# Recognition of Multilingual Text from Signage Boards

Poonam Salunkhe, Sreebha Bhaskaran, Amudha J
Department of Computer Science and Engineering
Amrita University
Bengaluru
India
poonamsalunkhe10@gmail.com, b_sreebha@blr.amrita.edu, j_amudha@blr.amrita.edu.

Deepa Gupta
Department of Mathematics
Amrita University
Bengaluru
India
g_deepa@blr.amrita.edu.

***Abstract*—This paper presents a text detection approach based on Stroke Width Transform (SWT) and methodology to extract letter candidates. Major application focus here is tourism industry and local transport, to help people to deal with different Indian languages which involve text associated with natural scenes in the local public places. Multilingual text system supports curvy letters with different scale, size, direction, font and language. This method can significantly perform for different Indian Languages and results are tested based on ICDAR 2015 dataset and different user Test dataset.**

***Keywords—Text detection, SWT, Connected components, Union find algorithm.***

## I. INTRODUCTION

Text is highly researched area in the computer vision application to model a smart self-learning system, which involve text associated with shop banners, highway and roadside sign boards or the text on local transport system, such a text provide significant clues about that environment. Text detection helps visually impaired and tourist to convey correct information in more understandable way to user. There are different tools and methods to detect and recognize text, most of these text detection methods are language specific but due to diversity in the Indian languages a specific approach is not suitable, most of the state transportation uses their local language on boards which is difficult for foreigners to commute so there is need to design even more stable multilingual text detection system. To overcome such an issue most generic multilingual text detection and recognition based on SWT method is implemented in this project, technique is designed for various Indian languages and their combination with English without any prior knowledge of words to be detected.

The method presented in this paper overcomes the problem of a distance of image captured; text at a very far distance of several feet is also detected and recognized. Natural scene text exists in several fonts and different writing styles based on language also most of the Indian languages consists of circular characters hence designed system aims specific to detect circular characters as well as English. Background of the text is also an important factor; background surface can be glass, metal, wood, cloth etc. designed method works ease for different surfaces, colors and textures. This technique also supports multi-scale fonts


Fig. 1. Natural scene text.

This paper is structured in four sections as follows: In section II, an overview of related work, different algorithms and techniques in text detection area are described. Referring to such a different methods text detection system is designed in section III. Which is based on SWT, it also uses algorithms like canny edge detection, connected component analysis, and several threshold techniques to obtained correct character from text present in the environment. Section IV summarizes overall experiments and results. Result shows the better performance in terms of precision and recall for different datasets and various Indian languages when compared with other techniques.

## II. RELATED WORK

In this section we focus on related methodology and research work in the area of text detection to design significant system, the MSER(Maximally Stable Extremal Regions) based method [1] uses Junction feature extract technique and SLAM(Simultaneous Localization And Mapping) which can reduce the number of region proposals of words from 779,500 to 54 in only around 100 ms, and scene priors reduce the number of ER (Extreme Region) method [2] from 45 to 6 only with the computational cost reduction in terms of surface normal estimation and filtering, SLAM for text can help provide robots with a higher level semantic understanding of the world necessary for more complex tasks. An important application for such a system would be as an aid to the visually impaired.

Canny edge detector approach [1], [2], [3], [4], [5] is a multistage algorithm which is used to detect a wide range of edges for smoothing image and eliminating the noise, edge strength. SWT used in [2], [3] is a combination of each and every pixel dense computing and estimation for far distant pixel. This information is used for comparison with many standard test cases involving fast computation, Union find method [2], [3], [4] is implemented for connected component analysis to extract letter candidate from detected region based upon probability estimation and certain rules. This algorithm outperforms another algorithm by 15 times speed up allows to process several types of background with various fonts texts and sets new benchmark database. The most awarded method by ICDAR-2013 context described in [6], [7] uses ERs with probability technique for many projection channels. Support Vector Machine (SVM) method used in [4] uses ERs per segment layer of image with homogeneous labeling. Even Letters with disconnected stroke line is also detected. The method used in [8] Gamma Correction Method (GCM) is a robust method for the segmentation and binarization of text features in complex color images, this method presents a behavioral study of the curves generated by the analyzed textural features and an effective sample method that retains the behavior of the curves. Accelerated Gamma Correction Method (AGCM) is more suitable for its integration in real-time applications which gives algorithm speed up which is best for real time application like text detection from Indoor Panorama Images.

