Simultaneous detection of vertical and horizontal text lines based on perceptual organisation

Claudie Faure [a], Nicole Vincent[b]
[a] CNRS-LTCI, TELECOM-ParisTech, 46 rue Barrault, 75634 Paris, Cedex 13, France
[b]CRIP5 - Université Paris Descartes, 45, rue des Saints-Pères, 75270 Paris Cedex 06, France

ABSTRACT

A page of a document is a set of small components which are grouped by a human reader into higher level components, such as lines and text blocs. Document image analysis is aimed at detecting these components in document images. We propose the encoding of local information by considering the properties that determine perceptual grouping. Each connected component is labelled according to the location of its nearest neighbour connected component. These labelled components constitute the input of a rule-based incremental process. Vertical and horizontal text lines are detected without prior assumption on their direction. Touching characters belonging to different lines are detected early and discarded from the grouping process to avoid line merging. The tolerance for grouping components increases in the course of the process until the final decision. After each step of the grouping process, conflict resolution rules are activated. This work was motivated by the automatic detection of Figure&Caption pairs in the documents of the historical collection of the BIUM digital library (Bibliothèque InterUniversitaire Médicale). The images that were used in this study belong to this collection.

Keywords: Historic documents, Figure-Caption pairs, Perceptual grouping, Text line detection

1. INTRODUCTION

Typography and layout are defined to facilitate the perceptual organisation leading to the salience of the page components. The spatial organisation helps the reader to detect the textual components at several levels (alphabetical symbols, words, lines ...). It is also responsible of implicit links between components leading to group a figure and its caption or a text area with its title. Human perception is considered as a source of inspiration for several methods in document image analysis. Banks of Gabor filters are used to mimic the human filtering mechanisms to discriminate textures. This approach is mainly used to detect text areas [2, 10]. For text line detection, low resolution images are used in [6] to reduce the text lines to linear segments as they are perceived when we wink. The Gestalt laws of perceptual organisation have been taken as a model to detect automatically meaningful components and spatial relationships in a document or in drawings [5, 12, 13]. In [7], the alignments of connected components are detected with the Hough transform, they are sorted according to the rule of proximity that is responsible of the salience for some alignments. Besides the approaches centred on grouping, other methods are aimed at detecting separators between components. Projections profiles or more elaborated methods to detect white separators have been proposed [1, 3, 11].

Text lines are the typical components of the written language. They are between the word and the text bloc levels. When they are well segmented, columns and margins are also well segmented. The RLSA algorithm is certainly the most popular to extract text lines. Nevertheless, it cannot answer all the problems encountered and several methods were proposed to increase the performance of text line extraction, among them the recent methods proposed for printed text [4, 8]. The detection of text lines is a challenging problem for the simulation of the human perception of global organisation from local information.

The proposed method starts from the local physical components of a page image and groups them according to the properties that enable a human reader to detect alignments of symbols. The main properties are the proximity, the similarity and the continuity of direction. The connected components (CCs) of the binarised image are early discriminated. A size criterion is used to interpret the greatest CCs as Graphics; they are labelled CCG and are discarded from the grouping process leading to text lines. The borders of the CCG bounding boxes are separators that text lines cannot straddle. The remaining CCs are labelled according to the position of their nearest neighbour in order to capture the proximity and alignment properties. They are the input of a rule-based incremental grouping process. This stepwise

method takes advantage of emerging organisation. Therefore, the spatial information involved in the grouping rules is not reduced to the local information between CCs. After each step of the grouping process, previously detected alignments are reinforced or eliminated. This enables to solve the conflicts that may appear when text lines are simultaneously detected in different directions.

