A tool for scanning document-images with a photophone or a digicam

Mohammed El Rhabi, Abdelilah Hakim, Zouhir Mahani

▶ To cite this version:

Mohammed El Rhabi, Abdelilah Hakim, Zouhir Mahani. A tool for scanning document-images with a photophone or a digicam. Computer Applications for Communication, Networking, and Digital Contents International Conferences, FGCN and DCA 2012, Dec 2012, Gangneug, South Korea. hal-01812625

HAL Id: hal-01812625

https://enpc.hal.science/hal-01812625

Submitted on 11 Jun 2018

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.

A tool for scanning document-images with a photophone or a digicam

Mohammed El Rhabi[1] Abdelilah Hakim[1] and Zouhir Mahani[2]

[1] Université Cadi Ayyad, Faculté des Sciences et Techniques - Guéliz (FSTG)
Laboratoire de Mathématiques Appliquées et Informatique
Bd. Abdelkrim El Khattabi , B.P. 618 Guliz, 40000 Marrakech- Maroc
elrhabi@gmail.com ; abdelilah.hakim@gmail.com,
http://www.fstg-marrakech.ac.ma/
[2] Université Ibn Zohr, ESTA, Laboratoire Matriaux, Systèmes et Technologies de l'information,
B.P: 33/S Agadir - Maroc,
zouhir.mahani@gmail.com,
http://www.esta.ac.ma/

Abstract. In this work, we propose a tool to scan a document-image acquired with a cameraphone. Firstly, we try to reduce the noise in the document-image. Then we build a new image by cropping or by perspective rectifying the denoised one. From this step, we can expect the document to a real quadrangle. The new document is analyzed and we try to find images, logo or non text element in the document-image with the aid of an image segmentation. At this stage, we provide deux parts of the document image: the text part and the "non text" part of the document-image (images, logos, non ...). The text part of the document-image is enhanced by an original pde's based model that we proposed. The "non text" document is enhanced by classical methods such as retinex processing. Then, we merge both parts of the document image by a poisson image editing. The effectiveness and the robustness of the proposed process are shown on numerical examples in real-world situation (images acquired from cameraphones).

1 Introduction

Computer vision based documents recognition could be an interesting way to dematerialize informations to manage clients and company's internal documents, offering enterprise wide fast access to business critical information while enhancing the achitecture in place. Typically the dematerialized document formats are PDF. Here, the problem could be separated as least in five steps: noise reduction, perpective rectifying, segmentation : text and non text parts of the document-image, text and image enhancement and a merge of enhanced parts. In the computer vision approaches, low cost cameras (webcam, cameraphone, ...) could introduce some distortions and noise artifacts, see [1] for an overview of document image degradation modeling.

Document images are supposed to be obtained from a mobile device - a cameraphone more exactly, but could also come from any digital camera. In the following, we will only suppose the device to be a cameraphone as this is the case where most of the problems occur.

The image restoration images by PDE's based models starts to produce signifiant results as images with the diffusion of documents to remove noise while preserving important information for readability ([2,3,5,6]) or to separate the back to front ink interferences ([7]) . These works have shown that could be improved up to 30% recognition rate of OCR by restoring the document images.

Generally speaking, an image u could be seen as the product of a reflectance v and the illumination effect I (see [8]). This model has been used in [9,10,11] to restore blurred barcode signals under nonuniform illumination.

This paper provide a tool to scan a document with cameraphone. After the segmentation process, the text enhancement is based on new method based on an isotropic ([5]) or anisotropic ([6]) diffusion to estimate the luminance (or the log of the luminance) of a document-image.

Our methods can produce three kinds of documents: black and white, grayscale or color ones. In all cases, input images are color pictures. Producing color documents should be seen as a different workflow, involving mostly the same algorithms as in the grayscale processing.

The paper is organized as follows : section 2 gives a brief description of a camera lens. Section **??** presents the non uniform illumination problem and defines a global model. Then two methods are given as solutions, namely a natural criterion afterward a pde's based model. Section **??** describes the method and shows how to implement it. Finally, some numerical results illustrate this work in Section 5.

2 Description

Most important image degradations are related to blur, poor contrast, Jpeg compression artifacts or even an inappropriate point of observation.

In figure 1, we show a flow chart of the processing sequence to be applied to restore document image.

Input image (top of the flow chart) is the color Jpeg image sent by the cameraphone user. At the very bottom of this chart, output images are grayscale images, a black and white one (fax purpose) and optionally a thumbnail of the grayscale image. All other images shown on the chart are temporary images.

