iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

MSER Based Object Character Recognition Technique

Bhaskar B.[1], Manjushree R M[2], Meeth Kumar R Kantesaria[3], Nagendra S [4]
[1,2,3,4]Dept. Of ECE SJB Institute of Technology

Abstract: Recognition of detected text from any image is challenging but it is essential to do so. In this paper, we propose an effective and accurate algorithm namely advanced 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 pre-processing 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, to select the text region we take it as an input of a novel optical character recognition (ocr) technique so that the text is usable. The evaluation result substantiates 70% accuracy.

I. INTRODUCTION

Text extraction and segmentation is an important phase in document recognition system and is widely used. As more and more information is transformed in digital form, visual texts are embedded in many forms of digital media such as images, documents, videos etc. This text present important information like names, locations, brands of products, name of books etc. which facilitate ease in indexing and understanding. The text embedded in media is of two types: Superimposed text and scene text. Superimposed text are also known as graphical text, is horizontally aligned and present in foreground with size, font spacing and orientation remaining constant within a text region, this type of text we do not deal with in this project. Whereas scene text is present in recorded scene or photograph with variable size, font, colour, alignment, lightening condition etc, we are interested in scenic text. Text detection and recognition can play vital role in day to

day life of humans and is a vital part of so many computer applications. More over text extraction plays a crucial role for blind people when they want to read the text presented in the scene images. It will enable in indexing the documents having high level of visual content by using text based search engines. It can help in developing systems to count characters, words in a document, identify the type of script, thereby helping in palaeography study. It can help in distinguishing between images and documents with embedded visual content and developing text conversion system, thus opening the possibility for more improved, efficient, fast and advanced systems.

Text detection, which is identification of textual regions and segmentation system which is segregating, each characters in the textual region is designed to elicit different types of text from arbitrary complex images, documents or videos. To segment various characters in the document, it is imperative first to identify all the manuscript regions in the document. Identifying text is in itself a complex process as text characteristics can vary in font, size, style, orientation, alignment, contrast, texture, and colour and background information. Correctness of text line segmentation directly affects accuracy of segmentation of word which subsequently affects the accuracy of character recognition. There are various techniques for text extraction and segmentation in images e.g. text segmentation based on texture aims at finding the difference between the text pixels and the background pixels. Text segmentation can be done in various methods. Text strings can be further be refined by applying certain constraints like edge detectors, region filtering, stroke width etc. Although many research efforts have been made to detect text regions in the images the paper proposes a novel text detection and segmentation algorithm which employs Maximally Stable Extremal Regions as basic letter candidates.

MSER helps in identifying character regions even when the image is of low quality. Following this the characters are segmented using connected component approach. The remainder of this paper is organized as follows.\

A. Edge-enhanced MSER

Maximal Stable Extremal Regions are used as a method of blob detection – blob detection refers to detecting regions in an image that differs in properties, such as brightness or colour. These blobs are typically found by searching for regions that maintain a consistent intensity when a wide range of thresholds is applied to an image. As the contrast of text within an image with its background is typically significant or the colour of every letters in the text stays consistent, MSER is a great choice for text

detection. However, MSER is sensitive to image blur, [1] suggested enhancing the text candidate by applying canny edge Detection on the original image. By finding the intersection of MSER and Canny edges, we combined the complimentary properties of canny edge detection and MSER region to achieve a better result. Fig 1 shows an image after applying edge-enhanced MSER. Then the edges are grown along the gradient and removed along the gradient in order to remove some of the non-text pixels in the MSER and edge intersection. The result of growing and removing the edges along the gradient are more segmented regions.

B. Geometric Filtering

The result of edge-enhanced MSER is a binary image where the foreground connected components are considered as letter candidates. We perform a set of simple geometric checks on each connected components to filter out non-text objects. First, very large and very small objects are rejected. We assume that the size of text within an image does not take up the whole image nor does it only make up a very small portion of the image. Then, since most letters have aspect ratio close to 1 [1], we reject connected components with very large and very small aspect ratio. Finally, we eliminate components that contain a large number of holes, because letters does not have many holes. Fig.2 shows the result of geometric filtering. Almost all of the non-text from edge-enhanced MSER objects had been eliminated.'

