IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Subtitle Extraction from Videos and Translation from English to Hindi

Abhinav Y. Watve, Madhuri A. Bhalekar
School of Computer Engineering and Technology, MIT World Peace University,
Pune, Maharashtra, India

***Abstract:*** In current situations, videos provide much help to better understand the information regarding any type of concept or any data relevant to the user. Therefore it becomes very important to make such content available to a wider audience along with audiences with disabilities and remove the barrier of languages. This can be achieved through the subtitles that are embedded inside the frames of the videos. However, these subtitles are limited by the language that is decided by the content creators. Solving this problem of restriction is an important subject of research. Therefore this paper includes the literature review of text extraction and text detection. The paper also proposes a method for extracting and translating the subtitles from videos. It also includes comparison between proposed method and an exhaustive method that does not include a support system and the proposed system that does. It shows how the proposed system goes around the problem of identifying the frames that contain subtitles and those that don't.

***Index Terms*** **- Text Extraction, Text translation, OCR**

## I. INTRODUCTION

Videos have texts in two types of forms [16]. First one being the Scene text, which is captured by the cameras while shooting the scenes. To give an example of this text, text written on the different buildings, banners and signs. The second type is the artificial text which is inserted artificially into the videos by using editing software. Usually the location and orientation of this text is fixed. General position being at the bottom part of the frames and having black border and white filled text.

Actually this artificial text can be further divided into captions and subtitles.

- Captions: Captions contain text which identify who is speaking in the current scene. Captions change their placement if they are obstructing any important content in the videos. Captions include not only dialogues but also other audio components like background noise of a train passing or crying of a baby.
- Subtitles: Subtitles contain only the dialogues that are being spoken by the person in the videos. They do not include any non-speech elements. They are usually synchronized with the media.

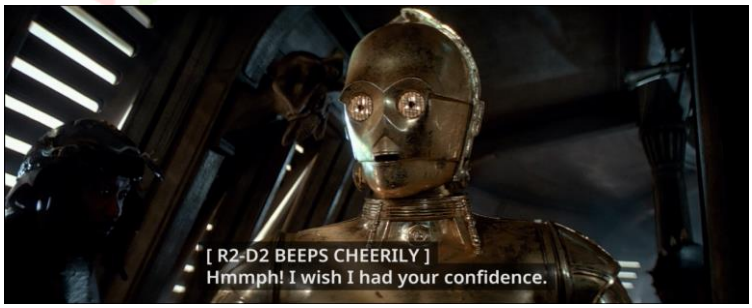


*Figure 1: Example of Captions*


*Figure 2: Example of Subtitles*

Though we are not going to discriminate between these two in this paper, it is important to understand the difference between the two. This difference can be further used. But in this paper the focus will be on the artificial text.

Also, video players provide facilities that external files can be used to display this artificial text over the videos. This text is not the artificial text that we are aiming for. The text that we are aiming for is the one which is embedded inside the frames of the videos.

## II. TECHNOLOGICAL DIFFICULTIES

*a)* Duration: The duration for which the subtitles are displayed is not fixed. This duration keeps on changing in single video let alone other videos. This makes it difficult to identify which frames in the video have the subtitles.

*b)* Background: As the text is embedded in the frames or the images, the background keeps changing throughout the video. This background of text has an impact on some of the techniques discussed further in the paper.

*c)* Standard: There is no universal standard that is followed for font, size, site and color of the text. There is a general norm that is followed regarding the color being white with black borders and site being the lower part of the frame.

*d)* Resolution: Resolution of the video has a great impact on the text extraction mechanism. If the resolution is too low that the text becomes unreadable then mechanism also fails to perform properly

## III. TEXT EXTRACTION APPROACHES

These different approaches can be differentiated based on how the artificial text is separated from the media content. Though there 4 approaches they can be mainly distinguished into 2 types [22]. First being the traditional approach which uses text features, texture and morphological operations to extract text. The second type is the smart way or the machine learning approach.

*a)* Region Based Approach: Numerous times the artificial text in a video has a consistent color and font. Also this color vastly differs from the media content. To segment this text from the image, color congregation and research analysis can be used.

