



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 10    **Issue:** V    **Month of publication:** May 2022

**DOI:** <https://doi.org/10.22214/ijraset.2022.42932>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Text Detection and Recognition: A Review

Prakhar Pathak<sup>1</sup>, Pulkit Gupta<sup>2</sup>, Nishant Kishore<sup>3</sup>, Nikhil Kumar Yadav<sup>4</sup>, Dr. Himanshu Chaundhary<sup>5</sup>

<sup>1, 2, 3, 4</sup>Dept. of Electronics & Communications Engineering, KIET Group of Institutions, (AKTU), Ghaziabad, India

<sup>5</sup>Assisant Prof. in Dept. of Electronics and Communication Engineering, KIET Group Of Institutions, (AKTU), Ghaziabad, India

**Abstract:** In this review paper we have done extensive reading of various research paper on Text Detection and Recognition from images by different authors of around the world. Each research paper deploys different algorithms and strategies for text detection and text recognition of image. At last, we have compared the Accuracy as well as Precision and Recall Rate of the various methods used in different research paper.

**Keywords:** Accuracy, Precision, recall rate, Digit recognition.

## I. INTRODUCTION

In modern era, the number of digital images capturing devices has been increased substantially. With the increase in these devices the text recognition and detection has turn out to be incredibly important. Text detection and recognition has got wide range of use such as in language translation, automatic license plate reading, document management and many more. In this review paper we explore different research papers which provide different methods for text detection and recognition.

In [1] CNN is used which is basically made up of 3 layers. The 1<sup>st</sup> two layers can be further classified into Convolution Layer ( $C_1$  &  $C_2$ ) & Sub-Sampling Layer ( $S_1$  &  $S_2$ ). Above the 1<sup>st</sup> layer, there is another layer called Input Layer which takes 3 input maps, where each input map corresponds to one color channel of image like RGB.  $C_1$  takes specific features directly from the three-color channels.  $S_1$  results in local averaging and sub-sampling operations in the previous layer  $C_1$  on a corresponding map. Outputs of  $S_1$  different feature maps are combined, allowing  $C_2$  to extract more information. The information transmission between  $S_2$  &  $C_2$  is same as that in  $S_1$  &  $C_1$ .

In [2] AdaBoost algorithm is used. It is a way by which we can aggregate a group of feeble classifiers & convert it into an efficient & a powerful classifier. It is followed by binarization which is a process. process by which a pixel image can be converted into a binary image. In addition to this, digits & letters can be detected via a connected component algorithm

In [3] the setup used in character detection & recognition is closely associated with a convolutional neural network [15].

In this, a new type of K-means algorithm of clustering is implemented to yield faster & simpler results in feature learning. The next step is of Feature extraction followed by text detector training. This is followed by the most important step which is Character Classifier Training.

For [4] The image is first collected using a smart device from the outside world, then the edges of a signboard are detected. The MSER is used to recognize text from a signboard. Final In order to classify and recognize text taken from natural sceneries or the outer world the Artificial Neural Networks is used.

In [5] We use both the MSER and colour clustering method for text detection. For the recognition step, a vertical projection of the text area template is used to segment the word picture. To address the problems of word recognition in nature images, we blend global and local characteristics for feature representation.

For [6] the stroke filter is used to obtain corners and then define a basic classification feature for creating a fast and effective text detector. This method has higher computational efficiency as it uses more simple and distinctive features, this algorithm applies SVM to detect text regions

In [7] method used is more robust than other texture based methods as corner response is very effective and reduces noises, Even in complex background it can still detect texts. In large font or small font it can easily detect text . Corner response is easy to compute as there is no need to know the exact location of the corner points, only the possible corner positions are required.

In [8] the method proposed focuses on the text detection and its segmentation in a complex images such as stills from a video. For the text detection it not only considers the density of high contrast points but also the variance of edgels. When a binary image is obtained then connected component analysis is done in order to get the potential text area. Then the obtained potential text area is separated from its background. This binarized image can now be efficiently processed by the OCR software for the text recognition. Thus in this way the accuracy of the overall text recognition increases.

