Text To Speech Using MSER and OCR

Kunche Satya Rahul[1] | Devulapalli Venkata Ramana[2] | Bikkina Rishitha[3] | Eduri Devisri[4] | DevarapalliAnudeep Reddy[5]

[1]Assistant Professor, Department of Electronics and Communication Engineering, Godavari Institute of Engineering and Technology (A), Rajahmundry, Andhra Pradesh, India

[2,3,4,5] UG Scholar, Department of Electronics and Communication Engineering, Godavari Institute of Engineering and Technology (A), Rajahmundry, Andhra Pradesh, India

Abstract: This paper presents an in the optical character recognition (OCR) system. It introduces general architecture of modern OCR systems, discussing each module in detail. Specific contribution of this paper is novelty of the character extraction and segmentation, by considering them as image features. MSER (Maximally Stable Extremal Regions) feature detector is applied, discussing numerical and practical restrictions for character segmentation and recognition. The neural network is trained for character recognition. Detection and recognition of text from any natural scene image is challenging but essential extensively for extracting information from the image. In this paper, we propose an accurate and effective algorithm for detecting enhanced Maximally Stable Extremal Regions (MSERs) as main character candidates and these character candidates are filtered by stroke width variation for removing regions where the stroke width exhibits too much variation. For the detection of text regions, firstly some preprocessing is applied to the natural image and then after detecting MSERs, an intersection of canny edge and MSER region is produced to locate regions that are even more likely to belong to text. Finally, the selected text region is taken as an input of a novel Optical Character Recognition (OCR) technique to make the text editable and usable

KEYWORDS: Maximum Stable Extremal Region, Optical Character Recognition, .Net Assembler

Check for updates

DOI of the Article: https://doi.org/10.46501/GIETEC20

Available online at: https://ijmtst.com/icetee2021.html

As per UGC guidelines an electronic bar code is provided to seure your paper

To Cite this Article: Kunche Satya Rahul; Devulapalli Venkata Ramana; Bikkina Rishitha; Eduri Devisri and DevarapalliAnudeep Reddy. Text To Speech Using MSER and OCR. *International Journal for Modern Trends in Science and Technology* 2021, *7*, pp. 112-117. https://doi.org/10.46501/GIETEC20

Article Info. Received: 18 May 2021; Accepted: 25 June 2021; Published: 30 June 2021

INTRODUCTION

Optical character recognition(OCR) represents numerous techniques for conversion of visual data of text (images, videos, scans, etc.) into machine-encoded character information. It is considered as the main technique for text computerization from a typewritten, printed or handwritten text. There are many needs for text computerization: electronic data searching, biometrics, identification, text storage optimization, machine translation, text-to-speech translation, data mining, etc. OCR is nowadays present in everyday use, in conversion of scanned documents to electronic document formats (doc, rtf, odt, pdf), post packages sorting, medical recipes processing, automated toll systems, etc. The concept of Maximally Stable Extremal Regions(MSERs) was proposed by Matas et al. MSERs denote a set of distinguished regions that are detected in gray scale image. All of these regions are defined by an extremal property of the intensity function in the region and on its outer boundary. MSERs have properties that form their superior performance as stable local detector. The set of MSERs is closed under continuous geometric transformations and is invariant to affine intensity changes. Furthermore MSERs are detected at different scales.

STRUCTURE OF PAPER

The paper is organized as follows: In Section 1, the introduction of the paper is provided along with the structure, important terms, objectives and overall description. In Section 2 we discussed literature survey. In Section 3 we discussed exciting method .In Section 4 detailed explain about proposed for MSER and OCR. In Section 5 text detection using MSER and stroke width. In Section 6 detailed information of recognition. Section 7 gives the conclusion of the entire process.

OBJECTIVES

The main objective of text to speech based on text detection and recognition using MSER and OCR. We have conducted the experiments on following objectives. To access with a user interface to select a image from files or webcam. After that the image is detected with text detection technique with various methods. After that the detected texts is recognized and get audio output. .

