Connected Component Based Approach for Text Extraction from Color Image

Kamrul Hasan Talukder
Computer Science & Engineering Discipline
Khulna University, Khulna
Bangladesh.
k.h.t@alumni.nus.edu.sg

Tania Mallick
Computer Science & Engineering Discipline
Khulna University, Khulna
Bangladesh.
mallick.tania@yahoo.com

***Abstract*---Image based text extraction is one of the fastest growing research areas in the field of multimedia technology. The extraction of text from a complex or more colorful images is a challenging problem. Text data present in images contains useful information for habitual explanation, indexing, and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. For fast extracting text from images, we have proposed a connected component based approach which identifies more accurately for small or large texts in the image. The text extraction process starts with conversion of the color image to gray scale image and then it converts the gray scale image into a binary image. Then each text region is marked and the text is extracted from the image. Finally, the extracted text is written into another gray scale image. The experimental results demonstrate that the performance of the proposed method is superior compared to some recent approaches.**

Keywords--- Image, image processing, text extraction, connected component.

I. INTRODUCTION

Due to the rapid improvement of information technology and internet communication, now-a-days people have been overwhelmed by the fast accumulation of digital information such as text, image, audio and video around the world. In real life situations, images contain much useful information for automatic annotation, indexing and structuring of images, document analysis, technical paper analysis, vehicle license plate extraction and object oriented data compression. Extracting and recognizing text in images has become a potential application in many fields like robotics, intelligent transport systems etc [1].The main modules of these types of applications are object localization, object extraction and text recognition [1].Text extraction from images are divided into two main types: single text string extraction and many text strings extraction.

Typical data/information extraction techniques are not appropriate for text extraction from images for many reasons. Firstly, different types of objects such as structural bars, company logos and smears may be mixed with characters. Secondly, background and text color may differ. Thirdly, the font style and size of the character may vary. Finally, there may be uneven lighting conditions in scene images [3].

Text extraction from an image is a classical problem in computer vision. Automatic recognition of text in images has been used in many applications [1]. Such applications are page segmentation, document retrieving, address block location, etc. Different techniques are used for image extraction such as- Content-based image/video, Scene text extraction etc.

Content-based image/video indexing is one of the typical applications of overlay text localization. Scene text extraction can be used in mobile robot navigation to detect text-based landmarks, vehicle license detection/recognition, object identification, etc. However variation of text due to difference in size, style, orientation, alignment, low image contrast and complex background make the problem of automatic text extraction extremely challenging.

A lot of research has been done on this field. After studying the related articles, it has been found that text and non-text classification is implemented by taking text area as a block or character by character [5].

In 2008, Chitrakala Gopalan and D. Manjula proposed a new approach based on a variation of contourlet transform on images [7]. In this method, texture details capture in different orientations at various scales and non subsampled contourlet transform (NSCT) is used [7]. They claimed that this work was much better when compared to CC or Edge based approaches [1]. But they accepted that it did not give satisfactory results for scene text images.

In the recent years, morphology based method have been used for text localization. In 2009,

J.Fabrizio, M. Cord, B. Marcotegui, proposed region based approach which seems to be more efficient. Such approach was ranked first during ImagEval campaign [8]. But the authors accepted that this method produced lot of false positives. In the same year, J.Sushma, M.Padmaja, proposed an edge based approach using mathematical morphology [11]. They used a feature map which is generated from images. Then dilation operation is applied to localize text. In that, edges are detected first. In the same year, Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen, presented a novel texture descriptor based on line-segment features for text detection in images and video sequences [13]. This method was applied to build a robust car license plate localization system.

In 2011, R. Chandrasekaran and RM. Chandrasekaran proposed a method for morphology based text extraction in images [1]. Their approach uses morphological operation for text localization and Support Vector Machine (SVM) for character recognition, along with some preprocessing and post-processing steps.

In 2012, Neha Gupta and V.K. Banga proposed a methodology for extracting text from images [2]. They use Discrete Wavelet Transform (DWT) for extracting text information from complex images. In the same year, Md. Shorif Uddin, Tanzila Rahman, Umme Sayma Busra and Madeena Sultana proposed another method using modified morphological filter along with an automatic clustering technique [3].

Here we have proposed a new method for text extraction based on connected component. The proposed method can extract text more accurately. The Recall Rate of Characters (RRC) of our method is 98.8% in average, which is better than Gu's method [14] and modified morphological method [3].

