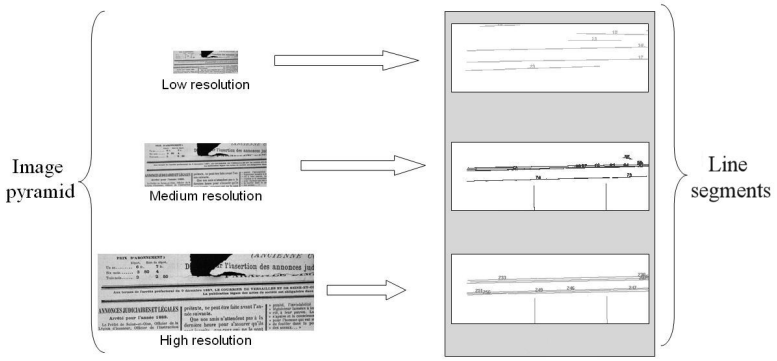


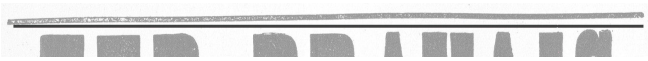
Fig. 3. Pyramid of images and associated extracted line segments stored in the same coordinate system

3.3 Adjustment of position

Using a single coordinate system is necessary but not sufficient to find the exact position of a ruling. Indeed, the difference of scale between resolutions may cause errors of quantification. An example is presented on figure 4: a line segment has been detected at low resolution (figure 4(a)); if we directly transpose it in the upper resolution (figure 4(b)), its position does not exactly match with the present black pixels of the image.

In order to solve this problem, we propose to introduce a new concept, the *abstract line*, and a new tool of *positioning*. The abstract line is a polygonal

(a) Line segment at low resolution



(b) Transposition at high resolution: quantification error

Fig. 4. Example of quantification error

approximation of a line segment. Thus, it can be manipulated independently of any resolution. This abstract line is mainly used by the positioning tool that aims at adjusting precisely the position of a line, according to the presence of black pixels.

An example of use is presented on figure 5, and enables to treat the quantification error presented on figure 4. First, the abstract line is used to define a zone of interest (figure 5(b)). Indeed, at this step, even if the position of the abstract line need to be precised, it contains interesting information about the thickness, the slope and the curvature of the searched ruling, provided by the perception at lower resolutions. The definition of a zone of interest enables to select a set of relevant black pixels (figure 5(c)). Then, the role of the positioning tool consists in adjusting the position of the abstract line according to the present black pixels (figure 5(d)).

(a) In black, line segment provided by a lower resolution: quantification error

(b) Definition of a zone of interest around the abstract line

(c) Selection of black pixels in the search zone

(d) Adjustment of the polygonal approximation according to present black pixels

Fig. 5. Principle of positioning tool

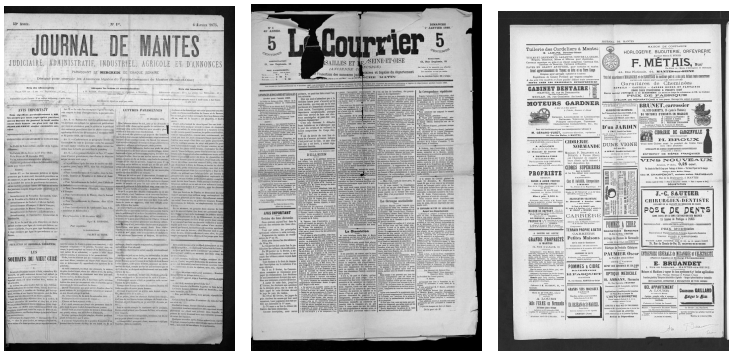
To sum up, this positioning tool enables to correct small quantification errors and to provide a precise localization of the abstract lines. It is one of the key elements that enables to progressively adjust the position of the ruling, taking into account the different levels of resolution.

4 Evaluation of the interest of perceptive vision

In order to show the interest of the perceptive vision, we apply our method for ruling recognition in ancient newspapers. More precisely, we propose to compare our perceptive approach with a monoresolution method.

4.1 Base of evaluation

We applied our method for the recognition of heterogeneous rulings in pages of old newspapers, dated between 1848 and 1944, provided by the *Archives Départementales des Yvelines*. Some examples of studied images are presented on figure 6. We manually labeled a ground truth of 4967 rulings in 157 pages of newspapers



(a) 1875 (b) 1888 (c) 1905

Fig. 6. Examples of studied old newspaper pages

The studied images have an initial size of 5400*8000 pixels. The medium resolution images have a size of 1350*2000 pixels. The low resolution images have a size of 330*500 pixels.

4.2 Metric

In order to evaluate the recognition rate, we have to make a correspondence, for each image, between the expected rulings of the ground truth, and the rulings found by the method. We classify the results according to four categories:

- *total recognition* is the number of expected rulings that have been properly recognized,
- *partial recognition* is the number of expected rulings that have been partially recognized, that is to say the recognized ruling is too short or too long,

– *omission* is the number of rulings of the ground truth that have not been detected,
– *noise* is the number or recognized ruling that does not correspond to any ruling of the ground truth.

