Acta Scientific COMPUTER SCIENCES

Volume 4 Issue 11 November 2022

Research Article

Analysis of Segmentation for Text Components in Context of Scene Images Utilizing Connected Components

Mr Jhuntu and Tathagata Roy Chowdhury*

Department of Computer Science and Engineering, St. Mary's Technical Campus Kolkata, Kolkata, West Bengal, India

***Corresponding Author:** Tathagata Roy Chowdhury, Department of Computer-Science and Engineering, St. Mary's Technical Campus Kolkata, Kolkata, West Bengal, India.

Received: September 23, 2022

Published: October 13, 2022

Abstract

Images of nature scenes that contain language contain useful information, such as text-based landmarks, etc. There are several steps involved in text extraction from scene photos. To get effective results, each stage is equally vital. The steps of text detection, localization, segmentation, and recognition are crucial. Due to differences in size, position, and alignment from one image to the next, extracting text from scene photographs is exceedingly challenging. The challenge of extracting text from scene photos is difficult because of all these issues. However, the placement of text in such images is arbitrary and is not restricted to any particular page layout. We used the dataset made available in conjunction with IIIT5K to assess how well the planned "segmentation of scene text images using connected component analysis" performed. Since each image comprises roughly four characters in various font styles and font sizes. A disjunct categorization does not seem to be possible, nevertheless, as even two segmentation procedures that are utterly unrelated can share traits that defy singular categorization 1. As a result, the classification being presented is one that takes into account an approach's emphasis rather than being a perfect split.

Keywords: Image Processing; Connected Components; Image; Binary Image; Segmentation; Grey Scale

Introduction

In digital image analysis, picture segmentation is a task of basic importance. It is the division of a digital image into several sections. The ability to separate text from visual documents has numerous useful applications, including document retrieval, object recognition, licence plate recognition, and more. In recent years, it has become a very prominent research area. Document analysis and recognition (DAR) has a long history in the science of computer vision, spanning more than 40 years. This period alone demonstrates the complexity of the subject. DAR has rapidly grown as a result of the continuous addition of new subfields for cutting-edge study. This increases interest in the evolution process. Technologies for photocopying, printing, scanning, and picture capturing are the biggest challenges in the sector. Therefore, the concept of fully comprehending the content of a document is astonishing and is still challenging for a machine to accomplish, in contrast to the human brain. Numerous other uses have been created by researchers as a result of technological developments like photocopying and Xerox, which changed the document industry. This ability is not available on the photocopier. Optical character recognition (OCR) engines may now successfully complete this task by producing documents.

In scene text pictures, text and image segmentation can be done in a number of ways, including the region-based technique and the texture-based approach. With the region-based technique, each pixel in the image is taken into account and given a specific region or item to belong to. This strategy essentially falls into two categories: connected component-based and edge-based. Either region-based or texture-based approaches can be used to identify the text-regions in a document image. However, they struggle with complicated images with irregular backgrounds. For instance, if a text string touches a graphical item in the original image, they might form one connected component in the resulting binary image.

The basic premise of edge-based algorithms is that text symbols often have stronger edges than noise, textured backgrounds, and other graphical elements. These top-down methods start by creating a binary edge picture with an edge detector, and then they use morphological operations or other algorithms, like run-length smoothing, to connect the adjacent edges. The candidate text areas, each of which represents either many merged lines or a graphical element, are connected elements of the resulting image. Then, by examining its vertical and horizontal projection profiles, each component is divided into smaller sections. Finally, each of the smaller regions satisfying particular heuristic constraints is labelled as text.

Discuss current related work here. Author Hongliang., *et al*. [1] describe in this study an effective method for locating, extracting, and identifying each segmented character from a licence plate. The suggested model may be broken down into four components: picture digitization, edge detection, character separation, and template matching. To remove non-license plate regions and increase solely the plate region, morphological techniques with structure elements (SE) were applied. Connected Component Analysis is used for character segmentation. Character recognition is accomplished using a correlation-based template matching technique. Chinese text recognition in photos and videos is shown by Bai., *et al.* in their article [2]. The approach differs from the one currently in use, which binarized text pictures and passed the findings to an OCR. The suggested method accomplishes recognition on grey pixels directly, followed by segmentation, generating a recognition graph, reading Chinese characters, and determining the beam search. The benefits are that it is independent of binarization performance, which is imperfect in practise and lowers OCR performance, and because grayscale images provide more information about the text, which enhances recognition rates. Using a discrete wavelet transform (D.W T) technique, Neha., *et al*. [3] extract text information from complex backgrounds. A grayscale or colour image could be used as the input. The edges on each sub-image are discovered using the Sobel edge detection method.