We can identify three major processing steps, which will be further explained with much more details :

- Denoising: if needed this processing step aims at reducing the noise of a color image. Here, we use for example the fast algorithm proposed in [4].
- Dewarping : if we can identify the document page borders in the image, we build a new image by cropping or by perspective rectifying the denoised one. From this step, we can expect the document to fit a real quadrangle (see section 3).
- Document-image segmentation : at this step, we try to separate text and "non text" parts in the document-image. This segmentation is obtained by a constrained watershed(see [12]).
- Text and "non text" enhancement (if applicable):
  - Text enhancement : this processing step works as a bad contrast or non uniform illumination correction of the whole document. We refer to [5,6]) where we describe our method to enhance text in document image by pde's based models.
  - "Non text" enhancement (images or logo) by a retinex processing (see [13]).
- The final document is obtained by the merge of both parts of the document (text and "non text" parts) by mean of a Poisson Image Editing (see [14])

For the sake of simplicity, some processings have been expunged here. Indeed, some features are measured and collected all over the processing sequence to provide some user notifications about the input image quality and so our ability to deal with it. These features come from the range of graylevels or the amount of noise in each color channel. They are all merged to give some clues about the input image quality.

3 Page border based dewarping : A way to post-correcting the observation point

It is particularly relevant to produce straight documents as opposed to inclined ones (see fig.2), photographers should not be stressed too much about perfectly framing the document. Even more, sending well aligned documents to faxes is important to enhance the final document legibility.

When page borders are visible in an image, we can use them as a guideline to correct the point of view and make a new image which simulates a perfectly straight document.

First, page borders need to be detected. To perform such a detection we need some image preprocessing techniques to obtain image edges (see fig.3).

As one can see, a first pass reduces information(fig. 3(a)) before extracting with *Canny* filtering the most important edges (fig. 3(c)) in the image. From here, these edges are only delivered as marked pixels.

One way to convert edge location to a line equation is achieved by using the *Hough transform* algorithm. The following equation describe a line :

$$ax + by + c = 0 \tag{1}$$

Where (x, y) are the coordinates of the points lying on the line which parameters are (a, b, c). But our problem is to find parameters only knowing pixel edge locations. The *Hough transform* proposes a way to exchange the unknowns. Making the assumption that (x, y) are known there are two unknowns [3] which describe all lines that meet the point (x, y).

(1) is not well adapted (reasons are outside the scope of the paper), and typical *Hough transform* adopts a polar line representation (see 4).

So, a point (x_0, y_0) in the image gives a set of lines represented as a sinus curve in the Hough space. Our Hough space behaves as an accumulator space, meaning for each point label as edge we accumulate the associated sinus curves in the Hough space. Peaks appear in the Hough space which are the detected lines we are looking for. Because a point in the Hough space represents the parameters of a line in the image.

Figure 5 represents the detected image edges , the associated Hough space representation, and fitted lines.

