

Methodology for eliminating plain regions from captured images

Shiva Shankar Reddy, Vuddagiri MNSSVKR. Gupta, Lokavarapu V. Srinivas,
Chigurupati Ravi Swaroop

Department of Computer Science and Engineering, Sagi RamaKrishnam Raju Engineering College, Bhimavaram, India

Article Info

Article history:

Received Dec 27, 2022

Revised Sep 29, 2023

Accepted Oct 31, 2023

Keywords:

Geometric properties

Maximally stable external regions

Optical character recognition

Scene images

Stroke width variation properties

Text detection

Text retrieval

ABSTRACT

Finding relevant content and extracting information from images is highly significant. Still, it may be challenging to do so because of changes within the textual contents, such as typefaces, size, line orientation, sophisticated backgrounds in images, and non-uniform illuminations. Despite these challenges, extracting content from captured images is still very important. Proficient textual content image recognition abilities extract text from the images to get over these issues. Despite the availability of several optical character recognition (OCR) techniques, this issue has yet to be resolved. Captured images with text are a rich source of information that should be presented so that viewers may make informed decisions. Because of this, it has become a complicated process to extract the text from an image because the text might be of poor quality, has a variety of fonts and styles, and occasionally have a complicated backdrop, among other things. Several approaches have been tried. However, finding a solution remains challenging. The maximally stable external regions (MSER) approach is developed to identify the text region in a picture. MSER is utilized to eliminate the plain regions outside the text and non-text areas using geometric features and stroke width variation qualities.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Shiva Shankar Reddy

Department of Computer Science and Engineering, Sagi RamaKrishnam Raju Engineering College

Bhimavaram, Andhra Pradesh, India

Email: shiva.srkr@gmail.com

1. INTRODUCTION

In recent years, text detection has developed into a significant issue. There have been advancements in computer vision and machine learning that have assisted text identification and retrieval. Moreover, expansion in application areas has come from this trend. Now that virtual cameras and smartphones are widely available; there has been a flurry of study towards extracting text from images taken with them. This is because text processing has attracted a lot of interest. Optical character recognition (OCR), format analysis, and pre-processing techniques all need text detection as an initial stage in processing scanned documents [1].

Only a tiny amount of work has been done on unconstrained camera-captured report pictures, although many solutions now focus on text or scenarios involving documents with constraints. Text identification in such inputs requires solving problems with curving content lines and natural (digital camera-captured) photographs. Despite this, most existing approaches focus on either/both of these scenarios and fail to adequately solve the text identification problem when presented with unrestricted inputs. The hidden texts inside the images provide essential information and make accusations regarding the photographs. Consequently, knowing what's happening behind the scenes is necessary for a person's ability as a computer. Some of the most recent advancements in machine learning focus on extracting features from unlabeled data

and demonstrating how these characteristics can be used to build classifiers capable of achieving high detection accuracy via massively parallel algorithms. The embedded texts in the images provide advanced semantic recordings of the environment [2]. Large amounts of information stored in databases and on the web will grow due to these records (Data can be images). Managing and recovering the resources should provide a formidable obstacle to developing practical solutions. Finding and separating text from images is difficult due to the wide variety of font types, sizes, and colours and the possibility of sharing a backdrop colour with the text. After these steps are completed, the text detection tool will transform the picture into legible text; nevertheless, the tool's performance is subpar when applied to complex images [3].

A process known as "Picture colourization" is used (Colour image) to create a coloured image from a black-and-white one, create a coloured image from black-and-white one, to create a coloured image from a black-and-white one. The process of colouring a picture, particularly the colour spectrum employed, has a far more significant impact on industries like astronomy, security photography, and optical microscopy. The quantity of data in most grayscale images is inadequate for accurate interpretation. So, colouring the photos adds the extra information needed to decode the semiotics of the picture [4].

2. RELATED WORK

This pattern recognition, which aims to determine how well-liked a text is based on its image, is still an active study area. There is more than one suggested method for fixing textual reputation issues, and they all take a somewhat different tack. To detect and extract textual information from brutal historical colour record photographs, Shivananda and Nagabhushan [5] proposed a helpful technique. The method uses the edge detection algorithm known as the Canny element detector. The operation will be carried out on the gathered edges. Therefore, it is inferred that the gaps between the modules should be reduced to a minimum. All the connected, holeless additives were taken out. The average variance of each component was calculated and interpreted to weed out anything else that didn't belong in the material. Noisy textual content sections were discovered and treated in the end to enhance the quality of the recovered foreground. The methodology for extracting text from graphical pictures was established by Hoang and Tabbone [6]. Sparse illustration framework creation is described using morphological component analysis (MCA). Transform and curvelet transform was the main foundations for two discriminative dictionaries.

To analyze the contents of a photograph, Angadi and Kodabagi [7] employed the discrete cosine transform (DCT) technique. This procedure involves fine-tuning the editing and parsing of nearby blocks of textual data and unprocessed regions to increase sensitivity and accuracy. Additionally, this approach detects nonlinear research subject areas and can be derived from images in other languages with minimal adjustment, and this entire flow is shown in Figure 1.

