An Improved Approach for Localization of Text Regions from Complex Document Images

Madeena Sultana, Sabrina Sharmin, Farhana Sabrina, Mohammad Shorif Uddin

Abstract—Text regions extraction from document images containing both texts and graphics is an important step of any optical character recognition system. This paper describes an improvement over existing methods for localization of text regions from document images. The improvement is achieved by accommodating distinctive features like regularity in frequency, orientation, width, area, spatial cohesion etc. to identify text blocks in a document image containing both text and graphics. Proposed technique is tested on MARG dataset of multiple layouts and large varieties of color document images collected from web. Experimental result confirms the improvement of extraction accuracy by suppressing the false alarms notably.

Keywords—Discrete wavelet transform (DWT), document image segmentation, Haar wavelet transform (HWT), Run length smoothing algorithm (RLSA), text localization.

1 Introduction

TEXT is usually the main source of information in documents and accurate text detection can greatly facilitate optical character recognition. Automatic recognition, reading, and storing information are the demands of modern technology. Therefore, text localization and extraction is a key area of research in document image analysis.

However, locating and extracting textual data is not an easy task. Since texts are often embedded in different font styles, sizes, orientations and colors against a complex background. Moreover, low contrast between the text and the complicated background often makes text detection extremely challenging. To address these problems, a large number of new methods for text localization, extraction and optical character recognition have been proposed recently. Some of the well known approaches are: (i) morphology based segmentation [1], ii) pixels based classification [2], iii) connected component based classification [3]-[5], iv) edge based segmentation [6], v) texture based segmentation [7], vi) frequency based classification [8]-[10], vii) run length based segmentation [11],[12], and viii) sparse representation based segmentation [13]. The survey papers [14]-[16] enlist more techniques for layout analysis of document images.

In text extraction process, the most important step is to find approximate locations of text lines in a gray-scale image. In this paper we address the problem of locating the textual data in an image. Our proposed system employs both connected component and discrete wavelets to localize text from complex document layouts.

The paper is organized as follows. Section 2 deals with the related work. Section 3 gives a step by step description of proposed method. Experimental results are illustrated in Section 4. Finally, conclusions and future works are summarized in Section 5.

2 Previous Work

Many researchers have been investigating various wavelet based techniques to retrieve textual information present in document and scene images. Li and Gray [8] used distribution characteristics of wavelet coefficients for document image segmentation. Liang and Chen [9] employed Haar wavelet transform to detect edges of candidate text regions. Kumar et al. [10] proposed globally matched wavelet filters and Markov random field (MRF) based processing for text extraction from document and scene text images. In 2004, Liang and Chen's proposed a simple approach and it performs well for separating captions and titles from video and scene images. However, it is often unable to differentiate between text and non-text components in document images and hence produces large false alarms especially when the layout is complex. In this paper, we introduce an improvement of Liang and Chen's segmentation algorithm to suppress false alarm and generalize it for separating text and non-text components from document images as well.

3 Proposed Method

We propose an improved and efficient method to extract text regions from document images containing both text

- *Madeena Sultana is with the Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh. E-mail: madeena.sultana@ulab.edu.bd.*
- *Sabrina Sharmin is with the Department of Computer Science and Engineering, State University of Bangladesh, Dhaka Bangladesh. E-mail: monika.329@gmail.com.*
- *Farhana Sabrina is with the Department of Electronics and Telecommunication Engineering. ULAB, Dhaka, Bangladesh. E-mail: farhana.sabrina@gmail.com*
- *Mohammad Shorif Uddin is the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. E-mail: shorifuddin@gmail.com.*

Manuscript received on 28 August 2012 and accepted for publication on 30 September 2012.

and graphics.

The whole text extraction process is divided into three distinct parts:

a. Candidate region extraction
b. Noise reduction
c. Text localization

Fig. 1 depicts the block diagram of our proposed method. We followed the method of Liang and Chen [9] for candidate region extraction. Then we added a noise reduction step to reduce false alarms and a connected component analysis step for localization of text regions. In our process, we have not considered the small text regions like the page number and vertical text lines.

