2017 14th IAPR International Conference on Document Analysis and Recognition

A document straight line based segmentation for complex layout extraction

Héloïse Alhéritière[1,2], Florence Cloppet[1], Camille Kurtz[1], Jean-Marc Ogier[2], Nicole Vincent[1]
[1] *Université Paris Descartes, LIPADE*
{*heloise.alheritiere, florence.cloppet, camille.kurtz, nicole.vincent*}*@parisdescartes.fr*
[2] *Université de La Rochelle, L3i*
{*heloise.alheritiere, jean-marc.ogier*}*@univ-lr.fr*

***Abstract*—Document layout extraction is a difficult step in the image interpretation process due to the high complexity of documents. The main challenge relies on the huge gap between both the physical and the logical structures of document images. In order to loose as few as possible information, most existing methods are working at pixel level. In this paper, we present a new framework for complex layout extraction based on features of high levels obtained from a document straight line based segmentation. We propose to capture the straight line segments thanks to a new transform integrating the local spatial organization of the segments contained in the document content. Such transform can be applied either on the foreground (related to the document content) or the background pixels, in order to take advantage of the duality of information present in both document parts. Experimental results obtained on the PRImA Layout Analysis dataset illustrate the robustness of our framework for the extraction of specific components of the document including text areas, images and separators.**

***Keywords*-Document layout extraction; page segmentation; straight line features; Local Diameter Transforms; complex layout; sequential labelling**

I. Introduction

Document image understanding aims to convert the information contained in a digitized document image into a more symbolic representation. Extracting the layout of a document is a difficult step in the recognition process due to the high complexity of the documents [12]. The success of this step directly impacts the performance of digitization systems, including OCR precision, document classification, document management system (DMS), and more generally the usability of the information extracted from the document in further recognition steps.

Among the different difficulties, there exists a huge gap between the physical and the logical structures of images of document. It is then difficult to accurately capture the relations between the image regions and the recognized layout structures. In some specific applications such as the authentication of documents, the paper document copies can be obtained after different phases along its life, including several scanning-printing steps. Then it is also necessary to ensure a sufficient level of stability of the layout extraction results with respect to the natural changes the support may encounter. The stability of the results may be impaired by image noise, layout variations or even artefacts produced during (pre-)processing steps.

Automatic approaches for document layout analysis generally involved different tasks of image analysis, ranging from image segmentation / classification to the representation of information extracted. The layout analysis problem aims to extract from the content of a given document various objects of interest (e.g. text, graphic, image, table). Most existing approaches focus on the pixel level to extract this information. In certain approaches, some importance is given to the background appearance, with the assumption that the large white rectangles also carry information. It makes more obvious the spots where the reader has to focus on. In fact there exists some duality property between foreground and background, and it seems relevant to take advantage of both sources of information before taking a decision.

In this paper, we propose an approach focusing on features of higher level than pixels, in particular straight line segments. They may be used to approximate filled forms, lines and drawings that constitute another level of document primitives from a topological point of view. Furthermore according to the length of the segments, some document parts can be discriminated. Thus a novel framework for complex layout extraction based on a document straight line segmentation is presented. Section II reviews some related works. Our first contribution (see Section III) is a new transform called the Local Diameter Transform (LDT) in order to integrate the local spatial organization of the segments. Our second contribution (see Section IV) is a novel framework for layout extraction. It relies on high level features, the straight line segments extracted thanks to the LDT. The originality of this approach is to extract the layout of a document by simultaneously considering information from the page layer and the background layer. Section V presents an experimental study of our approach on the PRImA Layout Analysis dataset. A conclusion that emphasizes our perspectives can be found in Section VI.

II. Related Work

The physical layout of a document includes the geometric layout that contains information about the type of content or media (e.g. text, graphic, image, table) and their respective spatial locations in the document. To extract the geometric layout of a document, two types of approaches coexist: methods without and with segmentation.

