IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Image Text Extraction and Text-to-Speech Conversion

[1]Dr.Sujatha. K., [2]Shruti P. Patil

[1]Professor, [2]Student

[1]Digital Communication Networking,

[1] Sharnbasva University, Kalaburagi, India

Abstract: Recent advances in the fields of image processing and natural language processing are aimed at creating smart systems that will improve people's quality of life. In this paper, an effective method for text detection and extraction from photographs, as well as text to audio conversion, is proposed. Gray scale conversion is used to improve the incoming image first. Following that, the improved image's text sections are discovered using the Maximally Stable External Regions (MSER) feature detector. To efficiently gather and filter text sections in a picture, use geometric filtering in conjunction with the stroke width transform (SWT). Geometric attributes and the stroke width transform are used to remove non-text MSERs. Individual letters/alphabets are then combined together to identify text sequences, which are ultimately fractured into words. Finally, the words are digitised using Optical Character Recognition (OCR). In the final stage, we convert the identified text to speech using our text-to-speech synthesiser (TTS). The suggested algorithm is put to the test on a variety of images, ranging from documents to nature settings. Promising findings have been presented, demonstrating the suggested framework's accuracy and robustness and encouraging its use..

***Index Terms* -** Maximally Stable External Regions (MSER), the stroke width transform (SWT), Optical Character Recognition (OCR), text-to-speech synthesiser (TTS)**.**

I. Introduction

Languages are the most ancient means of communication between humans, whether spoken or written. Visual writing in natural or constructed sceneries may hold very vital and useful information in the modern era. As a result, the researchers have begun digitising these photos, extracting and interpreting data using specialised methodologies, and then doing text-to-speech synthesis (TTS). It is done so that the information can be read aloud for the user's benefit and convenience. Text extraction and TTS can be used in tandem to assist people with reading problems and visual impairments in hearing printed material through a computer system. In this paper, we offer a novel text identification system based on connected component analysis and MSER methods for extracting CCs that are used as letter candidates. The geometric features and stroke width fluctuation of CCs that are likely to be characters are used to pick them. The picked items are subsequently sorted into text sequences that are then fractured into individual words. The words are recognised and retrieved using optical character recognition, and the extracted text is then translated to appropriate speech using a text-to-speech synthesiser. The suggested algorithm is put to the test on a variety of images, ranging from documents to nature settings. Promising results have been presented, demonstrating the suggested algorithm's accuracy and robustness, and encouraging its use in real-world applications.

II. RELATED WORK

[1]. Ranjit Ghosal, Ayan Banerjee ," An Improved Scene Text And Document Image Binarization Scheme",Recent Advances In Information Technology (RAIT) 2018. Identification of textual content parts have a vital have an effect on on sensible transport systems, file picture processing, robotics and content material primarily based photograph retrieval systems. So, an correct textual content identification technique is essential for textual content based totally scene picture processing duties such as OCR. Scene textual content photograph binarization performs an vital function in any textual content identification algorithm and consequently in the OCR performance. In this work a novel method to herbal scene textual content picture binarization with the aid of monitoring the textual content boundary primarily based on aspect and grey stage variance information. Further, damaged boundaries are linked to assemble the entire boundary map. Here, an adaptive threshold is decided primarily based on boundary part facts to binarize the photo effectively. Compared to different nicely regarded binarization methods, our approach has been proved extra advantageous in instances the place the herbal scene pics have low contrast, low resolution, non-uniform illumination and noise. Our experiments are performed on the datasets of ICDAR 2003 Robust Reading Competition, ICDAR 2011 Born Digital Dataset, Street View Text (SVT) Dataset, DIBCO dataset and our laboratory made Bangla Dataset. The experimental outcomes are satisfactory. [2]. Muhmmad Jaleed Khan, Naina Said, Aqsa Khan, Naila Rehman, Khurram Khurshid Automated Latin Text . Robust and accurate detection of text in natural scene images

