Pune Institute of Computer Technology Dhankawadi, Pune

A SEMINAR REPORT
ON

Text Detection and Recognition in Natural Images

Submitted By
Geet Choudhari
Rollno:31411
Class:TE-4

Under the guidance of
Dr.Sunil.D.Kale

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology Dhankawadi, Pune-411043

Pune Institute of Computer Technology
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Seminar report entitled

"**Text Detection and Recognition in Natural Images**"

Submitted By
Geet Choudhari
Rollno:31411

has satisfactorily completed a seminar report under the guidance Of Dr. Sunil. D. Kale towards the partial fulfillment of third year Computer Engineering Semester II, Academic Year 2019-20 of Savitribai Phule Pune University

Dr. Sunil. D. Kale
Internal Guide

Prof. M.S.Takalikar
H.O.D Department of Computer Engineering

Place:
Date:

Contents

Acknowledgement

I sincerely thank our Seminar Coordinator Prof. B.D.Zope and Head of Department Prof. M.S.Takalikar for their support.

I also sincerely convey my gratitude to my guide Dr.Sunil.D.Kale, Department of Computer Engineering for his constant support, providing all the help, motivation and encouragement from beginning till end to make this seminar a grand success.

1. ABSTRACT:

Text detection is one of the most challenging and commonly dealt application in computer vision. Detecting text Region is the first step of text recognition system called Optical Character Recognition (OCR). In this, we utilize Maximally Stable Extremal Regions to acquire very first text region candidates. This Process requires the Separations of text region from non-text region. Then possible regions are reduced in quantity by using geometric and stroke width properties.

Candidate regions are joined to obtain text groups. Finally Tesseract Optical Character Recognition engine is utilized as the last step to eliminate non-text group. For natural images and computer-generated images 82.7% precision and 52.0% f-accuracy; for computer-generated images 64.0% precision and 65.2% f-accuracy is achieved.

Keywords:

Text detection, Maximally Stable Extremal Regions, geometric and stroke width properties, non-text region elimination, Optical Character Recognition

2. INTRODUCTION

In recent years, use of multimedia technology has increased tremendously. In multimedia technology image is one of the important part and image can have different contents in it, such as face, human, scene, text, etc. Among all contents in images, text is found to be one of the most important features to understand the image contents. Text in images can be used as indexing purpose. The text information can be extracted in two stages: text detection and text recognition. Text detection detects the text regions as external regions of an image and in text recognition stage system retrieves the text information from these external regions

Optical character recognition (OCR) algorithms allow computers to analyze printed or handwritten documents automatically and prepare text data editable formats for computers to efficiently process them. Human eyes naturally recognize various patterns, fonts or styles. For computers, it is hard work to do. Any scanned document is a graphics file, i.e., a pattern of pixels. A computer localizes, detects and recognizes characters on an image and turns the image of paper documents into a text file. Then, it becomes possible to extract meaningful information. Texts in a machine-readable form can then be used for different purposes. They can be scanned in search of patterns and vital data, used to generate reports and draw up charts, distributed into spreadsheets, and more.

some **attributes** of the OCR tasks

- **Text density**: on a printed/written page, text is dense. However, given an image of a street with a single street sign, text is sparse.
- **Structure of text**: text on a page is structured, mostly in strict rows, while text in the wild may be sprinkled everywhere, in different rotations.
- **Fonts**: printed fonts are easier, since they are more structured then the noisy hand-written characters.
- **Character type:** text may come in different language which may be very different from each other. Additionally, structure of text may be different from numbers, such as house numbers etc.
- **Artifacts**: clearly, outdoor pictures are much noisier than the comfortable scanner.
- **Location**: some tasks include cropped/centred text, while in others, text may be located in random locations in the image.

2.1 MOTIVATION

Text Detection and Recognition in general have quite a lot of relevant application for automatic indexing and information retrieval such document indexing, context based image retrieval and licensed car plate recognition which further opens up the possibility for more improved and advanced facilities.