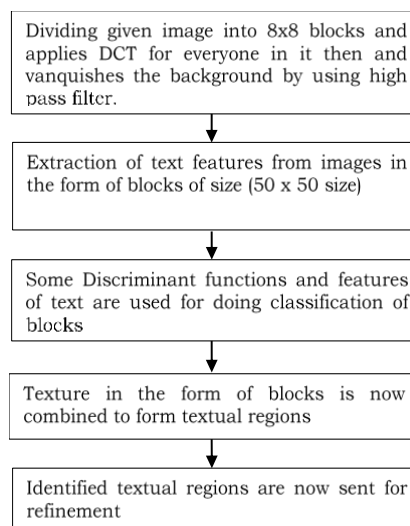


Figure 1. Extracting text regions from low resolution

Devareddi and Srikrishna [2] employed a long short-term memory (LSTM) model to sharpen blurry pictures at various resolutions. The resulting image was restored to its original shape at the decoder. Because

of this, it not only produces precise results but also decreases the time complexity involved. Combining convolutional neural networks (CNNs) with an existing system helps advance the fields of digital image processing, image interpretation, and image categorization. They also use tensor flow and an image data generator to eliminate the possibility of overfitting [8]. The achievement of the Devanagari script was also established by Sankaran and Jawahar [9]. They developed a strategy to do away with text extraction, a typical cause of high word mistake rates. Sundaresan and Ranjini [10] proposed discussing extracting English text from a comic book blob. Text detection and extraction from comedian pictures must be performed to preserve the text information and encoding throughout the phase change. The red, green and blue (RGB) pictures are transformed into binary images during this stage. Figure 2 shows how extensively blob size characteristics are employed in the identification process.

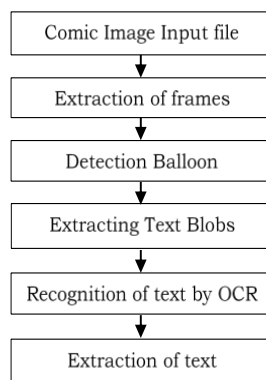


Figure 2. Two blobs extraction method for extracting digital English comic image

Automatic text area and identification in a coloured book is a method described by Sobottka *et al.* [11]. They apply a clustering rule set to simplify the pre-processing phase to minimize the number of colours transformed. Two main approaches may be used when extracting text. Top-down analysis is one method in which the image is divided into sections individually. Grover *et al.* [12] presented a fast and reliable method for extracting text-like areas from printed materials. It utilizes a differentiated text ranking strategy to screen for natural language edges and a thresholding approach for non-text edges. Attaching the remote candidate text before generating the line characteristic vector graph based mainly on the threshold map is necessary. Silpa and Rao [13] proposed an optimization technique for web data by using a genetic approach and also proposed ML models to investigate the behaviour of web users [14] and did work on the web data by using big data [15]. Bukhari *et al.* [16] suggested an approach for extracting colour text lines from grayscale camera-captured record images. This method employs local gradient paths as a quantitative indicator of textual significance and is grounded on real differential geometry. By applying multi-oriented smoothing, this line of grayscale text is improved. This information is used to spot the point where the strongest directional derivatives of a smooth image cross over to zero.

Using CNN, Shankar *et al.* [17] have developed a method to automatically create models for grayscale photos. A feature extractor was used to pull the features out of the encoder's output at the fusion layer. They determined that CNN would perform better than other methods for colourizing black-and-white photos, and this finding was based on the outcomes of their performance metrics. Using a fuzzy filter to efficiently eliminate noise from an image without altering the original image [18] will ensure the image's integrity is maintained. Arai and Tolle [19] have suggested that digital comics be read on a mobile device. Adding new features and fixing bugs to the existing interactive comics website are worthwhile endeavours instead of developing new ideas for cellular humour. It is proposed that an automated e-comic smartphone content versioning method be set to mechanically produce mobile comic content from a distant comic website repository. Figure 3 depicts the recommended procedure, the most efficient means of introducing real-time processes. Experiments showed that the flat comic body extraction method was 100 % accurate. The non-flat comic body extraction method was 91.48 % accurate, and processing time was reduced by 90% compared to the prior method.

Das *et al.* [20] used a technique for extracting compensatory characteristics that work globally and nearby. There are several technologies at play here, notably the ability to project six different shapes and curve four different shapes. To acquire neighbouring operations, we first partition the picture into nine equal

portions; then, we measure four gradient features per row for 36 operations. Each organization's personalities have unique characteristics based on a formula. Recognizing and extracting text from Gur and Zelavsky [21] posed little difficulties. Although OCR may help with many problems, human review is usually necessary to get the whole picture (OCR). Newer technologies use letter statistics and correlation coefficients to correctly identify distorted lyrics based on hazy, primarily dependent rules. Figure 4 shows the Rashi type faces the researchers focused on because they correlated with biblical commentary; these fonts are distinctly handwritten.