3.1 Candidate Region Extraction

According to the proposed method by Liang and Chen the candidate region extraction process has three major subsections.

a. Edge detection
b. Thresholding
c. Region detection

However, first of all, if the input image is colour, it is converted to an intensity image I by combining the RGB components of the original image as follows:

$$I = 0.299R + 0.587G + 0.114B \quad (1)$$

3.1.1 Edge Detection

The discrete wavelet transform has many applications in signal analysis and image processing. One important application among those is edge detection. We have used Haar wavelet transform (HWT) as it is simpler and operates fastest among all wavelets. Two-dimensional (2D) HWT decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH). We can obtain the following edge features of the original image from three detail components produced by 2D Haar (DWT).

a. HL sub-band detects vertical edges.
b. LH sub-band detects horizontal edges and
c. HH sub-band detects diagonal edges

For example, in Fig. 2 the gray-level image is decomposed into 2-D Haar DWT. From the three detail component sub-bands (LH, HL, and HH) in Fig. 2 the candidate text edges can be detected.

3.1.2 Thresholding

Thresholding is a simple technique for separating image objects from the background. Since, the intensity of the text edges is higher than that of the non-text edges we can preliminarily remove the non-text edges by selecting an appropriate threshold value for each sub-band. In this subsection, dynamic thresholding [9] technique is applied to calculate the target threshold value T. The target threshold value is determined by performing the following equations on each pixel of each sub-band:

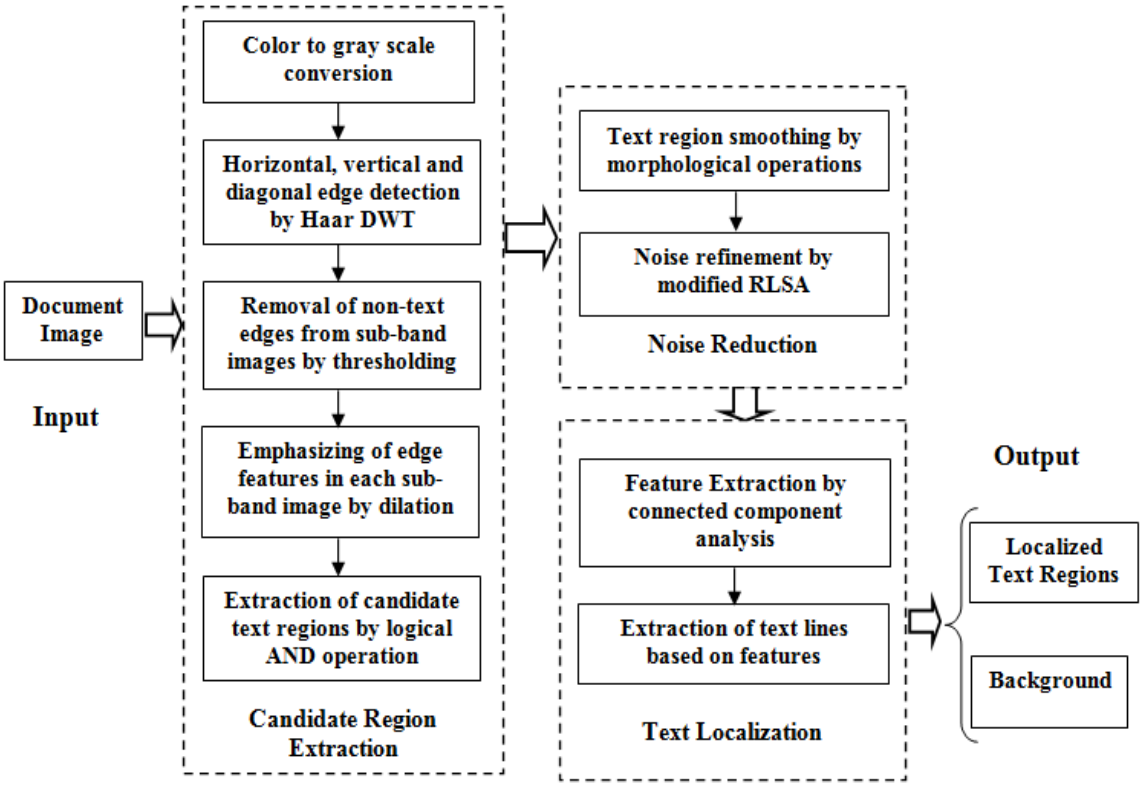


Figure 1: Block diagram of proposed model

$$T = \frac{\sum (subband \quad (i, j) \times e(i, j))}{\sum e(i, j)} \tag{2}$$

Where e(i,j) denotes intermediate sub-image and is calculated by the following equation:

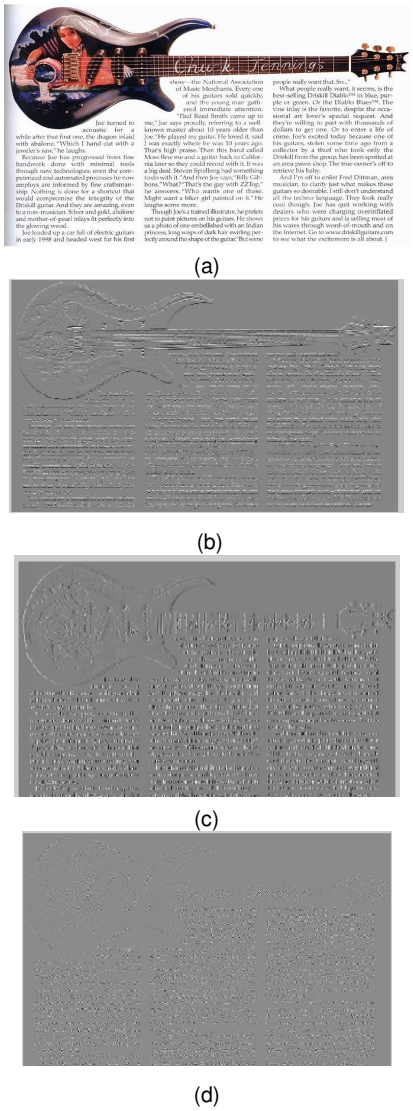
$$e(i, j) = Max(|g1 * *subband(i, j)|, |g2 * *subband(i, j)|) \tag{3}$$

$$g1 = [-1 \quad 0 \quad 1], g2 = [-1 \quad 0 \quad 1]^t \tag{4}$$

In (4), g1 and g2 are two mask operators and ** denotes two dimensional linear convolution operation.

$$b(i, j) = \begin{cases} 255, & if \quad subband \quad (i, j) > T \\ 0, & otherwise \end{cases} \tag{5}$$

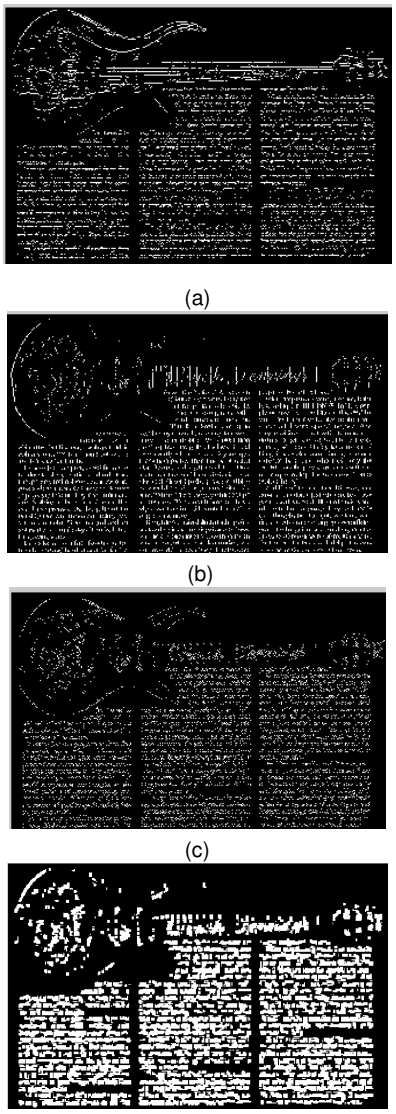
Using (2)-(4), the threshold value (T) is determined dynamically for different sub-bands and the binary image (b) is then obtained by comparing T with every pixel value of each detail components.