3. A SURVEY ON PAPERS

1. COCO-Text: Dataset for Text Detection and Recognition:

The dataset is based on the Microsoft COCO dataset [10] that annotates common objects in their natural contexts. Combining rich text annotations and object annotations in natural images provides a great opportunity for research in scene text detection and recognition. The dataset contains 63,686 images with 173,589 labeled text regions. For each text region, it provide the location in terms of bounding boxes, classifications in terms of legibility, category (machine printed or hand written) and script of the text, as well as transcriptions in case of legible text with western script.

2. Scene Text Detection and Recognition: The Deep Learning Era

This Paper is based on METHODS BEFORE THE DEEP LEARNING ERA brief glance retrospectively at algorithms before the deep learning era. More detailed and comprehensive coverage of these works can be found in [146], [166], [172], [184]. For text detection and recognition, the attention has been the design of features, most text detection methods either adopt Connected Components Analysis (CCA) [26], [57], [63], [110], [145], [168], [171] or Sliding Window (SW) based classification [19], [74], [152], [154]

3. Reading Text in the Wild with Convolutional Neural Networks:

In this paper we advance text spotting methods, stages and approaches are as follow, word bounding box proposal generation (Section 4), proposal filtering and adjustments (Section 5), text recognition (Section 6) and final merging for the specific task (Section 7).

4.PROBLEM DEFINITION AND SCOPE

Problem:

Text Detection and Recognition in Natural Images

Scope:

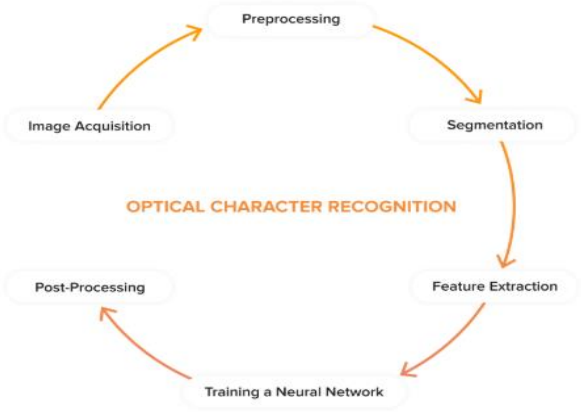
Text detection, segmentation and extraction from complex images can be applied to a variety of fields where the information needs to be analysed and understood. Some of these applications are given below

1. **Image understanding:** When images can be automatically understood and indexed by computer, the efficiency of running digital libraries and image database system will be greatly improved
2. **Vehicle testing:** Vehicle license and scene subtitles have many features in common, so text extraction can be used to supervise the traffic in real time. After text extraction from highway video flow, the traffic situation can be overseen and vehicle licenses can be recognized easily from traffic accidents, which can improve the efficiency of the transportation systems.

5.METHODOLOGY

Detection and recognition of text in an image are usually two different phases of the same application. The latter is handled by OCR engines. Yet in most of the cases, whole image includes too many non-text areas, which cause confusion for OCR algorithms. These non-text regions as seen in Figure (e.g.) reduce the performance of the applications. Therefore, eliminating these unwanted regions is necessary. As stated, natural images have too many light and shadow conditions in them, while computer-generated images tend to include wide variety of stylized fonts and colors. Both cases create challenging situations for separating non text regions. Although, these changeable environments make the task more difficult to solve with solely image processing steps, the proposed method still can be used as auxiliary steps with popular and robust deep learning-based techniques

The development process usually encompasses 6 steps needed to train an algorithm for efficient problem-solving with the help of optical character recognition.



1.Image Acquisition:

The first step is to acquire images and an OCR scanner should be able to threshold images. In other words, it should replace each pixel in an image with a black or a white pixel. It is a method of image segmentation.

