Text region Identification from Natural scene Images using Semi-supervised Technique

A thesis submitted in partial fulfilment of the requirement for the

Degree of Master of Computer Technology

of

Jadavpur University

By

Shiplu Das

Registration No. : 137118 of 2016-17

Examination Roll No. : M6TCT19019

Under the Guidance of

Dr. Nibaran Das

Associate Professor

Department of Computer Science and Engineering

Jadavpur University, Kolkata- 700032

India

May, 2019

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled **"Text region Identification from Natural scene Images using Semi supervised Technique"** has been satisfactorily completed by Shiplu Das (University Registration No. : 137118 of 2016-17, Examination Roll No. : M6TCT19019. It is a piece of work carried out under my guidance and supervision and be accepted in partial fulfilment of the requirement for the Degree of Master of Computer Application, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, Kolkata.

_____

Dr. Nibaran Das (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-700032

Countersigned

_____

Prof. Mahantapas Kundu
Head, Department of Computer Science and Engineering
Jadavpur University, Kolkata-700032

_____

Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology
Jadavpur University, Kolkata-700032

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **"Text region Identification from Natural scene Images using Semi supervised Technique is** a bonafide record of work carried out by Shiplu Das in partial fulfilment of the requirements for the award of the degree of Master of Computer Technology in the Department of Computer Science and Engineering, Jadavpur University during the period of January 2019 to May 2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

_____

Signature of Examiner

Date:

_____

Signature of Supervisor

Date:

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis entitled **"Text region Identification from Natural scene Images using Semi supervised Technique"** contains literature survey and original research work by the undersigned candidate, as part of her Degree of Master of Computer Technology.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name : Shiplu Das
University Registration No. : 137118 of 2016-17
Examination Roll No. : M6TCT19019

Thesis Title: **Text region Identification from Natural scene Images using Semi supervised Technique**

Signature:

_____

Date:

_____

ACKNOWLEDGEMENT

First and foremost, I would like to start by thanking God Almighty for showering me with the strength, knowledge and potential to embark on this wonderful journey and to persevere and complete the embodied research work satisfactorily.

I am pleased to express my deepest gratitude to my supervisor, **Dr. Nibaran Das**, Department of Computer Science and Engineering, Jadavpur University, Kolkata for his invaluable guidance, constant encouragement and inspiration during the period of my dissertation. I am highly indebted to **Jadavpur University** for providing me the opportunity and the required infrastructure to carry on my thesis.

I would also like to acknowledge all my lab mates especially **Ms. Kalpita Dutta** for helping me and motivating me constantly and spending such wonderful six months journey.

I am thankful to all the teaching and non-teaching staff whose helping hands have smoothed my journey through the period of my research.

Last but not the least, I would like to thank my family members, classmates, seniors and friends for giving me constant encouragement and mental support throughout my work.

Shiplu Das
University Registration No. : 137118 of 2016-17
Examination Roll No. : M6TCT19019
Master of Computer Technology
Department of Computer Science and Engineering
Jadavpur University

TABLE OF CONTENTS

1 **INTRODUCTION** 7

1.1 OVERVIEW OF TEXT DETECTION 9

1.2 RELATED REVIEW 101

1.3 TEXT DETECTION METHOD 13

1.4 BRIEF DESCRIPTION OF SOME POPULAR DATASETS 175

1.5 THESIS OBJECTIVE 14

2 METHODOLOGY 19

2.1 BRIEF INTRODUCTION OF MSER 19

2.2 MATHEMATICALLY DETAILS 20

2.3 IMPLEMENTATION OF MSER IN OUR WORK 21

3 EXPERIMENTAL RESULT AND ANALYSIS 23

3.1 DATASET COLLECTION 23

3.2 DATASET ANNOTATION 23

3.3 PROCEDURE 25

3.4 RESULT AND DISCUSSION 26

3.5 EXPERIMENTAL RESULT AND ANALYSIS 27

3.6 ANALYSIS 30

4 CONCLUSION AND FUTURE WORK 30

5 REFERENCES 32

CHAPTER ONE

INTRODUCTION

In today's world visual detection and recognition of text from an image is very demandable due to its application in content-based image retrieval, robotic navigation, automatic car number plate recognition, extracting information from passport or business card or bank statement, making editable the text of an image, text translation on mobile phones, etc. Text detection in Natural images has been increasingly popular because text images represent many technical and digital information. It is used in the different fields of computer vision and Pattern Recognition. One of the major medium of communications is Text. It can be traced in scattered form throughout the images, and it is available in different fonts, colors, and shapes. Text data present in images consist of useful information for automatic annotation and indexing. Whenever Extraction of this information is done, the process involves detection, localization, tracking, enhancement, and recognition of the text found in a given image. Text detection tecnique is used for text localization and recognition in an natuaral images. It also marks boundaries of the text areas. It is used to process images taken by a digital camera or a mobile phone and to read the content of each text in a natural image area into a digital format.

