



Text Detection from Natural Images using MSER Algorithm

Lokendra Singh Banafar¹, Dr. Lalita Gupta²

¹Lecturer, Electronics and Communication, Government Polytechnic College, Itarsi, M.P., India

²Professor, Electronics and Communication, MANIT, Bhopal, M.P., India

ABSTRACT

In present daily life text plays an important role in daily life because of its rich information that is why automatic text detection in natural images has many applications. But detecting the text from natural image is always a challenging problem. Due to the presence of variation in the background and as the size of the text also not fixed in case of natural images it is very difficult to identify the text accurately. Through tremendous efforts have recently been devoted in this research but still reading texts in unconstrained environment is still challenging and remain a problem. Today text detection finds many applications in various fields, including visual impairment assistance, tourist assistance, content based image retrieval and unmanned ground vehicle navigation. Today most of the images are taken from the camera and other handhold devices which are not fixed and sometimes due to the movement of the object the problem of blurring is observed which makes it even more difficult to detect the text from natural images. Here in this thesis an idea is proposed to detect and recognize the text contains in the image as the main problem in Computer vision is to separate the text from the background components. There are many methods which are still used to detect the text from the natural scene such as text detection using edge analysis, robust text detection, Real time text tracking, but none of them is promising. In this paper text detection is carried by using canny edge detection algorithm and MSER based method along with the image enhancement which results in the improved performance in terms of text detection. In addition, we improve current MSERs by developing a contrast Enhancement mechanism that enhances region stability of text patterns. to remove the blurring caused

during the capture of image Lucy Richardson de blurring Algorithm is used.

Keywords: Text Detection, Scene Test, MSER, OCR

I. INTRODUCTION

From the last decade the interest for the framework which can identify content from normal images has extraordinarily increased. Detection of content from characteristic picture is significantly more troublesome than distinguishing content from fax reports as common pictures does not contain altered foundation and shading combination. There are numerous promising work done in this field in such manner yet at the same time there is considerably more work to be finished. It is because the text data is easily recognized by machines and can be used in a variety of applications. Due to multifarious background, variations of size, font, orientation and color. Text detection in natural scene images has to be robustly detected before being recognized and retrieved. As an essential prerequisite for text-based image search, text within images has to be robustly located. Recognition of text from natural image is a part of computer vision system in which a methodology is adopted to make the computer system to detect the text like human eye. It includes several applications like automatic street sign translation, license plate recognition, text recognition for visually impaired people.



Figure 1: Natural image containing text.

For visual comprehension it gets to be imperative to automatically detect and perceive the content from normal images. Human insight results in the advancement of dialect, the psychological capacity to conceptualize, and communicate thoughts through language. Test frameworks are utilized to protect and communicate data and thoughts which exist subsequent to 6600BC. Today, it is required for the present day world to translated using printed signs and names which results in finding the content all through numerous pictures and recordings. So to permit the rich machines to speak with the individual a content identifying framework is required.



Figure 2 Example where text spotting is applied to detect text from natural image

So text spotting detection is the tools which allow us to design system which decode the text within the image and allow a greater understanding of media.

II. LITERATURE REVIEW

There are several methods proposed in past for the detection of text from images .These methods are capable in detecting and translating the text automatically. These techniques generally uses the pixels properties for the detection of characters, The distribution of edges, for example, is used in many text detection methods [7, 8, 9]. In these methods the edges are grouped together based on features such as size, color and aspect ratio. Texture is another commonly used feature for text segmentation [10, 11].

Many researchers working on text detection and thresholding algorithm with various approaches have achieved good performance based on some constraints.

2.1 Edge and gradient-based Text Detection

The work done by Jain and Bhattacharjee presented a simpler method for document image segmentation in which text regions are automatically identified in this paper the text region is considered as a textured region and nontext region such as blank spaces, graphic and pictures are considered as region with different textures. In this thesis a bank of Gabor filters ,fllowed by clustering to classify pixels. After this Wu et al [12] use a non-linear function of multiscale Gaussian derivatives to identify region f high energy. Similarly Thillou et al. Garcia and Apostolidisuses edge detectors and morphological operations to remove noise from the image.

