Novel Approach to Extract Lines of Documents for Blind and Impaired People

A. Gari
LIPI/ ENS
USMBA/ FEZ,
MOROCCO

G. Khassidi
LIPI/ ENS
USMBA/ FEZ,
MOROCCO

M. Mrabti
LIPI/ ENS
USMBA/ FEZ,
MOROCCO

ABSTRACT
Documents acquired by the blind and visually impaired people always have a degree of inclination, overlap between the characters and 3D geometric transformations. To solve these problems, firstly, a detection and correction of the tilt of the document is applied by the Hough Transform (HT) on the interest points detected on the document image. After separating the texts zones from the other objects, the partial projection is used to estimate the beginnings of lines for every text zone. Finally, extraction the lines of documents is done by following the baselines. The pro-posed method tested on the database used for the ICDAR2015 competition leads to better results for line extraction.

General Terms
Pattern Recognition.

Keywords
Blind and visually impaired people; Lines segmentation; Partial projection; Hough Transform; interest points.

1. INTRODUCTION
The goal of our research is to develop a portable system to assist blind and visually impaired people in their day-to-day routine and give them a wellness. Such a system can recognize texts, objects and faces, and convert texts to a speech. Actually, some portable systems/ devices are realized (OrCam, Doro 820 Mini Claria, Esight). Optical Character Recognition (OCR) is one of the most successful applications of automatic pattern recognition used by these systems. The majority of the existing OCRs present a high recognition accuracy, but in the presence of some geometric transformations like rotation and overlapping characters, the recognition rate decreases (Free OCR, Docs Matter, Goggles, OCR Instantly Free, Scan Text, Smart OCR: Text Miner, The simple OCR, etc). To solve these problems, a new approach is proposed in this paper based on inclination correction and lines segmentation.

The segmentation studies of text to lines are based on a decomposition of the image into connected components [1,2] or on a histogram partial projection [3,4]. Other authors use different methods of text line segmentation [5,6,7,8].

Most of these approaches present limitations to solve the problem of inclination, the overlapping characters and the 3D geometric transformation simultaneously. Our proposed method which allows tilt correction, text areas detection and lines segmentation is illustrated by the following synoptic schema:

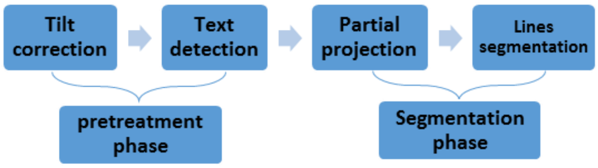


Fig 1: Synoptic schema of the proposed method

The paper is organized as follows: Section 2 treats the correction of inclination. Section 3 deals with the proposed line segmentation method. In section 4, the experimental results is presented and, finally, section 5 describes conclusions and future works.

2. CORRECTION OF INCLINATION
Images acquired by blind or partially sighted people are not generally aligned, and therefore have inclinations in relation to a given coordinate system. Even more, the outer edges of images can contain objects that influence the correction of inclination.

This correction is an essential step before the detection, segmentation and text recognition phases. It is performed using the HT (Hough Transform) [9]. However, the application of this transform on the entire image requires a long computation time. A method proposed in [10] consists of applying the HT on the points extracted by the Haar wavelets transform which reduces the number of pixels to be processed. The correction of inclination remains unsatisfactory.

In the proposed approach, the Hough transform is applied to the interest points which are chosen as the centroids of components connects (CC) Fig 1. Each gravity center (xi,) corresponds to a set of cells in the accumulator array H. Then, the rotation skew is the rotation corresponding to the maximum of the accumulated cell array.

$$[\rho,\theta] = \arg Max_{\rho_i,\theta_i} H(\rho_i,\theta_i)$$

with:

$\theta_i \in [0,180]$.

$\rho_i = x_i \cos\theta + y_i \sin\theta$.

y_i *and* y_i: *coordonnées de chaque pixel* p_i.

H: accumulated cell array.

The barycenter of a CC is calculated as follows:

$$\frac{\sum_{i=1}^{n} I_i(x,y) * p_i(x,y)}{\sum_{i=1}^{n} I_i(x,y)}$$

with:

$p_i(x, y)$: pixel coordinates.

$I_i(x, y)$: pixel intensity.

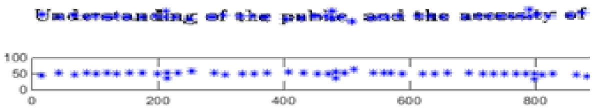


Fig 2: Representation of every CC by its centroids

This algorithm is tested on the database used by ICDAR2015 [11]. The obtained results compared to Canny and Haar methods show that the correctness factor of inclination approach is 97.54% Table 1.

