An Intelligent Sign board detection and recognition System on Computer vision and NLP: A Survey

Gunwanth[1], Anil Kumar Gupta[2], Saket Kumar Jha[3]
[1]*CSE with Specialization in AI and ML, Vellore Institute of Technology, Chennai, India*
[2]*Senior Member IEEE, Centre for Development of Advance Computing Pune, India*
[3]*Project Engineer, Centre for Development of Advance Computing Patna, India*

Abstract **- Parking signs acknowledgment and confinement intend to extricate and digitize exact on-road Parking limitations. The present visual information assortment, comment, and examination rehearses are still exorbitant, powerless to blunder, and awkward as performed physically. While online road level symbolism information bases contain refreshed all-encompassing pictures, all things considered, their potential for comprehension on-road Parking limitations at scale has not been completely investigated. The key favourable position of these information bases is that when the Parking signs are identified, precise geographic directions of the recognized signs are regularly naturally decided and imagined inside an equal stage. This paper assesses the machine of a computer vision strategy for Parking signs acknowledgment from road level symbolism pointed toward encouraging the Parking tricks of urban communities. Tricks such as vehicle to be parked at specific period of time at specific dates and theses specific conditions various. The likely recognitions of pictures from various perspectives are then consolidated to find the conditions of the signs on a guide. NLP (Natural Language Processing) is used as text extraction from the recognized parking signs and displays a result based on text recognition weather we can park or not on specific time and date. So, this exhibits the capability of utilizing road level pictures and flexibly a practical answer for digitizing at scale all parking signs to help drivers comprehend parking rules and keep away from fines.**

Index Terms **- computer vision, NLP (Natural Language Processing), text extraction, text recognition.**

I.INTRODUCTION

Optical Character Recognition is generally abridged as OCR. It is a direct result of this innovation that a gadget can distinguish the typeset naturally through an optical technique. We People get to know numerous items along these lines through eyes for example "optical system." In spite of the fact that the mind "sees" the info, the capacity to comprehend these signs fluctuate in each individual dependent on numerous elements.

OCR is an essential area of research in the domain of pattern recognition. The possibility of an OCR framework is to recognize letter sets, numeric, accentuation imprints, or unique characters, present in digital images, with no human contribution. A picture of each character must be changed over to suitable character code. This is accomplished through coordinating cycle between the removed highlights of given character's picture and the library of picture models. Preferably these highlights should be diverse for dissimilar character pictures for making it conceivable for the computer to take out the exact copy from the library with no mistake. Additionally, these highlights must be hearty enough with the goal that they may not get influenced by survey changes, commotions, changes in goal like components. Each OCR step is significant; the entire OCR cycle will fall flat if just a single its progression cannot deal with given picture effectively. To start with, if People read a page in a language other than our own, we may perceive the different characters.

The principal target in this paper is to perform extraction and detection of the text from parking sings. If we look at some metropolitan cites the parking signs or parking boards will be having fewer conditions so that they maintain proper help for the government to clean the area, fewer boards will contains specific conditions such as the vehicle should be parked from a given time on a given days, few may contain week days only few may contain weekends only, if we park the vehicle on a day that we shouldn't have parked we will be get a ticket and will be fined. So, our goal is to

get rid of these tickets by extracting text from the detected parking sign and develop a logic that the extracted text can show the exact result for the parking sign conditions given. This logic will satisfy for every parking signs specific conditions according to the date and time mentioned in the parking board. So, that the driver comprehends parking rules and keep away from getting a ticket or fines.

II.TEXT AND RECOGNITION

Parking signs is of different types, each sign board will vary with other, and each board will be having its own specific conditions with their own different style and each sign board will be having different colors. So, our training data should be in a way that contains all different styles of parking signs. For the testing purpose we are using SF Parking Signs dataset from Kaggle which is an open-source datasets library where we can use it for free. In a nutshell with the help of this dataset we can train our model to extract the text from the parking sign and the text will be cross checked with the logic and will be given a specific output. As of a survey looking at few algorithms that can be used which are listed below.