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.186

Among the methods without segmentation are the pixel-based classification methods and the layer approach classification methods seeking for a particular media. The first ones are either based on a homogeneity criterion or on specific features [18], [6] (e.g. colour, shape, texture). Their advantage is that there is no need of a priori information about the formatting of the documents or their content. However, they often require post-processing to reinforce the local uniformity of the extracted and labelled areas, as different labels are often given in neighboor pixels. Furthermore, these classification methods rely on a learning phase generally requiring a large document database. The layer approach methods are dedicated to a particular media (text [9], [3], tables [15], logos [11]). The notion of layer needs to be clearly defined. For example, does a signature represent a layer by itself or does it belong to the handwritten text layer? The advantage of these methods is that they extract information present in each layer independently. These different sources of information can then be aggregated in order to correct potential segmentation errors.

The methods relying on a segmentation step rely first on a splitting of the image and then on a labelisation of the regions. The labelisation step needs both the definitions of a suited feature and of a classifier. The segmentation-based methods can be grouped in three categories: top-down (i.e. recursive splitting from the whole image to the regions), bottom-up (i.e. iterative grouping of pixels or regions) and hybrid (i.e. a mixture of the two previous [17]). The main drawbacks of top-down techniques, such as the Projection Profile [10], recursive X-Y cut [13], [5], and run-length smoothing [16] methods, are their sensibility to the document orientation, and their ability to extract only rectangular regions. Bottom-up methods are generally based on local information from binarised document images. Such information can be the distance between black pixels that are in the foreground and grouped in connected components [19], [20], or the width of white areas that are in the background and can be considered as separators [1]. Most of the methods are based on criteria (contrast or colour homogeneity [4], texture [14], or stroke information [7], [8]) related to features that try to describe in a singular way the text regions to separate them from the others.

More recently, the information contained in the strokes has shown a great interest to develop methods dedicated to the extraction of text, lines, separators, etc. The estimation of the stroke width (Stroke Width Transform [7]) is efficient to extract text in images but is not able to correctly separate text, lines and separators. In this context, information concerning the orientation of the lines is crucial. Global transforms such as the Hough Transform are not well adapted since they only provide information about the segment directions. A similar information involving some directional study in different directions can be given for example by the Radon Transform, but it does not take into account the relative position of the shape pixels along a direction, and only gives a global information along the direction. Furthermore, the spatial organization of the segments in a local area is an important feature to discriminate text, drawings, images and separators. In [8], the separation between text and lines is based on a SVM classifier that integrates three sources of information: the relative thickness, the elongation and the compacity of the extracted components. However this method does not directly take into account the spatial organization in all the directions.

In this paper, we propose a new transform called the Local Diameter Transform in order to integrate the local spatial organization of the segments. Such transform can be applied either on the foreground (related to the document content) or the background pixels, to take advantage of the duality of information present in the two parts of the document. From a perceptual point of view, a document page can be viewed on one hand, as a lot of small objects (characters, parts of segments) that need to be grouped together in order to extract either words, lines, paragraphs, columns, or large and thin separators, or drawings or images. But on the other hand, it can also be considered as a set of delineated regions of interest (ROIs). This delineation phase can be done by detecting lines in the background, that are piecewise limited or tangent to these piecewise areas. The envelop of each ROI is then constituted by the set of the calculated tangents.

III. Local Transforms

We define in this section some transforms that give more local information about the spatial organization of the segments contained in a document content than traditional global transforms. Of course some wavelet transforms could be used but then the difficulty is to find the right decomposition level suited for a specific study. We then propose to adapt the neighbourhood level around each pixel according to the image content. As we focus on the straight lines contained in the image, we refer to the Radon Transform that we modify in order to obtain more localized information.

A. Local Radon Transform

An image is a function $f : \mathbb{R}\times\mathbb{R} \to V$ that associates with each point (x, y) a value $f(x, y)$ of the set V. We assume that the image is binary ($V = \{0, 1\}$), white is marked with 0, black with 1. The study set is assumed to be black.