Text detection tends to be quite challenging due to the uncontrolled image capturing process, variations of font style, size, color, orientation, contrast, context, geometric and photometric distortion of text in scenes, etc. In addition to that, text-like background objects, such as bricks, windows, and leaves often lead to many false positive in text detection. The major challenges [7] can be categorized into two types:

VARIETY OF SCENE TEXT:

Natural scene images usually contain entirely different fonts, colors, scales of texts instead of regular font, single color, and consistent size of texts.

INTERFERENCE FACTORS:

The background elements of the images like bricks, grasses, storefronts, street signs, and different types of signs are difficult to distinguish from the actual texts, and thus, confusion or error may arise. Text detection is challenging due to variability in some imaging condition, which are Noise, blur, low resolution, and obstruction in the text detection phase. Representation of text detection involves the way and manner of describing text and background in natural images.

1.1 OVERVIEW OF TEXT DETECTION:

Text detection is an crucial domain in computer vision. Text detection often preprocessing process before text classification. We need to classify the areas of text in my natural images. I need to set up a boundary so that the classifier knows which part of the image is the text. There are a various application of text detection-**Text Image Search, Land mark Identification.** Text detection is challenging due to variability in some imaging condition, which are noise, blur, obstruction, etc. So it is challenging to map the text in natural images.

Suitable text detection algorithm should be robust against variability .one of the most popular technique of text identification is using geometric properties of text to distinguish it from the background. Each of the languages has the stroke with the thickness of each or even letters remain practically the same and height of the characters and height of the letters remain in particular fashion to some extent. The problem with SWT (Stroke with Transform) that it cannot be so useful for all the texts which are not flat to be detected so easily. Moreover, it also is not so good for natural images which are corrupted with noise. So MSER is a method that detects the text region in natural image.

MSER (**Maximally Stable Extremal Region**) uses texture properties of the image to distinguish the text from the rest of the natural image. It has highly desirable properties such as invariance to monotonic intensity transformation and it has low computational complexity. It is very sensitive to blur and among all the different identification technique, it increases robustness. Text detection in natural images is done by locating text in bounding boxes [1][2]. Text extraction is done by finalizing the scene images in such a manner that all text pixels are foreground and the rests are background [3][4]. Text region proposal methods give multiple possible text bounding boxes [5][6].

.

1.2 RELATED REVIEW

In [8], everal research works have been proposed for image segmentation algorithms using MSER. Their purpose is to aim at segmenting out specific regions corresponding to user-defined objects. The abovementioned paper proposes a novel algorithm that is based on MSER. Its work is to segment natural images without user intervention and capturing multi-scale structure. This algorithm works by collecting MSERs and then partitioning the whole image plane by redrawing them in a specific order. Hierarchical morphological operations are developed in order to make noise free and smooth the region boundaries. Effects of different types of LOD control are demonstrated for image stylization to illustrate the effectiveness of the algorithm's multi-scale structure,

In [14], one of the important steps of scene text recognition system is Scene text detection, and it is a challenging problem also. The main challenges of scene text detection are different from general object detections that depend on arbitrary orientations, small sizes, and significantly variant aspect ratios of text in natural images. In the mentioned paper, an end-to-end trainable fast scene text detector is presented, namely TextBoxes++. It detects arbitrary oriented scene text with both high accuracy and efficiency in a single network forward pass. Only efficient non-maximum suppression is involved in this process.

The paper has evaluated the proposed TextBoxes++ based on four public data sets. TextBoxes++ performs better than all other competing methods in terms of text localization accuracy and runtime in all the experiments, to be more precise, TextBoxes++ achieves an f-measure of 0.817 at 11.6 frames/s for 1024 × 1024 ICDAR 2015 incidental text images and an f-measure of 0.5591 at 19.8 frames/s for 768× 768 COCO-Text images. When TextBoxes++ gets combined with a text recognizer, it significantly outperforms the state-of-the-art approaches for word spotting and end-to-end text recognition tasks on popular benchmarks. Code is available at https://github.com/MhLiao/TextBoxes_plusplus.

