Title Extraction from Book Cover Images Using Histogram of Oriented Gradients and Color Information

Yen Do, Soo Hyung Kim, In Seop Na*

School of Electronics & Computer Engineering, Chonnam National University
Gwangju, 500-757 Korea

ABSTRACT

In this paper, we present a technique to extract the title areas from book cover images. A typical book cover image may contain text, pictures, diagrams as well as complex and irregular background. In addition, the high variability of character features such as thickness, font, position, background and tilt of the text also makes the text extraction task more complicated. Therefore, we propose a two steps efficient method that uses Histogram of Oriented Gradients and color information to find the title areas. Firstly, text localization is carried out to find the title candidates. Finally, refinement process is performed to find the sufficient components of title areas. To obtain the best result, we also use other constraints about the size, ratio between the length and width of the title. We achieve encouraging results of extracted title regions from book cover images which prove the advantages and efficiency of the proposed method.

Keywords: Library Automation, Text Extraction, Histogram of Orientated Gradient, Localization, Connected Component, Color Clustering.

1. INTRODUCTION

Nowadays, pattern recognition and machine learning play increasingly important roles in handling diversity of features in large datasets. Among many applications, Optical Character Recognition (OCR) technique makes it possible to recognize the characters of the letter [1,2], or edit the text, search for a word or phrase, store it more compactly, display or print a copy free of scanning artifacts. Recognizing the title of the book is one of good applications of OCR system. Imagine a person who is in the library and wants to find all the description of an incidental book. He or she can take a photo of the book cover by a mobile phone. The image becomes the input of the OCR system. The application will recognize characters in the title, and display all the details of that book. In order to recognize characters in the book cover image, we have to extract the title regions. Although the work can also be done by barcode detection and extraction, but it has some limitations. If the book has its own International Standard Book Number (ISBN) and barcode, we can recognize the barcode to find the detail of the book. However, a book may appear without a printed ISBN and barcode if it is printed privately or the author does not follow the usual ISBN procedure, we could not solve the problem by barcode recognition. As the result, the problem should return to title extraction and recognition. Therefore, the accuracy of title

extractions plays an important role in the accuracy of recognition systems. Thus, to improve the accuracy of title recognition, we have to improve the accuracy of the title extraction. However, in order to get more accurate of title regions, we have to face with many difficulties, such as variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background [3]. As a matter of fact a typical book cover may contain text, pictures, diagrams, irregular background. Further, the text recognition task is complicated by the high variability of characters features such as dimension, thickness, font, background and position as well as the tilt of the text.

Several approaches have been presented in the last few years for the automatic localization of text in scene images. There have been attempts in the past for extraction of textual component of an image by analyzing the edges of candidate regions or homogeneous color/gray scale components that contain the characters [4-6]. Other method uses mathematical morphological technique for text extraction [7]. Nguyen Noi Bai, Kim Nam and Youngjun Song proposed a method can both deal with more than a specific language and extract text lines containing some wide inter-word gaps [13]. Another text extraction method is proposed by Chen and Ding et al. [8]. In this method, a bottom-up approach is followed by progressively merging image components at different levels based on the calculation of a quantitative measure [the MLC value]. A revised version of this method was analyzed in [9], which can deal with irregular regions. Jain and Yu in [10] have surveyed some OCR and page segmentation algorithms. In this paper they have also suggested use of traditional bottom-up approach

based upon connected component extraction to efficiently implement page segmentation and region identifying.

In this paper, we propose a method that uses the combination of Histogram of Oriented Gradients and color information to detect the title areas of the book cover. The method will be described in section 2. Some results of the proposed method will be shown in section 3. Section 4 will be reserved to conclude our solution and our achievement.