[1], [2], [3], [6], [8] have tested result and methodology with ICDAR database-2013, ICDAR database-2015 [10] and Google street view database [11]. ICDAR dataset contains images of various types like attonated scene characters, blurred images and local transport banner text. Test dataset is rotated by -45 to 45 degrees and skew, with f-measure 76 percent.

## III. THE PROPOSED METHOD

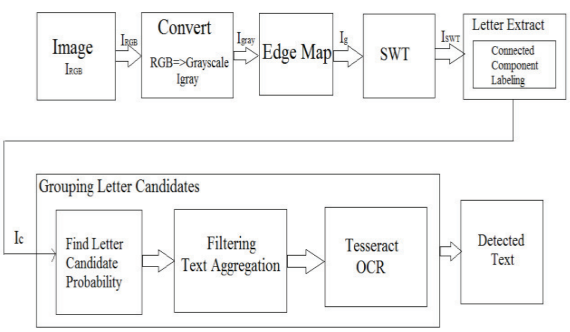



Fig. 2. System architecture, $I$:Image.

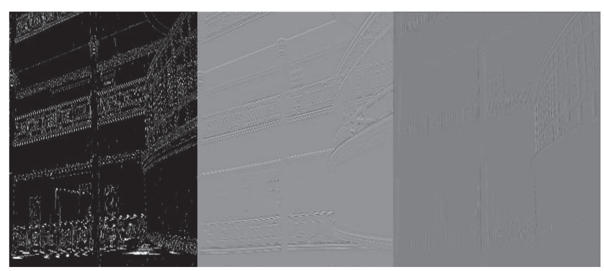
The flow of algorithms used in each block of system architecture is as shown in Fig. 2. Designed system takes input image of RGB *Irgb* type, which supports various image formats like JPEG, PNG. This image is further converted to a gray-scale image as shown in Fig. 3 and represented by *Igray*, this system can process input image of any size or scale, which is taken from different camera poses and datasets. In text detection algorithm we need to specify if a text is present over the darker background or lighter one depending upon +1 or -1 directions are set. SWT is then applied to obtained text boundaries. The output contains a matrix of each pixel representing Id associated with Connected Component (CC) analysis, obtained text region matrix is further processed by probability statistics and designed parameters are applied to extract text candidate and remove non-text region. OCR is the last step of the algorithm, Tesseract-OCR [12] is used to recognize detected text. The algorithm used in each block of system architecture is described as follows.

### *A. Edge Map*

Canny Edge detection algorithm is used for edge detection. Canny edge detector is a multistage algorithm that operates with different multi-scale edges, text in the image can be of any scale or font size and without any standard dimensions. Edge map algorithm takes gray-scale image refer to Fig. 3.b $I_{gray}$ as input and produces output edge map given by $I_g$ of similar dimension.

(a) (b)

Fig.3. (a) Original image. (b) Gray-scale image.



(a) (b) (c)

Fig. 4. (a) Edge map. (b)Horizontal gradient $G_x$. (c) Vertical gradient $G_y$.

First step of edge detection applied to an input image Igray is Gaussian blurring which is required to reduce noise and redundancies present in the captured image. Output edge map is as shown in Fig. 4.b for horizontal gradient $G_x$ and vertical gradient $G_y$ as shown in Fig. 4.c, Gaussian blur provides high magnitude weight to center pixels as shown below in equation (1). Gradient operator is used to find edge strength. The algorithm further uses Sobel operator to iterate over 2D image to find spatial gradient measurements with an image as follows.

Gradient magnitude:

$$\mathrm{M}(\mathrm{x},\mathrm{y}) = \sqrt{\mathrm{G}_\mathrm{x}^2 + \mathrm{G}_\mathrm{y}^2} \tag{1}$$

Direction angle:

$$A(x,y) = \tan^{-1}\frac{G_x}{G_y} \tag{2}$$

Nonmaxima Suppression technique reduces the thickness of edge, that is a strength of response around true edges. After the Sobel operation suppression technique selects the single maximum point across the width of an edge. This technique is used to suppress a number of pixels present in a non-text region adjacent to text candidate stroke boundary, which is set to value ’0’ to output perfect thin line. Thin line enhances the probability of extracting candidates with less error.

$$G_{nh}(x,y) = G_n(x,y) < T_h \tag{3}$$

$$G_{nl}(x,y) = G_n(x,y) > T_l \tag{4}$$

Edges are the large intensity gradient values than that of smaller intensity gradients values. Thresholding technique for Nonmaxima Suppression to categories the edge with upper or lower bound value to set uniform strong and weak set output matrix is as follow. Equation (3) will contain only strong edge point that of equation (4) contains weak edge point, with $T_h$ and $T_l$ in 2:1 and 3:1 ratio respectively.

