Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 3, Issue 2; January-March, 2016, pp. 94-97

http://www.krishisanskriti.org/Publication.html

A Review on Different Scene Text Detection Techniques

Mule S. S.[1] and Holambe S.N.[2]

[1]PG Student Department of ME (CSE) TPCT'S COE, Osmanabad, India
[2]Department of ME (CSE) TPCT'S COE, Osmanabad, India
E-mail: [1]ssmule137@gmail.com, [2]snholambe@yahoo.com

Abstract—*In recent years, with increasing popularity of portable devices for capturing images, visual processing, text extraction etc become key problems which gain the attention of many researchers. In this paper we have reviewed different techniques to find strings of characters from scene images. Text in an image may contain important information. We have reviewed different techniques like extraction of character string regions from scene images based on contours and thickness of characters, binarization and enhancement technique by a connected component analysis procedure, text detection from scene images by structure-based partition and grouping for text detection in images. Text extraction from image involves detecting the text from given image, finding the presence of text location, extraction, enhancement and recognition of text from the given image. Text extraction process becomes more challenging because of variations involved in text such as font style, size, orientation, alignment, reflections and illumination effect. Text reading in images is important step to achieve content retrieval from images. The content retrieve from image content useful information it acts as clue for many image based applications such as image understanding. This paper presents a short survey on various scene text detection methods suggested and implemented recently. Here we are also discussed general challenges for performing scene text detection.*

Keywords*: binarization, text detection, text extraction, text segmentation, scene image, character recognition*

1. INTRODUCTION

Retrieving texts in both indoor and outdoor environments provides appropriate clues for a wide variety vision related tasks. A text in any form or place contains more information related to the place and helps us to understand the objective more easily. The rapid growth in digital technologies and gadgets equipped with megapixel cameras and inventions of latest touch screen method in digital devices like PDA, mobile etc. increase the demand for information retrieval and it leads to many research challenges [7]. Text detection and segmentation from natural scene images are useful in many applications. Text reorganization from the detected text lines is a challenging problem due to the variety of colors, fonts, existence of complex backgrounds and the short length of the text strings. Extraction of text from scene image is concerned with extracting the relevant text data from a collection of images. Text data present in images contain useful information for indexing and structuring automatic annotation of images. Extraction of such type of information involves detection, localization which is used to determine the location of text in the image, tracking, extraction enhancement and recognition in which extracted text image can be transformed into text from a given input image.

The text information can be extracted in two stages: text detection and text recognition. Text detection detects the text regions as extremal regions of an image and in text recognition stage system retrieves the text information from these extremal regions [8]. However, variations of text due to differences in style, size, alignment and orientation as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging [7]. The main challenge in scene text detection is to design a system which is flexible to handle all variability in our daily life including scene text, several character fonts and sizes and inconsistency in imaging conditions. Proposed solutions for all text understanding steps must be context independent that means independent of lighting colors, scenes and all different conditions[10]. Text can be used to easily and clearly describe the contents of an image.

The algorithms of text extraction from images can be broadly classified under three types. They are gradient feature based, color segmentation based and texture analysis based. The gradient feature based algorithm is based on the idea that pixels which have high gradient are the candidates of characters since edges exist.

2. RELATED WORK

Chucai Yi and Yingli Tian proposed a new framework to extract text strings with multiple sizes and colors, and arbitrary orientation from scene images with a complex and cluttered background [1]. The proposed framework consists of two main steps 1) image partition to find text character candidates based on gradient features and color uniformity. In

this step Chucai Yi and Yingli Tian propose two methods to partition scene images into binary maps of non-overlapped connected components: gradient-based method and color-based method. 2) character candidate grouping to detect text strings base on joint structural features of text characters in each text string such as character sizes, distance between two neighboring characters and character alignment. In this step , Chucai Yi and Yingli Tian propose two methods of strucyural analysis of text strings: adjacent character grouping method and text line grouping method.

Hyung II Koo and Duck Hoon [2] proposed a scene text detection algorithm based on two machine learning classifiers : first classifier is used to generate candidate word regions and the other is used to filter out nontext regions. Connected components (CCs) in images are extracted by using the maximally stable extremal region algorithm then these extracted CCs are grouped into clusters so that candidate regions are generated. Then candidate word regions are normalized and it is determined that whether each region contains text or not. CC based method begin with CC extraction and localize text regions by processing only CC level information. A text/nontext classifier for normalized images is developed because the skew, scale and color of each candidate can be expected from CCs. This classifier is based on multilayer perceptrons and with a single free parameter recall and precision rates can be controlled.

