A Survey on Text Segmentation Methods for Digital Images

Vikas V Sonwane[1] N.M.Shahane[2]
[1]M.E. Student [2]Assistant Professor
[1,2]Department of Computer *Engineering*
[1,2]K. K. W. I. E. E. R., Nashik Savitribai Phule Pune University, Maharashtra, India

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 number of applications likes 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 color), some neighbor 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. Some recent methods based on distribution of pixel variations and image activity measures can able to precisely segment textual regions from BDCI.

Key words: Born-digital compound image, Text segmentation, Image activity measure

I. INTRODUCTION

Many needs result in use of images which are abundantly used in emails, webpage to attach or describe textual information. These are also used to beautify (e.g. titles, headings etc.) and to keep information hidden from end user (e.g. images embedded in spam mails to keep off text based filtering.) Past research has shown that web page has 17% amount of text which is embedded in image, while 76% of this 17% text; which is important part is absent anywhere in webpage excluding image. In this way semantic or meaningful text is embedded in image to define problem or abstract idea of text in webpage. (e.g. advertisements, headings)

Improved indexing, web content retrieval, enhanced content accessibility, filtering of contents; these applications are executed by an interesting task of extraction of text from BDCI. (e.g. spam email, advertisements etc). Many researchers have worked on extraction of text from complex compound image such as book cover and magazine cover also video frames and real scenes. Some work is dedicated to BDCI. Though BDCI may show similarity to complex text images or similar structures still BDCI possess few unique features. The low resolution BDCI is easy for online transmission and to get visuals on screen. Digital creation or imposition of text on images has setbacks like compression artifacts, anti-aliasing. Comparison shows that, illuminance and perspective transformations frequently show problems in higher resolution images which are captured by camera of real scenes. Therefore, method which is formulated for particular area may not work fine for other area.

Because of the unique characteristics of BDCI, it is hard to separate out text from such BDCI Images. A) As far as the properties of text character are concerns, it is generally appeared in different styles such as color, orientation and size. B) Most of the texts are embedded in the background in such a way that, it increases the possibility of text misclassification. C) The precision requirement for BDCI must be very accurate; as the amount of text in image form is not much high. Hence any artifact or changes over those texts are easily noticeable.

II. RELATED WORKS

Number of segmentation algorithms has been proposed so far by many researchers for separating the text component from other remaining components from compound images. Where compound image is a mixture of textual, pictorial and graphical components. JPEG-compliant method is proposed by Konstantinides and D. Tretter [13]. The basic motive behind this method is that, it allows effective variable quantization of compound images by computing the discrete cosine transforms activity (DCT). The DCT activity need to measure for each 8*8 block. JPEG method is able to effectively distinguished text part as well as an image part of a document. With the help of Discrete cosine transform activity for 16*16 block or macroblock, scaling factors are computed which has ability to compress text blocks at higher quality as compared to image block.

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 charactered level based or block level based 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 [9] particularly when there are large number of connected characters. The reason behind this, charactered level 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. 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 are 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 [10] 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 is 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 and pure text image has LZW format.

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] [12]. 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 300 dpi which is much higher than born-digital compound image which has resolution about 100dpi. B) The source of document images [8] 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, so misclassification of text on document images can be tolerated. But, text misclassification of text is easily noticeable in born-digital compound image.

S. Juliet and D. Florinabel [3] proposed algorithm for classification purpose. This algorithm is designed to classify the components from digital images like computer generated synthetic images, PPTs and magazine covers etc. The focused of this algorithm is to avoid the loss of integrality and visual quality of text components at the time of compression. With the help of discrete wavelength transform the algorithm then separate out the textual part from other remaining graphical or pictorial part of digital image. This method gives better results for classifying the text components provided that text may have various sizes or the different ways of arrangement.

Z. Li and M. Lyu [11] proposed a Novel merged-character recognition method which is based on forepart prediction. In order to recognize merged characters NSM-based recognition and character adaptive masking is proposed. There is a three type's classification between two involved characters: overlapped, linear and nonlinear. Linear type is handled well by existing segmentation methods; but, their capabilities of handling the other two types i.e. non-linear and overlapped are limited. NSM-based recognition method handles the linear, nonlinear and overlapped merging type's effecitiently.

Huan Yang and Shiqian Wu [15] consider the born-digital compound image as input; the initial stage tries to remove the pictorial components to help the TCC (Textual Connected Component) generation. The distribution of pixel variations is the most important difference between pictorial as well as textual components in BDCI (Born-digital Compound Images). Where the pictorial components are relatively very smooth as compared to textual and the textual components generally possess higher intensive variations, therefore one can be easily separate out with the help of distribution of pixel variation. Computation of local pixel variations in the image is done by recently proposed algorithm, LIAM (Local Image Activity Measure). It measures the local pixel variations in the image particularly for each pixel, so that one can highlight difference between pictorial and textual regions with respect to activity levels. There are four different kinds of activity measures are considered for each pixel in LIAM algorithm, which are one and two distance variation respectively. If the distance between two pixels are one then it is said to be one distance variation and if it is two it is called two distance variation. One distance variation in vertical, horizontal as well as in diagonal directions. And the same directions are further considered for two distance variation. First- and second-order texture information can be detected with the help of 1- and 2-distance variation, which are used to select adjacent pixels with more variations, so one can easily locate the edges in BDCI images.

Y. Chen and B. Wu [14] proposed method which applies a multi-plane segmentation technique to separate homogeneous objects including text blocks, non-text graphical objects, and background textures into individual object planes. It consists of two stages - automatic localized multilevel thresholding, and multi-plane region matching and assembling. Then a text extraction process can be performed on the resultant planes to detect and extract characters with different characteristics in the respective planes P. Shivakumara, T. Phan, and C. Tan [2] proposed a method for video text detection which is able to handle both graphics text and scene text. Our focus for scene text is text orientation because traditional methods only consider horizontal text. The Laplacian in the frequency domain helps to identify the candidate text regions; proposed method performs well for both horizontal and non-horizontal text.

III. CONCLUSION

This Paper provides review of different techniques of text detection and extraction from Born-Digital Compound Images. There are so many techniques which are used in text extraction but each method has its own significance to work with different properties of images like variation in font, orientation, alignment, color, style, size, texture, font size and provides satisfactory results. Very few methods are able to consider all the properties to make precise segmentation. Text detection and extraction in BDCI is very challenging task due to complexity of background. Many researchers can work on public database which is available online and give a better performance. On the basis of this review paper provide a light of different researchers which doing research on this area.

ACKNOWLEDGEMENT

I am glad to express my sentiments of gratitude to all who rendered their valuable help for the successful completion of the paper. I am thankful to my guide Prof. N. M. Shahane, Associate Professor, Computer Engineering, K.K.W.I.E.E.R., Nashik for his guidance and encouragement. His expert suggestions and scholarly feedback had greatly enhanced the effectiveness. I would also like to thanks to Prof. Dr. S. S. Sane, Head, Department of Computer Engineering, K.K.W.I.E.E.R., Nashik.

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] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Conf. Comput.Vis. Pattern Recognit., San Francisco, CA, USA, 2010.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] Y. Chen and B. Wu, "Text extraction from complex document images using the multi-plane segmentation technique," in Proc. IEEE Conf. Syst., Man, Cybern., Taipei, Taiwan, 2006.
[15] Huan Yang and Shiqian Wu, "Scale and Orientation Invariant Text Segmentation for Born-Digital Compound Images" IEEE Trans Cybernetics, vol. 45, no. 3, pp. 519-533 March 2015