The origin of this work is a collaboration with a digital library (BIUM: Bibliothèque InterUniversitaire Médicale [14]). The goal is to facilitate the task of the persons who perform the indexing and the storage of numerical data. Graphics are a kind of content that appears as a priority for the readers, they want to find specific figures, to locate them in the books and to visualise them. The archivists of the BIUM did not consider that the automatic detection of the figures in a book was helpful, their expected the detection of the figures with the associated captions. Therefore, our goal became the detection of Figure&Caption pairs in the document images. The pages of the BIUM collection may contain horizontal and vertical text lines. It is not possible to anticipate the direction of the caption lines from the dominant direction in a page. This observation motivated the presented study. The proposed method detects text lines without prior assumption on their direction. The collection of the BIUM contains ancient books that do not follow the rules of modern typography and layout. Many characters belonging to different lines are touching, leading (with RLSA) to merge the lines. The labelled CCs that overlap several text lines are early detected and discarded from the grouping process. We also observed that the caption of a figure might be included in the rectangular bounding box of the figure. This lead us to consider the CCs that are inside the CCG bounding boxes when performing text line detection.

The next section describes the preprocessing step. Section 3 explains how to label the CCs. The rules involved in the grouping process are given in Section 4. The procedures activated to detect and solve conflicts are in Section 5. Section 6 gives the performances of the text line detection method for 102 caption lines and 52 pages gathered from 22 books and containing horizontal and vertical text lines.

2. PREPROCESSING

Text line detection by grouping connected components (CCs) requires the binarisation of the original grey image to obtain the CCs. The binarisation is performed using an algorithm that is inspired from Niblack's formula [9]. It has the advantage to improve the binarisation of "white" and light page images by shifting down the binarisation threshold for these pages. The proposed formula is:

$$T = m + k\sqrt{\frac{\sum p_i^2 - m^2}{NP}} = m + k\sqrt{A}$$

where *NP* is the number of pixels in the image, *m* is the average grey value and *k* is a coefficient (k = -0,2).

The CCs of the binary image are sorted to discard the graphical CCs from the text line detection process. Graphics segmentation is a difficult problem. In this study, only a size criterion is used to label CCG the CCs that are interpreted as Graphics. Let BBX and BBY the border sizes of a CC bounding box. The average BBX and BBY in a page are respectively MCCX and MCCY. A CCG satisfies the condition:

((BBX > 10 * MCCX) OR (BBY > 10 * MCCY)) OR ((BBX > 5 * MCCX) AND (BBY > 5 * MCCY))

Text parts may be included into the rectangular bounding box of a CCG; this is the case in Fig. 2 but also for text in the cells of tables. This leads us to consider the CCs included in a CCG as potential text during the grouping process. The borders of the CCGs are separators that a text line cannot straddle. This condition enables the detection of text lines inside the bounding box of a CCG. It also reduces the risk to merge small CCs belonging to Graphics with text lines that are outside the CCG bounding boxes.

3. LABELLED CONNECTED COMPONENTS

Since the beginning of the written language, the evolution of the writing rules is aimed at facilitating the reading. Proximity is prior in the perception of salient alignments (rows or columns in Fig. 1a,b). Since the historical introduction (in the 6th century) of a white space between words greater that the space between characters, proximity became involved in several typographic rules to take advantage of the visual organisation principles. As an example, the recommendation to have a distance between the lines greater than the distance between the words of a line.

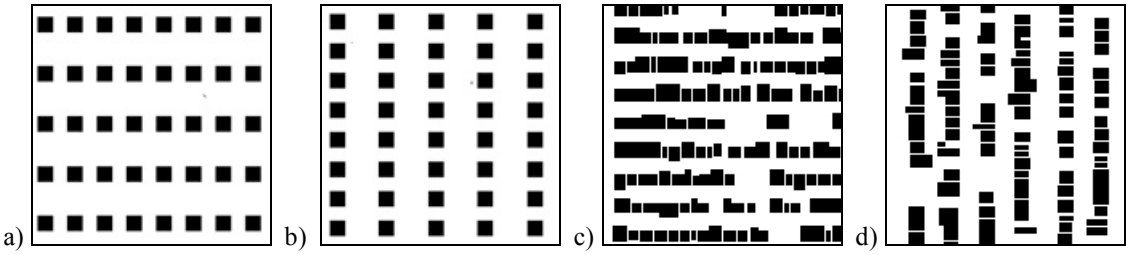


Fig. 1. a) The nearest neighbours of each black square belong to the perceived row (a) or column (b). Filled in black, the NNE bounding boxes in horizontal text lines (c) and the NNS in vertical text lines (d).