2. MATERIALS AND METHODS

The goal of our system is to localize text and extract the title of the book cover images. Our proposed method consists of two steps. First step, text localization is based on oriented gradient of the edge map and color information. The system first detects the text regions which consist of four sub steps. Each region is analyzed to determine whether the region corresponds to text or not. After that, at the second step, refinement process will be performed to extract the whole title of the book cover image. Fig.1 below shows the flowchart of our proposed system.
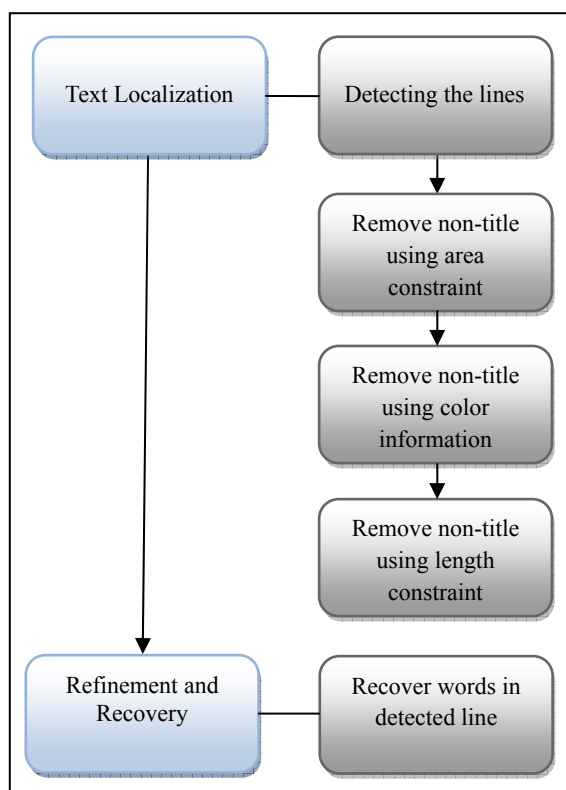

Fig. 1. Flow chart of the system.

2.1 Text Localization

2.1.1 Detecting the lines of text: Under normal circumstances, the title of the book should have been presented in a way easy to read with a certain contrast, a reasonable size, and an approximately horizontal orientation. First of all, the characters inside a word have tendency to be in horizontal orientation that is in a straight line or alignment. For this reason, the constraint of horizontal text orientation is worthy for the

title of the book cover image detection purpose. Using Histogram of Oriented Gradients (HOG) [11], local object appearance and shape within an image can be described by the distribution of edge directions. Consequently, we can predict the areas that are text candidates.

To calculate HOG, first divide the image into small connected regions, called cells. Then, for each cell we compile a histogram of gradient directions or edge orientations for all the pixels within the cell. Next step, to improve the accuracy, the local histograms are contrast-normalized. We now divide the image into larger regions, called blocks. For each block, a measure of the intensity across the block is calculated, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination or shadowing. After the contrast-normalized step, we should only take the candidates that have high value of horizontal oriented gradient. Building a common standard for every threshold is very difficult. With the purpose "extraction the title areas", every images have a similar size that can be anticipated. For this reason, some thresholds are appropriate with the purpose, for example, all the areas that have gradient orientation range between (0.5, 0.9) are selected since these elements have high probability of horizontal. Fig.2 shows the histogram of Oriented Gradient and the selected thresholds for text detection. Fig.3a shows an original book cover image and Fig.3b shows the result of using Histogram of Oriented Gradients features to detect the text candidates.
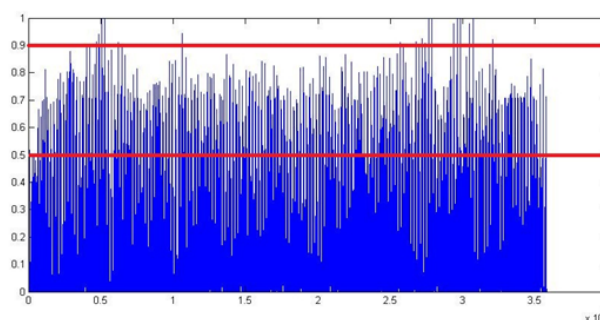

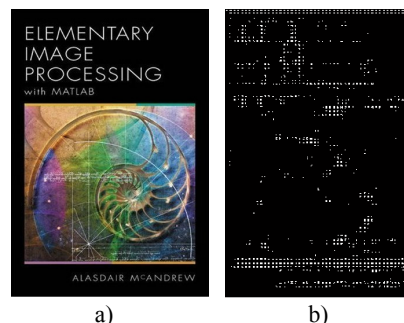Fig. 2. Histogram of Oriented Gradients and thresholds for text candidates.

a)   b)

Fig. 3. Process of detecting lines of text.
a) Input image;   b) Regions that are text candidates;

