Novateur Publication's
International Journal of Innovation in Engineering, Research and Technology [IJIERT]
ICITDCEME'15 Conference Proceedings
ISSN No - 2394-3696

ROBUST TEXT DETECTION AND EXTRACTION IN NATURAL SCENE IMAGES USING CONDITIONAL RANDOM FIELD MODEL AND OCR

Pratik Yadav
ME II Computer , Department of Computer Engineering., Pune University /
Vishvabahrti College Of Engineering, Ahmednagar,India pratik21287@gmail.com

Prabhudev Irabashetti
Assistant Professor, Department of Computer Engineering, Pune University
/ Vishvabahrti College Of Engineering, Ahmednagar, India irabashetti@gmail.com

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 includes stages character candidate extraction, text candidate construction, text candidate elimination & text candidate classification. Main limitations of this method are how to detect highly blurred text in low resolution natural scene images. The current technology not focuses on any text extraction method. In proposed system 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.

INTRODUCTION

TEXT in natural scene images contains valuable information .The text in the Natural scene image contains sign or letters which include Building Name, Street Name, Company Name, Commercial Advertisement, Announcement .This Information can be Useful for content based image and video applications, such as Content-based web image retrieval, video information retrieval. Text information as a main component of scene images, it usually 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 [4]. Also with the 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 camera image contains text base image .There for Text Detection in Natural Scene Image is important task .Maximally Stable External Regions (MSERs) based methods are used for this purpose [1] .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 fast and accurate MSERs pruning algorithm that enables us to detect most characters even when the image is in low quality [13]. Second, we use self-training distance metric learning algorithm that can learn distance weights and clustering threshold simultaneously; text candidates are constructed by clustering character candidates by the single link algorithm using the learned parameters. Third, we put forward to 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 [14]. 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.

RELATED WORK

major advantage of MSER-situated approaches over average related aspect centered ways roots in the usage of the MSERs algorithm for extraction the MSERs algorithm is able to become aware of most characters even when the picture is in low quality (low decision, robust noises, low contrast, etc.). However, one severe but not so apparent pitfall of the MSERs algorithm is that most of the detected MSERs are actually 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 [14]. MSER (maximal stable extermal region) method includes the following stages:

1. CHARACTER CANDIDATE EXTRACTION - In Character Candidate extraction pruning algorithms is used for character identification as each character is unique and different from other character. Parent –child relationship is from to calculate repeated text value Fast and accurate MSERs 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).

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. In short, 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-Linear Reduction Algorithm

1.2 TREE ACCUMULATION:

The tree accumulation algorithm is used when MSERs has more than one children. 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-Tree Accumulation Algorithm

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.
Given the top level cluster set {Ck}mk=1, we randomly initialize feature weights w and set C and M as

$$C = \{(\hat{u}k, \hat{v}k) = \text{argmin}_{u \in Ck, v \in C-k} d(u, v; w)\}mk=1,$$

$$M = \{(u*k , vk) = \text{argmin}_{u \in C1k, v \in C2k} d(u, v; w)\}mk=1,$$

Where C−k is the set of points excluding points in Ck; C1k and C2k are direct subclusters of Ck. Suppose _ is specified as the single-link clustering termination threshold. By the definition of single-link clustering, we must have

$$d(u, v; w) > \_ \text{ for all } (u, v) \in C,$$

$$d(u, v; w) \leq \_ \text{ for all } (u, v) \in M.$$

The above equations show that C and M can be corresponded as the positive and negative sample set of a classification problem, such that feature weights and threshold can be learned by minimizing the classification error.

3. TEXT CANDIDATE'S 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. We propose to 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. The following features are used to train the character classifier: text region height, width and aspect ratio, smoothness (defined as the average difference of adjacent boundary pixels' gradient directions) and stroke width features (including mean and variance of character stroke widths). Characters with small aspect ratios such as "i","j" and "l" are labeled as negative samples, as it is very uncommon that some words comprise many small aspect ratio characters. Given a text candidate T, let O(m, n; p) be the observation that there are m ($m \in N, m \geq 2$) character candidates in T, where n ($n \in N$, $n \leq m$) candidates are classified as non-characters by a character classifier of accuracy (on the validation set) p ($0 < p < 1$). The probabilities of the observation conditioning on T corresponding to text and non-text are $P(O(m, n; p)|\text{text}) = p^{m-n}(1-p)^n$ and $P(O(m, n; p)|\text{non-text}) = (1-p)^{m-n}p^n$ respectively.
Let P(text) and P(non-text) be the prior probability of T corresponding to text and non-text. By applying Bayes' rule,the posterior probability of T corresponding to non-text given the observation is P(non-text|O(m, n; p))= P(O(m, n; p)|non-text)P(non-text)P(O(m, n; p)), where P(O(m, n; p)) is the probability of the observation

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

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

4. TEXT CANDIDATES CLASSIFICATION- Text candidates corresponding to true texts are identified by the text classifier. An Ada Boost classifier is trained to decide whether an text candidate corresponding to the true text or not [8].

