Efficient Text Segmentation for Born-Digital Compound Images

SonwaneVikas V.
M. E. Computer (Student)
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik

Prof. Shahane N. M.
Associate Professor
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik

ABSTRACT
Images are important information carriers which are often used in email messages and web pages to attach textual information. In Born digital compound image (BDCI) text and graphics/pictures come together on digital devices having certain distinct characteristics like low resolution (easy for online transmission and to display on screen) and text is created digitally on image. Text from BDCI can be effectively adopted for large numbers of applications like to retrieve contents of web, to improve indexing, to enhance content accessibility and content filtering. There are several problems to distinguish texts from BDCI because, text appears in various styles (i.e. Orientation, size, and colour), some neighbour texts are connected, and some text characters are superimposed on pictorial region which may lead to misclassification. Although researchers have proposed many methods in which character-level and block-based objects are commonly assumed to separate text from compound images. But these methods failed to extract reliable features to detect all texts as well as to identify connected components. To address these issues, novel efficient algorithm Local Image Activity Measure (LIAM) and Scale and Orientation Invariant Grouping (SOIG) are proposed to assemble separated characters into Textual Connected Component (TCC). These algorithms arebased on distribution of pixel variations and mean intrastring distance to precisely segment textual regions from BDCI.

Keywords
Born-digital compound image, Text segmentation, Mean intrastring distance.

1. INTRODUCTION
Images are being used frequently in Web pages and email messages to attach textual information. These imagesare used as text in a different way, for example in order to beautify (e.g. titles, headings etc.), to attract attention (e.g. advertisements), to hide information (e.g. images in spam emails used to avoid text-based filtering); even to tell a human apart from a computer (CAPTCHA tests). Problem has been quantified in the past in terms of the use of image-text in Web pages [1]. Researchers has observed from past study for a particular webpage (which is mixture of text as well as an image), that amount of text presented in image form in a web page (17%), while an important fraction of this text (76%) is

not to be found anywhere else in the Web page [1]. Taking into account that the very text that is presented in image form is more often than not semantically important (i.e. titles, headings, advertisements), one can get a feeling of the importance of the problem.

Extracting of text from born-digital images would promote technology to work on number of applications such as retrieval of web content and improved indexing, content filtering (e.g. advertisements or spam emails), enhanced content accessibility, etc. Although researchers have done a lot of work on text extraction from complex images like video frames, book and magazine covers, real-scenes and little work has been published specifically focused on born-digital compound images. While born-digital compound images mayseems similar to other complex images [2] which are mixture of image and text part, butboth kind of images have their certain unique characteristics. Those distinct characteristics forced researchers to develop different methods for different kinds of images. Born digital images have low-resolution (made to be easily transmitted online and displayed on a screen), they often suffer from compression artifacts and severe anti-aliasing while text is digitally created (over imposed) on the image. For the sake of comparison, real scene images are high resolution images which is captured by camera that often present illumination problems and perspective transformations. Therefore, it is not necessarily true that methods developed for one domain would work in the other.

For better understanding of the concepts, it is divided into sections: Section 1 describes the introduction of the system and motivation of the recently proposed system. Section 2 describes the related work in which motivational survey about various methods of text detection, efficiency and drawbacks of previous system are discussed. Section 3 describes the detailed design of the recently proposed system. Section 4 presents the experimental results andtext segmentation on ICDAR Database Section 5 describes the conclusion.

2. LITERATURE SURVEY
Many researchers have proposed text segmentation algorithms to separate textual components from others in compound images [13]. They proposed a JPEG-compliant method that allows for the efficient variable quantization of compound documents. This method automatically detects the image-part and the text-part of a document by measuring DCT activity in each of the 8 * 8 blocks. Based on the DCT activity of a block or a macroblock (16 * 16 block), quantization scaling factors are derived that automatically adjust the quantization so that text blocks are compressed at higher quality than image blocks. C. Yao and X. Bai [1] used features at component level. These features are character descriptors which has ability to distinguished different characters. Here researchers have assumed character level based or block level based [3] objects for the process of detecting texts. But these features are not able to extract the prominent features in order to detect all the text from given input image. Given assumption may not performed well and it may lose the integrity of text

characters particularly when there are large number of connected characters. The reason behind this, characterlevel based or block level based objects are not enough for varied sizes of characters and features of characters.