II. TEXT EXTRACTION FROM IMAGE

In this section, we have discussed about text extraction process from image which is based on [14].The whole text extraction process is divided into two distinct parts:

i. Primary process
ii. Text extraction process.

i. *Primary Process*

The decomposition procedure is implemented by the following morphological algorithm, which is applied to produce binary images.

$$X_i = [(X_0 - X_0 \circ r_iB_{Disk}) - (X_0 - X_0 \circ r_iB_{Disk})]_B - X'_{i-1} \quad (1)$$

$$X'_j = \bigcup X_i,\ X'_1 = \emptyset \qquad 0< j\leq i$$

$$X_i = \left| (X_0 \bullet r_iB_{Disk} - X_0) - (X_0 \bullet r_iB_{Disk} - X_0) \right|_B - X'_{i-1} \quad (2)$$

$$X'_j = \bigcup X_i,\ X'_1 = \emptyset \qquad 0< j\leq i$$

Where, X_0 is the original gray scale image, X'_i is the decomposed binary image, and B denotes threshold operation in a defined value. Equation (1) is used to decompose the images where characters are brighter than the background; if the characters are darker than background, (2) will be applied. The procedure of this processing is start with r_iB_{Disk}, a series of sub-images X'_i are produced in a recursive manner. The processing will be stopped when the image $X_0 \circ r_iB_{Disk}$ or $X_0 \bullet r_iB_{Disk}$ have no characters remained.

ii. *Text Extraction Process*

This is the most important part of this method. Extraction processing is divided into 3 distinct steps:

1. Feature emphasis
2. Character extraction
3. Noise reduction / refinement

1) *Feature Emphasis*

In this stage, the decomposed sub images are processed by a morphological filter to reduce noises and emphasize character region.

$$E_i = \begin{cases} (((X'_i \circ r_{i-1}B_{Disk}) \oplus r_{i+1}B_{Disk}) \circ r_{2i}B_{Disc}) \times X_0 & i\leq 10 \\ (X'_i - X'_{10}) \times X_0 & i>10 \end{cases} \quad (3)$$

$$E_i = \begin{cases} ((X'_{ki} \circ r_{i-1}B_{Disk}) \oplus r_{i+1}B_{Disk}) \circ r_{2i}B_{Disc} & i\leq 10 \\ (X'_{ki} - X'_{k10}) & i>10 \end{cases} \quad (4)$$

2) *Character Extraction*

Since character regions are the main component in E_i, those hold the peak values in the histogram. The peak values which are bigger than the average of all peak values are searched and is the selected peak is used as threshold for E_i to extract characters from it. The extracted characters are in Hi.

3) Noise Reduction/Refinement

The extracted characters are broken in Hi and there remain diverse noises. A morphological filter derived from conditional dilation is implemented to refine the extracted characters.

$$R_{i0} = H_i \circ r_{i-1}B_{square}$$
$$R_{in} = (R_{i(n-i)} \oplus r_5 B_{Disc}) \cap |X_0|_B \quad (5)$$
If $R_{ik} = R_{i(k-i)}$ then stop.

Finally, sub-images R_{ik} are united to obtain the entire resultant image X_r denoted by

$$X_r = X_{r-1} \cup R_{ik} \quad (6)$$

III. Proposed Method

The text extraction from an image is a sequential process. In our method, text extraction from an image is divided into three steps: i) Pre-processing the input image ii) Marking each text region iii) Text extraction process. The input of our proposed method is a color image. Thus some preprocessing of the input image is required. The image is first divided into two sub-images. If the image is not divided into sub images, some text which is very small in size is missing and some noise remains in the final extracted text image. Then two sub images are converted into two gray scale images which are then converted into two binary images. We applied the text extraction process for each sub images and then composed two sub-images. Finally, the extracted text is written into another gray scale image. If the input image is a gray scale image then the preprocessing steps are not required. The block diagram of our proposed method is shown in Fig.1:

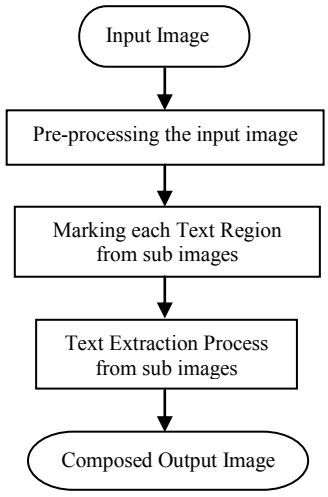

Fig 1: Block diagram of proposed method

The block diagram of pre-processing the input image is shown in Fig.2:

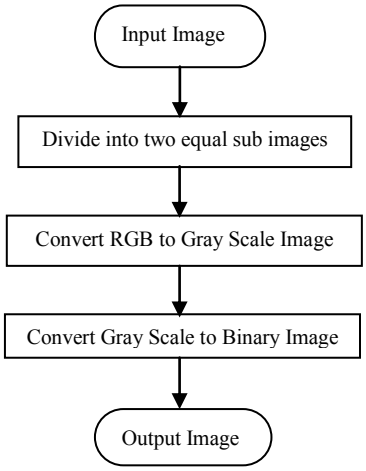

Fig 2: Block diagram of pre-processing the input image

A. Explanation of Our Proposed Method

- *Preprocessing the input image*

At first, we have divided the input image into two equal sub-images. Then we converted two sub images into two gray scale images. The RGB components are combined to give an intensity image. Usually, color images are normally captured by the digital cameras. The pictures are often in the Red-Green-Blue color space. Intensity image Y is given by:

$$Y = 0.299R + 0.587G + 0.114B$$

Image *Y* is then processed with binary image.

