2011 International Conference on Document Analysis and Recognition

Text/Graphics Segmentation in Architectural Floor Plans

Sheraz Ahmed*†, Markus Weber*, Marcus Liwicki*, Andreas Dengel*†
* *German Research Center for AI (DFKI GmbH)*
Knowledge Management Department,
Kaiserslautern, Germany
{firstname.lastname}@dfki.de
† *Knowledge-Based Systems Group,*
Department of Computer Science, University of Kaiserslautern,
P.O. Box 3049, 67653 Kaiserslautern

***Abstract*—In this paper, we propose an improved method for text/graphics segmentation. Text/graphics separation is a crucial preprocessing step in document analysis before further analysis and recognition can be applied. Our proposed system extends the method of Tombre et al. with a number of improvements to make it more suitable for architectural floor plans. A crucial novel preprocessing step is the detection and removal of walls before the actual segmentation. Furthermore, text components are then extracted by analyzing connected components and even considering text overlapping with graphics. Finally, a smearing approach is used to remove noise and extract the final text components. Evaluation results over the series of 90 floor plans which has also been used in reference work shows that our method has a recall of almost 99 % and a precision greater then 97 %.**

I. Introduction

Text/graphics segmentation is a useful preprocessing step in the document analysis chain. The aim of this process is to extract two separate layers, one containing just graphical information, the other just containing textual information. Textual information in technical drawings is important for further analysis, especially when semantics of room should be detected. For other steps in the analysis, e.g., the structural analysis of graphical elements, text might lead to false interpretations. Thus text/graphics segmentation is a useful preproccessing step in the document analysis chain. The aim of text/graphics segmentation is to extract two separate layers, one containing just graphical information, and one containing textual information, respectively.

The graphics recognition community has already put a lot of effort into research on text/graphics segmentation. In general, different methods have been proposed to work in different scenarios. In the case of document images it is used to separate text lines from the images. For natural images the focus is to extract the text from the remaining scene. In technical drawings it is used to separate technical drawing symbols from the text labels. In each domain different rules has been imposed to get the better results.

In this paper the focus is text/graphics segmentation as a preprocessing unit of the recognition of specific technical drawings, i.e., on the recognition of architectural floor plans. The final aim of the floor plan recognition system is to generate a semantic representation of the rooms including the room functions [1]. As floor plans often contain textual information labeling the type of the room, the automatic extraction of textual information is an important step for further analysis using optical character recognition (OCR). Note that the focus of this paper is the actual text/graphics segmentation. The complete floor plan analysis system is presented in another paper.

II. Related Work

Wahl et al. [2] have proposed a method for block segmentation and text extraction in mixed text/image documents. This method has shown promising results for document images. For instance, it is used successfully for text line segmentation and image block separation. Various improvements have been proposed for this approach. However, the focus of all of these approaches lies on general document image analysis.

For the purpose of the more specific technical drawings, Fletcher and Kasturi [3] proposed a method for separation of text strings from mixed text/graphics images which is based on connected component analysis. A minor drawback of their method is that text touching graphics is marked as a graphics component rather then text. It gives the promising results on the images where no text is touching the graphics. However, in most of the technical drawings and maps images text and graphics overlay. On the other hand side, simplicity is a major advantage of this approach.

In [4] a vector-based segmentation of text connected to the graphics is performed. The focus of work was engineering drawings. This method is based on growing individual character box regions which are then merged into text boxes. Finally, the text box is re-segmented into character boxes.

A text/graphics separation method for overlapping text from graphics was proposed in [5]. They start with preprocessing to separate the solid graphical components and remove all the dashed lines. This method also based on connected component analysis where a size filter is used for

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.153

marking components as either text or graphics. They applied this method to images of maps.

Tombre et al. [6] proposed an improvement for the above-mentioned method introduced by [3]. [6] proposed additional filters to be applied on connected components. In combination to area/ratio filter, size and shape filters are introduced. Furthermore, they split image in to three layers, i.e., text, graphic, small elongated components layer. The third layer is used by text string extraction and for finding the dashed lines. After text separation Hough transform is used to group characters in to strings. This method improved the results of [3], but still some touching characters were marked as graphics components.