In [13] as per the research, most state-of-the-art text detection methods are specific to horizontal Latin text. They are not fast enough for real-time applications. The research introduces Segment Linking (SegLink), as an oriented text detection method. The main idea behind this is to decompose text into two locally detectable elements, namely segments and links. A segment serves as an oriented box that covers a part of a word or text line; a link connects two adjacent segments; it implies that they belong to the same word or text line. Both the elements are detected densely at multiple scales by an end-to-end trained, fully-convolutional neural network. Final detections are done by combining segments that are connected by links. SegLink improves in terms of dimensions of accuracy, speed, and ease of training compared to the previous methods. It achieves an f-measure of 75.0% on the standard ICDAR 2015 Incidental (Challenge 4) benchmark that goes above the previous best by a large margin. Besides, SegLink is able to detect long lines of non-Latin text without modification, such as Chinese.

In this chapter, some fundamental concepts and related works that are useful to understand the methods of text localization and recognition are introduced and briefly described.

1.3 TEXT DETECTION METHODS

In Region-Based, The objective of segmentation is to divide a natural scene images into multiple regions[23]. Technique such as thresholding achieve this goal by looking for the boundaries using some segmentation methods[22] between regions based on discontinuities in color properties or grayscale properties. Region-based segmentation is an important technique that is used for determining the region directly. These region-based algorithms can be classified into two main classes:

MERGING ALGORITHMS:

Under these algorithms, neighboring regions are compared and if they are close enough in some property and after that, they are merged.

SPLITTING ALGORITHMS:

In these algorithm large regions which are not uniform in nature, they are broken up into smaller areas in order to make them uniform properly and this algorithm is also a combination of splitting and merging. There are some criterion applied to decide whether the regions have to be merged or split into multiple regions. This criterion defined by using different statistical approaches such as variance on the measured property (intensity, color, mean, etc.) of the regions and standard deviation. The two works that are worth mentioning due to their particular novelty occurs. [9] and Wang et al. [10].

Pan et al. [9] proposed a method to accurately localize natural texts in natural scene images in Computer Vision. They built a text map with the help of a text region detector based on which text components can be segmented by local banalization method. A Conditional Random Field (CRF) is a good model that is classified into label the components as “Text” or “Non-Text.” This proposed method that is used for performance comparing with the existing ones on ICDAR 2003 competition dataset.

The second algorithm is worth analyzing is of Wang et al. [10] where two systems are developed to solve the end-to-end problem for word recognition in computer vision and Pattern Recognition. The first one represents a two-stage pipeline that is

consisting of text detection algorithm followed by Optical Character Recognition (OCR) engine. The second one is a system of generic object recognition.

This algorithm can able to overcome all the limitations of region-based methods such as (i) high computational complexity, (ii) long training times and (iii) false-positive detection errors as many regions of natural scene images are difficult to classify from the text components.

In connected component cased, there are some connected component-based approach is usually used in text characters that have the same geometric properties. For example, the color of a text character remains the same for the all whole letter, the characters are usually placed on a high contrast background to increase readability. So, researchers have been done over the years to propose different techniques on text detections.

Maximally Stable Extremal Region (MSER) [11] to identify the text components from a natural scene image In Maximally Stable Extremal Region Based.

MSER is a method for blob detection [12] in Natural images to compute several co-variant regions from a given gray image called MSER.

MSER (**Maximally Stable Extremal Region**) uses texture properties of the image to distinguish the text from the rest of the natural image. It has highly desirable properties such as invariance to monotonic intensity transformation and it has low computational complexity. It is very sensitive to blur, and among all the different identification technique, it increases robustness.

Stroke Width Transform (SWT) [15] is another algorithm that is quite popular among text detection method. Suitable Text detection algorithm should be robust against variability and one of the most popular techniques of text recognition is using geometric properties of text to distinguish it from the background. If it check out the alphabets, it will find that most languages have common pattern. Each of the languages has the stroke with the thickness of each or even letters remain practically same and Height of the Characters and height of the letters remain in particular fashion to some extent. The problem with SWT (Stroke with Transform) that it cannot be so

useful for all the texts which are not flat to be detected so easily. Moreover, it also is not so good for natural images which are corrupted with noise.

1.4 BRIEF DESCRIPTION OF SOME POPULAR DATASETS

There are several datasets [16] are used in Computer vision but it is related to text detection which are as follows-

CHARS74K DATASET [20] was developed by de Campos et al [17]. This dataset contains 7705 English character images (A-Z, a-z and 0-9, in total 64 classes present) and 3345 Kannada character images (647 classes), which were manually segmented from 1922 scene text images.

