

A Survey on Scene Text Detection and Text Recognition

Priyanka Patil¹, S I Nipanikar²

Student, Department of E&TC, PVPIT, Pune, Maharashtra, India¹

Professor, Department of E&TC, PVPIT, Pune, Maharashtra, India²

Abstract: Now a day's reading words in unimpeded and noisy images is a not an easy problem. Also extracting words or text from natural scenes or videos is very difficult thing. So we should have some mechanism to recognize the text from images and videos. In this paper an effort has been taken to focus the progress made so far in the character feature of scene text recognition system and an overview of technological prospective of scene text recognition systems are discussed. In Text recognition, feature extraction required much attention because recognition performance heavily depends on it.

Keywords: Text Recognition, Text Detection, Character Features, Feature Extraction.

I. INTRODUCTION

Now a day's everybody want to give a unique name with the unique font style to their shops, institutes, buildings, hotels, restaurants to attract people. And it is very difficult task to understand the same. Because the intensity of text is affected by shadow, low lights, high lights and the weather conditions, language of the text and orientation. Reading character from photographs is again a challenging thing, as it has very low visibility and needed to have some system on place which should make such text readable to everybody. Scene text recognition method is the answer to all these challenges. We should have some technologies which can recognition the text from natural scene images and videos. Text recognition from any natural scenes images and videos is application of image processing technique. Text recognition is depends on the pattern of image text. For that text recognition, localization and understanding is the core methods used.

We widely started using cell phones and development for the same is growing like anything. It has become the need of individual to use cameras to capture information. There are many applications running over cell phone to understand captured text information, but to understand text from natural scene images or videos is little tricky job. So text recognition from such scenes using mobile application is one good option.

Text detection and recognition in general have quite a lot of relevant functions for automatic information retrieval such document indexing, content-based image retrieval text or character recognition.

II. BLOCK DIAGRAM

Following diagram shows the process of Scene Text Recognition system.

Text Detection: Text detection is the method used to detect the Text or character area from original scene image. There may be scattered or noisy images where it is difficult to understand the text area. This method helps to identify such areas.

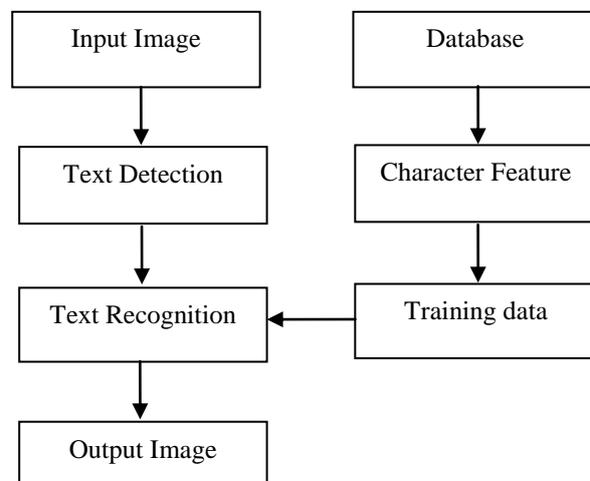


Fig1. Block Diagram of Scene Text Recognition System

Database & Character features: Database and Character features will have the collection of pattern data of characters and numbers also there features like shapes, angles and structures which is included in training dataset to verify the text against with available patterns

Training dataset: Training dataset is created using more than 200 text fonts of computer and additional characters to by the use of Probable occurrences of English language words

Text Recognition: once we have detected or extracted area of target image, we can recognize the characters or text with the help of training data. The detected image area compared by available training dataset which outputs the exact character written in image

III. RELATED WORK

Methods used for Text Detection:

The text detection stage seeks to detect the presence as well as to understand the text in a given image. Kim selected in 1996 using an automatic scene text detection

and context based indexing, a frame from shots detected by a scene-change detection method as a candidate containing text [1]. Very low threshold values are needed for scene-change detection because the target area of image relative to the whole image is usually small. This approach is very sensitive image change detection. Image text frame selection is performed at an interval of 1 to 2 seconds for caption text in the detected scene image frames. This will be very simple and efficient answer for video indexing applications that only need key words from video clips, rather than complete character text.