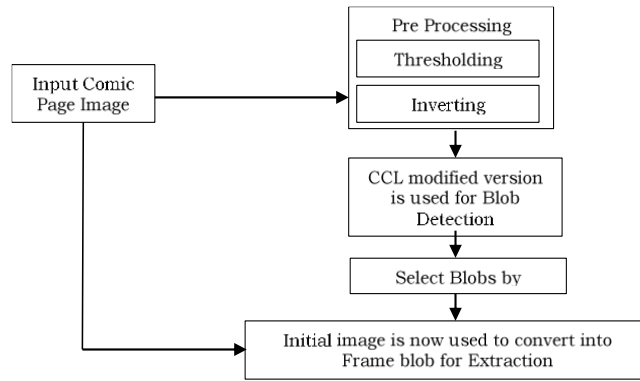


Figure 3. Automatic E-comiccontentadaptations

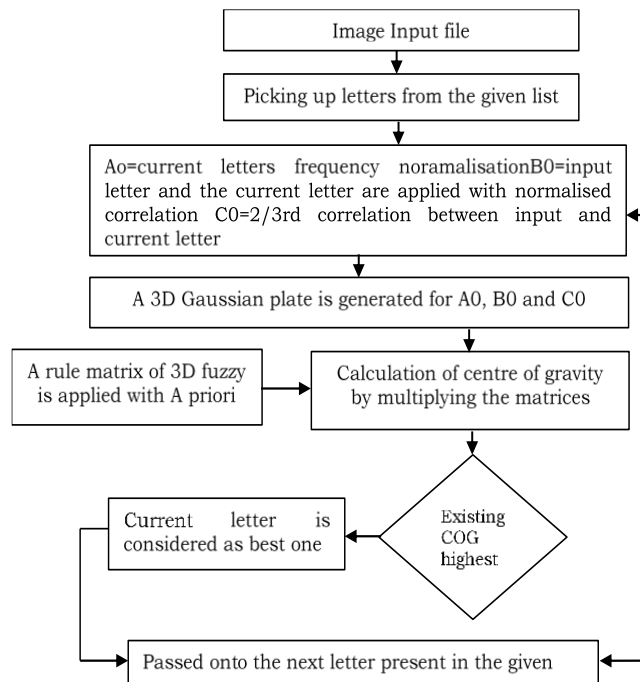


Figure 4. Retrieving content from Rashi semi-cursive handwriting via fuzzy logic

Yang *et al.* [22] developed a new, original method regarding the wavelet filter. Figure 5 illustrates the possibilities for real-time computing and cellular applications enabled by this method's rapid strategy transformations. The individualized approach was tested on a library of complicated scene photographs. Yin *et al.* [23] offer a unique maximally stable extremal regions (MSER)-based scene text identification strategy. First, they present a method that prunes MSERs rapidly and precisely, allowing us to find and categorize most characters despite poor image quality. Finally, they present a classifier that approximates a nominee's posterior text probability and rejects high-risk candidates. Shankar *et al.* [24] worked on facial recognition using Beizercurves and portable grey map (PGM) images by removing Noise reduction [25].

Chen and Yuille [26] recognize and analyze text in natural images. This programme helps the blind and visually impaired navigate a city. They start by gathering photographs taken in cities by the blind and seeing things. They manually mark and delete dataset text. Next, the text areas are examined to see whether the picture characteristics are legitimate texts. A hybrid method for scene text localization that includes local information into an applicable CC-based procedure is provided by Pan *et al.* [27]. Notably, the binary contextual connection is integrated into a conditional random fields (CRF) model where both the supervised learning parameter and the unary relationship play crucial roles. The approach works for messy or ordered situations with open or regular layouts. Regional information helps segment and analyze text and classify non-text items using Devedreddi *et al.* [28] worked on image segmentation using scanned documents with neural networks.

Merino-Gracia *et al.* [29] built a portable text recognition platform using MSERs for real-time text detection. Using MESRs' hierarchical structure, the suggested technique refines the prior real-time algorithm to create more safe zones than the adaptive threshold approach. It outperforms previous algorithms and retains International Conference on Document Analysis and Recognition (ICDAR) text detection efficiency. In the future, they plan to explore character recognition without third-party apps and expand on their cascade filters to improve speed. Pixels in an image may be segmented on a variety of semantics, such as their membership in a particular class or instance. They introduced masked-attention mask transformer (Mask2Former) (panoptic, instance or semantic) to tackle any image segmentation issue. Among its major components is masked attention, which extracts localized characteristics by confining cross-attention in expected mask areas. It reduces research effort by three times and outperforms specialist designs on four prominent datasets [30].

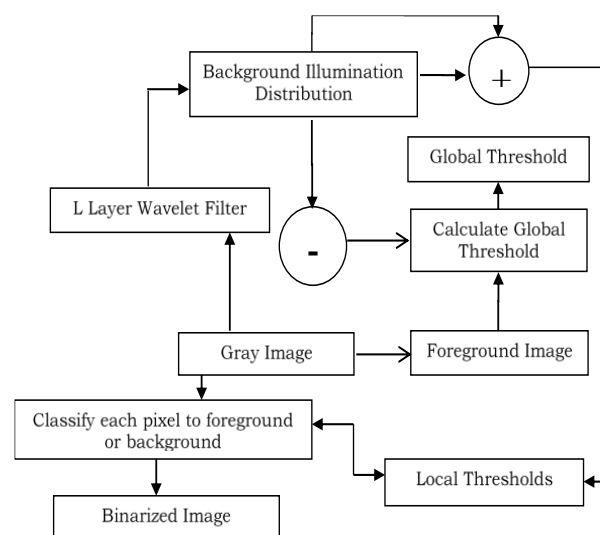


Figure 5. Adaptive binarisation method for the complex scene images

3. METHODOLOGY

Here, techniques include a few steps written in Figure 6. Maximally Stable External Regions (MSER) is a set of rules that detects the text region inside the photograph and exists in the result. MSER detects main text areas from natural scene images. However, a few non-text regions were detected in the next step. MSER uses two vital houses to remove non-text content in a region by photograph first is geometric properties, followed by stroke width. By the non-textual content, characters are merged by a group of words. Finally, the OCR detects the text contained in every bounding box.