---

[3] (a, b, c) are related to each other, so formally speaking only two unknowns remained
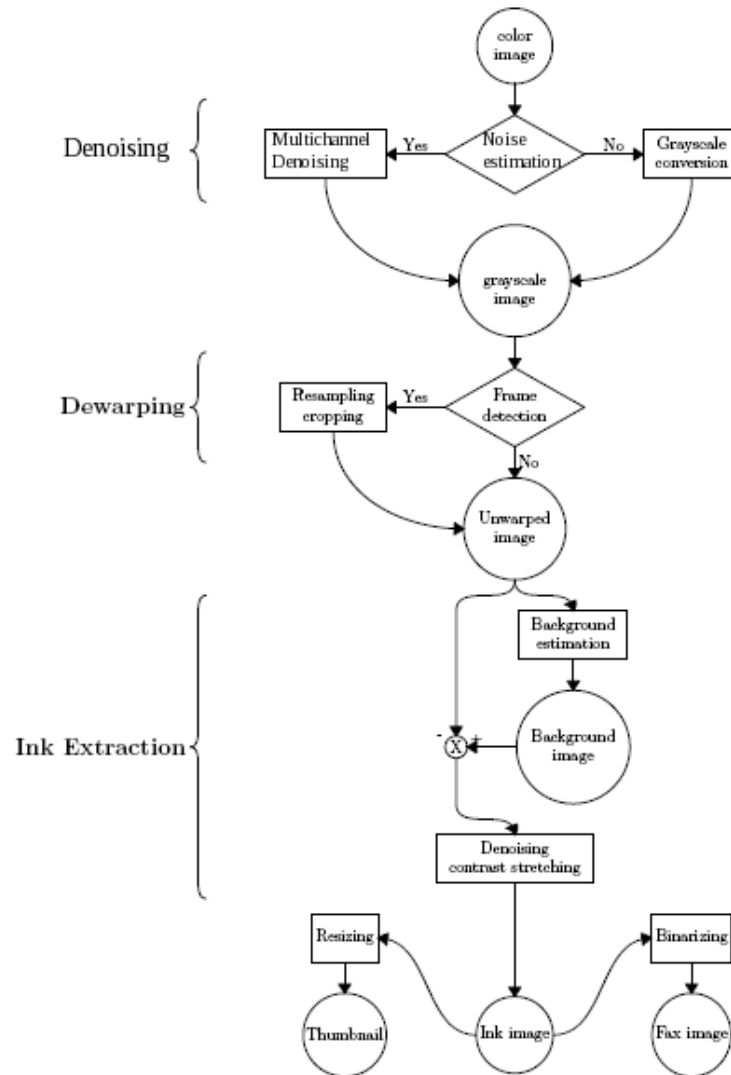
Fig. 1. Global flow chart of the image processing involved in the document image processings improvement from $qipit^{\mathrm{TM}}$. Images are represented like circles, while processing steps are associated with quadrangles. Decisions look as diamonds. A color image is provided as input (top of the chart), and one may expect three output images (very bottom of the chart): a grayscale one, a fax image (black and white) and optionally a thumbnail of the grayscale image. Major processing steps are deblurring, dewarping, and ink extracting. Minor steps deal with user notifications (not mentioned here) or binarization.
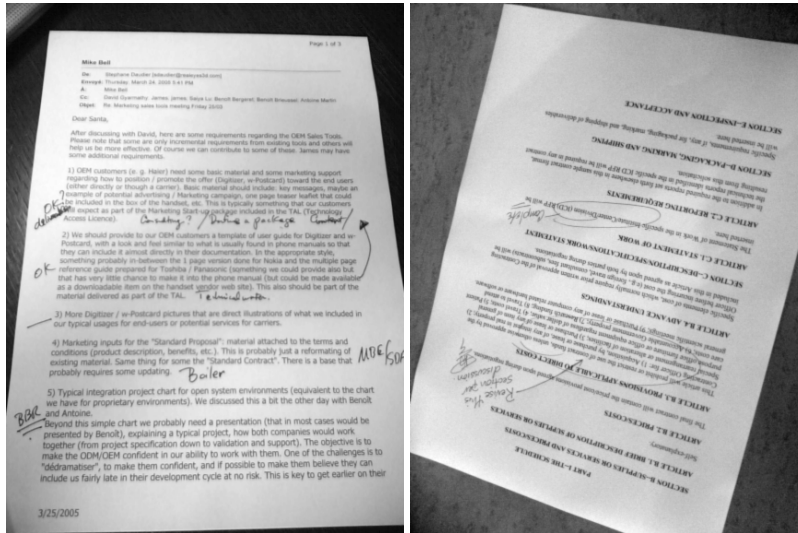
Fig. 2. Typical documents seen under an inclined point of view.
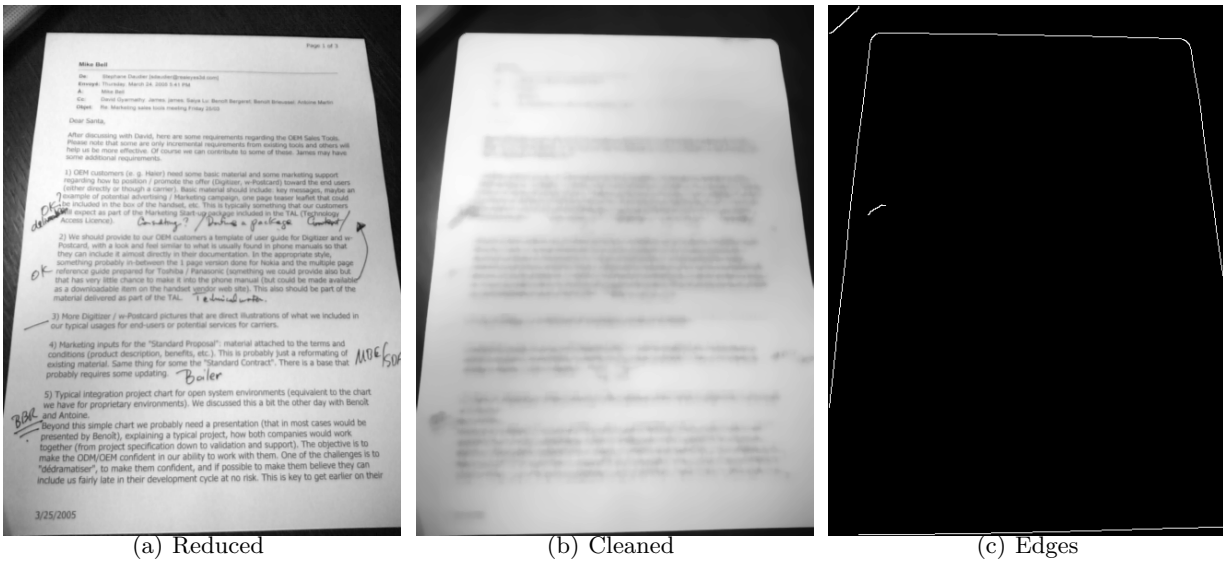
(a) Reduced   (b) Cleaned   (c) Edges
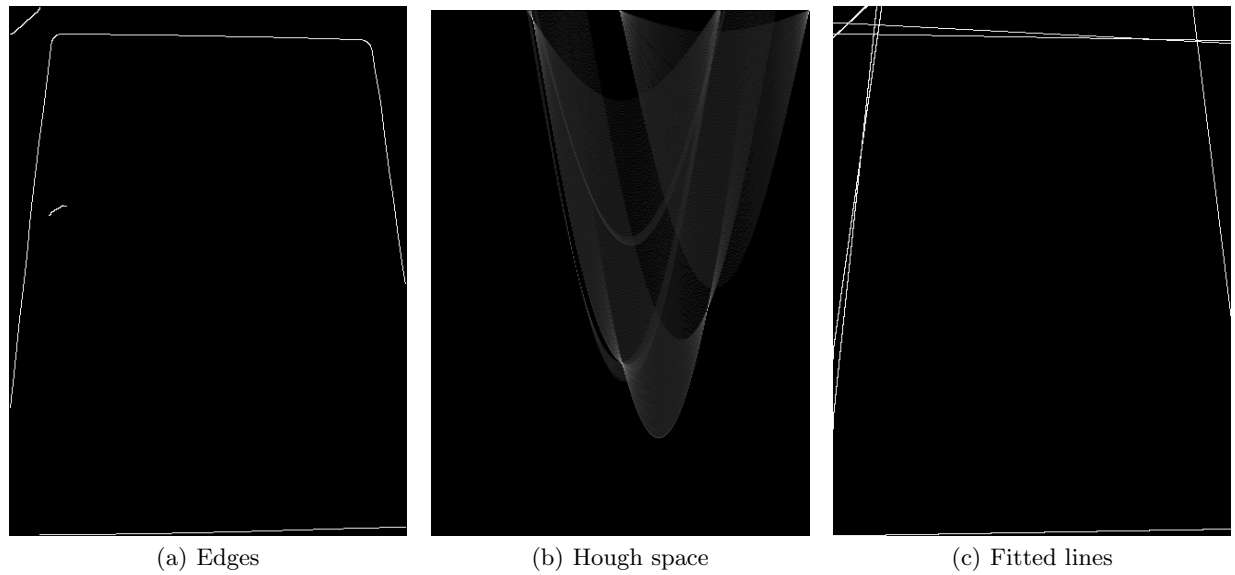
Fig. 3. Preprocessings before detecting page borders.