and document images is a very challenging and common research problem. Over the past few decades, a variety of algorithms for text detection in images have been developed but there is still need for more robust and accurate text detection methods. In this work, we have proposed an accurate and robust text detection framework in which canny edge detection, maximally stable extremal regions and geometric filtering are employed in combination to efficiently collect and filter letter candidates in an image. Subsequently, individual letter patches are grouped to detect text sequences, which are then fragmented into isolated word patches. Finally, optical character recognition is employed to digitize the word patches. The proposed algorithm is tested on images representing different scenarios ranging from documents to natural scenes. Promising results have been reported which prove the accuracy and robustness of the proposed framework and encourage its practical implementation in real world applications Robust and correct detection of textual content in herbal scene pictures and report pics is a very difficult and frequent lookup problem. Over the previous few decades, a range of algorithms for textual content detection in pix have been developed however there is nonetheless want for greater sturdy and correct textual content detection methods. In this work, we have proposed an correct and sturdy textual content detection framework in which canny side detection, maximally steady extremal regions and geometric filtering are employed in mixture to successfully gather and filter letter candidates in an image. Subsequently, person letter patches are grouped to observe textual content sequences, which are then fragmented into remoted phrase patches. Finally, optical personality consciousness is employed to digitize the phrase patches. The proposed algorithm is examined on pictures representing exclusive eventualities ranging from archives to herbal scenes. Promising consequences have been mentioned which show the accuracy and robustness of the proposed framework and inspire its sensible implementation in actual world purposes

III. Proposed Work

In this work, we recommend a sturdy MSER technique to extract the text from images. The MSER areas are areas that have a distinctly awesome depth in contrast to their heritage contrast. They are retrieved via a procedure of trying severa thresholds. The areas that maintain steady shapes over a extensive vary of thresholds are selected. Segmenting the textual content from a scene by way of MSER intensively helps in similarly processing of picture for detecting textual content regions. Once the MSER areas are detected these vicinity are similarly processed the usage of geometric properties, linked aspects and stroke width variant. Once the textual content areas are detected, the different non-text areas are removed. MSER is well suited with textual content due to the regular colour and excessive distinction with the background, which collectively provide us secure depth profiles. However it is incredibly probably that a quantity of non-text areas that are steady are additionally selected. Geometric houses such as eccentricity, bounding field ,solidity, euler quantity are additionally taken into consideration for detection of textual content regions. Connected aspects inside a location are additionally viewed for detecting the vicinity of interest. To take away the non-text areas the stroke-width is considered. Text characters have a tendency to have little version when it comes to stroke widths of the traces and curves, whereas non textual content areas show a excessive stroke width variance. So the areas that have excessive stroke width version are eliminated as they greater in all likelihood to be non-text regions. The detected textual content areas bear OCR (Optical Character Recognition) for digitizing the textual content areas and to realize and extract the textual content from image. Finally the detected textual content is transformed to speech the usage of text-to-speech synthesizer. In our work we make use of the Microsoft text-to-speech device handy for Windows.

IV. System Architecture

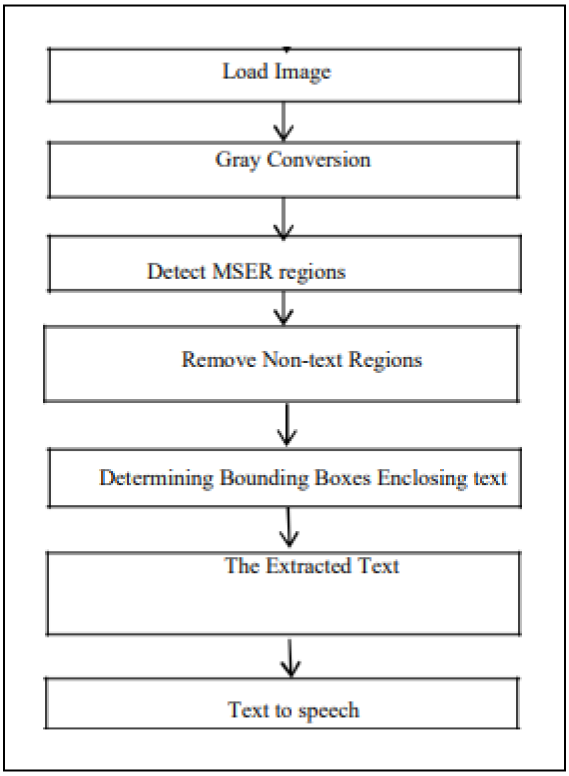


Figure 1: System Architecture of Proposed system

Architecture consists of :