2.1.2 Removing non-title areas using area constraint: At the first step, RGB components of the input color image are combined to give an intensity image Y as: $Y = (R + G + B)/3$,

where R, G and B are the red, green and blue components of the input color image, respectively. The color components may different in a text region (and/or its surrounding background) while having an almost constant intensity. So, the intensity image Y is processed in the next steps of the algorithm rather than the color components R, G and B. If the input image is originally a gray-level image, the image is processed directly starting to next step. At the second step, our system uses the Canny operator for edge detection. All the pixels will be grouped into connected components. We later may refer to these connected components as regions or areas. The reason for using edge information is that a common title usually has clear and strong edge. Among all the edges, only the regions that are text candidates - resulted from the previous step are kept. Then at the next step, our system will remove the regions may contain some too large and/or too small connected components since they could not be the title of the book cover image.

Generally, the regions that are too large might be the border or a straight line that covers the whole image. Moreover, the regions that are too small might be the background but not the title regions. As the title of the book cover usually has big size but not too big and too small, we choose the areas inside the two values: The first value is the average area of all regions. The second value equals four times the first value. Our system first calculates the areas of all the connected components in the edge image. Only the connected components (CC) that are satisfied the oriented gradient in the previous step, are then processed. If the CC has area greater than the average area and less than four times the average area, it is then selected as the title candidate. The range from Average Area to 4*Average Area is appropriate with the purpose of title extraction. We show the area chart of all connected components in this Fig. 4 below. Fig.5b shows the edge map after applying Canny operator. Fig.5c shows the result after applying area constraint.
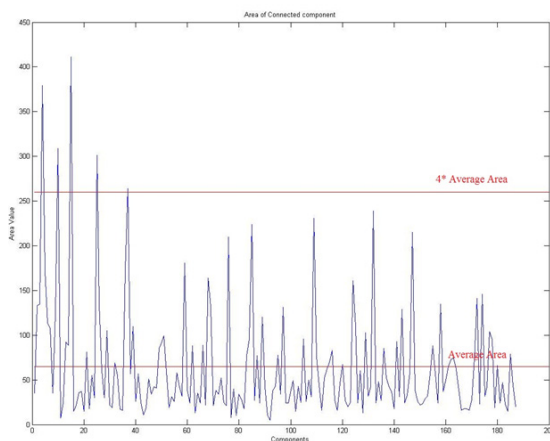

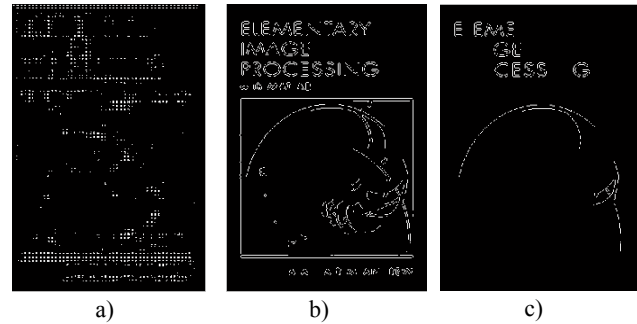Fig. 4. Area chart of connected components.

Fig. 5. Process of removing non-title areas using area constraint. a) Result of previous step;
b) After applying Canny operator; c) After applying area constraint.

2.1.3 Removing non-title using color information: We believe that most text strings originally have uniform color. After satisfying the oriented constraint and the area constraint, the remaining connected components contains many candidates for title regions or the numbers of connected components that belong to title regions are larger than the numbers of connected components belong to background. Accordingly, in this step, the remaining regions are divided into just two clusters: title (with same color) and background. The cluster that has larger number of connected components is considered to be the title candidates. Fig.6b shows the title candidates after applying k-Means clustering.
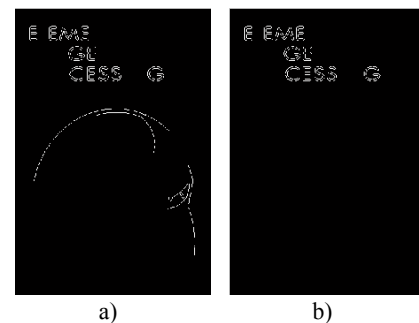

Fig. 6. Process of removing non-title using color information.
a) Result of previous step;
b) After applying K Means clustering.

