A Text Detection Technique Applied in the Framework of a Mobile Camera-Based Application

Silvio Ferreira — Ph.D. Student
Vincent Garin — M.S. Student
Bernard Gosselin — Ph. D., Assistant Professor
Faculté Polytechnique de Mons, TCTS Labs, Belgium,
silvio.ferreira@tcts.fpms.ac.be bernard.gosselin@fpms.ac.be

Abstract

Recent advances in mobile devices allow us to address many new challenging problems. One of them is automatic text recognition for embedded platform. This paper describes an innovative text detection system in the context of an embedded camera-based application. We propose a method to identify text regions inside an image, to correct orientation problems and to analyze document layout. Text areas are isolated with a texture segmentation approach. Due to mobile conditions, text orientation and perspective must be corrected. First a fuzzy estimation of text orientation is computed quickly. If text is too much distorted, the text perspective is corrected by using a line segmentation method in two steps. Finally the layout of the document is computed in order to deliver the reading order of the document. This language-free system has been developed with special attention to computational performances. The experimental results have proven that the method is effective and realistic.

1. Introduction

One of the most fascinating frontier projects in the field of artificial intelligence is machine understanding of text. Commercial solutions combining a scanner and a computer currently exist and have proved to be efficient. But through the recent developments in the segment of mobile devices like personal digital assistants (PDA) or smartphones a broad range of new applications are emerging. These mature hardware technologies introduce new opportunities to automatically recognize document images taken in mobile conditions with a camera-based system instead of a scanner. These new devices could be helpful for specific user groups such as blind and visually impaired people or tourists in foreign countries. Textual information is everywhere in our daily life and having access to it is essential for them to improve their autonomy and their integration. Our application aims at overcoming these barriers by offering to blind and visually impaired people a mobile access to textual information (signs, books, menus, etc.)

In our application, a blind user first takes a picture with a mobile device such as a smartphone or a PDA. Then, our system automatically detects text areas in the picture and delivers the layout of the document. Finally, text can be transformed into speech signal.

But the specific conditions of this application imply several major constraints:

- *Text image deterioration*: a document image acquired on a camera brings text detection and characters segmentation problems. Solutions need to be found to take care of the poor quality of image sensors (up to now available with this kind of devices), image stabilization and unknown lighting conditions.
- *Low computational resources*: the use of a mobile device such as a PDA or a smartphone limits the CPU and the memory resources. These constraints force our algorithmic techniques to be efficient and well optimized in order to achieve an acceptable execution time.

Three key software technologies are required for this system: text detection, optical character recognition (OCR) and text to speech synthesis (TTS). Figure 1 illustrates several examples of the realistic images database we have created in close collaboration with a group of blind users.

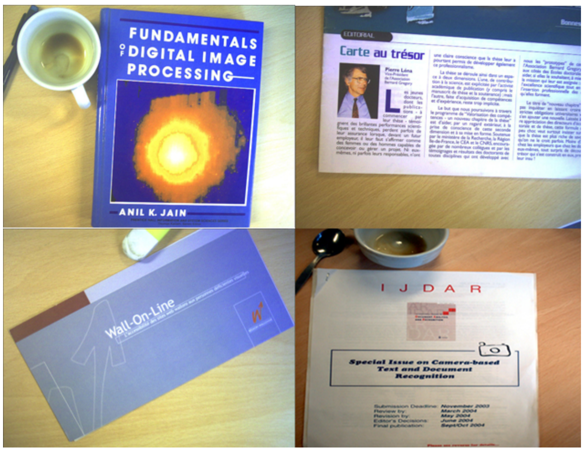

FUNDAMENTALS OF DIGITAL IMAGE PROCESSING
ANIL K. JAIN
Carte au trésor
Wall-On-Line
I J D A R
Special Issue on Camera-based Text and Document Recognition

Figure 1. Samples of our images database

In this paper, we describe the current state of our text detection system, from capture of images to document layout analysis necessary in our framework for the determination of the reading order. For each step, we have either adapted traditional techniques of the text recognition field or developed new ones. This paper is organized as followed: section 2 provides an overview of the text regions detection system and the approach we follow. Considering that this part has already been detailed in a previous publication [1], this work is briefly described. This paper mainly details the methodology to deal with orientation and perspective problems in section 3 and the document layout analysis in section 4. These automatic sub-systems are especially dedicated to mobile camera-based applications. Finally section 5 concludes the paper.

