International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Recognition of Text in Natural Images

Adiba Maniyar[1], Sujata Desai[2]

[1]*Student, Department of Computer Science and Engineering, BLDEA College of Engineering and Technology, Karnataka, India*

[2]*Professor, Department of Computer Science and Engineering, BLDEA College of Engineering and Technology, Karnataka, India*

Abstract - Text in an image provides vital information for interpreting its contents, and text in a scene can aid a variety of tasks from navigation to obstacle avoidance and odometry. Despite its value, however, detecting general text in images remains a challenging research problem. Motivated by the need to consider the widely varying forms of natural text, we propose a bottom-up approach to the problem, which reflects the characterness of an image region. In this sense, our approach mirrors the move from saliency detection methods to measures of objectness. In order to measure the characterness, we develop three novel cues that are tailored for character detection and a Bayesian method for their integration. Because text is made up of sets of characters, we then design a Markov random field model so as to exploit the inherent dependencies between characters. We experimentally demonstrate the effectiveness of our characterness cues as well as the advantage of Bayesian multicue integration. The proposed text detector outperforms state-of-the- art methods on a few benchmark scene text detection data sets. We also show that our measurement of characterness is superior than state-of-the-art saliency detection models when applied to the same task.

Key Words: Characterness, scene text detection, saliency detection.

1. INTRODUCTION

In recent years, use of multimedia technology has increased tremendously. In multimedia technology image is one of the important part and image can have different contents in it, such as face, human, scene, text, etc. Among all contents in images, text is found to be one of the most important features to understand the image contents. Text in images can be used as indexing purpose.

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].Retrieving the contents from images is very challenging because of image quality and background noise.

There are different kinds of images that have text as its part with background, such as document images, scene images and born-digital images. In which scene images are often

taken by cameras. The digital cameras and camera phones enables acquisition of image and video materials containing scene text like street signs, advertisements, billboards, or restaurant menus, but these devices also introduce new imaging conditions such as sensor noise, viewing angle, blur, variable illumination, uneven lighting, lower resolution, etc.

Text images can be classified into three types.

A) Document images: Document images are nothing but image-format of the document[6]. Document images can have text and graphics. This type of images are generated by scanners or camera phones, which acquire printed documents, historical documents, handwritten documents, books, etc.[7]. In which, the image is transformed from paper-based documents into image-format for electric read. In the early stage of text extraction, there is only focus on document images.

B) Scene images: Scene images contain the text, such as the advertising boards, banners, which is captured by the cameras; therefore scene text appears with the background part of the scene[6]. These types of images are very challenging to detect and recognize, because the backgrounds are complex, containing the text in different sizes, styles and alignments. Also, scene text is affected by lighting conditions and perspective distortions. The current OCR software cannot handle complex background interferences and non-orienting

C) Born-digital images: Born-digital images are generated by computer software and are saved as digital images. Compared with document images and scene images, there are more defects in born digital images, such as more complex foreground/background, low resolution, compression loss, and severe edge softness. Therefore, during text extraction, it is difficult to distinct the text from the background[6].

D) Heterogeneous text images: This image have all kinds of images such as scene text images, caption text, document image and born-digital image[7].

2. LITERATURE SURVEY

C. Yi and Y. Tian have proposed that text information in natural scene images serves as important clues for many image-based applications such as scene understanding, content-based image retrieval, assistive navigation, and automatic geocoding. However, locating text from a complex background with multiple colors is a challenging task. In this paper, we explore a new framework to detect text strings with arbitrary orientations in complex natural scene images. Our proposed framework of text string detection consists of two steps: 1) image partition to find text character candidates based on local gradient features and color uniformity of character components and 2) character candidate grouping to detect text strings based on joint structural features of text characters in each text string such as character size differences, distances between neighboring characters, and character alignment. By assuming that a text string has at least three characters, we propose two algorithms of text string detection: 1) adjacent character grouping method and 2) text line grouping method. The adjacent character grouping method calculates the sibling groups of each character candidate as string segments and then merges the intersecting sibling groups into text string. The text line grouping method performs Hough transform to fit text line among the centroids of text candidates. Each fitted text line describes the orientation of a potential text string. The detected text string is presented by a rectangle region covering all characters whose centroids are cascaded in its text line. To improve efficiency and accuracy, our algorithms are carried out in multi-scales. The proposed methods outperform the state-of-the-art results on the public Robust Reading Dataset, which contains text only in horizontal orientation. Furthermore, the effectiveness of our methods to detect text strings with arbitrary orientations is evaluated on the Oriented Scene Text Dataset collected by ourselves containing text strings in non horizontal orientations.[1]