ICDAR-2003 DATASET was proposed by Simon Lucas and his colleagues for the ICDAR-2003 Robust Reading Competition [18]. This dataset was used in an unchanged form in the ICDAR-2005 Robust Reading Competition. This Dataset Contains 258 training and 251 testing images with words and characters and consists of bounding boxes and their text content. 1157 word and 6185 character images.

ICDAR-2011DATASET [21] was developed by taking all images from the ICDAR-2003 dataset, it is used for removing images with no text and adding several new images and splitting them again into a testing subset. ICDAR-2011DATASET was first used in the ICDAR-2011 Robust Reading Competition and then subsequently in computer vision. There are different datasets, and evaluated on the 2011 testing set but the testing set contains many images from the joint training set.

ICDAR-2015 DATASET was collected by people who wearing Google Glass devices [19] and walking in Singapore and other cities and then subsequently by selecting and annotating only images with text. This dataset was introduced in the ICDAR-2015 Robust Reading Competition [20] to address the problems in a Dataset which is ICDAR 2003/2011 datasets. It is used in Incidental Scene Text Challenge. The dataset

contains 1670 .Images with 17548 annotated words. 1500 images are publicly available Street View Text (SVT) Dataset

ICDAR-2015 DATASET [22] was developed by Wang and Belongie.The data was collected by asking annotators to collect images in the Google Street View application. The dataset contains mostly used in business names and business signs and the business names by looking up businesses close to the GPS position of the image. This dataset contains 350 images (100 training and 250 testing images) of 20 different cities and 725 labeled words.

Above mentioned datasets there are few more types of datasets related to text detection. They are [16] (a) COCO-TEXT DATASET, (b) IIIT DATASETS, (c) KAIST DATASET, (d) NEOCR DATASET.

Figure: 1 ICDAR2003 DATASET SAMPLE

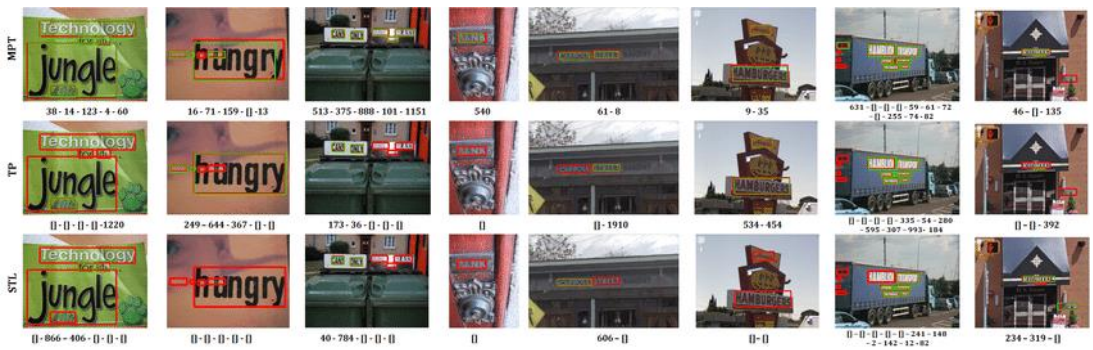


Figure: 2 SVT and ICDAR2015 DATASET SAMPLE

Figure: 3 ICDAR-2011 DATASET SAMPLE

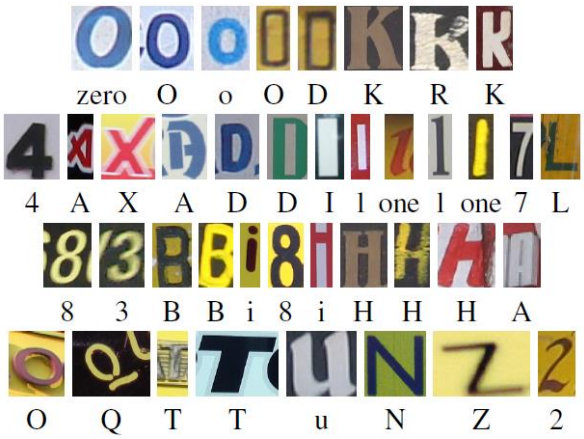


Figure: 4 CHARS74K DATASET SAMPLE

There are so many Data Set are available in scene text detection. But we have created our own dataset using some new technique.

1.5 THESIS OBJECTIVE

This thesis paper has been classified into five main chapters. The first chapter presents a rough idea about the domain of text detection, different challenges and the motivation behind the present work. The second chapter highlights the evolution of different text detection methods and provides a literature survey of the novel approaches utilized in the avenue of text detection. Also a brief introduction of the datasets has also been included in this Paper. The third chapter provides detailed of study of text detection using Maximally Stable Extremal Region (MSER) method. The proposed methodology has been mathematically established for the localization of the texts. The fourth chapter highlights the efficiency of the method in localizing texts in natural images. The corresponding challenges related to our work and the advantages of using this method is also described in this Paper. The last chapter concludes and discusses relevant future scope in scene text understanding in natural Images.