A. Gray scale conversion The grayscale photograph is represented by way of the usage of eight bits value. The pixel price of a grayscale picture stages from zero to 255. The conversion of a colour picture into a grayscale photograph is finished by way of changing the RGB values (24 bit) into grayscale values (8 bit). One approach of changing RGB to grayscale is to take the common of the contribution from every pixel (R+G+B)/

B. MSER Regions: Maximally secure extremal areas are used as a approach of blob detection in images. MSER areas are related areas characterised with the aid of nearly uniform depth for the duration of a vary of thresholds. The chosen areas are these that keep unchanged over a giant set of thresholds.

a. Edge: Edge is a team of factors having robust gradient magnitude in an image.

b. Corner (or Point of Interest): Corner is a team of factors having a excessive stage of curvature in the gradient in an image.

c. Region: A place is a contiguous set of adjoining pixels.

d. Blob (or Region of Interest): Blob is the location in which some houses (color, brightness, etc.) are invariant or barely variant in an image, i.e. factors in a blob are similar.

e. Boundary: Boundary of a place is the team of pixels neighboring at least one pixel of that vicinity however now not a section of that region.

f. Extremal Region: If all the pixels in a location have values larger than (or smaller than) that of the boundary, the location is referred to as extremal region.

g. Maximally Stable Extremal Region (MSER): An extremal place is termed as maximally steady when its version w.r.t. a given threshold is minimal.

h. Load Image Gray Conversion Detect MSER areas Remove Non-text Regions Determining Bounding Boxes Enclosing textual content

C. Connected components: Connected factors of an picture are the areas which have non-stop pixels inside that region. The pixels in the linked elements are linked to every different thru both 4-pixel, or 8-pixel connectivity.

D. Geometric homes The following geometric residences are taken into consideration: a. Bounding Box: Bounding containers are rectangular bins created round the location of interest. It includes all the pixel values inside the enclosing boundary. b. Eccentricity: The eccentricity is the ratio of the distance between its primary axis size and the foci. The cost ought to be between zero and 1. An ellipse is stated to be circle if its eccentricity fee is zero whereas if the eccentricity fee is 1 then the ellipse is a line segment. c. Solidity: Solidity additionally regarded as convexity of an photo is the region of the photograph divided with the aid of location of its convex hull. It is the percentage of the pixels in the convex hull that are existing in the area to the real variety of pixels in the picture d. Extent : Extent of an picture is described as the ratio of the pixels in the photograph to the variety of pixels in the complete bounding container in that image. e. Euler : Euler variety is described as the whole wide variety of pixels in the photo minus the variety of holes in that region. Holes in a vicinity shows there are no pixels in the region. We can use both four or 8-connectivity.

E. Stroke width radically change A stroke in an picture is a non-stop band of a almost consistent width. As the title suggests stroke width version calculates the width of the most in all likelihood stroke containing the pixel for every pixel in that stroke

F. OCR OCR stands for Optical Character Recognition. As the identify suggests OCR is used to notice the ordinary human readable language which can also be current in the shape of textual remember current in picture or any files or pdf documents and convert it into editable formats.

G. Text to speech A text-to-speech (TTS) machine converts the regular human readable language textual content into speech

V. RESEARCH METHODOLOGY

In this section, we describe the textual content detection algorithm that is MSER (Maximally Stable Extremal Region) algorithm. MSER is a technique for textual content detection, blob detection in images. The MSER algorithm extracts quantity of co-variant areas from image. We first outline the notion of stroke and then give an explanation for the stroke width transformation. MSER is primarily based on the concept of taking areas which remain almost the identical via a broad vary of thresholds. All the pixels above or equal to a given threshold are black and all the pixels beneath a given threshold are white. MSER makes use of two vital residences to get rid of non-textual content areas from picture first is Geometric Properties and any other is Stroke Width Variation Properties. To use MSER algorithm we first summarized the frequent attributes of textual content as a) Text in photo continually incorporates a lot of edges; b) The width of textual content is large than height; c) Text is bounded in size; d) Text has one of a kind texture however this texture is irregular; The drift chart of the algorithm is proven :