*B. Stroke Width Transform.*

Character candidate possesses the property of constant stroke band of fixed width, formed by connecting the likely pixels. After edge detection algorithm SWT operator form a group of pixels which are most likely pixels of constant stroke band. This concept is used to produce stroke width matrix $I_{SWT}$ from an input Edge map $I_g$.

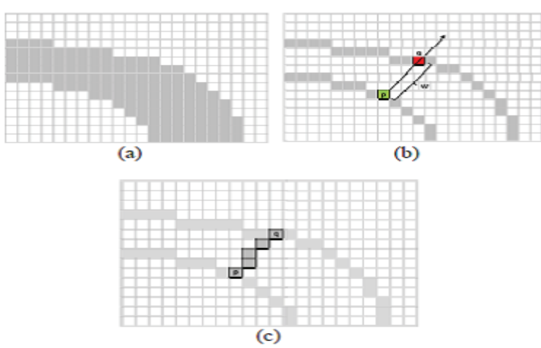



Fig. 5. Stroke Width Transform technique.

All pixel of $I_g$ at the initial instant are considered to have infinite stroke width, boundary detection is for black stroke on a white surface or vice versa. Referring to equation (5) and (6) from each boundary pixel ‘p’ send a ray ‘r’ along the text gradient direction, this leads to finding another boundary pixel ‘q’.

$$r = p + n * g_p \tag{5}$$

$$r = q + n * g_q \tag{6}$$

Equation (5) contains $g_p$ and equation (6) contains $g_q$ are the gradient direction for ‘$p$’ pixel and ‘$q$’ pixel respectively, also n>0 gives a number of pixels in gradient direction. SWT follow certain rules like while iterating through each pixel in an algorithm if pixel neighbor do not contain likely pixels then that ray is discarded as shown above in Fig. 5.a, it shows black stroke over white surface, after edge detection with Fig. 5.b SWT is iterated as shown in Fig. 5.c to form continuous likely bounded pixels hence stroke of letter candidate at the output matrix$I_{swt}$.

*C. Letters Extract*

To extract character candidates CC labeling method is implemented after obtaining stroke width transform matrix $I_{swt}$. Union find is a method to find CC labels of likely intensity pixels among stroke region area. Pixels which are connected to its root parent node is assign the least value of $I_d$ that is nothing but $I_d$ number further process by OCR as a letter candidate.

Rules and conditions for character extraction are as follows: Aspect ratio is set to fixed bounding range to filter out non text component which are either small or large. Large variance in the stroke width is not acceptable, as character is formed by constant width. Component diameter to the median stroke ratio must be lesser than threshold set. Ignore too large and small component. Pixel bounding range is applied for total number of character pixels to the number of pixels in the bounding box of the component, which help to reject less number of pixels count region. Matrix I contains final output for letter extraction.


Fig. 6. Text detection output.

Accepted font heights are between 10px and 300px and height > max letter height or height < min letter height is valid condition, Reject CC with highest stroke width variance. The threshold condition of variance for SWT to the mean SWT > 0.5 is a valid condition. Aspect ratio must be between 0.1 and 10 that is if aspect ratio < 0.1 and aspect Ratio > 10 is a valid Case considering Stroke width > height*2.5, these are the conditions for valid character.

*D. Grouping character candidates into text region*

After obtaining CC labels from $I_c$ matrix it is important to remove all other non-textual region and output should contain only textual region without noise, to do this certain aspect ratio, variance and standards are set in letter extract process. Rules to form a word out of letters and text line are described here.

Two or more character candidates which are forming word should have similar character height and width, that is height and width ratio between characters must be less than 2.5, because word contains combination of lower and caps characters. Distance between letters is bounded by three times the max width of a character in the word. Average color of letter candidates is a reference to bind letters of the word. We estimated final text based on probability threshold with an equation as follow. Equation (7) states probability threshold that is $Prob(x, y)$ matrix is updated with true value.

$$Prob(x, y) > PROB_{THRESHOLD} \quad (7)$$

$$Added(currIndices) = Prob(x, y) \quad (8)$$

With this estimated result we obtain update final matrix Iout. The final step also has provision to select particular text region in form of rectangle selection box to process only text present within a rectangle and reduce the computation and processing time of the algorithm. Final output result for text detection after algorithm processing is shown in Fig. 6, this output matrix is further processed by the tesseract OCR, and respected text is recognized.

