

A Review Paper on Text Detection, Classification and Segmentation

Dr. Poonam Patel¹, Shweta Gandevia²

^{1,2}SDJ International College, Plot no. 166, Opp. IDI, B/h Someshwara Bungalows, Vesu, Surat, 395007, India

Abstract: This paper analyzes, compares, and identifies the different stages in the process of text detection and recognition. To detect text from the images, some basic image processing techniques are used. That includes pre-processing, extraction or text localization, classification and character detection. There are two methods for this problem is stepwise method and integrated method. Sub problems are highlighted including text localization, verification, segmentation and recognition from image. This survey provides a fundamental comparison and analysis of the remaining problems in the field.

Keywords: Text detection, Segmentation, Extraction, SVM

1. Introduction

The problems of text detection and recognition in images and video have received increased attention from past many years. Text is one of the most expressive means of communication, and can be embedded into documents or into scenes as a means of communicating information. This is done in the way that makes it “noticeable” and/or readable by others. The text has an important role in our life for communication purpose. Text means either scene text or caption text. Text detection is to identify text from given input images. *Text detection is the method of discovering areas in a picture wherever, text is present.* There are many techniques exists for the text detection [1]. Retrieving text from both in within and external is difficult due to variety of backgrounds. Text detection can be easily done with the help of image retrieval algorithm. Fig.1 shows the basic text detection method in image processing technique.

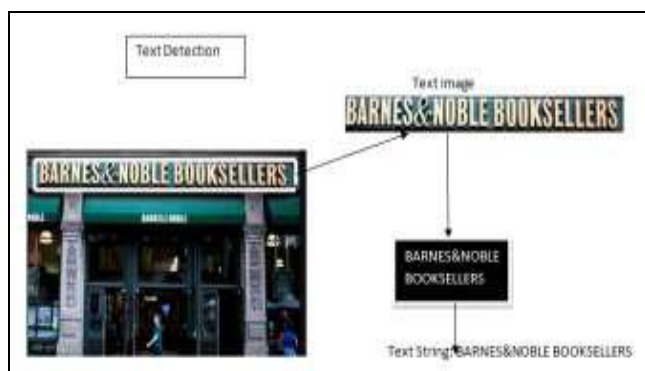


Figure 1: Sample screenshot of Automated Text detection System

Currently text detection faces difficulties like background complexity, different direction of the text, complexity in backgrounds, diversity of scene text and interference factors etc. In order to overcome these difficulties, CNN (Convolution neural networks) and related methods are used [2], and also in some cases [3] and valley and ridge techniques are used.

Text detection and recognition from video captions and web pages is also getting attention. Voluminous work has been done in the field of text detection and extraction from natural scenes images. Various optical character recognition techniques are also available. Still problem of text detection and recognition is not thoroughly solved. Segmentation and extraction of text from natural scene images is still very difficult to achieve.

Text detection and recognition system have a number of important application areas, including advanced driver assistance systems, road surveying, and autonomous vehicles. This paper studies various stages in process of text detection and recognition and analyses and compares different approaches used to undergo these stages. It presents importance of every processing stage and advantages, disadvantages and applications of approaches used by various contributors to solve these problems. Various applications of text detection, recognition and classification are also reviewed in this paper.

1.1 The Challenges of Text Detection

However, reading text from different images is a difficult task. The main text detection challenges are categorized as follows [4].

Table 1: Challenges in Text Detection and Recognition

Category	Sub-category
Environment	Scene complexity, Uneven lighting
Image acquisition	Blurring/degradation, Perspective distortion
Text content	Variation of aspect ratio, Multi-oriented/curved text, Variation of fonts, Multilingual environments
Diversity of natural images	Different font style, different size, unique color, unique font alignment etc.
Complexity in backgrounds	Background includes rocks, grasses, bricks and sign boards are there, and that leads to more complexity for identifying text
Inference factors	Inference factors are blurring, noise and has low resolution of input images

2. Methodologies

Volume 9 Issue 2, February 2020

www.ijsr.net

Licensed Under Creative Commons Attribution CC BY

The process of detecting and recognizing text is divided into text detection stage and recognition stage [5]. Text detection deals with finding text region from input image, whereas recognition deals with converting obtained text into characters and words. Methods used for this purpose are categorized as stepwise methods and integrated methods. Stepwise methods have separate stages of detection and recognition and they proceed through detection, classification, segmentation and recognition. Integrated methods have information sharing amongst detection and recognition stages and these methods aim at recognizing words from text available.