3.1. Objectives

To achieve a personalized user experience the user to retrieve the text from an image in a customized manner. So, the user feels more contended. To indulge more users, the embedded text extraction system, the user can quickly get the exact text from any image. So, more users start using the service.

Step 1: Detection of text regions: The MSER algorithm finds text regions from all the areas in the given photo. They exist many non-text regions, and textual content can be eliminated. MSER first converts the colouration image into a grayscale photograph to detect textual content areas.

Step 2: Removing non-text regions involves the geometric properties: The MSER set of rules chooses most of the textual content and detects other stable regions within the photograph that aren't textual content. We can use a rule-based approach to locate and put off non-text areas present inside the photo. The usage of geometric homes filters the image's non-textual content areas. Besides, we are ready to maintain the system learning method by educating the main text on geometric things, and the learning technique gives the best.

Step 3: On removing the non-text regions based on stroke width, an unusual metric was used between the text and non-text. The main variance is to be occurred by the width of the curves, and the contours made up of person are contained by non-text and a little bit by text regions.

Step 4: The outcomes of merging text regions are done by personal textual content characters. Permits recognition of the absolute phrases on a photo, which contain more meaningful records than just the specifications of the man or woman.

Step 5: By detecting the textual content areas, use the OCR characteristic to understand the textual content within every bounding box.

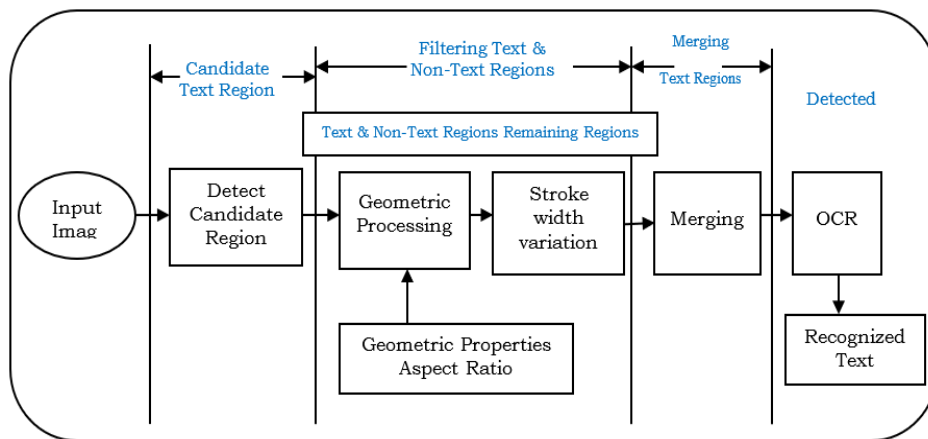


Figure 6. System architecture

3.2. Existing system

Different imaginative challenges have been attained in the past few years by growing attention to text-based records in photographs. Extracting the text from an image could be challenging as it can contain different variations in size and style and sometimes have massive text in the picture and some blurry backgrounds with contrast lighting. The existing system is not that efficient for extracting the text from an image, and it contains the following steps:

Step-1: Detecting features

Step-2: Finding and fixing using localization

Step-3: Tracking

Step-4: Text extraction and improvement

Step-5: Final text recognition is done using the function OCR

Disadvantages of the existing system: Due to complex backgrounds and variations of font, size, colour and orientation, textual content in natural scene images has to be robustly detected before being diagnosed and retrieved.

3.3. Proposed System

As the First step, we apply the MSER algorithm to find the images' textual contents and predict the results. This algorithm and text parts in the image also detect some unwanted regions (Thus may contain symbols or be a complete non-text region). Removal of such regions will be further done in the following steps. The proposed algorithm mainly uses two critical geometric properties and stroke width variation properties, which are used to eliminate most of the plain regions from the image. After removing the plain and unwanted regions like non-text regions from the image, we combine all the identified characters to form phrases or combined text words. The merged content is finally recognized by OCR, which will identify the text within the bounding box. Later this will be displayed in the form of output to the user.

4. IMPLEMENTATION

Image converters may be thanks to executing particular properties on entering photos, containing textual and non-textual content to take out valuable statistics from the pictures or beautify the pictures. An image converter may be a signal converter in which input takes the image, and the output will be the photographer's features/capabilities related to that image. In our day-to-day life, image converter is a fast-growing technology. Image converter involves the following three points: i) Importing the photograph; ii) Analyzing and arranging the photograph; and iii) Output where the result is also altered picture or report based on image evaluation.

Texts that can be seen are used often throughout the day and provide an essential function as a representation of the language used by individuals. Things in our immediate environment appear to be communicated by the text. Images with accompanying text preserve some approximation of the image's semantics for future reference. We also examine and use the MSER algorithm and its governing principles. The MSER algorithm can identify every letter in a picture, regardless of the image's resolution.

4.1. Definitions of the phrases of textual content retrieval

In Text Document the characters had some smooth history, like a few scanning papers in a file text. The Text texture includes the pixels and the textual media content. The Local history nearby background carries a particular textual content region. Near the historical past, there is a mass of embedded text in the snapshots.

4.2. Application of the textual content recognition

Extra records are stored virtually nowadays in various forms, including snapshots. The file contains many characters that may control that particular character. A wealth of information is included in the words that appear in the image. The photographs include names, places, dates and hours, product manufacturers, street signs and symptoms, which might assist in understanding the image. Words recognition from images aims to extract text from complex images.

