International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Improved Scene Text Recognition Using Fuzzy Based Method

Anju Maria Babu[1], Radhika K. T[2]

[1, 2]MEA Engineering College, Calicut University, Vengoor, Perinthelmanna, Malappuram, Kerala, India

Abstract: ***Scene Text Recognition plays a major role in finding vital and valuable information from the scene. Text recognition involves detection, localization, tracking, extraction, enhancement and recognition of the text from the given image. However these text characters are difficult to be detected and recognized due to their deviation of size, font, style, orientation, alignment, contrast, varying color, textured background. Several techniques have been developed for efficient extraction of the text from an image. Among these techniques most efficient one is Scene Text Recognition utilizing Structure-Guided Character Detection and Linguistic Knowledge. Since the character models make utilization of both the local appearance and global structure information, the detection results are more reliable. However this method cannot detect text of varying font size. To overcome this drawback we have incorporate a fuzzy based segmentation.***

Keywords: Part-based tree-structured models (TSMs), posterior Probability, Fuzzy based segmentation.

1. Introduction

Text Extraction from image is concerned with extracting the actual text data from images. Rapid development of digital technology has resulted in digitization of all categories of materials. Lot of resources are available in electronic medium. Many existing paper-predicated accumulations, historical manuscripts, records, books, journals, scanned document, book covers, video images, maps, manuscripts, pamphlets, posters, broadsides, newspapers, micro facsimile, microfilms, university archives, slides and films, book plates, pictures, painting, graphic materials, coins and currency, stamps, magazines, clipping files, edifying, TV programs, business card, magazines, advertisements, web pages, commixed text-picture-graphics regions etc are converted to images. These images present many challenging research issues in text extraction and recognition. Text extraction from images have many utilizable applications in document analysis, detection of vehicle license plate, maps, charts, diagrams etc. keyword predicated image search, identification of components in industrial automation, content predicated retrieval, name plates, object identification, street signs, text predicated video indexing, video content analysis, page segmentation, document retrieving, address block location etc.

The text which is present in the scene can be captured when the image or video is shot, since Scene text appears within the scene of the recording devices. Scene texts occur naturally as a component of the scene and contain consequential semantic information such as advertisements that include artistic fonts, names of streets, institutes, shops, road signs, traffic information, board signs, nameplates, pabulum containers, cloth, street signs, bill boards, banners, and text on conveyance etc. Scene text extraction can be utilized in detecting text-predicated landmarks, conveyance license detection and object identification rather than general indexing and retrieval. It is arduous to detect and extract since it may appear in a virtually illimitable number of poses, size, shapes and colors, low resolution, complex background, nonuniform lightning or blurring effects of varying lighting, complex background and transformation, unknown layout, uneven lighting, shadowing and variation in font style, size, orientation, alignment & complexity of background

2. Background

The text understanding system consists of four stages: text detection, text localization, text extraction, text recognition. This system receives an input in the form of an image or video. The images can be in color or gray scale, un-compressed or compressed, orientation. This problem can be divided into the following sub-problems: (i)detection (ii)localization (iii) extraction and enhancement, and (iv) recognition(OCR).

Text detection consists of determination of the presence of text in a given image. Text localization is the process of determining the location of text in the image. Although the concrete location of text in an image can be getting, the text still needs to be segmented from the background to make possible its recognition. In stage of text extraction where the text components in images are segmented from the background. The text region conventionally has low-resolution that's why enhancement of the components is essential. After, the extracted text images can be converted into plain text utilizing OCR technology.

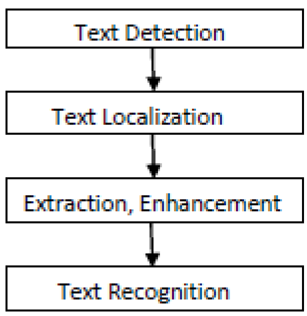

Figure 1: Text Extraction Process from Scene Images

3. Related Works

Michael R. Lyu and Jiqiang Song [1] presented a comprehensive method for multilingual video text detection, localization and extraction. This method is capable of handling multilingual text because it depends on language independent characteristics. The text detection is carried out by edge detection, local thresholding, and hysteresis edge recovery. The coarse-to-fine localization scheme is then performed to identify text regions accurately. The text extraction consists of adaptive thresholding, dam point labeling, and inward filling. This method is capable of handling multilingual video texts accurately and is robust to various background complexities and text appearances. The main drawback of this method is that it cannot detect motion texts due to the assumption of stationary text. However, this drawback can be overcome by using a signature-based text-region tracking algorithm to the multi-frame verification process. Another disadvantage of this method is that it cannot localize nonhorizontally aligned text.