Our method is based on perceptual organisation. We propose to early discriminate the CCs according to the proximity property. For each CC, the closest CC is searched on its right, as well as the closest CC under it. A CC is labelled NNE (E for East) when its nearest neighbour is on its right; it is labelled NNS (S for South) when its nearest neighbour is under the CC. The CCs on the right of a page have no East neighbour; they are labelled NNR (R for right). The CCs at the bottom of a page have not South neighbour, they are labelled NNB (B for Bottom). The neighbourhood considered to label a CC is limited to East and South, the reading order adopted in occidental writings. The North and West neighbours are ignored for the spatial labelling. This has the advantage to simplify the data representation and to avoid coping with ambiguous situations such as: CC2 is the East nearest neighbour of CC1 but CC1 is not the West nearest neighbour of CC2. The disadvantage is an incomplete knowledge on spatial relationships between the CCs. It may be a source of errors (discussed further) that are scarcely encountered.

The spatial labelling associates a visual property to the CCs leading to differentiate them in an early stage of the process. The labelled CCs give a first information on the local direction of text. Fig. 1c and 1d show that the row and column organisation can be perceived when the NNEs of horizontal lines and the NNSs of vertical lines are displayed. The NNE density is greater in the horizontal text areas while the NNS density is greater in the vertical text area, as seen in Fig. 2b,c.

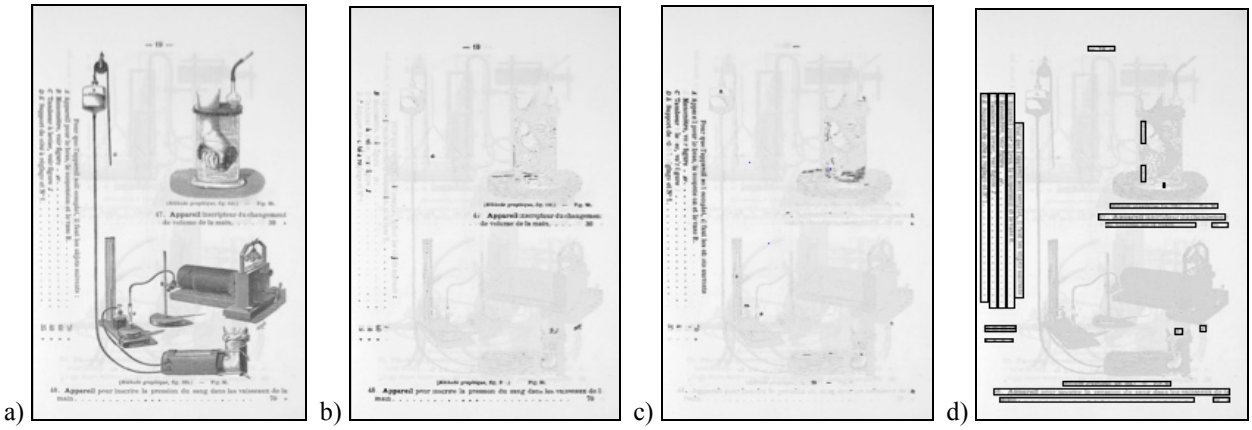


Fig. 2. a) Original. b) CCs labelled NNE c) CCs labelled NNS d) Detected text lines.

Touching characters that belong to different text lines are responsible of line merging with RLSA. They are early detected to be discarded from the grouping process. The CCs that are the nearest neighbour of more than one CC along a direction are searched. If a CC is the nearest neighbour of more than one NNE (or NNS), its label becomes EM (or SM). These labels exclude the CCs from the grouping process and reduce the risk of merging text lines. See examples in Fig. 3 where the borders of the detected text lines are visualised and the touching characters are displayed in black. The early detection of touching characters is efficient. Nevertheless, it has a drawback resulting from limiting the search of the CC's nearest neighbours to the East and South directions. This prevents to detect a CC overlapping several text lines when it is the leftmost or the topmost CC in a page. As a consequence, errors in text line segmentation may be encountered, see an example in Fig. 3c.

