Text Localization and Recognition in Natural Scene Images

Mona Saudagar¹, S. V. Jain²

Student, Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India¹

Assistant Professor, Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India²

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 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
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. Multi-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.

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.

### III. PROPOSED APPROACH

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input Color natural scene image</td>
</tr>
<tr>
<td>2</td>
<td>Text Localization and Extraction</td>
</tr>
<tr>
<td>3</td>
<td>Text Image Binarization</td>
</tr>
<tr>
<td>4</td>
<td>Text Alignment and Character Extraction</td>
</tr>
<tr>
<td>5</td>
<td>Text Recognition</td>
</tr>
</tbody>
</table>

Figure 1: Input Image
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.
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


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.