Starting from the definition of the Radon Transform, the objective is to propose a "local" version of this definition. Let $\theta \in \mathbb{R}$ be an orientation angle, and $\rho \in \mathbb{R}$ a distance from the origin. The Radon Transform of an image f can be expressed as

$$R(f)(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\times \delta_0(\rho - x.\cos(\theta) - y.\sin(\theta))\,\mathrm{d}x\,\mathrm{d}y \quad (1)$$

where δ_0 is the Dirac distribution in 0. By considering a parametric representation of a line defined by (ρ, θ) such as

$$\begin{cases} x = \rho\cos(\theta) - t\sin(\theta) \\ y = \rho\sin(\theta) + t\cos(\theta) \end{cases} \quad t \in [-\infty, +\infty] \tag{2}$$

the Radon Transform of an image f can be expressed as

$$R(f)(\rho,\theta) = \int_{-\infty}^{+\infty} f(\rho\cos(\theta) - t\sin(\theta), \rho\sin(\theta) + t\cos(\theta))\ \mathrm{d}t \tag{3}$$

In the plane (ρ, θ) the maxima of the transform indicate the presence of a dominant direction.

The definition that we propose is relative to each point P of the image, we denote (x_0, y_0) its coordinates. The lines that pass through this point are characterized by pairs (ρ, θ) where ρ and θ are linked by the relation $\rho(\theta) = x_0\cos(\theta) + y_0\sin(\theta)$. Thus, on each line characterized by $(\rho(\theta), \theta)$ passing through the point P, the parameter of the point P is defined as $t_0 = \frac{x_0 - \rho\cos(\theta)}{-\sin(\theta)} = \frac{y_0 - \rho\sin(\theta)}{\cos(\theta)}$.

At each point, we define a *local* Radon Transform by

$$LR(f)(\theta, x_0, y_0) = \int_{-\infty}^{+\infty} f(\rho(\theta)\cos(\theta) - t\sin(\theta), \rho(\theta)\sin(\theta) + t\cos(\theta))\delta_0 \times \left(\int_{t_0}^{t} f(\rho(\theta)\cos(\theta) - u\sin(\theta), \rho(\theta)\sin(\theta) + u\cos(\theta)) - 1\ \mathrm{d}u\right)\ \mathrm{d}t \tag{4}$$

where LR gives the maximum length of the segment passing through P in the direction θ.

B. Local Diameter Transform

In many applications, the length of a segment is not relevant in an absolute way but only relatively to the size of the image. Then, based on the local Radon Transform, the local diameter at a point is measured by evaluating the length of the largest segment passing through this point and contained in the set of pixels labelled 1. Thus only one value at each point corresponding to the maximum length of the segment is kept, independent of its direction. This is what we call the Local Diameter Transform (LDT) defined as:

$$LDT(f)(x,y) = \max_{\theta\in[0,\pi]} LR(f)(\theta, x, y) \tag{5}$$

Obviously, depending on the application, the importance of the segment must be estimated relative to the dimensions of the document page D. By denoting $diam(\Delta_f, \theta)$ the diameter of the image definition set Δ_f in the direction θ, we then define in each pixel, the relative local diameter by:

$$RLDT(f)(x,y) = \max_{\theta\in[0,\pi]} \frac{LR(f)(\theta, x, y)}{diam(\Delta_f, \theta)} \tag{6}$$

and the Relative Local Diameter Transformation of f is $RLDT(f)$. An example of applying RLDT to an image can be seen in Fig. 1(b).

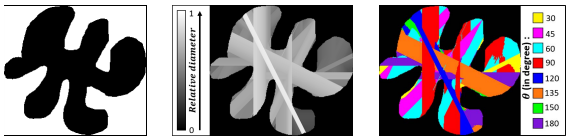
In some cases, for instance if the value of the Distance Transform (pixel to contour distance) at a pixel is not high, then the RLDT value offers a local linearity information. The two transforms, Distance Transform and RLDT, provide complementary local information about the shape.

C. Relative Local Orientation Transform

From a perceptual point of view, the direction of a set is perceived locally as a function of the direction of the largest segment passing through the considered point. Based on the previous notations, the transform in orientation is then defined in each pixel by:

$$RLOT(f)(x,y) = \arg\max_{\theta\in[0,\pi]} \frac{LR(f)(\theta, x, y)}{diam(\Delta_f, \theta)} \tag{7}$$

where we rely on the LR as the perception depends on the relative length of the segments. The Relative Local Orientation Transformation of f is $RLOT(f)$ (see Fig. 1(c)).