In the recent years many OCR system have been developed which gives high accuracy of about 90%. But their performance degrades significantly when they are introduced to bad or degraded quality image. So, in [9] We adopt a different method and extract important geometric characteristics from the digital image. Each character are described by its unique features like shape, corners etc. Each feature generated in this image describes the character it represents. Even if some of the features is missing, the rest of the feature represents the character. This approached has originated from the concept of occluded-object recognition.

The achievement of text recognition is largely reliant on text detection first. Hence in [10], the proposed method has an output which gives us Text detection method that is both quick and trustworthy. The image processor used here is known as stroke width transform ,as it changes the picture information from having the colour value per pixel to most likely stroke width for each pixel. The proposed method is able to detect text regardless of their font, size, language and direction. The accuracy of OCR is less when applied on a natural scene image, this can be because of several reasons like camera quality and several natural scene images are clicked from mobile devices, this further reduces the accuracy of OCR. OCR systems are mostly designed to recognize scanned text and As a result, segmentation is required to distinguish the text region from the background pixels. This is simple in scanned images but not in natural scene images. Also natural scene images have several image noise such as blur, color noise, etc. At last, the natural scene images are not as well structured as scanned images, and it has a different geometry and structure. Images have less text and have varying features. So to counter these problems stroke width transform as an image operator was proposed in this algorithm. The movement of algorithm is in this way

Firstly an edge map of image is made with the help of sobel edge detection, canny edge detection can also be used. Then stroke width is stored for each pixel according to likely strokes. Then we find letter candidates within the output of SWT. Then geometric filtering takes place to remove possible non-text components. Then each character component is aggregated according to their height to get the detected text lines now finally this is divided into detected words and a mask is generated.

In the modern world the use of captcha has increase significantly as they can separate a real user and a automated user as captcha are easy to understand by humans but it is difficult for a machine to recognise as they are distorted, merged and bent. But in some situations, we need to recognise captcha by a machine. For instance, when a login of app or website is being tested before launching it for public the tester needs to manually enter captcha when they want test a particular feature this becomes a cumbersome task for him. So, we need to recognise captcha by machine itself. So, in [11] a method is demonstrated for captcha recognition. This involves Pre-Processing followed by segmentation there after recognition is being performed.

## II. LITERATURE REVIEW

### A. Automatic Scene Text Recognition using CNN.

In this, a CNN is used which is basically made up of 3 layers (from top to bottom): Feature Extraction Level, Feature Combination Level & Classification Level. The 1<sup>st</sup> two layers can be further classified into Convolution Layer ( $C_1$  &  $C_2$ ) & Sub-Sampling Layer ( $S_1$  &  $S_2$ ). Above the 1<sup>st</sup> layer, there is another layer called Input Layer which takes 3 input maps, where each input map corresponds to one color channel of image like RGB etc. The input to this layer is the image itself.  $C_1$  takes specific features directly from the three color channels. In the preceding layer  $C_1$  of the matching map,  $S_1$  results in local averaging and sub-sampling processes. Different feature maps' outputs of  $S_1$  are combined, allowing  $C_2$  to extract more information. The information transmission between  $S_2$  &  $C_2$  is same as that in  $S_1$  &  $C_1$ . The back-propagation algorithm performs the stage of training as described in [13]. This paper, an automatic identification system for a complex color scene text image is being proposed. Due to supervised learning feature, our system doesn't require any tunable attribute & considers both geometrical properties & color distributions of characters. As future work, it is planned to do word recognition via statistical modelling in a more general network.

### B. Detecting and Identifying text in Natural Scenes.

This can be achieved using AdaBoost Algorithm. It's a technique by which feeble classifiers are combined in order to make an efficient & a strong classifier. The feeble classifiers are build using probability distributions. This can be done via a log-likelihood ratio also. Another class of features is also available which are quite cumbersome. These checks are based on intensity gradient, intensity & gradient histogram direction. The third-class features are mathematically costlier than the tests done previously. They are based on performing of edge detection, thresholding of intensity gradient, followed by linking of edge. The next step is of binarization. It is done by applying adaptive binarization [14 ] detected by AdaBoost classifier. In addition to this, digits & letters can be detected via a connected component algorithm [15], ultimately, enabling us to identify their size & spacing between them. At last, to extended text regions, usage of additional OCR program is implemented. It is done so in order to identify the text but also discard false positive cases.