2. Text detection system

Traditionally, document images are scanned with a flatbed, sheet-fed or mounted imaging device. However digital cameras have shown their potential as an alternative imaging device. But camera-based images require specific processing. The first step is detection and localization of the text regions. The idea is to locate the text elements without necessarily recognizing them, cut them out of the image, determine the reading order and eventually correct their perspective problems. Our system captures images with a resolution of 1280 * 1000 pixels and a focus fixed at a distance of 30 cm in order to be able to enclose an A4-sized document. Figure 1 illustrates the main steps of our text detection process.

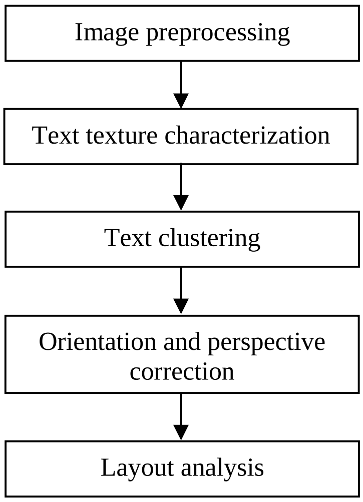


Image preprocessing
Text texture characterization
Text clustering
Orientation and perspective correction
Layout analysis

Figure 2. Overview of text detection system

Most of the previous researches about text detection focus on extracting text from video. Techniques applied to images or video key frames can broadly be classified as edge ([2], [3], [4]), color ([5], [6]), or texture based ([7], [8], [9]). Each approach has its advantages/drawbacks concerning accuracy, efficiency and computational requirements.

Our text detection technique is based on a texture segmentation approach. Text regions inside the image are considered as a textured region to isolate. Non-text contents in the image, such as blanks, pictures, graphics and other objects in the image must be considered as regions with different textures. The human vision can quickly identify text regions without having to recognize individual characters because text has textural properties that differentiate it from the rest of a scene.

Before characterization, images require pre-processing operations. Firstly original images are downsampled for the whole text detection process (due to computational restrictions). This reduction of pixels is obligatory mainly to reduce later the execution time of k-means clustering. A contrast adjustment is then operated in order to normalize global lighting conditions.

Our method for text texture characterization is based on Gabor filters which have been used earlier for a variety of texture classification and segmentation tasks [9]. The features are designed to identify text paragraphs. None of them will uniquely identify text regions. Each individual feature will still confuse text with non-text regions but a bank of filters will complement each other and allow identifying text unambiguously. We associate to the bank of filters a partially redundant feature, a local edge density measure based on Sobel filters. This feature improves the accuracy and robustness of this method while reducing false detections.

We use a reduced K-means clustering algorithm to cluster feature vectors. In order to reduce computational time, we apply the standard K-means clustering to a reduced number of pixels and a minimum distance classification is used to categorize all surrounding non-clustered pixels. Empirically, the number of clusters (value of K) was set to three, value that works well with all test images. The cluster whose centre is closest to the origin of feature vector space is labelled as background while the furthest one is labelled as text. This is because of larger answers to high spatial frequency filters in the text areas. Several results of text clustering and text region isolation are shown on figure 3.