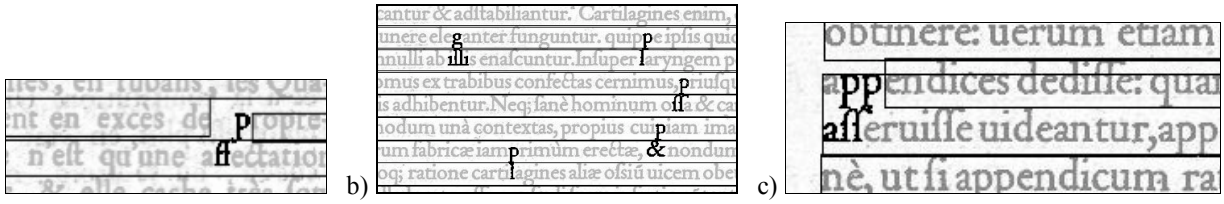


Fig. 3. Detected text lines. In black, the CCs that are the nearest neighbour of several CCs and correspond to touching characters belonging to different text lines.

Several authors proposed to eliminate the small components, such as dot, before starting text line detection. Despite these dots may be noisy components, it appears that they are important to ensure the continuity of the text lines. A text line may contain a sequence of dots (see an example in Fig. 2) or a sequence of small character fragments resulting from the binarisation. The very small CCs are labelled DOT, they are not labelled according to their nearest neighbour. Therefore, they cannot instantiate a new alignment but they can be included into an alignment to expand it. A CC is labelled DOT if the both sides of its bounding box are smaller than three pixels.

Once the CCs are labelled, the perceptual grouping process can start.

4. PERCEPTUAL GROUPING

The main goal of this perceptually based method is to detect vertical and horizontal text lines without prior assumption on the line directions. The detection of vertical and horizontal text lines arose as a prior task when we started to automatically extract the Figure&Caption pairs in the books of the BIUM collection. Several methods, such as RLSA, can segment vertical or horizontal text lines, with eventually a strong angular deviation [4], but they require to initiate the detection process with a direction and implicitly suppose that the text lines in a page have the same direction (horizontal or vertical). They do not provide a way to merge the results obtained after performing independently text line detection in vertical and in horizontal directions.

The rule-based grouping process has three main steps. The first one groups into alignments the closest CCs according to their labels. The second and the third steps extend the alignments by adding CCs or by merging alignments. Each alignment is associated with a confidence value incremented at each step of the grouping process. As the confidence level of the alignments increases, the tolerance of the grouping criteria involved in the rules increases. Conflicts may appear when detecting simultaneously text lines in different directions. Each grouping step is followed by the activation of a conflict resolution procedure that can set to zero the confidence value of some alignments.

The actions performed during the grouping steps are:

- the creation of a new alignment,
- the inclusion of a CC into an alignment,
- the merging of two alignments.

They are performed if conditions are satisfied. Global information is calculated and involved in the grouping conditions. The average height and width of the CC bounding boxes (MCCX and MCCY) are updated after the preprocessing, the CCG are discarded to evaluate them. The average height and width of the detected alignements (MALX and MALY) are calculated.

Conditions are defined to reduce the risk of error when including a CC into an alignment and when merging two alignments. We call them "typographic conditions" because they try to capture the typographic conventions: the length of the space between the words is related to the height of the characters, the sizes of the characters do not present strong variations in a text line. Other works, among them [4, 8], introduced the same kinds of conditions when grouping the CCs. They only take into account the position and the size of CCs, limiting to a local context the conditions to satisfy. In our case, the spatial context involved in the typographic conditions is determined by the size and the position of the emerging alignments. The typographic conditions are:

TC1: HCA0 = the height of the current alignment,

HCA1 = the height of the alignment after inclusion of a CC or after merging two alignments.

The height of a horizontal alignment is the length of the Y-border of its bounding box, while the height of a vertical alignment is the length of the X-border of its bounding box.