C. Yi and Y. Tian [5] proposed grouping and structure-based partition for locating text from a complex background with multiple colors. Researchers have also introduced textual objects at string level. For these objects they have proposed novel method which is based on partition of image and grouping of connected components. In first step they have used color as well as gradient features for selecting the potential text characters from connected components [9]. To combine the potential text into string (text string) grouping of character is employed. But the criterion is, there should be alignment of at least three characters. The final outcome shows partition (which is based on color) is always superior as compared to gradient based partition. Limitation is, it takes more time for detecting text from each color layer.

E. Haneda and C. Bouman [4] proposed MRC (Mixed Raster Content) gives particular framework for compression of document at a higher rate as compared to previous existing image compression algorithms. The key to this compression method is to separate the background as well as foreground layers. These layers represented with the help of binary mask which is computed by thresholding. The outcome of this MRC method is strongly dependent on segmentation algorithm for computing the binary mask. This algorithm performs better and achieves higher accuracy for detecting text but it comes with false rate i.e. it detects the image components which are non-textual or may be pictorial component. These methods result in over-connection (text characters are highly superimposed with background can lead to text misclassification) and over segmentation (in which characters are belonging to same string but still separated to each other i.e. it forms two separate components of same string), or under segmentation (characters are actually from different strings but it forms or connect to a single string.) problems.

W. Ding, Y. Lu, and F. Wu [9] proposed a fast compression algorithm in which BDCI image is decomposed to four different types of 16*16 macroblocks by combining the gradient and color information. They have proposed algorithm for fast block classification (BFC) and four different algorithms for image compression are designed, so that each algorithm can be uniquely applied to each distinct category. To code the pure picture as well as text images this BFC algorithm performs better than any other existing algorithms. Where pure picture image has JPEG format [11] and pure text image has LZW format [12].

Z. Pan and H. Shen [6] studied various visual attributes of textual blocks which includes gradient distribution and relationship of adjacent pixels. Here 16*16 blocks are divided into pictorial as well as textual blocks. The properties which are derived from block based methods are not enough to classify all blocks efficiently. There are some methods introduced by researchers for detecting text from document images [7]. The technique which is used for detecting text from document image cannot be directly used for detecting text from BDCI images. Because each category have their distinct characteristics. A) Resolution: Resolution of document image is about 300dpiwhich is much higher than born-digital compound image which has resolution about 100dpi. B) The source of document images are generally scanned papers or documents in order to capture paper-based information where born-digital compound image is created digitally with the help of computers. C) Precision requirement for processing: As document images contain large amount of text, somisclassificationof text on document images can be tolerated. But, text misclassificationof text is easily noticeable in born-digital compound image.

3. IMPLEMENTATION DETAILS

3.1 Advantages

The recently proposed methods Local Image Activity Measure (LIAM) and Scale and Orientation Invariant Grouping(SOIG) have several advantages over existingone:

1)These methods are better enough to detect texts with arbitrary orientations as well as scales with complex background in born-digital compound images.

2)These methodscan well preserve the text characters integrality avoiding over and under segmentation of textual regions.

3)For characters which are superimposed on pictorial regions, recently proposed method avoids such over-connection problem to avoid the problem of text misclassification.

3.2 Text Segmentation

The detailed flow of Text Segmentation is described in Fig.1. First step is to collect different dataset which will be used during testing of system. Dataset contains various Born-Digital compound images with arbitrary scales and orientation.