4.3 Using only one resolution

Our first experiment consists in extracting the rulings using only one resolution of the image. Thus, we apply our line segment extractor at this given resolution and try to gather them into rulings, without using dedicated thresholds. We propose to compare the obtained results for three different resolutions. This experiment is illustrated on figure 7. The obtained results are presented on table 2.

Method Resolution	Monoresolution High	Monoresolution Medium	Monoresolution Low
Total recognition	69.1%	52.6%	5.6%
Partial recognition	21.2%	8.5%	2.4%
Omission	9.7%	38.8%	92.0 %
Noise	93.9%	16.6%	2.5 %
Time per image	10.9 sec	3.3 sec	2.6 sec

Table 2. Comparison of three monoresolution approaches on 4967 rulings extracted from 157 newspaper pages

Thanks to high resolution (figure 7(a)), we obtain 69.1% of complete recognition of rulings. Thus, this resolution enables the recognition of many rulings but is not able to deal with discontinuities: many rulings are only partially recognized, some thick speckled lines are forgotten. On the opposite, the low resolution (figure 7(c)) enables to detect completely thick rullings but omits all the thin rulings. The medium resolution (figure 7(b)) presents intermediate results but that are not better: only 52.6% recognition.

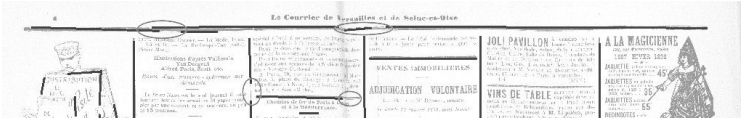
As a conclusion, the obtained results show that, if we have to deal with only one resolution, the best recognition is obtained with the initial image. Now, we will compare this results with our perceptive method.

4.4 Using perceptive vision

The figure 7(d) shows the good recognition of rulings thanks to the perceptive vision. The table 3 present the results obtained with our perceptive method, faced with the best results obtained in monoresolution, *ie* with high resolution.

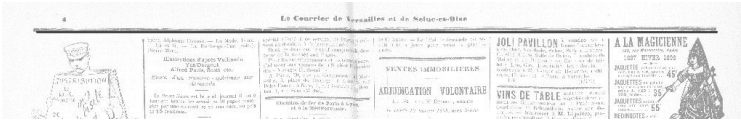
The results show the significant interest of our perceptive method for ruling recognition. Thus, with the perceptive approach, 94.4% rulings are entirely recognized, againts 69.1% in monoresolution.

The computing time is bigger for the perceptive method, due to the need to apply the line segment extractor at three resolutions. However, the large increase of recognition rate compensates for the longer running time.

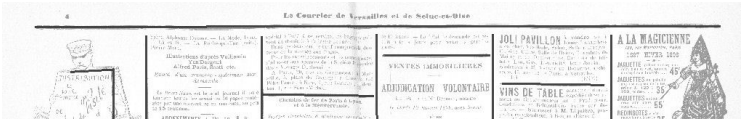


(a) Rulings built with the single high resolution: over-splitting (circled)

(b) Rulings built with the single medium resolution: over-splitting (circled)



(c) Rulings built with the single low resolution: omission of many thin rulings



(d) Rulings built thanks to perceptive approach: good recognition

Fig. 7. Comparison of rulings built with different methods: interest of the perceptive approach

Method	Monoresolution (high)	Perceptive vision
Total recognition	69.1%	94.4%
Partial recognition	21.2%	3.0%
Omission	9.7%	2.6%
Noise	93.9%	31.1%
Time per image	10.9 sec	15.4 sec

Table 3. Comparison of monoresolution approach with our perceptive method, on 4967 rulings extracted from 157 newspaper pages

The weak omission rates with perceptive method (2.6% *vs* 9.7%) is due to the better recognition of thick rulings and speckled rulings that are well perceived thanks to the use of low resolution.

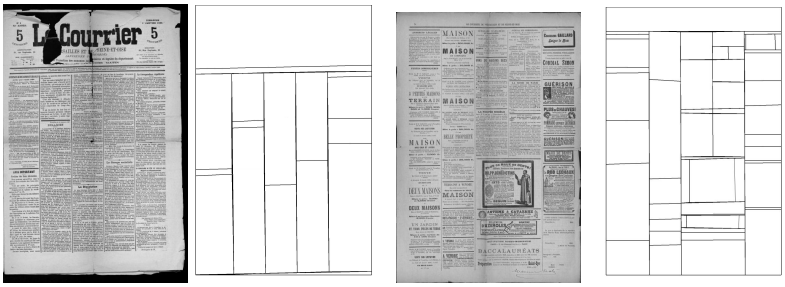
The small partial recognition rate for perceptive approach (3.0% *vs* 21.2%) shows the interest to be guided by hypotheses that are emitted at low resolution, in order to combine line segments at high resolution and to form a single ruling.