Yi-Feng Pan, Xinwen Hou and Cheng-Lin Liu [2] propose a hybrid approach to detect and localize texts in the natural scene images. The subsisting methods of text detection and localization can be roughly categorized into two groups: region-based method and connected component (CC)-based method. Region-based methods uses texture analysis to detect and localize text regions.It extract local texture information to accurately segment candidate component. These methods can detect and localize texts accurately even when images are strepitous. But the speed is relatively slow and the performance is sensitive to text alignment orientation. The CC-based method filter out non-text components and localize text region accurately. Since the number of segmented candidate components is relatively minuscule, CC-based methods have lower computation cost and the located text components can be directly utilized for recognition. Although the subsisting methods have better localization performance, there still remain several problems to solve. For region-based methods, the speed is relatively slow and the performance is sensitive to text alignment orientation. On the other hand, CC-based methods cannot segment text components accurately without prior erudition of text position and scale. To overcome the above difficulties, Feng Pan, Xinwen Hou, and Cheng-Lin Liu present a hybrid approach, combining the better features of both region-based and CC-based methods to robustly detect and localize texts in natural scene images. Since local region detection can robustly detect scene texts even in noisy images, a text region detector is used to estimate the probabilities of text position and scale, which avail segment candidate text components with an efficient local binarization algorithm. To combine unary component properties and binary contextual component relationships, a conditional random field (CRF) model with supervised parameter learning is used. Determinately, text components are grouped into text lines/words robustly with an energy minimization method.

Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang, Song Gao and Zhong Zhang [3] Given a text image as input, Then the tree-structured models is used to get the character detection results, predicated on which we get the potential character locations. Based on this potential character location, a CRF model is created. Character detection scores are acclimated to define the unary cost and language model is utilized to define the pairwise cost. The final word recognition result is acquired by minimizing the cost function.

Cun-Zhao Shi, Chun-Heng Wang, Bai-Hua Xiao, and Song Gao, Jin-Long HuKuang [4] presented a scene text recognition using structure-guided character detection and linguistic knowledge. In order to model each category of characters part-based tree structure is used. The detection results are more reliable, since the character models make utilization of both the local appearance and global structure informations. For word recognition, we cumulate the detection scores and language model into the posterior probability of character sequence from the Bayesian decision view. By maximizing the character sequence posterior probability via Viterbi algorithm the final word-recognition result is obtained. An input text image is given as input. A part-based TSM is used to detect the character-specific structures, predicated on which the potential character locations are obtained. Then candidate detection scores are converted into posterior probabilities via confidence transformation. For word recognition, detection scores and language model are cumulated into the posterior probability of the character sequence from the Bayesian decision view. Bigram, trigram, and even higher order language model could be incorporated. The final word-recognition result is obtained by finding the most probable character sequence via Viterbi algorithm. The disadvantage of this method is that it cannot detect text of varying font size.

4. Proposed Work

This work introduce a new methodology called Fuzzy based image segmentation for variable font sized text which will improve text extraction from scene images. The objective of segmentation is to process the image easily. It will also reduce the computational complexity. This proposed segmentation method consist of two process: splitting and merging. Splitting is based on region based segmentation and merging is based on novel fuzzy based method. Figure 2 shows the proposed system architecture diagram. Given a text image as input. Then it undergo following processes.

4.1 Edge Detection

Edge detection along with the connected component labeling is used for segmentation in the proposed methodology, where Sobel edge detection technique is used for edge detection.

4.2 Image Dilation

Image dilation is applied to connect the broken edges. Dilation is performed prior to the fuzzy merging just to minimize the computational efforts.