2.2 Color-Based Text Detection

Gaoet al. Use color based text detection, assuming that character regions can be segmented with a Gaussian mixture model into foreground and background components. Alternately, Zhang and Chang segment an input image using the mean-shift algorithm and then label each region as text or non-text with a Markov field. Attempts to binarize images without regard for character identity are prone to poor performance when images have low resolution and/or low contrast. Instead, our approach throughout this thesis will be to detect and recognize characters based directly on the image, rather than an intermediate, uninformed binarization.

2.3 Methods based on component analysis

It is one of the first method for text detection where a local adaptive thresholding in grey scale image is used. In this method simple alignment rules are used for grouping individual characters. This method assumes that background is uniform due to this it can

be used for detecting license plate and uniform background document but not natural image.

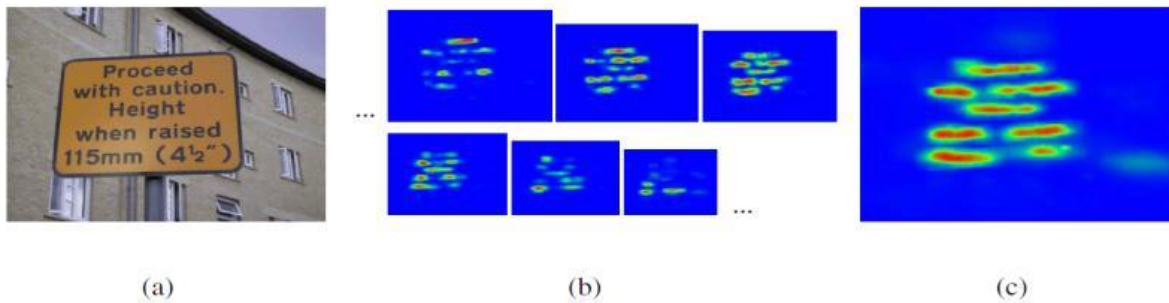


Figure 3: (a) Natural image (b) Text confined image for natural image. (c) Text confined image for natural image. [10]

This method does not have great application in text detection from natural image as it cannot handle illumination variations in background.

2.4 Stroke width transform

An image operated stroke width transform (SWT) was introduced by Epshtein et al. It estimates stroke width for each pixel in the image and finds edges using a Canny detector. But this method also fails for low light and blurred images.

The stroke width transform is an image operator which computes per pixel the most likely stroke width containing the pixel. The output of the SWT is an image of size equal to the input image size, with width estimation for each pixel.

2.5 Sliding window approach based Text

An early histogram based global thresholding Otsu's method is widely used in many applications. Text detection and binarization method is proposed for Korean sign board images using k-means clustering. But finding a best value for 'k' to achieve a good binary image is difficult in images with complex background and uneven lighting.

The linear Niblack method was proposed to extract connected components and texts were localized using a classifier algorithm. Four different methods were suggested to extract text, depending on character size.

In the work of Wu et al. a method was proposed to clean up and extract text using a histogram based binarization algorithm. The local threshold was picked at the first valley on the smoothed intensity histogram and used to achieve good binarization results. A thresholding method was developed using intensity and saturation features to separate text from background in color document images.

2.6 Previous work

2.6.1 Text Localization in Scene Images by Using Character Energy and MSER

The method for Text Localization in Scene Images by Using Character Energy and MSER is done by JUN baba and AKHIHIRO Yamamoto of Kyoto University. It used text region segmentation which allows people to accurately extract true text information from cluttered background. The second step used by JUN baba and AKHIHIRO is contrast-enhanced MSERs.

2.6.2 Robust Text Detection in Natural Scene Images

Xu-Cheng Yin, Member, IEEE, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao presented a new method for text detection known as robust text detection in natural scene images. It used MSER as character candidates where distance weight and clustering threshold are learned automatically by a novel self-training distance metric learning algorithm.

Character candidates extraction. Character candidates are extracted using the MSERs algorithm; most of the repeating components are removed by the proposed MSERs pruning algorithm by minimizing regularized variations.

Text candidates construction. Distance weights and clustering threshold are learned simultaneously using the proposed metric learning algorithm; character candidates are clustered into text candidates by the single-link clustering algorithm using the learned parameters. Text candidates elimination. The posterior probabilities of text candidates corresponding to non-texts are estimated using the

character classifier and text candidates with high non-text probabilities are removed.

Text candidates classification. Text candidates corresponding to true texts are identified by the text classifier. An AdaBoost classifier is trained to decide whether a text candidate corresponds to the true text or not. These are the four steps suggested by the Xu-Cheng Yin, Member, IEEE, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao for text detection.