Table 1. Correctness factor of inclination by 3 different methods

Method	Percentage of tilt correction	Execution time (s)
CANNY	50.63%	6.75
HAAR	82,39%	0.48
Centroids of CC.	97.54%	1.08

3. THE PROPOSED LINE SEGMENTATION METHOD

3.1 Text Detection

This phase starts by the text detection followed by the global binarization, morphological treatments and finally local binarisation.

The text detection is done by applying the algorithm of LeBourgois [12] described as follows:

$$I(x, y) = \sum_{i=-t}^{t} \frac{\partial I}{\partial x}(x + i, y) \ , t \in \mathbb{N}$$

C.Wolf [12] showed that the horizontal Sobel filter gives better results, and the parameter t is selected such that 2t+1=13. Normalization by 13 improves the results of text detection [13].

The global binarisation is based on OTSU method [14]. It performs a global automated threshold in the form of the histogram of image. This iterative algorithm, calculates the optimal threshold T by minimizing the intra-class variance(σ_w).

$$\begin{cases} \min_{T \in [0,255]} (\sigma_w^2(T)) \end{cases}$$

with :

$$\sigma_w^2(T) = \omega_1(T) \times \sigma_1^2(T) + \ \omega_2(T) \times \sigma_2^2(T)$$

$\omega_1(T)$and$\omega_2(T)$are the probability of being in class 1 and class 2 respectively.

$\sigma_1^2(T)$ ($\sigma_2^2(T)$) is the variance of class 1(class 2).

Morphological treatments are used to extract rectangles encompassing text boxes and to reduce noise. The remaining errors will be corrected by the squareness method:

$$\begin{cases} \dfrac{logical\_area(C_i)}{real\_area(C_i)} < 70\% C_i = 0 \\ \dfrac{logical\_area(C_i)}{real\_area(C_i)} \geq 70\% C_i = C_i \end{cases}$$

Where $C_i$is a connected component.

Finally, the extraction of the text from the detected zones is realized by local binarisation using the NICK algorithm [15]:

$$T_{NICK} = m + k \sqrt{\frac{\sum p_i^2}{Nbr\ pix}}$$

with:

T_{NICK}: the calculated threshold.

m: the average value of the pixels into the sliding window.

k: Niblack factor k=0.4.

$Nbrpix$: the number of pixels in the window.

p_i : the intensity of each pixel in the sliding window.

Results of detection of the text zones is shown in Fig 3. As we can see, a zone can contain one or many lines.

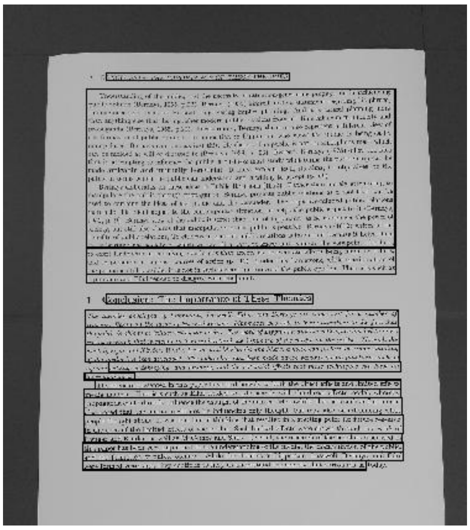


Fig 3: Results of detected text zones

3.2 Base Line Segmentation

This method starts by applying the partial projection to determine the initial points of the contour tracking. The line segmentation begins from the middle of two successive lines$L$and$L + 1$, and follows the external edge of related components$C_i$from left to right until it touches a black pixel. The tracking correction is done by taking into account the convex hull of$C_i$as described next.

3.2.1 *Partial projection*

The partial projection is a popular method based on the histogram computing [4,16,17]. It is used in the segmentation lines to determine the beginning, the end, the height, the number of lines in the text zone $Z_i$and the distance between the line L and the line $L + 1$. In this case, the width of columns can be variable and depends on the size of the text zone. An illustration example in Fig 4 is given and the results are presented in Table 2.

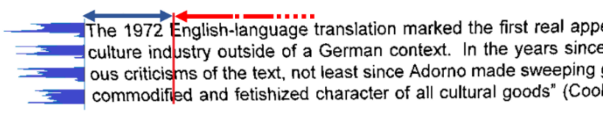


Fig 4: Partial projection on a part of a text zone

Table 2. Results of partial projection

Beginning	End	Height	Distance
42	80	39	23
103	140	38	22
162	199	38	22
221	258	38	-

3.2.2 Segmentation between two lines

Notations:

Z_i : text zone, with i=1:n number of zones in a document.