*b)* Texture Based Approach: In a video, the style of the artificial text is most of the time consistent. Hence, this text space is often treated as a distinct texture. Then, we are able to use the features of this texture to style applicable texture segmentation algorithms to extract the subtitle text. Common ways for texture analysis include Gaussian Filter, Wavelet Transform and space variance, Gabor Filter etc.

*c)* Feature based Approach: The artificial text embedded within the frames have robust edge features. This approach initially finds out the edges of the total image, then connects these features into pieces with the strategy of smoothing filtering or morphological dilation. Finally, filter text using specific knowledge of the text.

*d)* Machine Learning Approach: This approach analyzes the frames and finds the patterns and makes use of this pattern to predict the unknown information. Accurately get the parameters of the model by the method of example learning, it upgrades the dependability of the test results.

Though the techniques can be categorized into 4 approaches, as technologies proceed these approaches can be seen overlapping. For better results of text detection and text extraction the techniques developed used multiple approaches for text detection as well as text extraction. By using multiple approaches the process becomes a lot complex but the results become better. Also time required for technique using multiple approaches was more than simple. The machine learning approach did the comprehension of the rules all on it's own. This makes the time requirement for this approach the longest.

## IV. LITERATURE REVIEW

Even though the aim of the project is to draw out the text from the videos, it is not the sole method that is required to retrieve the text from the video. There are many preprocessing steps that are required to successfully extract text from the videos. The whole text extraction process can be divided into 2 simple steps, first one is text detection and second one is text extraction.

*a)* Text Detection: Text detection can be said to be the preprocessing step that is important for better text recognition in later stages. During the literature review, we found many different techniques that use different methods for text detection.

Paper proposes [1] to use the Marr-Hildreth edge detector to divide the image according to the edges in the images. After this basic global thresholding is done. Following the basic thresholding optimum global thresholding is done using Otsu's method. This leads us to detect the text in frames. Paper proposes [3] to separate the background of scene images and text by making the use of SR based morphological component analysis. From training samples, background and dictionaries of text are learned. In the text and background separation stage these dictionaries are used. By solving the equations once dictionaries are learned separately, the coding coefficients of a query sample can be obtained. In the reconstructed text picture the gravity points are calculated and connected with the equivalent points in the original picture. The candidate text areas are then boxed. By joining the nearest rectangles the final text area is obtained.

Another one such paper proposes [7] extracts frames from the video. Roberts edge detection operator is applied on the same frame in two different formats. First one is directly applying the operator and second is rotating the frame 180 degree and then applying the edge detection operator. These two edge maps are integrated together using the logical and operator. After making the use of morphological operators, pixels of the text are joined in a group and remove the pixels that appear isolated from the text. Around the text area forming a text box, a bounding box is constructed.

To recall additional text by touching on alternative frames, paper [8] proposes to use spatial temporal video text detector which means temporal relationship in between back to back video frames. Text trailing, text quality rating and trailing modules are combined by Text recommender into a network. This FREE is needed to identify the text region with the very best quality score in a tracked text stream in distinction to previous methods going through every text region from the tracked stream. A novel framework is planned to discover multi-color and multi size captions from low quality video with quick speed. At the start, edge maps are computed by the sobel operator on 4 orientations and 2 polarities. For correct text localization, perform stroke density analysis and nearest character grouping. Finally use a texture analysis algorithmic program on the detected text blocks.

In the paper [20], Input image is passed to a text region detector that detects text and non-text regions. Conjointly text confidence and scale information is calculated. Then scale adjustive local binarization is applied to come up with candidate text elements. The Connected Component Analysis (CCA) stage Conditional Random Field (CRF) model combines the single element properties and binary contextual element relationship is employed to separate out text and non-text elements. At the ultimate stage near text elements are connected with a learning based MST algorithm and the middle line or words are cut off with an energy minimization model to cluster text combining into text lines or words.

A novel framework is proposed in paper [11] to identify multi-color and size captions from lower quality video with quicker speed. At the start, on four orientations and two polarities edge maps are calculated by the sobel operator. After that perform stroke density analysis and adjacent character grouping for correct text localization. Finally use a texture analysis algorithmic program on the detected text blocks. The plan of action planned by another such paper [14] uses multiple steps to discover text within the images. At the beginning, the product of Laplacian and sobel operation is calculated and is known as Laplacian-sobel product. Then Bayesian classifier is employed to sort true text pixels and non-text pixels.The text candidates are picked by crossed output of the bayesian classifier with canny operation of input image. Then the Boundary Growing Technique is used to eliminate false positives and discover the text from the images.

