IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-5 E-ISSN: 2347-2693

Optimized Solution for Efficient Detection of Text from Images

Aarti Arjun Andhale[1*] and Rishikesh Yeolekar[2]

[1*,2] *Department of Computer Engineering, MITCOE, Pune 38, India*

www.ijcseonline.org

Received: May/02/2015 Revised: May/10//2015 Accepted: May/24/2015 Published: May/30/ 2015

Abstract— Text detection and recognition in camera captured images have been considered as very important problems in computer vision community. Text detection and recognition is a hot topic for researchers in the field of image processing. Text detection and extraction is performed in a four-step approach that consists of the pre-processing which include binarization and noise removal of an image, image segmentation using connected component analysis, feature extraction using variance generation and finally classification by choosing a threshold value of variance property. The goal of the project is to develop an Android-platform based text detection application that will be able to recognize the text captured by a mobile phone camera. Optical character recognition (OCR) methods recognize the characters and can be really useful when you have got a paper document you want in digital, editable form. Character which can be used to assist a wide variety of applications, such as image understanding, image indexing and search, geolocation or navigation, and human computer interaction. Optical character recognition is very important technique that is used for recognition of characters and it is very useful when we want our paper document in digital form and with the help of this technique we can edit our form.

Keywords— Pre-processing, Segmentation, Optical Character Recognition (OCR)

I. INTRODUCTION

Text extraction techniques are widely studied because text embedded in images and videos provides important information. We can recognize the text with the help of Optical Character Recognition techniques. It allows a machine to recognize characters through optical mechanisms. Extracted text can be used in variety of applications such as banking applications, Translation camera, Reading data entered in forms(in Hospitals etc). Every such application relies on a Textual Information Extraction (TIE) system which can efficiently detect, localize and extract the text information present in the natural images. OCR is becoming an important part of modern research based computer applications. Especially with the advent of Unicode and support of complex scripts on personal computers, the importance of this application has increased. It is because text information is easily recognized by machines and can be used in a variety of applications. Some examples are aids for visually impaired people, translators for tourists, information retrieval systems in indoor and outdoor environments, and automatic robot navigation. Although there exist a lot of research activities in this field, scene text detection is still remained 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 [1].

II. RELATED WORK

In paper [1] authors proposed a new scene text detection algorithm based on two machine learning classifiers: one allows us to generate candidate word regions and the other filters out nontext ones.

In paper [2] author proposed Snooper text algorithm for the detection of text embedded in images or videos of urban scenes. In paper [3] authors presented a new approach for detection and extraction of text data from both scanned document images and scene images. In paper [4] authors presented a hybrid approach to robustly detect and localize texts in natural scene images. In paper [5] authors proposed a novel text detection algorithm that extracts six different classes' features of text, and uses Modest AdaBoost with multi-scale sequential search.

In paper [6] authors propose a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. In paper [7] authors proposed a unified framework for multi oriented text detection and recognition. In paper [8] authors proposed a multi-level MSER technology that identifies the best-quality text candidates from a set of stable regions that are extracted from different color channel images.

In paper [9] author presented one technique that was used for detection and removal of text from images. The system detects text using morphological operations, connected component labeling and a set of selection criteria which helps to filter out non text regions. So, the resultant image is the image with only texts. Text Inpainting is done in two steps. The first step detects the text region automatically, without user interaction and in the second step; the text is removed from the image using exemplar based Inpainting algorithm.

In paper [10] authors presented a simple, efficient, and less costly approach to construct OCR for reading any document that has fix font size and style or handwritten style. To achieve efficiency and less computational cost, OCR in this paper uses database to recognize English characters which makes this OCR very simple to manage.

In paper [11] author proposed one method that combines the functionality of Optical Character Recognition and speech synthesizer. The objective is to develop user friendly application which performs image to speech conversion system using android phones.

III. METHODOLOGY

In this section, we present the main ideas and details of the proposed algorithm. Specifically, we give an overview of the proposed framework.