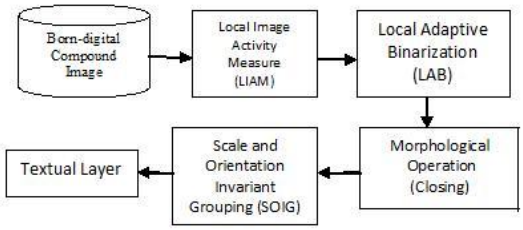

Fig.1. Block diagram of text segmentation

3.2.1 Local Image Activity Measure

Consider the born-digital compound image is given as input; LIAM[13] tries to remove the pictorial components in initial stage which would help to generate effective TCC (Textual Connected Component) in later stage. The most important difference between textual and pictorial is the distribution of pixel variations in BDCI. Where the pictorial components are relatively very smooth as compared to textual ones and the textual components generally possess higher intensive variations, therefore it becomes easy to distinguish. One can be easily separate out those two regions with the help of distribution of pixel variation. Computation of local pixel variations in the image is done by recently proposed algorithm, LIAM. It measures the local pixel variations in the image particularly for each pixel, so it becomes easy to highlight difference between pictorial and textual regions with respect to activity levels. After calculating activity values for each pixel, it has been observed that textual component would always possess higher activity value than pictorial components. The output of LIAM is an image has some limitation. Because, output part of an image contains some pictorial regions as well. The reason behind some pictorial regions in output is that, there could be some pictorial components which might have activity values same or very

nearer to the activity values of textual components. The activity value of a pixel $\{P_{i,j}\}$ is calculated as follows:

$\{P_{i,j}\}$ Where i=1,……., M ; j=1,……,N;

$L(P_{i,j})=\alpha V_1(P_{i,j})+(1-\alpha)\, V_2(P_{i,j})$ ……(1)

Where,

$V_1(P_{i,j})=(P_{i,j}-P_{i-1,j-1})^2+(P_{i,j}-P_{i+1,j+1})^2+$
$(P_{i,j}-P_{i-1,j+1})^2+(P_{i,j}-P_{i+1,j-1})^2$ …….....(2)

$V_2(P_{i,j})=(P_{i-1,j-1}-P_{i+1,j+1})^2+(P_{i-1,j+1}-P_{i+1,j-1})^2$ ……..(3)

Where V1 and V2 are the1- and 2-distance variations in diagonal direction.

3.2.2 Local Adaptive Binarization

The document images generally have uniform contrast distribution. Due to uniform nature of document images single threshold value is sufficient to convert the image into binary form. This is called Global Thresholding. But in BDCI there is more intensive variations with added background noise and illumination may exist. Therefore, Image would be no longer uniform;hence single threshold value would not besufficient. Hence in such cases global thresholding method would not be useful. LocalThresholdingmethod needs to apply. It involves calculation of local mean of the neighbouring pixels in image. Based on the value of local mean threshold value for each pixel is decided. Local Adaptive Binarization outputs Binary image which is denoted by Tbin.

3.2.3 Morphological Operation

Now consider the result of local adaptive binarization method i.e. binary image T (bin), where small sizes characters are much closer to nearby characters or partially connected, while others adjacent characters are still having specified distance. Morphological closing operation [8] is used with minimum scale structuring element (se) of size 3*3 for partially connected characters. Instead of applying SOIG operation for nearby characters which would unnecessary increase the cost of computation, simple closing operation is used for constructing the potential TCCs (Textual Connected Components).Closing operation [8] helps to connect partially connected characters to fully connected characters. Morphological closing is calculated as follows:

$T_{close} = (T_{bin} \oplus se)\ominus$ ……..(4)

3.2.4 Scale and Orientation Invariant Grouping(SOIG)

Morphological operations have formed TCC's of partially connected or nearby characters. To construct the remaining TCC for those characters which are not included in closing operation and having specified distance among adjacent characters new method need to use. To construct TCC for such distributed characters in the binary image previous methods were using fixed connection scale. It means once adjacent characters satisfies the fixed connection scale [10] they would be considered in a particular set of TCC. But distance between each two characters or strings may get changed according to character size variation, it means fixed connection scale would fail to generate TCC. In order to address all such issues, recently proposed method SOIG [13]has enough capability to construct efficient TCC's as compared to existing method. In this process each character is considered as a component and centroid of each component is computed. Then it considers mean intrastring distance as connecting scale to connect adjacent characters. Mean intrastring distance is the closest Euclidian distance between the centroids of two adjacent characters. Once the adjacent characters satisfies the mean intrastring distance then they will be added in TCC set, then it needs to go on checking further adjacent characters whether it satisfies the condition. If it doesn't satisfies the condition that component will be considered as an isolate component (ex. 'a' character).Finally, TCC set would contain all the TCC for each string. For example if "Digital Image Processing" is a text within image. For this text final 3 TCC's would generate TCC1 for Digital, TCC2 for Image and TCC3 for Processing.

