International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Robust Text Detection and Extraction in Natural Scene Images using Conditional Random Field Model and OCR

Pratik Yadav †* and Prabhudev Irabashetti†

†Computer Engineering, Pune University, Ahmednagar, Maharashtra, India

Accepted 12 Dec 2015, Available online 14 Dec 2015, **Vol.5, No.6 (Dec 2015)**

Abstract

In Natural Scene Image, Text detection is important tasks which are used for many content based image analysis. A maximally stable external region based method is used for scene detection .This MSER based method has stages character candidate extraction, text candidate construction, text candidate elimination & text candidate classification. In this systems the method are not focus on how to detect highly blurred text in low resolution natural scene images. The current technology not any text extraction method provided. In proposed system by using Connected Component analysis a Conditional Random field (CRF) model is used to assign candidate component as one of the two classes (text& Non Text) by Considering both unary component properties and binary contextual component relationship. For this purpose we are using connected component analysis method. The proposed system also performs a text extraction using OCR.

Keywords: *Maximally stable Extremal Region, Text candidates Construction, Text Candidate Elimination, text Candidate Classification, Connected Component Analysis, Optical Character recognition.*

1. Introduction

Text in natural scene images contains valuable information this text in the Natural scene image contains sign or letters which include Building Name, Street Name, Company Name, Commercial Advertisement, and Announcement. This Information can be Useful for many applications such as image retrieval and Video retrieval, content based image and video applications, such as Content-based web image retrieval and video information. As Text information is a main component of scene images, it provides an important clue for scene understanding. Given the vast number of text-based search engines, retrieving image using the embedded text offers an efficient supplement to the visual search systems. Also now a day wide use of smart phones and rapid development of the mobile internet, it has become a living style for people to capture information by using of Cameras embedded in mobile terminals. The text base image Natural scene image is captured in camera. Recognition of text in natural scene images is becoming a prominent research area due to the widespread availability of imaging devices in low-cost consumer products like mobile phones. Maximally Stable Extremal Regions (MSERs) based methods are used for this purpose (Xu-Cheng Yin *et al*, 2014). An extremal region is a connected component of an image whose pixels have either higher or lower intensity than its outer boundary pixels. Extremal regions of the whole image are extracted as a rooted tree. An extremal region is in fact a set of pixels After Detection of this text the next parts comes how to extract this text. Automatic extracting text from scene text an image after text detection is still a problem.

The main difficulty while extracting the text is high variability of text appearance, for instance, variation in color, changing font style, size and different languages. Apart from that the problems are complex background, uneven illumination and blur make the problem of scene text extraction much more challenging. Researchers have reported many methods to solve this problem, and some can give good results. Form those methods we use Text Extraction optical character recognition method is used. We use a character classifier to estimate the posterior probability of text candidate corresponding to non-text and eliminate text candidates with high non text probability, which helps to build a more powerful Text classifier. As a result of advanced background, and variations of font, size, color and orientation, text in natural scene pictures should be robustly detected before being recognized and retrieved.

2. Related Work

Major advantage of MSER-situated approaches over average method is MSERs algorithm is able to become

*Corresponding author: **Pratik Yadav**

aware of most characters even when the picture is in low quality (low decision, robust noises, low contrast, etc.). However, there is difficulty for MSERs algorithm with repeating with each different. Repeating MSERs are not easy for the latter character candidates grouping algorithm, as a consequence many of the repeating MSERs, apart from the MSERs that without doubt correspond to character, have got to be eliminated before being fed to the person a grouping algorithm. MSER (maximal stable extremal region) method includes the following stages (Pratik Yadav *et al*, 2014)

2.1 Character candidates extraction

In Character Candidate extraction pruning algorithms is used for character identification as each character is unique and different from other character. Here Parent–child relationship is from to calculate repeated text value Fast and accurate pruning algorithm that enables us to detect characters even when the image is in low quality. This pruning algorithm has two a methods (i.e. two algorithm).First when MSER has only one type of text (one child) and when MSER has different type of text (a to z more than one child).

2.1.1 Linear Reduction

The linear reduction algorithm is used in situations where MSERs has only one child. The algorithm chooses from parent and child the one with the minimum variation and discards the other. This procedure is applied across the whole tree recursively. As shown in fig 2.1.1 in given a MSERs tree, the procedure returns the root of the processed tree whose linear segments are reduced. The procedure works as follows. Given a node *t*, the procedure checks the number of children of *t*; if *t* has no children, return *t* immediately; if *t* has only one child, get the root *c* of child tree by first applying the linear reduction procedure to the child tree.

```
procedure LINEAR-REDUCTION(T)
    if nchildren[T] = 0 then
        return T
    else if nchildren[T] = 1 then
        c ← LINEAR-REDUCTION(child[T])
        if var[T] ≤ var[c] then
            link-children(T, children[c])
            return T
        else
            return c
        end if
    else                                  ▷ nchildren[T] ≥ 2
        for each c ∈ children[T] do
            link-children(T, LINEAR-REDUCTION(c))
        end for
        return T
    end if
end procedure
```