$$\rho = \frac{\rho_0}{sin(\theta - \phi)}$$
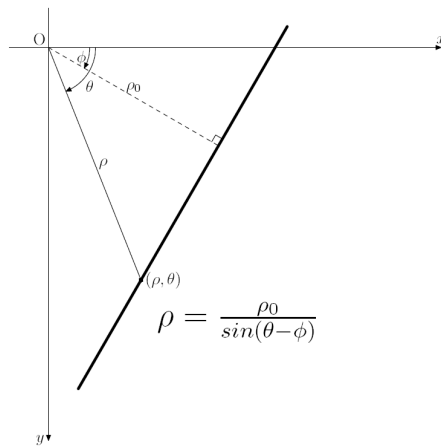
Fig. 4. Polar line representation - the couple of values (ρ, θ) describe a line in polar coordinates.

(a) Edges   (b) Hough space   (c) Fitted lines

Fig. 5. Hough transformation to detect lines

Several lines are detected but we still don't know which ones constitute the page borders. So all sets of four lines are tested if they are good candidates to represent the page borders. Unrealistic sets are immediately rejected and a score is given to the other ones. This score is built from the assumptions that the good lines have extremities closed two by two and they constitute a quadrangle big enough in the image. Figure 6 shows selected edges from previous line detection:

(a) Selected edges   (b) Part of the score came from distances between lines extremities
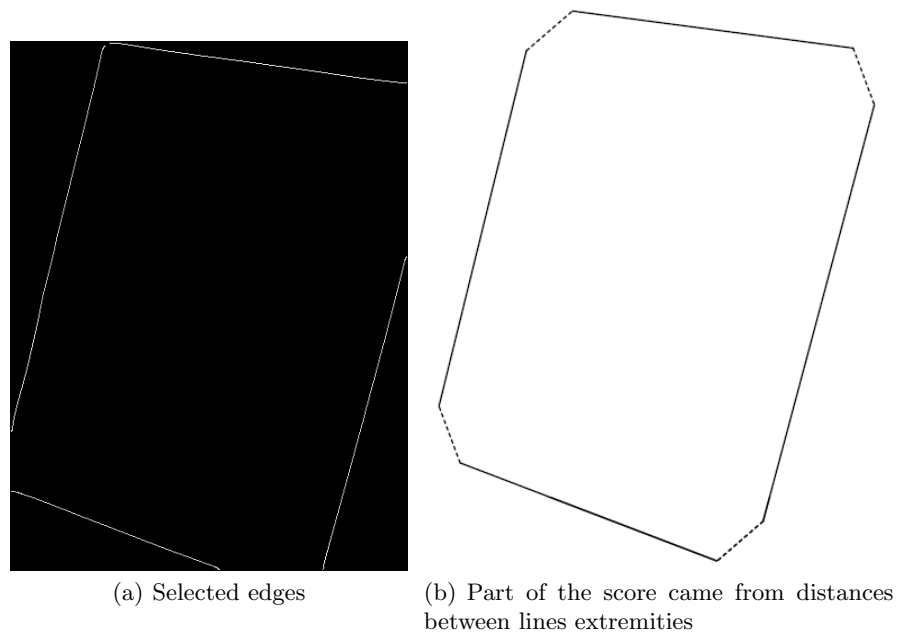
Fig. 6. Selected edges (fig.6(a)) lead to the computation of a score. This score is partly based on the cumulative distance between segments end points (dashed lines in fig. 6(b))