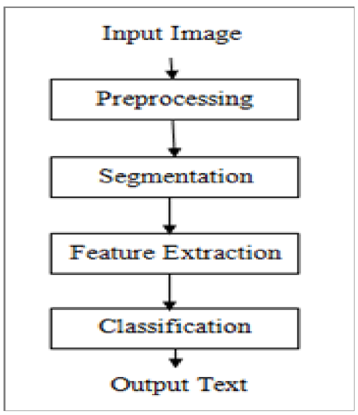


Fig 1.Optical Character Recognition [12]

A. Pre-processing:

Pre-processing steps are necessary to improve the performance and make the process efficient to the time. This includes gray-scaling and binarization of image and filtering to remove noise.

1. Gray-scaling: The given image is multicolor RGB image, in which text may not be separated from the background. For gray-scaling, these values are added in a proportion of Red: 30%, Green: 59% and Blue: 11% to get the gray scaled equivalent of that particular pixel.

2. Binarization: This converts gray-scale image into binary image i.e. containing only Black (0) and White (1) pixels. Gray-scaling gives a threshold for binarization of image. This process noticeably separate (or distinguish) text from image background. This is done by comparing each pixel value to a threshold value (that lies between black and white) and setting that pixel value to black or white as its consequence.

3. Filtering: Any image taken from camera contained noise such as blurred image, high frequency noise and white noise. To improve image quality and for further processing on image, Gaussian low pass filter is used.

It has following properties: 1) Gaussian smoothing is very effective for removing Gaussian noise. 2) They are linear low pass filters. 3) Rotationally symmetric (perform the same in all directions). 4) The degree of smoothing is controlled by σ (larger σ for more intensive smoothing) .

B. Segmentation:

The objective of the segmentation is to extract each character from the text present in the image.

C. Feature extraction:

In this stage, the features of the characters that are crucial for classifying them at recognition stage are extracted. Following are the techniques for features extraction Template matching and Transformations.

D. Classification:

The classification is the process of identifying each character and assigning to it the correct character class [12].

Optical Character Recognition is very important technique that is used for text detection from an image. OCR (Optical Character Recognition) also called Optical Character Reader is a system that provides a full alphanumeric recognition of printed or handwritten characters at electronic speed by simply scanning the form. More recently, the term Intelligent Character Recognition (ICR) has been used to describe the process of interpreting image data, in particular alphanumeric text [12].

IV. SCENE TEXT DETECTION VIA CONNECTED COMPONENT CLUSTERING AND NONTEXT FILTERING

A. Overview:

This is a new scene text detection algorithm based on two machine learning classifiers: one allows us to generate candidate word regions and the other filters out nontext ones. To be precise, we extract connected components (CCs) in images by using the maximally stable external region algorithm. These extracted CCs are partitioned into clusters so that we can generate candidate regions. Unlike conventional methods relying on heuristic rules in clustering, we train an AdaBoost classifier that determines the adjacency relationship and cluster CCs by using their pair wise relations. Then we normalize candidate word regions and determine whether each region contains text or not. Since the scale, skew, and color of each candidate can be estimated from CCs, we develop a text/nontext classifier for normalized images [13].

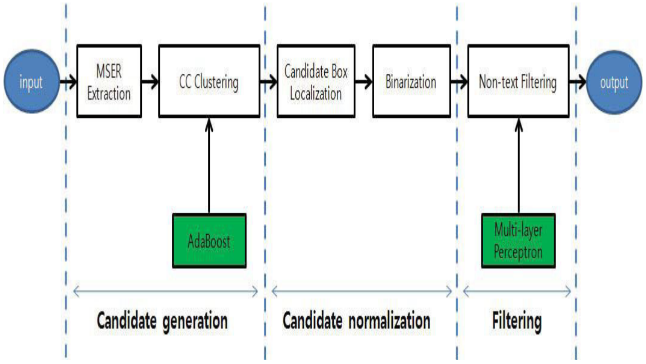


Fig 2: Block Diagram [13]

B. Steps:

The proposed framework consists of three steps Candidate Generation, Candidate normalization, filtering.

1) Candidate Generation: For the generation of candidates, extract CCs in images and partition the extracted CCs into clusters, where clustering algorithm is based on an adjacency relation classifier.

2) Character Normalization: In this first normalize the candidate and then binarize.

3) Filtering*:* Develop a text/nontext classifier that rejects nontext blocks among normalized images.

C. Advantages and Disadvantages:

- It shows good performance
- Small computation cost

V. SNOOPERTEXT: A TEXT DETECTION SYSTEM FOR AUTOMATIC INDEXING OF URBAN SCENES

SNOOPERTEXT is an original detector for textual information embedded in photos of building façades (such as names of stores, products and services) that we developed for the iTowns urban geographic information project. SNOOPERTEXT locates candidate characters by using toggle-mapping image segmentation and character/non-character classification based on shape descriptors. The candidate characters are then grouped to form either candidate words or candidate text lines. These candidate regions are then validated by a text/non-text classifier using a HOG-based descriptor specifically tuned to single-line text regions [14].

SnooperText detector has following five stages:

1. Image Segmentation
2. Character filtering
3. Character grouping
4. Text filtering
5. OCR

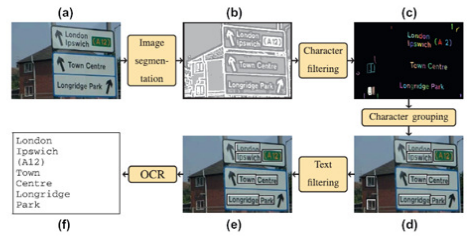


Fig 3. Overall diagram of the SnooperText detector (a)-(e) and the OCR step (e) & (f)[14]

Our system consists of three stages as shown in Fig. At the pre-processing stage, a text region detector is designed to detect text regions in each layer of the image pyramid and project the text confidence and scale information back to the original image, scale-adaptive local binarization is then applied to generate candidate text components. At the connected component analysis (CCA) stage, a CRF model combining unary component properties (including the text confidence) and binary contextual component relationships is used to filter out non-text components.

At the last stage, neighboring text components are linked with a learning-based minimum spanning tree (MST) algorithm and between-line/word edges are cut off with an energy minimization model to group text components into text lines or words [15].

Fig 4: Some text regions missed by SNOOPERTEXT

VI. DETECTION OF SCENE TEXT BASED ON MACHINE LEARNING CLASSIFIERS

In this paper, authors proposed one system which is based on machine learning perspective. It consists of three parts and they are listed as candidate generation, candidate normalization, and non-text filtering.

1. CANDIDATE GENERATION:

In the process of candidate generation, the extraction of Connected Components (CCs) takes place followed by the partitioning of the extracted CCs into clusters, based on an adjacency relation classifier.

a) CC Extraction:

Among a number of CC extraction methods, we have adopted the MSER algorithm 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. We can find out brighter CC's by assigning random colors to them.

b) Building Training Sets:

Training sets are built based on pair wise relations between CCs. The following are the cases for a CC pair.

1) $C_i \in T, C_j \in T, C_i \sim C_j$
2) $C_i \in T, C_j \in T, C_i \sim C_j, t(C_i) = t(C_j)$
3) $C_i \in T, C_j \in T, C_i \sim C_j, t(C_j) \neq t(C_j)$
4) $C_i \in T, C_j \in N$
5) $C_i \in N, C_j \in N$.

Here Ci and Cj represent the connected components, T and N represents the text and nontext sets. t(Ci) and t(Cj) represents the text line for that particular connected component. Ci ~ Cj represents Ci is adjacent to Cj. Based on the above observations, we build training sets. Corresponding to the case (1) and corresponding to the case (3) or (4) a positive set and a negative set are built respectively by gathering the samples. Samples corresponding to other cases are discarded.

c) AdaBoost Learning:

We train an AdaBoost classifier with the collected samples which gives us information whether (Ci, Cj) is adjacent or not. AdaBoost algorithm is a machine learning algorithm. It is an algorithm for constructing a strong classifier as linear combination. Strong classifier is built by the combinations of many weak classifiers.

d) CC Clustering:

The AdaBoost algorithm yields a function. We use this function in clustering decisions. We can find all adjacent pairs by evaluating all possible pairs. Based on these adjacency relations, set of connected components is partitioned in to a set of clusters. After clustering, we have discarded clusters having only one CC.

2. CANDIDATE NORMALIZATION:

a) Geometric Normalization:

By the process of normalization we mean geometrically transforming the image into a standard form and shape.

b) Binarization of the images:

Now we perform the binarization of all the corresponding regions **.**There is only two possible values for each pixel in a binary image. Typically the two colors used for a binary image are black and white.

3. *NONTEXT FILTERING*:

Finally we develop a text/nontext classifier that rejects nontext blocks among normalized images to identify only the text in the images [16].

VII. ALGORITHM

The formulation of the algorithm is given below, using the following abbreviations: {*c*: Color, *p*: Position, *s*: Size, *t*: Text, 0 *t:* The initially detected text}.

STEP1: Detect the initial text cues, i.e., determine regions such that TextCue(c, p, s, t) = True 0. For each region, let SR(c p s t) = TextCue (c p s t);

STEP2: Determine the search region SR(c, p, s, t_0)in the neighborhood of *SR*0(c, p, s, t_0); If for each(c, p, s, t_0), there is no *SR* (c, p, s, t_0), Such as SR$(c, p, s, t_0) \supset SR0(c, p, s, t_0)$,

Let *SR*(c, p, s, t) =SR0(c, p, s, t_0), Go toSTEP5; otherwise, continue;

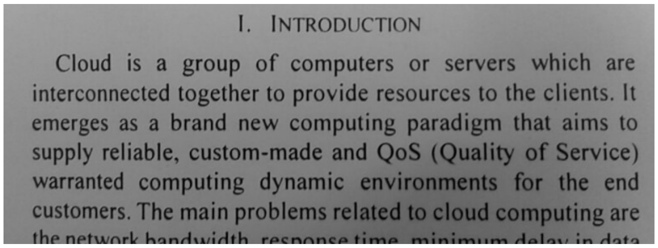
STEP3: Perform color modeling in SR(c, p, s, t_0), under the layout constraints, to extract characters or text regions in SR(c, p, s, t_0);

STEP4: Update the SR$(c, p, s, t_0)$to SR(c, p, s, t), *t* is the character/text in SR(c, p, s, t_0)extracted inSTEP3.Delete the search regions that have no characters or text in it, and let SR0(c, p, s, t_0)=*SR*(c, p, s, t_0), go to STEP2;

STEP5: Perform layout analysis on *SR*(c, p, s, t), output the detected text regions.

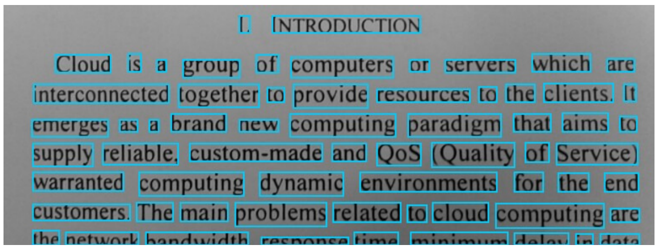
VIII. EXPERIMENTAL RESULTS & ANALYSIS

Performance of each algorithm are evaluated and compared in terms of precision rate and recall rate. False positives are the non-text regions in the image and have been detected by the algorithm as text regions. False negatives are the text regions in the image and have not been detected by the algorithm.


I. INTRODUCTION

Cloud is a group of computers or servers which are interconnected together to provide resources to the clients. It emerges as a brand new computing paradigm that aims to supply reliable, custom-made and QoS (Quality of Service) warranted computing dynamic environments for the end customers. The main problems related to cloud computing are the network bandwidth, response time, minimum delay in data

Fig 5: Before OCR


I. INTRODUCTION

Cloud is a group of computers or servers which are interconnected together to provide resources to the clients. It emerges as a brand new computing paradigm that aims to supply reliable, custom-made and QoS (Quality of Service) warranted computing dynamic environments for the end customers. The main problems related to cloud computing are the network bandwidth, response time, minimum delay in data

Fig 6: After OCR

Recognized text from image:

l. INTRODUCTION

Cloud is a group of computers or servers which are interconnected together to provide resources to the clients. It emerges as a brand new computing paradigm that aims to supply reliable, custom-made and 005 (Quality of Service) warranted computing dynamic environments for the end customers. The main problems related to cloud computing are
the nthm-k handuiirhh rncnnne- oim- minimum Anlau in data

IX. CONCLUSION

In this paper, we have discussed different text detection techniques that can be used for text detection from an image. With the help of this proposed application we can improve the recognition rate and also we can improve the recognition rate for vertical or extremely tilted text.

ACKNOWLEDGMENT

I would like to thanks Prof. Rishikesh Yeolekar for his enthusiastic encouragement and continuous support. I extend my thanks to Prof. Jayshree Ghorpade mam for her patient guidance and useful critiques of this work.

REFERENCES

[1] Hyung Il Koo, Member, IEEE, and Duck Hoon Kim, Member, IEEE," Scene Text Detection via Connected Component Clustering and Nontext Filtering", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 6, JUNE 2013

[2] Rodrigo Minetto ,Nicolas Thome , Matthieu Cord , Neucimar J. Leite and Jorge Stolfi ," SnooperText: A text detection system for automatic indexing of urban Scenes" Computer Vision and Image Understanding 2013

[3] Vandana Gupta and Kanchan Singh,"A Novel Approach for Detection and Extraction of Textual Information from Scanned Document Images and Scene Images", International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 10, October 2013

[4] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, Senior Member, IEEE," A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 3, MARCH 2011

[5] Jung-Jin Lee∗, Pyoung-Hean Lee∗, Seong-Whan Lee∗, Alan Yuille∗† and Christof Koch," AdaBoost for Text Detection in Natural Scene" International Conference on Document Analysis and Recognition

[6] Chucai Yi, *Student Member*, *IEEE and* YingLi Tian, *Senior Member*, *IEEE,* Aries Arditi," Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons" IEEE/ASME Transactions on Mechatronics

[7] Cong Yao, Xiang Bai, *Member, IEEE*, and Wenyu Liu, *Member, IEEE,"* A Unified Framework for Multi-Oriented Text Detection and Recognition", IEEE TRANSACTIONS ON IMAGE PROCESSING 2014

[8] Shangxuan Tian, Shijian Lu, Bolan Su and Chew Lim Tan," Scene Text Segmentation with Multi-level Maximally Stable Extremal Regions"

[9] Khyati Vaghela and Narendra Patel," AUTOMATIC TEXT DETECTION USING MORPHOLOGICAL OPERATIONS AND INPAINTING", International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 5, May 2013

[10] Shalin A. Chopra, Amit A. Ghadge, Onkar A. Padwal, Karan S. Punjabi and Prof. Gandhali S. Gurjar," Optical Character Recognition", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014

[11] Ravina Mithe, Supriya Indalkar and Nilam Divekar," Optical Character Recognition", International Journal of Recent Technology and Engineering (IJRTE) Volume-2, Issue-1, March 2013

[12] Prof. Amit Choksi, Nihar Desai, Ajay Chauhan, Vishal Revdiwala and Prof. Kaushal Patel," Text Extraction from Natural Scene Images using Prewitt Edge Detection Method", International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 12, December 2013

[13] Fathima A Muhammadali," Survey on Localizing Text in Scene Images" INTERNATIONAL JOURNA L FOR RES EARCH IN AP PL I ED SC IENC E AND ENGINEERING TECHNOLO GY (I JRAS ET) Vol. 2 Issue V, May 2014

[14] Rodrigo Minetto ,Nicolas Thome , Matthieu Cord , Neucimar J. Leite and Jorge Stolfi ," SnooperText: A text detection system for automatic indexing of urban

Scenes" Computer Vision and Image Understanding 2013

[15] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, Senior Member, IEEE," A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 3, MARCH 2011

[16] k. Sruthi nivetha, m. Surya, saany varghese, 4s.r.vidhya, 5p. Venkateswara rao," detection of scene text based on machine learning classifiers" Proceedings of 5th IRF International Conference, Chennai, 23rd March. 2014, ISBN: 978-93-82702-67-2

AUTHORS PROFILE

Miss. Aarti Arjun Andhale received the Bachelor's degree in Computer Engineering from MIT Collegeof Engineering and pursuing final year Master's degree in Computer Engineering from MIT College of Engineering, Pune.

Prof. Rishikesh Yeolekar is presently working as an Assistant Professor in the Department of Computer Engineering, MIT College of Engineering, Pune. He completed his B.E and M.Tech in Computer Engineering Department. He has 10 years of experience and His areas of interest are Networking, Operating System.