International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey Paper on Scene Text Detection Methods

Manjushree Kaware[1], Priti Saktel[2]

[1]M.tech. Scholar, Department of Computer Science & Engg, GHRIETW, Nagpur, Maharashtra, India

[2]Assistant Professor, Department of Computer Science & Engg, GHRIETW, Nagpur, Maharashtra, India

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. Extraction of text information from images or scene involves detection, localization, tracking, segmentation, enhancement and character recognition. But because of variations involved in text such as font style, size, orientation, alignment, reflections and illumination effect, as well as low image contrast and complex background details make text extraction process more challenging. A large number of methods have been proposed to address this problem but still none of them are perfect. This paper presents a short survey on various scene text detection methods suggested and implemented recently. General challenges for performing scene text detection are also discussed.***

Keywords: text extraction, text detection, localization, segmentation, character recognition.

1. Introduction

Retrieving texts in both indoor and outdoor environments provides appropriate clues for a wide variety of 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 invention of latest touch screen method in digital devices like PDA, mobile, etc. increase the demand for information retrieval and it leads to many new research challenges [5]. Detection of text and segmentation from natural scene images are useful in many applications. Recognizing text 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. Text data present in images contain useful information for indexing, and structuring, automatic annotation of images. Extraction of such type of information involves detection (The text detection stage detects the presence of text in a given input image), localization which is used to determine the location of text in the image & generating bounding boxes around the text, tracking (To reduce the processing time for localization), extraction (In this text components are segmented from the background.), enhancement (To magnify small text at a higher resolution), and recognition in which extracted text image can be transformed into text of the text from a given input image. 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].

Text can be used to easily and clearly describe the contents of an image. Many varieties of applications are found in current studies that uses extracted text. Recently developed mobile banking application that is provided by the banking institutions facilitates the customers to carry out the transactions even on passing the image of the cheque to the server. Every such type of application depends on a Textual Information Extraction (TIE) system which can proficiently detect, localize and extract the text information present in the images. Textual Information Extraction system mainly has two phases in which text detection and localization is done in the first phase and text recognition is performed in the second phase in which detected text regions are specified to the OCR which recognizes the characters and gives the textual output [8]. 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 through uneven lighting, aliasing and shadowing. Proposed solutions for all text understanding steps must be context independent that means independent of lighting colors, scenes and all different conditions [9].

2. Related Work

Hyung Il Koo and Duck Hoon [1], 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. 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.

Boris Epshtein , Eyal Ofek, Yonatan Wexler [2], proposed a novel image operator that is used to find out 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.

X. Chen, J. Yang, J. Zhang, A. Waibel [3], 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 proposed a local intensity normalization method to effectively handle lighting variations, a Gabor transform is used to find local features and then for feature selection a linear discriminate analysis (LDA) method is applied. They have utilized this approach in developing a Chinese sign system, which can automatically detect and recognize Chinese signs as input from a camera, and can translate the recognized text into English. The procedure can extensively improve text detection rate and optical character recognition (OCR) accuracy.

K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon [4], 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.

Y. Pan, X. Hou, and C. Liu [5], 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 relationships with the unary component properties are incorporated in the CRF model. The proposed hybrid approach needs further improvements because this approach fails on some texts that are difficult to segment. The speed of the proposed hybrid approach need to be accelerated further. Text recognition should be included with text localization to complete the need of text information extraction as well.

3. Scene Text Detection Methods

A. Region-Based Methods

Region-based methods exploit the properties of the color or gray-scale in a text region or their differences with the corresponding properties of the background. In these methods, pixels having certain similar properties are grouped together. The methods in this category suffer from low computation speed but can simultaneously detect texts at any scale and are not limited to horizontal texts. These methods are divided into two sub-approaches: connected component (CC)-based and edge-based. These two approaches follow a bottom-up fashion in which sub-structures, such as CCs or edges are identified and then these sub-structures are merged to spot bounding boxes for text. [1], [7].

[a] CC-based Methods

A bottom-up approach is used in the CC-based methods by grouping small components into successively larger components until all regions are recognized in the image. A geometrical analysis is required to combine the text components using the spatial arrangement 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 methods is degraded while dealing with the texts in complex background [6].

[b] Edge-Based Methods

Along with the several textual properties in an image, edge-based methods focus on the 'high contrast between the text and the background'. Several heuristics are used to filter out the non-text regions after recognizing and combining the edges of the text boundary. Generally, an edge filter is used for the edge detection and a morphological operator is used for the merging stage [8].

B. Texture-Based Methods

Texture based method is a feature based algorithm in which gray-level co-occurrence matrix which is used to calculate the features such as homogeneity, contrast and dissimilarity. Texture-based methods 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, etc. can be utilized to detect the textural properties of a text region in an image. The methods in this category also having certain limitations including big computational complexity because of the need of scanning the image at several scales, inability to detect sufficiently slanted text [7].

C. Stroke Width Transform (SWT)

So, 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 the 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. Conclusion

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

pros and cons of each of the methods used and the challenges that are faced for scene text detection.

References

[1] Hyung Il Koo and Duck Hoon Kim, “Scene Text Detection via Connected Component Clustering and Nontext Filtering”, IEEE Transactions On Image Processing, Vol. 22, No. 6, June 2013.
[2] 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.
[3] 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.
[4] 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.
[5] 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.
[6] Yao Li and Huchuan Lu, “Scene Text Detection via Stroke Width” International Conference on Pattern Recognition (ICPR 2012) November 2012.
[7] K. Jung, “Text information extraction in images and video: A survey,” Pattern Recognit., vol. 37, no. 5, pp. 977–997, May 2004.
[8] 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 (IJETAE), Volume 2, Issue 6, June 2012.
[9] 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