III.PROGRESS IN TEXT EXTRACTION FROM IMAGES

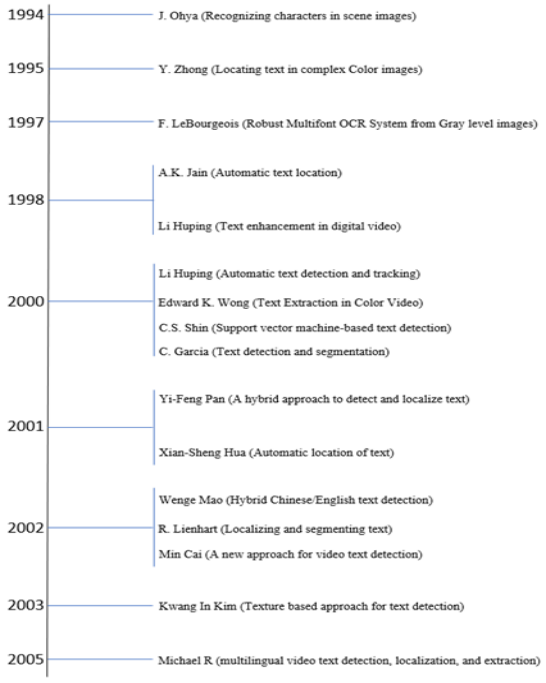

Fig. Chronological progression of text detection

Ohya et al. (1994) introduced a four-stage strategy: in the first stage binarization was done dependent on nearby thresholding, at that point the character part was distinguished utilizing dim level contrast, in the third stage acknowledgment of characters is performed by coordinating them with the standard examples put awayin an information base, and in the last tage unwinding activity is performed to modernize the similitudes. In their paper the characters, multi-fragment characters were removed and remembered, they likewise dealt with characters of changing, sizes, positions, and text styles under fluctuating lighting up conditions in scene text pictures. The paper inferred that parallel division is inadmissible for video records, due to the presence of a few items in it that too with various dark levels and due to the presence of commotion (that too at high levels) and light varieties. Moreover, this methodology works with text arrangement and shading limitations. The creators performed investigates 100 pictures and achieved 85.4% of the review rate had been gotten. While the character acknowledgment pace of 66.5% was gotten.

F. LeBourgeois (1997) limited content in complex grayscale text pictures. After pre-handling, picture inclinations were spread the level way. Associated segments were found in the subsequent picture to restrict text areas into text lines. Text lines were then portioned to get the characters separately by finding valleys in both even also, vertical projection profiles.

Jain and Yu (1998) applied a CC-based technique after pre-processing, which included bit dropping, shading grouping, multi-esteemed image deterioration, and forefront image age. They diminished the pieces of the 24-bit shading images to as low as 6-cycle images and afterward utilizing bunching calculation quantized the images. Initially, Info image was decayed into numerous closer view images and afterward text limitation was performed on every one of these closer view images. Parallelly, CCs were created for each forefront image by utilizing a square nearness chart. The confined text parts of person frontal area images were then consolidated into a solitary yield image. The calculation was tried with different kinds of images, for example, paired record images, web images, filtered shading images, and video imhongages. Just even and vertical text was separated, while the calculation was not tested on slanted text and for inadequate shading histogram.

Huping Li et al (1998) introduced a framework for the location and following of text in computerized video consequently. A half breed wavelet/neural organization-based technique was utilized to identify text locales. The following module utilized SSD based image coordinating to locate an underlying position at that point shape-based adjustment was used to refine the coordinating positions. The outcomes proposed that text improvement was important for sensible outcomes.

Li et al. (1999) took a shot at text following methodology for looking over text, printed text and subtitle text. They utilized the SSD (entirety of squared distinction) for an unadulterated translational movement model. This model diminished the computational intricacy as it was based coordinating at multi-goal. To balance out the following cycle, they utilized text forms, in more perplexing movements. For marginally bigger text block they utilized shrewd administrator to produce energy maps. Notwithstanding, since an unadulterated translational model was applied, they infer that procedure utilized in the paper was not appropriate to deal with varieties in scale and pivot.