(a) (b) (c)

Figure 1. Transforms: (b) RLDT and (c) RLOT calculated on an original image (a) by taking into account 8 orientations.

IV. A Mixed Decomposition of Content Page and Background

As mentioned in Section II, we are both interested in extracting information from the content page layer and the background layer. Our method is a mixed approach based on high level features, as they are more structured than the pixels, that is to say the straight line segments extracted thanks to the transforms presented in Section III. This method is considered as mixed, as it makes simultaneously the segmentation and the labelisation of the image regions.

A. Method Overview

The proposed local transforms (see Section III) enable to characterize the lengths of the straight line segments of a document image D, their directions and the dominant segments. They will be applied both on the binarised image I (see Section IV-C) and in the background image $\bar{I}$ (see Section IV-B). In I, the text can be characterized by several short segments whereas the materialised separators can be characterized by long segments. Besides, in $\bar{I}$ the long straight lines can be either in the inner part of the background far from the ROIs or they can be tangent to them, or in image regions wrongly rendered by binarisation process. Thus, these long segments, as they are parts of the ROI envelop, make possible the characterization of the ROIs, that might be textual areas or images. The aggregation

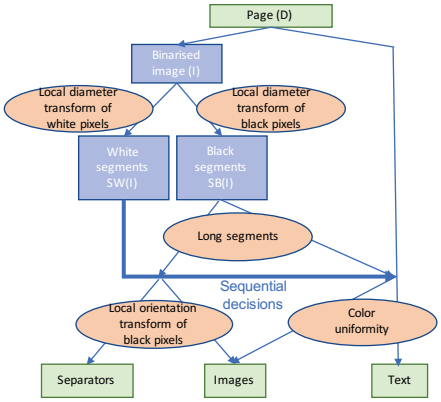


Figure 2. Overview of the method (input / outputs are depicted in green).

of the information extracted both from I and $\bar{I}$ allows the geometric layout extraction that is presented in Section IV-D. Fig. 2 illustrates the method workflow.

B. Background Decomposition

As mentioned above, the ROI envelops of a document page can be extracted from $\bar{I}$. Let us note $RLDT(\bar{I})$ the transform of $\bar{I}$. The grey levels of this transformed image are indicating the maximum length of a straight line segment containing the pixel. As main regions are sought for, only long segments and their spatial organisation will be kept by applying $Th^t(RLDT(\bar{I}))$[1]. In our case we have experimentally chosen t equal to 0.07. It means that these segments represent at least 7% of page size in the direction of the segment. The connected components of the complementary of this image are candidate ROIs. Among these candidates, empty ROIs generated by the fact that more lines than needed to define the envelop are extracted or that some segments are longer than their informative parts, which are tangent to the ROI, are suppressed. The resulting ROI set is denoted image $SW(I)$ thereafter. Considering image of Fig. 3(a), binarised in Fig. 3(b), $SW(I)$ is illustrated in Fig. 3(c) in which ROIs are depicted in white. Finally, the ROIs have to be labelled. This can be done according to characteristics of the document page content.

C. Content Page Decomposition

The information about the length of straight segments allows to address the problem of labelling textual regions and materialised separators. Indeed, from a perceptual point of view, a separator is characterised in $RLDT(I)$ by long straight line segments whereas textual part is characterized by short straight line segments. In a practical way, materialised separator candidates are modelled as straight lines with length longer than 10% of the page size in the direction of the segment (see Fig. 3(d)). In order to get results invariant with respect to a small rotation of the document page, the $RLDT$ operator is applied to the morphological dilation (with a three pixel wide square structuring element, designated by operator Di) of I. Then the materialised separator candidates are the connected components of $RLDT(Di(I))$.

For the textual part candidates, with a bottom-up approach, we consider all pixels belonging to a maximum straight line segment with length less than 2% of the page size in the direction of the segment (see Fig.3(e)). In this process some parts of the writing are lost, and the selection of these pixels (named set K) is used as a seed to recover all connected components in I that contain some elements of K. The more the small segments cover the connected component in I, the more we are confident in the fact that a text part has been found. Then a degree of confidence doC for each connected component C to be a text part can be measured by the percentage of its elements in K as

$$doC_K(C) = \frac{Area(C \cap K)}{Area(C)} \quad (8)$$

The natural extension of doC yields to the definition of $doC_K(I)$. Finally, the text candidate part is defined as $Te(I) = Th^t(doC_K(I))$. In our case, t has been taken equal to 0.85. It allows the detection of good textual candidates.