Fig 2.1.1 Linear Reduction Algorithm

2.1.2 Tree Accumulation

The tree accumulation algorithm is used when MSERs has more than one children. As shown in fig 2.1.2 in short, given a MSERs tree, the procedure returns a set of disconnected nodes.

```
procedure TREE-ACCUMULATION(T)
    if nchildren[T] ≥ 2 then
        C ← ∅
        for each c ∈ children[T] do
            C ← C ∪ TREE-ACCUMULATION(c)
        end for
        if var[T] ≤ min-var[C] then
            discard-children(T)
            return T
        else
            return C
        end if
    else                                  ▷ nchildren[T] = 0
        return T
    end if
end procedure
```

Fig 2.1.2 Tree Accumulation Algorithm

2.2 Text candidate construction

A single-link clustering algorithm is used to calculate Distance weights and clustering threshold values of the character. In the case of single-link clustering, the two clusters whose two closest members have the smallest distance are merged in each step. A distance threshold can be specified such that the clustering process is terminated when the distance between nearest Clusters exceed the threshold. The resulting clusters of single-link algorithm form a hierarchical cluster tree or cluster forest if termination threshold is specified Clustered into text candidates by the single-link clustering algorithm using the learned parameters. To learn the distance function the Strategy follows as learn the distance function by minimizing distance between two points. One the strategy of metric learning is to learn the distance function by minimizing distance between point pairs in C while maximizing distance between point pairs in M, where C specifies pairs of points in different clusters and M specifies pairs of points in the same cluster. In single-link clustering, clusters are formed by merging smaller clusters; the final resulting clusters will form a binary Cluster tree, in which non-singleton clusters have exactly two direct sub clusters. By minimizing regularized variations Character Candidate are extracted with the help of MSER algorithm.

2.3 Text candidates elimination

In this Step the probabilities of Text Candidates related with text and non text are calculated and text with high non text probabilities is removed. As it is hard to train an effective text classifier using such an unbalanced

database, most of the non-text candidates need to be removed before training the classifier. In my System use a character classifier to estimate the posterior probabilities of text candidates corresponding to non-text and remove text candidates with high non-text probabilities.

$$P(O(m, n; p)) = P(O(m, n; p)|\text{text})P(\text{text})+P(O(m, n; p)|\text{non-text})P(\text{non-text}).$$

The candidate region is rejected if P (non-text| $O(m, n; p)) \geq \varepsilon$, where ε is the threshold.

2.4 Text candidates classification

Text candidates corresponding to true texts are identified by the text classifier. An AdaBoost classifier is trained to decide whether a text candidate corresponding to the true text or not.

3. Proposed Work

The steps of Character candidate extraction, text candidate construction, text candidate classification are followed as same in proposed system. In text candidates classification in my system conditional Random Field Model is used instead of Adaboost classifier. In maximally stable extremal region (MSER) method not any text extraction method is given in my system I use Optical Character Extraction method is used. It can do single character to character text extraction which has been detected.

3.1 Text candidate classification

Text candidates corresponding to true texts are identified by the text classifier. Instead of AdaBoost classifier in proposed method we are using connected component analysis.

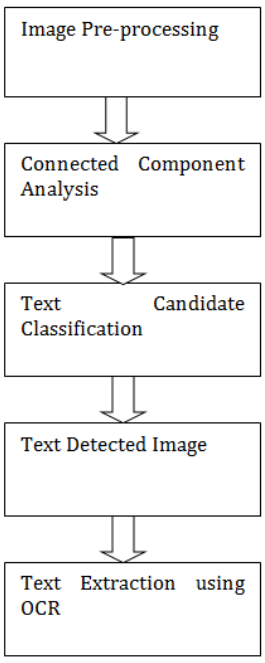

Fig 3.1 Proposed System

3.1.1 Connected Component Analysis

Connected-component labeling (alternatively connected-component analysis, blob extraction, region labeling, blob discovery, or region extraction) is an algorithmic application of graph theory, where subsets of connected components are uniquely labeled based on a given heuristic. Connected-component labeling is not to be confused with segmentation. Connected-component labeling is used in computer vision to detect connected regions in binary digital images, although color images and data with higher dimensionality can also be processed. When integrated into an image recognition system or human-computer interaction interface, connected component labeling can operate on a variety of information. Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered, and tracked. Connectivity checks are carried out by checking neighbor pixels' labels (neighbor elements whose labels are not assigned yet are ignored), or say, the North-East, the North, the North-West and the West of the current pixel (assuming 8-connectivity). 4-connectivity uses only North and West neighbors of the current pixel. The following conditions are checked to determine the value of the label to be assigned to the current pixel (4-connectivity is assumed).