LITERATURE SURVEY

Many methods have been proposed to measure and enhance the Perceived standard of quality. The computerization widespread, digitalization of text data imposed OCR to wider community of users. US Armed Forces Security Agency developed a system for the conversion of printed messages into the machine language for computer processing, called “Gismo”. Later, in 1950’s, the same inventor founded. Evaluations of Mikolajczyk and Schmid , as well as Fraundorfer and Bischof revealed that the Maximally Stable Extremal Region (MSER) detector from Mataset al. Performs best on a wide range of test sequences.MSERs denote a set of distinguished regions, which are de-fined by an extremal property of its intensity function in the region and on its outer boundary. In addition, MSERs have all the properties required of a stable local detector.This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms.

EXISTING METHOD

Mathematical Morphology is a theory and technique for the analysis and processing of geometrical structures.Mathematical morphology is commonly applied to digital images, but it can employed as well on graphs, mesh and solid spatial structures.A shape(in blue) and its morphological dilation(in green) and erosion(in yellow) by a diamond shaped structuring elements.Mathematical morphology was originally developed for binary images and was later extended to gray scale image. Mathematical morphology is mainly based on erosion, dilation, opening and closing.

EROSION:

Erosion is a basic operator in the area of mathematical morphology.

The basis effect of erosion on a image is to erode away the boundaries of regions of foreground pixels, white pixel.

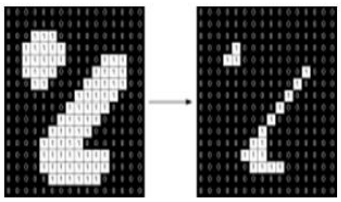


Figure3.1:Erosion

DILATION:

Dilation is a basic operator in the area of edge detection in mathematical morphology.

The basic effect of dilation on a digital image is to gradually enlarge the boundaries of region of foreground pixels.

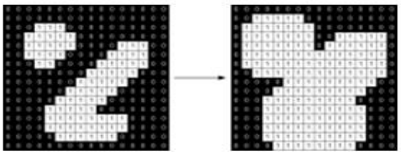


Figure3.2:Dilation

OPENING:

Opening is derived from the fundamental morphology operation of erosion and dilation.

It is tends to remove some of the foreground pixel from the edges of the region in digital image.

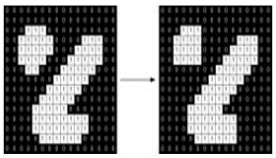


Figure3.3:Opening

PROPOSED FOR MSER AND OCR

Process of optical character recognition is complex and sequentially conditioned. Typical OCR system provides textual data at the output, for image (scan) data at the input (Fig 1). As the quality of an input image is unknown in advance, OCR system has to preprocess the input image in order to reduce noise and improve the image quality. OCR system module charged for preprocessing must guaranty quality of the image required by the next module. When the image can be considered of satisfying quality, the series of morphological operations are applied in order to extrude structured formations from it. The task of the next system module is to distinguish character data between all structured data. At this point, character data are extruded from a text. The final module provides recognition of a character, assigning its predefined class of the output. Connected to characters, those types of structural elements are unwanted, because they change a shape of the character element, producing more false positives in the character recognition. These types of elements are removed with morphological closing performed with structural elements similar to non-desired type (line, dot) or during the process of character segmentation.

The main contribution of this paper is efficient tracking of MSERs through an image sequence. Including information from the MSER detection result of the preceding image, the computational time and, in addition, the stability of the track can be improved significance. This section describes the main ideas behind the tracking concept. The algorithm starts with the analysis of the entire image It at time t, which results in a detection of MSERs for this image. Then every detected MSER of image It is tracked by performing the two following steps on the image It+1.First, a region of interest (ROI) of predefined size, centered around the center of mass of the MSER to be tracked, is propagated to the next frame. If a motion model is available it can be incorporated here. Then the component tree for this ROI is built in quasi-linear time by the algorithm presented in section 2. Second, the entire tree is analyzed and the node which best fits to the input MSER is chosen as the tracked extremal region representation. It has to be pointed out, that this step considers all extremal regions, not only maximum ones, which is the reason for the increased stability of the tracking algorithm.