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

Figure 2: Haar discrete wavelet transform: a) Original image, b) Horizontal sub-band (LH), c) Vertical sub-band (HL), d) Diagonal sub-band (HH) image.

In Figs. 3(a), 3(b), and 3(c) show horizontal, vertical, and diagonal sub-band images after thresholding, respectively.



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

Figure 3: Binary Images: a)Horizontal threshold image, b)Vertical threshold image, c)Diagonal threshold image, d)after AND operation.

3.1.3 Region Detection

Different morphological operators are used to connect isolated candidate text edges in each binary image of the detail sub-band components. In this paper, we used 5×7 pixels for horizontal operator, 5×5 pixels for diagonal operator and 7×5 pixels for vertical operator. The operators are determined through experimentation of a wide range of document images. Since vertical edges, horizontal edges and diagonal edges are intermixed in text regions we can detect candidate text regions by logical AND operation of the dilated binary images. Fig. 3(d) depicts the image after AND operation.

3.2 Noise Reduction

Noise reduction is accomplished by the following two

steps:

a. Text region smoothing by morphological operation
b. Noise refinement by Run-length Smoothing Algorithm (RLSA)

Before noise refinement the following morphological operations are performed for smoothing the candidate text regions.

$$X_1 = ((X_0 \bullet B_{Square}) \circ B_{Square}) \quad (6)$$
$$X_2 = ((X_1 \bullet B_{Rectangle}) \circ B_{Rectangle}) \quad (7)$$

Where, X_0 is the binary image containing candidate text regions along with false alarm, $\circ$ and $\bullet$ denote morphological opening and closing respectively. These operations produce smooth text lines by removing narrow bridges between adjacent text regions and small spurious non-text regions, and fill in narrow holes in regions. The resulting image shows which blocks in the output image are informative.

At this stage, several closely parallel edges may be falsely detected as text. Therefore, we incorporated horizontal and vertical projection profile analysis with (Run-length Smoothing Algorithm) RSLA to separate the true texts from the candidate ones. In this RLSA approach, candidate text regions are refined based on empirical threshold values, Tx and Ty calculated for the horizontal and vertical projections, respectively. RLSA then eliminates horizontal and vertical white runs whose lengths are smaller than the corresponding threshold value.

$$T_x = \frac{1}{4} \times \text{Mean of (horizontal projection profile)} \quad (8)$$

$$T_y = \frac{1}{3} \times \text{Mean of (Vertical projection profile)} \quad (9)$$

Fig. 4 illustrates the steps of our RLSA.

3.3 Text Localization

At this stage, the non-text components are then removed using various characteristic features of texts. Our text localization process has two major steps:

a. Extraction of features by connected component analysis
b. Localization of text regions based on features

In document images a number of lines typically have approximately same height and width. The small regions with irregular shapes usually belong to the non-text regions. Based on these heuristic rules we processed candidate text regions as text lines. We computed various features such as: average area, width, and height of each connected component and removed regions having too small width and area comparing to the average value. At the end, only those regions in the final image are retained, which have area close to a rectangle and width greater than half of the mean width of all candidate regions. This process prunes the non-text regions mostly and filters out promising candidates. The resultant image of localized texts is shown in Fig. 5.