- *Marking Each Text Region*

In this step, the input is a binary image which is obtained from the previous step. Then, the text extraction process is in progress of two sub images. At first each text elements are bounded one by one separately. All connected components are then bounded by a green line.

- *Text Extraction Process*

In this step, we read two sub images and extract the text from that image. Then we write two sub images into another jpeg file which contains only text region and some noise. Then we compose two sub images and display the final image in gray scale format and the original image in RGB format.

The pseudocode of our text extraction process is given below:

```
[L Ne]:= bwlabel  (Imagen);
Propied:= regionprops (L,'BoundingBox;
Hold the screen sometime;

Repeat until n= size (propied,1)
rectangle('Position',propied(n).BoundingBox,'EdgeC
olor','g','LineWidth',1)
End loop

hold off
Pause 1 second;
Repeat until n= Ne
      [r,c]:= find(L=n);
      n1:=Imagen(min(r):max(r),min(c):max(c));
      show the image (n1);
      Pause 0.1 second;
      Write the image in a file;
 End loop
```

IV. Result Analysis

To assess the performance of our method for extracting texts from images, we used total 25 images .The images have been taken from another published paper from web [3]. Gray scale and color cover images of books/journals/magazines, business card, color map, billboard are used in our experiment to display the efficiency of our method. We have focused on this type of image because they have variations in font size, style and background color. In our experiment the total text extraction process for an image is given below:

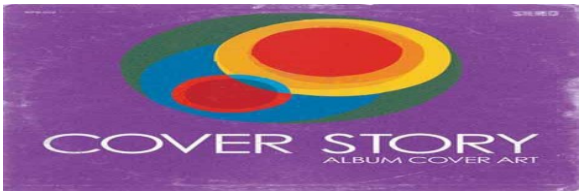

Fig 3: Original image

In pre processing steps the following two sub images are found in our method.

Fig 4: Sub image 1

Fig 5: Sub image 2

Fig 6: Image after extraction process from sub-images

Fig 7: Final output image

A. *Sample Images of Different Inputs And Outputs*

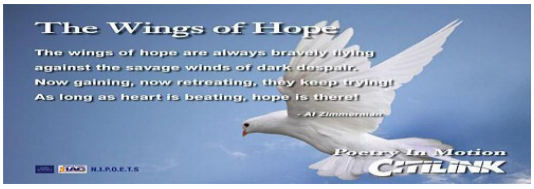

Fig 8: Different Inputs of our method

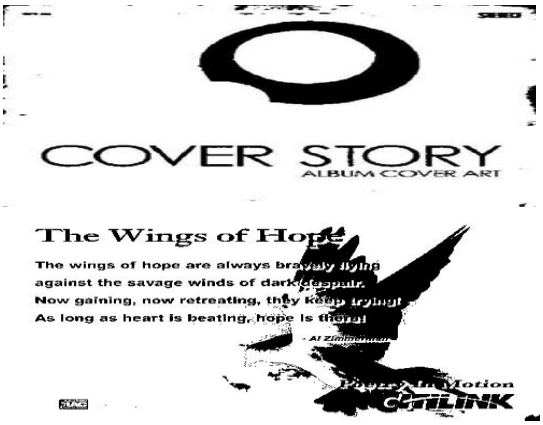

Fig 9: Outputs of our method for different images

Above figures show that using our method the resultant images contain more character than [14] and [3]. For quantification of accuracy, recall rate of characters are calculated to evaluate performance. Recall rate of characters (RRC) is defined as follows:

RRC = (No. of extracted characters/ No. of characters in image) × 100.

Table I shows the result of extraction performance using our method and modified morphological method and Gu's method, which confirms the superiority of our technique.

TABLE I. EXTRACTION PERFORMANCE FOR CHARACTERS

	Our Method	**Modified Morphological Method[3]**	**Gu's Method [14]**
Total Character	420	420	420
Extracted Characters	408	396	375
RRC (Accuracy)	98.8%	94.29%	89.28%

The following graph compares the performance among modified morphological method [3], Gu's method [14] and our proposed method.