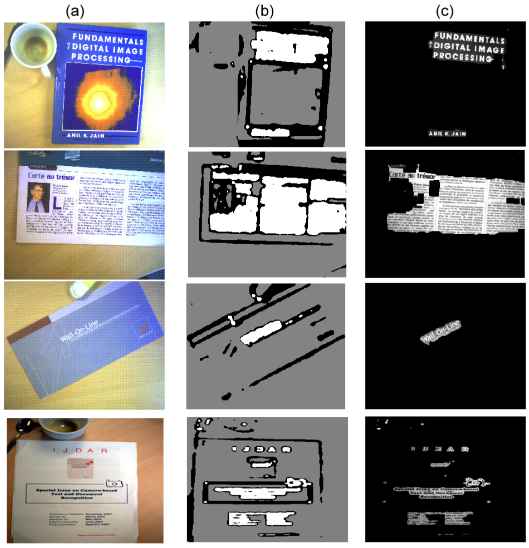


Figure 3. (a) Original images (b) Three-classes clustering (c) Text region

3. Orientation and perspective correction

3.1. Overview

At this step of the system, the image will be treated differently according to the estimation of the global orientation of the document. This fairly accurate estimation allows deciding if the perspective correction module must be applied. Indeed, due to computational efficiency, the whole perspective correction is applied only when this estimation of orientation is larger than an absolute tolerance margin. We have noticed in practise an average of 10% of non-horizontal text images taken by blind users. It confirms the need to apply perspective correction only when it is necessary to boost the average execution time. A margin of 5° is tolerated which ensures an efficient line segmentation and OCR without reorientation. This approximate orientation estimation must be computed quickly. Figure 4 schematizes the orientation and perspective correction system.

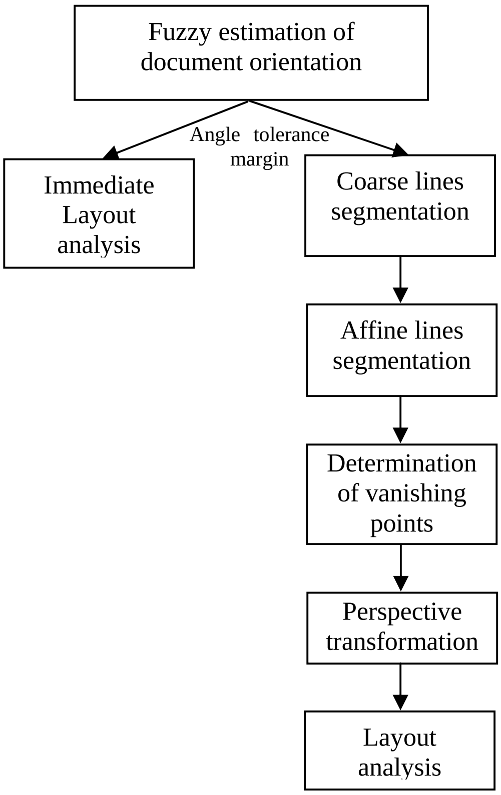

Fuzzy estimation of document orientation
Angle tolerance margin
Immediate Layout analysis
Coarse lines segmentation
Affine lines segmentation
Determination of vanishing points
Perspective transformation
Layout analysis

Figure 4. Scheme of perspective correction system

When the document is considered to have its perspective to be corrected, the main idea is to identify and segment text lines and then to determinate the horizontal vanishing point. The position of this point allows resolving partially unknown parameters of the perspective transform to apply. The vertical vanishing point is difficult to estimate accurately but correcting only the horizontal vanishing point already gives good results.

3.2. Approximate document orientation

Our approach is based on the theory of illusory clues [10]. When a document is captured by a camera at an unknown angle, it is of course impossible to establish a priori what is horizontal. However, given the usual layout of western-style writing, the horizontal direction is reflected in the image in the dominant direction of illusory lines. A preprocessing stage binarizes the input text areas computed during the text detection step, turning them into ‘blobs’ representing either single characters or (portion of) words or lines (cf. figure 5). An interesting advantage of this binarization is to analyse only pixels previously classified as text and using an independent threshold for each text box. This method allows taking care of local gradient of luminosity in the image such as shadows, etc.

Finally the approximate global orientation is estimated with the orientation of the major axis of the blobs which are the most elongated and then the most representative. This fast estimation of orientation is performed in about 1 second with a classical PDA (CPU Intel XScale 400 MHz, 64 MB Ram). Figure 5 illustrates the fast binarization phase.