The three main region properties that are used to filter the connected components by the typical text attributes are the eccentricity, area, and solidity of the object. MSER regions can usually be drawn with a surrounding ellipse, so the eccentricity of the ellipse is used to determine the aspect ratio. The area, or the number of pixels in a region, is used to remove the very large and very small objects that are assumed to be non-text components. The solidity or proportion of pixels in the convex hull that are also in the region, of letters is typically low, so the connected components with a high solidity can be rejected. The connected components that were removed tend to have an eccentricity greater than 0.995, an area less than 80 pixels or greater than 1000 pixels, and solidity under 0.4.

C. Stroke Width Filtering

Characters in text usually have a similar stroke width or thickness. This text property can be used to continue to filter out more non-text regions remaining after filtering the edge enhanced MSER by the geometric properties of the connected components. In order to find the stroke width of each object, we use a distance transform approach. The distance image can be found by setting the intensity of each pixel within a text candidate region to the distance from that pixel to the nearest background pixel. The stroke width information can then be derived from the distance image.

Since the stroke width of characters in a line of text are consistent among characters, regions with too much variations in object stroke width can be assumed to be non-text and removed from the candidate regions. A text candidate region is removed when the stroke width standard deviation divided by the stroke width mean is greater than 0.35 (std/mean > 0.35).

D. Text Region Bounding Box

Once the connected components are removed until remaining components are mostly text characters, the individual characters have to be merged into a single connected component. This is accomplished by performing morphological operations on the connected components until the outliers are rooted out. Morphological closing, which is essentially dilation followed by an erosion, is used to enlarge the boundaries of the foreground to shrink any small holes in the region. Then morphological opening is applied to the image by performing erosion followed by dilation. The opening removes any stray foreground pixels in the background. The bounding box is then found for each region with an area greater than a specified threshold (5000 pixels). Each of these bounded regions can then be passed through to an OCR system to recognize the text in each region.

II. METHODOLOGY

The above figure shows the steps involved in order to detect and recognized printed text on the spine of the book.

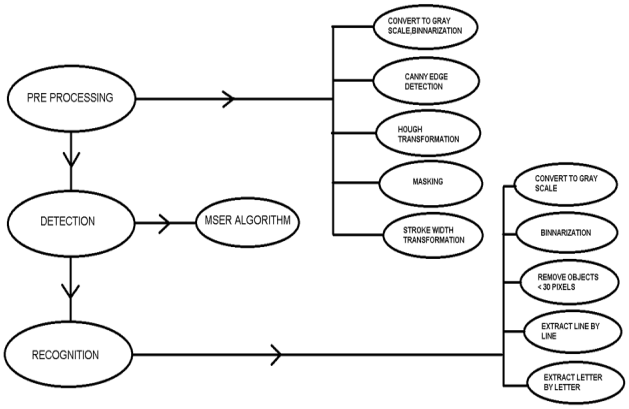
The following are the steps involved.

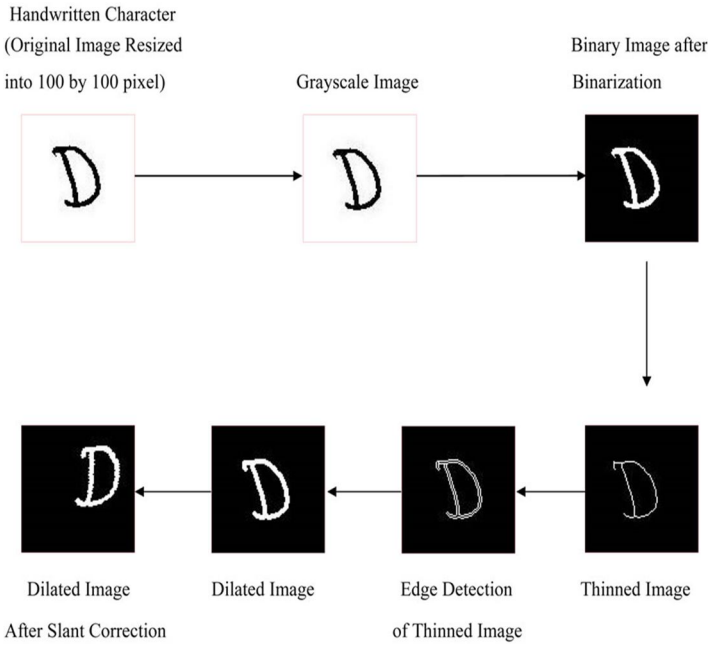
Pre-processing

Detectio

Recognition

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)





These three are the main steps that help us in achieving are cognized text, under each step are few other steps that's necessary for the processing of the image. They are as follows.