The outcome is seen as forming an edge map. The edge map is subjected to morphological operations in the following stage, and further thresholding is used to boost efficiency. According to Dutta., *et al*. [4], the method is based on edge map selection and gradient information. The approach begins by determining the gradient of the image and then improving the gradient information. The improved gradient image was binarized in the following phase, and edges were chosen by intersecting the edge map with the enhanced gradient image's binary information. A clever edge detector is employed to produce the edge map. The chosen edges are then morphologically dilated, opened, and used for text areas using the appropriate structuring components.

An study of the projection profile is done to determine where the text regions' borders are. Grayscale transformation and smoothing with a median filter were utilised as preprocessing procedures by Sivasankaran., *et al*. Weak edges are removed using the Canny edge detection and Gaussian filter methods. Text part is retrieved further using dilatation and related component labelling algorithms. The proposed study by Angadi., *et al*. [6] utilises the discrete cosine transform and is based on texture analysis (D.C.T). This eliminates comparable backdrop using a high pass filter. Following the application of the resulting texture characteristics to each 50*50 block of the input, discriminant functions are used to determine which blocks of strong text are present. Finally, the extracted text sections are appended to the recognised text blocks.

Wei., *et al.* [7] the authors used a pyramidal concept to detect text in video images with variations of background, size of text font and colour. In the first step, two downsized images are obtained from the original image. Then, the gradient difference is calculated for three differently sized images. k-means clustering procedures are applied to separate the pixels. Next, determine the boundaries of candidate text regions using projection profile analysis. Finally, text candidates are identified using two verification phases. One is geometric properties. Another is text candidate using D.WT. To reduce the number of dimensions of these features principal component analysis is used. SVM is used to classify the text and non-text. The literature on text segmentation is broad in scope but there appears to be very little literature on using machine learning techniques on this subject. The text segmentation algorithm should be able to adapt and learn, but this limits the method's applicability because learning typically takes a lot of time and training data. M. M.

Haji and S. D. Katebi provide a straightforward method for creating training data from manually segmented photos and then applying a Naive Bayes Classifier (NBC), which is quick in both the training and application phases, to solve these issues. Punam Thakare has suggested A Study of Image Segmentation and Edge Detection Techniques. In this work, various picture segmentation methods, including edge-based, region-based, and integrated methods, were reviewed. The findings demonstrate that the type of image and its underlying truths have an impact on the recognition rate [10].

Danial Md Nor, Rosli Omar, and M. Zarar M. Jenu suggest a method for extracting textual data. The issue of textual information extraction from photographs of presenting scenes is addressed in this work. Three processes make up the suggested method for extracting textual information: image segmentation, text localisation and extraction, and optical character recognition. The quality of the OCR images or documents has a significant impact on the outcomes [11]. A method developed by Neha Gupta and V.K. Banga for extracting text from photos, such as document images, scene images, etc. Discrete wavelet transform (DWT) is used in this study to separate text content from complicated images. A colour or grayscale image may be used as the input image. If the image is colour image, then preprocessing is required. For extracting text edges, the sobel edge detector is applied on each sub image. The resultant edge so obtained are used to form an edge map. Morphological operations are applied on the processed edge map and further thresholding is applied to improve the performance [12].

Related data set

ICDAR 2003

Images were captured with a variety of digital cameras by each of the Essex authors. Cameras were used with a range of resolution and other settings, with the particular settings chosen at the discretion of the photographer. To allow management of the ground truthing or tagging of the images, and with a view to possible future tagging jobs, we implemented a web based tagging system. This operates along similar lines to the OpenMind concept. People working as taggers can log in to the system from anywhere on the Internet using a Java (1.4) enabled web browser. On logging in, a Java Applet window appears and presents a series of images. The tagger tags each image by dragging rectangles over words, and then typing in the associated text. The applet then suggests a possible segmentation of the word into its individual characters, which the tagger can then adjust on a character-by-character basis.