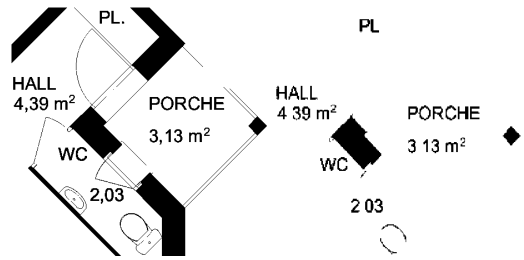
In [7] an improvement for Tombre's approach has been proposed by using color information to separate touching text from the graphics. After separation of text/graphics Hough transform is used to remove the lines from the image. Finally, for grouping the characters into words Pyramid segmentation has been applied. This method can be used where the text and graphics are more or less occurring in different colors.

[8] introduced a hierarchical method for segmenting text areas in natural images. The basic assumption is that text is written with a contrast color on a uniform background. Segmentation process starts with finding the text background areas. In each segmented region, the presence of text is tested afterwards.

[9] used color information coupled with the graph representation. Initially, a color model is computed from the color properties of the image. Then image contours are extracted using edge detection. Finally, connected components of the contour image are classified according to the graph representation. Structural training is used to learn the text and graphics diversity. They also based their method on the assumption that text is not touching graphics components.

[10] introduced an approach for text/graphics segmentation. This technique is based on the sparse representation framework and two appropriately chosen discriminative dictionaries. Using each dictionary, sparse representation of one type of signal and non-sparse representation for other type of signal is achieved. Finally, text graphics separation is achieved by promoting sparse representation of input image in these two dictionaries. [10] claimed that their approach could be used in any domain. However, an adoption to floor plans would require additional effort in order to benefit from specific properties of floor plan images.

Each of the methods above has a different complexity level and was proposed for different domains. In this paper we have chosen the method of Tombre [6] as a baseline method, because it produces already reasonable results on floor plan images. We improved the method of [6] by first segmenting the walls from the floor plan image. This segmentation is very important because if any part of the walls is marked as text will lead to severe errors during later analysis. Furthermore, we split the image into two instead of three layers and eliminate the need of using the additional threshold introduced in [6] and no Hough transform is used for string extraction. Finally, we use smearing for retrieving the text strings and recovering characters touching graphical components.



(a) (b)

Figure 1: Detail of a floor plan (a) and the text extraction result if no wall segmentation was performed.

III. Text/Graphics Segmentation Approach

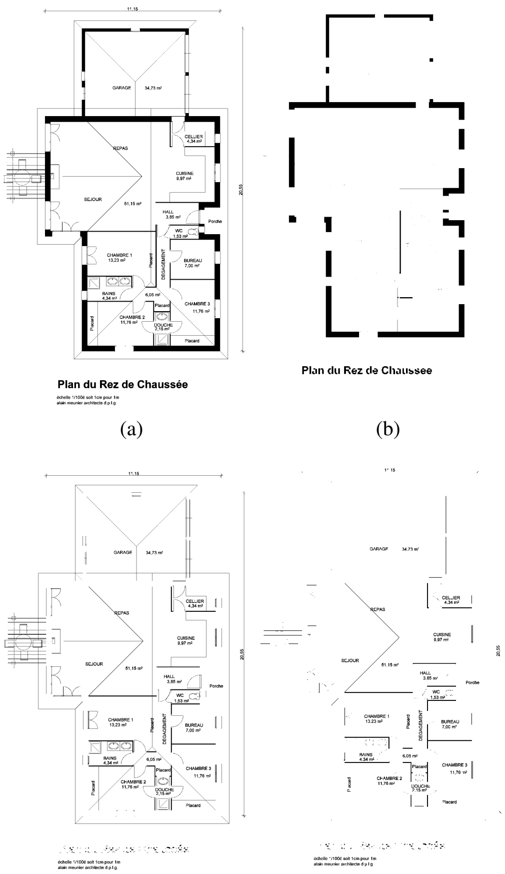
The process of text/graphics segmentation analyzes the floor plan image and converts it into two images. The former image is the text image that contains only the text components. The latter one is the graphics image containing all building elements of the floor plan.

Before performing text/graphics segmentation, it is necessary to first segment all the external walls from the image. This segmentation is very important, because if text/graphics segmentation is applied directly on the original image some of the external walls would be marked as a text, creating errors during a later structural analysis. Figure 1b shows a wall of Figure 1a that has been incorrectly marked as text component (close to the label "WC").