2.1.4 Removing non-title using length constraint: After removing non- title using color information, sometimes the noise still appears if the color of the text is the same color with the noise regions. Thus, our system has one more step to eliminate these noises, by using the length constraint. We first compute the height of all connected components. Then applying heuristics include the aspect ratio and size of each connected components, the system could filter out non-title regions. The system just keeps the candidate title regions by filtering out non-character connected components. The title candidates that satisfy these three conditions as in Eq. (1) are kept.

$$\begin{cases} threshold = \min(\overline{height}, \dfrac{L}{25}) \\ \quad h > threshold \\ \quad h < \dfrac{L}{3} \end{cases} \qquad (1)$$

Where $\overline{height}$ is the average height of all the characters in the line, L is the height of the book cover image, h is the height of a connected component. Due to the arrangement of composition, the height of the characters is normally not higher than one third of the total height of the book cover image. Accordingly, we have the maximum threshold as one third of the total height of the book cover image. Besides, the height of the characters should not be too small. Here we choose the threshold by finding the minimum value between one twenty fifths of the book's height and the average height of all the characters in the line. Correspondingly, the minimum threshold is chosen that way. Therefore, in this step we keep only the components that have the height between the minimum threshold and maximum threshold. Fig.7b shows the result after eliminating the non-title regions by length constraint.
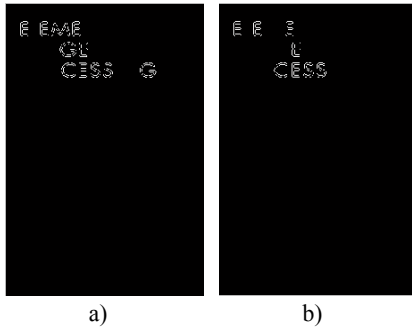

Fig. 7. Process of removing non-title using length constraint.
a)   Result of previous step;
b) after applying length constraint.

2.2 Refinement and Recovery

The final step of the system is refinement. Because the title candidates obtained from the previous step are not sufficient or the obtained results do not contain the whole title of the book cover image, the purpose of this step is to try to recover all characters of the titles. In order to obtain that whole line, we first calculate all the bounding boxes of the connected components. At this time, we know the candidate title in a line. Indeed, all the characters in the same word if they have the y-coordinate nearly the y-coordinate of the detected characters. Thus, we know the average height of the characters in a line as well as the top left and bottom left coordinate of the detected character in the line. If the bounding box satisfies two conditions, it could be in the text line. Equation (2) shows the first condition. The constraint is about the height of the characters in a line. Normally, the height of characters is not larger than two times the average height. The height is also greater than half times the average height. Figure 3 shows us the constraint for the height and coordinate of the characters in a line. The second condition is about the y coordinate of the bounding box. The top left y coordinate of a CC should be greater than the y coordinate of the top left and smaller than the

y coordinate of the bottom left. Top left y coordinate equals to the average value of y coordinate in the line minus half times the average height of y. Bottom left equals to the average value of y coordinate in the line plus 3/2 times the average height. Equation (3) shows this constraint.

$$\frac{1}{2}\overline{height} \le height \le 2\overline{height} \qquad (2)$$

$$\overline{y} - \frac{1}{2}\overline{height} \le y \le \overline{y} + \frac{3}{2}\overline{height} \qquad (3)$$

where $\overline{height}$, $\overline{y}$ is the average height of the characters in the line and the average value of y coordinate in the line, respectively. Fig.8 shows an example of the height and coordinate of the characters in a line.
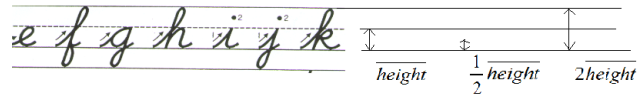

Fig. 8. Example of the height and coordinate constraint of the characters in a line.