$L_{i,j}$: line which belongs to Z_i, with j=1: l number of lines in Z_i .

$beg_{i,j}$: initial value of the $tracker_{i,j}$ wich represents the middle of lines $L_{i,j}$ and $L_{i,j+1}$.

k : number of columns in Z_i .

$tracker_{i,j}$: segment two lines $L_{i,j}$ and $L_{i,j+1}$ of a zone Zi.

Dist_haut (Dist_bas): distance between the pixel $tracker_{i,j}(x, y)$ and the min_rows(max_rows) of C_i, Fig 5.

Dist_ligne : half distance between 2 lines $L_{i,j}$ and $L_{i,j+1}$ of Z_i.

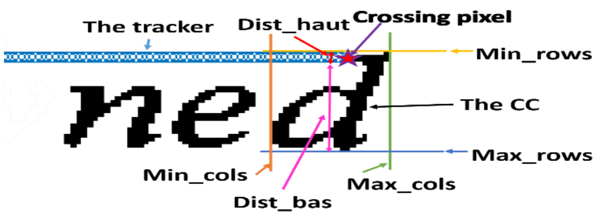


Fig 5: Boundaries values extraction of the CC

Proposed algorithm:
The proposed approach for segmentation lines is described in Fig 6:

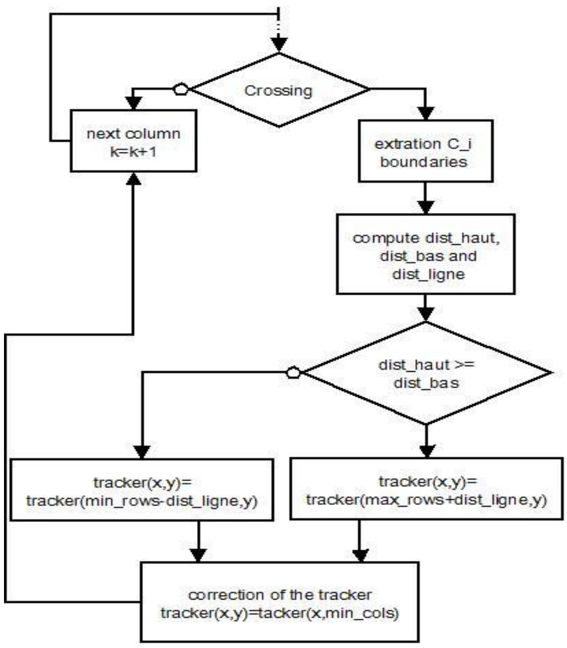


Fig 6: Correction process of the tracker

In the absence of overlapping between two successive lines, the tracker is shown as in Fig 7:

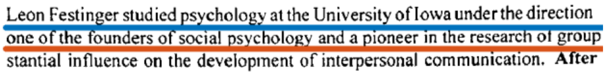


Fig 7: Example of segmentation without collision

In the case of a crossing between the tracker and C_i, the decision to move-up or down can be taken depending on the comparison and calculation of two distances, Dist_haut and Dist_bas, then, the correction process of the tracker will be applied as figured in Fig 6.

$$Dist\_haut = crossing\_pixel - \min(rows)$$

$$Dist\_bas = crossing\_pixel - \max(rows)$$

An example of the tracker rectification is presented in Fig 8.

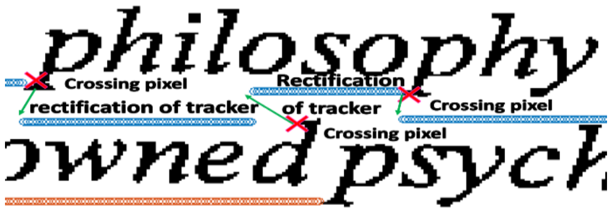


Fig 8: Rectification of the tracker

4. EXPERIMENTAL RESULTS

To illustrate the effectiveness of the developed method, the ICDAR2015 database "Competition on Smartphone Document Capture and OCR (SmartDoc)" [11] is used.

Tests were prepared after a manual calculation of zones and lines of many documents. They were performed on a PC equipped with a microprocessor Core I5 -GHz 2.67 and 4 GB RAM under Windows 7. The application is developed with MATLAB R2014b. In the detection of mono or multi lines zones, we achieved in our tests a detection around 938 and 191 zones respectively. The percentage of lines in multi-lines zone rate is 56.13%.