When a frame is detected, we only need to perform the perspective correction of the image. Perspective correction is performed by using *bicubic* interpolation, and is a real perspective transformation (homography) hence respectfull of original document proportions.

4 Separation Text - Non Text in the document-image

In this section, we need a tool to localize uniform background image regions from previous ink extraction method. Such a tool should be a constrained watershed algorithm (see [12]). Here, images can be viewed as a topographical relief, which are flooded progressively (see figure 7). In practice, we consider gradients magnitude of the image. In figure 8, we describe the first step of this process, we remove the text from the document image in order to obtain only the non text part of the document. Figure 9, we apply a filtered constrained watershed with the rational assumption that the biggest uniform region is probably the background of the document-image.
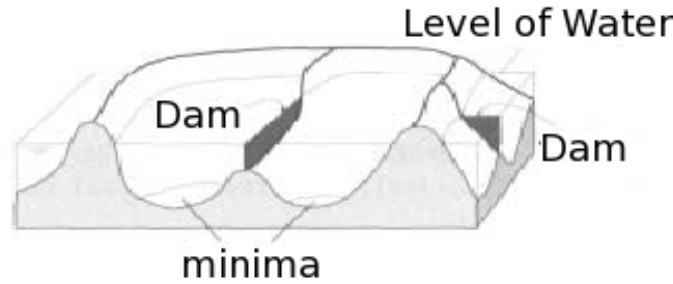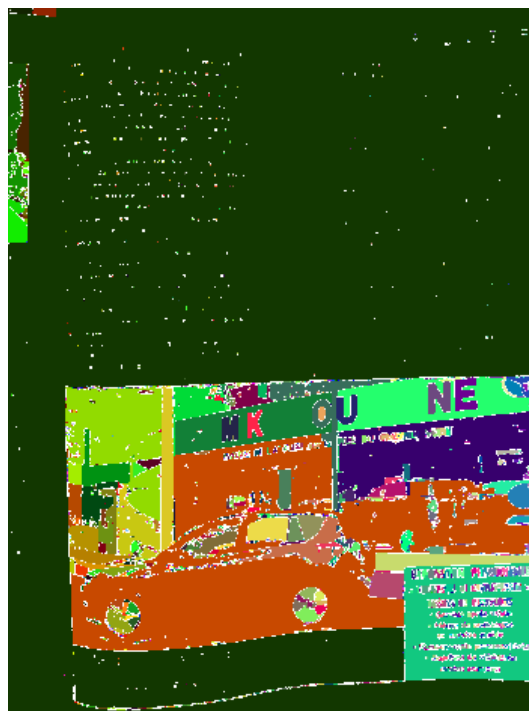

Fig. 7. Watershed principle.

(a) Original

(b) Result of new background estimation (only dark ink- obtained by calssical mathematical morphology tools)
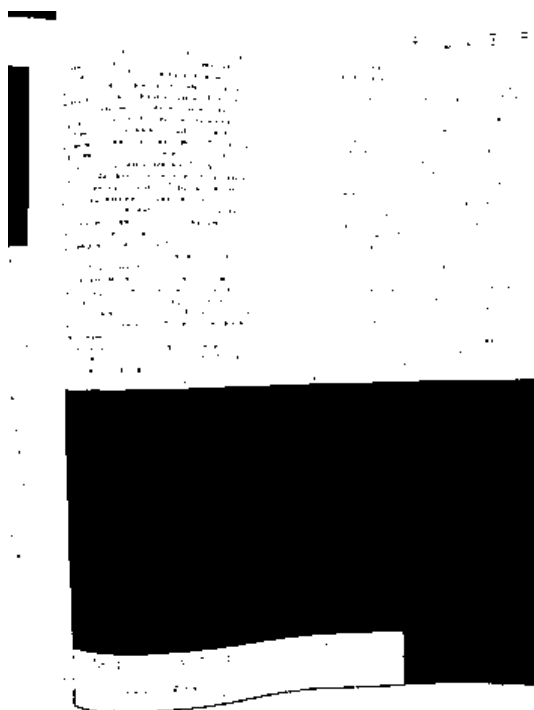
Fig. 8. Original document is seen under some classical mathematical operations.