2.PreProcessing:

The goal of preprocessing is to make raw data usable by computers. The noise level on an image should be optimized and areas outside the text to be removed. Preprocessing allows obtaining a clean character image to yield better results of image recognition

3.Segmentation:

The process of segmentation is aimed at grouping characters into meaningful chunks. There can be predefined classes for characters. So, images can be scanned for patterns that match the classes.

4.Feature Extraction:

This step means splitting the input data into a set of features, that is, to find essential characteristics that make one or another pattern recognizable. As a result, each character gets classified in a particular class.

5.Training a Neural Network:

Once all the features are extracted, they can be fetched to a neural network (NN) to train it to recognize characters. A training dataset and the methods applied to achieve the best output will depend on a problem that requires an OCR-based solution

6.Post-Processing:

This stage is the process of refinement as an OCR model can require some corrections. However, it isn't possible to achieve 100% recognition accuracy. The identification of characters heavily depends on the context. The verification of the output requires a human-in-the-loop approach.

```
Detect the Text from Given Images:

Text Detection in images:
PUNÉ INSTITUTE OF
COMPUTER TECHNOLOGY
SOCIETY FOR COMPUTER TECHNOLO0Y & RESEANCH
Society for Computer Technology & Research's
PUNE ISTITUTE OF COMPUTER TECHN
SN 27, DHANKARAD PUNE-44 NDIA
O PARKING
```

6.RESULTS:

1. Implementation results:


```
>>> client = vision.ImageAnnotatorClient()
>>> file_name = 'text1.png'
>>> image_path = f'.\VisionApiDemo\images\{file_name}'
>>> with io.open(image_path,'rb') as image_file:
...     content = image_file.read()
...
>>>         # create image instances
>>> image = vision.types.Image(content=content)
>>>
>>>         #to create response and json
>>> response = client.text_detection(image=image)
>>> texts = response.text_annotations
>>>
>>> df = pd.DataFrame(columns=['locale','description'])
>>>
>>> for text in texts:
...     df = df.append(
...         dict(locale=text.locale,
...             description=text.description
...         ),
...         ignore_index=True
...     )
...
...
>>> print(df['description'][0])
MES WATER AUTHORITY
SEWER OUTFALL
EXTENDS 283 FT.
RIVERWARDS
FROM THIS NOTICE
```

2.Discussion:

Maximally Stable Extremal Regions (MSER) is used to get the outline of the text in a proper form and to avoid the broken or non-uniform alignment of pixels. It is a method that used to extract number of covariant regions in an image. The MSER not only represents the text region, it also represents the background region as well, which is helpful to classify the text into words or meaningful sentences

Using MSER, regions are detected. Geometric properties are used to filter out non-text region using threshold. Some of the geometric properties that are used in this system include:

Eccentricity: It is used to count the circular nature of the given regions
Solidity: It is the proportion of the pixels in the raised structure area that are likewise in a given region. It is calculated by (Area / raised area)
Extent: The location and size of the rectangle
Euler Number: It is a feature of the binary image.

7.CONCLUSION AND REFERENCES

Conclusion:

Text detection is applicable in real world scenarios like optical character recognition, artificial intelligence, distinguish between human and machine inputs and spam removal. Text detection is the process of locating areas in an image where, a meaning full text is occurred. Variation in environment in which the image is captured makes it a difficult process

The characters which are identified are classified in to meaning full word or sensible sentences. The identified words are chained together and checked weather they form the meaning full sentence. In this work, a system to detect text and classify of the same is presented. The accuracy of the proposed system is 82.31%.

References:

1. Amritha S Nadarajan, Thamizharasi A, “A Survey on Text Detection in Natural Images", International Journal of Engineering Development and Research (IJEDR), ISSN: 2321-9939, Volume.6, Issue 1, pp.60-66, January 2018.
2. Tridib Chakraborty et al, (2017), Text recognition using image processing, International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017, 765-768
3. Automatic Text Detection and Classification in Natural Images by C.P. Chaithanya, N. Manohar, Ajay Bazil Issac