International Journal on Applications in Information and Communication Engineering
Volume 1: Issue 8: August 2015, pp 46-50. www.aetsjournal.com ISSN (Online) : 2394-6237

Scene Text Recognition By Character Descriptor And Structure Configuration

Kausalya Banu R. , Dr.S.Saravanakumar,

Abstract— Optical Character Recognition is the technique which is used for recognizing printed or written text by the computer. With the increase in multimedia contents, this OCR plays an important role in video sequences. Text characters and strings in natural scene can provide valuable information for many applications. Extracting text directly from natural scene images or videos is a challenging task because of diverse text patterns and variant background interferences. In general, text displayed in the videos can be classified into scene text and overlay text. Here Scene text in videos plays an important role in understanding and gives more information about it. In this paper, a new approach for recognizing the scene text is done. The goal is to increase the recognition rate of scene text and its accuracy. The proposed method is based on feature detection by Maximal Stable Extremal Regions and Harris corner. The performance of proposed method with others algorithms is compared.

Keywords— Maximal Stable Extremal Regions, Connected Components, Rectangular Bounding Box, Harris corner.

I. INTRODUCTION

Optical Character Recognition is a process that can convert text, present in digital image, to editable text. It allows a machine to recognize characters through optical mechanisms. The output of the OCR should ideally be same as input in formatting. The process involves some pre-processing of the image file and then acquisition of important knowledge about written text. That knowledge or data can be used to recognize characters. OCR is becoming an important part of modern research based computer applications. Optical Character Recognition deals with the problem of recognizing optically processed characters. Optical Character Recognition is performed off-line after the writing or printing has been completed, as opposed to on-line recognition where the computer recognizes the characters as they are drawn. Both hand printed and printed characters may be recognized, but the performance is directly dependent upon the quality of the input documents. In general, text displayed in the videos can be classified into scene text and overlay text. Scene text occurs naturally in the background as a part of the scene, such as the advertising boards, banners and so on. In contrast to that, overlay text is superimposed on the video scene and used to help viewers understanding.

A. PROBLEM DESCRIPTION

Kausalya Banu R. Anna University Regional Centre, Coimbatore, India(Email: kausalyaciet@gmail.com)

Dr.S.Saravanakumar. Associate professor,Anna University Regional Centre, Coimbatore, India (Email: sskaucbe@gmail.com)

Scene Text detection still remains as a challenging problem. This is because scene text images usually suffer from photometric degradations as well as geometrical distortions so that many algorithms faced the accuracy and/or speed (complexity) issues. Extracting scene text is a challenging task due to two main factors: Cluttered Backgrounds with noise and non-text outliers. Diverse text patterns such as character types, fonts, and sizes. The frequency of occurrence of text in natural scene is very low, and a limited number of text characters are embedded into complex non-text background outliers. Background textures, such as grid, window, and brick, even resemble text characters and strings. Although these challenging factors exist in face and car, many state-of-the-art algorithms have demonstrated effectiveness on those applications, because face and car, have relatively stable features. For example, a frontal face normally contains a mouth, a nose, two eyes, and two brows as prior knowledge. However, it is difficult to model the structure of text characters in scene images due to the lack of discriminative pixel-level appearance and Structure features from non-text background outliers. Further, text consists of different words where each word may contain different characters in various fonts, styles, and sizes, resulting in large intra-variations of text patterns.

B. PREVIOUS WORK

Chen, J.X et al [], present an approach to automatic detection and recognition of signs from natural scenes, and its application to a sign translation task. The approach embeds multi-resolution and multi-scale edge detection, adaptive searching, color analysis, and affine rectification in a hierarchical framework for sign detection, with different emphases at each phase to handle the text in different sizes, orientations, color distributions and backgrounds. They use affine rectification to recover deformation of the text regions caused by an inappropriate camera view angle. They use a local intensity normalization method to effectively handle lighting variations, followed by a Gabor transform to obtain local features, and finally a linear discriminant analysis (LDA) method for feature selection.

Coates et al [], applied methods developed in machine learning–specifically, large-scale algorithms for learning the features automatically from unlabeled data–and show that they allow us to construct highly effective classifiers for both detection and recognition to be used in a high accuracy end-to-end system. Feature learning algorithms have enjoyed a string of successes in other fields (for instance, achieving high performance in visual recognition and audio recognition).