N. B. Ali Mosleh and A. B. Hamza have put forward a text detection method based on a feature vector generated from connected components produced via the stroke width transform. Several properties, such as variant directionality of gradient of text edges, high contrast with background, and geometric properties of text components jointly with the properties found by the stroke width transform are considered in the formation of feature vectors. Then, k-means clustering is performed by employing the feature vectors in a bid to distinguish text and non-text components. Finally, the obtained text components are grouped and the remaining components are discarded. Since the stroke width transform relies on a precise edge detection scheme, we introduce a novel bandlet-based edge detector which is quite effective at obtaining text edges in images while dismissing noisy and foliage edges. Our experimental results indicate a high performance for the proposed method and the effectiveness of our proposed edge detector for text localization purposes.[2]

Y. Li and H. Lu have proposed that a novel text detection approach based on stroke width. Firstly, a unique contrast-enhanced Maximally Stable Extremal Region (MSER) algorithm is designed to extract character candidates. Secondly, simple geometric constrains are applied to remove non-text regions. Then by integrating stroke width generated from skeletons of those candidates, we reject remained false positives. Finally, MSERs are clustered into text regions. Experimental results on the ICDAR competition datasets demonstrate that our algorithm performs favorably against several state-of-the-art methods.[3]

X. Li, Y. Li, C. Shen, said that Salient object detection aims to locate objects that capture human attention within images. Previous approaches often pose this as a problem of image contrast analysis. In this work, we model an image as a hyper graph that utilizes a set of hyper edges to capture the contextual properties of image pixels or regions. As a result, the problem of salient object detection becomes one of finding salient vertices and hyper edges in the hyper graph. The main advantage of hyper graph modeling is that it takes into account each pixel's (or region's) affinity with its neighborhood as well as its separation from image background. Furthermore, we propose an alternative approach based on center-versus-surround contextual contrast analysis, which performs salient object detection by optimizing a cost-sensitive support vector machine (SVM) objective function. Experimental results on four challenging datasets demonstrate the effectiveness of the proposed approaches against the state-of-the-art approaches to salient object detection.[4]

R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk have proposed that detection of visually salient image regions is useful for applications like object segmentation, adaptive compression, and object recognition. In this paper, we introduce a method for salient region detection that outputs full resolution saliency maps with well-defined boundaries of salient objects. These boundaries are preserved by retaining substantially more frequency content from the original image than other existing techniques. Our method exploits features of color and luminance, is simple to implement, and is computationally efficient. We compare our algorithm to five state-of-the-art salient region detection methods with a frequency domain analysis, ground truth, and a salient object segmentation application. Our method outperforms the five algorithms both on the ground truth evaluation and on the segmentation task by achieving both higher precision and better recall.[5]

3. PROPOSED MODEL

Figure 3 shows the flow chart of the proposed method. The proposed system is designed as follows.

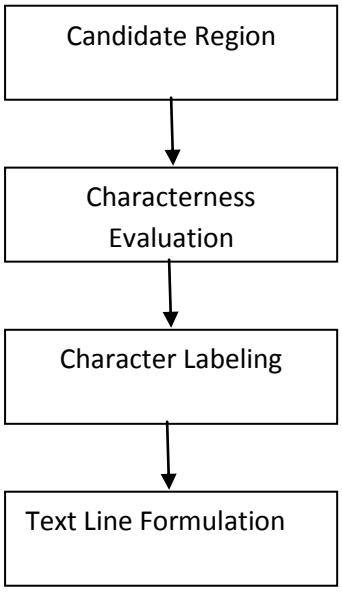

Figure 3: Flow Chart of Proposed Model

Candidate Region Extraction: MSER [48] is an effective region detector which has been applied in various vision tasks, such as tracking [49], image matching [50], and scene text detection [38], [41], [46], [51], [52] amongst others. Roughly speaking, for a gray- scale image, MSERs are those whose shape and size remain relatively unchanged over a set of different intensity thresh- olds. The MSER detector is thus particularly well suited for identifying regions with almost uniform intensity surrounded by contrasting background.

Characterness Evaluation: For characterness evaluation, three novel cues are proposed.

A) Stroke Width (SW): Stroke width has been a widely exploited feature for text detection [34], [36], [44], [45]. In particular, SWT [34] computes the length of a straight line between two edge pixels in the perpendicular direction, which is used as a preprocessing step for later algorithms [44], [55], [56]. In [45], a stroke is defined as a connected image region with uniform color and half-closed boundary. Although this assumption is not supported by some uncommon typefaces, stroke width remains a valuable cue.

B) Perceptual Divergence (PD): Color contrast is a widely adopted measurement of saliency. For the task of scene text detection, we observed that, in order to ensure reasonable readability of text to a human, the color of text in natural scenes is typically distinct from that of the surrounding area.