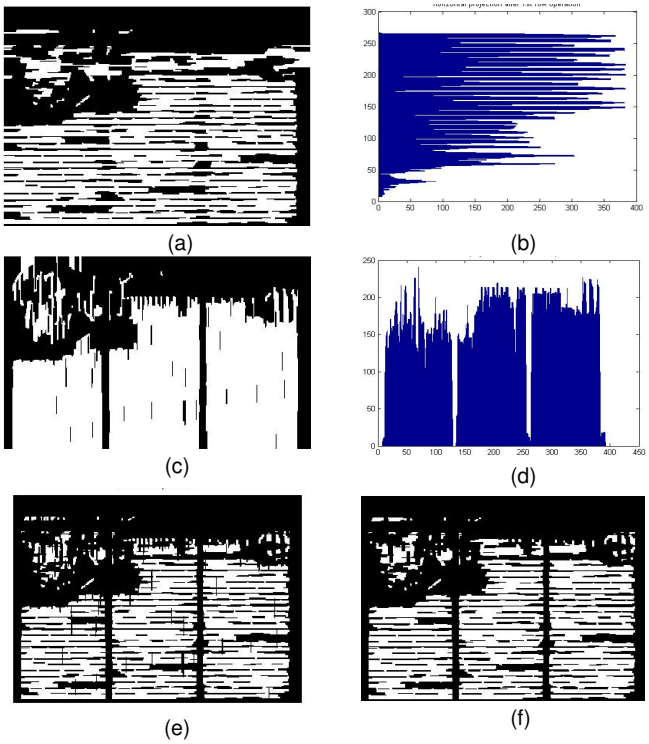


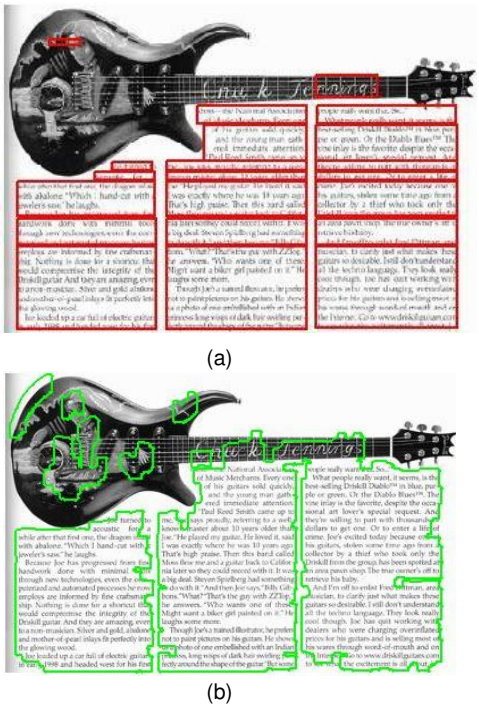
Figure 4: RLSA: a) Row (horizontal) operation, b) Horizontal projection profile, c) Column (vertical) operation, d) Vertical projection profile, e) AND operation, f) Again horizontal, vertical and AND operation.

Figure 5: Text Localization: a) Input Image, b) The localization of text using proposed method

4 Experimental Results and Discussions

To evaluate the performance of proposed approach, we have selected 35 downloaded images of books, journals, and magazines from the Internet containing complex backgrounds, graphics, different font sizes, and overlapping styles as the experimental data set. To demonstrate

the efficiency of our method we have also tested proposed method in another database which contains 45 document images from MARG [17] dataset. The latter dataset is created by randomly picking 5 images from each of nine classes of the page layouts of MARG. Our proposed method performed equally well for regular and irregular layouts along with complex background. Fig. 6 compares the resultant images by proposed method and Liang and Chen's method. In Fig.6 it is visually evident that resultant image obtained by our method can locate text region more accurately and contains less false alarms than Liang and Chen's method. Fig. 7 represents some results of applying our method on MARG dataset.



(a) (b)

Figure 6: Text localization of a downloaded image by: a) proposed method, b) Liang and Chen's method

For quantification of accuracy of localization, rate of accuracy is calculated to evaluate the performance. Rate of accuracy of text localization is defined as follows:

$$Accuracy\ Rate = \frac{Total\ \ localized\ \ text\ \ regions}{Total\ \ text\ \ regions} \times 100 \quad (10)$$

Table 1 shows the accuracy rate and false alarms obtained by proposed method and Liang and Chen's method [9], respectively for downloaded images. Images containing total 99 text regions we obtained an accuracy of 96% by the proposed method.