Pre-processing

Convert to grey scale binarization

Canny edge detection

Rough transformation

Masking

Stroke width transformation

DetectionMSER

Recognition

onvert to grey scale

BinarizationRemove objects <30 pixelsExtract line by lineExtract letter by letter

And then finally the printed text is detected and recognized in the above pattern.

A. Pre-processing

The above image shows the various steps in the pre-processing first the image is converted to a grey scale image, this is done

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

because if we use RGB image we would need more space to store the pre-processed image

as the size of the image increases the processing also takes longer. Using a RGB image would take 66%more time for processing so instead we use grey scale which has only 256 levels which represent different composition of the image with different shades of grey.

Then canny edge detection is used to detect the edges of the printed text in the image.

Canny edge detection algorithm is also known as the optimal edge detector. Canny's intentions were to enhance the many edge detectors in the image.

The first criterion should have low error rate and filter out unwanted information while the useful information preserve.

The second criterion is to keep the lower variation as possible between the original image and the processed image.

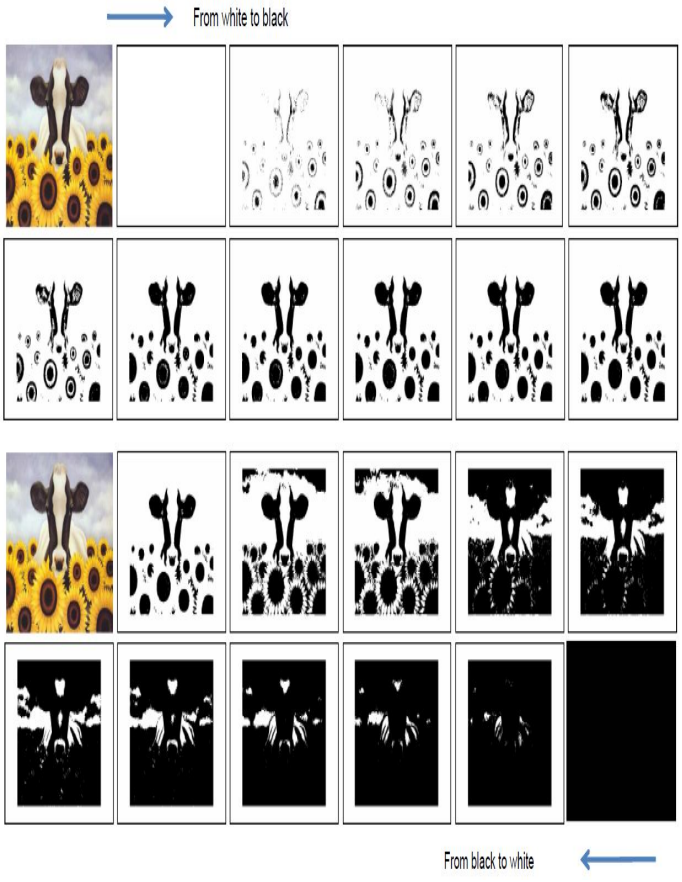
Third criterion removes multiple responses to an edge.

Based on these criteria, the canny edge detector first smoothes the image to eliminate noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum using non-maximum suppression. The gradient array is now further reduced by hysteresis to remove streaking and thinning the edges.

Then Hough transformation is used to isolate and extract the features of the edge detection. For example in an image we apply canny edge detection, then the edges present in the image are detected but in order to use the information of the detected text we need to extract the edges detected and eliminates the unwanted part of the image. Thus Hough transformation is used to extract the detected text from the image.

Then masking is done in order to remove the unwanted part of the image.

The stroke width variation is then used whose functionality is mentioned above.



B. Detection

We use MSER technique to detect the characters from the given image. What happens within the MSER technique is mentioned

below.

C. Sweeping image thresholds

Apply a series of thresholds – one for each grey scale level.

Threshold the image at each level to create a series of black and white images.

One extreme will be all white, the other all black. In between, blobs grow and merge.