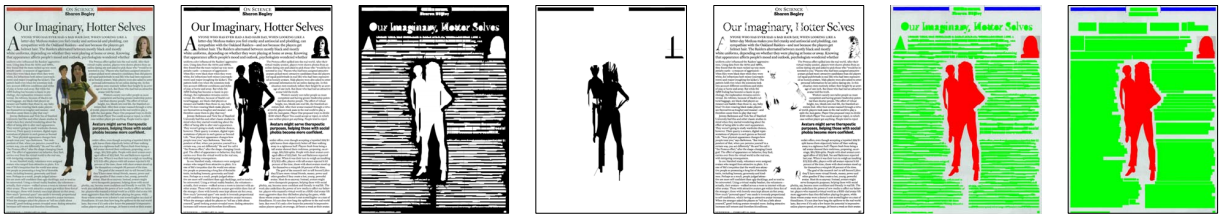
To sum up, as a result of the decompositions presented in Sections IV-B and IV-C, three kinds of information are now available: $SW(I)$, an approximation for the document page segmentation, $RLDT(Di(I))$ indicating separators or more generally graphical parts and $Te(I)$ indicating text candidates. Aggregating these three sources of information will allow a refining of the region contours, and/or a reinforcement in the confidence of the labelling.

D. Layout Extraction

The connected components R_n in $SW(I)$ will receive some labels when a confidence level high enough is reached. This will be done using several sources of information extracted: from the initial image D; from the local transformations of the binary image, $RLDT(I)$, $RLDT(Di(I))$ and $RLOT(Di(I))$; from the knowledge of the already labelled elements. Starting from $SW(I)$ the set of connected components is iteratively decreasing till no more components are to be labelled. The updated set is then named $G(I)$.

First we focus on long segments because we are confident on their presence, the elements R_n of $SW(I)$ are selected according to the segment length criterion, this is deduced from $RLDT(Di(I))$. The easiest conclusion is when the direction of the present segments is unique. The RLOT enables to get the information. If all values in the restriction of $RLOT(Di(I))$ to $(R_n \cap Th^{10}(RLDT(Di(I))))$ are the same, it means the long segments have all the same direction. Then, the region should be either a materialised separator,

[1]We note respectively Th_t and Th^t the operators enabling to threshold a grey level image either considering the pixels with values respectively inferior to t or superior to t.



(a) (b) (c) (d) (e) (f) (g)

Figure 3. Worflow images: (a) Example of PRImA document image D; (b) Binarised image I; (c) Background analysis: $SW(I)$; (d) Long straight line segments (black); (e) Short straight line segments (black); (f) labelled image of D and (g) post-processed labelled image where text lines are aggregated in paragraphs: text (green), image (red) and separator (blue).

or a reverse video text, which appears in white within a significant black box in the binarised image. These two labels can be easily discriminated from the height to width ratio. For materialized separators this ratio is high. Otherwise the region containing reverse video text is labelled as text.

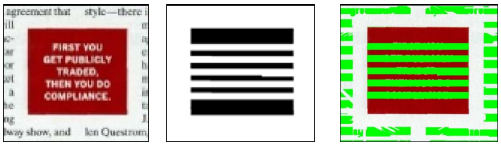
Now $G(I)$ can be updated. In all remaining R_n regions, different directions of segments are present. When long segments with several directions are present, the region can be labelled as image. This occurs when in the region $R_n \cap Th^{10}(RLDT(Di(I)))$ is not empty. The region is surely composed of an image or some graphical part but they are not distinguished in this study and are both labelled as image. Of course, to smooth the regions, the non bounded connected components of the complementary of the images are merged with the image region and labelled as image. The $G(I)$ set has now to be updated and the new labelled regions are excluded.