CHAPTER TWO

METHODOLOGY

Here, in this chapter, we are trying to describe the whole proposed methodology through the following steps-

- Maximally Stable Extremal Region (MSER) detection

2.1 BRIEF INTRODUCTION OF MAXIMALLY STABLE EXTREMAL REGION (MSER)

MSER [16] [17] is a popularly used method for blob detection in Natural Images. MSER (**Maximally Stable Extremal Region**) uses texture properties of the image to distinguish the text from the rest of the natural image. It has highly desirable properties such as invariance to monotonic intensity transformation and it has low computational complexity. It is very sensitive to blur, and among all the different identification technique, it increases robustness. Text detection in natural images is done by locating text in bounding boxes [1][2]. Text extraction is done by finalizing the scene images in such a manner that all text pixels are foreground and the rests are background [3][4].Text region proposal methods give multiple possible text bounding boxes [5][6]. In Some many Natural images is stable over a huge range of threshols in certain region that is of interest since they possess the following properties. Those are Invariance to affine transformation of natural images intensities of Text identification process and Adjacency preserving of covariance, stability and Multi-scale detection.

If we see a sequence of mages I_t with frame t corresponding to threshold t, we have to see first a black image after that see the white spots corresponding to local intensity

minima will be appear then grow increases. So white spots will eventually merge, until the whole image is white. The word extremal refers to the property that all pixels are inside the Maximally Stable Extremal Region have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the addressable element of the screen i.e. Pixel on its outer boundary.

2.2 MATHEMATICAL DETAILS

Let us consider a Natural scene image I(x), and $x \in \Lambda$ is a real function of a finite set that is Λ with a topology τ. Elements of Λ are called pixels that is smallest addressable element of the screen.

For simplicity method, let us $\Lambda = [1, 2, \ldots, N]^n$ and the topology τ induced by the 4-way or 8-way neighborhoods in sample image , but we do not restrict ourselves to n = 2.

Now we have to consider a level set S(x), $x \in \Lambda$ of the image I(x) is the set of pixels that have intensity which is not greater than I(x), so can write the following form-

$$\boldsymbol{S}(\boldsymbol{x}) = \{\boldsymbol{y} \in \Lambda : \boldsymbol{I}(\boldsymbol{y}) \leq \boldsymbol{I}(\boldsymbol{x})\} \qquad \text{.....Equation 1}$$

Considering a path $(x_1,,,,,,,x_n)$ is a continuous sequence of pixels which is addressable element of the screen (i.e. such that x_i and x_{i+1}) are 4-way or 8-way neighbors for (i = 1, . . . , n − 1). Now A connected component C that is in the set Λ is a subset $C \subset \Lambda$ for which each pair $(x_1, x_2) \in C^2$ of pixels. That is connected by a path fully contained in connected component C.

So the connected component C is maximal if any other connected component C´ containing C is equal to connected component C. An extremal region R is a maximal connected component of a level set S(x). We denote by R (I) that is the set of all extremal regions of sample image I.

STABILITY CRITERIA-

Considering all extremal regions R (I), we are interested in the ones that satisfy certain stability criteria. This criterion will be introduced next.so Let the level I(R).That is of the extremal region R be the maximum natural image value attained in the region R, i.e.

$$\boldsymbol{I(R) = \sup_{x \in R} I(x)} \qquad \text{.....Equation 2}$$

Now, an extremal region R of a one-dimensional image I(x) is shown in equation 2. And the two corresponding extremal regions are considered as $R_{+\triangle}$ and $R_{-\triangle}$.

Let $\triangle > 0$. Let $R_{+\triangle}$ be the smallest extremal region that contains iR and the stability of an extremal region R is the inverse of the relative area of the region R when the intensity level is increased $\triangle$. It has intensity which exceeds of at least $\triangle$ the intensity of R, i.e.

$$\boldsymbol{R_{+\triangle} = argmin\{|Q|: Q \in R(I), Q \supset R, I(Q) \geq I(R) + \triangle\}} \qquad \text{.....Equation 3}$$

Similarly, let $R_{-\triangle}$ be the most significant extremal region containing R .R has intensity which is exceeded by at least $\triangle$ by R, i.e.