PROPOSED WORK

All the three step of Character candidate extraction, text candidate construction, text candidate classification is followed as same in proposed system. In text candidates classification conditional Random Field Model is used instead of Ada boost classifier.

TEXT CANDIDATE CLASSIFICATION: Text candidates corresponding to true texts are identified by the text classifier. Instead of Ada Boost classifier in proposed method we are using connected component analysis.

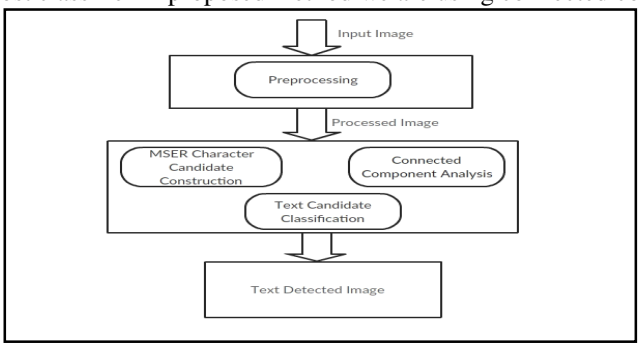

Fig: Flow of System

CONNECTED COMPONENT ANALYSIS

Connected-component labeling (alternatively connected-component analysis, blob extraction, region labelling, 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 labelling 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 labelling 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 neighbouring elements of the current element
    2. If there are no neighbours, uniquely label the current element and continue
    3. Otherwise, find the neighbour with the smallest label and assign it to the current element
    4. Store the equivalence between neighbouring 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. Relabelled the element with the lowest equivalent label

OPTICAL CHARACTER RECOGNITION (OCR)-

Optical character recognition (OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text. It is widely 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 common 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. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

RESULTS

Following figure shows the graph of time complexity of maximally stable external region using Ada Boost classifier and by using Conditional Random Field Model. From the graph it is clearly concluded that that required for text detection by using Conditional Random filed model is less as compare to Ada Boost classifier Method.

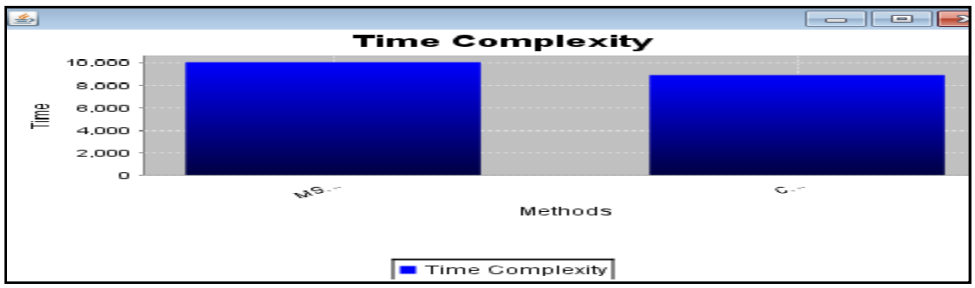

Fig-Graph of Time Complexity

CONCLUSION AND FUTURE WORK

We have a tendency to reward a hybrid method to localize scene texts by mistreatment group action space info right into a potent CC-established procedure. The binary discourse side relationships, what is more to the single part

homes, square measure constitutional during a CRF model, whose parameters square measure jointly optimized by mistreatment supervised learning. Our experimental outcome valid that the proposed methodology is mighty in free scene text localization in a range of respects: first vicinity-established experience may well be terribly helpful for text part segmentation and analysis.

In this our project 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

[1] Xu-Cheng Yin, Member, IEEE, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, "Robust Text Detection in Natural Scene Images", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 5, MAY 2014

[2]Y. Zhong, H. Zhang, and A. Jain, "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, MAY 2014

[3] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. Image Process., vol. 9, no. 1, pp. 147–156, Jan. 2000.

[4] J. Weinman, E. Learned- Miller , and A. Hanson, "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, Oct. 2009.

[5] X.-C. Yin, H .-W. Hao, J. Sun, and S. Naoi, "Robust vanishing point detection for mobile cam-based documents," in Proc. ICDAR, Beijing, China, 2011, pp. 136–140.

[6] P. Shivakumara,Q. P. Trung, and C. L. Tan, "A laplacian approach to multi-oriented text detection in video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412–419, Feb. 2011.

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

[8] 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.

[9] K. Kim, K. Jung, and J. Kim, "Texture-base approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[10] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. CVPR, San Francisco, CA, USA, 2010, pp. 2963–2970.

[11] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sept. 2011.

[12] C. Mancas-Thillou and B. Gosselin, "Color text extraction with selective metric-based clustering," Comput. Vis. Image Und., vol. 107, no. 1–2, pp. 97–107, 2007

[13] Pratik Yadav, "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, November 2014, Pg. 213-216.

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