Mr. Smith and Kanade in 1995 defined a scene-change based on the difference between two successive frames and then used this scene change information for text detection[2]. They achieve an accuracy of 90% in scene-change detection.

Gargi et al. in 1998 all performed text detection using the assumption that the number of intracoded blocks in P and B different frames of an mpeg compacted video increases, when a text caption appear. [3]

Lim et al. in 2000, made a simple assumption that text usually has a higher intensity than the background [4]. Researchers figured out the number of pixels that are lighter than a defined exponential value and exhibited a significant colour difference relative to their neighbourhood, and regarded a frame with a large number of such pixels as a text frame. This method is extremely simple and fast. However, problems can occur with colour-reversed text.

Chukai y in July 2014 layout based detection is mainly focuses on colour decomposition and horizontal alignment, whereas it decomposes image in different colour's layers, and boundary clustering algorithm is used to uniformity of colour[11]. And all colour layers are horizontally aligned, by analysing geometrical properties of boundaries for text character detection.

Neuman has proposed, maximally stable extremal region detector is method is used to extract text characters from scene image to group the connected components [5]. False positives can be removed by the additional checks. The MSER methods helps to detect the text characters from images, even if they are low resolution, low contrast or noisy.

Priyanka N Guttedar in May 2015 proposed Cluster detection method. In this method layout analysis of colour conversion is performed on image and to search image region of text string[5], k-mean clustering is used which partitioned the dataset in clusters with respective define distance, using grouping of same number of pixels in an image.

Methods used for Text Recognition:

Xuejian Rong¹, Chucai Yi², Xiaodong Yang¹ and Yingli Tian proposed Scene text recognition based on multi frames technique by tracking text regions in videos taken using moving camera [11]. First a single framework of scene text is presented in multiple frames using scene text character (STC) for prediction of characters. And to configure model to words conditional random field (CRF) is used. Feature representation scene text character (STC) is done by use of Fisher vector, Sample SHIFT descriptor

and Fisher Vector and the datasets of extracted text information will be collected from scene videos. Text detection is carried in first frame of a target image or video. And it will generate initial bounding box. Text will be recognizing once CRF model will be applied on the same. This system is more compatible with the video or image based mobile applications. This system improved the performance of Scene text recognition process. Conditional random field (CRF) will give better results only if the size of image or video is increased Designing a robust fusion methods for multiple frames to increase the STC score and improve the scene text recognition performance at Word level text.

Priyanka N Guttedar, Pushpalata S in 2015 mentioned, text character is very important to provide valuable information [6]. For problem of character segmentation need to detect text. Therefore in this paper scene text detection and text recognition both are playing very important roles. Text detection deals with colour conversion and k-means clustering algorithm used for the same. Clustering algorithm is defined as, grouping same no of pixel from their distance measure and also colour conversion is important because of all know text character are generally localize with colour from background and colour clustering is efficient for binavigation. To recognize text, support vector machine classifier is used. SVM is a binary classifier. K-means clustering method for scene text recognition is simple and faster method. It is focused on determined the text from an image and to produce more accurate result. Need to improve accuracy rate for detection of text recognize for image as well as videos also.

Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang, Song Gao in 2013 has mentioned about part based tree structure to represent all type of character which can detect and recognize at the same time [7]. A framework prepared to recognize words, which has detection score and linguistic knowledge incorporated with the help of Conditional random field (CRF) model. There are nodes to target locations and these nodes are connected to their horizontal distance. This system recognizes the text from unconstrained scene images that to a high accuracy. This system can recognize texts from low resolution images. Also this method can help to language translation, understanding texts, sign readings and navigate the direction. It may reduce the processing time for recognizing text from scene.