K. Subramanian, P. Natarajan, M. Decerbo, D.. Castanon[3] approached the text-localization problem using a CC-based approach by first detecting character strokes and then a threshold and stroke-width which are used for character segmentation are estimated by the detected stroke. The sensitivity of the detection algorithm to three key parameters is evaluated against four matrices : stroke precision, character recall, word recall and computing time. Character detection algorithm is not capable to perform well on italic fonts or when characters of a word are encrusted together. The performance of the system can be improved by working directly on color space to detect character strokes.

Basilios Gatos proposes a novel methodology for text detection in natural scene images [4]. The proposed methodology is based on an efficient binarization and enhancement technique followed by a suitable connected component analysis procedure. Image binarization successfully processes natural scene images having shadows, non-uniform illumination, low contrast and large signal-dependent noise. Connected component analysis is used to define the final binary image that mainly consists of text regions.

X. Chen, J. Yang, J. Zhang, A. Waibel [5] combined 1) multiresolution and multiscale edge detection 2) adaptive searching, 3) color analysis 4) affine rectification in a hierarchical framework for sign detection with different priority at each phase to handle the text in different orientations, sizes, color, distributions and backgrounds. They used affine rectification to improve deformation of the text regions caused by an inappropriate camera view angle. They extracted features from an intensity image directly rather using binary information for OCR. They have utilized this approach in developing a Chinese sign system, which can automatically detect and recognize Chinese sign as input from a camera and can translate the recognized text into English. The procedure can extensively improve text detection and optical character recognition (OCR) accuracy.

Boris Epshtein, Eyal Ofek, Yontan Wexler [6], proposed a novel image operator that is used to find the value of stroke width for each image pixel and exhibit its use in natural images for text detection. The proposed image operator is local and data dependent which makes it fast and robust enough to reduce the need for multi-scale computation. Its simplicity allows the algorithm to detect texts in many fonts and languages. The grouping of letters can be enhanced by considering the directions of the improved strokes and curved text lines can be detected as well.

Y. Pan, X. Hou and C. Liu [7], proposed a hybrid approach to localize scene texts by integrating region information into a robust CC-based method. Parameters of a conditional random field (CRF) model are jointly optimized by supervised learning and the binary contextual component properties are incorporated in CRF model. The proposed hybrid approach needs further improvements because this approach fails on some texts that are difficult to segment.

3. SCENE TEXT DETECTION TECHNIQUES

Text detection process consists of five stages which are text region detection, text localization, tracking, character extraction and text recognition. Among these stages first two (text region detection, text localization) stages are more important and they are more difficult to implement. The goal of the two stages is to generate accurate bounding boxes of all text objects in images and provide a unique identity to each text. In this section, the recent techniques focused on text detection are reviewed.

3.1 Region-Based Technique

Region-based techniques use the properties of the color or gray scale or alignment in a text region or their differences in properties of the background. In this technique, pixels having certain similar properties are grouped together. For region-based technique, the computation speed is low but can detect texts at any scale. This method can be further divided into two sub-approaches: edge-based and connected component (CC)-based. These two approaches follow a bottom-up fashion in which sub-structures, such as CCs or edges are identified and then these substructures are merged to spot bounding boxes for text [2], [9].

3.1.1 CC-based Technique

Connected component based methods use bottom up approach in which grouping smaller components are done into larger components until all regions are identified in the image. A geometrical analysis used to identify text components and group them using the spatial arrangements of the components so as to filter out non-text components and mark the boundaries of the text regions. The advantage of CC-based methods is that they have lower computational complexity. The performance of the CC-based method is degraded while dealing with the texts in complex background [11].

3.1.2 Edge-Based Technique

The edges of the text boundary are identified and merged, and then several heuristics are used to alter out the non-text regions. Among the several textual properties edge based focus on 'high contrast between the text and the background'. Usually, an edge filter is used for the edge detection and morphological operator is used for the merging stage [12].