In future, it will involve alternate text reading software. Though the algorithms of OCR are quite effective, hence they can't be improved or modified. Instead, reading algorithms will be developed based on deformable templates. One more advantage these algorithms offer is they make the use of generative models & can be used on image intensity without using binarization.

### C. Text Detection And Character Recognition In Scene Images With Unsupervised Feature Learning

The setup used in character detection & recognition is closely associated with a convolutional neural network [16].

The first major step is of Learning of Feature. The important part of this system is unsupervised learning algorithms' usage. In this, a new type of K-means clustering is used to yield faster & simpler results. The next step is of Feature extraction. To evaluate feature representation for any 32 x 32 image, we evaluate the illustration for each & every 8 x 8 sub-patch, producing a 25 x 25-by-d representation. The third step is Text Detector Training. For this purpose, a binary classifier is trained for distinguishing the 32 x 32 windows that contains characters from those which don't contain. This extraction is done from ICDAR 2003 dataset. Later, feature extraction method is used for converting each image into a 9-dimensional feature vector.

The ultimate step is of Character Classifier Training. For this operation, a predetermined-size, 32 x 32 pixels input image is implemented to images of text in a set of labelled train and test datasets. It can produce large number of features. But with this, it'll be useful if we have more data. To satisfy the needs, synthetic examples are added which are ICDAR training samples' duplicates, having arbitrary deformities & image filters used. This paper introduces a text detection & identification system which is scalable feature learning algorithm based. It applies on characters present in images of natural scenes. It's clearly visible that accuracy increases with number of learnt features.

### D. Text Detection and Recognition in Natural Scene Images.

In this research paper a technique is developed to identify and detect text in scene photos. The MSER method is unaffected by changes in the region's perspective, size, or lighting. However, when dealing with blurred pictures or parts that are consistently brighter than the backdrop, the method's effectiveness will suffer dramatically. So, we use both the MSER and colour clustering method for text detection.

Clustering reduces the number of colours in the images then the connected component analysis is performed. So, by clustering the The whole character region in the picture is given the same colour. For filtering out the non-text area from the image obtained after clustering they have used saliency filter with some past data about the image.

The MSER algorithm can detect texts with a high brightness contrast against the backdrop. MSER turns the picture to a grayscale image before converting it to a series of binary images using a continuous stream of thresholds. Constant areas arise, expand, and solidify when the brightness threshold value is increased or decreased. Area dissimilarity among the two different boundaries can be treated as stable, if it doesn't exceed a certain prescribed value.

For recognition stage technique grounded on the vertical projection of the To segment word graphics, text area blueprint is used. Character spacing is done here by calculating troughs of vertical projection curve. To address the problems of word recognition in nature images the blend of global and local characteristics for feature representation. For testing this technique the public text detection ICDAR 2003 competition dataset to is used to method's performance. Precision rate and recall rate are the assessment criteria, which may be represented as:

$$p_r = c_t / |E|, r_r = c_t / |T| \quad [17]$$

where  $p_r$  s equal to the total number of estimated targets  $c_t$  divided by the number of properly estimated targets by recall correctly estimates IEI. Recall rate  $r_r$  is equivalent to properly approximation of the goal number  $c_t$  divided by the amount of goals in the original picture.  $f$  represents combination of recall rate and precision rate. [14] It can be observed that the algorithm has an accuracy rate of 68 percent and a recall rate of 60 percent, which is significantly higher than previous algorithms. It can be shown that this strategy, which combines local HOG features with global vertical histogram features, outperforms the usual traditional method.

### E. Signboard Detection and Text Recognition Using Artificial Neural Networks.

The image is first collected using a smart device from the outside world, then the edges of a signboard are detected. Then the text detection and analysis step follows . The text may be recognized in two languages: Urdu and English. Detecting text is a difficult task because the writing on the signboard may be blurry, have broken ligature, and vary in size, color, resolution, form, texture, backdrop, text geometry, lighting issues, and so forth. The text detection is performed using the following equation

$$H_{r1} = \sum f_{i,j}$$

The MSER is used to recognize text from a signboard. Final In order to classify and recognize text taken from natural sceneries or the outer world the Artificial Neural Networks [18],SVM classifier[19], HOG method [20] is used . Experiments using this method is performed on 500 images containing natural scenes in which 70.2% of items where correctly identified and recognized.