4.3. Scientific validation

In report text processing, the challenge is extracting text from random history. Snapshot text may also vary in size and typeface style. As a matter of honesty, we enlarge the images to conform to the standard format of the archival text. An example of text processing is shown in Figure 7. We selected an Input image for the text processing, as shown in Figure 7(a). After the selected image was pre-processed, an output image was generated, as shown in Figure 7(b).



Figure 7. Validation of Images as shown in; (a) Text processing image as an input image, and (b) After pre-processing image as an output image

- The text detection algorithm

Image text and blob identification may be possible with the help of the maximally stable extremal region (MSER) method. The MSER rule removes a picture's various covariant areas (Text regions). Let's start with a quick review of the stroke idea and then go on to the width transformation. MSER relies on locations sharing similar lifestyles over a wide range of thresholds. If the pixel value is more than the specified threshold, it is interpreted as black; if it is less than the specified threshold, it is interpreted as white. Geometric properties and stroke width variation properties are used in the MSER method to eliminate the text and non-text portions, respectively. To begin using the MSER algorithm, we evaluate the text properties as follows: i) Text in images often contains a large number of corners; ii) The text width is often more

significant than the text height; iii) The text dimension is often bounded; and iv) This property's text has a unique function in that the texture is odd.

The following Figure 8 depicts the MSER algorithm's flowchart. MSER algorithm identifies all language regions as well as certain non-textual content regions. Geometric properties are added to photographs to exclude non-textual areas. If geometric properties cannot eliminate all non-textual content areas, use the stroke width variation properties on photographs to eliminate any remaining non-textual content areas.

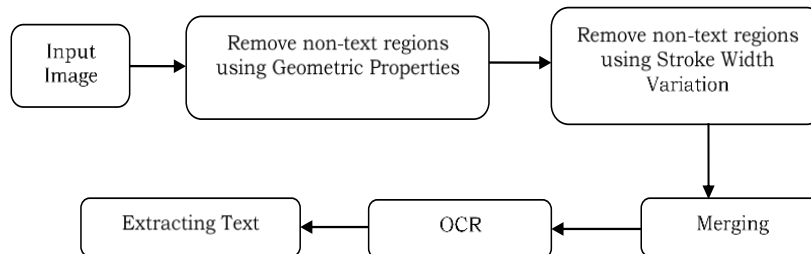


Figure 8. Flowchart of the MSER algorithm

4.4. Geometric properties

The MSER component locates and names many images in a given text. But other than writing, there was found to be other stuff. Locations without words put a premium on geometric qualities like simplicity and unpredictability. Since almost all portions of an image that contain non-textual information were eliminated, geometric properties were used. Any non-text portions that are still present after applying the Geometric Characteristics are extended to the Stroke width difference properties. The entire process was written in Algorithm 1.

ALGORITHM 1: Finding stroke width

```

Input: Binary Image BW
Output: Stroke width image SW
D= Distance Transform (BW); D=round (D);
For p=each foreground pixel in D do
    PVal=D(p);
    Lookup(p)=p's 8 neighbours whose value < PVal;
End for
MaxStroke=max(D);
For Stroke=MaxStroke to 1 do
    StrokeIndex=find(D==Stroke);
    NeighborIndex=Lookup (StrokeIndex);
    While NeighborIndex not empty do
        D(NeighborIndex)=Stroke; NeighborIndex=Lookup
(NeighborIndex);
    End While
End for
Return SW=D;
  
```

4.5. Stroke width variation properties

It's possible to utilize these settings to eliminate everything except text. A character's form and level of intricacy may be determined by the stroke width, which is a design parameter. Stroke distance is not as malleable outside of text areas. Defining a minimum width value will allow us to filter unnecessary data. In a pixel-based display, black indicates that the pixel is higher than the preceding one, whereas white indicates the opposite.

4.6. Stroke width filtering

In character recognition, stroke width is the distance between two stroke edges perpendicular to the centerline. The stroke width is the distance between two text aspects on an orthogonal axis. Due to the inconsistencies in the stroke algorithm that define single characters, non-text sections often alternate between solid and patterned backgrounds.

4.7. Converting a group of letters into texted lines

Here, isolated letters are joined together to form words, sentences, and even whole lines of text. Converting text to outlines creates artwork (shapes) and eliminates font information. This makes the content uneditable but eliminates font mistakes so we can print it correctly. These statistics are much easier to comprehend and express than individual characters.

- Working process

This method shows that finding recognized areas in photographs, including text, is a regular job accomplished on unformed scenes. The term "unformed scenes" describes visual representations of inconsistent or incoherent text. To alert a vehicle to an avenue sign, you may, for instance, routinely walk through and recognize textual content in recorded video. In contrast to the concentrated scenes, which employ recognized scripts in which the position of textual material is known in advance, this one doesn't. Optical character recognition may be used for many activities (OCR) when the scene is divided into sections. In this scenario, the text detection ruleset finds numerous candidates in the text neighbourhood and gradually eliminates them until only the most likely candidates remain.

Step 1: Detects text regions using MSER.

Locating functional regions inside texts is made easy using the MSER method. Consistent colouring and thorough text comparison provide steady intensity profiles, making it a good choice for text. Identifying and visualizing all image areas is possible using the MSER attributes encountered during travel. Due to the abundance of text fields, searching for non-textual indicators in the visual information is necessary.