The method planned within the paper [18] 1st performs edge detection on the image to detect the edges. The edges typically demonstrate the strikes to show the text. Symmetry may be derived from gradient direction. Character elements within the corresponding text line reveal symmetry in between intra and inter character parts and constant stroke width. The planned technique traverses geometrical properties of sub-graphs fashioned by triangulation, like edge strengths and density for text candidates to eliminate false ones, which ends in potential text candidates. It's true that multi-level decomposition in a pyramid structure is a stimulating plan for enduring with multi-sized or multi-font texts. Hence, it proposes multi-scale integration in a pyramid structure on totally different scales to extract the complete text line.

The paper [13] uses blob feature detection technique. Foremost image is binarized and candidate text features are extracted. These extracted blobs are allotted a possibility of being text supported on histograms of geometric properties of blobs learned from training data. The connected neighbour blocks are joined into super blobs based on similarity measures. These superblobs are then classified into text and non-text categories. Another paper [19] proposes Harris corner detection detection, corners are drawn out from the image. After this, morphological dilation is performed to join the separate corner points that are near to one another. Then CART is employed to sort the text and non-text.

*b)* Text Recognition: The paper [2] selects new individuals based on the selected threshold value of the fitness function. Identification of key points is performed using the SIFT algorithm. After that local enhancement operations are performed on the image. Then on the image morphological operations are performed. Final output images with text regions extracted are obtained. Another paper [4] proposes their own text detection and recognition technique. The input video is split into frames. Then they're converted into gray scale images. After this, sobel and canny edge detectors are used to draw out edges from the images. After performing dilations on these images, it joins the character contours of every text line. The dilated regions are sent to seek out whether text-like regions are text or not. Once these regions are known, the images are passed to optical character recognition software for recognizing the texts.

The method planned by paper [5] uses the sobel edge detector to calculate the edges within the image and the edge which has a value greater than hundred and fifty will be considered edge from the subtitle. The range between the upper and lower bounds, the left and right boundary can be found in frames. The upper and lower boundaries can be found based on the statistics. The possibility that there are more frames without subtitles among samples than those with if the samples are not sufficient. Subtitles are typically filled with light colors such as white. By adjusting the color gradient of the gray scale image, the tone of the text area

and the background area can be separated easily. By executing binarization subtitles become easily visible. This image is passed to OCR for text recognition.

Unified Bayesian based framework planned by paper [6] has both trailing based text detection and trailing based text recognition from complicated videos, totally different from standard ways of solely tackling one amongst two tasks individually. Additionally proposes a novel trailing by detection approach for text trailing different from standard ways only focusing on region matching. Another contribution being trailing based text detection and tracking based text recognition approach. Lastly, a new sensible dataset for text detection and recognition from net videos.

The Gamma correction methodology proposed by paper [9] suppresses the non-text background details within the image by applying precise gamma value and removing non-text regions. The algorithm foretolds the gamma value by using texture measure without knowing any details about the imaging device. By applying this gamma value to the image, background suppressed images are going to be achieved. Gray level co-occurrence matrix for every image is computed to extract the textural features after changing into gray images. To work out the gamma value, three rules are defined. Background suppressed image is then converted into a grayscale image. To come up with a threshold value and apply it to make output images, Otsu's thresholding algorithm is employed.

The paper [10] uses a neatrest of connected elements within the image determined by the Voronoi regions of their centroids. To draw out text within the images based on the GMM of 3 neighboring characters described, we consider each character in an image as connected elements in its binary image and perform morphological closing operations on the binary image. Label every connected element based on GMM neighbour characters. Another paper [12] firstly uses a canny edge detector to identify edges of the text and differentiate it from the background of the text. Then the task of spatial localization is based on edges' distribution and corners identified in images. Then caption segmentation is completed to find the caption region. Then industrial Hanwang OCR software was used to identify the images in binary format created by the proposed method.