4.3 Feature Extraction

The merging of character candidates relies on number of factors. Four features are extracted for the decision of joining objects as words or sentences. These features are color, height, position, and distance. The feature values that are extracted from the objects are explained as follows:

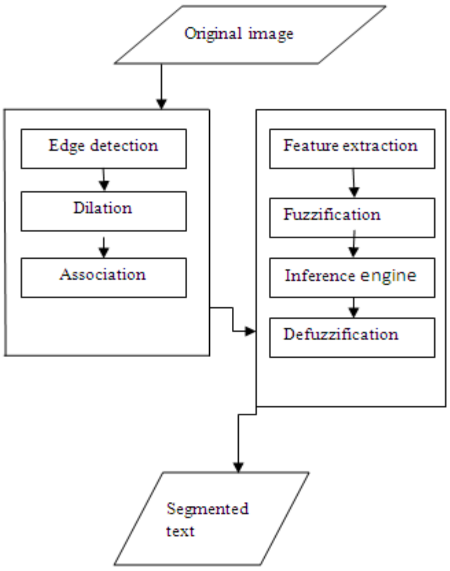

Figure 2: Architecture of proposed system

4.3.1 Color

Color is taken as the parameter to join two text objects. Color of the characters of a single word or sentence is mostly the same. If the colors of two text objects are similar, then these objects can be the candidates to merge. In order to get the degree of similarity, difference between the two colors is calculated.

4.3.2 Height

Difference of height is taken as the second input parameter for fuzzy system. Only objects with similar heights should be merged because characters of the same word or sentence have the same or similar heights. Difference of heights of two objects is measured as follows

$$\Delta Ht_{i,j} = \frac{|Ht_i - Ht_j|}{Ht_i} \qquad (1)$$

Where Ht_i and Ht_j are the heights of i^{th} and j^{th} objects respectively.

4.3.3 Position

Position of two objects should be the same for merger. This merging process is proposed for horizontal text only as most of the caption text is horizontally aligned. This can be expanded to other directions by considering positions at different angles. Consider

$$\Delta Pos_{i,j} = \frac{|pos_i - pos_j|}{Ht_i} \qquad (2)$$

Where Pos_i and Pos_j are the bottom coordinates of bounding boxes of i^{th} and j^{th} objects.

4.3.4 Distance

Characters of same work or sentences are placed closely. The distance between characters varies with the variation in font size and is highly dependent upon the heights of characters.

4.5 Fuzzification

This makes use of fuzzy rule base and linguistic rules. This module encompasses individual rule-based inference with union combination, min implication, min operator for t-norm, and max operator of s-norm.

4.6 Fuzzy Inference Engines

This makes use of fuzzy rule base and linguistic rules. This module encompasses individual rule-based inference with union combination, min implication, min operator for t-norm, and max operator of s-norm.

4.7 Defuzzification

It maps fuzzy value in to real world value.

5. Result

This paper extract and recognize text of varying font size from the scene images using fuzzy based segmentation. Recall, precision and accuracy are the main parameters used to evaluate the proposed system. The Precision, Recall and Accuracy values of extracted text from scene images by the Structure guided character detection and linguistic knowledge and proposed method are shown in Table 2.It can be observed that proposed method outperforms.

Table 2 provides Precision, Recall and Accuracy value for the Structure guided character detection method and fuzzy based method. According to this table, both the three values are PSNR improved in proposed model (Fuzzy based segmentation method) as compared with the other method.

Table 2: Comparison table

Method	*Recall*	*Precision*	*Accuracy*
Structure guided	0.8878	0.8867	88.764
Fuzzy based	0.9455	0.9487	94.6667

6. System Evaluation

This section gives the comparison graph for structure guided character detection method and Fuzzy based segmentation system. Figure 3 shows the Comparison of Recall for Structure guided character detection method and fuzzy based method in text extraction from scene images. Each figure shows the superiority of proposed system. They attain more than 0.9 in both precision and recall compared to other method. And obtained 95% accuracy for segmentation of text objects.