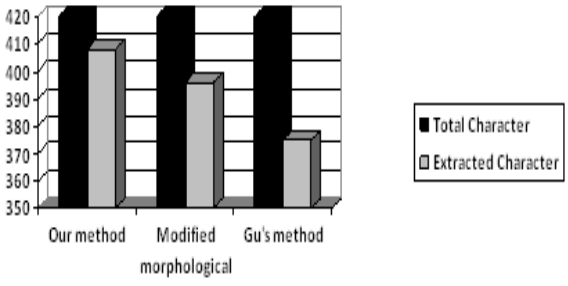

Fig 10: Performance comparison among our proposed method; modified morphological method and Gu's method for extracted character.

The superior performance of the proposed method is evident in the above graph. We notice that the improved performance is consistent across different input images. In terms of extraction time our total process takes about 1.5 minutes for Fig 3 on a Pentium® Dual-Core CPU E5300 @ 2.60 GHz, 2.60 GHz processor, 988 MB of RAM running on Microsoft Windows XP Service Pack 2.

V. CONCLUSION

Now a days, there is increasing demand of text information extraction from image. So, many extracting techniques for retrieving relevant information have been developed. Moreover, extracting text from the color image takes time that leads to user dissatisfaction. In this paper we have proposed a method to extract the text from image which extracts text more accurately. The proposed method is tested with various types of images, both images with caption text and scene text. Using our method it is possible to extract information within short time. Although, our connected component based approach for text extraction from color image method has several features than existing method but it becomes less effective when the text is too small. In this case, the text region is not clearly visible or the color of the text is not visible clearly.

ACKNOWLEDGEMENT

First of all we are grateful to Almighty God to give us the opportunity and capability to complete this research. Then we remember our parents whose are always thinking our successive. While accomplishing this work we faced so many problems and to solve the problem we got help from various sources. We would like to express our sincere regard all the teachers of CSE Discipline, Khulna University for their co-operation, suggestion, guidance, supervision and continuous encouragement.

REFERENCES

[1] R. Chandrasekaran and RM. Chandrasekaran , "Morphology based Text Extraction in Images", in proceeding of IJCST Vol 2, Issue 4, ISSN:0976-8491(Online)| ISSN:2229-4333(Print) Oct-Dec. 2011.

[2] Neha Gupta and V.K. Banga," Image Segmentation for Text Extraction", in proceeding of 2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012), Singapore April 28-29, 2012.

[3] Md. Shorif Uddin, Tanzila Rahman, Umme Sayma Busra and Madeena Sultana, "Automated Extraction of Text from Images using Morphology Based Approach ", in proceeding of international journal of electronics & informatics, Vol.1, No.1, 14th August, 2012 .

[4] Lukas Neumann, Jiri Matas, "A method for text localization and recognition in real-world images", ACCV, Nov. 8-12,2010.

[5] C.P. Sumathi, T. Santhanam, N. Priya, "Techniques and Challenges Of Automatic Text Extraction In Complex

Images: A Survey", in Proceeding Of Journal Of Theoretical And Applied Information Technology, 31st January 2012. Vol. 35 No.2.

[6] Jung, K., Kim, K.I., Jain, A.K.: "Text information extraction in images and video: a survey", Patt. Recognit. 37(5), 977–997,2004.

[7] Chitrakala Gopalan, D. Manjula, "Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images", International Journal of Computer Science and Engineering , 2-4, 2008.

[8] J.Fabrizio, M. Cord, B. Marcotegui, "Text extraction from street level Images", CMRT09. IAPRS, Vol. XXXVIII, Part 3/ W4 3-4 September, 2009

[9] Lin-Lin Shen Zhen JI, "Gabor Wavelet Selection and SVM Classification for Object Recognition", Acta Automatica Sinica, Volume 35, Issue 4, Pages 350-355, April, 2009.

[10] S.Abirami, Dr. D.Manjula, "A Survey of Script Identification techniques for Multi- Script Document Images", International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.

[11] J.Sushma, M.Padmaja, "Text detection in colour Images", lAMA, IEEE, 2009.

[12] G.Rama Mohan Babu, P. Srimaiyee, A. Srikrishna, "Text extraction from Hetrogenius Images using Mathematical Morphology", Journal of Theoretical and Applied Information Technology, 2010.

[13] Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen , "Robust Car License Plate Localization using a Novel Texture Descriptor", Advanced Video and Signal Based Surveillance,2009.

[14] Lixu Gu, Naoki Tanaka , R.M.Haralick and Toyohisa kaneko, "The Extraction of Characters from Scene Image Using Mathematical morphology", IAPR Workshop on Machine Vision Applications, vol. 2, pp. 12-14, 1996.