Condition: HCA1 < 1.5 HCA0

TC2: D = the distance between the current alignment and the CC to be included,

Condition: D < 3 MCCX for horizontal alignments, D < 3 MCCY for vertical alignments.

TC3: D = the distance between the current alignment and the CC to be included or the alignment to be merged.

HCA = the height of the current alignment.

Condition: D < 2 HCA

The borders of the CCG are involved in the conditions to satisfy when expanding an alignment. The inclusion of a CC into an alignment or the merging of two alignments is not allowed if the resulting alignment straddles a separator, this is the separator condition (SC).

The three grouping steps are presented with the labels used for the horizontal alignments. For the vertical alignments, the procedure is similar; NNE is replaced by NNS and NNR by NNB.

1. Grouping the CCs according to their labels.

The labelled CCs give reliable information about the direction of the perceived alignment. Horizontal alignments are built from NNE and NNR. Let CC1 a NNE belonging to an alignment, if not, CC1 initiates a new horizontal alignment. Its nearest neighbour CC2 is added to the current alignment containing CC1 if TC1, TC2 and SC are satisfied. If CC2 is labelled NNE or NNR, then the procedure continues with CC1 = CC2. If not, the grouping is stopped for this alignment. Figures 4a show examples of alignments that are obtained after this step.

2. Grouping by proximity and continuity of direction

The previously detected alignments are extended along their principal direction according to the proximity of the CCs. The first (leftmost or topmost) and last CCs of each alignment have a closest neighbour along the principal direction of the alignment. If the neighbour is not already included into an alignment, then it will be included into the current alignment. If it is included into an alignment that has the same direction as the current alignment, then the two alignments are merged. The inclusion of a CC and the merging of two alignments must satisfy TC1, TC3 and SC.

3. Grouping by continuity of direction

This step is a more tolerant version of the previous one: the proximity criterion is no more restricted to the neighbour of the first and last CCs of each alignment. An alignment is expanded by merging it with an overlapping alignment or with its closest alignment if both have the same principal direction. The alignments are also expanded along their principal direction by including the nearest CCs that are not already included into an alignment. TC1, TC3 and SC must be satisfied to expand an alignment.

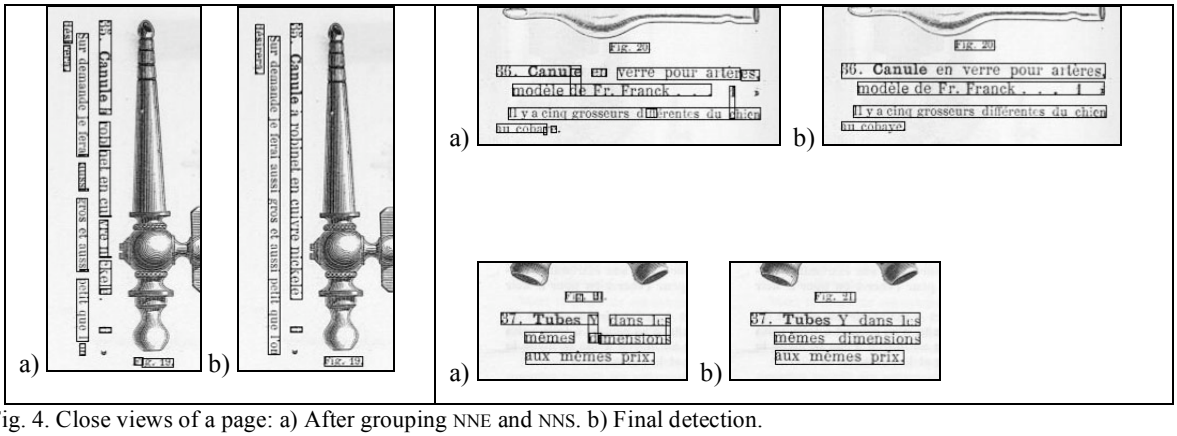


Fig. 4. Close views of a page: a) After grouping NNE and NNS. b) Final detection.