TEXT DETECTION FOR MSER AND OCR

Our text detection method is slightly different from the traditional Maximally Stable Extremal Regions (MSERs) method. We proposed an enhanced MSER detection technique to locate the position of text in the image and Optical Character Recognition (OCR) technique is applied to this selected text part of the image. The complete process of our detection The proposed enhanced MSER based method of text detection includes the following steps:

5.1.1 MSER region detection: Normally, text characters usually have consistent color. So we start to find the text by selecting the regions of similar intensities by using MSER region detector

regions = detectMSERFeatures(I) returns an MSERRegions object, regions, containing information about MSER features detected in the 2-D grayscale input image, I. This object uses Maximally Stable Extremal Regions (MSER) algorithm to find region Connected component structure, returned as a structure with four fields. The connected component structure is

useful for measuring region properties using the regionprops function.

MSER Region Detection

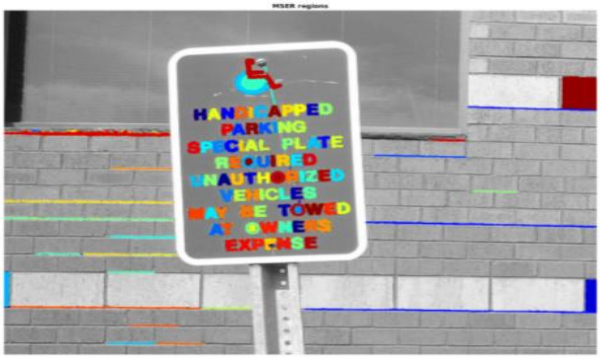


Figure5.1:Text Detection

5.1.2. Intersection of canny edge with MSER region: Canny edge detection algorithm performs a high response to edge detection. And intersection of MSER and canny edge produce the region that is likely be text. By using the region properties, some connected component can be removed. According to the variation of different front, image size, or languages the filtering thresholds.

Figure5.2:MSER edge canny

5.1.3. Visualization of text candidate's stroke width: Character in almost all language have a similar thickness throughout or stroke width. After this step the region where the stroke width contains too much variation is eliminated.

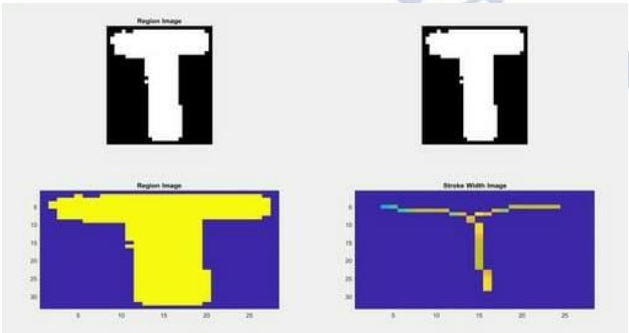


Figure5.3:Visualization of Letter Stroke

5.1.4. Text candidate after stroke width filtering: Non text region can be eliminated by determining a large variation in stroke width.

5.1.5. Image region under mask created by joining individual characters: Then the individual component is merged to compute a bounding box of text region. Morphological closing is done here algorithm.

5.1.6. Text region: Finally text region of the image is detected efficiently. text part of the natural scene image is shown which is found by our algorithm.

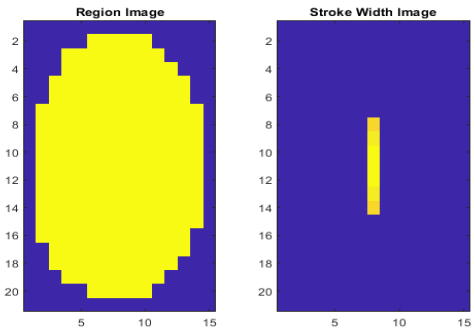


Figure5.4:Stroke Width Variation

5.1.7. Optical character recognition technique applied on the text part: To recognize the text Optical Character Recognition technique is activated on the achieved.