(a) Segmentation obtained by an unconstrained watershed

(b) Segmentation obtained by a constrained watershed

(c) Segmentation obtained by an constrained watershed : The biggest uniform region is probably background

(d) Final segmentation obtained by an constrained watershed after an appropriate filtering

Fig. 9. Separation text-non text information in a document-image.

5 Numerical results

In this section, we present some simulation results showing the performance of the presented tool. We show the ability of the proposed process to successfully scan document-image acquired by camera in a tough environment.

We show how the step of segmentation is important is we would like to scan complex document-images, for example document acquired by a photophone from magazines or when we we have document-images containing text and images.

In all examples, the original images are acquired from a Sony Erickksson K800i (3.2 megapixels).

Figure 11 shows an example of the dewarp-text enhancement by methods proposed in this paper. 0ne can see that page corners are not necessary in the image, only edges are required to detect frames.

Figure 12 shows the performance of our process to provide scanned quite complex documents. Here, after a segmentation which separates the text and images in the document-image, the texts in the document are enhanced by the method that we have proposed in [6]. As regards the "non text" part of the document, the images are enhanced by a retinex processing.

In figure 13, the acquired document needs all processes, namely a dewarp, a segmentation to separate text and image in the document, an enhancement for each parts of the segmented document-image., then a Poisson Image Editing (see [14]) in order to merge both parts and to reconstruct a coherent document. This test is very interesting and our process has not only correctly cropped the document but the image enhancement give a good result.
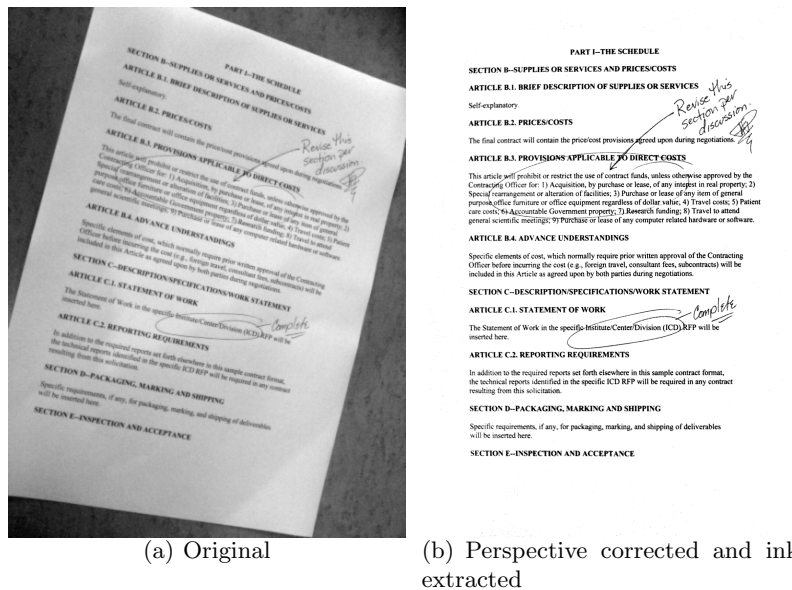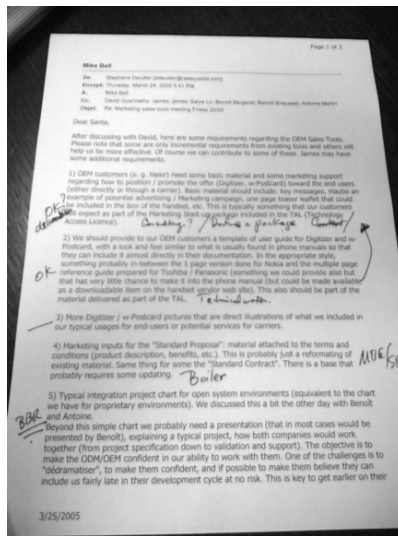

(a) Original
(b) Perspective corrected and ink extracted

Fig. 10. Example - Notice that page corners are outside the bounds of the image.

In figures 14, 15 and 16, we present some results for blackboard or whiteboard images. This use-case could be interesting for students and professionals who want to capture quickly informations with their cameraphones. For example students could have the slides of their course or when the meeting is over, but the whiteboard is chock full of lists, notes, diagrams, and to-do's you didn't have time to jot down yet. Instead of manually transcribing it all, you could pull out your cameraphone and snap a picture. Even if these images are very noisy (due to the acquisition process) our process gives good results, but in term of readabilty/printability.

6   Conclusion

In this paper, we present a tool for scanning document-images. We have described five steps to obtain a correct scan. So far we have shown that our process remains robust even in rought environment and in low lightning conditions. Moreover, the resulting process depends only on few fixed parameters. Moreover, we proved that this process allows to estimate a scanning version of a document-image acquired from a cameraphone in a robust and reproductible way.