Fig.9b shows the result of this refinement step. The title is recovered from the candidate text regions and extracted from the image. Fig.9c shows the result of text extraction in the original book cover image.
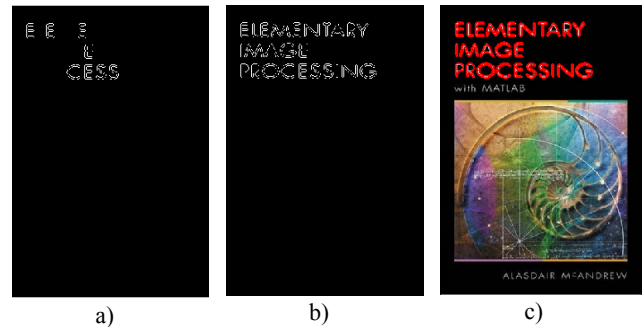

Fig. 9. Process of refinement and recovery.
a) Result of previous step;   b) After refinement process;
c) Final result.

3. EXPERIMENTAL RESULTS

Using our system, we can find the promising text candidates. Figure 8 shows the input and the results of the proposed system. We can see many title candidates which are described by the red color in the figure are correctly detected.
First of all, characters with variance of height can be detected by using our system. Fig.10a shows that the heights of all the characters in the title areas are approximately the same. However, in Fig.10b and Fig.10c, the characters in the line have variety heights. The top left y-coordinates or bottom left y-coordinates of the characters are also different. For both two situations, our proposed system can detect and extract all characters of the title book in the corresponding line. In case book cover images have tilted alignment characters, such as in

Fig.10d, 10e, 10f, our proposed system can also find the title regions. Our proposed system also works well with book cover images that have complex background, such as in Fig.10g, 10h, 10i, 10k.

Beside the advantage of title detection correctly as shown in Fig.10, the method still has some errors in finding redundancy candidates as show in the Fig.11b and Fig.11a. In addition when the color of the title is different, the method could not detect the completely title characters in the cover book images. The result is shown in Fig.11c. However, the error is acceptable since it is quite small compared with the whole cover book image and the title regions that are detected.

We take an experimental with the dataset of 100 cover book images. Table 1 below shows the experimental results. To compare the effectiveness of our proposed method, another text extraction is also implemented with the same database. This method seeks to find the value of stroke width for each image pixel, and use this information on the task of text detection in natural images [12]. However, this is appropriate with the purpose text extracting, but not the title extracting. Therefore, with this method, we expand the topic as text extracting. Therefore, we consider if this method can detect exactly characters in the cover images as successful (not only title areas are also acceptable). If no any characters are detected or the detected characters are not sufficient, we consider this as Deficiency. The results are considered as Wrong candidates when the method detects non-text as text area. Lastly, the redundancy is not important with text extraction, because if the method detects some redundancy, it should be classified as wrong candidates. Therefore, we do not concern about this factor. Using the method in [12], 61% of the results have wrong candidates and 17% of the results are deficiency. Only 22% of the results are correctly detected. Using our proposed method, more than 86% of the cover book images title areas are correctly detected. The rate of successfully detected title regions is very high. The number of redundancy title areas is 7%, namely 7 images. This may happen when the height of author name characters is close to the height of title areas. We also find out that 5% of the cover book images are wrongly detected for title candidates. In addition, the system cannot detect all parts of the title in 2%. This may happen when the title characters has more than two colors, the system will only keep the majority color, therefore lack of candidates for the title characters.
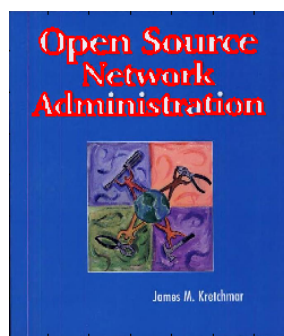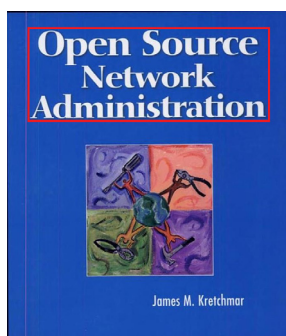
From table 1, we can see that encouraging results can be obtained by using our method. It proves the effectiveness of our proposed method.
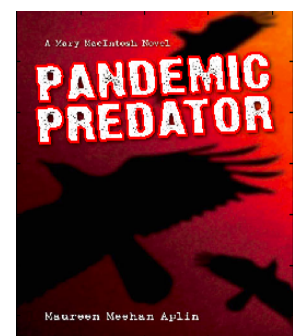

a)

b)

c)

d)

e)   f)

g)   h)