The smaller noise rate with perceptive method (31.1% *vs* 93.9%) is due to the prediction/verification strategy that enables to validate the presence of a ruling only if it has been perceived at last on two resolutions. The remaining noise is mainly due to vertical line segments in capital letters of the title. The introduction of more precise knowledge about newspaper pages will enable to take this into account.

5 Application to a real problem

The obtained results show that a perceptive approach enable to detect more reliable rulings. These rulings can be used as a base for document structure recognition.

As an example of application, we introduces our perceptive ruling extractor into a generic method for document structure recognition, DMOS-P [1] [6]. Thus, we realized a grammatical description of the structure of newspaper pages, and its decomposition into boxes. The terminals of this grammar are the rulings. Thanks to this description in DMOS-P method, we obtain a system that is able to segment newspaper pages into boxes (figure 8).



(a) Example of first page (b) Extracted boxes (c) Example of last page (d) Extracted boxes

Fig. 8. Segmentation of a newspaper page using rulings detected with perceptive vision

We evaluate this method on newspaper pages dated between 1859 and 1944, provided by the *Archives Départementales des Yvelines*. More precisely, we noticed that some of the pages were more complicated than other. Thus, in first pages of the newspapers (figure 8(a)), the rulings have a regular thickness,

whereas the last pages (figure 8(c)) contains advertisement with varied kinds of rulings. Consequently, we have created two bases of evaluation: one base of 179 first pages, in which we have built a ground-truth of 4148 boxes, and a second base of 79 last pages with 3480 boxes.

On these two bases, we propose to compare the segmentation into boxes when the rulings are obtained in monoresolution, and when the rulings are obtained with perceptive vision. The obtained results on the two bases are presented on tables 4 and 5.

Version	Boxes	Over-segmentation	Under-segmentation
Monoresolution	4148	10.5%	10.7%
Perceptive approach	4148	10.2%	7.9%
Improvement		-	**33%**

Table 4. Extracting rulings on 179 first pages (good quality rulings)

Version	Boxes	Over-segmentation	Under-segmentation
Monoresolution	3480	17.1%	11.3%
Perceptive approach	3480	13.7%	6.2%
Improvement		**20%**	**45%**

Table 5. Extracting rulings on 79 last pages (difficult rulings)

The results show that using the perceptive vision for ruling detection improves structure analysis. The improvement are more significant on the last pages. Indeed, on these pages, the rulings are particularly difficult to recognize. On these base, thanks to the rulings extracted with perceptive vision, under-segmentation decreases by 45% while over-segmentation decreases by 20%.

6 Conclusion

In this paper, we have presented a new method for rulings recognition in ancient or damaged documents. This method is generic and does not require specific a priori knowledge on dimensions of the rulings, contrary to the approaches found in the literature.

We have shown that the combination of different points of view is necessary in heterogeneous documents because thick or double rulings are not well perceived at high resolution, whereas thin rulings are not detected at low resolution. Consequently, we proposed a prediction/verification strategy to combine

the different perceptions. This works leads to the development of a new positioning tool that enables to set up a correspondence between line segments obtained at different resolutions, and to deal with quantification errors.

The obtained results show the interest of each resolution for a good ruling recognition. The perceptive strategy enables a large increase of rulings recognition. At last, we have also demonstrated that using more reliable detected rulings improve the analysis of more complex structures.

Thanks to the interesting results, this method has led to an industrial transfer to Evodia company, who has treated more than 45,000 newspaper pages thanks to our perceptive method.

References

1. B. Coüasnon. DMOS, a generic document recognition method: Application to table structure analysis in a general and in a specific way. *International Journal on Document Analysis and Recognition, IJDAR*, 8(2):111–122, 2006.
2. B. Gatos, S. L. Mantzaris, K. V. Chandrinos, A. Tsigris, and S. J. Perantonis. Integrated algorithms for newspaper page decomposition and article tracking. In *ICDAR '99: Proceedings of the Fifth International Conference on Document Analysis and Recognition*, page 559, Washington, DC, USA, 1999. IEEE Computer Society.
3. B. Gatos, S.L. Mantzaris, and A. Antonacopoulos. First international newspaper segmentation contest. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, page 1190, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
4. K. Hadjar, O. Hitz, and R. Ingold. Newspaper page decomposition using a split and merge approach. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, page 1186, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
5. Osamu Hori and David S. Doermann. Robust table-form structure analysis based on box-driven reasoning. In *ICDAR*, pages 218–221, 1995.
6. Aurélie Lemaitre, Jean Camillerapp, and Bertrand Coüasnon. Multiresolution cooperation improves document structure recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 11(2):97–109, November 2008.
7. I. Leplumey, J. Camillerapp, and C. Queguiner. Kalman filter contributions towards document segmentation. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR'95)*, pages 765–769, 1995.
8. F. Liu, Y. Luo, D. Hu, and M. Yoshikawa. A new component based algorithm for newspaper layout analysis. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, page 1176, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
9. D. Xi and S. W. Lee. Extraction of reference lines and items from form document images with complicated background. *Pattern Recognition*, 38(2):289–305, February 2005.