Unfortunately, one caveat is that these systems have often been too computationally expensive, especially for application to large images

Epstein B et al [], proposed the detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired, automatic geocoding of businesses, and robotic navigation in urban environments. The stroke width transform (SWT) is a local image operator which computes per pixel the width of the most likely stroke containing the pixel. The output of the SWT is a image of size equal to the size of the input image where each element contains the width of the stroke associated with the pixel. The grouping of letters and detection of curved text lines are not possible using this method.

Kumar S et al [], extracted the textual areas of an image using globally matched wavelet filters. A clustering-based technique has been devised for estimating globally matched wavelet filters using a collection of ground truth images. They have extended text extraction scheme for the segmentation of document images into text, background, and picture components (which include graphics and continuous tone images). Multiple, two-class Fisher classifiers have been used for this purpose. They also exploit contextual information by using a Markov random field formulation-based pixel labeling scheme for refinement of the segmentation results. But automatic document analysis is not possible here and for new applications they are not applicable.

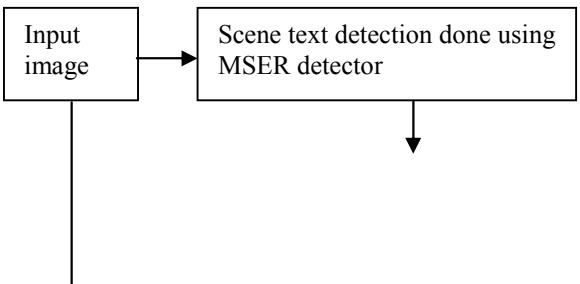
II. PROPOSED WORK

Proposed work uses the connected components method for finding the text region. Normally the text present inside the image or video frame will have same color regions and non-text regions are detected while going for color based region detection. Hence MSER region detection is done so that only features alone are detected. The block diagram for scene text recognition using character descriptor and structure configuration is explained in Figure 1.

A. SCENE TEXT DETECTION

For the generation of candidates, we extract CCs in images and partition the extracted CCs into clusters, where our clustering algorithm is based on an adjacency relation classifier. In this section, we first explain our CC extraction method. Then, we will explain our approaches

(i) to build training samples,
(ii) to train the classifier, and
(iii) to use that classifier in our CC clustering method.



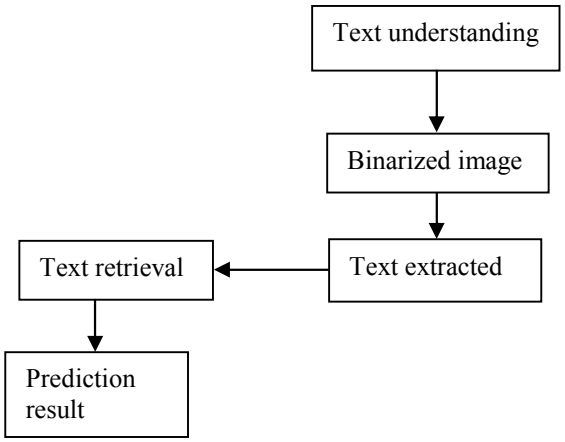


Figure 1.Block diagram of scene text recognition by character descriptor and structure configuration

B. CC EXTRACTION BY MSER DETECTOR

Among a number of CC extraction methods, MSER algorithm is used because it shows good performance with a small computation cost. This algorithm can be considered as a process to find local binarization results that are stable over a range of thresholds, and this property allows us to find most of the text components.

ALGORITHM:

1.first the input image is taken and converted into gray image.

2. Then this gray image is applied by MSER region detection method where only features are detected from stroke components. The below Equation 1 is used for finding the MSER regions.

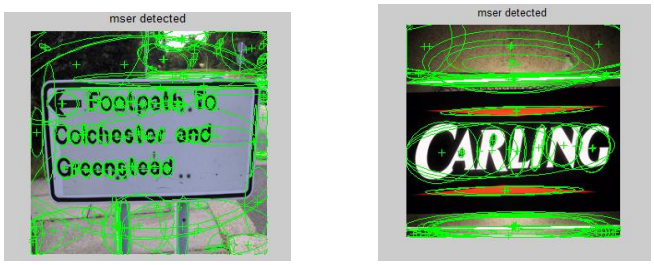
$$Qi = |qi - \Delta \backslash qi + \Delta| / |qi| \quad (1)$$

Where Q (D, D is a region, such that either for all $P \in q$ $q \in \delta q$: $i(p) \succ i(q)$;

Max intensity regions $q1, q2, \ldots qi-1, qi$ is sequence of nested extremal regions $(qi \in qi+1)$. The input images are shown in Figure 2(a) and (b). Respective MSER regions are shown in Figure 3(a) and (b).