3.3 Mathematical Model

Recently proposed system takes an image from a set of N images as input, where N is number of images in database. Set of operations are then performed on those images. The output of previous function will act as an input to the next function.

The recently proposed system S is defined as follows

S = {I, F, O}

I denote a set of N BDCI and it is defined by,

I = {I1, I2...,IN }

Ii denotes an input image.

O denotes output of the system.

F denotes set of function. It is denoted by,

F = {F1, F2, F3, F4}

F1 = It is a function of computing local image activity measure of each pixel. This function operates on input image Ii and gives output as normalized activity map (Nmap).

F1 (Ii)=Nmap: …….(5)

F2= It is a function Local Adaptive Binarization which operates on Nmap and computes the binary image Tbin.

F2(Nmap)=Tbin…….(6)

F3=It is a function of morphological closing operations. This function operates on highly connected characters. It results fully connected characters from partially connected character.P denotes partially connected characters.Tclose denotes logical image of fully connected characters.

F3(P)=Tclose…….(7)

F4= It is a function of a SOIG algorithm which is used to generate final TCCs. It is operated on remaining components of Tclose, which is denoted by T'close. Iclose is an output of SOIG

F4(T'close)=>Iclose……(8)

4. RESULTS AND DISCUSSION

4.1 Performance Measures

The harmonic mean of p and r is denoted as standard F-measure *f* .It is defined as the following:

$f = 2 \times p \times r/(p + r)$ ..……..(9)

Where*p*and *r* s Precision and Recall. It is defined as the following:

p= TP/ (TP+FP) , r= TP/(TP+FN) ……(10)

TP is the number of correctly detected textual pixels. FP is the number of pixels which are pictorial pixels, but detected as textual ones.FN is the number of pixels which are textual pixels, but misclassified as pictorial ones.

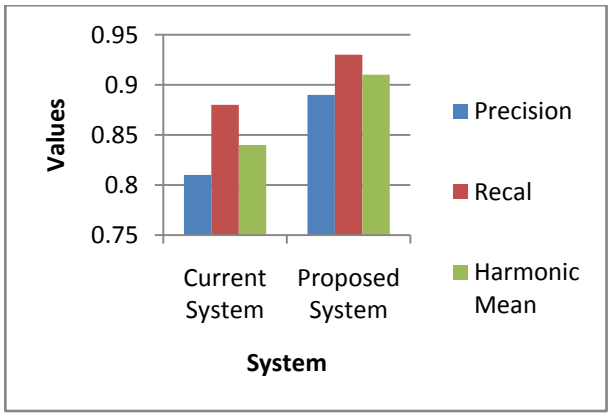

Fig.2.Comparative results between recently proposed methods and Existing Method

Table 1.Comparision Table

Performance Parameter	Precision	Recall	Harmonic Mean
Current System	0.81	0.88	0.84
Recently proposed System	0.89	0.93	0.91

4.2 Text Segmentation on ICDAR Database

For testing of text detection algorithm ICDARdatabase is commonly used. Most of the characters from this database are horizontally distributed and some of characters are superimposed on pictorial background. The testing of recentlyproposed method is doneon some images of the ICDAR database and compared with TCC-basedsegmentationframework. From table 1, it is clear that recently proposed method gives best harmonic mean by increasing Precision and Recall values. It indicates higher the value fewer texts are misclassified as a pictorial component. Fig.3 shows input image from ICDAR database. On input image LIAM and then binarization is applied. Its output is shown in Fig.4. It can be observed from Fig.4 there are some pictorial regions due to their higher activity values. Then closing and SOIG algorithm is used which outputs the TCC (Textual Connected Components).

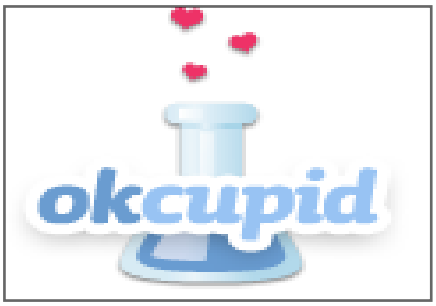

Fig.3. Input image(BDCI)

Fig.4.Image after LIAM and Binarization method

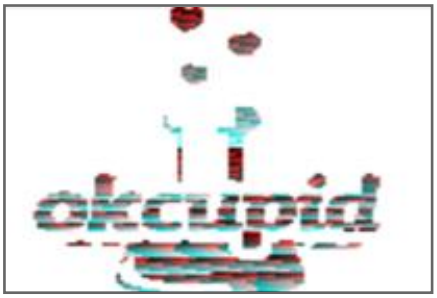

Fig.5.Textual Connected Components

Fig.6.Output Image

All image part including textual as well as pictorial which satisfies the condition of constructing connected components is shown in Fig.5. Fig.6. demonstrates the textual part of an image enclosed by rectangular box.