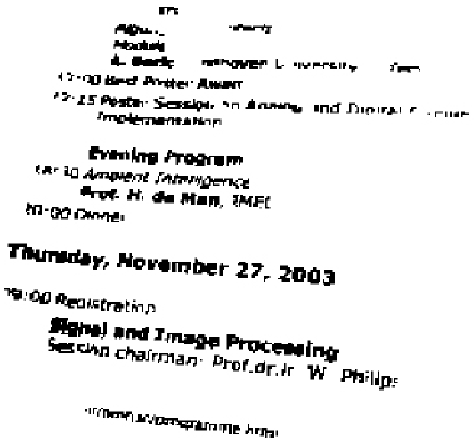


Figure 5. Binarization used for fuzzy estimation of document orientation using 'blobs' properties

3.3. Coarse and affine lines segmentation

As previously mentioned, the lines segmentation procedure is performed in two levels.

First, a reorientation is operated with the previous fuzzy estimation of text orientation. We can then compute the vertical profile of binarized text zones and operate a first coarse segmentation (cf. image (b) of figure 7) with the detection of gaps in the vertical profile.

This first lines segmentation is not accurate enough in all cases, especially when text suffers from a perspective problem. It is why a second local method is used to segment line after line. For each line previously detected with the vertical profile, a diffusion cone starts on the first blob detected. The line detection evolves incorporating the first blobs detected into the cone. When a new blob is added into the line, the properties of the diffusion cone (orientation and aperture) are adapted with the position, the size and the orientation of the last incorporated blob. This local segmentation method enables taking into account text lines with perspective problems. The algorithm is illustrated in figure 6 and in image (c) of figure 7. Figure 6 shows the diffusion of the cone across one line.

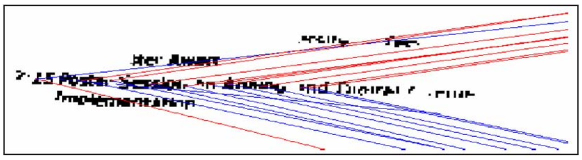
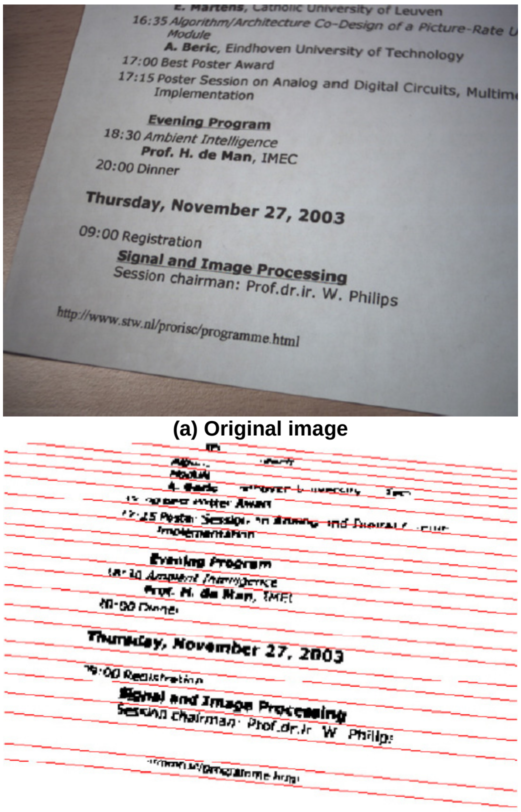


Figure 6. Affine lines segmentation

The orientation and perspective correction sub-system performs well in about 80% of cases of images with perspective problems.