(a) (b)

Figure 2. Examples of scene texts



(a) (b)

Figure 3.MSER detected output from gray image

C. TEXT UNDERSTANDING

Text understanding is to find the text information from the natural scene image or still images. Hence Harris corner detector is used to find the corners and junctions of the scene text as well as the other non-text regions. This is expressed in the below Equation 2.

$$E(u,v) = \sum_{x,y} w(x,y)\,[\,I(x+u,y+v) - I(x+y)]^2 \quad (2)$$

Where $w(x,y)$ is the window function

$[\,I(x+u,y+v)]$ is the shifted intensity

$I(x,y)$ is the intensity

The window function taken here is rectangular window. The Harris corner detection output is shown in Figure 4. (a) and (b)

(a) (b)

Figure 4. Harris corner detected outputs

III. SCENE TEXT EXTRACTION

A. BINARIZATION

Now take the input image and convert into gray image. This gray image is binarized by finding the edges of the image .Common edge detector is used and binarized. The image is binarized as white and black colors, and then the sample pixels are alone taken. These sample pixels are the ones which are non-edge pixels which are negative to binarized image. The output of these are given in Figure 5 (a) and (b)

(a) (b)

Figure 5.Binarized image and Sample pixels

B. TEXT LOCALIZATION REGION

The text region is localized by using the detection of features of Harris detection and MSER detection. The sample pixels which are taken are used for rectangular bounding box. By using this rectangular boundary condition, the exact text region is found and can be detected. Hence the localized text region shown in Figure 6.

Figure 6.Text localized region

Connected components:

The Binarized image is taken which is the edge detection. Canny edge detection is done here. Then the connected components technique is used.

Rectangular bounding box:

The rectangular bounding box is applied to the connected components result.

C. ADABOOST LEARNING

Sample images are taken and train the adaboost classifier. This classifier will check whether the given $(c_i, c_j) \in C \times C \,(i = j)$ is adjacent or not. 6-dimensional feature vectors consisting of five geometrical features and one color-based feature is used. All of geometric features are designed to be invariant to the scale of an input image. Adaboost clustering algorithm yields a function $\varphi: C \times C \rightarrow \_R$ and we use this function in binary decisions $\varphi(ci, cj) > \tau 1 \Longleftrightarrow ci \sim cj$ with a threshold $\tau 1$.

By using the features of the MSER region detector which acts as the binarization outputs, however input image is binarized by estimating the text and background colors. Considering the average color of CCs as the text color and the average color of the entire block is consider as background color. Thus output for text extraction is shown in Figure 7 (a) and (b).

Figure 7. Extracted output from binary image

IV. TEXT/NONTEXT CLASSSIFICATION

In order to get final results just like in Figure 7(a) and (b), a text/non text classifier that reject non blocks from the normalized block is developed. In the classification, sophisticated techniques such as cascade structures are adopted, since the number of samples to be classified is usually small. One possible approach to this problem is to split the normalized images into patches covering one of the letters and develop a character/non-character classifier. The binarized image is taken and applied to the classifier. That is, trying to develop MSER-based CC extractors

Yielding individual characters (i.e., high precision and high recall). On the other hand, we mainly focus on retrieving the text components as much as possible. As a result, redundant and noisy ccs could be involved in finding clusters. The output for classifier is shown in Figure 8 (a) and (b).

(a) (b)

Figure 8.Classifier output

V. DISCUSSION

A text/non-text classifier based on normalized gray-scale images (without binarizaton) is developed, because gray scale images seem to be more informative. To be precise, adoption of the AdaBoost learning method is made and gradient features. However, experiments have shown that this approach yielded almost the same performance, with a considerable amount of overhead in training. Usage of both classifiers in a row (neural network with binary images and Adaboost with gray images) is tried and, however, noticeable gains are found.