3.2 Texture-Based Techniques

Texture based method is a feature based approach which involves the construction of gray-level co-occurrence matrix. This matrix is used to calculate the features like homogeneity, contrast, dissimilarity and which are the results for feature extraction in texture based technique. Texture-based techniques make use of the observation that text in images have distinct textural properties that distinguish them from the background. The techniques based on Wavelet, Gabor filters, FFT, spatial variance can be utilized to detect textural properties of a text region in an image. This technique has certain limitations including big computational complexity because of the need of scanning the image at several scales, inability to detect sufficiently slanted text [9].

3.3 Stroke Width Transform (SWT)

A new method called stroke width transform (SWT) is used in order to overcome the limitations of the previous methods such as high computational complexity and the difficulty to select best features for scene text detection. Stroke width transform converts value of each color pixel into the width of most likely stroke and then neighboring pixels with approximately similar stroke width are merged into the connected components so that the resulting system will be able to detect text regardless of its font, scale and direction.

4. APPLICATIONS

Text detection from images finds many useful applications in document analysis, vehicle license plate extraction, content based image retrieval, text-based image indexing, video content analysis, industrial automation etc and many applications have become realities in recent years [13]. Educational and training video and TV programs such as news contain mixed text-picture-graphics regions. Region classification is helpful in object-based compression.

5. GENERAL CHALLENGES

A lot of approaches have been developed on text detection in real application. But a fast and robust algorithms for detecting text under various conditions need to be further investigated. To develop a fast and robust text detection algorithm is a non trivial task since there exists such difficulties as:

- Text may be embedded in complex background
- It is difficult to find effective features to discriminate text with other text-like things, such as leaves, window curtains or other general textures
- Text pattern varies with different font-size, font-color and languages
- Text quality decrease due to noise
- It is difficult to detect text of arbitrary orientations

6. CONCLUSION

In this paper we present short review on various techniques used for scene text detection. We have mainly discussed three techniques namely region based technique, texture based technique and Stroke Width Transform (SWT). Stroke width transform technique is recently useful to solve the problems of scene text detection related to the large variations in character size, font, texture, color etc. We also discussed the challenges that are faced for scene text detection.

REFERENCES

[1] Text String Detection from Natural Scenes by Structure-based Partition and Grouping by Chucai Yi and YingLibTian, Senior Member, IEEE

[2] Hyung II Koo and Duck Hoon Kim, "Scene Text Detection via Connected Clustering and Nontext Filtering', IEEE Transactions On Image Processing, Vol. 22, No. 6, June 2013.

[3] K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon, "Character-Stroke Detection for Text-Localization and Extraction", International Conference on Document Analysis and Recognition (ICDAR), 2005.

[4] Towards Text Recognition in Natural Scene Images by Basilios Gatos, Ioannis Pratikakis and Stavros Perantonis.

[5] X. Chen, J. Yang, J. Zhang and A. Waibel, "Automatic Detection and Recognition of Signs from Natural Scenes" IEEE Transactions on Image Processing VOL. 13, No. 1, January 2004.

[6] B.Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2963-2970

[7] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. On Image Processing, vol. 20, no. 3, pp. 800, Mar. 2011.

[8] Xiaobing Wang, Yonghang Song, Yuanlin Zhang, "Natural scene text detection in multi-channel connected component segmentation", 12th International conf. on Document Analysis and Recognition, pp. 1375-1379, 2013.

[9] K. Jung, "Text information extraction in images and video: A survey," Pattern Recognit., vol. 37, no. 5, pp. 977-997, May 2004.

[10] Ms. Saumya Sucharita Sahoo, "Review of Methods of Scene Text Detection and its Challenges," International journal of Electronics and Communication Engineering (IJECET), Volume 5, Issue 1, January (2014), pp. 74-81

[11] Yao Li and Huchuan Lu, "Scene Text Detection via Stroke Width" International Conference on Pattern Recognition (ICPR 2012) November 2012.

[12] M. Swamy Das, B. Hima Bindhu, A. Govardhan, "Evaluation of Text Detection and Localization Methods in Natural Images," International Journal of Emerging Technology and Advanced Engineering (IETAE), Volume 2, Issue 6, June 2012.

[13] Karin Sobottka, Horst Bunke and Heino Kronenberg, " Identification of Text on Colored Book and Journal Covers", Document Analysis and Recognition, 20-22 1999