Extremal regions have the important properties, which the set is closed under continuous (and thus projective) transformation of image coordinates, i.e. it is affine invariant and it doesn't matter if the image is warped or skewed.

Because the regions are defined exclusively by the intensity function in the region and the outer border, this leads to many key characteristics of the regions which make them useful. Over a large range of thresholds, the local binarization is stable in certain regions, and have the properties listed below:

Invariance to affine transformation of image intensities

Stability: Only region whose support is nearly the same over a range of thresholds is selected.

Multi-scale detection both fine and large structure is detected.

III. MSER REGIONS

The MSER detector using a union-find based approach. The detector selects the salient regions among the extremal regions of an image I, defined by the extremal property of the image intensity function f on the region outer boundary. Among these regions, minimal (resp. maximal) extremal regions are such connected regions R of the image that all the elements on the outer boundary have strictly greater (resp. smaller) intensity than all the adjacent region elements. We denote by R_k a minimal extremal region with the maximal intensity level k of all the elements in the region. Similarly, R_k denotes a maximal extremal region of minimal intensity level k among the region elements. Minimal extremal regions are nested for increasing k, i.e. R_k, R_l for any $k < l$, and similar relation holds true for maximal extremal regions. A function q(), calculated for the elements of a nested sequence, indicates the rate of growth of a region with the change of intensity. The minimal extremal regions R_k where this function reaches a local minimum are labelled maximally stable (minimal) extremal regions, and similarly for the local minima of the stability function q() on the nested sequences of maximal extremal regions. In order to speed up the computation as well as resolve certain ambiguities from the original definition, implementations in popular computer vision libraries, as well as our implementation, use a simplified version of the stability function:

$$q'(\mathcal{R}_k) = \frac{|\mathcal{R}_{k+\Delta} \setminus \mathcal{R}_k|}{|\mathcal{R}_k|}.$$

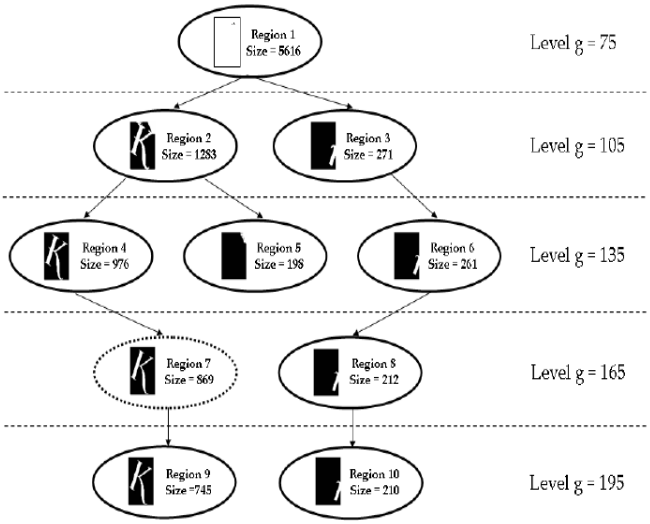
where. | | denotes cardinality. The parameter Δ is the parameter of the method. A larger Δ parameter requires the region to be stable through a greater range of grey levels. The region $R_k+\Delta$ are determined from the sequence of nested regions to be the largest region such that:

$$d(\mathcal{R}_k, \mathcal{R}_{k+\Delta}) \leq \Delta.$$

The distance between the regions d() is defined hereafter depending on the region type. Additional parameters to control the size of the detected regions, limit the maximal allowed value of the stability function as well as restrict the detections to only sufficiently different nested regions are also introduced to the detection method.

After which a component tree is created by processing the image at different thresholds and selects a particular picture according to the above method. A component tree looks something like this.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