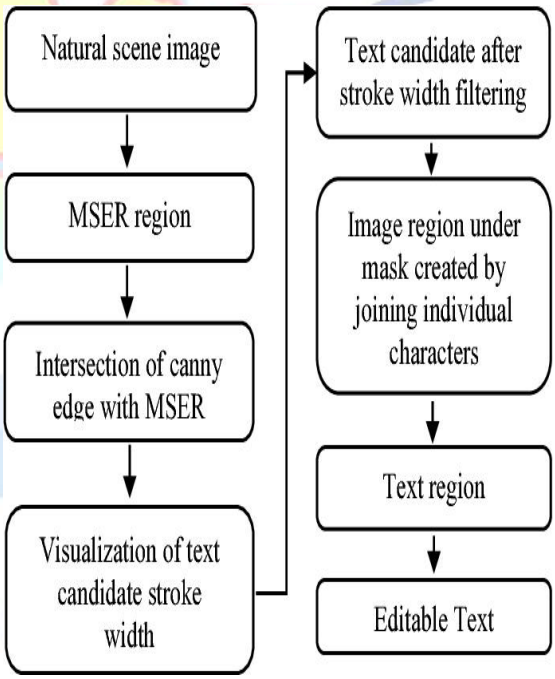


Figure5.5:Flowchart of MSER

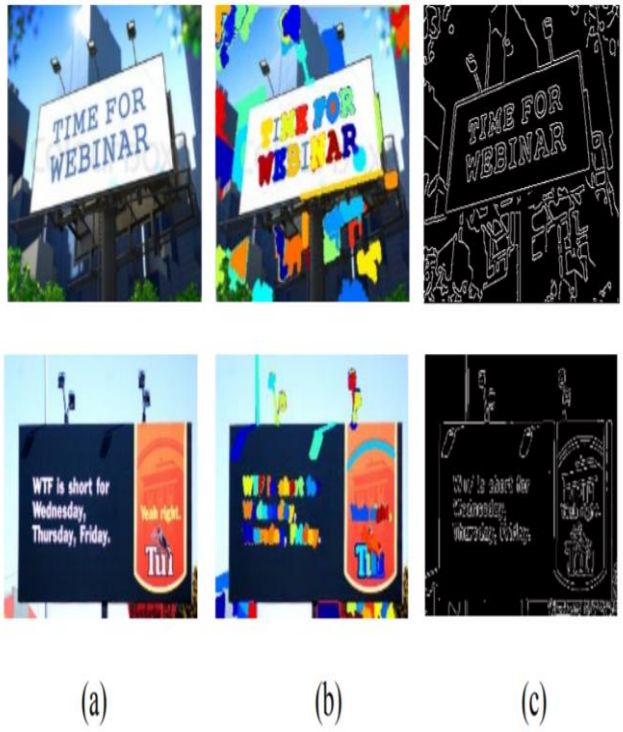


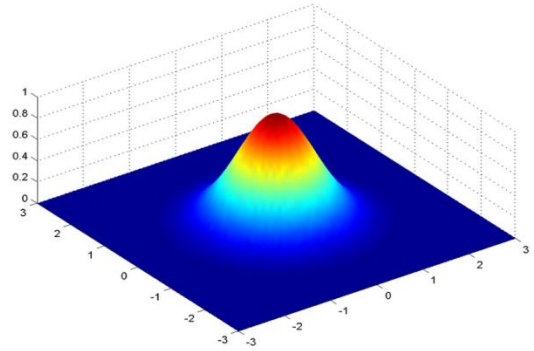
Figure5.6 MSER Text Detection Process

TEXT RECOGNITION FOR MSER AND OCR

Recognition process is employed to our selected text region from the natural scene image. We train a character recognizer which can recognize 10 digits (0 to 9), 52 English letters (26) for upper case and 26 for lower case) total 62 characters. Two dataset containing image patches of complete and full text characters is used to train. For different font style character is described helps to get more accurate results in recognition. For the description of text character a novel character descriptor model is proposed which combined MSER detector to find out stroke components and Random detector to extract preset number in random pattern.

A. Gaussian Mixture Model (GMM)

Gaussian mixture model is normally used to fit a vector of unknown parameter. In our algorithm, Dense Detector produces a uniform 8×8 key points array and Random Detector produces 64 key points randomly. We use 8 Gaussian distribution for clustering. For building GMM, firstly calculate K centers of the Histogram of Oriented Gradients (HOG) descriptor, where K means clustering (K=8).