VI. EXPERIMENTAL RESULTS

In experiments, the feature detector called MSER and color based clustering algorithm called boundary clustering algorithm is compared. Performance Comparison between MSER Detector and Boundary Clustering Algorithm the comparison between maximal stable extremal region and boundary clustering algorithm is done in this section. Here the boundary clustering algorithm is based on color. It segment the regions based on colors and sometimes it detect the non-text regions since colors of non-text may be as same as the text colors, where as in MSER detection they use the features alone, hence the detection of only text region is easier and faster . The Figure 8 shows the output of boundary clusteringalgorithm.

Figure 8. Boundary Clustering Algorithm Output

Figure 9(a) and (b) shows the region detected based on boundary clustering algorithm. Figure 10 shows the MSER detected output which takes only the features of the scene image hence the text region localization becomes easier and faster.

Figure 10. MSER detected output

VII. CONCLUSION

The scene text which is present in the background of the video frame gives the information for better understanding to the viewers, additional information is given. This helps in Automatic bank cheque processing, Vehicle license recognition, Recognition of compact disk labels, Image indexing and retrieval. Semantic video indexing, summarization, video surveillance and security, multi-lingual video information access. The text is detected by first features detection by MSER detector and then given to the connected components method. The future work is to recognize the scene text using the template matching technique and to improve the accuracy rate of text detection and to add lexicon analysis to extend our system to word-level recognition.

REFERENCES

[1] Beaufort R and Mancas-Thillou C (2007), weighted finite-state framework for correcting errors in natural scene OCR,' in Proc. 9th Int. Conf. Document Analysis Recognition.pp. 889–893

[2] Campos T. de, Babu B., and Varma M. (2009), 'Character recognition in natural images,' in Proc. VISAPP.

[3] Chen, J.X. Yang, Zhang.J, and Waibel.A (2004), 'Automatic detection and recognition of signs from natural scenes,' IEEE Transactions on Image Processing, Vol. 13, No. 1, pp. 87–99.

[4] Chucai Yi and Yingli Tian, (2014) 'Scene Text Recognition by Character Descriptor and Structure Configuration' IEEE transactions on image processing, Vol. 23, No. 7.

[5] Coates et al A. (2011), 'Text detection and character recognition in scene images with unsupervised feature learning,' in Proc. ICDAR, pp. 440–445.

[6] Dalal. N and Triggs. B, (2005) 'Histograms of oriented gradients for human detection,' in Proc. IEEE Conference Compute. Vis. Pattern Recognition., pp. 886–893.

[7] Epstein B, Ofek E., and Wexler Y. (2010), 'Detecting text in natural scenes with stroke width transform, ' in Proc. CVPR, pp. 2963–2970.

[8] Felzenszwalb P.F, Girshick R. B., McAlester D., and Ramanan D (2010). 'Object detection with discriminatively trained part-based models,' IEEE Trans. Pattern Analysis. Mach. Intell., Vol. 32, No. 9, pp. 1627–1645.

[9] Hyung II Koo, and Duck Hoon Kim (2013), 'Scene Text Detection via Connected Component Clustering and Non text Filtering:' IEEE transactions on image processing, Vol. 22, No. 6.

[10] Kumar S, Gupta R, Khanna N, Chaudhury S, and Johsi S. D. (2007),'Text extraction and document image segmentation using matched wavelets and MRF model,' IEEE Transactions Image Processing., Vol. 16, No. 8, pp. 2117–2128.

[11] Liu C, Wang C., and Dai R. (2005), 'Text detection in images based on unsupervised classification of edge-based features,' in Proc. Int. Conf. Document Analysis and Recognition, Vol. 2, pp. 610–614.

[12] Neumann L and Matas J (2012), 'Real-time scene text localization and recognition,' in Proc. IEEE Conf. Compute. Vis. Pattern Recognition. pp. 3538–3545.

[13] Shi C, Wang C, Xiao B, Zhang Y, Gao S, and Zhang Z, (2013) 'Scene text recognition using part-based tree-structured character detection,' in Proc. CVPR, Vol. 3, No. 5 pp. 2961–2968.

[14] Weinman J. J., Learned-Miller E., and Hanson A. R (2009). 'Scene text recognition using similarity and a lexicon with sparse belief propagation,' IEEE