Another paper [15] uses a connected element approach. Initially some preprocessing is needed for text extraction. The image is split into 2 subimages. These 2 are then converted into gray scale images which are then converted to binary images. Then the planned text extraction method is applied on sub images. This extracted text is then written into another gray image.

The paper [16] initially takes the frames and passes them to a text region detector to identify candidate regions. Perform morphological operations subsequently to get rid of noise. When the text area is identified, Connected component analysis(CCA) is performed to identify text in frames both horizontally and vertically. In CCA the CRF is employed to sort the candidate area into 2 categories, text and non-text. Then ANN is trained to classify the text and non-text parts. For character recognition, this extracted text is passed to the OCR. Ultimately characters are sorted into words and in turn lines by employing the horizontal and vertical bounding box distances by building a minimum spanning tree.

The paper [17] proposed uses morphological operations and is firstly used to identify the candidate regions. A bounding box is employed to get rid of the noise in the images, after this use the edge ratio to cut off noises with no strong edges in distinction. Employing a fill hole method to deal with the background enclosed by the text regions. Using thresholding alternative areas are removed from the image except the text. Then a Canny edge detector is employed to decrease noise and draw out edges from the image. To identify text, raster scan is employed together with alternative supplementary techniques.

Along with these different papers that proposed different text detection as well as text extraction methods some papers that survey such papers were also referred.[21][22]

## V. RELATED TECHNOLOGIES

The technologies used for subtitle extraction and translation are mainly OCR and the library for translation. The main purpose for using these technologies will be discussed in the following points.

*a)* Optical Character Recognition: OCR stands for Optical Character Recognition. OCR can be used to identify text that is either handwritten or electronic. OCR can be classified into 4 different stages, which are as follows, pre-processing, text detection, text extraction and post processing.

Preprocessing is required so that the process of OCR gives better results. It can include techniques like binarization, noise reduction, line removal, layout analysis, segmentation and normalization. Sometimes one of these techniques is while other times multiple methods are used in conjunction with each other. Text detection and text extraction can be part of text recognition. First, the text needs to be detected for the text to be extracted. Each OCR method has their own detection and extraction method. Also, these OCR methods have different datasets for training which leads to differences in results.

Post Processing is a kind of optimization of output acquired from text recognition. There are different techniques like constraining output by using lexicon, levenshtein distance, etc. Although preprocessing, text recognition and post processing are included in OCR, in actuality, preprocessing and post processing are complementary processes required to increase the output of the actual process of OCR which is text recognition. The OCR which we are using in this project is Keras OCR which is free to use and is made available as a library on github.

Measuring the efficiency of the text extraction process can be done by counting the the characters or words that are extracted against the characters or words that are actually present in the video.

*b)* Translation Library: Translation is the process of converting the meaning represented in one language into another. This process becomes difficult because each language has its own flavor and it's very difficult to relay the same flavor in another language. To make this translation process more efficient and reliable translation libraries are developed. Libraries are developed by researchers to provide easy access to features not easily available to other researchers or other students.

Different types of methods are used for language translation in NLP (Natural Language Translation). They can be divided into 4 basic models, Statistical Machine Translation (SMT), Rule Based Machine Translation (RBMT), Neural Machine Translation

(NMT), and Hybrid Machine Translation (HMT). The library that we are using, which is googletrans, uses NMT model without further specification.

Translation can't be measured quantitatively like other processes of computation. Therefore, it needs to be measured qualitatively. This measurement is usually done by humans as they can understand the natural flavors of languages. Quality of translation can be measured by following these three conditions. First one is 'Translation accuracy'. This term refers to conveying correct meaning from one language to another. However, authority regarding the meaning being conveyed rests with the owner of the content.

Second one is 'Translation Structure'. This term refers to using accurate structure of the required language that is the grammar. There is little room for creativity when accompanied by broken grammar. This is more so for languages that are more sensitive like Chinese or Arabic. Third one is 'Creativity in Translation'. This includes taking into consideration the culture of the language being translated into.

*c)* Support System: Support system is just a name given by us to the system that supports the main system of subtitle extraction and translation. The technical part will be explained in later stages, so let's get the theory down first.

This support system uses the audio component of the video used. In any video with subtitles, most of the time it includes dialogues. We tried to separate these dialogues from the background silence. By using these separated dialogues we calculated the dialogue durations to point out the frames which may contain the subtitles. These frames are then used for subtitle extraction and translation which is our main system.