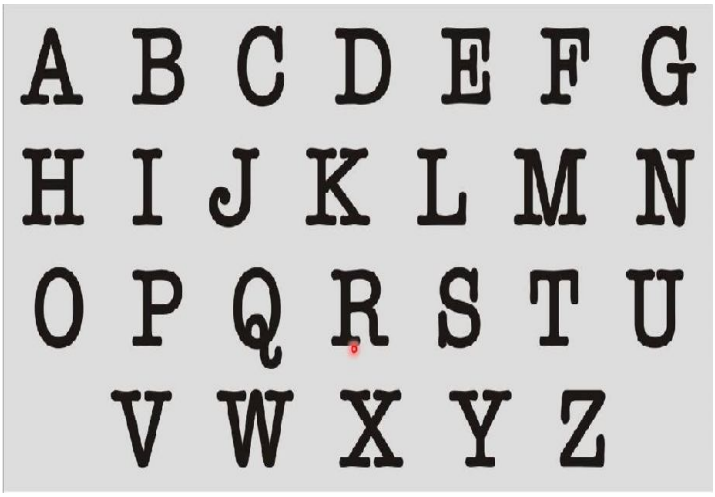


After detection of the characters present in the given image it is time for the characters to be recognized.

A. Recognition

A novel OCR technique is used for the recognition of the text, the steps which occur in the OCR are mentioned below.

Read image and show image: Image given as input for text recognition is read and displayed.

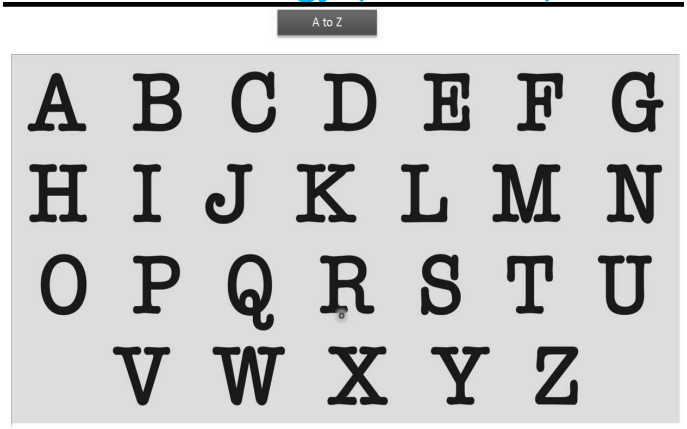


Convert to grey scale: a grey scale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort are composed exclusively of shades of grey, varying from black at the weakest intensity to white at the strongest

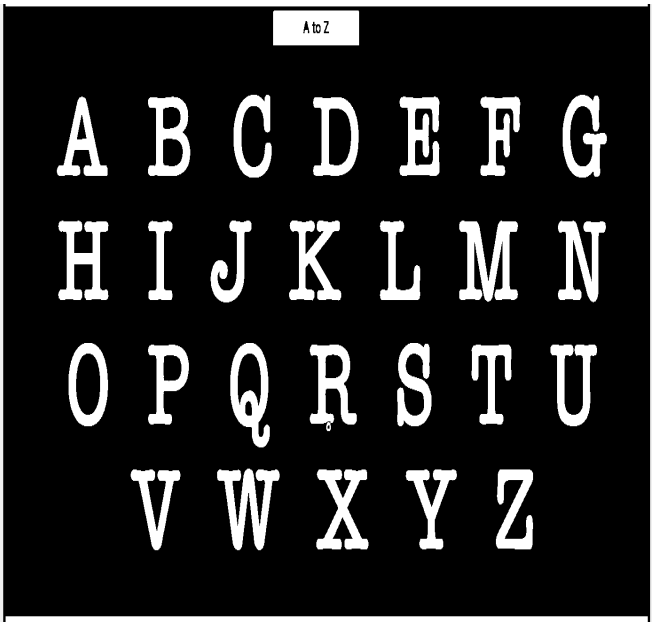
Grey scale images are distinct from one-bit bi-tonal black-and-white images, which in the context of computer imaging are images with only two colours, black and white (also called bi-level or binary images). Grey scale images have many shades of grey in between.

Grey scale images are often the result of measuring the intensity of light at each pixel in a single band of the electromagnetic spectrum (e.g. infrared, visible light, ultraviolet, etc.), and in such cases they are monochromatic proper when only a given frequency is captured. But also they can be synthesized from a full colour image; see the section about converting to grey scale.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



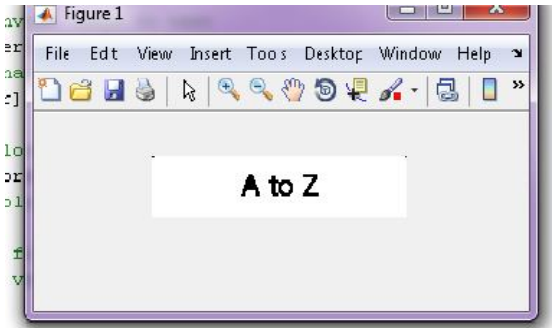
1) *Convert to B/W:* A black and white image is a digital image that has only two possible values for each pixel. Typically, the two colours used for a binary image are black and white, though any two colours can be used. The colour used for the object(s) in the image is the foreground colour while the rest of the image is the background colour.
2) In the document-scanning industry, this is often referred to as "bi-tonal".