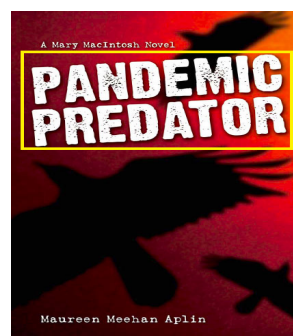i)   j)
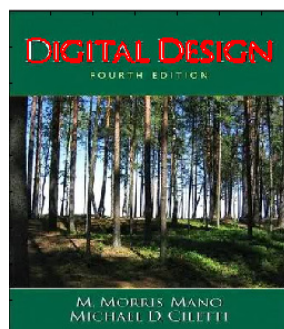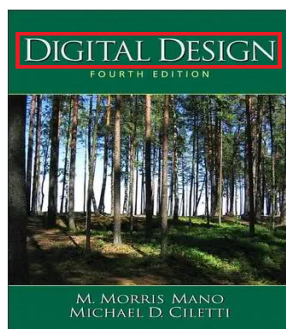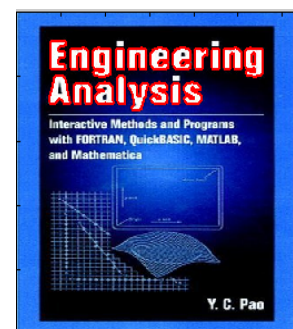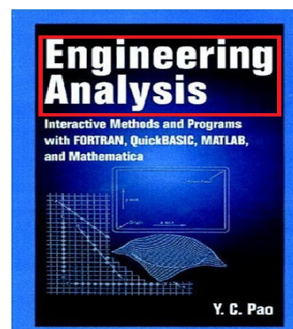
Fig. 10. Inputs and outputs of our proposed system.
Two first rows: variety height of characters in book cover images; Third row: different angle tilt in book covers images;
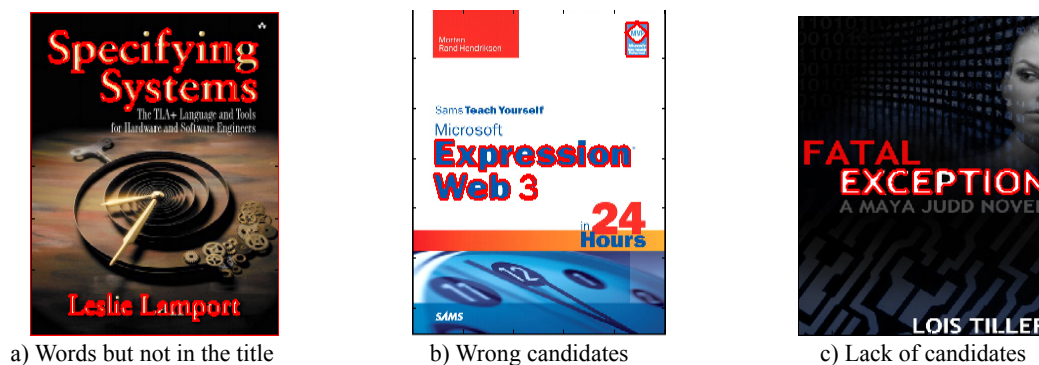Two last rows: book cover image with complex background.

a) Words but not in the title   b) Wrong candidates   c) Lack of candidates

Fig. 11. Example of incorrect detections.

Table 1. PERFORMANCE OF TITLE EXTRACTION

|		Incorrect		
	Correct	Redundancy	Wrong candidates	Deficiency
Our method	86%	7%	5%	2%
[12] SWT method	22%	~	61%	17%

4. CONCLUSIONS

Text extraction in general and title extraction in particular can be used in many applications and brings us many benefits. One of its most important applications is be the output of the pre-processing in recognizing the title of the books in the library. Results of our proposed method can be further used in the OCR systems and hence play an important role in increasing the accuracies of the system. The better result of title regions that are detected, the better accuracies of the OCR system we can obtain. After recognizing the book, all the information of the recognized book will be displayed, accordingly, the users can easily get all the information that they want to know about the book.