1. Iterate through each element of the data by column, then by row (Raster Scanning)
2. If the element is not the background
   1) Get the neighboring elements of the current element
   2) If there are no neighbors, uniquely label the current element and continue
   3) Otherwise, find the neighbor with the smallest label and assign it to the current element
   4) Store the equivalence between neighboring labels

On the second pass

1. Iterate through each element of the data by column, then by row
2. If the element is not the background
   1) Relabeled the element with the lowest equivalent label

3.2 Optical Character Recognition (OCR)

Optical character recognition (OCR) is the electronic conversion of images of typed, handwritten or printed text into machine-encoded text. It is used as a form of data entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation. It is a very popular method of digitizing printed texts so that it can be electronically edited, searched, stored more

compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data and text mining. Optical Character Recognition is a field of research in pattern recognition, artificial intelligence and computer vision.

3.3 Output & Results

As shown in fig3.3.1 the screenshot of proposed system is given. The text in Natural scene image is first detected by using connected component analysis .Then character recognition is done in image .In another fig graph of time complexity of maximally stable extremal region using AdaBoost classifier and by using Conditional Random Field Model. From the graph (fig 3.3.2). It is concluded that time required for text detection by using Conditional Random filed model is less as compare to AdaBoost classifier Method.

Fig 3.3.1 Output of Proposed System

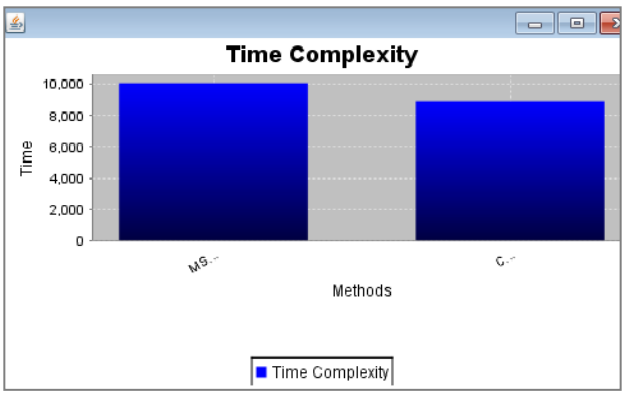

Fig 3.3.2 Graph of Time Complexity

Conclusion and Future work

By Using Maximally extremal stable region method and using conditional random field model text can be detected in Natural scene image. It is fast and accurate method than previous MSER method.

Also the current technology focuses on Optical Character text extraction Method. In the Future work we use a dataset of images to recognize text rather in future we can do it runtime i.e. images taken from webcam containing text we can use to recognize also in addition to that we can develop the same for portable devices like a camera, mobiles it will give more real time results. Also we can improvise it by doing text recognition from videos containing text.

References

Xu-Cheng Yin, *Member, IEEE,* Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao (May 2014), Robust Text Detection in Natural Scene Images, IEEE transactions on pattern analysis and machine intelligence, Vol. 36, No. 5

Pratik Yadav (November 2014), Robust Text Detection and Extraction in Natural Scene Images Using Conditional Random Field Model and OCR, IJRIT International Journal of Research in Information Technology, Volume 2, Issue 11, Pg. 213-216.

Pratik Yadav and Prabhudev Irabashetti (13th 14th March 2015),Robust Text Detection and Extraction in Natural Scene Images Using Conditional Random Field Model and OCR, cPGCON_2015, Paper Id 920, MET's Institute of Engineering Nasik

Y. Zhong, H. Zhang, and A. Jain (May 2014), Automatic caption localization in compressed video, IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 4, pp. 385–392, Apr. 2000. 982 IEEE transactions on pattern analysis and machine intelligence, Vol. 36, No. 5

J. Weinman, E. Learned- Miller , and A. Hanson (Oct. 2009), Scene text recognition using similarity and a lexicon with sparse belief propagation, IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1733–1746.

X. Chen and A. Yuille (2004), Detecting and reading text in natural scenes, in Proc. IEEE Conf. CVPR, vol. 2. Washington, DC, USA, pp. 366–373.

J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, Adaboost for text detection in natural scene, in Proc. ICDAR, Beijing, China, 2011, pp. 429–434.

Authors Profile

Pratik Ashok Yadav is Pursuing Master of Engineering in Computer Engineering from Vishwabharti Academy College of Engineering, Ahmednagar. He has three Year of teaching experience as Lecturer in Polytechnic College.

Irabashetti Prabhudev is working as Assistant professor in Vishwabharti Academy College of Engineering, Ahmednagar in Computer Engineering Department. He Has Completed master of Engineering in Computer Engineering