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

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 proposed 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.

$$\theta_i = \arg \max_{\rho, \theta} H(\rho, \theta)$$

with:
$$\theta_i \in [0, 180]$$.
$$\rho_i = x_i \cos \theta + y_i \sin \theta$$.
$$y_i, x_i: coordonnées de chaque pixel p_i.$$
$$H: accumulated cell array.$$

The barycenter of a CC is calculated as follows:
$$barycenter = \frac{\sum_{i=1}^{n} l_i(x,y) \cdot p_i(x,y)}{\sum_{i=1}^{n} l_i(x,y)}$$

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage of tilt correction</th>
<th>Execution time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANNY</td>
<td>50.63%</td>
<td>6.75</td>
</tr>
<tr>
<td>HAAR</td>
<td>82.39%</td>
<td>0.48</td>
</tr>
<tr>
<td>Centroids of CC</td>
<td>97.54%</td>
<td>1.08</td>
</tr>
</tbody>
</table>

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(\( \sigma_w \)).

\[
\min_{\{T \in [0,255]\}} \left( \sigma_w^2(T) \right)
\]

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.

\[
\sigma_i^2(T) (\sigma_w^2(T)) \text{ is the variance of class } 1 \text{ (class } 2).\]

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

\[
\begin{align*}
\frac{\text{logical}_i}{\text{real}_i} & < 70\% C_i = 0 \\
\frac{\text{logical}_i}{\text{real}_i} & \geq 70\% C_i = C_i
\end{align*}
\]

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.

\( Nbr \ pix \): 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.

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 \( L+1 \), 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 3: Results of detected text zones

Fig 4: Partial projection on a part of a text zone
### 3.2.2 Segmentation between two lines

**Notations:**

- \( Z_i \): text zone, with \( i = 1:n \) number of zones in a document.
- \( L_{ij} \): line which belongs to \( Z_i \), with \( j = 1: l \) number of lines in \( Z_i \).
- \( beg_{ij} \): initial value of the tracker \( T_{ij} \) which represents the middle of lines \( L_{ij} \) and \( L_{ij+1} \).
- \( k \): number of columns in \( Z_i \).
- \( Dist_{haut} \) (\( Dist_{bas} \)): distance between the pixel tracker \( (x,y) \) and the min_rows(max_rows) of \( C_i \), Fig 5.
- \( Dist_{ligne} \): half distance between 2 lines \( L_{ij} \) and \( L_{ij+1} \) of \( Z_i \).

**Proposed algorithm:**

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

1. **Extraction of zones:**
   - Compute \( Dist_{haut} \) and \( Dist_{bas} \) of zones.
   - Compute \( Dist_{ligne} \) of the lines.

2. **Correction process of the tracker:**
   - Initial value of the tracker \( T_{ij} \).
   - Compute \( Dist_{haut} \) and \( Dist_{bas} \) of the lines.
   - Correction of the tracker by applying the algorithm.

### Table 2. Results of partial projection

<table>
<thead>
<tr>
<th>Beginning</th>
<th>End</th>
<th>Height</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>80</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>103</td>
<td>140</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>162</td>
<td>199</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>221</td>
<td>258</td>
<td>38</td>
<td>-</td>
</tr>
</tbody>
</table>

### 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 - 2.67 GHz 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 rate of 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 lines. The effectiveness of our algorithm is tested on 3636 images which present different inclinations. An example is given in Fig 9.

### 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 overlapping 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.
6. REFERENCES


[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


Fig 9: Example of the proposed line segmentation