The problem of Text detection is still open in realistic images. Till date Presented text detection approaches or techniques perhaps are divided into two major classes. First is connected component-based method (Y. Zhong et al. 1995)

(A.K. Jain, B. Yu 1998), and another is texture-analysis-based method (H. Li et al 2000) (X. Gao, and X. Tang 2000). The second category algorithms obtain regions for text detection by examining the spatial allocation of edges/boundaries or

consistent colour/grayscale segmented text components. For example, Y. Zhong et al (1995) extracted connected components of repetitive colour text which follow some size constraints and horizontal alignment constraints. Further, Texture analysis-based methods could be separated into top-down approaches and bottom-up approaches. C. Garcia and Apostolidis (2000) located horizontal text in colour images. Edge pixel magnitudes and locations are determined in each colour plane. They selected Text regions by identifying high edge density areas and high variance of edge orientation. This prevented incorrect identification of regions with simple edges uncharacteristic of text. Morphological operations were performed to remove singletons and nonhorizontal regions. localization was performed by finding connected components.

Edward K. Wong et. al (2000) proposed a vigorous calculation for extraction of text in Shading Video. The calculation works by discovering potential text line sections from check lines which are level. Perceived fragments of text line are extended or gotten together with line sections of text from check lines which are nearby it to make squares of bigger text, which are at that point subject to refinement and sifting. Text pixels within blocks of text are then found by using associated segment and bicolor bunching investigation. The calculations of morphological goal upgrade and morphological form smoothing are then applied to the distinguished twofold texts for improving their visual quality. The calculation actualized has speedy time of execution and is practical in text recognition in a tough situation a few cases like scenes with foundation having significantly texture and furthermore little text scenes.

C.S. Shin et.al (2000) talked about a strategy which depends on help support vector machine. Textual data within casings of video are amazingly useful for depicting the video outlines content, since they engage search which depends on free and catchphrase text. In this paper, the issue of area of text in advanced video as texture order which is administered, and use uphold support vector machine (SVM). Not under any condition like strategies for discovery of text, there is no joining of any of the express plans of extraction of texture highlights. Or maybe, the estimations of dark level of crude pixels are explicitly urged to the classifier. This relies upon the observation that SVM has capacity of learning in space which is of high measurement and consolidate plan of extraction of highlights in its own engineering.

Kim et al. (2001) utilized help vector machines (SVMs) for investigating the textural properties of text in images. SVMs functioned admirably even in this high-dimensional space and can join a component extractor inside their design. After texture characterization utilizing a SVM, a difficult investigation was performed to separate text lines.

Chen et al. (2001) utilized the shrewd administrator to identify edges in an image. Just one edge point in a little window was utilized in the assessment of scale and direction to decrease the computational unpredictability. The edges of the text were at that point improved utilizing this scale data.

Morphological expansion was performed to interface the edges into groups. A few heuristic information, for example, the even vertical perspective proportion and stature, was utilized to Modify out non-text groups. Two gatherings of edge-structure Adjust, and a stripe-structure Change and a neural network were utilized to gauge the size of the edge pixels in view of these Modifies' yields. The edge data was at that point upgraded at a suitable scale. In that capacity, this outcomes in the
end or obscuring of structures that do not have the particular scales. The text restriction was applied to the upgraded image. The creators utilized a business OCR bundle (Type Per user OCR bundle) after size standardization of singular characters into 128 pixels utilizing bilinear addition.

Xian-Sheng et al (2001) proposed another programmed text area approach for recordings. Most importantly, the corner purposes of the chose video outlines were recognized. In the wake of erasing some segregated corners, they blended the remaining corners to shape up-and-comer text areas. The districts were then disintegrated vertically and evenly utilizing edge guides of the video edges to get up-and-comer text lines. At last, a text box confirmation step dependent on the highlights gotten from edge maps was taken to essentially lessen bogus cautions. Trial results demonstrated that the new text area plot proposed in this paper was precise.

R. Lienhart and A. Wernike (2002) proposed a novel technique for restricting and dividing text in complex images and recordings. Text lines were distinguished by utilizing a complex-esteemed multilayer feed-forward organization prepared to distinguish text at a fixed scale and position. The organization's yield at all scales and positions was coordinated into a solitary text-saliency map, filling in as a beginning stage for competitor text lines. On account of video, these competitor text lines were refined by abusing the transient repetition of text in video. Confined text lines were at that point scaled to a fixed stature of 100 pixels and fragmented into a double image with dark characters on white foundation. For recordings, fleeting excess was abused to improve division execution. Information images and recordings can be of any size because of a genuine multiresolution approach.