B. Character stroke configuration

In the view of pixel-level prospective, stroke is defined as a region bounded by two parallel boundary segments. A character structure consists of multiple strokes. The distance between the two parallel boundaries is called stroke width.Same class character has the same stroke. We use a stroke alignment method for estimating the average value of stroke configuration . Thus our proposed method can detect text for any kind of fonts, styles, and sizes.

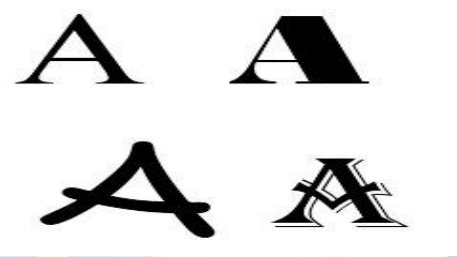


Figure6.1:Letter Stroke Width

RESULTS AND EVALUATION

In this section, to ensure the fairness of each evaluation. All the text detection and recognition using MSER and OCR is verified. All methods are implemented on a Windows 10 personal computer and performed in MATLAB R2021a software. We performed the conversion of blur image to clear image and after process continued.

Field	Description
Connectivity	Connectivity of the MSER regions. Default: 8
ImageSize	Max Size of 1239x949.
NumObjects	Number of MSER regions in

Field	Description
	1239.
PixelIdxList	1-by-NumObjects cell array containing NumObjects vectors. Each vector represents the linear indices of the pixels in the element's corresponding MSER region.

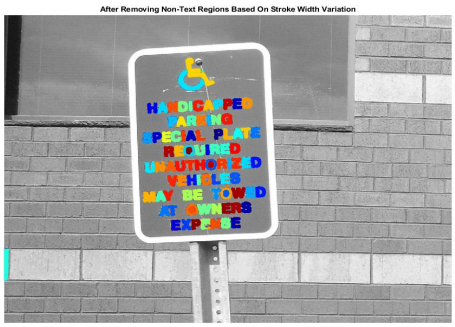


Figure: Input image

```
ans =
    'HANDICIXPPED
     PARKING
     SPECIAL PLATE
     REQUIRED
     UNAUTHORIZED
     VEHICLES
     MAY BE TOWED
     AT OWNERS
     EXPENSE
```

Figure: Output image

CONCLUSION

This paper provides an overview of OCR system based on neural network. Novel approach to character extraction trough MSER feature extractor is presented, making process of character extraction invariant to affine transformation and rapid illumination change. Non-character objects were removed, in order to reduce problem of character recognition to classification. Neural network was used for character recognition. Neural network was trained with back-propagation algorithm for character recognition.

REFERENCES

1. G. Zhou, Y. Liu , Z. Tian, and Y. Su, "A new hybrid method to detect text in natural scene," 18th IEEE Int. Conf. Image Processing (ICIP), pp. 2605 – 2608, Sep. 2011.
2. L. Gómez and D. Karatzas, "MSER-Based Real-Time Text Detection and Tracking," 22nd International Conference on Pattern Recognition (ICPR), pp. 3110-3115, 2014.
3. K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp. 1631 – 1639, Dec. 2003.
4. H. Takahashi and M. Nakajima, "Region graph based text extraction from outdoor images," in Proc. 3rd ICITA, vol. 1. Jul. 2005, pp. 680–685.
5. C. Liu, C. Wang, and R. Dai, "Text detection in images based on unsupervised classification of edge-based features," in Proc. 8th Int.Conf. Document Anal. Recognit., 2005, pp. 610–614.
6. K. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.
7. H. Takahashi and M. Nakajima, "Region graph based text extraction from outdoor images," in Proc. 3rd ICITA, vol. 1. Jul. 2005, pp. 680–685.
8. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, T. Kadir and L. Van Gool: "A Comparison of Affine Region Detectors"; International Journal of Computer Vision, Volume 65, Numbers 1-2 / November, 2005, pp 43-72.
9. Donoser, M. and Bischof, H. Efficient Maximally Stable Extremal Region (MSER) Tracking CVPR, 2006.
10. C. Yi and Y. Tian, Text string detection from natural scenes by structure based partition and grouping, IEEE Trans. Image Process., vol. 20, no.9, pp. 25942605, Sep. 2011.