Step2: Removing the non-textual regions by using geometric properties.

The MSER algorithm and its criteria choose the most relevant text and find several solid regions in the image that are not text. Using rule-based characteristics, we can strip out the irrelevant parts. The best results are usually achieved by combining the two approaches. Apply region properties to degree a few of these properties and then do away with areas primarily based on their property values. Several geometric homes can be appropriate for distinguishing between textual content, and a non-textual content region includes:

- Aspect ratio: It is the width-to-height ratio.
- Eccentricity: It depends on the circular nature of the given areas.
- Euler wide variety: It is a feature of the binary photograph
- Extent: The place and length of the rectangle.
- Solidity: The percentage of the pixels within the raised shape area are likewise in a given place. It is calculated by $\text{Area}/\text{raised area}$

Step 3: Use stroke width variation to exclude non-textual regions from a picture.

Stroke width properties are sometimes used to distinguish text and other content further. The stroke width quantitatively measures the total area covered by a series of strokes or curves. There may be few possible stroke interpretations in textual sections, but many in non-textual information areas. Analyzing the stroke width of each MSER area found might provide light on how to filter out irrelevant content portions.

Step 4: Merging of textual regions for text detection results.

At this stage, a person's textual material creates all detection outcomes. Finding adjacent textual sections and then forming a bounding container around them is one method for organizing textual character areas into text traces. Increase the boundary bins that were previously calculated using area props to find neighbouring regions. A result of this is that it causes the borders of adjacent text sections to get boxed.

4.8. Connected component analysis

Once area surroundings have been detected, eliminating areas that are not separated is often beneficial. A named unit is any group of pixels not distinguished by a border. A joint component is a connection between the pixels of a maximal position. Segmentation is beneficial for an image to form a distinct subset of the collection of components for specific image processing systems.

5. RESULTS AND DISCUSSION

This developed system automatically describes the text of images, which may be challenging. It could provide more accurate and compact text present in photos when the image is given as input to the system and is explained by considering an image, as shown in Figure 9. The next step is to transform it into a grayscale picture, which means that the input colour image is turned into a black-and-white image, as shown in Figure 10(a), by feature extraction through MSER, as shown in Figure 10(b).

The removal of such non-text regions depends on the geometric properties of the regions that contain both text and non-text. One of the regions in random is automatically considered, and stroke width variation is

applied to it. Later that region is converted into a skeleton image to classify the characters easily, as shown in Figure 11(a). An inverted region image is shown in Figure 11(b). After that, a stroke is applied to an image, as shown in Figure 11(c) and from that, we obtained a skeleton image, as shown in Figure 11(d). From the image, i.e. Figures 10(a) and 10(b), one of the characters named D is recognized, as shown in Figure 11.



Figure 9. Consider the image as the Input image



Figure 10. Transformation of images as; (a) image to black and white

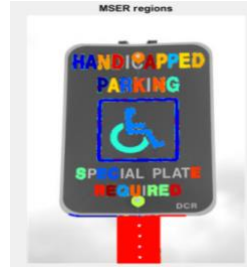


Figure 10. Extraction as; (b) Feature Extraction through MSER

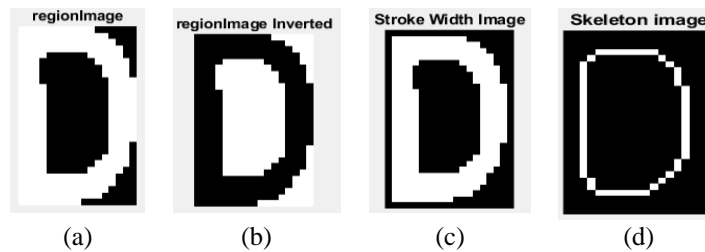


Figure 11. Removal of non-text regions as; (a) Region image, (b) Inverted region image, (c) Stroke width image and (d) Skeleton image

We applied a subplot between the region image and the stroke width image, as shown in Figure 12. We obtain the following picture after extracting all non-text regions from the image using stroke width variation, as shown in Figure 13. Expanding bounding boxes such that each character will be under each bounding box, as shown in Figure 14.

After expanding the boxes, we finally detected text by developing the individual bounding boxes to identify the complete word, as shown in Figure 15. As a result of the entire process, the output screen is shown in Figure 16. Finally, we can easily extract the text after eliminating plain regions on images.

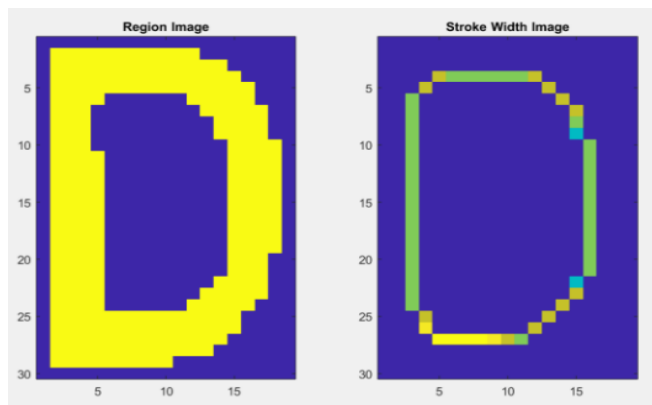


Figure 12. Apply sub plot between the region image and stroke width image



Figure 13. Extracting all non-text regions from the image using stroke width variation



