ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Text Localization and Recognition in Natural Scene Images

Mona Saudagar[1], S. V. Jain[2]

Student, Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India[1]

Assistant Professor, Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India [2]

Abstract: The extraction of text in an image is a classical problem in the computer vision. With the increasing popularity of practical vision system text recognition in natural scenes becomes a critical task. Text data present in images and video contain useful information for automatic annotation, image indexing. But as natural scene images contain complex background, multiple fonts and orientations, and different alignments make the problem of automatic text extraction extremely challenging. This paper proposes the approach for automatic detection of text from images and explains the methodology to extract and recognize multi-oriented text in natural scene images.

Keywords: Text Localization, Text extraction, Text recognition, Sobel Mask, Binarization, Bounding Box.

I. INTRODUCTION

Among all the contents in images, text information is very important, since it can be easily understood by humans and computers. Text in image contains useful information which helps to acquire the overall idea behind the image. Lot of text detection and recognition systems are considered for horizontal or near horizontal texts but detecting texts of random orientations from images have become a challenging task. Detecting texts of random orientations from images is a challenging problem due to the multiple fonts, different sizes, various orientations and alignment, reflections, shadows, the complexity of image background. Text detection and segmentation from natural scene images are useful in many applications. Text in images can be basically distinguished from is background on the basis of its high contrast or colour. But in natural scene images due to the low contrast or uniform colour it is difficult to segment text from non-text content. And due to this, normal document OCR does not give accurate recognition results. Extraction and recognition of text from various types of images are very effectual in text based application like: Video and image database retrieval, Image annotation, Data mining, Detection of vehicle license plate. Indexing images or videos requires information about textual content in it. In the same way, natural scene based text explores automatic detection of street name, location, traffic warning and name of commercial goods. Texts in an image directly carry high-level semantic information about a scene, which can be used to assist a wide variety of applications, such as image search and indexing, navigation, and human computer interaction.

II. RELATED WORK

The methods which are available for text detection in natural scene images are mainly based on texture or region or on both texture and region which is named as hybrid methods.

Texture based methods [3, 6, 7, 8, 9] usually treat the pattern of text appearance as a special texture. Techniques used in these methods include Fourier transform, wavelet decomposition, and combination of wavelet and moments with the help of a classifier to classify text and non-text candidates. The main problems of texture based methods lie in the training of large number of samples and features that heavily depend on the classifier in use. Many of the methods works well for horizontal and near-horizontal straight lines but not curved lines.

The next category is region-based methods [2, 4, 8, 9]. It requires heuristics and parameters setting. This method is affected by the size of the text and does not provide good results when background and textual contents are closer.

The third category, namely hybrid methods, [5] proposes both texture and region based methods for text detection. Many of the Hybrid methods are suited to horizontal and some near-horizontal straight text lines images. Such methods work well when all text line components have the same orientations. Any arbitrary text lines such as arc, circular, S, or Z shaped text lines characteristics is no truer as character components on a curved line. Due to such multiple orientation styles text detection has become more challenging problem and demands a new robust text detection technique.

The major categories of text detection method are (a) connected component-based [15, 16] (b), texture-based methods [17, 18], and edge and gradient based methods [19, 20]. Connected component based methods expect that the character must be of proper shape; therefore the method may not be suitable for scene text detection in images with complex background or low contrast images. For complex background or low contrast images texture

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

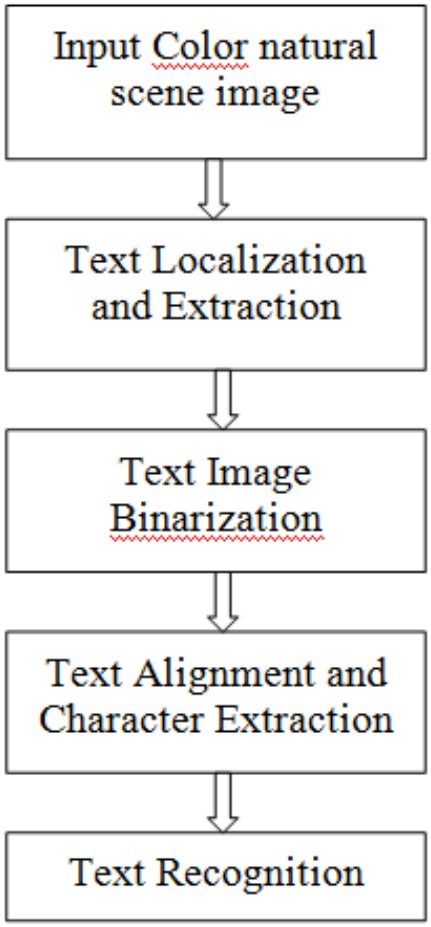
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

based methods provide better results as compared to connected component based methods. But it is very difficult to define texture property for scene text detection because sometimes background has the texture similar to text. For scene text detection the edge and gradient feature based method are good in terms of efficiency and some extent to complex background. But these methods suffer from setting threshold values at several stages of the algorithms.