## IV. EXPERIMENTS AND RESULTS

In this section, we presented the experimental results of text detection and recognize for many test cases which involve images from ICDAR 2015 database, images captured from the mobile camera of various size and background. Different test images involve natural scenes, shop banners, public transport services and many tourism related images. The main contribution of this paper is to process natural scene captured images and extract character candidate. It also has facility to select text region present in the scene and then process it for faster computations and better algorithm optimization. The following Table of images consists of some parts of our work for text detection results using different test cases.

We can estimate from several test cases that the proposed method overcomes the problem of a distance of image captured that is text at a very far distance is also detected and recognized, as shown in Fig. 7.1.a text at a distance of 45feet is also detected and output is given in Fig. 7.1.b, Text in Fig 7.1.a is white color text present over dark surface, we have tested result cases for this text from several distances and camera angle and method works for most of the cases also we can specify if background is darker or lighter. Fig 7.2.a is an example of dark illumination of surrounding, here image is captured in low light luminance and its respected detected output is displayed in Fig 7.2.b, this promises that method works in different environment light change. Example Fig. 7.3.a contains images from ICDAR dataset, which consists of text of various scales and fonts over different surface with several rotation it is an example of large amount of text information of approx 120 characters out of which 75 characters are successfully detected as shown in Fig. 7.3.b. Fig 7.4.a is an example of skewed text here image is captured from skewed lens angle of camera and output is successfully detected text as shown in Fig 7.4.b. Fig. 7.5.a contains multi-scale text with various fonts and this text is present over glass surface with no skew; independent to these surface conditions like glass surface text is detected with all characters as shown in Fig 7.5.b. Fig 7.6.a shows ICDAR 2015 dataset image, here text is only small part of captured image, which is detected and shown in Fig 7.6.b. it also shows blur text region which is identified. The methodology implemented in this paper still need to overcome cases like trademarks with a blurred region, severe tilted, large skew in the captured image.

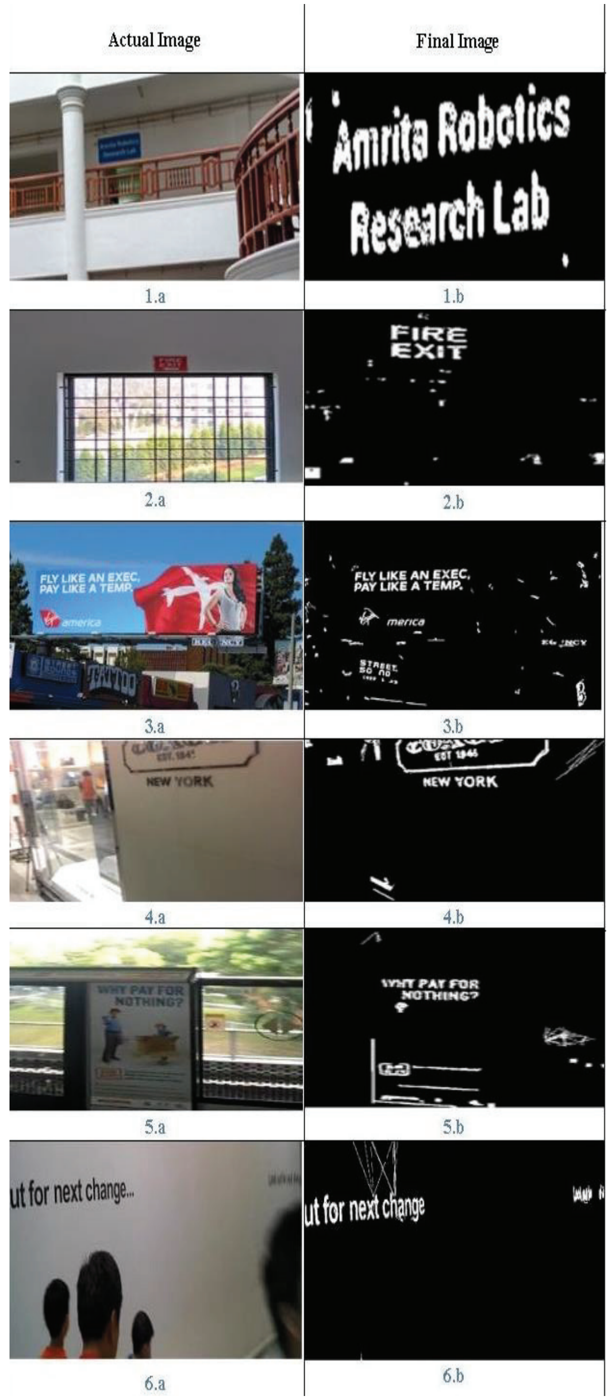



Fig. 7. Result of ICDAR 2015 and Test dataset.