$$\boldsymbol{R_{-\triangle} = argmax\{|Q|: Q \in R(I), Q \subset R, I(Q) \leq I(R) - \triangle\}} \qquad \text{..... Equation 4}$$

Consider the area variation, so formula of area variation

$$\boldsymbol{\rho(R;\triangle) = \frac{|R_{+\triangle}| - |R_{-\triangle}|}{|R|}} \qquad \text{..... Equation 5}$$

The region R is maximally stable [28] if it is a minimum for the area variation, in the following sense: $\rho(R;\triangle)$ is smaller than $\rho(Q;\triangle)$ for any extremal region Q immediately containing R.As we increase $\triangle$ fewer and fewer regions are detected

until finally at △=160.MSER are controlled by a single parameter △ which controls how the stability is calculated.

An extremal region R immediately contains another extremal region Q if R ⊃ Q and if R´ is another extremal region with R ⊃ R´ ⊃ Q, then R´ = R. because the base set Λ is finite.

2.3 IMPLEMENTATION OF SEMI SUPERVIED TECHNIQUE IN OUR WORK

Semi-supervised learning is a class of machine learning process and techniques that is used to make unlabeled data for training.Typically main function of that process is a small amount of labeled data with a large amount of unlabeled data conversion. This technique learning falls between unsupervised learning and supervised learning.Most of the machine-learning researchers have found unlabeled data when used in conjunction with a small amount of labeled data.It can produce improvement in learning accuracy. The cost of this process associated with the labeling process. It may render a fully labeled training set infeasible.In this technique, whereas the acquisition of unlabeled data is relatively so much inexpensive. For this situation, semi-supervised technique can be of great practical value. Semi-supervised tecnique process is also of theoretical interest in machine learning and pattern recognition.

Different classifiers, we now consider how we can apply them in a semi-supervised technique. We use L that denotes the set of labeled training images examples, and U refers to the set of unlabeled training images examples. So our training images have associated tags, but that our final task must classify images that does not have such tags. This technique learning falls between unsupervised learning and supervised learning. So proceed by learning images a first classifier on the labeled images examples in L, and then have to use predict the class labels for the unlabeled examples in U.

Estimate the class labels for the natural scene images in U. we train a visual-only SVM classifier fv in all training images examples in L ∪ U. In practice, however, the joint

classifier is not perfect because we have two alternative approaches to leverage the predictions way to classifier joint on the unlabeled images examples in U.

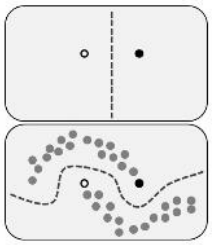
We add the examples which is confidently classified that is use by MKL classifier and fall outside the margin. i.e.

$$|\mathrm{fc}(x)| \geq 1,$$

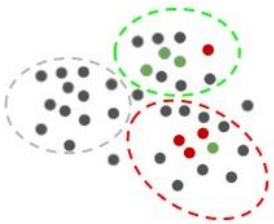
Then adding all examples in U.The observation that these are precisely the examples of natural images that would not change the MKL classifier because if they were included the training data for this.Our second alternative process is motivated by the observation that gives information from the MKL classifier. So this tecnique that we use when training the final visual classifier which is the sign of the examples selected from U..

This technique learning falls between unsupervised learning and supervised learning .Most of the machine-learning researchers have found unlabeled data, when used in conjunction with a small amount of labeled data.It can produce improvement in learning accuracy.The cost in this process associated with the labeling process.It may render a fully labeled training set infeasible.

Semi Supervied learning:



Classification



Clustering

An image-only classifier has 2 options:

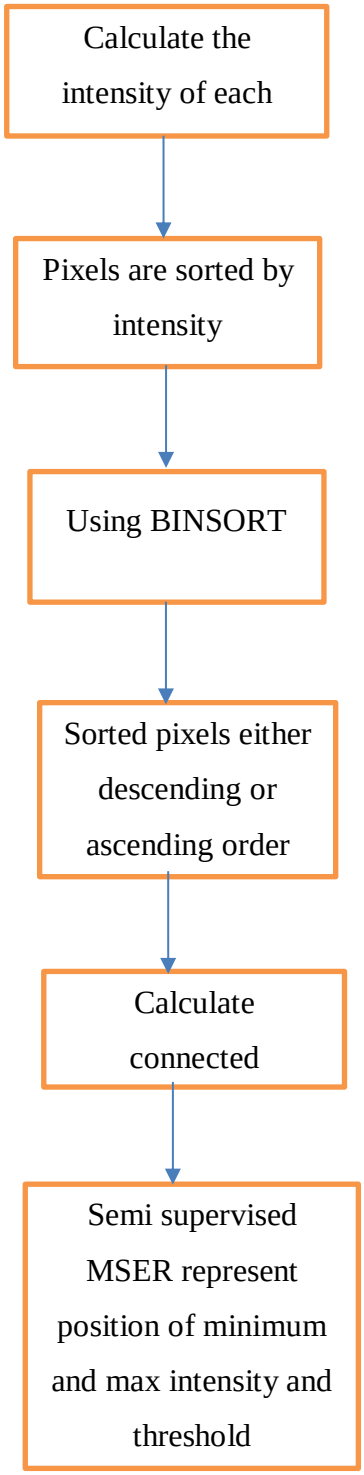
1 SVM: Full form of SVM is support vector machine.Use unlabeled data with label from sign of MKL score in the natural images