The regions R_n that comprise enough text part candidates, $SB(I)$, will be labelled as text. The density of text is considered as a confidence given to the label "text". The confidence degree is expressed as $doC_{SB(I)}(R_n)$. The regions defined in $Th^t(doC_{SB(I)}(G(I)))$ are labelled as text and taken off $G(I)$. We chose t equal to 0.80. We may notice that $SB(I)$ was a good approximation of text parts. In our process, we recover the initial components present in the binary image.

At the end of the process, $G(I)$ contains some garbage parts. In order to label these remaining elements R_n in $G(I)$ we are referring to the initial image where colour information is available. When the region colour is uniform, we assume the region can be labelled as text, this should be large text, otherwise it is labelled as image.

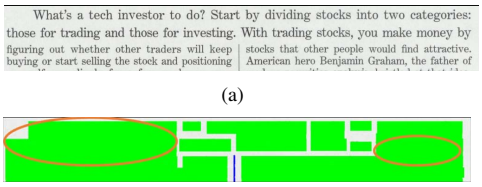
V. Experimental Results

From a qualitative point of view, we observe from Fig. 3(f) that our framework has been efficient to extract specific components of the document, including text areas images and separators. The text has been extracted at the text line level providing a precise description of the document content. The method also enables to detect both text and reverse video text as illustrated in Fig. 4 where the text zone is in white, limited by long segments (here in black).



(a) (b) (c)

Figure 4. Reverse video text: (a) initial image; (b) Long straight line segments (black); (c) extracted text zones (green).



(a) (b)

Figure 5. Post processing to aggregate text lines: (a) initial image (b) examples of final text zones violating the reading order (orange circles).

We also observe that the method is also suited for images of any shape, as the lines in different directions are used as the envelops of the ROIs. Besides the method seems efficient in order to find the materialised separators. Then, the document description will enable to give some semantics to many document parts such as tables or graphs, which could be tackled in future works.

This fine behaviour of our method has also been evaluated in a quantitative way. For this purpose, its performance has been compared to the method of Felhi et al. [8], which was itself compared with methods involved in the ICDAR page segmentation competition [2]. The competition offered a publicly available benchmark composed of a dataset (called PRImA), including 55 document images with their associated ground truth. Each document image is composed of text paragraphs, separators and images. To provide a fair comparison, we have made use of the PrimA evaluation tool. It assumes the text paragraphs are extracted and it gives penalty when the reading order is violated; it takes as input the labelled image and the binarised image of the document.

Otsu method was used to binarise the 55 document images. Both $RLDT$ and $RLOT$ were computed according to 8 orientations (30°, 45°, 60°, 90°, 120°, 135°, 150°,

Table I
EVALUATION RESULTS: PRIMA MEASURE COMPARISON FOR DIFFERENT COMPONENTS OF THE DOCUMENT LAYOUT.

	Image (%)	Separator (%)	Text (%)
Dice	36.46	27.06	40.02
Fraunhofer	61.83	84.51	82.37
REGIM-ENIS	54.42	74.53	15.44
Tesseract	52.95	69.42	73.24
Felhi et al.[8]	67.96	78.46	89.53
Proposed Method	95.3	94.3	81.35

180°). Then the proposed method allows the extraction of images, separators and text lines. In order to extract paragraphs instead of lines, a post-processing step, based on the extraction of limits of text lines, was developed. These limits are selected from the long segments present in the background, only in the vertical direction, more precisely obtained from $Th^7(RLDT(\bar{I}))$ image. This enables to comfort alignments and to group similar lines. Of course some errors remain. In Fig. 5(b), some words of the upper long lines are wrongly considered to belong to the two paragraphs below (see orange circled areas).

Table I presents the results of 6 methods, 4 from the competition [2], the method of Felhi et al. [8], and our method. These results show that our method is more efficient for the labelling of separators and images. Concerning text areas, the performance is slightly lower. The main errors are due to paragraph over-segmentation. In fact 93.18% of the text regions are correctly labelled (by only considering the organization in lines). Thus, the post-processing step aggregating lines into paragraphs needs to be improved to avoid violating the reading order as in Fig. 5(b).