(a) Original
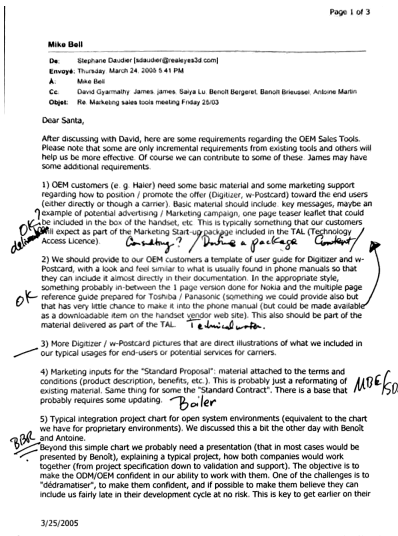
(b) Perspective corrected and ink extracted

Fig. 11. Original document is seen under some perspective which makes it slanted. Text lines become parallel in the corrected document.

(a) Example 1 : original image

(b) Result without segmentation

(c) Result after segmentation

Fig. 12. Original image acquired from a Sony Ericsson K800i
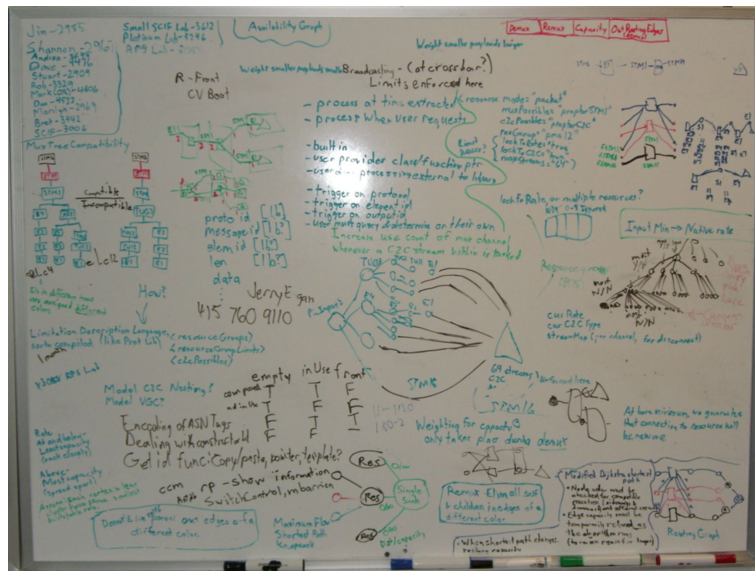
(a) Example 2, original image

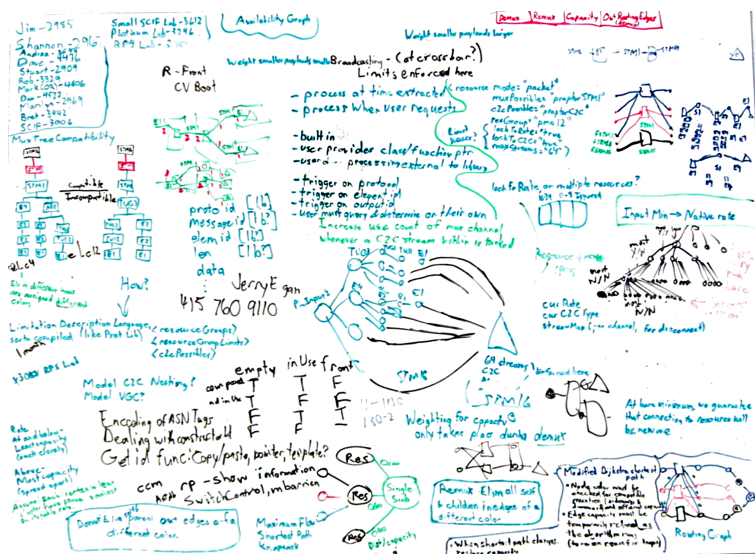(b) Result without segmentation but with dewarp

(c) Result with segmentation and dewarp

Fig. 13. Original image acquired from a Sony Ericsson K800i
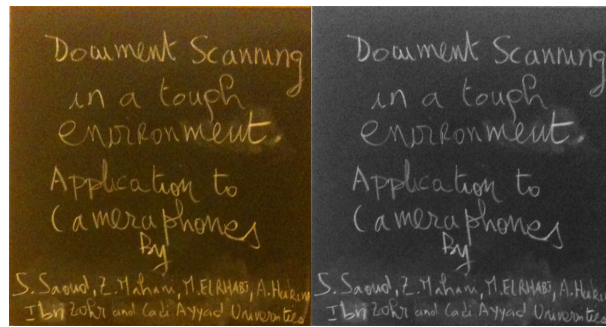
(a) Example 3 : original image
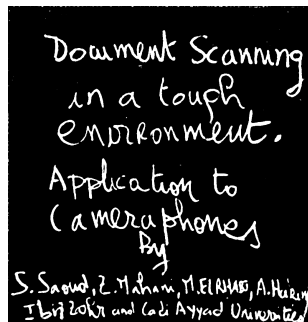