3.4. Vanishing point determination and perspective transform

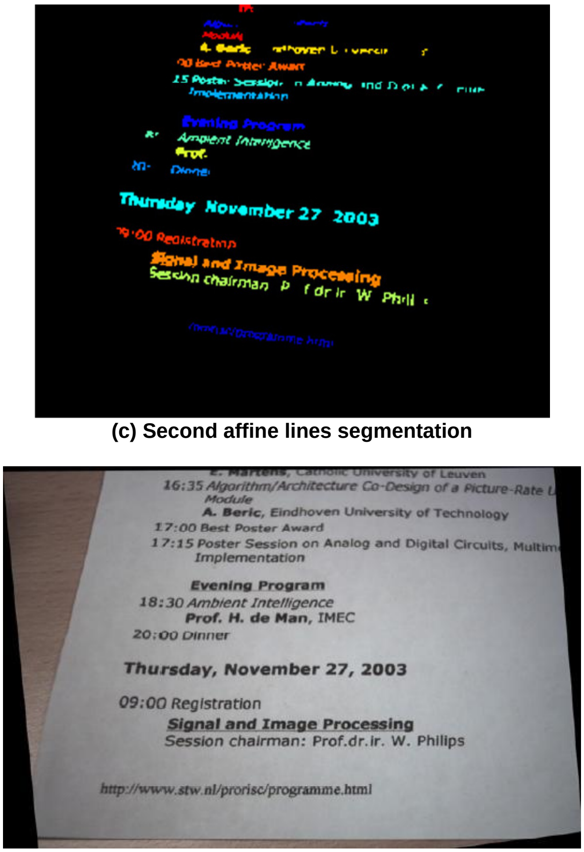
In non-horizontal images the text lines converge in a point called the horizontal vanishing point. We use a fast approximation method to estimate its position. We can then resolve partially the perspective transform. The mathematical development is detailed in [11]. The result is illustrated in image (d) of figure 7.



E. Martens, Catholic University of Leuven
16:35 Algorithm/Architecture Co-Design of a Picture-Rate U
Module
A. Beric, Eindhoven University of Technology
17:00 Best Poster Award
17:15 Poster Session on Analog and Digital Circuits, Multim
Implementation
Evening Program
18:30 Ambient Intelligence
Prof. H. de Man, IMEC
20:00 Dinner
Thursday, November 27, 2003
09:00 Registration
Signal and Image Processing
Session chairman: Prof.dr.ir. W. Philips
http://www.stw.nl/prorisc/programme.html

(a) Original image

(b) First coarse lines segmentation


E. Martens, Catholic University of Leuven
16:35 Algorithm/Architecture Co-Design of a Picture-Rate U
Module
A. Beric, Eindhoven University of Technology
17:00 Best Poster Award
17:15 Poster Session on Analog and Digital Circuits, Multim
Implementation
Evening Program
18:30 Ambient Intelligence
Prof. H. de Man, IMEC
20:00 Dinner
Thursday, November 27, 2003
09:00 Registration
Signal and Image Processing
Session chairman: Prof.dr.ir. W. Philips
http://www.stw.nl/prorisc/programme.html

(c) Second affine lines segmentation

(d) Final result of perspective correction

Figure 7. Main steps of perspective correction

Even without the determination of the vertical vanishing point, this method performs satisfying results in about 80% of non-horizontal document images but the entire algorithm requires an execution time of about 5 seconds with a classical PDA.

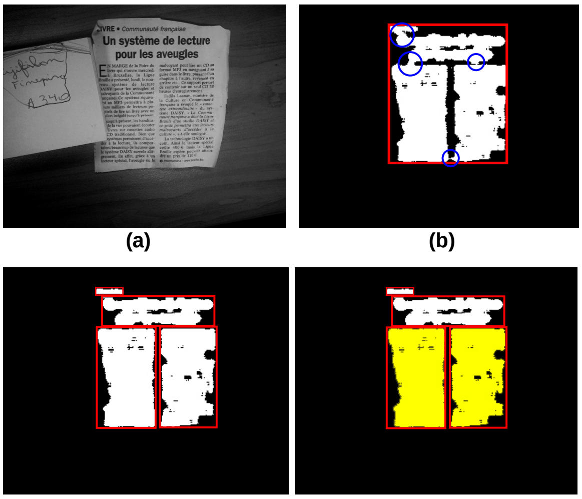
4. Document layout analysis

A document image contains important information in the geometric arrangement of the text zones on the page – the page layout. The layout of a document is the result of the application of complex, interactive rules about where to place text on the page. But almost all layouts tend to be composed of a number of recurring primitives, text lines, paragraphs and columns. We call the extraction of these primitives physical document layout analysis. The extraction of higher-level properties of a document like titles or authors is referred as logical document layout analysis [12].