The region's slant and rotation can also be changed by the tagger. The marked rectangles are all transmitted back to a server and recorded in a database when the tagger clicks submit after finishing an image. The number of taggers that should tag each image is one of the system's parameters. If there was a lot of tagging effort available, we could send each image to numerous taggers and then accept all the images with tags that were largely in agreement. However, this is a bit of a waste of the tagging effort as checking an image takes far less time than tagging it. As a result, we implemented a two-tier system of taggers and checkers, where the job of a checker was to approve a set of tags.

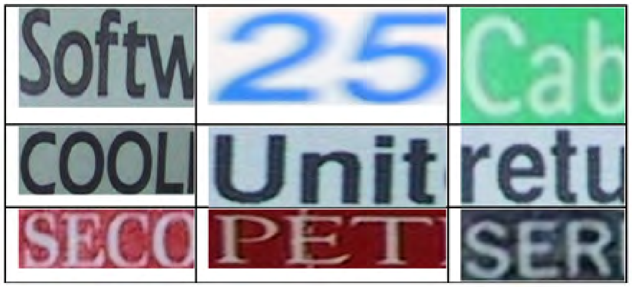

Figure 1: The dataset of ICDAR 2003.

SVHN data set

SVHN is a real-world image dataset for developing machine learning and object recognition algorithms with minimal requirement on data preprocessing and formatting. It can be seen as similar in flavor to MNIST (e.g., the images are of small cropped digits), but incorporates an order of magnitude more labeled data (over 600,000 digit images) and comes from a significantly harder, unsolved, real world problem (recognizing digits and numbers in natural scene images). SVHN is obtained from house numbers in Google Street View images.

10 classes, 1 for each digit. Digit '1' has label 1, '9' has label 9 and '0' has label 10.

73257 digits for training, 26032 digits for testing, and 531131 additional, somewhat less difficult samples, to use as extra training data.

Comes in two formats:

- Original images with character level bounding boxes.
- MNIST-like 32-by-32 images centered around a single character (many of the images do contain some distractors at the sides.

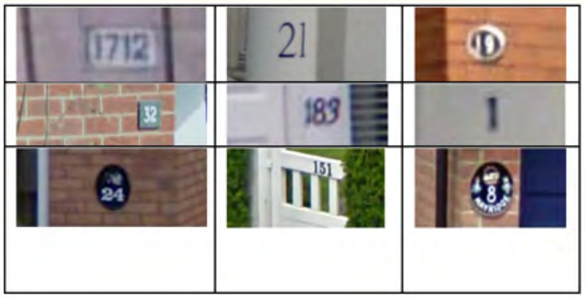

Figure 2: The data set of SVHN.

Segmentation

The process of dividing a picture into a number of related groupings of pixels is known as image segmentation. Image analysis, object representation, visualisation, and many other picture processing tasks frequently need the segmentation of an image into meaningful components. Since there are so many different segmentation techniques that have been developed throughout the years, categorization is required in order to appropriately display the techniques here. However, a disjunct categorization does not appear to be feasible since even two quite unlike segmentation techniques may share characteristics that resist singular categorization 1. Therefore, rather than being a precise split, the categorization that is being offered is a categorization addressing the emphasis of an approach. The following categories are used:

- **Threshold based segmentation:** Histogram thresholding and slicing techniques are used to segment the image. They may be applied directly to an image, but can also be combined with pre- and post-processing techniques.
- **Edge based segmentation:** With this technique, detected edges in an image are assumed to represent object boundaries, and used to identify these objects**.**
- **Region based segmentation:** Where an edge based technique may attempt to find the object boundaries and then locate the object itself by filling them in, a region based technique takes the opposite approach, by (e.g.) starting in the middle of an object and then "growing" outward until it meets the object boundaries.
- **Clustering techniques:** Although clustering is sometimes used as a synonym for (agglomerative) segmentation techniques, we use it here to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. In this context, clustering methods attempt to group together patterns that are similar in some sense. This goal is very similar to what we are attempting to do when we segment an image, and indeed some clustering techniques can readily be applied for image segmentation.
- **Matching:** When we know what an object we wish to identify in an image (approximately) looks like, we can use this knowledge to locate the object in an image. This approach to segmentation is called matching.