2 LSR: Full form of Least-squares regression.That is of MKL scores using the visual kernel using KPCA projection.

First, most of the text regions are detected using Maximally Stable Extremal Region. It works well for text regions because the high consistent color and high contrast of text able to stable intensity of the images.

First calculate the intensity of each pixel that mean pixels are sorted by intensity approaches. Then sort the all pixels using BINSORT since the sort can be implemented as BINSORT. Place those sorted pixels either descending or ascending order or the list of connected component. Maximally stable Extremal region is represented by position of the local intensity minimum or maximum and threshold. next find connected component and area maintained by UNION find algorithm. next find the stable area of the Text in natural scene Images.

Block Diagram of Semi-supervised MSER Tecnique:



CHAPTER THREE

EXPERIMENTAL RESULT AND ANALYSIS

3.1 DATASET COLLECTION:

We have collected 3120 RGB images from natural scene using MI note 4 mobile camera.and create 424 ground truth images randomly. The images are natural images such as heading of any shop, hospital name, road sign, banner, car number plate, house name and also other natural images. Where the text portions have different orientation and different font style. To make the dataset usable to any research related work, we have annotated those data.

3.2. DATASET ANNOTATION:

Step 1. We have manually draw the region of interest i.e. text region using MATLAB application and save the coordinates (upper left corner x, y position and height, and width) of that rectangle box. All the coordinates are stored in a matrix format.

Step 2. We have written all the coordinate values into an xml file. The XML file also contains respective image name, respective text name.

Step 3. Using this XML draw the rectangle box over the respective region.

(a)

(b)

(c)

Figure:10 Collected Images

(a)

(b)

(c)

Figure:11 Localized ground truth image

3.3 PROCEDURES:

Preprocessing stage:

First, we have enhanced contrast of our dataset images using histogram equalization method.

Detect text regions using region detection method

In this paper we have used gradient based maximally stable extremal region (MSER) method to detect text region over the image. Then convert the image region into binary form and find the connected component of those region. We have filtered non text region using selected connected components. Also draw a bounding box on respective connected components. Where the bounding box contain text and non-text both region. We have cropped those box regions.

Feature calculation

We have calculated histogram of oriented feature from each box region.

Classification using semi supervised SVM.

We have applied semi supervised algorithm to classify text and non-text regions.

3.4 RESULTS AND DISCUSSIONS:

In this paper, we have used our own developed dataset. Figure 5 describes some examples of captured images, the images after contrast enhancement and also text region detected image using MSER. Figure 6 describes some original images and also display their detected region before filtering nontext regions.

FIGURE: 5 Sample images taken from the developed datasetimage:

Figure: 6 The Sample images after Contrast enhancement

Figure: 7 Detected regions in the sample images after using semisupervised MSER

3.5 Experimental Result And Analysis:

Figure:8 Sample images taken from the developed original mage

Figure:9 Detected regions in the sample images after using MSER method.

Accuracy is an important part of classification project. It differntiates the classified image to other data source and it is considered to be accurate or ground truth data.We calculate an accuracy by number of correct prediction divided by the total no of prediction.In our experiment we have used the same accuracy metric on pixel level.So for our model the accuracy metic will be :

Accuracy=[Total no of predicted pixel /Total no of pixels]

In our experiment we have used 35 images for Training purposes and 100 images for testing.We have got 70.34% accuracy in pixel level from the text region applying on 100 test images.

3.6 Experimental Result And Analysis:

In this paper we have used our own dataset. Original Images(Figure:8) describes some example of original image, contrast enhanced image and also MSER region detected image(Figure:9). Detected Region Images describes some original images and also display their detected region before filtering non text regions.