2.1 Stepwise method

Stepwise methodologies have four primary steps: localization, verification, segmentation, and recognition. The localization step coarsely classifies components and groups them into candidate text regions, which are further classified into text or non-text regions during verification. The underlying assumption is that various text regions might be regarded as a kind of uniform pattern, therefore, there must exist properties or features that are invariant over this pattern.

The segmentation step separates the characters so that exclusive, accurate outlines of image blocks remain for there

cognition step. Finally, the recognition step converts image blocks into characters. In some approaches, the verification and/or segmentation step could be ignored, or additional steps might be included to perform text enhancement and/or rectification.

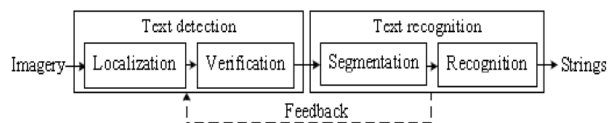


Figure 2: Framework of Stepwise methodology.

2.2 Integrated method

With an integrated methodology, character classification responses are considered the primary cues, and shared with detection and recognition modules [15]. Using character classification responses as primary features requires the discrimination of characters from the background as well as from each other, which is a complex multiclass problem. Solutions require not only robust character recognition models but also appropriate integration strategies, such as holistic matching, i.e., “word spotting” [17], joint optimization [16] and/or decision delay [18]. Word spotting looks to match specific words in a given lexicon with image patches by character and word models.

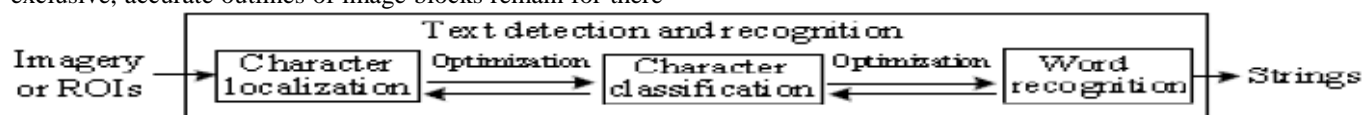


Figure 3: Framework of Integrated methodology

3. Important stages of text detection and recognition

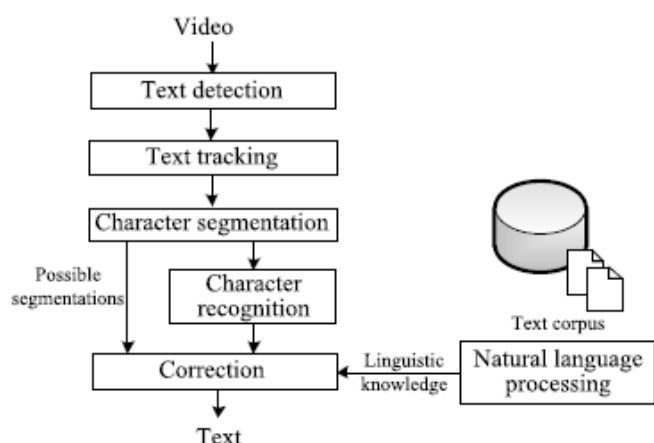


Figure 4: Flowchart of the stepwise video text recognition approach with detection, tracking, segmentation, recognition and language processing [6]

3.1 Text detection and localization

Text detection deals with detecting presence of the text in the input image whereas text localization localizes position of the text and forms groups of text regions by eliminating maximum of the background. Text detection and localization process is performed using connected component analysis or region based methods.

3.2. Classification

After text detection and localization stage output may contain non-text regions along with text regions as false positives. Classification stage verifies text regions and eliminates non-text regions using classification algorithms. This stage can also be called as verification. Classification algorithms are either supervised or unsupervised.

Supervised algorithms know properties of text such as color, size, texture, etc. before classification. Unsupervised algorithms do not have prior knowledge about text features. Supervised classification algorithms need training before classification. These algorithms undergo training to be able to extract features of the text to be classified and use these features in classification phase.

Unsupervised classification algorithm does not undergo training. They extract features of text during the classification phase only unlike supervised classification and they use features extracted in previous classification for next one. This is similar to adaptive learning.

3.3. Segmentation

Segmentation process is used to separate text from background and to extract bounded text from image.