Experimental Results

Segmentation experimental result

For evaluating the performance of the proposed the "segmentation of scene text images using connected component analysis", we used the dataset made available with the occasion of IIIT5K. As each image contains approximately four characters of different font styles of different font size. The result of the system for 300 images is given in the table (1).

Data set	No of input images	No of segmented images	Percentage
IIIT5K	300	243	81%
ICDAR2003	12897	10935	84.78%

Table 1

Limitation

In this paper, we have verified different alphabetical order having different names those are segmented properly, when we are using Uppercase Syntax images, we are not facing any problem, but when the lowercase checked some of them running properly, but particularly 'i' and 'j' they have connected two dots up of the letters that's can't be segmented and they show very different components, a dot component and another design components.

And also in similar way if two or more components are merged or Overlap each other, then it also not possible to segment again cause as they are merged they looses their original shape and design.

Conclusion

The proposed method, which differs from existing algorithms, is called "Segmentation for Text Components in context of Scene images utilizing Connected Components". ICDAR2003 and IIIT5K datasets were used. We tested our method using more than 300 scene photos, and the segmentation rate was 81.00%. We also tested the ICDAR, and the rate was approximately 84.78%. However, because to the variety of scene photographs, segmentation works best for pictures with a straightforward font style, a moderate range of intensity, and a straightforward background. The focus of future study will be on achieving improved accuracy and addressing complicated background and light intensity variance.

Bibliography

1. Hongliang Bai and Liu Changping. "A hybrid license plate extraction method based on edge statistics and morphology". Pattern Recognition, ICPR. Proceedings of the 17th International Conference on. IEEE 2 (2004).

2. Bai Jinfeng., *et al*. "Chinese Image Character Recognition Using DNN and Machine Simulated Training Samples". Artificial Neural Networks and Machine Learning–ICANN Springer International Publishing, (2014): 209-216.

3. Gupta Neha and V K Banga. "Image Segmentation for Text Extraction". Proceedings of the 2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012), Singapore, April 28-29. (2012).

4. Dutta A., *et al*. "Gradient based Approach for Text Detection in Video Frames 1 (2009).

5. Sivasankaran V., *et al*. "Recognition of Text in Mobile Captured Images Based on Edge and Connected Component Hybrid Algorithm". *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)* 3.6 (2014): 358.

6. Angadi S A and M M Kodabagi. "Text region extraction from low resolution natural scene images using texture features". Advance Computing Conference (IACC), IEEE 2nd International. IEEE (2010).

7. Wei Yi Cheng and Chang Hong Lin. "A robust video text detection approach using SVM". *Expert Systems with Applications* 39.12 (2012): 10832-10840.

8. D Doermann., *et al*. "Progress in camera based document image analysis". in Proc. 7th Int. Conf. on Document Analysis and Recognition 1 (2003): 606-687.

9. S M Lucas., *et al*. "ICDAR 2003 robust reading competitions". in Proc. 7th Int. Conf. on Document Analysis and Recognition 2 (2003): 682-687.

10. Trier and A Jain. "Goal directed evaluation of binarization methods". IEEE Trans. Pattern Anal. Machine Intell, 17 (1995): 1191-1201.

11. C Wolf., *et al*. "Text localization, enhancement and binalization in multimedia document". in Proc. 16th Int. Conf. on Pattern Recognition 2 (2002): 1037-1040.

12. S Wu and A Amin. "Automatic thresholding of gray-level using multi-stage approach". in Proc. 7th Int. Conf. on Document Analysis and Recognition 1 (2003): 493-497.

13. A Miene., *et al*. "Extracting textual inserts from digital videos". in Proc. 6th Int. Conf. on Document Analysis and Recognition (2001): 1079-1083.

14. A Sato. "A learning method for definite canonicalization based on minimum classification error". in Proc. 15th Int. Conf. on Pattern Recognition 2 (2000): 199-202.

15. M Mori. "Video text recognition using feature compensation as category-dependent feature extraction". in Proc. 7th Int. Conf. on Document Analysis and Recognition 2 (2003): 645-649.

16. P Simard., *et al*. "Efficient pattern recognition using a new transformation distance". Advances in Neural Information Processing Systems, Morgan Kaufmann, 5 (1993): 50-58.

17. T Wakahara., *et al*. "Affine-invariant recognition of gray-scale characters using global affine transformation correlation". IEEE Trans. Pattern (2001).