Texts with very large font size are treated as graphics by proposed method and, therefore, can not be localized or partially localized. However, from Table 1 and resultant images it can be concluded that proposed method has better performance in terms of both accuracy rate and suppressing false alarms for localization of texts in document images.

TABLE 1
PERFORMANCE COMPARISON OF TEXT REGION LOCALIZATION

	Proposed method	Liang and Chen's method
Total text regions	99	99
Total extracted text regions	95	81
False alarms	191	714
Accuracy rate	96%	82%

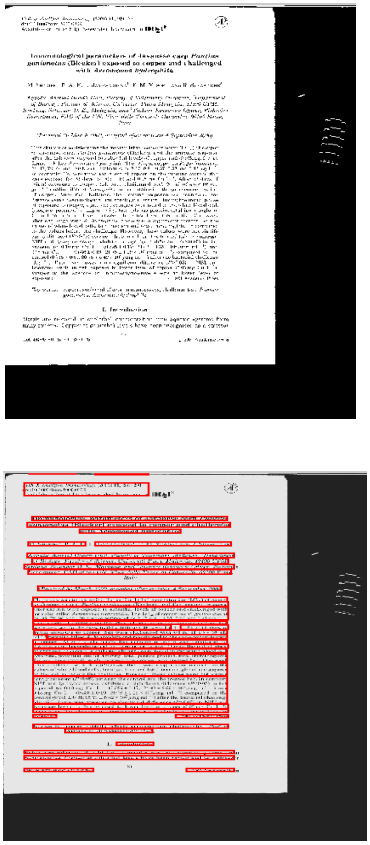


Figure 7: Text localization by proposed method: Sample output document images from MARG dataset.

5 CONCLUSIONS

Text extraction from document images is a challenging task because of the complex background and multi-resolution criteria. Moreover, degradations introduce during scanning or copying a paper document. This paper presents an efficient and simple method to locate texts in documents. To improve the accuracy we modified Liang and Chen's approach by accumulating RLSA with connected component analysis. Our experimental results show that, along with improving accuracy, our method reduces false alarms from resultant images. Moreover,

compared with other methods our technique relied on adaptability of predefined text region features. However, successful detection with our method dropped down significantly under cases when textual regions are vertical or scattered and intertwined heavily with irregular graphical blocks and backgrounds. Future work, involves dealing with this problem for more accurate detection mechanism.

REFERENCES

[1] Hasan, M. Y. and Lina J, Y. K," Morphological text extraction from image," *IEEE Transactions on Image Processing*, vol. 9, no.11, pp.1978 - 1983.

[2] Moll, M. A., Baird, H. S., and An, C., "Truthing for pixel-accurate segmentation," *Document Analysis Systems, the Eighth IAPR Int. Workshop*, pp. 379-385, Sep. 2008.

[3] L. A. Fletcher and R. Kasturi,"A robust algorithm for text string separation from mixed text/graphics images," *IEEE Transactions on PAMI*, vol. 10, no. 6, pp. 910-918, 1988.

[4] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, "A Robust algorithm for Text detection in images," *Proc. of the 3rd International Symposium on Image and Signal Processing and Analysis*, 2003.

[5] Bukhari, S. S., Shafait, F., and Breuel, T. M., "Document image segmentation using discriminative learning over connected components," *Proc. 9th IAPR Workshop on Document Analysis Systems*, pp. 183-190, 2010.

[6] M. Ly u, J. Song, M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 243 –255, 2005.

[7] A.K. Jain, Y. Zhong, "Page segmentation using texture analysis, "*Pattern Recognition*," vol. 29, no. 5, pp.743–770, 1996.

[8] J. Li and R. M. Gray, "Context based multi-scale classification of document images using wavelet coefficient distribution, " IEEE Trans. Image Process., vol. 9, no. 9, pp. 1604 –1616, Sep. 2000.