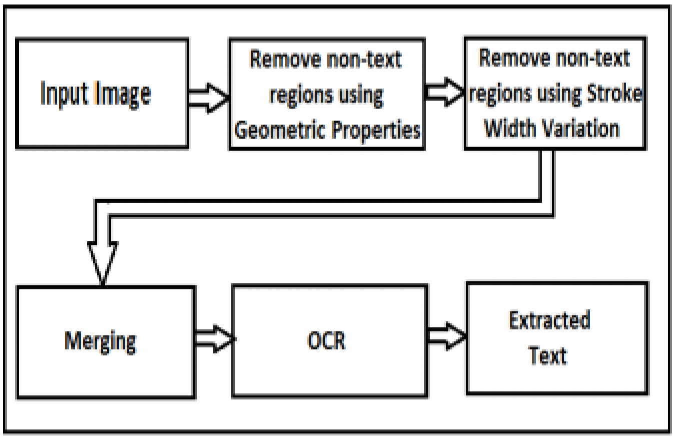
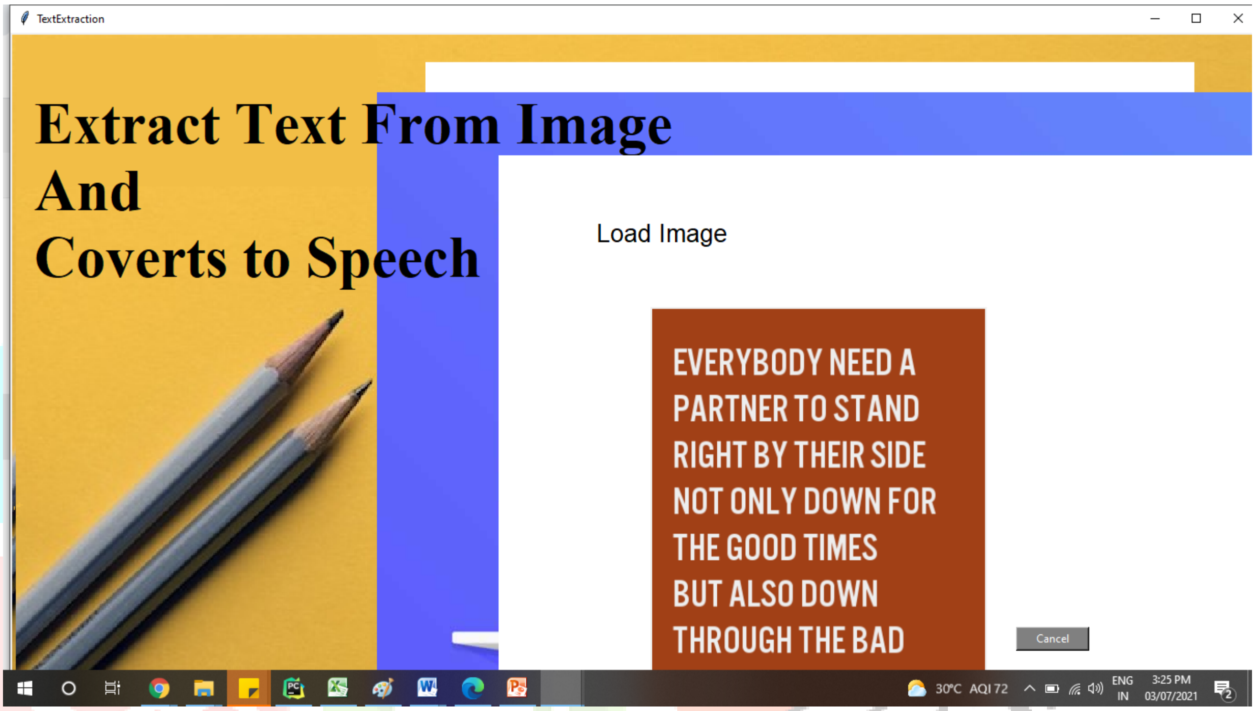

Input Image
Remove non-text regions using Geometric Properties
Remove non-text regions using Stroke Width Variation
Merging
OCR
Extracted Text