#### F. Fast and Effective Text Detection

Stroke filter is used on the test image and then stroke strength is characterized in four different directions which are left diagonal, horizontal, vertical, right diagonal respectively. Initially, the original image is change into grayscale picture. Stroke filter response is calculated as

$$R_{\alpha,l,w}(x,y) = \frac{|u_1 - u_2| + |u_1 - u_3| - |u_2 - u_3|}{\max(\sigma_1, 2)}$$

With a Width\*Height sliding window, the SVM detects possible text blocks. Then for each sliding window features are extracted and sliding windows are classified in text and without text regions by SVM. Then all text blocks are expressed as binary mask image. The intensities that are closer to the middle of the sliding window comprise of higher weights.

VAP & HAP are defined for characterization of spatial distribution of strokes . Every sliding window is vertically divided in 8 rectangular regions in the vertical stroke map and in each of these regions the VAP is calculated as

$$VAP[i] = \sum_{(x,y) \in S_i} R_{\frac{\pi}{2}}(x,y)$$

Every sliding window is horizontally divided in 4 rectangular regions in horizontal stroke map and in each of these regions the HAP is calculated as

$$HAP[j] = \sum_{(x,y) \in M_j} R_0(x,y)$$

Hence a 24 dimensioned feature vector represents the region covered by sliding window. Pixels of sliding window are called as text pixels when output of SVM is positive. So, a binary mask image is created in which the black regions are non-text blocks and white region consists of text.

The white region is a polygon which is then divided in several rectangular regions, if the horizontal distance between two rectangles is found to be less than one-sixth of their total width then the rectangles are joined. If two rectangles area adjacent vertically and one has its width exceed by four-fifth of the other rectangles width, then merge these rectangles and the width is chosen according to the rectangle which has greater width, height becomes the sum of individual heights of rectangles.

Then we finalize text line verification, by extracting a six-dimensional feature for each text line and is verified by new SVM classifier.

#### G. A Novel Text Detection and Localization Method Based on Corner Response

For this method, the text regions are found in images using corner response. Start by computing the corner response in the multiple scale. A corner has a very large curvature in region boundary. A corner can be located by computing local maxima of corner response. Some of the advantages of using corner response are that it just detects the possible regions which can have corner and it could get a repeated value for every pixel which makes it easier for the upcoming procedures. Based on Corner Response we generate the candidate text region. First possible corner regions are divided into small blocks of 8\*8. Then for each block of corner response mean intensity value is computed. A limit is set for block and in case the condition is satisfied then that region is considered as text containing region. The limit here is very less as several pixels in corner response are zero. With that no text regions are lost and all the noise can be removed by following methods. The text is often uniform and different colored as compared to other regions and the gray value deviation in text containing regions are also less as compared to background. So, with help of this the background noises are removed. After removing noises, the text regions are detected after verifying, but its shape is not regular, therefore for each component that is connected, the region is enlarged by 4 pixels along the border, after that to locate the regions a bounding box is used. Then a gauss filter is used to smoothen the curve that is generated by summation of the intensity in Corner response for every row and column in any bounding box. After that a limit(or threshold) is set that is used for locating the text containing region. In this method if the threshold is put to 30% the result is better.

#### H. Text Detection and Segmentation in Complex color Images

For this research paper a new method is being suggested for A text identification and segmentation technique intended specifically for use on colour photos with complicated backgrounds. Here the main purpose of this method is to reduce the amount of false alarms and efficiently binarize the discovered text regions so that typical OCR software can process them more accurately and hassle-free.

For the purpose of finding a possible text area We discovered that a region has a high chance of harbouring characters not only if the density of high contrast points is great, but also if the variance of edgels is high. ( A pixel in an image that is recognised as the edge of something) orientations is high. High contrast spots must be discovered within a particular granularity range. We use the fast recursive deriche edge detector [23], which has a parameter cv for regulating the level of granularity at which edges must be identified. Using Deriche Filter we find Zero Crossing Edgels which resembles local maximum point of the obtained curve in magnitude along the edgel directions. On applying the high threshold value T1 on these zero crossing edgels having some specific magnitude a We get a binary picture with a lot of high contrast spots.