Figure 14. Expanding bounding boxes



Figure 15. Apply sub plot between the region image and stroke width image

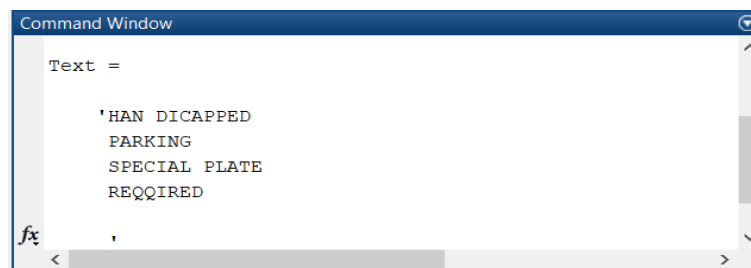


Figure 16. Output screen of detected text

6. CONCLUSION

We used the MSER method in conjunction with the geometric and stroke width variation features to determine the legibility of handwritten text in an image. A good textual reputation performance may be found in MSER. We have utilized MSER to identify the text regions and OCR to recognize the characters inside them. Extension to Video: The approaches used in this exercise are intended to identify text in photographs. Video frames may depend on one another, which must be acknowledged. If a frame's body contains a sentence, that exact phrase will probably appear in nearly the same spot in the next frame. Multi-Script Scene Textual content understanding: Numerous nations, including India, use a variety of scripts because of the country's large population and geographical diversity. Identifying many of these scripts is an ongoing problem in the written text domain, and recognizing scripts in scenes is significantly more difficult.





REFERENCES

- [1] B. Li, X. Qi, T. Lukasiewicz, and P. H. S. Torr, "Controllable text-to-image generation," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [2] R. B. Devareddi and A. Srikrishna, "Review on content-based image retrieval models for efficient feature extraction for data analysis," in *Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022*, Mar. 2022, pp. 969–980, doi: 10.1109/ICEARS53579.2022.9752281.
- [3] N. Vo *et al.*, "Composing text and image for image retrieval-an empirical odyssey," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2019, vol. 2019-June, pp. 6432–6441, doi: 10.1109/CVPR.2019.00660.
- [4] X. Li, J. Yang, and J. Ma, "Recent developments of content-based image retrieval (CBIR)," *Neurocomputing*, vol. 452, pp. 675–689, Sep. 2021, doi: 10.1016/j.neucom.2020.07.139.
- [5] N. Shivananda and P. Nagabhushan, "Separation of foreground text from complex background in color document images," in *Proceedings of the 7th International Conference on Advances in Pattern Recognition, ICAPR 2009*, Feb. 2009, pp. 306–309, doi: 10.1109/ICAPR.2009.26.
- [6] T. V. Hoang and S. Tabbone, "Text extraction from graphical document images using sparse representation," in *ACM International Conference Proceeding Series*, Jun. 2010, pp. 143–150, doi: 10.1145/1815330.1815350.
- [7] S. A. Angadi and M. M. Kodabagi, "Text region extraction from low resolution natural scene images using texture features," in *2010 IEEE 2nd International Advance Computing Conference, IACC 2010*, Feb. 2010, pp. 121–128, doi:




- 10.1109/IADCC.2010.5423026.
- [8] S. S. Reddy, M. Gadiraju, and V. V. R. Maheswara Rao, "Analyzing student reviews on teacher performance using long short-term memory," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 96, Springer Nature Singapore, 2022, pp. 539–553.
- [9] N. Sankaran and C. V. Jawahar, "Recognition of printed devanagari text using BLSTM neural network," *Proceedings - International Conference on Pattern Recognition*, pp. 322–325, 2012.
- [10] M. Sundaresan and S. Ranjini, "Text extraction from digital English comic image using two blobs extraction method," in *International Conference on Pattern Recognition, Informatics and Medical Engineering, PRIME 2012*, Mar. 2012, pp. 449–452, doi: 10.1109/ICPRIME.2012.6208388.
- [11] K. Sobottka, H. Bunke, and H. Kronenberg, "Identification of text on colored book and journal covers," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1999, pp. 57–62, doi: 10.1109/ICDAR.1999.791724.
- [12] S. Grover, K. Arora, and S. K. Mitra, "Text extraction from document images using edge information," 2009, doi: 10.1109/INDCON.2009.5409409.
- [13] N. Silpa and V. V. R. Maheswara Rao, "Machine learning-based optimal segmentation system for web data using genetic approach," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 11, pp. 3552–3561, 2022.
- [14] N. Silpa and V. V. R. Maheswara Rao, "Enriched big data pre-processing model with machine learning approach to investigate web user usage behaviour," *Indian Journal of Computer Science and Engineering*, vol. 12, no. 5, pp. 1248–1256, Oct. 2021, doi: 10.21817/INDJCS/2021/V12I5/211205050.
- [15] N. Silpa and V. V. R. Maheswara Rao, "A complete research on techniques & technologies of big web data preparation to web user usage behaviour," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 2356–2367, Nov. 2019, doi: 10.35940/ijrte.B1269.0982S1119.
- [16] S. S. Bukhari, T. M. Breuel, and F. Shafait, "Textline information extraction from grayscale camera-captured document images," in *Proceedings - International Conference on Image Processing, ICIP*, Nov. 2009, pp. 2013–2016, doi: 10.1109/ICIP.2009.5413799.
- [17] R. S. Shankar, P. Neelima, V. Priyadarshini, and S. R. Chigurupati, "An approach to classify distraction driver detection system by using mining techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 3, pp. 1670–1680, Sep. 2022, doi: 10.11591/ijeecs.v27.i3.pp1670-1680.
- [18] M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 11843–11866, May 2018, doi: 10.1007/s11042-017-4834-3.
- [19] K. Arai and H. Tolle, "Automatic e-comic content adaptation," *International Journal of Ubiquitous Computing*, vol. 1, no. 1, pp. 1–11, 2010.
- [20] R. L. Das, B. K. Prasad, and G. Sanyal, "HMM based offline handwritten writer independent english character recognition using global and local feature extraction," *International Journal of Computer Applications*, vol. 46, no. 10, pp. 975–8887, 2012.
- [21] E. Gur and Z. Zalevsky, "Retrieval of Rashi semi-cursive handwriting via fuzzy logic," in *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, Sep. 2012, pp. 354–359, doi: 10.1109/ICFHR.2012.262.
- [22] J. Yang, K. Wang, J. Li, J. Jiao, and J. Xu, "A fast adaptive binarization method for complex scene images," in *Proceedings - International Conference on Image Processing, ICIP*, Sep. 2012, pp. 1889–1892, doi: 10.1109/ICIP.2012.6467253.
- [23] X. C. Yin, X. Yin, K. Huang, and H. W. Hao, "Robust text detection in natural scene images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 970–983, May 2014, doi: 10.1109/TPAMI.2013.182.
- [24] S. Shankar, J. Raghavani, P. Rudraraju, and Y. Sravya, "Classification of gender by voice recognition using machine learning algorithms," *Journal of Critical Reviews*, vol. 7, no. 9, pp. 1217–1229, Jun. 2020, doi: 10.31838/jcr.07.09.222.
- [25] R. S. Shankar, V. S. Raju, K. V. Murthy, and D. Ravibabu, "Optimized model for predicting gestational diabetes using ml techniques," in *Proceedings of the 5th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2021*, Dec. 2021, pp. 1623–1629, doi: 10.1109/ICECA52323.2021.9676075.
- [26] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, doi: 10.1109/cvpr.2004.1315187.
- [27] Y. F. Pan, X. Hou, and C. L. Liu, "A hybrid approach to detect and localize texts in natural scene images," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 800–813, Mar. 2011, doi: 10.1109/TIP.2010.2070803.
- [28] R. B. Devareddi, R. S. Shankar, K. V. Murthy, and C. Raminaidu, "Image segmentation based on scanned document and hand script counterfeit detection using neural network," in *AIP Conference Proceedings*, 2022, vol. 2576, doi: 10.1063/5.0105808.
- [29] C. Merino-Gracia, K. Lenc, and M. Mirmehdi, "A head-mounted device for recognizing text in natural scenes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7139 LNCS, Springer Berlin Heidelberg, 2012, pp. 29–41.
- [30] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2022, vol. 2022-June, pp. 1280–1289, doi: 10.1109/CVPR52688.2022.00135.

BIOGRAPHIES OF AUTHORS






Shiva Shankar Reddy     is an Assistant Professor at the Department of Computer Science and Engineering in SagiRamaKrishnam Raju Engineering College, Bhimavaram, Andhrapradesh, INDIA. He is pursuing PhD degree in Computer Science and Engineering with a specialization in Medical Mining and Machine Learning. His research areas are Image Processing, Medical Mining, Machine Learning, Deep Learning and Pattern Recognition. He published 30+ papers in International Journals and Conferences. S.S. Reddy has filed 05 patents. His research interests include Image Processing, Medical Mining, Machine Learning, Deep Learning and Pattern Recognition. He can be contacted at email: shiva.csesrkr@gmail.com.



Dr. Vuddagiri MNSSVKR Gupta    holds Master of Science (M.Sc.), M.Tech. Degree in Computer Science and Technology, Ph.D. in Computer Science and Engineering, besides several professional certificates and skills. He is currently Professor in the Department of Computer Science Engineering at SRKR Engineering College, Bhimavaram. Andhra Pradesh, India. In addition, he is serving as PG Coordinator in Computer Science and Engineering. His research areas of interest include Data mining, Image Processing, Medical Mining and Machine Learning, He has written 30+ International Journals and conferences. He can be contacted at email: guptavkrao@gmail.com.



Lokavarapu V. Srinivas    is an Assistant Professor in the Department of Computer Science and Engineering in Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India. He Received B.Tech. and M.Tech. Degrees from Andhra University, Visakhapatnam. Currently He is pursuing Ph.D. in Computer Science and Engineering at Andhra University. His Research areas are Machine Learning, Edge Computing, Natural Language Processing and Image Processing. He can be contacted at email: srinivas.srkrce@gmail.com.



Chigurupati Ravi Swaroop    is Assistant Professor at SagiRamakrishnam Raju Engineering College, Department of Computer Science and Engineering, India. He received B.Tech. degree in SagiRamakrishnam Raju Engineering College, Department of Information Technology in 2012. He holds an M.Tech. degree in SagiRamakrishnam Raju Engineering College, Department of Information Technology, in 2018. His research areas are Image Processing, Bioinformatics, Machine Learning, Deep Learning, and Data Mining. He can be contacted at email: raviswaroop.chigurupati@gmail.com.