Figure 2: Flow of the system

The flowchart suggests the real drift of MSER algorithm. MSER algorithm become aware of all textual content areas alongside it additionally become aware of some non-textual content regions. To do away with that non textual content areas Geometric Properties are utilized on that image. If Geometric Properties are unable to put off all non- textual content areas then practice Stroke Width Variation on that photograph to do away with last non textual content regions. Geometric homes MSER detects nearly all textual content areas from photo however alongside it additionally observe some non -text regions. To take away these non -textual content areas first we practice Geometric Properties on image. Geometric Properties detects the non- textual content areas from photo and put off these regions. The non-textual content areas which are no longer eliminated in Geometric Properties for these areas we use the Stroke Width Variation Properties. Stroke width variant residences The Stroke Width Variation Properties are additionally used to get rid of the non- textual content regions. Stroke Width is a measure of the curves and strains that make up character. Text areas have little stroke width variation, the place as non- textual content areas have large variations. To take away the non- textual content areas the usage of stroke width we require thresholds. All the pixels above or equal to a given threshold are black and all the pixels under a given threshold are white. Grouping letters into textual content strains At this point, the person detected letters are grouped into phrases or textual content lines. The grouping of letters contains greater significant facts than simply the man or woman letter. For example, recognizing the string 'HELP' vs. the set of character characters {'L','H','P','E'}, the place the which means of the word is misplaced barring the right ordering

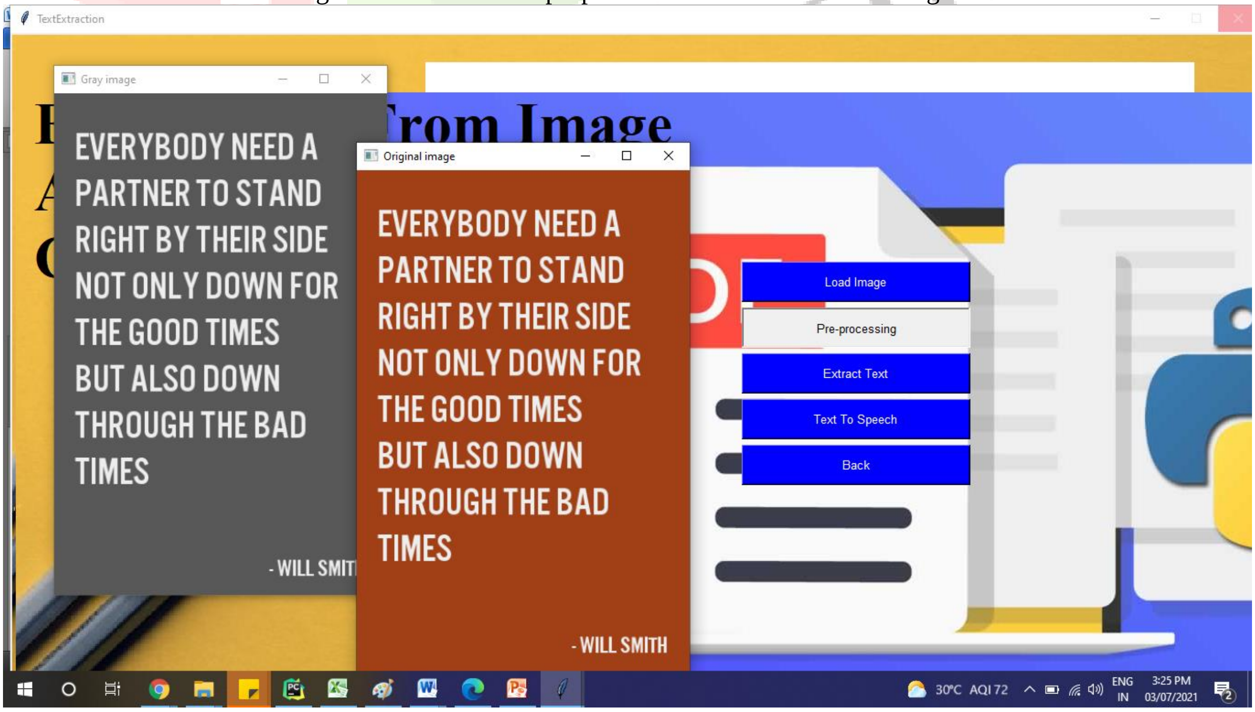
IV. RESULTS AND DISCUSSION



EVERYBODY NEED A PARTNER TO STAND RIGHT BY THEIR SIDE NOT ONLY DOWN FOR THE GOOD TIMES BUT ALSO DOWN THROUGH THE BAD

Figure 3: Input Image

Used to Read the image for text extraction purpose. This module reads the images for text extraction.