The eight-connected active pixels are then grouped together using a connected component analysis, and the resultant connected component is known as a text blob. Following that, text blobs in a certain neighbourhood are merged using adaptive row and column distance tolerance, for example, if the top and bottom lines of two blobs are aligned. Then these text blobs are tested for their content by setting horizontal alignments as constraints .When a text blob is chosen, it must not include more than 25% of the maximum number of text points discovered in a single line of the blob. Thus in this way the potential text area is detected

Now the identified potential text area has to be separated from the background for which the colour quantization process is used. This divides each text region into four leading colours, assuming that the text and backdrop will be totally separated by assigning one or more of these four colours to them. The backdrop colour is chosen because it has the highest number of occurrences. A binary picture is created for each combination of the remaining non-background colours (maximum of 7 possibilities). By analysing the periodicity of their vertical profile's local maxima, all of these binary pictures associated to one probable text areas are identified as text or non-text.

For testing images from various video frames of MPEG videos were used. The video material was provided by by the Institut National Audiovisuel, France and by ERT TV, Greece. The test data contained 200 jpeg images having different background. On using the proposed method 93% line were detected and was binarized with success rate of 82% .

#### I. Word Recognition in a Segmentation-free Approach to OCR

The notion of occluded-object recognition supports segmentation-free method to OCR proposed in this study. These methods look for the largest subset of object characteristics in a picture. The item is recognized when a subset of consistent characteristics is discovered. This process is repeated until all of the items in the image have been identified and segmented out. It even works even if the object is distorted.

Four steps involved in this process are: (1) extracting the features, (2) To match the features, (3) weighted voting, and (4) recognizing the words.

There are two possibilities. For starters, two separate letters in certain typefaces may have similar or almost identical designs. Except for the small slanted tip at the top, "I" in Times typeface is quite identical to "l". To avoid getting any such problems, we should be careful in selecting characters. For recognizing words lexicon based we follow the three steps which are: (1) selecting candidate words from the lexicon, (2) to find any instance of candidate word and (3) to check for the strength of the candidate word. As gaps in words are considerably larger than in character, they may be identified with far more accuracy than character breaks. Another reason is that work of word recognition is much easier than word segmentation.

The OCR system that is free from segmentation was tested on several pages scanned from magazines and various other newspapers. It shows that on an average 80% of the words were correctly recognized. The results from the experiment shows that OCR system free from segmentation can easily recognize words which rather a standard OCR system would face difficulty.

#### J. Detecting Text in Natural Scenes with Stroke Width Transform

The algorithm used for text detection in this research paper is on the basis of the SWT (Stroke Width Transform). It is basically the local operator which computes the stroke which contains the pixel for each pixel within an image. Output of SWT is a picture of similar size as the original image which contains the width of stroke of each pixel. Initially value of every element is set as infinity. Then to recover strokes canny edge detector is used, sobel edge detector can also be used for edge detection however using the former produces smooth edges so it is preferred.

Now for each edge pixel a gradient direction is computed, for example for a pixel  $p_1$ , a gradient direction  $d_p$  is considered,  $d_p$  should be roughly at  $90^\circ$  to the boundary stroke (or edge stroke). We keep moving to the direction of  $d_p$  by each pixel until we reach another edge pixel ( $q_1$ ). Now the gradient direction of the edge pixel  $q_1$  is computed if the direction  $d_q$  is facing on the other side to direction of  $d_p$  then every element of the resultant image along the segment  $[p_1, q_1]$  is assigned  $|p_1 - q_1|$  width.