2.6.3 Text Detection using Real Time Text Tracking

Carlos Merino-Gracia, Majid Mirmehdi in this paper Adaptive thresholding is applied to binarise the input image and retrieve a set of connected component (CC) regions. A tree is then constructed to represent the topological relationship between these CCs. A key step of this algorithm is the hierarchical filtering of the tree nodes, based on the assumption that in natural scene images with text, structural elements (such as sign borders, posters frames etc.) The advantage of this algorithm is that the performance is very fast due to the use of text tracking.

2.6.4 Text detection and recognition in natural scene with edge analysis

Chong Yu, Yonghong Song, Quan Meng, Yuanlin Zhang, Yang Liu propose a candidate edge recombination method in which the whole edge is separated into edge segments and then depending upon the colour and attribute these edges are again merged.

Edge segmentation: Here Canny edge operator is used for the purpose of edge detection. The problem in this paper is that in case of complex background the text edges could be connected with surrounding background.

III. ALGORITHM USED IN TEXT DETECTION

Edge detection is an operation in computer vision system which identifies the sharp change in the image pixel. By identifying the edges present in the image

we can greatly reduce the amount of data to be processed. There are several different edge detection algorithms but here we are focusing mainly on the algorithm developed by John F. Canny in 1986. In spite of one of the oldest methods of edge detection it is one of the standard edge detection methods and still used by the researchers.

The Canny edge detection is the most widely used edge detection algorithm to locate sharp intensity changes which is used to detect object boundaries in any image. In the Canny edge detection method the algorithm classifies the pixel as an edge if the gradient magnitude of the pixel is larger than those of the pixels at both its sides in the direction of maximum intensity change.

The Canny Edge Detection Algorithm

The algorithm runs in 5 separate steps:

1. **Smoothing:** Blurring of the image to remove noise.
2. **Finding gradients:** The edges should be marked where the gradients of the image have large magnitudes.
3. **Non-maximum suppression:** Only local maxima should be marked as edges.
4. **Double thresholding:** Potential edges are determined by thresholding.
5. **Edge tracking by hysteresis:** Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

3.1 Maximally Stable External Regions

MSER regions are connected areas characterized by almost uniform intensity, surrounded by contrasting background. They are constructed through a process of trying multiple thresholds. The selected regions are those that maintain unchanged shapes over a large set of thresholds. For color images MSER algorithm replaced thresholding of the intensity function with agglomerative clustering, which is based on the colour gradients.



Figure 4 Examples of MSER region.

3.2 MSER Algorithm

MSER is a method for blob detection in images. The MSER algorithm extracts from an image a number of co variant regions, called MSERs: an MSER is a stable connected component of some gray-level sets of the image. MSER is based on the idea of taking regions which stay nearly the same through a wide range of thresholds. – All the pixels below a given threshold are white and all those above or equal are black. – If we are shown a sequence of threshold images with frame t corresponding to threshold t , we would see first a black image, then white spots corresponding to local intensity minima will appear then grow larger. These white spots will eventually merge, until the whole image is white. – The set of all connected components in the sequence is the set of all extreme regions. Optionally, elliptical frames are attached to the MSERs by fitting ellipses to the regions. Those regions descriptors are kept as features. The word external refers to the property that all

pixels inside the MSER have either higher (bright external regions) or lower (dark external regions) intensity than all the pixels on its outer boundary. This operation can be performed by first sorting all pixels by gray value and then incrementally adding pixels to each connected component as the threshold is changed. The area is monitored. Regions such that their variation wrt the threshold is minimal are defined maximally stable: – Let's make all the pixels below a threshold white. The others black. Considering a sequence of threshold images with increasing thresholds sweeping from black to white we pass from a black image to images where white blobs appear and grow larger by merging, up to the final image. – Over a large range of thresholds the local binarization is stable and shows some invariance to affine transformation of image intensities and scaling.