Integrated method which focuses on word matching/ recognition often combine or replace complex segmentation stage with recognition stage however stepwise methods undergo segmentation to obtain precisely extracted characters which are fed to recognition stage. Binarization & character segmentation are few of the segmentation algorithms

Binarization converts color or gray-scale image into black and white image. To achieve good segmentation result irrespective of dark or bright text or background uses adaptive thresholding for binarization [7].



Figure 5: Binarization of text from street view [9]

Character segmentation is process of converting text into multiple sets of single characters. It is suitable in case of degraded text or connected characters. Gradient vector flow based method is used in [8], which is applied directly on grayscale images eliminating need of binarization. It initially identifies candidate cut pixels from the characters and then uses two pass path finding process that finds out potential cuts in forward pass and verify true cuts and remove false cuts in backward pass.

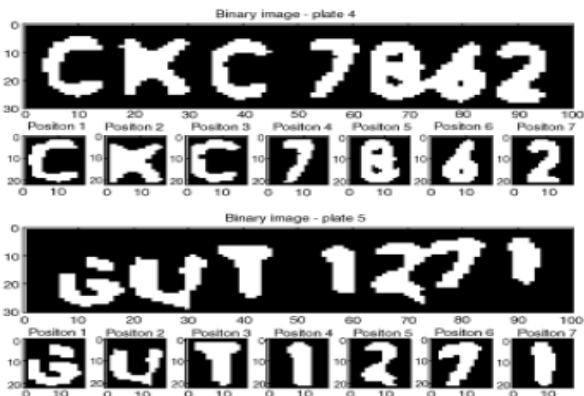


Figure 6: Results of character segmentation of degraded characters [10]

3.4 Text Recognition

Text recognition stage converts images of text into string of characters or words. It is important to convert images of text into words as word is an elementary entity used by human for his visual recognition. Different approaches of recognition are character recognition and word recognition. Character recognition methods divide text image into multiple cut-outs of single characters. Separation between adjacent characters is very important for these methods.

Character recognition methods divide text image into multiple cut-outs of single characters. Separation between adjacent characters is very important for these methods.

Character recognition approach using Optical Character Recognition module (OCR) is used in [11], where initially images are segmented into k classes followed by binary text image hypothesis generation which passes through connected components analysis and gray scale consistency constraint module before getting fed to OCR. Support Vector Machine (SVM) based classifier is used for character recognition in [12]. SVM gives good support for multi-class classification, which is tested in [12] on Indic language Kannada which has total 578 characters formed by altering 34 base consonants using 16 vowels. Many of these characters falls in similar classes which has made use of SVM based layered classification approach easier.

Word recognition uses character recognition outputs along with language models or lexicons to recognize words from text image. It can be used in case of degraded characters. For applications with limited number of word possibilities in input images, word recognition is better approach than character recognition. Wachenfeld, et al. used graphical approach for word recognition in [13]. During segmentation stage, when each character is segmented, a hypothesis graph is formed to represent every segment and each path of graph is an ordered segment sequence which leads to formation of words from characters.

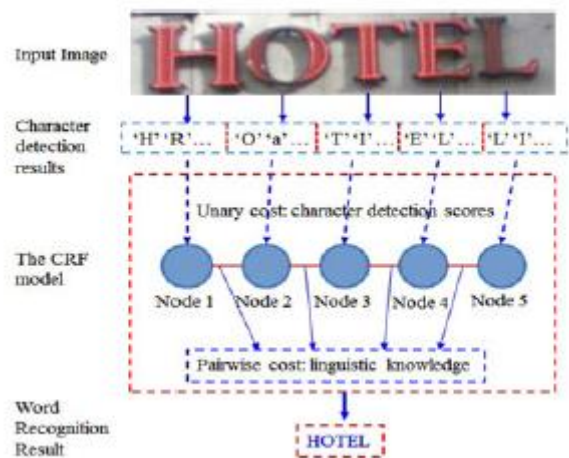


Figure 7: Word recognition with a CRF model [14]

4. Applications

- 1) Text detection and recognition is used in industries for reading package labels, numbers etc.
- 2) It is used to retrieve video captions as well as specific text contents from web pages.
- 3) It is used for automatic number plate recognition at toll booths as well as for street boards reading purpose in case of unnamed vehicles.
- 4) Text detection and recognition has very important application in form of assisting blind or visually impaired people for reading, making their daily life easy.
- 5) It is also used in automatic cheque signature reading.