External walls are removed by successive morphological binary erosion (3 times) followed by successive morphological binary dilation (4 times) of the image with a 3×3 square mask. The result of this process is an image where only the thick components remain. This image is than overlaid with the original image and the intersection is considered to contain the thick walls. Note, that this process does not only remove the thick wall components but also the main title text of the floor plan as it contains thick characters. However, this title can be recovered at a later step.

Figure 2a shows an example image of a floor plan. After the removal of external walls, which are illustrated in Figure 2b, the remaining image contains only the text, medium lines, and thin lines. This remaining image is illustrated in Figure 2c).

In floor plans it happens quite often that thin lines overlap with text components. This overlapping makes it difficult to separate text from the graphics components, especially when the statistics of connected components are used for



(a) (b) (c) (d)

Figure 2: Original floor plan (a), the detected thick walls (b), the intermediate image after wall segmentation (c), and the image after thin line removal (d).

Figure 3: Text touching a thin line

text extraction. Figure 3 illustrates a detail of a floor plan where a thin line touches a text component.

These thin lines are removed from the image, so that all of the text components can be extracted. Morphological binary opening is performed to remove these thin lines first, with a horizontal mask, later, with a vertical mask. The result after thin line removal is illustrated in Figure 2d. This image is then further processed for extracting text components.

The text components are now extracted by performing connected component analysis on the segmented image. Connected component analysis not only locates the black connected components in the image, but also computes their sizes/areas, densities, dimensions, bounding boxes, widths, heights, and centroids. After completion of connected component analysis, a histogram is plotted for the size of all connected components. The purpose of plotting this histogram is to analyze the distribution of different components with respect to their sizes, especially the most populated area. Let A_{mp} be the size of components in the most populated area of the histogram and A_{avg} be the average size of the connected components.

The main idea behind finding A_{mp} and A_{avg} is to define the size filter. This size filter is used to separate graphics components from text components. This separation is based on two thresholds T_1 and T_2. T_1 is a size threshold and is set to $T_1 = N \times max(A_{mp}, A_{avg})$. T_2 is the maximum elongation threshold and is set to $T_2 = max(AVG_{height}, AVG_{width})$, where AVG_{height} is average height and AVG_{width} is average width of connected components. The value of size factor $N = 5.0$ is selected empirically. The other two thresholds T_1 and T_2 are calculated automatically.

The filter is then applied to mark all of the connected components as text satisfying the following criteria:

- The area is less than T_1,
- $\frac{height}{width}$ is in the range of $[\frac{1}{T_2}, T_2]$, and
- both, height and width are lower than $\sqrt{T_1}$.

Note, that the other threshold, i.e., T_4 specified by [6] is not used, because we are splitting the image into two layers, whereas T_4 was required for a third layer which is not needed in our approach. An example text image obtained after application of the size filter is shown in Figure 4a. Note that in this intermediate text image are still many small and large components, which are not text.

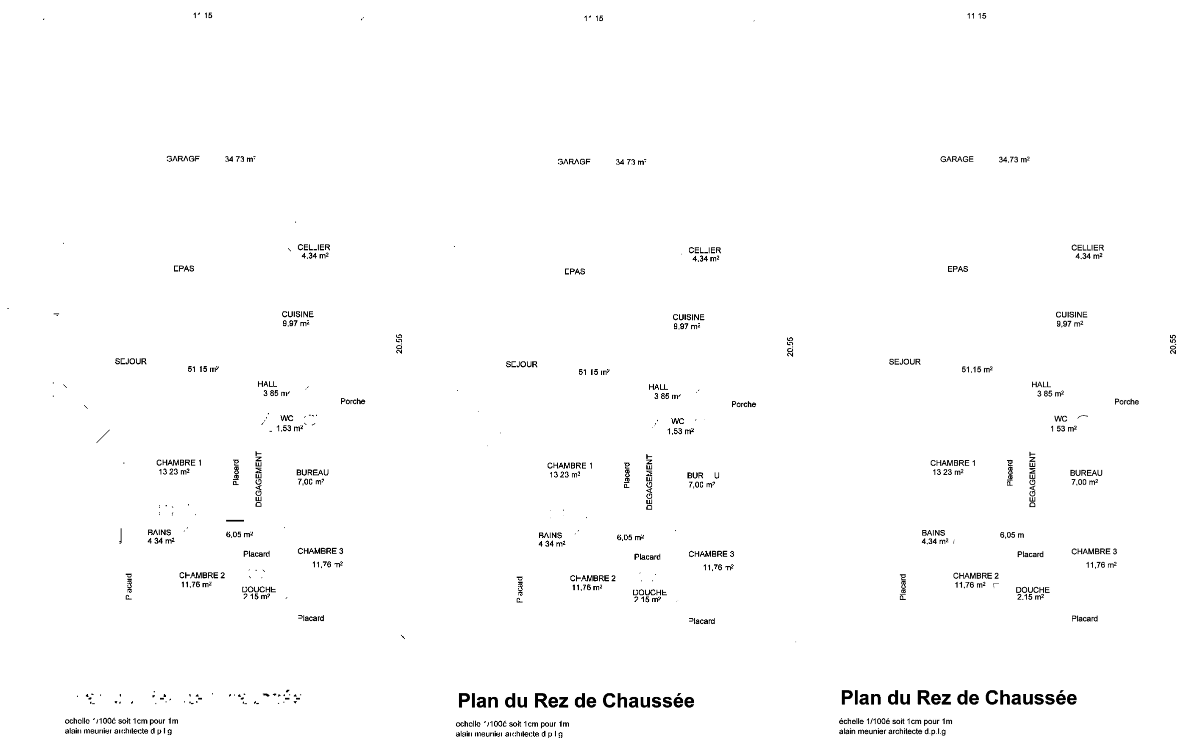
To get rid of these unnecessary components, noise removal is performed on the intermediate image. Noise removal starts by analyzing the remaining connected component of the intermediate text image. Let $Height_{avg}$ be the average height and $Width_{avg}$ be the average width of the connected components. Horizontal and vertical smearing is performed on the intermediate text image to merge the components nearby. From the smeared image, all small connected components are removed. A component is regarded as small if it satisfies any of the following criteria:

- Height is less then $Height_{avg}$.
- Width is less then $Width_{avg}$.
- Density is less then density threshold T_d.
- Area is very small.

The value $T_d = 30\%$ is selected emipirically.

Smearing is very important here, because if small components were removed directly from the intermediate text image many components would be removed incorrectly. After removing all of the small connected components from the smeared image, the content from the intermediate text image is restored for all the remaining bounding boxes of the smeared image.

At this step, all small components are removed, but large non-text components still remain and have to be filtered out.



(a) (b) (c)

Figure 4: Text/graphics segmentation of the floor plan illustrated in Figure 2: intermediate text layer after the first size filtering (a), after noise removal and main title-reconstruction (b), and after the final retrieval of missing small text components.

In order to remove these large components the width- and height-histograms of the remaining connected components are computed. The width and height at the maximum of these histograms are taken as width and height thresholds. Now, those components are removed which have a width or height greater than twice the thresholds. The remaining image contains nearly only text components. However, still the title text and text touching graphics are missing, which will be recovered in the following steps.

The floor plan title is retrieved as follows. Horizontal and vertical smearing is performed on the walls image, which has been retrieved above (see Figure 2b). Now, all connected components having a density less than an empirically defined density threshold (T_{d2}) and a width not in the range of the empirically defined width interval threshold (T_w) are removed. The remaining components are copied to the text image. The values of the thresholds T_w and T_{d2} are set to $[200 - 600]$ and to 70%, respectively. Figure 4b shows the output where the noise removed image and the floor plan text image are combined. It can be seen that still some characters are missing which were touching graphical components before, e.g., the second 1 of the value 11 on the top and the "E" appearing as "F" in the word "GARAGE" on the top.

For retrieving those characters, which were touching graphics in the original, a text string extraction is performed. For text string extraction the current version of text image is duplicated. Horizontal smearing is performed on one copy and vertical smearing on the other. Consequently, taking a union of both smeared images combines both images. Now the connected components on the combined smeared image are investigated. The portion of the original floor plan image, which appears in the bounding box of each connected component, is copied to a blank image. The resulting image finally contains all the text components, plan title, and even missing characters, which were touching graphics.

Figure 4c shows the final result after recovering missing text.

IV. Evaluation

Our system is evaluated using a data set of original floor plans from the period of more then ten years. This data set was primarily introduced for floor plan analysis in [11] and contains 90 floor plan images.[1] Unfortunately, no performance of the text/graphics segmentation process has been mentioned in these considered references [6] and [11].

[1]The actual image size is 2479×3508. For making the analysis process more efficient, isotropic down scaling to 1413×2000 has been applied.