[9] Chung-Wei Liang and Po-Yueh Chen, "DWT based text localization," *International Journal of Applied Science and Engineering*, vol. 2, no. 1, pp.105-116, 2004.

[10] Sunil Kumar, Rajat Gupta, Nitin Khanna, Santanu Chaudhury, and Shiv Dutt Joshi, "Text extraction and document image segmentation using matched wavelets and MRF model", *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2117-2128, Aug. 2007.

[11] O. Okun, D. Doermann, and M. Pietikainen, "Page segmentation and zone classification: the state of art," Technical Report LAM-TR-036, CAR-TR-927, CS-TR-4079, University of Maryland, College Park, Nov. 1999.

[12] Hung-Ming Sun, "Enhanced constrained run-length algorithm for complex layout document processing," *International Journal of Applied Science and Engineering*, vol.3, pp.297-309, April. 2006.

[13] Ming Zhao, Shutao Li, James Kwok, "Text detection in images using sparse representation with discriminative dictionaries", *Image and Vision Computing*, vol.28, pp.1590–1599, 2010.

[14] Y.Y. Tang, S.W. Lee, C.Y. Suen, "Automatic document processing: a survey," *Pattern Recognition*, vol. 29, no. 12, pp. 1931–1952, 1996.

[15] Nawei C hen, Dorothea Blostein, "A survey o f document image classification: problem statement, classifier architecture and performance evaluation," *International Journal on Document Analysis and Recognition (IJDAR)*, vol.10, pp.1–16, 2007.

[16] K. Tombre, S. Tabbone, L. Pélissier, B. Lamiroy, and P. Dosch, "Text/graphics separation revisited", *Proc. Document Analysis Systems*, pp.200-211, 2002.

[17] MARG document image dataset. Retrieved on 1 June 2011.
http://marg.nlm.nih.gov/gtintro.asp

Madeena Sultana received her M.S. and B.Sc. in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, in 2011 and 2008, respectively. In October 2010, she joined the Dept. of Computer Science and Engineering (CSE) at the University of Liberal Arts Bangladesh. She was a Lecturer at the Dept. CSE of Northern University Bangladesh from June, 2008 to July, 2009. She has published more than 10 papers in peer-reviewed journals and conference proceedings. Her research interests include Digital Image Processing, Computer Vision, and GPU Computing. She is a member of the International Association of Computer Science and Information Technology (IACSIT).

Sabrina Sharmin received her B.Sc degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka in 2009. She is now pursuing her MS at the same university. She has been doing her research in the field of Digital Image Processing with special concentration on physical layout analysis of document images.

Farhana Sabrina received her B.Sc (Engg.) in Electrical and Electronic Engineering from Chittagong University of Engineering & Technology (CUET), Chittagong in 2008. She joined in the department of Electronics & Telecommunication Engineering at University of Liberal Arts Bangladesh (ULAB) in 2010 and continuing the job. Previously she worked in Warid Telecom Bangladesh Limited (presently known as Airtel Bangladesh) and Premier University, Chittagong. She is a member of IEEE and IEB.

Mohammad Shorif Uddin received his PhD in Information Science from Kyoto Institute of Technology, Japan, Masters of Education in Technology Education from Shiga University, Japan and Bachelor of Science in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET). He joined the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka in 1992 and currently serves as a Professor of this department. In addition, he is serving as an Adviser of School of Science and Engineering, ULAB. He began his teaching career in 1991 as a Lecturer of the Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET). He undertook postdoctoral research at Bioinformatics Institute, A-STAR, Singapore, Toyota Technological Institute, Japan and Kyoto Institute of Technology, Japan. His research is motivated by applications in the fields of computer vision, pattern recognition, blind navigation, bio-imaging, medical diagnosis and disaster prevention. He has published a remarkable number of papers in peer-reviewed international journals and conference proceedings. He holds two patents for his scientific inventions. He received the Best Presenter Award from the International Conference on Computer Vision and Graphics (ICCVG 2004), Warsaw, Poland. He is the co-author of two books. He is also a member of IEEE, SPIE, IEB and a senior member of IACSIT.