In our framework we try to extract the physical layout analysis of the unknown document. The layout analysis is performed to organise text boxes for a logical reading order. Layout analysis provides in this manner a reading position to every text box. They are later processed independently by the the OCR system. We make assumptions of occidental writing systems. Text is read from top to bottom and from left to right. Another assumption consists in the major presence of document images composed of traditional class of layouts like Manhattan textual layouts which are fully described in [13]. Briefly the document page must be composed mainly of blocks of text lines and symbols and lines must share a common orientation.

Our document analysis sub-system is designed with special care to computational performances. It is why the algorithm uses previously computed information (results of text detection module). Text detection areas (binary images) will be transformed in text boxes and labelled in order.

Firstly an iterative procedure of columns and paragraphs separation is applied based on the morphological profiles of every binary text box. The separation can be achieved precisely with the detection of gaps in the vertical and the horizontal text class profiles of each textbox. These gaps correspond to undesirable 'bridges' which link columns and paragraphs together (cf. blue circles in image (b) of figure 8). This type of analysis allows fast computational performances due to the use of one-dimensional signals and binary sub-images.


LIVRE • Communauté française
Un système de lecture pour les aveugles

(a)

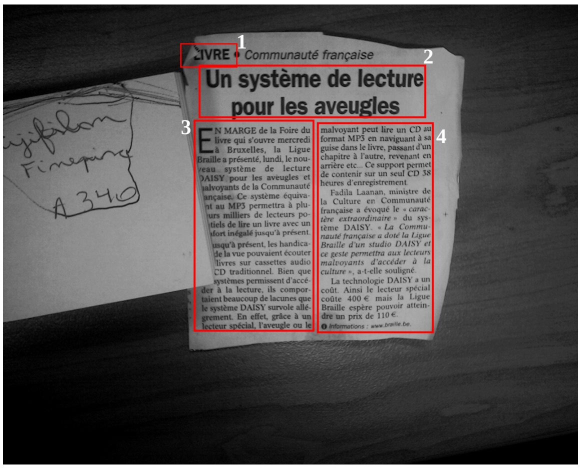
(b)

(c)

(d)

Figure 8. (a) Original image (b) Illustration of text regions to separate (c) Results after iterative columns & paragraphs separation (d) Illustration of detection of columns

Now that paragraphs and columns are separated, we can detect the columns (cf. image (d) of figure 8). The method to identify two or more columns takes into account the ratios of vertical overlay between textboxes (to detect horizontal alignment) and their relative distance. Finally the reading order is decided between the boxes from top to bottom taking into account detected columns. This method of layout analysis is performed on a PDA in about one second. One final result is shown on figure 9. The layout analysis sub-system has separated and given a reading order to four text boxes.



LIVRE • *Communauté française*

Un système de lecture pour les aveugles

EN MARGE de la Foire du livre qui s'ouvre mercredi à Bruxelles, la Ligue Braille a présenté, lundi, le nouveau système de lecture DAISY pour les aveugles et alvoyants de la Communauté ançaise. Ce système équivant au MP3 permettra à pluurs milliers de lecteurs potiels de lire un livre avec un fort inégalé jusqu'à présent.

usqu'à présent, les handicade la vue pouvaient écouter livres sur cassettes audio CD traditionnel. Bien que systèmes permissent d'accéder à la lecture, ils comportaient beaucoup de lacunes que le système DAISY survole allégrement. En effet, grâce à un lecteur spécial, l'aveugle ou le malvoyant peut lire un CD au format MP3 en naviguant à sa guise dans le livre, passant d'un chapitre à l'autre, revenant en arrière etc... Ce support permet de contenir sur un seul CD 38 heures d'enregistrement.

Fadila Laanan, ministre de la Culture en Communauté française a évoqué le « *caractère extraordinaire* » du système DAISY. « *La Communauté française a doté la Ligue Braille d'un studio DAISY et ce geste permettra aux lecteurs malvoyants d'accéder à la culture* », a-t-elle souligné.

La technologie DAISY a un coût. Ainsi le lecteur spécial coûte 400 € mais la Ligue Braille espère pouvoir atteindre un prix de 110 €.