3) Remove all object containing fewer than 30 pixels: This is done so as to obtain only the region of interest. Here other than the text part other pixels are removed.
4) Store matrix word from image: Obtained text part of the image is converted into matrix form and stored.
5) Opens text.txt as file for write: Text.txt file is the file that contains the text that is extracted from the image, hence wen the text region is obtained it is written on to this file.
6) Load templates: Templates of the letters of all fonts possible are loaded so as to recognise any font given for recognition.
7) Compute the number of letters in template file: program always keeps a count of templates loaded so that the comparison with the text detected becomes easier for recognisation.
8) Separate lines in text
9) Label and count connected components
10) Extract text lines

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

B. Example



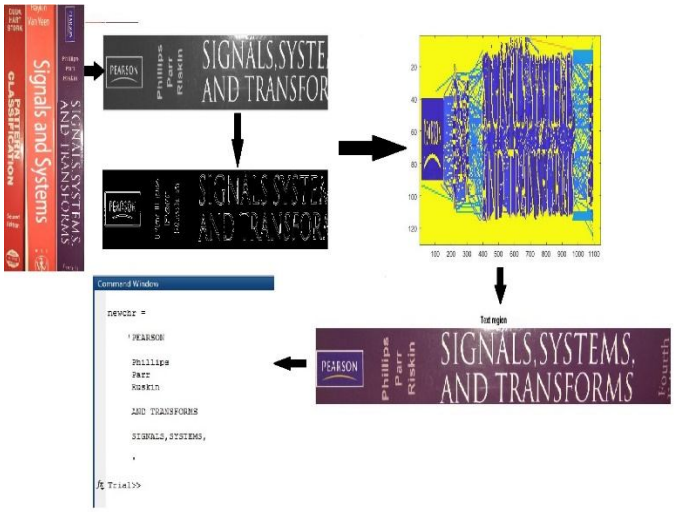
1) Resize letter (same size of template): so that it matches with the template size and becomes easier to compare and recognised.
2) Convert image to text
3) Letter concatenation
4) When the sentences finish, breaks the loop

IV. RESULT

The algorithm has been tested on fragmented characters, blur characters, smeared or broken characters and also on the image which contain graphics and overlapping character.

The below image shows the process of selecting the image needed, from which data is to be extracted is selected properly. The image is then taken through the process to obtain the final extracted text from the image.

Here in the below image, we have considered 'Signals and Systems' text book to be the reference image. From the set of the three images, the signals and systems book was selected properly and the output was observed to be the same text as that of the text which was seen on the book.



V. CONCLUSION

In this paper, we have discussed complete methodology to implement character recognition, it has been observed that template matching a feature extraction method when applied with MSER. We observed that feature extraction is probably the single most important factor for gaining high accuracy. Moreover, the pre-processing methods discussed in this paper also remove noise at great extent. MSER methodology to take advantage of multi-channel data. During this, methodology is intended to handle the text detection in document pictures. As an outlook, we plan to carry out more evaluations to find an accurate mechanism for application to our case study.

REFERENCES

[1] Santosh Kumar Henge and B. Rama "Comparative Study with Analysis of OCR Algorithms and Invention Analysis of Character Recognition Approached Methodologies" 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016
[2] Abubakar Muhammad Ashir and Gaddafi Sani Shehu "Adaptive clustering algorithm for optical character recognition" ECAI 2015 - International Conference –