The alignments in Fig. 4.b result from the three grouping steps but also from the additional procedures that are presented in the next section. They are performed after each grouping step to detect conflicts and to activate resolution rules.

5. CONFLICTS

5.1 Location conflict

A conflict is detected if the bounding box of a CC is inside the bounding box of an alignment and the CC is not in the CC list of this alignment. Including the missed CC into the CC list of the alignment solves the conflict. Another case of conflict is detected if the bounding box of an alignment (Aint) is inside the bounding box of another alignment (Aext). Including the CCs of Aint into the CC list of Aext solves it; the confidence of Aint is set to zero. This conflict is solved under the condition that the height of Aext is lower than 1.5MALY for a horizontal alignment (lower than 1.5MALX for a vertical alignment). This kind of conflict is often encountered and solved for characters with two components such as i, j, é, ... or split characters.

5.2 Line conflict

A conflict between lines is detected when a vertical and a horizontal alignments are crossing (see examples in Fig. 4.a). When a conflict is solved, the confidence of one of the conflicting alignments is set to zero. The decision depends on the current set of detected alignments. A conflict is solved by using voting rules involving the labels of the CCs that belong to the alignments.

Let VAL be a vertical alignment. The number of CCs in VAL = *numV*, the number of NNR in VAL = *numR*.

Let HAL be a horizontal alignment crossing VAL. The number of CCs in HAL = *numH*, the number of NNB in HAL = *numB*.

For each VAL, the set of HAL is searched. Let *n* be the number of HAL in this set:

$$numH = \sum_{i=1}^{n} numH_i \qquad \text{and} \qquad numB = \sum_{i=1}^{n} numB_i$$

The confidence value of the vertical alignment is set to zero if the set of HAL is not empty ($n \neq 0$) and if:

($numV > 0$) and ($numH > 0$) and (($numH$ - $numB$) > (2 + ($numV$ - $numR$)).

The procedure is similar for detecting suspect horizontal alignments. The voting rule becomes:

($numV > 0$) and ($numH > 0$) and (($numV$ - $numR$) > (2 + ($numH$ - $numB$))

5.3 Text/Graphics conflict

Despite text parts may be found inside CCG bounding boxes, they are more often outside. According to that and to reduce the number of false text lines in CCG, the tolerance for the alignments in CCG is lower than outside the CCG. Let BBX and BBY the border lengths of the alignment bounding box. The confidence of the alignment inside a CCG is set to zero if one of these condition is satisfied:

- BBX < MCCX or BBY < MCCY
- BBX < BBY for a horizontal alignment, BBY < BBX for a vertical alignment
- BBY > 5 MCCY for a horizontal alignment, BBX > 5 MCCX for a vertical alignment

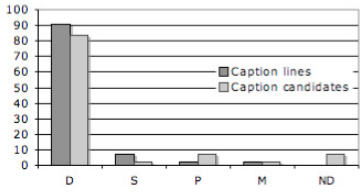
6. RESULTS

The results are given for a set of 52 images of pages gathered from the BIUM collection. The layout of these pages is well appropriate to evaluate the detection of text lines without prior assumption on their direction: all of the pages contain both horizontal and vertical text lines and also one or several Figure&Caption pairs. To evaluate the robustness of the method, the pages were chosen in several books, which have different typographic and presentation styles. A total of 22 books is represented in the data set. These images are available on the web site of the BIUM digital library [14]. The documents were type printed during the 19th century.

Examples of text line detection are given in Fig. 2d, 3 and 4b. The simultaneous detection of vertical and horizontal text lines was a first objective that has arisen when starting the detection of Figure&Caption pairs. The presented evaluation measures the efficiency of the proposed method for the detection of the first caption lines. The number of Figure&Caption pairs is 102 with 31 horizontal captions and 71 vertical captions. The results on caption line detection are reported in Table 1. Split lines are most often observed in skewed pages or skewed lines. The detection of a global or local skew is not performed here. In Fig. 5a, the first caption line (vertical) is detected in two parts. Because of the skew of the caption lines, the grouping criteria TC1 is not satisfied, therefore, the two parts are not merged into a single line. The results show that the proposed method is a valuable contribution to detect Figure&Caption pairs when the direction of the caption lines cannot be predicted and may be horizontal or vertical.