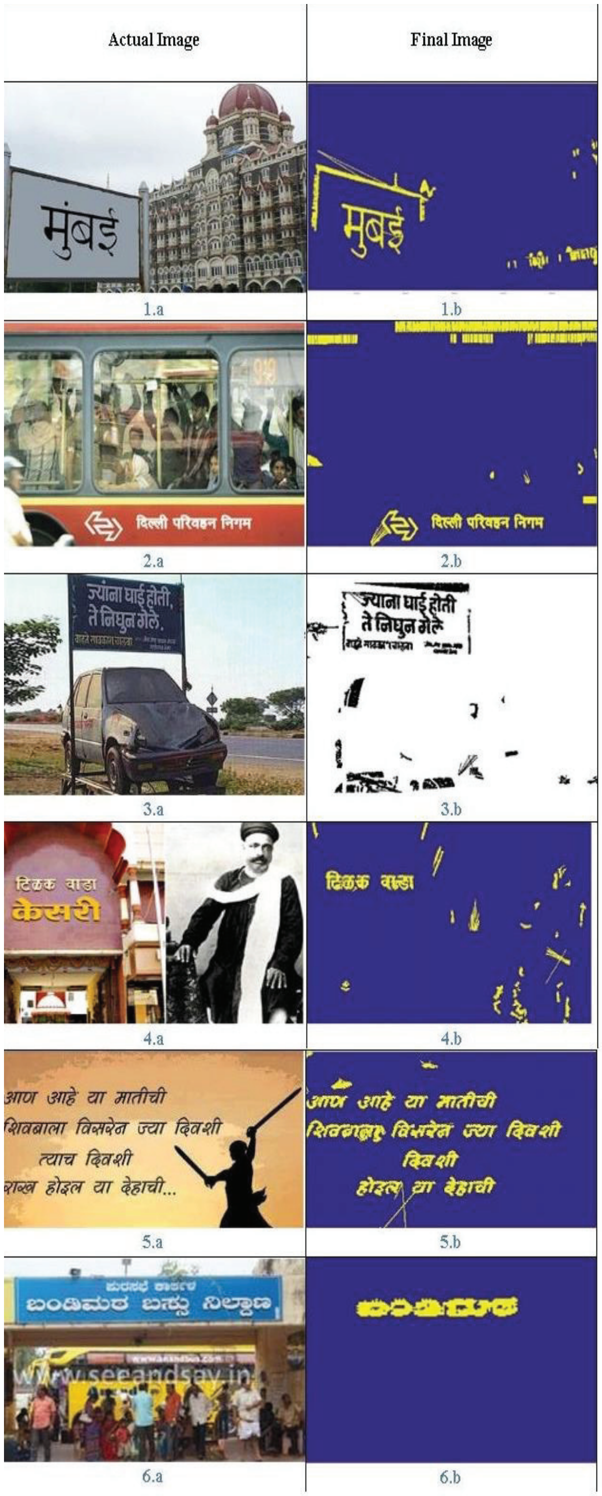



Fig. 8. Result of Indian Language for- Hindi, Marathi, Kannada Language.


Fig. 9. Result of English and Indian Mixed Language.

Experimental results for the Indian Languages like Hindi, Marathi, Kannada are shown in Fig. 8. Most of the Indian language characters are curvy, for example Fig. 8.2.a contains

Hindi language text on bus transport system which is successfully detected as shown in Fig. 8.2.b. Methodology implemented is most suitable for mixed languages for example Fig. 9 shows results for mixed language, Fig. 9.1.a is an example of Marathi and English language, Fig. 9.3.a contains text information from Mumbai local transport system, we have tested these test data result for several highway side sign boards, shop banners etc. which contains combination of English, Hindi and Marathi. Test image Fig. 9.5.a is an example of Marathi text on shop banner present in public crowd. Overall results for several datasets and its precision and recall are listed in Table -I.