CHAPTER FOUR

CONCLUSION AND FUTURE SCOPE

Text detection from scene images is a complex Computer Vision task is being studied by many research laboratories and international companies for its importance. And it is use in newly developed technologies, such as automated driving and automated locating of information from visual data. Unfortunately, till now no method proposed in literature achieves semi-supervised text detection rates that are even remotely comparable to human observers' performances. Among different text detection methods in this field is still very strong, especially on standard datasets like the ones from the national and International Conference on Document Analysis and Recognition (ICDAR). As described in this thesis, we have presented here an approach for text detection from natural scene images using Maximally Stable Extremal Region. Maximally Stable Extremal Region (MSER) to obtain state-of-the-art accuracy rates for text detection and identification from natural scene images. In this solution uncommon text fonts that are typically filtered out as noise elements by competing approaches are correctly retained and recognized.

To improve our accuracy and also recognize the text regions. The positive side of these process is In MSER detection, we seek a range of threshold. This threshold leaves watershed effectively unchanged because the threshold are highly unstable.Importantily this process is very fast in practice. Finally we remark that Maximally Stable Extremal Region can be defined on any natural image whose pixel values are from a totally order set approaches.

In Future work, we trying to proceed fully automatic projective reconstruction of 3D scene images. Secondly we have to find properties of robust similarity measurement area and their selection which based on statistical properties of the text.

CHAPTER FIVE

REFERENCES

[1] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 34, no. 3, pp. 2963–2970, Mar. 2010.

[2] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Deep Features for Text Spotting," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, 2014, pp. 512–528.

[3] D. Karatzas *et al.*, "ICDAR 2015 competition on Robust Reading," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2015, vol. 2015–Novem, pp. 1156–1160.

[4] Y. Zhu, C. Yao, and X. Bai, "Scene text detection and recognition: recent advances and future trends," *Frontiers of Computer Science*. 2016.

[5] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.

[6] Y.-F. Pan, X. Hou, and C.-L. Liu, "Text Localization in Natural Scene Images Based on Conditional Random Field," *2009 10th Int. Conf. Doc. Anal. Recognit.*, pp. 6–10, 2009.

[7] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3538–3545.

[8] L. Neumann and J. Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in *2011 International Conference on Document Analysis and Recognition*, 2011, pp. 687–691.

[9] L. Gomez and D. Karatzas, "A fast hierarchical method for multi-script and arbitrary oriented scene text extraction," *Int. J. Doc. Anal. Recognit.*, vol. 19, no. 4, pp. 335–349, 2016.

[10] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2013 document image binarization contest (DIBCO 2013)," in *2013 12th International Conference on Document Analysis and Recognition*, 2013, pp. 1471–1476.

[11] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 1–20,

2016.

[12] I.-S. Oh, J. Lee, and A. Majumder, “Multi-scale image segmentation using MSER,” in *International Conference on Computer Analysis of Images and Patterns*, 2013, pp. 201–208.

[13] K. Wang, B. Babenko, and S. Belongie, “End-to-end scene text recognition,” in *2011 International Conference on Computer Vision*, 2011, pp. 1457–1464.

[14] S. Tian, S. Lu, B. Su, and C. L. Tan, “Scene text segmentation with multi-level maximally stable extremal regions,” in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 2703–2708.

[15] B. Shi, X. Bai, and S. Belongie, “Detecting oriented text in natural images by linking segments,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2550–2558.

[16] M. Liao, B. Shi, and X. Bai, “Textboxes++: A single-shot oriented scene text detector,” *IEEE Trans. image Process.*, vol. 27, no. 8, pp. 3676–3690, 2018.

[17] L. Neumann, “Scene text localization and recognition in images and videos,” Department of Cybernetics Faculty of Electrical Engineering, Czech Technical, 2017.

[18] T. E. De Campos, B. R. Babu, M. Varma, and others, “Character recognition in natural images.,” *VISAPP (2)*, vol. 7, 2009.

[19] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, “ICDAR 2003 robust reading competitions,” in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, 2003, pp. 682–687.

[20] R. Nagy, A. Dicker, and K. Meyer-Wegener, “NEOCR: A configurable dataset for natural image text recognition,” in *International Workshop on Camera-Based Document Analysis and Recognition*, 2011, pp. 150–163.

[21] S. Saini and C. Marawaha, “Comparative study of text detection in natural scene images,” in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2016, pp. 1981–1985.

[22] M. Guillaumin, J. Verbeek, and C. Schmid, “Multimodal semi-supervised learning for image classification,” in *2010 IEEE Computer society conference on computer vision and pattern recognition*, 2010, pp. 902–909.

[23] https://en.wikipedia.org/wiki/Semi-supervised_learning