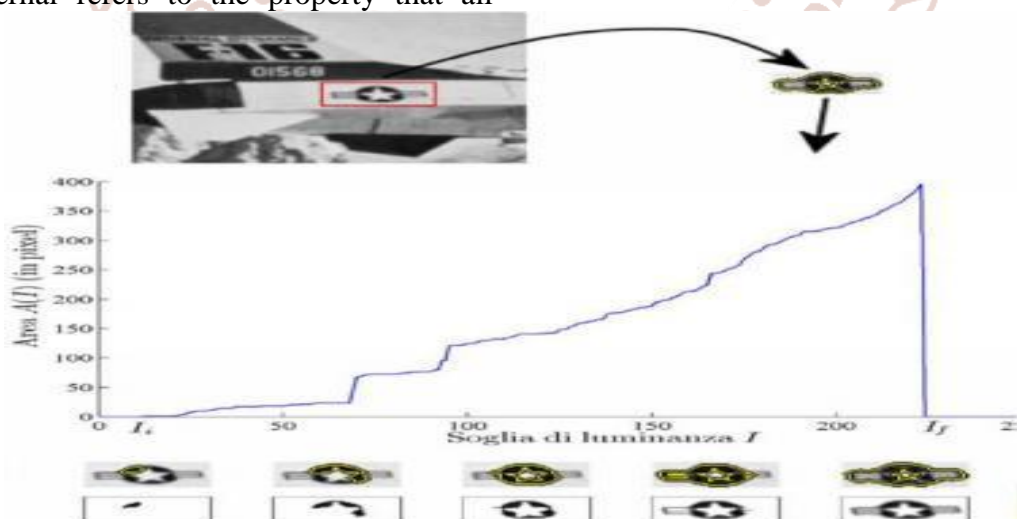


Figure 5 Region detected for different threshold area detected is different in MSER

3.3 Lucy Richardson de blurring Algorithm

Due to the large use of handheld camera and smart phone, image blurring becomes an important problem. There are two types of image motion blurring

observed i.e. due to the motion of camera or due to the motion of object. The blurring, or degradation, of an image can be caused by many factors: Movement during the image capture process, by the camera or, when long exposure times are used, by the subject,

Out-of-focus optics, use of a wide-angle lens, atmospheric turbulence, or a short exposure time, which reduces the number of photons captured, Scattered light distortion in confocal microscopy.

In general a Blurred image can be modeled as

$$I = L \otimes k + n$$

In which I represent the blurred image; L is latent unblurred image; K is point spread function; and N is noise and the operator represent the convolution. Richardson-lucyalgorithm[26] is well known iterative deconvolution method for image deblurring.

A more clearly deblurred image is generated for each time of iteration. After several iterations, a sharp and unblurred result image can be obtained. According to Bayes' theorem, (1) can be transform into

$$L = \arg \max P(L/I) \propto P(I/L) P(L)$$

Then the formulation of Richardson-lucy algorithm can be defined as

$$L^{t+1} = (k * \frac{I}{L \otimes k}) L^t$$

The above formula is used for the deblurring of the image in Richardson-Lucy Algorithm.



Figure 6: Example of ringing Artifact (a)original image (b)Blurred image.(c) Blurred image after using Richardson-Lucy Algorithm

IV. PROPOSED WORK

A novel Connecting Character based text recognition and extraction algorithm is designed which uses Maximally Stable Extremely Regions (MSER) for test candidate recognition and extraction from traffic signs. Despite their auspicious properties, MSER has been conveyed to be delicate towards blurred Image. To allow for detecting small letters in images of limited resolution or blurred Image, the

complimentary properties of Lucy-Richardson Algorithm and canny edge Algorithm is used. Further geometric filtering and pairing is applied to efficiently obtain more reliable results. Finally, texts are clustered into lines and additional checks are performed to eliminate false positives.

The proposed algorithm, illustrated in Figure 1, is divided into two basic steps i.e. text recognition and text elimination.