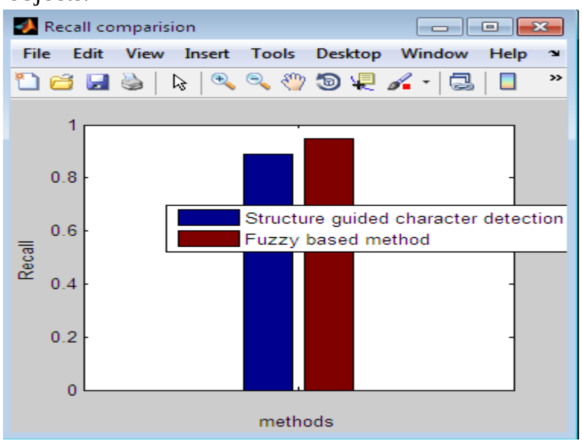

Figure 3: Comparison of Recall for Structure guided character detection method and fuzzy based method in text

extraction from scene images. In this figure, blue colored stem represent proposed model.

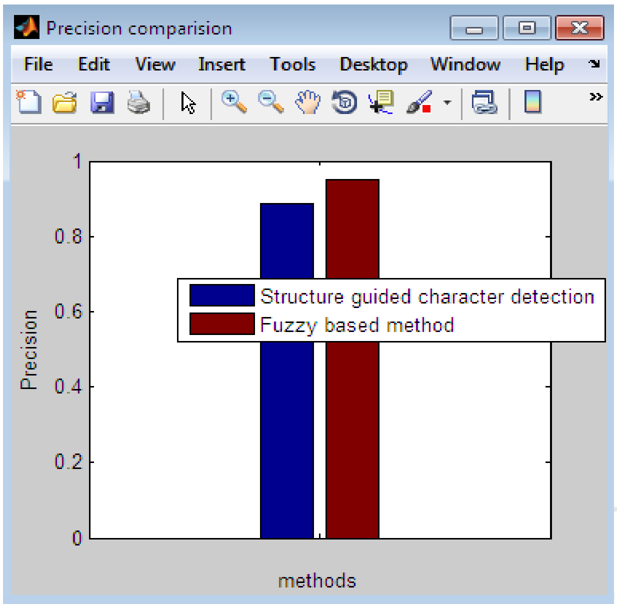

Figure 4: Comparison of Precision for Structure guided character detection method and fuzzy based method in text extraction from scene image

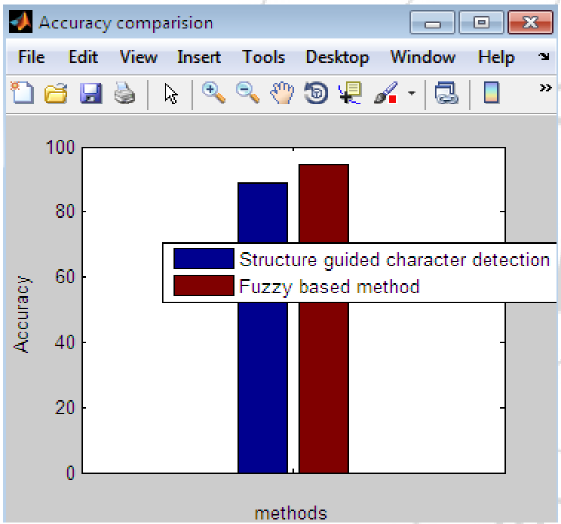

Figure 5: Comparison of Accuracy for Structure guided character detection method and fuzzy based method in text extraction from scene images.

7. Conclusion

In this paper, we propose an effective scene text-recognition method incorporating Fuzzy based segmentation. By using this method we extract text from varying font sized scene text image in accurate and faster manner. We report comparison of this proposed system and other most effective scene text recognition method [4] on Precision, Recall and Accuracy. The result show that our method outperforms.

References

[1] M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction, "*Circuits and Systems for Video Technology, IEEE*

[2] Y.-F. Pan, X. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images", *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 800-813, 2011

[3] 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

[4] C. Shi, C. Wang, B. Xiao, S. Gao, and J. Hu, "Scene text recognition using structure guided character detection and linguistic knowledge", 2013

Author Profile

Anju Maria Babu received the B Tech degrees in Computer Science and Engineering from kerala University in 2013 and pursuing, M Tech degrees in Computer Science and Engineering from Calicut University in 2015.