	Number	Percentage (%)
Ground Truth Text	$21,737$	100
Clearly Readable	$21,164$	97.4
Difficult to Read	327	1.5
Missing	246	1.1
Noise Error	255	1.2

Table I: Text/graphics segmentation results

However, manually inspecting the resulting images suggests that our approach works better on the data set than previous approaches, especially when considering the text portions which touch thin lines (those were completely removed by previous approaches). As generating the ground truth for the method in [6] takes more time than a simple manual inspection, therefore we have chosen the criteria for the manual inspection which is very objective.

In order to assess the extraction results, we categorized our resulting characters into three categories: *clearly readable*, indicating that the characters were perfectly retrieved, *difficult to read*, indicating that minor parts are missing (not influencing an OCR) or thin symbol structures overlay the text in the results, *missing*, indicating the character in the floor plan did not appear in the results, and *noise*, indicating that a non-text component appears in the retrieval results. These values have been determined by visual inspection of an independent person.

Table I shows the results of the text/graphics segmentation. Our system has recall of almost 99 % with the precision of more then 97 %.

V. Conclusion and Future work

In this paper we have proposed a method for text/graphics separation for architectural floor plan images. We have evaluated the system on floor plan images, which are publicly available. Our method has a recall of almost 99 % and a precision greater then 97 %.

The high recall value is a result of wall removal and missing characters extraction using smearing. The high precision value is a result of noise removal using the connected components statistics. This combination of high precision and recall makes it feasible to use our text/graphics separation as a preprocessing of floor plan analysis systems.

While our system achieves already very good results, there is still room for improvement.

Currently, we take the bounding boxes of each connected component into account to recover missing characters. This leads to errors on diagonal text components (which rarely appear in the considered data set). Instead of that we could use the best fitting rectangle. To make the system more robust, we are working on incorporating different wall designs for wall removal step.

A further idea is to remove the thin straight lines permanently on the final by applying methods for straight line removal in text images. Another interesting topic for future research is to further analyze the connected component and to estimate the values of all thresholds dynamically.

Acknowledgment

The authors of this paper thank Joseph Llados and the members of the CVC for providing us with the floor plans. Our work was financially supported by the ADIWA project.

References

[1] M. Weber, M. Liwicki, and A. Dengel, "a.SCAtch - A Sketch-Based Retrieval for Architectural Floor Plans," in *12th Int. Conf. on Frontiers of Handwriting Recognition.*, 2010, pp. 289–294.

[2] F. M. Wahl, K. Y. Wong, and R. G. Casey, "Block segmentation and text extraction in mixed text/image documents," *Computer Graphics and Image Processing*, vol. 20, no. 4, pp. 375–390, 1982.

[3] L. Fletcher and R. Kasturi, "A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 910–918, 1988.

[4] D. Dori and L. Wenyin, "Vector-based segmentation of text connected to graphics in engineering drawings," in *Advances in Structural and Syntactical Pattern Recognition*, ser. LNCS, 1996, vol. 1121, pp. 322–331.

[5] R. Cao and C. L. Tan, "Separation of overlapping text from graphics," in *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*, 2001, pp. 44–48.

[6] K. Tombre, S. Tabbone, L. Plissier, B. Lamiroy, and P. Dosch, "Text/graphics separation revisited," in *Document Analysis Systems V*, ser. Lecture Notes in Computer Science, D. Lopresti, J. Hu, and R. Kashi, Eds. Springer Berlin / Heidelberg, 2002, vol. 2423, pp. 615–620.

[7] P. P. Roy, J. Llados, and U. Pal, "Text/Graphics Separation in Color Maps," *International Conference on Computing: Theory and Applications*, vol. 0, pp. 545–551, 2007.

[8] S. A. R. Jafri, M. Boutin, and E. J. Delp, "Automatic text area segmentation in natural images." in *ICIP'08*, 2008, pp. 3196–3199.

[9] R. Raveaux, J.-C. Burie, and J.-M. Ogier, "A colour text/graphics separation based on a graph representation," in *ICPR*, 2008, pp. 1–4.

[10] T. V. Hoang and S. Tabbone, "Text extraction from graphical document images using sparse representation," in *9th Int. Workshop on Document Analysis Systems*. New York, NY, USA: ACM, 2010, pp. 143–150.

[11] S. Macé, H. Locteau, E. Valveny, and S. Tabbone, "A system to detect rooms in architectural floor plan images," in *9th Int. Workshop on Document Analysis Systems*. New York, NY, USA: ACM, 2010, pp. 167–174.