C) Histogram of Gradients at Edges (eHOG): The Histogram of Gradients (HOGs) [58] is an effective feature descriptor which captures the distribution of gradient magnitude and orientation. Inspired by [35], we propose a characterness cue based on the gradient orientation at edges of a region, denoted by eHOG. This cue aims to exploit the fact that the edge pixels of characters typically appear in pairs with opposing gradient directions [35]

Character Labeling: We cast the task of separating characters from non-characters as a binary labeling problem.

Text Line Formulation: The goal of this step, given a set of characters identified in the previous step, is to group them into readable lines of text. A comparable step is carried out in most region-based text detection approaches [34], [44], In this work, we introduce text line formulation. Specifically, two normalized features (characteristic scale and major orientation [44]) are exploited to group regions into clusters via mean shift. For each cluster with at least two elements, we group elements into text lines based on their spatial distance measured by Euclidean norm.

4. CONCLUSION AND FUTURE SCOPE

In this work, we have proposed a scene text detection approach based on measuring 'characterness'. The proposed characterness model reflects the probability of extracted regions belonging to character, which is constructed via fusion of novel characterness cues in the Bayesian frame-work. We have demonstrated that this model significantly outperforms the state-of-the-art saliency detection approaches in the task of measuring the 'characterness' of text. In the character labeling model, by constructing a standard graph, not only characterness score of individual regions is considered, similarity between regions is also adopted as the pairwise potential. Compared with state-of-the-art scene text detection approaches, we have shown that our method is able to achieve accurate and robust results of scene text detection.

This work can be extended by extracting the text from natural scene images. It allows separating text from background. It has lot of scope in document image processing.

REFERENCES

[1] 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, Apr. 2011

[2] N. B. Ali Mosleh and A. B. Hamza, "Image text detection using a bandlet-based edge detector and stroke width transform," in Proc. Brit. Mach. Vis. Conf., 2012, pp. 1–12.

[3] Y. Li and H. Lu, "Scene text detection via stroke width," in Proc. IEEE Int. Conf. Pattern Recognit., Nov. 2012, pp. 681–684.

[4] X. Li, Y. Li, C. Shen, A. Dick, and A. van den Hengel, "Contextual hypergraph modeling for salient object detection," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2013, pp. 3328–3335.

[5] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1597–1604.

[6]Jian Zhang, Renhong Cheng, Kai Wang, Hong Zhao, "Research on the text detection and extration from complex images", Fourth International Conference on Emerging Intelligent Data and Web Technologies. Vol. 10, 2013, Page no. 708-713.

[7] C.P. Sumathi, T. Santhanam, G.Gayathri Devi, "A Survey On Various Approaches Of text Extraction In Images", International Journal of Computer Science & Engineering Survey (IJCSES). Vol.3, August 2012, Page no. 27-42. [3]

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

[9] J. Zhang and R. Kasturi, "Text detection using edge gradient and graph spectrum," in Proc. IEEE 20th Int. Conf. Pattern Recognit., Aug. 2010, pp. 3979–3982.

[10] J. Zhang and R. Kasturi, "Character energy and link energy-based text extraction in scene images," in Proc. Asian Conf. Comput. Vis., 2010, pp. 308–320.

[11] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 2609–2612.

[12] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in Proc. Asian Conf. Comput. Vis., 2010, pp. 770–783.

[13] 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., Jun. 2012, pp. 1083–1090.

[14] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4256–4268, Sep. 2012.

[15] H. Koo and D. Kim, "Scene text detection via connected component clustering and non-text filtering," IEEE Trans. Image Process., vol. 22, no. 6, pp. 2296–2305, Jun. 2013.
[16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. Brit. Mach. Vis. Conf., 2002, pp. 384–393.

[17] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2006, pp. 553–560.

[18] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in Proc. IEEE 11th Int. Conf. Comput. Vis., Oct. 2007, pp. 1–8.

[19] L. Neumann and J. Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in Proc. IEEE Int. Conf. Document Anal. Recognit., Sep. 2011, pp. 687–691.

[20] S. Tsai, V. Parameswaran, J. Berclaz, R. Vedantham, R. Grzeszczuk, and B. Girod, "Design of a text detection system via hypothesis generation and verification," in Proc. Asian Conf. Comput. Vis., 2012, pp. 1–12.

[21] N. B. Ali Mosleh and A. B. Hamza, "Image text detection using a bandlet-based edge detector and stroke width transform," in Proc. Brit. Mach. Vis. Conf., 2012, pp. 1–12.

[22] J. Pan, Y. Chen, B. Anderson, P. Berkhin, and T. Kanade, "Effectively leveraging visual context to detect texts in natural scenes," in Proc. Asian Conf. Comput. Vis., 2012, pp. 1–20.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 886–893.