The Precision and Recall of text line segmentation using our proposed method is summarized in Table 3. The undetected or misdetected lines are due to the algorithm used in the detection of zones. The effectiveness of our algorithm is tested on 3636 images which present different inclinations. An example is given in Fig 9.

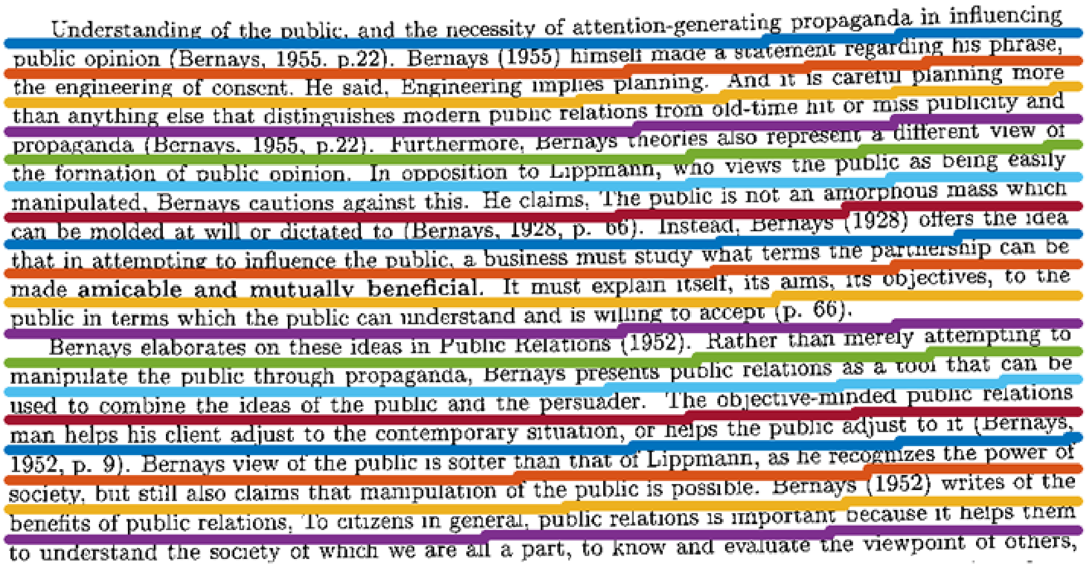
Table 3. Precision and Recall using text-line segmentation algorithm.

№ of lines	% of mono lines in zones	% of multi lines in zones	Precision	Recall
2138	43,87%	56,13%	97,58%	95,01%

5. CONCLUSION AND PERSPECTIVES

In this article, an original method had been proposed, which aims to extract the lines from documents, and it was tested on the ICDAR2015 database. Initially, a detection and correction of inclination is applied by HT on centroid of CC. For the lines segmentation, the proposed algorithm based on the partial projection is applied. The successive application of the partial projection after the inclination correction leads to better results with high robustness and performance in the case of overlap-ping lines.

The rate of precision and recall, shows the performance of our approach. The next step of work is to improve the extraction rate and generalize the approach to other languages.



Understanding of the public, and the necessity of attention-generating propaganda in influencing public opinion (Bernays, 1955, p.22). Bernays (1955) himself made a statement regarding his phrase, the engineering of consent. He said, Engineering implies planning. And it is careful planning more than anything else that distinguishes modern public relations from old-time hit or miss publicity and propaganda (Bernays, 1955, p.22). Furthermore, Bernays theories also represent a different view of the formation of public opinion. In opposition to Lippmann, who views the public as being easily manipulated, Bernays cautions against this. He claims, The public is not an amorphous mass which can be molded at will or dictated to (Bernays, 1928, p. 66). Instead, Bernays (1928) offers the idea that in attempting to influence the public, a business must study what terms the partnership can be made amicable and mutually beneficial. It must explain itself, its aims, its objectives, to the public in terms which the public can understand and is willing to accept (p. 66).