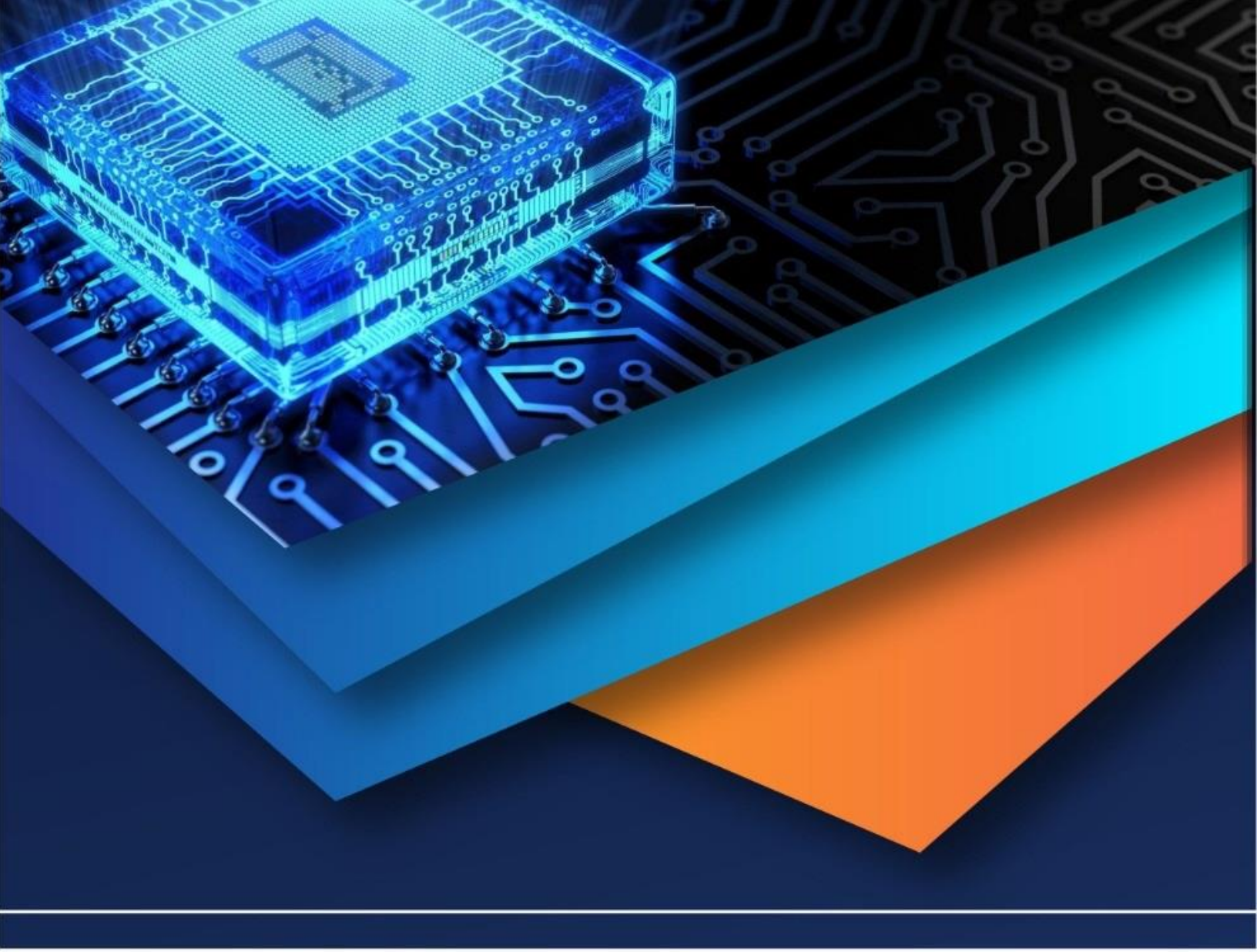
Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

7th Edition Electronics, Computers and Artificial Intelligence 25 June -27 June, 2015, Bucharest, ROMÂNIA

[3] M. SHRIDHAR and A. BADRELI "High accuracy character recognition algorithm using fourier and topological descriptor" University of Windsor, Windsor, Ontario, Canada N9B 3P4 (Received 1 November 1983; in revised form 20 January 1984 ; received for publication 1 February 1984

[4] Maher Ahmed and Rabab Ward, Fellow "A Rotation Invariant Rule-Based Thinning Algorithm for Character Recognition" 16 march 2012

[5] Li shan shan, Li Yi-min "A Recognition Algorithm for Similar Characters on License Plates Based on Improved CNN" 12 June 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY