

Text Detection Tracking and Recognition in image using Connected Components

D.Kavya¹, A.Mohan Krishna²

¹B.TechInformation Technology
GMR Institute of Technology
Rajam, India
Kavyadevarasetty989@gmail.com

²B.TechInformation Technology
GMR Institute of Technology
Rajam, India
asapu.mohankrishna@gmail.com

Abstract: The study of image text is presently in wide demand because an image is a main source of afferent data in our lives. Text is a prominent and direct source of information in an image. But is a great challenge for us in order to detect, extract and recognize the text by seeing an image due to the difference in size, style, orientation, robustness to background complexity, text degradation and distortion, text variations, and moving objects. There are major contributions in order to detect, track, and recognize the text in an image. First one is a generic framework for text extraction. Second includes of methods and systems are summarized, compared and examined and final one is related applications, prominent challenges and future directions for text extraction. In order to extract the text, we have an algorithm based on two machine learning techniques(classifiers). One to make candidate word regions, other eliminates non-text regions. These educed CCs are divided into small parts known as clusters so that we can produce candidate regions. In additional to MSER algorithm we have AdaBoost classifier that discover the closeness relational and cluster CC's by their pair wise relations. Defines whether each region contains text or not.

Keywords: generic framework, candidate word regions, connected components(CC), maximally stable external region(MSER), clusters.

1. INTRODUCTION

Text Detection and Recognition and Tracking has emerged as an important problem in the past few years. Text in a image consists of valuable information and is used in many text-based image applications, such as text-based web image search, phone based text study and finding and tracking. Due to composite backdrop, and inequality of font size and color, text in natural scene images has to be esoteric detected before being admitted and recover. Surviving methods for text detection can be divided into three groups: Connected component based methods, and AdaBoost Classifier methods, Clustering techniques. Connected component based methods extract CCs from images by mapping components followed by combining character candidates into text; some other checks may be performed to eliminate the non-text regions. AdaBoost technique is for connecting the clusters by their pair wise relations. MSER algorithm is used for checking weather the connected clusters are connected correctly or not.

Literature survey

In this paper an algorithm based on two machine learning techniques is presented in-order to extract the text from an image. One is to generate the candidate text regions, and the other one eliminates the non-text regions. We withdraw connected components(CC's) in images by using the maximally stable external region(MSER) algorithm.

There are three major contributions in-order to detect, track, and recognize the text in a scene or in a image. First one is a generic framework for text extraction. Second consists of methods, systems are summarized, compared, analyzed and final one is related applications and future directions for text extraction. This method is developed by state estimation into a connected component (CC)-based framework. We extract CCs with the maximally stable external region algorithm. Finally, a text-line/non-text-line classifier is trained and non-text candidates are filtered out with the classifier. A fast and successful pruning algorithm is plan to withdraw Maximally Stable External Regions (MSERs) as character candidates using the scheme of decreasing regular differences. The text applicants with high non-text prospect are removed and texts are discovered with a text classifier.

2. RELATED WORK

Almost all of the text detection algorithms are to be classified into two types: text-based and connected component (CC) – based. Text-based technique view text as a special text that is evolved from the text image background. As collate to this method, the CC-based near remove regions from the image and uses geometric control to filter out non-text regions. After the CC extraction, thos method eliminates non-text part. The region-based methods have absorbed on binary classification that consists text verses non text part of a small image parts. In this method,

firstly an image is separated and word text components are removed. After eliminating the non-text regions through different techniques. Recently, Maximally Stable External Regions (MSER) algorithm have become one of the commonly used region detecting skill because of their high replay and also because they are somewhat correlated to many other commonly used detectors. MSER based methods have demonstrated very promising complementary performance in many real projects.

3. Text Detection, Tracking and Recognition

In image data grasping, spatial-temporal analysis and multi-frame integration are common plans. Compared to video, the temporal information is helpful in improving text detection and recognition. Correspondingly, using spatial and temporal information acquired from multiple frames, image text tracking, tracking based detection, and tracking based recognition methods are comprehensively surveyed and highlighted in this section. The categorizations of these methods are summarized, where text tracking methods are divided into subgroups based on typical tracking strategies, while tracking based detection and recognition techniques are categorized with multi-frame integration fusion strategies.

3.1 TEXT TRACKING

The goal of text tracking is to continuously determine the location of text across multiple dynamic images. Text tracking is useful for verification, integration, enhancement and speedup in image text detection and recognition. In recent years, a variety of text tracking methods have been explored in the literature. We roughly categorize these text tracking methods into two groups: tracking with detection and refinement by recognition. For tracking with detection, objects or detected text positions are used to track text across consecutive frames. The methods in this group can be further divided into several subgroups based on the typical tracking strategies i.e., text tracking with template matching, with particle filtering, and with tracking-by-detection.¹We also discuss text tracking in the compressed domain and text tracking with other techniques.

3.1.1 TRACKING WITH TEMPLATE MATCHING

The template matching method attempts to answer some variations of the following question: Does the image contain a specified view of some features and if so, where? As one of the conventional methods of text tracking, it implements tracking by seeking the most similar region in the image compared with a thesaurus image. Therefore, feature representation, and similarity calculation and (matching) search are the two important factors for designing a template matching algorithm.

3.1.2 TRACKING WITH PARTICLE FILTERING

Particle filtering, also known as the Sequential Monte Carlo Method, is a nonlinear filtering technique that recursively

estimates a system's state based on available observations. It solves the problem that state variables do not follow a Gaussian distribution. Particle filters are popular in computer vision, especially for object detection and tracking. Thus, they are widely used in video text tracking. The main steps in particle filtering include feature extraction, the observation model and the sampling scheme. Here, we summarize video text tracking techniques with particle filtering into these subgroups.

3.1.3 TRACKING WITH TRACKING-BY-DETECTION

The tracking-by-detection method associates detected results across successive frames to create the appearances of objects, namely, estimating the tracking trajectories using text detection results. When Compared to other tracking methods, tracking-by-detection successfully solves the re-starting the problem even when the object is unfortunately lost in some images. It also eliminates immoderate model gather due to the near aspects of unlike objects. Here, we discuss tracking-by-detection methods in a view with the features used for region matching moving from simple to complicated, e.g., edge maps, character strokes and MSERs.

3.2 TRACKING BASED DETECTION

Conventional methods for image text detection mainly focus on detecting text in each individual frame or in some key frames. However, these methods cannot obtain high detection accuracy due to complex backgrounds, poor contrast, and degraded text quality caused by lossy compression. Thus text tracking techniques are introduced in the detection process to reduce false alarms and improve the accuracy of detection. We call these strategies tracking based detection methods. Commonly, these methods can be categorized into temporal-spatial information based methods and fusion based methods. The former methods directly use temporal or spatial information to remove noise. The latter methods merge detection and tracking results over multiple frames, or combine detection and tracking results in a single frame to improve the detection power.

3.2.1 TEMPORAL-SPATIAL INFORMATION BASED METHODS

The secular and structural information is directly used to lessen false alarms in image text extraction, such as the time period of the text, the interval between the starting frame and the ending frame of the same text.

3.3 TRACKING BASED RECOGNITION

In the literature, most proposed methods for text recognition mainly focus on recognizing text in a single image. OCR engines and other commercial recognition softwares have been widely used for recognizing text in images. However, they cannot always achieve good performance in each individual image frame due to complex backgrounds, low contrast, and poor resolution. Fortunately, because tracking techniques have been introduced in image text extraction methods and the same text usually has the

same ID label in tracking results, MFI methods can make use of temporal redundancy to obtain a frame with a cleaner background, higher contrast, and better resolution. This is a fast and effective way to improve the accuracy of text recognition.

3.3.1 IMAGE ENHANCEMENT

Image enhancement technology uses selection-based or integration-based strategies to obtain a high-resolution image from the text regions that contain the same text. The selection-based methods choose the text region with highest resolution from the text regions in consecutive frames. The integration-based methods, in contrast, obtain a high-resolution image through image fusion, which mainly includes multi-frame averaging, time-based minimum/maximum pixel value searching, and Boolean “And” operations.

4. PROBLEM DOMAIN

Although many research efforts have been made to detect text regions from natural scene images, more robust and effective methods are expected to manage the difficulties in size and background. The executed framework is able to accurately detect text clusters in different sizes, orientations, colors and some variations of sizes shape of attachment bases. Compared with the existing methods which focus on individual analysis of single character, the text string structure is more powerful to distinguish background connections from text. Also used to notice whether the connected components belong to text nature or unforeseen sounds.

5. PROPOSED METHOD

The proposed method focus on image text detection from different text images. After the detection the obtained text image which is free from complicated background, variation in font, size and orientation, color twist and noise. This method consists of three steps they are Candidate generation, Candidate normalization and non-text eliminating. The first method is based on popular CC based procedures which consists of a MSER based and an AdaBoost-based CC clustering method. The maximally stable external region (MSER) algorithm is unedited to scales changes and related power changes and other blocks in our method are also designed to be different to these changes. In our method, both problems are labeled based on machine learning techniques, so that our method is largely free from problem solving. In training, we have selected efficient features and trained the classifier with the AdaBoost algorithm. We apply the CC clustering method to fresh CC sets that usually contains many non-text CCs and some candidate may communicate to non-text clusters. From that normalized image we built a binary image. Filtering means eliminating a text/non-text classifier in an images.

5.1 CANDIDATE GENERATION

In the discussed method for producing the characters, extract components in images and divide the extracted components into set of pieces, using the clustering algorithm is based on the adjacency relation classifier. After the detection the required text in an image which is free from complicated background in size and color differences. The proposed method consists of some steps. They are candidate generation, normalization and non-text eliminating.

5.1.1 CCs EXTRACTION

From all the component methods, we have taken the MSER algorithm because it shows good performance with the less cost. This algorithm can be considered as a process to find local binarization results that are fixed over a range of entry, and this property allows us to find most of the text components.

5.1.2 AdaBoost CLASSIFIER and CC CLUSTERING

For these operations we shall define some local properties of CCs. All of these features are informative and consider each feature as a weak classifier. From these weak classifiers, build a strong classifier with the AdaBoost learning algorithm. The AdaBoost is easy to execute and known to show good performance in many request.

5.2 CANDIDATE NORMALIZATION

After CC clustering there are set of pieces. In this section, we will normalize corresponding parts for the authentic text/non-text classification.

5.3 TEXT/NON-TEXT CLASSIFICATION

In order to get final results we develop a text/non-text classifier that unaccepted non-text parts among organized images. In our classification, we do not acquire experienced techniques such as cascade structures, since the number of samples to be classified is usually small. One possible proceed towards to this problem is to separate the organized images into spot covering one of the letters and develop a character/non-character classifier. However, character segmentation is not an easy problem and there are examples where this approach may fail. Finally, decision results for all square blocks are integrated so that the original block is classified. In this section, we first present our training method that allows us to have a textness value for each square. Then, we explain our text/non-text classification method for normalized images.

6. RECENT TECHNOLOGY

Recent technology used is Single Link Clustering algorithm. It is based on grouping clusters in bottom-up fashion, at every step merging of two pices contains the nearest pair of elements are not belonged to the same piece each other. A disadvantage of this method is that it lean to produce long thin clusters in which close elements of the

same pieces have small distances, but elements at opposite ends of a pikes may be much far from each other than two elements of other pieces. This may lead to trouble in solving classes that could usefully separate the data.

Natural Scene Images”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 5, MAY 2014

7. FUTURE WORK

Our future work will focus on developing learning based methods for removing text from composite backgrounds and text localization for OCR recognition. We also attempt to improve the effective planning and transfer the algorithms into a navigation system prepared for the finding of visually damaged people.

8. CONCLUSION

We have presented in this paper an improved text string detection method which can effectively detect text from the document background. Due to the unknown text arrival and complex backgrounds, text detection in natural scene images is still an unsolved problem to locate text regions fixed in those images. We have grant a novel image text detection algorithm with the help of two machine learning classifiers: one is for candidate generation and the other one eliminates the non-text parts.

REFERENCES

- [1] Y. Zhu, C. Yao, and X. Bai, “Scene text detection and recognition: Recent advances and future trends,” *Frontiers Comput. Sci.*, vol. 10, no. 1, pp. 19–36, Feb. 2016.
- [2] Q. Ye and D. Doermann, “Text detection and recognition in imagery: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1480–1500, Jul. 2015
- [3] H. I. Koo and D. H. Kim, “Scene text detection via connected component clustering and nontext filtering,” *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2296–2305, June 2013.
- [4] X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, “Robust text detection in natural scene images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 970–983, May 2014.
- [5] H. Zhang, K. Zhao, Y.-Z. Song, and J. Guo, “Text extraction from natural scene image: A survey,” *Neurocomputing*, vol. 122, pp. 310–323, Dec. 2013.
- [6] Xu-Cheng Yin; Ze-Yu Zuo; Shu Tian; Cheng-Lin Liu”Text Detection, Tracking and Recognition in Video: A Comprehensive Survey”, Year: 2016, Volume: 25.
- [7] Hyung Il Koo”Text-Line Detection in Camera-Captured Document Images Using the State Estimation of Connected Components”, VOL. 25, NO. 11, NOVEMBER 2016.
- [8] Xu-Cheng Yin, Member, IEEE, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao “Robust Text Detection in