Table 1. Detected caption lines and selection of candidate caption lines

102 caption lines	Detected (D)	Split (S)	Partially (P)	Merged (M)	Missed (ND)
Caption lines	91	7	2	2	0
Caption candidates	84	2	7	2	7



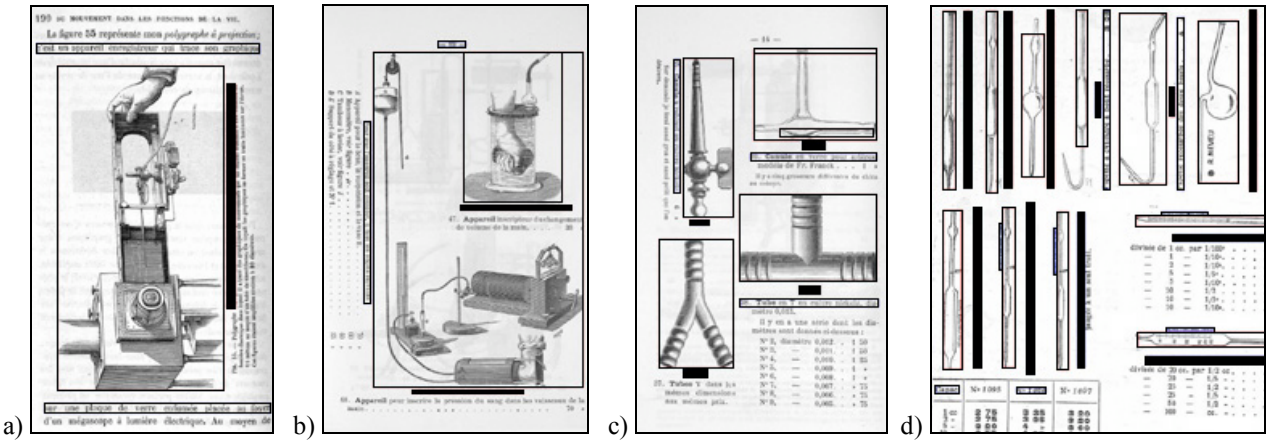


Fig. 5. Caption candidates with their associated graphics. The text lines filled in black are the first caption lines belonging to the set of caption candidates

The detection of Figure&Caption pairs requires a spatial reasoning, which is not in the scope of this paper and is planned for a further development. Nevertheless, a first attempt was made to extract caption candidates from the detected text lines in a page. Captions are associated with Graphics. The Graphics involved in the search of caption candidates are the CCG but also the Graphics obtained by merging the CCGs with overlapping CCs and CCGs. The spatial relationship between text lines and Graphics are considered to build the candidate set of Figure&Caption pairs. A proximity criterion between Graphics and text lines is suitable but the caption is not always the closest text line of the Graphics. An alignment criterion is added. It is defined as the vertical or horizontal alignment of the Graphics and text line centres (a tolerance equal to 1/10 of the greatest bounding box border is introduced to accept the alignment of the centres). A text line is included in the candidate caption set if the proximity or the alignment condition, or the both are satisfied. Fig. 5 gives examples of candidate caption lines with the associated Graphics. Table 1 shows the numbers of caption lines that are in the candidate sets. The detection of text lines in several parts (split lines) leads to partially detected caption lines: only the part closest to the graphics is retained as a candidate caption (see an example in Fig. 5a).

7. CONCLUSION

The need to simultaneously detect vertical and horizontal text lines was established as a goal prior to the automatic extraction of Figure&Caption pairs in ancient books of digital libraries. Pages may contain text lines type printed in different directions. The proposed method detects text lines without prior assumption on their direction and retains the

most salient character alignments. The results obtained for the detection of caption lines lead us to adopt this method with the aim of extracting Figure&Caption pairs. Nevertheless, some improvements are necessary to reduce the lines that are merged or split and to better control conflict resolution. The current results encourage the involvement of perceptual organisation in document image analysis. Better performances are expected by increasing the spatial context taken into account to segment the text lines. Text bloc information has already been introduced in the ongoing work. A careful evaluation of the updated method has to be conducted before concluding on its performances.