VI. CONCLUSION

Layout extraction in document is still a hot topic and we have proposed here a unified description of the different media contained in a document. Considering in a dual way the content and the background, we are able to better describe the shapes. The use of different length straight lines enables to have a multi level approximation of the elements contained in the document. Unlike other approaches, no learning phase is necessary to label the ROIs, as our method is based on a description feature. As short term perspective, we plan to propose a strategy to automatically adapt the choice of the different thresholds involved in the methodology. In future works, it will be also interesting to aggregate different sources of information to build some even higher level elements such as tables or graphs. Finally, in order to get rid of the binarisation errors, improvements can be done on the binarisation step. An other way could be to generalise the proposed transforms for colour images.

ACKNOWLEDGMENT

This work was supported by the French *Agence Nationale de la Recherche* under Grant ANR-14-CE28-0022.

REFERENCES

[1] A. Antonacopoulos. Page segmentation using the description of the background. *Comput Vis Image Underst*, 70(3):350–369, 1998.
[2] A. Antonacopoulos, S. Pletschacher, D. Bridson, and C. Papadopoulos. ICDAR 2009 page segmentation competition. In *Proc. ICDAR*, pages 1370–1374, 2009.
[3] S. Bukhari, A. Azawi, M. Ali, F. Shafait, and T. Breuel. Document image segmentation using discriminative learning over connected components. In *Proc. DAS*, pages 183–190, 2010.
[4] S. Chuai-Aree, C. Lursinsap, P. Sophatsathit, and S. Siripant. Fuzzy c-mean: A statistical feature classification of text and image segmentation method. *Int J Uncertainty Fuzziness Knowledge Based Syst*, 9(6):661–671, 2001.
[5] D. Coppi, C. Grana, and R. Cucchiara. Illustrations segmentation in digitized documents using local correlation features. *Procedia Comput Sci*, 38(1):76–83, 2014.
[6] M. Cote and A. B. Albu. Texture sparseness for pixel classification of business document images. *Int J Doc Anal Recognit*, 17(3):257–273, 2014.
[7] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *Proc. CVPR*, pages 2963–2970, 2010.
[8] M. Felhi, S. Tabbone, and M. V. O. Segovia. Multiscale stroke-based page segmentation approach. In *Proc. DAS*, pages 6–10, 2014.
[9] S. Hamrouni, F. Cloppet, and N. Vincent. Handwritten and printed text separation: Linearity and regularity assessment. In *Proc. ICIAR*, pages 387–394, 2014.
[10] O. Iwaki, H. Kida, and H. Arakawa. A document image segmentation classification and recogntion system. In *Proc. CSMC*, pages 759–763, 1987.
[11] V. P. Le, M. Visani, D. C. Tran, and J. M. Ogier. Logo spotting for document categorization. In *Proc. ICPR*, pages 3484–3487, 2012.
[12] G. Nagy. Twenty years of document image analysis in PAMI. *IEEE Trans Pattern Anal Mach Intell*, 22(1):38–62, 2000.
[13] G. Nagy and S. Seth. Hierarchical representation of optically scanned documents. In *Proc. ICPR*, pages 347–349, 1984.
[14] S. S. Raju, P. B. Pati, and A. G. Ramakrishnan. Gabor filter based block energy analysis for text extraction from digital document images. In *Proc. DIAL*, pages 233–243, 2004.
[15] F. Shafait and R. Smith. Table detection in heterogeneous documents. In *Proc. DAS*, pages 65–72, 2010.
[16] F. Shih, S. Chen, D. Hung, and P. Ng. A segmentation method based on office document hierarchical structure. In *Proc. ICSI*, pages 258–267, 1992.
[17] R. Smith. Hybrid page layout analysis via tab-stop detection. In *Proc. ICDAR*, pages 241–245, 2009.
[18] R. Vieux and J. Domenger. Hierarchical clustering model for pixel-based classification of document images. In *Proc. ICPR*, pages 290–293, 2012.
[19] F. M. Wahl, K. Y. Wong, and R. G. Casey. Block segmentation and text extraction in mixed text/image documents. *Comput Vision Graph*, 20(4):375–390, 1982.
[20] A. Winder, T. L. Andersen, and E. H. B. Smith. Extending page segmentation algorithms for mixed-layout document processing. In *Proc. ICDAR*, pages 1245–1249, 2011.