Min Cai et al. (2002) proposed another productive video-text-identification approach that was fit for recognizing text in a mind-boggling foundation, furthermore, was strong for text dimension, textual style tone, and language. To start with, it changed over the video image into edge map utilizing a shading edge identifier (J. Fan et al 2001) and utilized a low worldwide limit to sift through certainly non-edge focuses. At that point, a specific neighbourhood thresholding was performed to rearrange the complex foundation. An edge-quality smoothing administrator and an edge grouping power administrator were intended to feature those regions with high edge quality or edge thickness, for example text applicants.

Wenge Mao et al. (2002) proposed a multiscale texture-based strategy utilizing nearby energy investigation for mixture text recognition from the edges. The neighbourhood energy variety was determined in a nearby locale utilizing wavelet coefficients of image. The pixels of the foundation and non-textual districts had low nearby energy varieties when contrasted with the textual locales. This separation was used in this paper then thresholding was applied upon the images and text identification was finished utilizing associated part investigation and mathematical separating.

Kwang kim et al. (2003) offers a novel based method for detecting texts in images this will be helpful in analysing textural properties of texts they used SVM (Support Vector Machine).

Research	Methodology	Method Used	Accuracy
[1]	Character recognition	introduced a four-stage strategy which dealt with recognizing characters of changing, sizes, positions, and text styles	85.4%
[2]	Text Location detection	Texture analysis-based methods could be separated into top down approaches and bottom-up approaches	84%
[3]	Object Character Recognition	Delt with complex Gray scale images to restrict text areas into text lines	95%
[4][12]	Text detection	introduced a framework for the location and following of text in computerized video consequently	92.8%
[5]	Text Extraction	calculation for extraction of text in Shading video	88%
[6]	Text detection	discussed a method which is based on support vector machine for textual information inside of frames in video	94.5%
[7]	Text detection	Text detection was implied with the help of support vector machine and adaptive mean shift algorithm	71.5%
[8]	Text segmentation	a novel technique for restricting and dividing text in complex images and recordings	69.5%
[9],[13]	Text detection	video-text-identification approach so that fits the text identification	90%
[10]	Text detection	A multiscale texture-based strategy utilizing nearby energy investigation for mixture text recognition from the edges	88.1%
[11]	Text location detection	Identification of the text as the video will be break down into frames of images	94.7%
[14]	Text detection and segmentation	They selected Text regions by identifying high edge density areas and high variance of edge orientation	93%
[15]	Text location detection	The detection of the text that where in videos as break down into frames	90.5%

Table. Accuracy results for text detection practices.

IV.SIGNBOARD DETECTION

Angela Tam, Hua Shen, Jianzhuang Liu, and Xiaoou Tang (2003) proposed an image processing technique where all the images were turned into grey scale to detect the text in boards. Edges are found by noise reduction with smoothing and media filters and sober edge detection then Hough transform is applied for the output filtered image. The detection of text in signboards was successful with an accuracy of 83.33%.

Xiong Changzhen, Wang Cong, Ma Weixin, Shan Yanmei (2016) proposed a model to detect all major categories in China. It takes only 51ms average speed to detect each frame with a resolution of 640*480. The use of Faster R-CNN model by end-to-end training method with the help of pretrained model ImageNet the initialization of weights was done. Three more models were trained VGG16, VGG_CNN_M_1024 and ZF. Then the three model is tested by the testing dataset (4706 images). The Mean Average Precision(mAP) and the average test time in different number of iterations for three trained model were noted for the comparison the final result is 99.62,99.24 and 99.33% was noted.

V.CHALLENGES

A study of the literature reveals that Most methods have been developed to extract text from complex colors images and have been extended for application to video data. However, these methods do not take advantage of the temporal redundancy in video. The foremost challenge in extracting news ticker from the video is to separate out frames from the video. A small video clip of say 4 - 5 MB size may have more than 3000 frames in it. All the frames may not be containing the news ticker, as for a while the video screens of news channels do not have news ticker. Extracting the relevant frames is a big issue. Extracting text from the frames is also a challenge because of background contrast and colour bleeding.

Text quality is decreased due to the presence of noise and image encoding and decoding procedures.

- Unspecified text colour: text can have random and nonuniform colour.
- Unknown text size, position, orientation, and layout: captions lack the structure usually associated with documents.
- Unconstrained background: the background can have colours similar to the text colour. The background may include streaks that like that of character strokes.
- Colour bleeding: lossy video compression may cause colours to run together.
- Low contrast: low bit-rate video compression can result into undistinguishable contrast between character strokes and background.

Albeit much exploration work has been accomplished for text in English, Arabic or Chinese Artificial text in Videos, very little research work has been done on recordings containing Punjabi Gurmukhi text in it.

VI.CONCLUSION & FUTURE WORK

In this work, we have done the survey of Text extraction and recognition from images. The extraction of frames out of the videos itself was a challenge. Text is to be extracted based on segmentation of word and character extraction from text Line. The main future scope of our work is to develop a methodology to extract the text from the video with increased recognition rate, since in video images the resolution is so low, recognizing compound characters becomes more and more difficult. So, that will be the main point of interest for further modification in near future.

REFERENCES

[1] J. Ohya, A. Shio, S. Akamatsu (1994). Recognizing characters in scene images. IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (2), 214–224.

[2] Y. Zhong, K. Karu, and A.K. Jain, “Locating text in complex color images,” Pattern Recognition, vol. 28, no. 10, pp. 1523-1535, 1995.

[3] F. LeBourgeois, “Robust Multifont OCR System from gray level images,” in Proc. of International Conference on Document Analysis and Recognition, vol. 1, pp. 1-5, 1997.

[4] Li Huping, Doermann D and Kia O., “Automatic text detection and tracking in digial video,” IEEE Transactions on Image Processing, vol. 9, no. 1, pp. 147-156, 2000.

[5] Edward K. Wong, Minya Chen, “A Robust Algorithm for Text Extraction in Color Video,” in

Proc. of IEEE International Conference on Multimedia and Expo, vol. 2, pp. 797-800, 2000.

[6] C.S. Shin, K. I. Kim, M.H. Park, H. J. Kim, "Support vector machine-based text detection in digital video, "in Proc. of IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X, vol.2, pp. 634-641, 2000.

[7] Kwang In Kim, Keechul Jung, and Jin Hyung Kim, "Texture based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp. 1631-1639, 2003.

[8] R. Lienhart, A. Wernike, "Localizing and segmenting text in images and videos," IEEE Transactions on Circuits and Systems for video Technology, vol. 12, no. 4, pp. 256-268, 2002.

[9] Michael R. Lyu, Jiqiang Song and Min Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 2, pp. 243-255, 2005.

[10] Wenge Mao, Fu-lai Chung, Kenneth K.M.Lam and Wan-chi Siu, "Hybrid Chinese/English text detection in images and video frames," in Proc. of 16th International Conference on Pattern Recognition, ICPR IEEE Computer Society, pp. 1015-1018,2002.

[11] A.K. Jain, B. Yu, "Automatic text location in images and video frames," Pattern Recognition, vol. 31, no. 12, pp. 2055–2076, 1998.

[12] Li Huping, Doermann D and Kia O., "Text enhancement in digital video," in Proc. of SPIE, Document Recognition IV, pp.1–8, 1998.

[13] Min Cai, Jiqiang Song, and Michael R. Lyu, "A new approach for video text detection," IEEE International Conference on Image Processing, vol. 1, pp. I-117 - I-120, 2002.

[14] C. Garcia and X. Apostolidis, "Text detection and segmentation in complex color images," in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2326-2329, 2000.

[15] Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, Hong-Jiang Zhang, "Automatic location of text in video frames," in Proc. Of the 2001 ACM workshops on Multimedia: multimedia information retrieval, pp. 24-27, 2001.

[16] Angela Tam, Hua Shen, Jianzhuang Liu, and Xiaoou Tang. "Quadrilateral signboard detection and text extraction" In Proc. Of the 2003 in ResearchGate publication.

[17] Changzhen, Xiong; Cong, Wang; Weixin, Ma; Yanmei, Shan (2016). "A traffic sign detection algorithm based on deep convolutional neural network". IEEE International Conference on Signal and Image Processing (ICSIP). pp., 676–679. doi:10.1109/SIPROCESS.2016.7888348