Now the upcoming step is to find and assemble the pixels into potential letter candidates. For that, two neighbouring pixels having same stroke width may be grouped together. Here two neighbouring pixels are grouped together if their Stroke width ratio is not more than 3.0, this guarantees that stroke with smooth edges will also be assembled. To aid both light coloured texts with darker background and darker texts with light coloured background the algorithm is once applied along  $d_p$  then along  $-d_p$ . Now for each connected component stroke width variance is calculated and whichever connected component's variance is too big that is rejected. These rejected areas which are hard to differentiate from text can be removed by doing this. Long and Narrow components can be generated by many natural processes can be mistaken for being text, so to remove these components the value of aspect ratio is limited only from 0.1 to 10. Other problems are components that can be surrounding text like signs on a signboard, this is mitigated by limiting only two other components inside the bounding box of a component. At last components which are too big or too small in size are also ignored. The text component was only accepted if the font height is between 10 to 300 pixels. This usage of font height helps in detecting connected fonts like Arabic. After all this process the left-out components are considered to be letter candidates. Next is the task to collect these letter candidates to find out text and words.

It is a fact that single letter is not usually present in images, by this logic we can remove the randomly scattered noise. Also, the text appears in a straight form, text in a singular line might have similarities like same font size and style and other distinguishing features. Also, similarities in text spacing between letters and lines. The ratio of median stroke width of two components should be less than 2.0 and ratio of height of two components should also be less than 2.0. The distance between letters should not be more than three times of the width of the widest component. Also letters in same words are considered to be written in same colour that is also taken into consideration. Now chains are formed by clustering the candidate pairs by above considered parameters. Chains only has one pair of letter region candidate and it is merged if they have similar direction and share an end. When chains can no longer be merged the process ends. Every chain of at least three letter candidates is taken as a text line and lastly the text line is separated into words using a measure of horizontal distance between successive letters and calculate the limit that divides inter word and intra word letter distances.

#### K. Segmentation of Connected Characters in text-based CAPTCHAs for Intelligent Character Recognition

A method is designed in this paper, to recognize CAPTCHA by machine. For achieving this pre-processing is performed which is followed by segmentation and then by Recognition.

The pre-processing phase can be divided into 4 sub-phases which are gray scale conversion, binary conversion, cropping & thinning. It should be ensured that these 4 sub-phases need to be performed in this sequence only which is original image  $\rightarrow$  grey scale image  $\rightarrow$  Binary Image  $\rightarrow$  Cropped image Thinned image. The Gray image to binary image conversion is done via Otsu threshold method [13]. While the thinning operation of the cropped image is done via Zhang's thinning algorithm [25], so that its skeleton can be obtained.

The data can be broadly into two categories which are

- 1) Open characters
- 2) Closed Characters

On the basis of number of pixels, we can classify any given character whether it falls under closed characters category or open characters category. However, the problem of over-segmentation may arise along with complex ligatures between closed characters (a ligature is a character formed by joining a couple of letters like  $\text{\ae}$ ). This problem can be overcome by two methods; first one is threshold horizontal distance calculation for upcoming segmentation column within that text in order to determine True Segmented Columns (TSCs) & another method is neural classifier implementation for correct and incorrect segments. This over-segmentation can be further reduced by training an artificial neural network (ANN) along with back propagation to reduce load of classifying incorrect & correct segments.

In above step of segmentation, we calculated the size of all segments. If they are found to be more than some threshold value, there is a possibility that there are double character segments. Else it is considered as right segment comprising of a character & identified via the help of a neural classifier.

The experiments involving this method is performed on CAPTCHA from various sources Taoba, MSN and ebay with a precision of 51.3, 27.1 and 53.2 on them respectively

### III. CONCLUSION

In [1], the accuracy of text recognition is 84.53%. The accuracy varies from 67.86 percent for blurred images to 93.47 percent for clear images. For [2], the text recognition accuracy for all 281 text regions correctly detected is found to be 93%. In all the text regions 97.2% of visible text is detected. In [3], there are a total of 49200 images. It has 5198 test characters divided in 62 classes which consists of 26 capital letters, 26 small letters & 10 digits. The accuracy of text recognition is calculated to be 81.7%. In [4], after evaluation the precision rate reach is 68%, the recall reach rate is 60% and the accuracy of text recognition is 53.2%. In [5], the technique involves signboard detection, the accuracy of signboard detection is 85%, the text detection accuracy is 80% and the accuracy of text recognition is 70.2%. In [6], for text detection and localization, approximately 13 images are pre-processed per second. Recall rate is 91.1% and the accuracy of text detection is calculated as 95.8%. In [7], Approximately 70 images are pre-processed per millisecond for text detection. The recall rate is 96.3% and the accuracy of text detection is 95.86%. In [8], the dataset has 200 jpeg images in which 50 does not contain text and other have 480 lines of readable text. The algorithm has a text detection accuracy of 93% and it can binarize them with a readability rate of 82%. In [9], only usage of outline features was made, and after eliminating those features whose checkability was lesser than 5% there were 383 lines left. After evaluating the algorithm gave an accuracy of 80% for text recognition.