ACKNOWLEDGEMENT

REFERENCES

[1] Do Yen, S. H. Kim, S. C. Park, Ha Le, Y. J. Chen, S. H. Jeong, I. S. Na (2012), "Generation of Training Database Using a Noise Model for OCR Systems", in *The 6th International Conference on Ubiquitous Information Management and Communication, Kuala Lumpur, 2012.*(CD-Pub.)

[2] Ha Le, S. H. Kim, S. C. Park, Do Yen, Y.J. Chen, S. H. Jeong, I. S. Na (2012), " Automatic Generation of Database to Train New Fonts for OCR Systems" in *The 2012 International Workshop on Advanced Image Technology IWAIT2012, Ho Chi Minh City, 2012.* pp. 178-182.

[3] Kee chul Jung, K. I. Kim, Anil K. Jain.(2004),"Text Information Extraction in Images and Video: A Survey", *Pattern Recognition*, Vol. 37 No.5, pp. 977-997.

[4] C. J. Park, K. A. Moon, O. W. Geun, and H. M. Choi(2000), "An efficient extraction of character string positions using morphological operator", *in Proc. IEEE Int. Conf. Systems, Man, Cybernetics, 2000,* vol. 3, pp. 1616–1620.

[5] Z. Yu, K. Karu, and A. K. Jain (1995),"Locating text in complex color images," *in Proc. 3rd Int. Conf. Document Analysis and Recognition,1995*, pp. 146–149.

[6] D. Chen, H. Bourlard, and J. P. Thiran (2001), "Text identification in complex background using SVM,*" in Proc. IEEE Computer Soc.Conf. Computer Vision and Pattern Recognition, 2001*, pp. 621–626.

[7] Yassin M. Y. Hasan and Lina J. Karam. (2000),"Morphological Text Extraction from Images", *IEEE Transaction on Image Processing*, Vol. 9, No. 11, pp.1978-1983.

[8] M. Chen and X. Ding(2000),"Analysis, understanding and representation of Chinese newspaper with complex layout," *in Proc. 7th IEEE Int. Conf. Image Processing, Vancouver, 2000, BC,* Canada, vol. 2, pp. 590–593.

[9] A. Antonacopoulos, B. Gatos, and D. Bridson (2005),"ICDAR 2005 page segmentation competition," *in Proc. ICDAR,2005,* Seoul, Korea, pp.75–80.

[10] A. K. Jain and B. Yu (1998), "Document representation and its application to page decomposition", *IEEE Trans. Pattern Anal. Mach. Intell.,* Vol. 20, No. 3, pp. 294–308.

[11] Navneet Dalal, Bill Triggs, "Histograms of Oriented Gradients for Human Detection". *Computer Vision and Pattern Recognition,* 886-893, 2005.

[12] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform", in Proc. CVPR, 2010, pp.2963-2970.

[13] Nguyen Noi Bai, Kim Nam and Youngjun Song "Extracting curved text lines using the chain composition and the expanded grouping method", in Korea information processing society, vol. 14-B, No.6, pp.453-460, Oct. 2007.
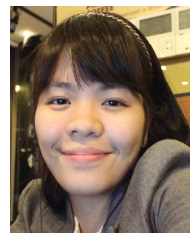
Yen Do
She received her B.S. degree in Information Technology from Hanoi University of Science and Technology, Vietnam in 2009. Since 2011, she has been a master process student in the Department of Computer Science, Chonnam National University, Korea. Her main research interests include pattern recognition, document image processing and ubiquitous computing.

Soo-Hyung Kim
He received his B.S. degree in Computer Engineering from Seoul National University in 1986, and his M.S. and Ph.D degrees in Computer Science from Korea Advanced Institute of Science and Technology in 1988 and 1993, respectively. From 1990 to 1996, he was a senior member of research staff in Multimedia Research Center of Samsung Electronics Co., Korea. Since 1997, he has been a professor in the Department of Computer Science, Chonnam National University, Korea. His research interests are pattern recognition, document image processing, medical

image processing, and ubiquitous computing.

In Seop Na
He received his B.S., M.S. and Ph.D. degree in Computer Science from Chonnam National University, Korea in 1997, 1999 and 2008, respectively. Since 2012, he has been a research professor in Department of Computer Science, Chonnam National University, Korea. His research interests are image processing, pattern recognition, character recognition and digital library.