6) Automatic document scanning is another application of text recognition.

5. Conclusion

Text detection is applicable in real world scenarios like optical character recognition, artificial intelligence; distinguish between human and machine inputs and spam removal. Text detection is the process of locating areas in an image where, a meaning full text is occurred. Variation in environment in which the image is captured makes it a difficult process.

References

- [1] Y. Li, W. Jia, C. Shen, and A. van den Hengel, "Characterness: An indicator of text in the wild," *IEEE Trans. ImageProcessing (TIP)*, vol. 23, pp. 1666–1677, 2014.
- [2] Shao Y., Wang C., Xiao B., Zhang Y., Zhang L., Ma L. (2010) Text Detection in Natural Images Based on Character Classification. In: Qiu G., Lam K.M., Kiya H., Xue XY., Kuo CC.J., Lew M.S. (Eds) *Advances in Multimedia Information Processing - PCM 2010*. PCM 2010. Lecture Notes in Computer Science, vol 6298. Springer, Berlin, Heidelberg.
- [3] B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 2010, pp. 2963-2970.
- [4] A. Vinciarelli, "A survey on off-line word recognition," *Pattern Recognit.*, vol. 35, no. 7, pp. 1433–1446, 2002.
- [5] Qixiang Ye, David Doermann, "Text Detection and Recognition in Imagery: A Survey" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- [6] K. Elagouni, C. Garcia, and P. Sbillot, "A comprehensive neuralbased approach for text recognition in videos using natural language processing," in *Proc. ACM Conf. Multimedia Retrieval*, 2011.
- [7] W. Kim and C. Kim, "A New Approach for Overlay Text Detection and Extraction from Complex Video Scene," *IEEE Trans. Image Processing*, vol. 18, no. 2, pp. 401-411, 2009.
- [8] T. Phan, P. Shivakumara, B. Su and C.L. Tan, "A Gradient Vector Flow-Based Method for Video Character Segmentation," in *Proc. IEEE Int'l Conf. Document Analysis and Recognition*, pp. 1024-1028, 2011.
- [9] J.J. Weinman, Z. Butler, D. Knoll, J. Feild, "Toward Integrated Scene Text Reading," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 3, no. 2, pp. 375-387, 2014.
- [10] S. Nomura, K. Yamanak, O. Katai, H. Kawakami, T. Shiose, "A Novel Adaptive Morphological Approach for Degraded Character Image Segmentation," *Pattern Recognition*, vol. 38, no. 11, pp. 1961-1975, 2005.
- [11] D. Chen, J.M. Odobez, H. Bourlard, "Text Detection and Recognition in Images and Video Frames," *Pattern Recognition*, vol. 37, no. 3, pp. 596-608, 2004.
- [12] K. Sheshadri, S.K. Divvala, "Exemplar Driven Character Recognition in the Wild," in *Proc. British Machine Vision Conference*, pp. 1-10, 2012.
- [13] S. Wachenfeld, H. Klein and X. Jiang, "Recognition of Screen-Rendered Text," in *Proc. IEEE Int'l Conf. Pattern Recognition*, pp. 1086-1089, 2006.1733-1746, 2009.
- [14] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, "Scene text recognition using part-based tree-structured character detections," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2961–2968.
- [15] W. Wu, D. Chen, and J. Yang, "Integrating co-training and recognition for text detection," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2005, pp. 1166–1169.
- [16] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolution neural networks," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2012, pp. 3304–3308.
- [17] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1457–1464.
- [18] T. Yamazoe, M. Etoh, T. Yoshimura, and K. Tsujino, "Hypothesis preservation approach to scene text recognition with weighted finite-state transducer," in *Proc. Int. Conf. Doc. Anal. Recognit.*, 2011, pp. 359–363.
- [19] L. Neumann and J. Matas, "On combining multiple segmentations in scene text recognition," in *Proc. IEEE Int. Conf. Doc. Anal. Recognit.*, 2013, pp. 523–527.
- [20] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multiscale representation for scene text recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4042–4049.
- [21] X. C. Yin, X. Yin, K. Huang, and H. Hao, "Robust text detection in natural scene images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 970–983, May 2014.
- [22] J. H. Seok and J. H. Kim, "Scene text recognition with a Hough forest implicit shape model," in *Proc. IEEE Int. Conf. Doc. Anal. Recognit.*, 2013, pp. 599–603.