In [10], the ICDAR dataset, 258 images are present in training set & 251 images are present in test set, which comprise to a total of 509 images. These images are full-colour images and with their size varying from 307-by-93 to 1280-by-960 pixels. The precision from proposed method is 0.73 and the recall is 0.60 with a f-measure equal to 0.66. In this method to check the importance of stroke breadth transform and geometric filtering, the algorithm is run in two configurations, in first configuration all stroke breadth values who are less than infinity, are set to 5, with this the precision became 0.66 and the recall dropped to 0.55. In second configuration the geometric filtering was removed so the precision and recall both dropped to 0.65 and 0.50 respectively. In [11], the proposed algorithm was tested on three different datasets namely MSN, eBay and Taobao the overall precision was 0.27, 0.53 and 0.51 respectively. The success recognition rate for Taobao was 0.96, 0.91 for MSN and 0.97 for eBay. Similarly, segmentation success rate was 0.59 for Taobao, .51 for MSN and .62 for eBay. The number of images in Taobao was 1000, 500 in MSN and 1000 in eBay data set.

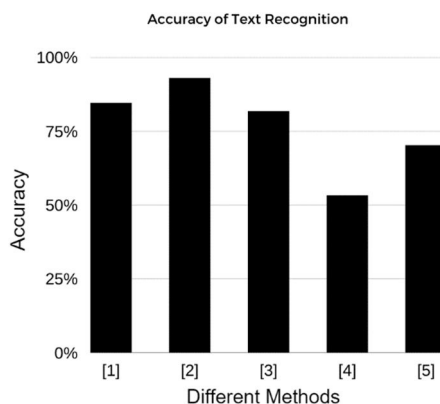


Fig. 1. Accuracy of text recognition as calculated in different methods.

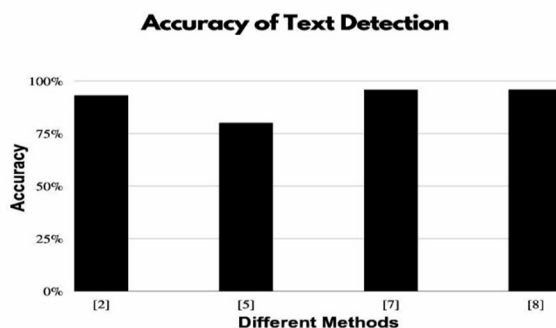


Fig. 2. Accuracy of text detection as calculated in different methods.

Different methods	Precision Rate	Recall Rate
[4] (MSER)	0.64	0.65
[6]	0.91	-
[10] second config.	0.65	0.50
[10] proposed method	0.73	0.60