Jerod J. Weinman, Member in 2009, IEEE, Erik Learned-Miller, Member, IEEE proposed, Gabor model method is used to recognize the text from a natural scene image. Gabor-based model is an appearance model or a linguistic model[8]. This is related to frequency and letter case, similarity model, and lexicon model to perform scene character recognition. For feature extraction process Gabor filter is used. Gabor filter is very effective in extracting the character from scene image. This paper highlighting the text recognition method with anyplace in this world or environment like different banners by using probabilistic graphical mode for Scene text recognition and the information provided by top-down and bottom up on the

same time. It will help to reduce the different errors with no loss in accuracy. Because of text recognized from an image similarity come with each other and reduces the character recognition error and as well as lexicon reduces word recognition error by more than 30 to 40% and sparse belief propagation reduces the lexicon words with Lossless accuracy

Anju Maria Babul, Radhika K. has proposed, newly identified method called FUZZY based image segmentation for improvement in performance of Scene text recognition for variable font sized text accurately and faster manner [9]. It processes image very easily. Firstly to detect the edges of target image, sobel edge detection technique is used, after that to minimize the computation effort image dilation is used to connect the broken components in the image. There are four main features which should be extracted while merging the characters, distance, position, height and colour. And Fuzzy linguistic rules have been applied on this character. This will lead to extract the character and recognize it of varying font size from a natural scene image. This method is very effective to detect variable font sized text that to accurately and faster manner.

Luka's Neumann Ji ri Matas in 2012 mentioned about, real time scene text localization and recognition method and which is depend on extremal regions (ER) [10]. Recognition and localization of text or Character in real-world (scene) images is a known problem which has been receiving a significant attention since it is an important part or component in a number of computer vision based applications like searching images by their textual part or component, for that each extremal region characters are estimated by the use of Novel feature. A highly efficient search with feedback is used to extract the words from Extremal regions. Positioning of character detection issues will help to the real time performance. Recognizing labels on business in map applications of computer or assisting visually in paired. This method is robust against noise and low contrast of characters.

Chukai Y. in July 2014 proposed, scene text detection and scene text recognition are two important phases'. For scene text detection layout based and horizontal alignment scheme used to detect image region or text region [11]. And in another phase means in text recognition, text understanding and text retrieval to recognize text from text detection information using Character descriptor and stroke configuration methods. Character descriptor is a combination of feature detector and descriptor. Dense, Random, Harris, MSER these four feature extraction method defined character classes and this character class is defined because of stroke configuration. Performance improvement of scene text detection. This method also support to combining low level feature descriptors. Scene text detection process is improved.

IV. CONCLUSION

In this paper, we defined different research paper's surveys. And to conclude the same we found that, there are different approaches mentioned in all papers for

character segmentation in natural scene image and each one of them is providing the improved performance with more accuracy. The sections mentioned above are all the compared results of methodologies. To provide better text detection we applied MSER method algorithm and for Text recognition we applied character descriptor method with approach text extraction algorithm, which we found the more relevant method

REFERENCES

- [1] H. K. Kim, Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database, *Journal of Visual Communication and Image Representation* 7 (4) (1996) 336-344.
- [2] M.A. Smith and T. Kanade, Video Skimming for Quick Browsing Based on Audio and Image Characterization, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995.
- [3] U. Gargi, S. Antani, and R. Kasturi, Indexing Text Events in Digital Video Database, *Proc. of International Conference on Pattern Recognition*, 1998, vol. 1. pp. 1481-1483
- [4] Y. K. Lim, S. H. Choi, and S.W. Lee, Text Extraction in MPEG Compressed Video for Content-based Indexing, *Proc. of International Conference on Pattern Recognition*, 2000 pp. 409-412.
- [5] Xuejian Rong1, Chucai Yi2, Xiaodong Yang1 and Yingli Tian, SCENE TEXT RECOGNITION IN MULTIPLE FRAMES BASED ON TEXT TRACKING
- [6] Priyanka N Guttedar, Pushpalata S. Scene Text Recognition in Mobile Application using K-Mean Clustering and Support Vector Machine
- [7] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z.Zhang, "Scene text recognition using part-based tree structured character detection", in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Conference, 2013, pp.2961-2968
- [8] J. J. Weinman, E. Learned-Miller, and A. R. Hanson, "Scene text recognition using similarity and a lexicon with sparse belief propagation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 31, no. 10, pp. 1733-1746, Oct. 2009.
- [9] Anju Maria Babul, Radhika K. - Improved Scene Text Recognition Using Fuzzy Based Method
- [10] Neumann, L., Matas, J.: Real-time scene text localization and recognition. In *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, California, US, IEEE (2012) 3538-3545
- [11] Chucai Yi, "Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration," *IEEE Trans. Image Process.*, vol. 23, no. 7, July2014, pp.2972-2982.