Bernays elaborates on these ideas in Public Relations (1952). Rather than merely attempting to manipulate the public through propaganda, Bernays presents public relations as a tool that can be used to combine the ideas of the public and the persuader. The objective-minded public relations man helps his client adjust to the contemporary situation, or helps the public adjust to it (Bernays, 1952, p. 9). Bernays view of the public is softer than that of Lippmann, as he recognizes the power of society, but still also claims that manipulation of the public is possible. Bernays (1952) writes of the benefits of public relations. To citizens in general, public relations is important because it helps them to understand the society of which we are all a part, to know and evaluate the viewpoint of others,

Fig 9: Example of the proposed line segmentation

6. REFERENCES

[1] Likforman-Sulem, L., and Faure, C.1994. Une Méthode de Résolution des Conflits d'Alignements pour la Segmentation des Documents Manuscrits. In: 3ème Colloque National sur l'Ecrit et le Document. CNED'94, pp. 265-272. French, Rouen

[2] Likforman-Sulem, L., and Faure, C. 1994. Extracting Text Lines In Handwritten Documents By Perceptual Grouping. In: Advances in Handwriting and Drawing : A Multidisciplinary Approach, Faure, C., Keuss, P., Lorette, G., Winter, A., pp. 21-38. Paris

[3] Venturelli, F. 1997. Successful Technique For Unconstrained Hand-Written Line Segmentation. In: Progress in Handwriting Recognition, World Scientific, pp. 563-568

[4] Bennasri, A., zahour, A., and taconet, B. 1999. Extraction Des Lignes d'un Texte Manuscrit Arabe. In : Vision Interface. VI'99, pp. 41-48. Canada, Trois-Rivières

[5] Roy, P. P., Pal, U., and Lladós, J. 2008. Morphology Based Handwritten Line Segmentation Using Foreground and Background Information. In: International Conference on Frontiers in Handwriting Recognition.ICFHR,pp. 241-246. Montreal

[6] Ramappa, M. H., and Krishnamurthy, S. 2012. Skew Detection, Correction and Segmentation of Handwritten Kannada Document. In: International Journal of Advanced Science and Technology.vol. 48, pp. 71-88

[7] Ouwayed, N., and Belaid, A. 2009. une Approche Génerale pour l'Extraction des Lignes des Document Arabes Anciens Multi Orienté. In: 12eme Colloque International sur le Document Electronique. Montréal

[8] Ouwayed, N., and Belaid, A. 2009. Separation of Overlapping and Touching Lines within Handwritten Arabic Documents. In: Computer Analysis of Images and Patterns, 13th International Conference. CAIP,pp. 237-244. Münster

[9] Abdat, A., Khorissi, N., Sehad, A., Mezai, L., Benselama, Z. A., and Namane, A. 2005. Detection de l'Inclinaison de l'Ecriture dans les Documents Arabes Imprimes. In: Colloque National sur le Traitement du Signal et ses Applications. CNTSA'05. Guelma

[10] Choi, H., and Baraniuk, R. G. 2001. Multiscale image segmentation using wavelet-domain hidden Markov models.In:IEEE Transactions on Image Processing. Vol. 10, pp. 1309 – 1321

[11] Burie, T. C., Chazalon, J., Coustaty, M., Eskenazi, S., Luqman, M. M., Mehri, M., Nayef, N., OGIER, J.-M., Prum, S., and Rusinol, M. 2015. Competition on Smartphone Document Capture and OCR (SmartDoc). In 13th International Conference on Document Analysis and Recognition. ICDAR2015

[12] Wolf, C., Jolion, J. M., and Chassaing, F. 2001. Détection et Extraction de Texte de la Vidéo.In : 7èmes Journées d'Etudes et d'Echanges, COmpression et REprésentation des Signaux Audiovisuels. CORESA 2001, pp. 251-258. Dijon

[13] Gari, A., Khaissidi, G., Mrabti, M., Chenouni, D., and El Yacoubi, M.: OCR for Blind and Visually Impaired. 2015. In: Advanced Information Technology, Services and Systems AIT2S2015. pp. 331-335. SETTAT

[14] Otsu, N.: A Threshold Selection Method from Gray-Level Histogram. 1979. In: IEEE Transactions on System Man Cybernetics. Vol. 9, pp. 62-66

[15] Khurshid, K., Siddiqi, I., Faure, C., and Vincent, N. 2009. Comparison of Niblack inspired Binarization Methods for Ancient Documents. In: SPIE Proceedings, Document Recognition and Retrieval XVI.Vol. 7247

[16] Sakhi, O. B: Segmentation of heterogeneous document images. 2012. an approach based on machine learning, connected components analysis, and texture analysis.Université Paris-Est. PARIS

[17] Garg, R., and Garg, N. K. 2014. An algorithm for Text Line Segmentation in Handwritten Skewed and Overlapped Devanagari Script.In: International Journal of Emerging Technology and Advanced Engineering.*IJETAE*, vol. 4(5), pp. 114-118