Table 1: Containing data on recall and precision rate in different methods

### REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (*references*)
- [2] Zohra Saidane and Christophe Garcia, Orange, "Automatic Scene Text Recognition using CNN", Labs 4, rue du Clos Courtel BP 91226 35512 Cesson Sevign ´e Cedex – France
- [3] Xiangrong Chen and A. L. Yuille, "Detecting and reading text in natural scenes," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, pp. II-II, doi: 10.1109/CVPR.2004.1315187.
- [4] Xiaoming Huang, Tao Shen, Run Wang, Chenqiang Gao Chongqing, "Text Detection and Recognition in Natural Scene Images", Key Laboratory of Signal and Information Processing, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
- [5] "Signboard Detection and Text Recognition Using Artificial Neural Networks" Muhammad A.Panhwar School of Electronic Engineering, Beijing University of Posts & Telecommunications, Beijing, China, Sijjad A. Khuhro School of Computer Science and Technology University of Science and Technology of China Anhui, Hefei, 230026, China. Kamran A. Memon, Adeel Abro, Deng Zhongliang School of Electronic Engineering, Beijing University of Posts & Telecommunications, Beijing, China, Saleemullah Memon School of information & communication engineering, Beijing University of Posts & Telecommunications, Beijing, China
- [6] Xiaojun Li, Weiqiang Wang, Shuqiang Jiang, Qingming Huang, Wen Gao, "FAST AND EFFECTIVE TEXT DETECTION", Graduate University of Chinese Academy of Sciences, Beijing, China Key Lab of Intell. Info. Process., Inst. of Comput. Tech., Chinese Academy of Sciences, Beijing China Institute of Digital Media, Peking University, Beijing, China
- [7] Li Sun, Guizhong Liu, Xueming Qian, Danping Guo, "A NOVEL TEXT DETECTION AND LOCALIZATION METHOD BASED ON CORNER RESPONSE", School of Electronics and Information Engineering, Xi'an Jiaotong University, 710049, China
- [8] C. Garcia and X. Apostolidis, "TEXT DETECTION AND SEGMENTATION IN COMPLEX COLOR IMAGES", Institute of Computer Science Foundation for R.esearch and Technology-Hellas P.O.Box 1385, GR 711 10 Heraklion, Crete, Greece E-mail: {cgarcia,hapostol}@csi.forth.gr
- [9] C. H. Chen and J. L. DeCurtins, "Word Recognition in a Segmentation-free Approach to OCR", Information, Telecommunications, and Automation Division SRI Intemational, 333 Ravenswood Avenue, Menlo Park, California 94025, E-mail: chen@erg.sri.com [decurtin@erg.sri.com](mailto:decurtin@erg.sri.com)
- [10] Boris Epshtein, Eyal Ofek, Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform", Microsoft Corporation
- [11] Rafaqat Hussain & Hui Gao & Riaz Ahmed Shaikh, "Segmentation of connected characters in text-based CAPTCHAs for intelligent character recognition."
- [12] S. Kopf, T. Haenselmann, and W. Effelsberg. Robust character recognition in low-resolution images and videos. Technical report, Department for Mathematics and Computer Science, University of Mannheim, April 2005
- [13] W. Niblack. An Introduction to Digital Image Processing. pp. 115-116, Prentice Hall, 1986.
- [14] T. Pavlidis. Structural pattern Recognition. Springer-Verlag, Berlin-Heidelberg-New York. 1977
- [15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard & L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541-551, 1989
- [16] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition, " in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1457-1464.
- [17] Shi, B., Bai, X., & Yao, C. (n.d.). An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition.
- [18] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, 13, 18–28.
- [19] Sun, D., & Watada, J. (2015). Detecting pedestrians and vehicles in traffic scene based on boosted HOG features and SVM. In *2015 IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings* (pp. 1–4). <https://doi.org/10.1109/WISP.2015.7139161>
- [20] R. Dericbr. Using Canny's criteria to detect ail optiirial edge detctor recursively irnplement,ccl. *Int. bournat of Coinputer Vision*, 2:167- 187,1982
- [21] M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. In *CVPR*, 2007.
- [22] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR*, 2008
- [23] Y. Liu, L. Jin and C. Fang, "Arbitrarily Shaped Scene Text Detection With a Mask Tightness Text Detector," in *IEEE Transactions on Image Processing*, vol. 29, pp. 2918-2930, 2020, doi: 10.1109/TIP.2019.2954218.
- [24] Y. Cao, S. Ma and H. Pan, "FDTA: Fully Convolutional Scene Text Detection With Text Attention," in *IEEE Access*, vol. 8, pp. 155441-155449, 2020, doi: 10.1109/ACCESS.2020.3018784.
- [25] L. Cao, H. Li, R. Xie and J. Zhu, "A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3," in *IEEE Access*, vol. 8, pp.176924-176934,2020,doi: 10.1109/ACCESS.2020.3025221.
- [26] X. Rong, C. Yi and Y. Tian, "Unambiguous Text Localization, Retrieval, and Recognition for Cluttered Scenes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 3, pp. 1638-1652, 1 March 2022, doi: 10.1109/TPAMI.2020.3018491.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)