(b) Color Ink extracted and dewarped

Fig. 14. Original acquired from a Casio Computer co. ltd-Digital still camera

(a) Example 4 : original blackboard-image

(b) The (original) cropped image

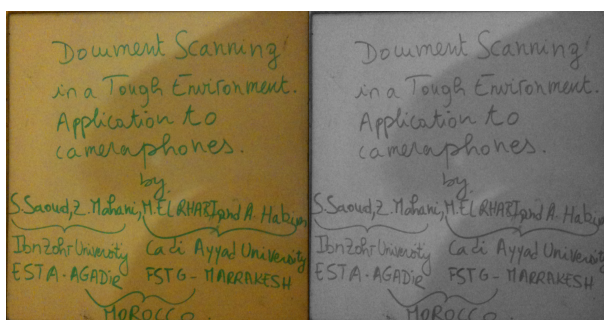(c) the (original) cropped image in grayscale

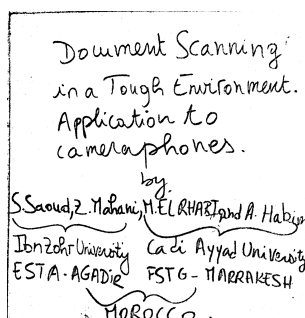(d) Text enhancement : estimated reflectance ; model **??**

Fig. 15. Original image acquired from an Iphone 3GS: Blackboard test

(a) Example 5 : original Whiteboard-image

(b) The (original) cropped image

(c) The (original) cropped image in grayscale

(d) Text enhancement : estimated reflectance ; model **??**

Fig. 16. Original image acquired from an Iphone 3GS: Whiteboard test

References

1. Henry S. Baird, *The State of the Art of Document Image Degradation Modeling*, In Proc. of 4 th IAPR International Workshop on Document Analysis Systems, Rio de Janeiro, pp 1-16, 2000.
2. I. Nwogu, Z. Shi and V. Govindaraju, *PDE-Based Enhancement of Low Quality Documents*, In Proceedings of the Ninth international Conference on Document Analysis and Recognition - Vol.01, pp.541-545 (2007).
3. F. Drira, F. Le Bourgeois and H. Emptoz, *Document images restoration by a new tensor based diffusion process : Application to the recognition of old printed documents*, 10th International Conference on Document Analysis and Recognition (ICDAR09), Barcelone, pp. 321-325, 2009.
4. A. Chambolle, *An algorithm for total variation minimization and applications*, Special issue on mathematics and image analysis, J. Math. Imaging Vision, Vol. 20, N.1-2, pp 89-97, 2004.
5. Z. Mahani and J. Zahid and S. Saoud and M. El Rhabi and A. Hakim, *Text Enhancement by PDE's Based Methods*, Lecture Notes in Computer Science, 2012, Volume 7340, Image and Signal Processing, pp 65-76, 2012.
6. S. Saoud and Z. Mahani and M. El Rhabi and A. Hakim, *Document Scanning in a tough environment : application to cameraphones*, International Journal of Imaging & Robotics (IJIR), accepted, 2012.
7. R. F. Moghaddam and M. Cheriet, *RSLDI: Restoration of single-sided low-quality document images*, Pattern Recognition, Special Issue on Handwriting Recognition, no. 42, pp. 3355-3364 (2009).
8. B. K. Horn, Robot Vision, MIT Press, 1986.
9. J. Kim and H. Lee, *Joint nonuniform illumination estimation and deblurring for bar code signals*, Optic Express, Vol. 15, issue 22, pp. 14817-14837, 2007.
10. L. Dumas, M. El Rhabi and G. Rochefort, *An evolutionary approach for blind deconvolution of barcode images with nonuniform illumination*, IEEE Congress on Evolutionary Computation, pp. 2423 - 2428, 2011.
11. M. El Rhabi and G. Rochefort *Method of restoring a blurred image acquired by means of a camera fitted to a communication terminal*, Realeyes3D SA, patent : http://www.wipo.int/patentscope/search/en/WO2009112710, 2009.
12. R. Beare, A Locally Constrained Watershed Transform, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, pp. 1063-1074, 2006.
13. Zia-ur Rahman and Glenn A. Woodell, *Retinex processing for automatic image enhancement*, Journal of Electronic Imaging, Vol. 13, pp. 100-110, 2004.
14. P. Pérez, M. Gangnet and A. Blake, *Poisson image editing*, ACM Transactions on Graphics (SIGGRAPH'03), 22(3):313-318, 2003.