EVERYBODY NEED A PARTNER TO STAND RIGHT BY THEIR SIDE NOT ONLY DOWN FOR THE GOOD TIMES BUT ALSO DOWN THROUGH THE BAD TIMES

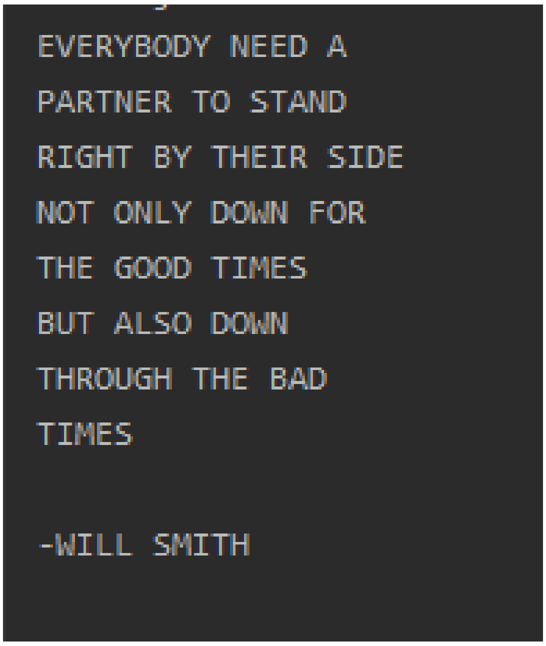
- WILL SMIT

EVERYBODY NEED A PARTNER TO STAND RIGHT BY THEIR SIDE NOT ONLY DOWN FOR THE GOOD TIMES BUT ALSO DOWN THROUGH THE BAD TIMES

- WILL SMITH

Figure 4: Converts to Grayscale

Used to convert the image in the form of gray color.



```
EVERYBODY NEED A
PARTNER TO STAND
RIGHT BY THEIR SIDE
NOT ONLY DOWN FOR
THE GOOD TIMES
BUT ALSO DOWN
THROUGH THE BAD
TIMES


-WILL SMITH
```

Figure 5: Text Extraction

This module extracts the text and displays as shown above.

V. CONCLUSION

Nowadays, there is growing demand of textual content facts extraction from image. So, many extracting methods for retrieving applicable records have been developed. Moreover, extracting textual content from the shade photograph takes time that leads to person dissatisfaction. In this paper we have proposed a technique to extract the textual content from picture which extracts textual content greater accurately. Using our approach it is feasible to extract facts inside brief time. Although, our linked element primarily based method for textual content extraction from coloration photograph approach has countless points than current technique however it turns into much less high quality when the textual content is too small and if the textual content location is now not genuinely seen or the shade of the textual content is now not seen virtually. In future, this work can be prolonged to become aware of the textual content from video or actual time evaluation and can be robotically documented in Word Pad or any different editable structure for in addition use.

REFERENCES

[1]. Ranjit Ghosal, Ayan Banerjee ,” An Improved Scene Text And Document Image Binarization Scheme”,Recent Advances In Information Technology (RAIT) 2018.

[2]. Muhmmad Jaleed Khan, Naina Said, Aqsa Khan, Naila Rehman, Khurram Khurshid Automated Latin Text .

[3]. Satish Kumar, Sunil Kumar , Dr. S, Gopinath “ Text Extraction From Images”, International Journal Of Advanced Research In Computer Engineering & Technology, June 2012.

[4]. Nitin Sharma And Nidhi , “Text Extraction And Recognition From The Normal Image Using MSER Feature Extraction And Text Segmentation Methods.” Indian Journal Of Science And Technology May 2017.

[5]. Amani Jamal, Noora Alhindi, Raghdah Nahhas, Somayh AlAmoudi “Image Assistant Tools For Extracting, Detecting, Searching Images And Texts”.2019.

[6]. Saeed Mian Qaisar†, Raviha Khan, Noofa Hammad “Scene To Text Conversion And pronounciation For Visually Impaired People” 2019.

[7]. B.Gatos, I.Pratikakis, K.Kepene And S.J. Perantonis, ”Text Detection In Indoor/Outdoor cene Images” 2005.

[8]. Kamrul Hasan Talukderr, Tania Mallick “Connected Component Based Approach For Text Extraction From Color Image”, International Conference On Computer And Information Technology (ICCIT) 2014. [9]. Jaswant P ,Anusuya S, Anil Kumar M, Dhikhi T ,” Enhanced Mser For Text Extraction”, International Journal OfComputational Intelligence And Informatics ,March 2016