Multi-oriented text has only been partially addressed in [21, 22] where the algorithm is limited to caption text and a few selected directions. Recently, Shivakumara et al. [6] have addressed this multi-oriented issue which is based on Laplacian and skeletonization methods. But this method cannot detect text of arbitrary orientation. That method does not work well for an image with curve text. Muti-oriented text recognition described in [14] has shown good results but the method does not consider complete natural scene image. Described method considers only text portion as input and recognition is performed. This work does not involve text localization and extraction

Lot of work is done in field of text recognition but available methods are suitable to horizontally aligned text and do not work well to recognize text of arbitrary orientation from complete natural scene images.

III. PROPOSED APPROACH



Text recognition from natural scene images involves three major phases; Text Localization i.e., where text is actually located in an image, Text Extraction i.e., separation of text from background and third is Text Recognition i.e., actual textual content. Therefore to recognize multi-oriented text from natural scene image a different approach is proposed and these three major phases are divided into five steps as shown in following flow diagram. It explains the methodology to extract and recognize the text from natural scene images.

IV. IMPLEMENTATION DETAILS

A. *Dataset*

To carry out the work we are taking images from two different datasets namely: MSRA-TD500 and NUS, where text appears in different orientations like horizontal, non-horizontal, curved etc. MSRA-TD500 dataset is popular data for scene text detection and recognition as it contains varieties of images that contains both horizontal and non horizontal text images with variety of background complexity. However, the orientation of the text is limited to horizontal and near horizontal straight and there was no curved text. To test effectiveness of the proposed method in terms of orientation and contrast variation, we use NUS data that contains curved text and their resolution was low because of video data.

B. *Text Localization and Extraction*

Following figures 1-8 shows the phenomenon of text localization and extraction. We start finding text edges to get actual textual area. Figure 2 is the obtained by applying sobel mask on input image 1, because among all gradient operators sobel operator gives good edge detection results. Area other than textual content form borders. Then result of figure 2 is dilated as shown in figure 3 and holes are filled and borders are cleared as shown in figure 4 to get text area as close as possible and to isolate the non text area. This is how we localizes the text as shown in figure 5 and extracted the text from complete natural scene image as shown in figure 6.

C. *Text Image Binarization*

If we improve binarization for the segmented text lines in video and natural scene images, recognition rate can be improved. There various methods for binarization; Niblack, Sauvola, OTSU, SWT, Wavelet-Gradient fusion [13]. Here we have used OTSU method to binarize the extracted text because it uses global thresholding and gives good binarization results as compared to other binarization methods. We get the binarized image as shown in figure 7 of figure 6.

Figure 1: Input Image

DOI 10.17148/IJARCCE.2015.4413

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Figure 2: Binary gradient mask

Figure 3: Dilated gradient mask

Figure 4: Binary image with filled whole and cleared borders

Figure 5: Text Localization

Figure 6: Text Extraction

Figure 7: Binarized text image

Figure 8: Text image with Bounding Boxex

D. *Text Alignment and Character Extraction*

In this step, from the binary text line image the foreground pixels of text lines are chosen. Here, each character is one component. Bounding box is drawn to each component as shown in figure 8. As text lines can be curved as shown in figure 9 and non-horizontal, centroid of bounding boxes can be taken and joined. Then this chained path can be then aligned straight and each bounding box will get extracted and character will be recognized. Results for figure 9 are shown in figures 10-16.

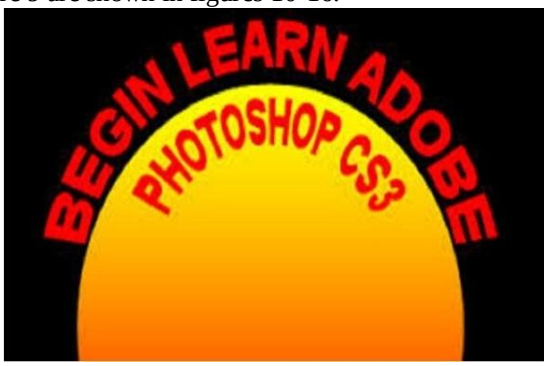


Figure 9: Input Image

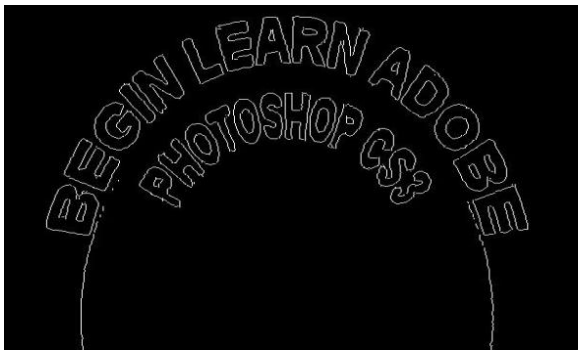


Figure 10: Binary gradient mask

DOI 10.17148/IJARCCE.2015.4413

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

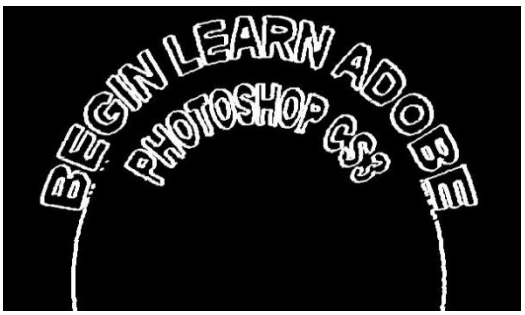


Figure 11: Dilated gradient mask

Figure 12: Binary image with filled whole and cleared borders

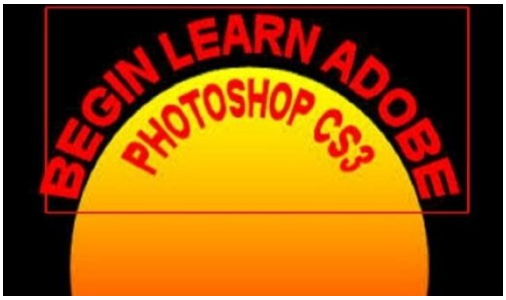


Figure 12: Text Localization

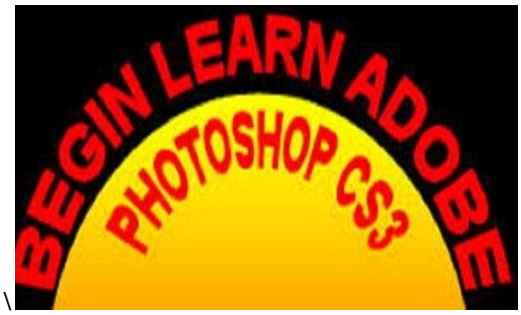


Figure 14: Text Extraction

Figure 15: Binarized text image

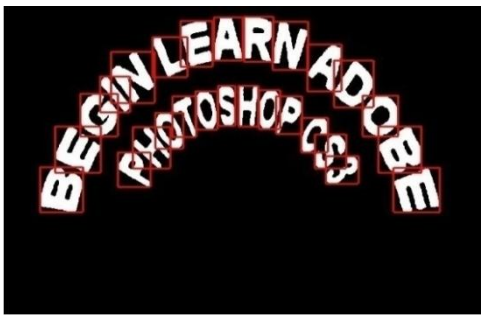


Figure 16: Text image with Bounding Boxex

V. CONCLUSION & FUTURE WORK

Detecting and recognizing text from natural scene images is still an unsolved problem because sometimes images are of low contrast or of complex background containing text of various orientation styles and text can be of different font types and sizes. Future work is to get the path of bounding boxes and correctly align the text of different orientation in natural scene images horizontally to improve recognition rate. In recognition phase, classification errors can be found and those errors can be caused due to ambiguous characters, such as {L, I}, {O, D}, {h, n}, {e, c} etc. Therefore, further improvements can be made to recognize these characters correctly, so that accuracy can be increased.

ACKNOWLEDGMENT

We authors are grateful to the Head, Computer Science and Engineering Department and the Principal, Shri Ramdeobaba College of Engineering and Management for providing adequate facilities to conduct a research. We hereby thank the authors listed in the references for the valuable information and survey statistics.

REFERENCES

[1]. K. Jung, K. I. Kim and A. K. Jain, Text information extraction in images and video: a survey, Pattern Recognition, 2004, 977 - 997.
[2]. T. Q. Phan, P. Shivakumara and C. L. Tan, Detecting text in the real world, In Proc. ACM Multimedia, 2012, 765-768.
[3]. X. Chen and A. Yuille, Detecting and reading text in natural scenes, In Proc. CVPR, 2004, 366 - 373.
[4]. L. Neumann and J. Matas, Real-time scene text localization and recognition, In Proc. CVPR, 2012, 3538 - 3545.
[5]. Y.-F. Pan, X. Hou and C.-L. Liu, A hybrid approach to detect and localize texts in natural scene images, IEEE Trans. IP, 2011, 800 - 813.
[6]. P. Shivakumara, T. Q. Phan and C. L. Tan, A Laplacian approach to multi-oriented text detection in video, IEEE Trans. PAMI, 2011, 412 - 419.
[7]. C. Yi and Y. Tian, Text string detection from natural scenes by structure-based partition and grouping, IEEE Trans IP, 2011, 2594 - 2605.
[8]. C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, Detecting texts of arbitrary orientations in natural images, In Proc. CVPR, 2012, 1083 - 1090.
[9]. P. Shivakumara, A. Dutta, C. L. Tan and U. Pal, Multi-Oriented Scene Text Detection in Video based on Wavelet and Angle Projection Boundary Growing, Multimedia Tools and Applications, 2013, 1 – 25.
[10]. N. Sharma, P. Shivakumara, U. Pal, M. Blumenstein and C. L. Tan, A new method for arbitrarily-oriented text detection in video, In Proc. DAS, 74 - 78.
[11]. P. Shivakumara, T. Q. Phan, L. Shijian and C. L. Tan, Gradient Vector Flow and Grouping Based for Arbitrarily-Oriented Scene Text Detection in Video Images, IEEE Trans. CSVT, 2013, pp 1729-1739.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

[12]. H. I. Koo, "Scene text detection via connected component clustering and nontext filtering," IEEE TIP, 2013, 2296 – 2305.
[13]. S. Roy, P. Shivakumara, P. P. Roy and C. L. Tan, "Wavelet-Gradient-Fusion for Video Text Binarization" In Proc. ICPR, 2012, pp. 3300-3303.
[14]. S. Roy, P. P. Roy, P. Shivakumara, G. Louloudis, C. L. Tan, HMM-based Multi Oriented Text Recognition in Natural Scene Image©2013 IEEE.
[15]. A. K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition, 1998, pp. 2055-2076.
[16]. K. Jung and J. H. Han, "Hybrid Approach to Efficient Text Extraction in Complex Color Images", Pattern Recognition Letters, 2004 pp. 679–699.
[17]. K. L Kim, K. Jung and J. H. Kim. "Texture-Based Approach for Text Detection in Images using Support Vector Machines and Continuous Adaptive Mean Shift Algorithm" IEEE Trans. on PAMI, 2003, pp. 1631-1639.
[18]. M. Anthimopoulos, B. Gatos and I. Pratikakis, "A Two-Stage Scheme for Text Detection in Video Images", Image and Vision Computing, 2010, pp. 1413-1426.
[19]. D. Chen, J.M. Odobez and J.P. Thiran, "A Localization/Verification Scheme for Finding Text in Images and Video Frames based on Contrast Independent Features and Machine Learning", Signal Processing: Image Communication, 2004, pp. 205-217.
[20]. A. Jamil, I. Siddiqi, F. Arif and A. Raza, "Edge-based Features for Localization of Artificial Urdu Text in Video Images", ICDAR, 2011, pp. 1120-1124.
[21]. H. Tran, A. Lux, H. L. T. Nguyen and A. Boucher, "A Novel Approach for Text Detection in Images using Structural Features", ICAPR, 2005, pp. 627-635.
[22]. J. Zhou, L. Xu, B. Xiao and R. Dai, "A Robust System for Text Extraction in Video", ICMV, 2007, pp. 119-124.

BIOGRAPHIES

Mona Saudagar has received her B.E, degree in computer science and engineering in 2012. She is a pursuing Masters in Technology in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management, Nagpur-440013. Her areas of interest include Image Processing, Pattern Recognition and Artificial Intelligence.

Professor Shweta Jain received the Masters in Technology in Computer Science and Engineering from Nagpur University in 2009 as a first merit holder. She is currently Assistant professor in computer science and engineering department at Shri Ram deobaba College of Engineering and Management Nagpur. She has a total teaching experience of around 13 Years. Her research interests include Pattern Recognition, Digital Image Processing and Machine Learning.