4. CONCLUSION

There are different techniques of text detection and extraction from Born-digital compound images. In most of these methods, block-based or character-level objects are commonly assumed to detect texts. Text detection and extraction in BDCI is very challenging task due to complexity of background. LIAM method considers the distribution of

pixel variations to separate text part from image part which considers four activity measures in horizontal, vertical and diagonal direction. One and two distance variations in diagonal direction are further considered to calculate activity value for each pixel. LAB method involves computation oflocal mean of the neighbouring pixels in image which achieves better results as compared to other existing thresholding methods.For local regions some characters of small sizes are much close to nearby characters or partially connected. In such regions morphological closing operation results fully connectedcharacters from more adjacent partially connected characters. SOIG algorithm assembles separated characters into TCC.It can be seen from Table 1 average precision is 0.81. It is not affecting very much thecharacter recognition using Optical Character Recognition(OCR) tool because incorrectly detected pixels which lowersthe precision which is not going to change characteristic shape.

The recently proposed system is focused on detecting text from born digital compoundimages. The system gives satisfactory results for simple images to complex images. But the typical kind of images containing scanned signature is also fails along withexisting system to detect the text. Also it does not detect the overlapping or too adjacenttext. Also if the text is in vertically tilted fashion with large vertical angle the systemfaces difficulties in detecting them. Hence, more investigations and experimentations using different parameterssetting is required.

6. REFERENCES

[1] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI, USA, 2012.

[2] P. Shivakumara, T. Phan, and C. Tan, "A Laplacian approach to multioriented text detection in video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412-419, Feb. 2011.

[3] S. Juliet and D. Florinabel, "Efficient block prediction-based coding of computer screen images with precise block classification," IET Image Process., vol. 5, no. 4, pp. 306-314, Jun. 2011.

[4] E. Haneda and C. Bouman, "Text segmentation for MRC document compression," IEEE Trans. Image Process., vol. 20, no. 6, pp. 1611-1626, Jun. 2011.

[5] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594-2605, Sep. 2011

[6] Z. Pan, H. Shen, and Y. Lu, "Brower-friendly hybrid codec for compound image compression," in Proc. IEEE Symp. Circuits Syst., Rio de Janeiro, Brazil, 2011.

[7] D. Karatzas, S. Mestre, J. Mas, F. Nourbakhsh, and P. Roy, "ICDAR 2011 robust reading competition challenge 1: Reading text in borndigital images (web and email)," in Proc. Conf. Document Anal. Recognit., Beijing, China, 2011.

[8] N. Francisco, N. Rodrigues, and E. Silva, "Scanned compound document encoding using multiscale recurrent patterns," IEEE Trans. Image Process., vol. 19, no. 10, pp. 2712-2724, Oct. 2010.

[9] W. Ding, Y. Lu, and F. Wu, "Enable efficient compound image compression in H.264/AVC intra coding," in Proc. IEEE Conf. Image Process., San Antonio, TX, USA, 2007.

[10] J. Song, Z. Li, M. Lyu, and S. Cai, "Recognition of merged characters based on forepart prediction, necessity-sufficiency matching, and character-adaptive masking," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 35, no. 1, pp. 2-11, Feb. 2005.

[11] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," IEEE Trans. Image Process., vol. 14, no. 8, pp. 993-1005, Aug. 2005.

[12] K. Konstantinides and D. Tretter, "A JPEG variable quantization method for compound documents," IEEE Trans. Image Process., vol. 9, no. 7, pp. 1282-1287, Jul. 2000.

[13] Huan, Yang and Shiqian Wu "Scale and Orientation Invariant Text Segmentation for Born-Digital Compound Images" IEEE Trans.CyberneticS, vol. 45, no. 3, March 2015