Informations : www.braille.be.

Figure 9. Layout analysis result

In our images database, the layout analysis sub-system has an estimated efficiency of about 90%.

5. Conclusion

This paper has described a text detection and document analysis system in the context of a camera-based text recognition application. The method has been designed in the context of providing mobile access to textual information for blind and visually impaired people.

The initial work has focused on an adapted technique for the separation of a document image into regions of text.

The text detection method based on texture segmentation has been tested with a variety of printed documents from different sources. Performances are acceptable although misclassifications occur occasionally. These misclassifications errors are often detected later during the optical character recognition step and then rejected.

The main contribution of this work consists in a new efficient approach to perform page deskewing and document analysis. The system performs orientation and perspective correction only when required after a first fast fuzzy estimation of text orientation. Results of this algorithm are promising but a reduction of its computational complexity need to be realised when the whole process of page deskewing is operated. Our document analysis system aims at extracting the physical layout of the document and deciding the logical reading order. Its results and computational performances are satisfactory.

The algorithms have been designed using a realistic images database (pictures were taken by blind users). Due to this methodology, further improvements would consist in the adaptation of this system to reference images databases (ICDAR images databases for example) in order to compare it performances to other techniques and obtain more quantitative results.

6. Acknowledgements

This project is called Sypole and is funded by Ministère de la Région wallonne in Belgium.

7. References

[1] S. Ferreira, C. Mancas-Thillou, B. Gosselin,"From Picture to speech: an innovative OCR application for embedded environment", Proc. of the 14th ProRISC workshop on Circuits, Systems and Signal Processing (ProRISC), 2003

[2] P. Clark and M. Mirmehdi, “Recognizing text in real scenes”, International Journal on Document Analysis and Recognition, Springer Berlin Heidelberg, August 2002, vol.4, pp.243-257

[3] J. Ohya, A. Shio and S. Akamatsu, “Recognizing characters in scene images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, February 1994, vol. 16 n°2 pp.214-220

[4] M. Pietikäinen and O. Okun, “Text extraction from grey scale page images by simple edge detectors”, Proc. of the 12th Scandinavian Conference on Image Analysis, June 2001, pp.628-635

[5] W.-Y. Chen and S.-Y. Chen, “Adaptative page segmentation for color technical journals’cover images”, Image and Vision Computing, Elsevier Science, 1998, vol.16, pp.855-877

[6] Y. Zhong, K. Karu and A.K. Jain, “Locating text in complex color images”, Pattern Recognition, 1995, vol.28, n° 10, pp.1523-1535

[7] V. Wu, R. Manmatha and E.M. Riseman, “Textfinder: an automatic system to detect and recognize text in images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Nov. 1999, vol.21, n° 11, pp.1224-1229

[8] H. Li, D. Doermann and O. Kia, “Automatic text detection and tracking in digital video”, IEEE Transactions on Image Processing, January 2000, vol.9, n°1, pp.147-156

[9] A.K. Jain and S. Bhattacharjee, “Text segmentation using Gabor filters for automatic document processing”, Machine Vision and Applications, Springer-Verlag, 1992, vol.5, pp.169-184

[10] Maurizio Pilu, “Extraction of illusory linear clues in perspectively skewed documents”, IEEE Computer Vision and Pattern Recognition Conference, Kauai, December 2001

[11] G. Fangi, G. Gagliardini, E.S. Malinverni, “Photointerpretation and small scale stereoplotting using digitally rectified photographs by geometrical constraints”, in

Int. Archives of Photogrammetry and Remote Sensing, vol. XXXIV, part. 5/C7, Germania, 2001, pp. 160-167

[12]T.M. Breuel, “A Review of Branch-and-Bound Algorithms for Geometric and Statistical Layout Analysis”, Huitième Colloque International Francophone sur l’Ecrit et le Document CIFED 2004, La Rochelle, 21-25 juin 2004

[13] D.J. Ittner and H.S. Baird, “Language-Free Layout Analysis”, Proc. IAPR 2nd International Conf. on Document Analysis & Recognition, Tsukuba Science City, Japan, October 1993, pp.336-340