The application of this work is to provide an assistive technology that can reduce the load of archivists who manually extract the metadata of the document images. The automatic detection of Figure&Caption pairs was the origin of this study. Despite the fact that a great number of Figure&Caption pairs can be found with proximity and alignment criteria, most difficult cases require the application of sophisticated spatial reasoning, that is another perspective of this work.

8. REFERENCES

[1] Antonacopoulos, A., "Page Segmentation Using the Description of the Background," Computer Vision and Image Understanding. Special Issue on Document Image Understanding and Retrieval, 70(3), 350-369 (1998).

[2] Ar, I., and Karsligli, M.E., "Text Area Detection in Digital Documents Images Using Textural Features," CAIP, LNCS 4673, Springer-Verlag, 555-562 (2007).

[3] Chang, T.-Y., Takiguchi Y., Okada M., "Physical Structure Segmentation with Projection Profile for Mathematic Formulae and Graphics in Academic Paper Images," Proc. of the 9th Inter. Conf. on Document Analysis and Recognition, Curitiba, Brazil, IEEE Computer Society, 392-396 (2007).

[4] Cao, H., Prasad, R., Natarajan, P., and MacRostie, E., "Robust Page Segmentation Based on Smearing and Error Correction Unifying Top-down and Bottom-up Approaches," Proc. of the 9th Inter. Conf. on Document Analysis and Recognition, Curitiba, Brazil, IEEE Computer Society, 1193-1197 (2007).

[5] Galindo, D., Faure, C., "Perceptually-Based Representation of Network Diagrams," Proc. Inter. Conf. on Document Analysis and Recognition, Ulm, Germany, 352-356 (1997).

[6] Lemaitre, A., Camillerapp, J., "Text Line Extraction in Handwritten Document with Kalman Filter Applied on Low Resolution Image," Proc. of DIAL'06 the Second International Conference on Document Image Analysis for Libraries, 38-45 (2006).

[7] Likforman-Sulem, L., Hanimyan, A., Faure, C., "A Hough Based Algorithm for Extracting Text Lines in Handwritten Documents," Proc. Inter. Conf. on Document Analysis and Recognition, ICDAR'95, Montréal, Canada, 774-777 (1995).

[8] Makridis, M., Nikolaou, N., and Gatos, B., "An Efficient Word Segmentation Technique for Historical and Degraded Machine-Printed Documents," Proc. of the 9th Inter. Conf. on Document Analysis and Recognition, Curitiba, Brazil, IEEE Computer Society, 178-182 (2007).

[9] Niblack, W., [An Introduction to Image Processing], Prentice-Hall, Englewood Cliffs, NJ, 115-116 (1986).

[10] Raju, S.S., Pati, P.B., Ramakrishnan, A.G., "Gabor Filter Based Block Energy Analysis for Text Extraction from Digital Document Images," Proc. of DIAL'04 the First International Conference on Document Image Analysis for Libraries, IEEE Computer Society, 233-244 (2004).

[11] Ramel, J.-Y., Busson, S., Demonet M.-L., "AGORA: the Interactive Document Image Analysis Tool of the BVH Project," *Proc. of DIAL'06 the Second International Conference on Document Image Analysis for Libraries,* 145-155 (2006).

[12] Saund, E. and J. Mahoney, J., "Perceptual support of diagram creation and editing," In International Conference on the Theory and Applications of Diagrams, 424-427 (2004).

[13] Sarkar, P., Saund, E., "Perceptual organization in Semantic Role Labeling," Symposium on Document Image Understanding Technology, College Park, Maryland, (2005).

[14] BIUM: Bibliothèque InterUniversitaire Médicale, Paris, http://www.bium.univ-paris5.fr/histmed/medica.htm