TABLE I: Result of Proposed system

| **Test Cases** | **Metrics Table** | | |
|---|---|---|---|
| | ***Precision*** | ***Recall*** | ***F-measure*** |
| Indian-language | 0.71 | 0.65 | 0.7 |
| ICDR-2015 and Test dataset | 0.91 | 0.88 | 0.9 |
| Overall | 0.87 | 0.76 | 0.8 |

Experimental results are summarized in Table-I, measuring parameters are precision, recall, and f-measure. False Positive (FP) is a non-text region which is detected as text candidate and False Negative (FN) is a actual text but considered as a background in processing. Based on this false positive, false negative and actual numbers of character measuring calculations are done. The result is verified with natural Scene text with Indian languages like Hindi, Marathi, Kannada etc, and natural scene text with ICDAR 2015 dataset and mobile captured images.

## V. CONCLUSION

In this paper we have implemented the methodology to detect text in English and Indian languages. This technique supports different font and scale for text associated with a natural scene. We obtained efficient results which decrease the False character detection. Designed techniques are also tested on various datasets which yields output with very good precision and recall. Further improvement to this is text detection for skewed letters, GPU acceleration to reduce timing need. In the future improvement, labeling algorithm can be made even more efficient by adding several optimization techniques, better labeling method of components could improve the detection of characters and will allow us to use harsher thresholds. This way instead of manually specifying OCR to be used more intelligent OCR select method can be implemented; also we could get better results for circular text, which tends to be dismissed as noise due to the grouping of the letters. This would allow us to identify curvy letters better, such as Arabic fonts or cursive handwriting and any other language.

## References


[1] Hsueh-ChengWang1, Chelsea Finn2, Liam Paull1 and Michael Kaess3 Ruth, "Bridging Text Spotting and SLAM with Junction Features," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Congress Center Hamburg Sept 28 - Oct 2, 2015. Hamburg, Germany, pp. 3701-3708

[2] Abdelhamid Mammeri, Azzedine Boukerche and El Hebri Khiari, "MSER-Based Text Detection and Communication Algorithm for Autonomous Vehicles," DIVA Strategic Research Network University of Ottawa Ottawa, Ontario Canada, IEEE Symposium on Computers and Communication (ISCC) 2016.

[3] Do Kyoung Shin1, "Text Detection in vedio sequence using 1-D DCT," Department of Radiation Oncology 1Case Western Reserve University, Cleveland, USA IEEE ISCE 2014.

[4] Jianjun Chen, "Gray-Scale Morphology Based Image Segmentation and Character Extraction Using SVM," Depar`tment of Intelligent Systems Design Engineering Toyama Prefectural University, Noboru Takagi Department of Intelligent Systems Design Engineering Toyama Prefectural University, IEEE 46th International Symposium on Multiple-Valued Logic, 2016

[5] Boris Epshtein , Eyal Ofek and Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform"

[6] Lukas Neumann and Matas, "Real-Time Lexicon-Free Scene Text Localization and Recognition," IEEE TRANSACTIONS ON Pattern Analysis and machine Intelligence, VOL. 38, NO. 9, SEPTEMBER 2016, pp. 1872-1885

[7] Yuan Liu, Kao Zhang, Jian Yao and Tong He, "An Effective Method for Text Detection from Indoor Panorama Images Using Extremal Regions," 2015 IEEE International Conference on Information and Automation Lijiang China, pp. 781-786

[8] Rostom Kachouri, Christian Medina Armas and Mohamed, "Gamma Correction Acceleration for Real-time text extraction from complex colored images," ICIP 2015 IEEE, pp. 527-531

[9] Deepa Gupta, M. L. Leema. 2013. "Improving OCR by Effective Pre-Processing and Segmentation for Devanagari Script: A Quantified Study", Journal of Theoretical and Applied Information Technology. Publisher: ARPN. Vol 52 No.2, page 142-153, 2013.

[10] ICDAR2015Datasetlink: http://rrc.cvc.uab.es/?com=introduction.

[11] Google Street View Dataset Link: http://ufldl.stanford.edu/housenumbers/, http://crcv.ucf.edu/data/.

[12] Tesseract OCR download link: https://sourceforge.net/projects/tesseract-ocr/

[13] Amudha J and Naresh Kumar, "Gradual transition detection based on fuzzy logic using visual attention system advances in intelligent informatics," pp111-122,2014.