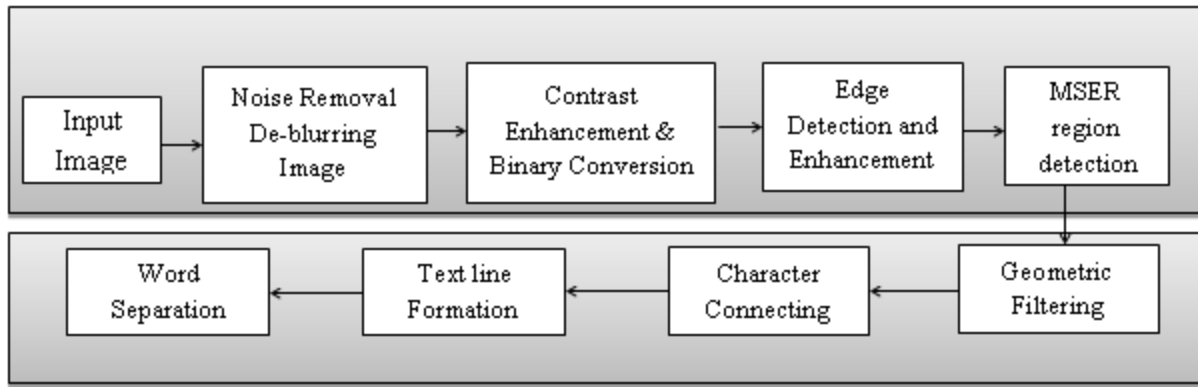


Figure 7: The flow chart of proposed Algorithm

V. RESULT AND ANALYSIS

In the result part the performance of the algorithm used is assessed experimentally. Here we also check the accuracy of the method for different test images. Here we use 4 different pattern images and the performance of system when applied to that images.

Test Image:

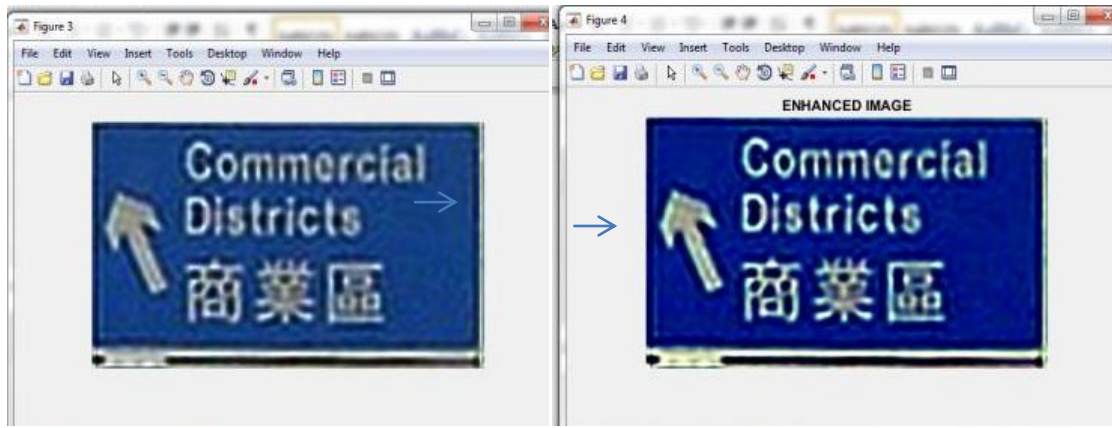
ORIGINAL IMAGE



CROPPED IMAGE

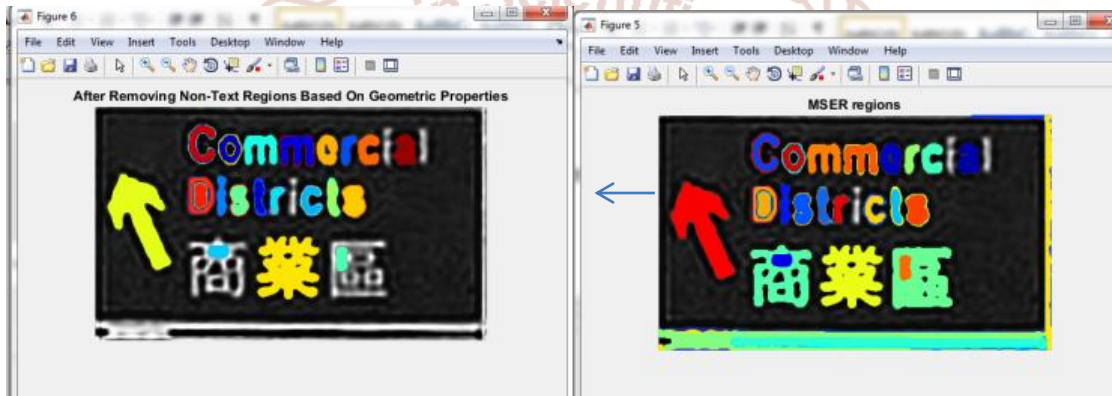


DEBLURRING THE IMAGE CONTRAST ENHANCEMENT IMAGE

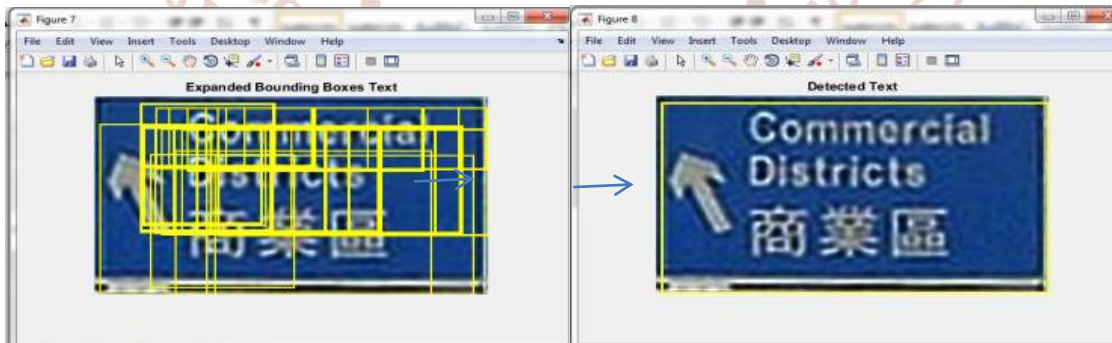


NON TEXT REGION REMOVAL

MSER REGION DETECTION

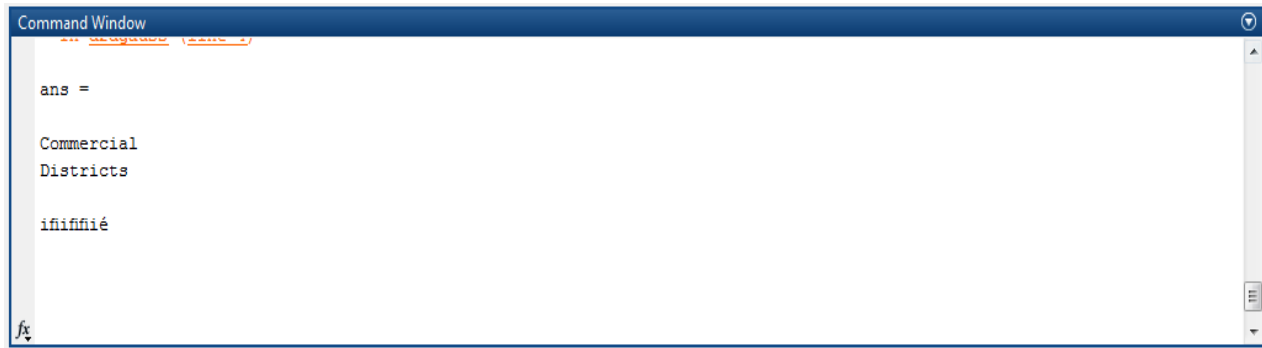


BOUNDING BOX DETECTED TEXT



OUTPUT





VI. CONCLUSION

In this thesis we present MSER based text recognition method. Here the image which is to be processed is first applied to as a input to the Lucy Richardson de blurring Algorithm which remove the blurring if present in the image after this image enhancement is done so that the intensity levels of the edges is increased the output from the image enhancement system is then feed to the canny edge detection system which detect the edges and then MSER region is used to separate the text from the background. After this various filters and text line method is used to make it easy for the user or system to identify the text part easily. The method using MSER Algorithm gives 100% accuracy when it comes to even background but when the intensity difference between text and the background reduces the accuracy of the system also reduces still when you compare it with the previous algorithm it gives you far better results.

REFERENCES

1. B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In "CVPR 2010", pages 2963-2970, (6, 2010).
2. Otsu, N., "A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man, and Cybernetics*", Vol. 9, No. 1, pp. 62-66 (1979).
3. Text-Attentional Convolutional Neural Network for Scene Text Detection Tong He, Weilin Huang, Member, "IEEE", Yu Qiao, Senior Member, IEEE, and Jian Yao, Senior Member, IEEE, (2007)
4. Robust Text Detection in Natural Scene Images Xu-Cheng Yin, Member, IEEE, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, "IEEE Transactions on Systems", june, (2008)
5. Text detection and recognition in natural scene with edge analysis Chong Yu, Yonghong Song, QuanMeng, Yuanlin Zhang, Yang Liu, "IEEE Transactions on Systems", june, (2012)