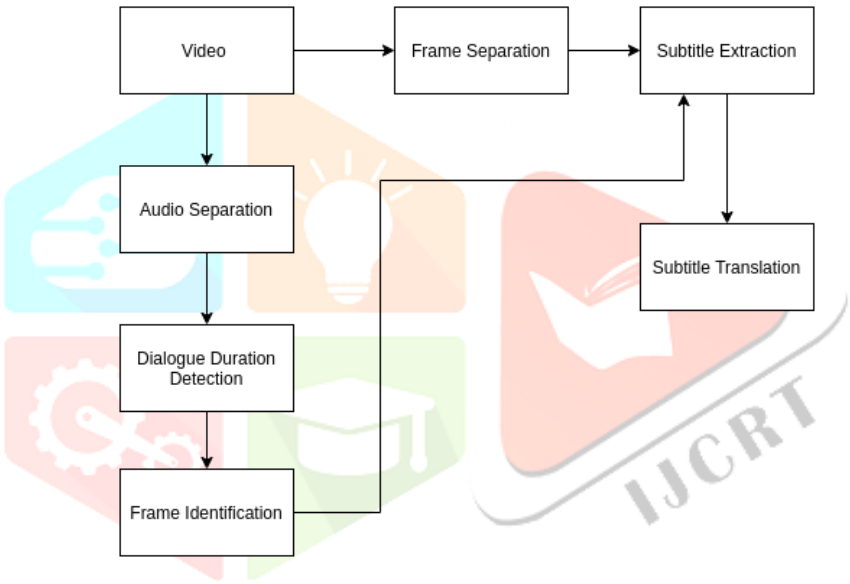


*Figure 3: Project Architecture*

## VI. PROPOSED SYSTEM

Proposed system can be divided into three parts. First part is the support system that identifies the frames that need to be scanned. The second part is a subtitle extraction system that performs text extraction operation on the specific frames using the OCR technology. The third part is the text translation which uses the pre-developed library by Google.

*a)* Subtitle Extraction: Subtitle Extraction is the process of retrieving texts from frames of video in question. This process includes the use of OCR software. While we are discussing the topic of OCR, let's discuss why we decided to use OCR rather than to develop our own special method for doing the same task. Following are some of the reasons that we decided to use OCR over a self-developed method.

A pre-developed OCR model gives better accuracy than a model that was developed by amateurs in the field. Such pre-developed OCR models are optimized so they give faster outputs. It also gives faster updates to OCR for any requirements requested by customers. It also has better results as it is trained on a lot more data than the one which we could have developed.

In this project the OCR that we are using is Keras-OCR. It is a packaged and flexible version of CRAFT text detector and Keras CRNN recognition model. The most basic reason for choosing this OCR is that it is free to use and it performs very well even against OCRs developed by Google and Amazon. This ocr ignores the punctuation and letter case because of the In *Fig.3* we can see the architecture of the project. It shows the full process of how the subtitles are extracted and translated. The Subtitle extraction process includes sub processes like Frame separation, Subtitle Extraction, Subtitle set creation.

*b)* Subtitle translation: Subtitle translation is the process of translating the subtitles extracted from the frames of the video. This step is the last of the project and also the one where the final output of the project becomes available. In this project we are using googletrans library. It is a free and unlimited python library that implements Google Translate API. It uses Google Translate Ajax API to make calls like detect and translate.

In this project, we directly use the library to translate the subtitles. It gives quick output and also very easy to integrate it into the project. It is also flexible as it provides many different options for changing the languages for detecting as well as translating. It also provides over 100 languages for translation. It can also detect language given as input if it is supported by the library.

*c)* Support System:

I. Audio Separation: During this first step, the audio part of the video is separated from the video. The reason for doing this is that the library that we use to perform operations on the audio can't be used for video directly. The process of separation creates a new audio file. This file is passed to the next process for further operation.

II. Dialogue Duration Detection: Though this part is called dialogue duration detection, it is actually detection of silence from the audio clip. By detecting silence we also identify part of the audio which may contain the subtitles. To make sure that selected part contains subtitles there are multiple variables that need to be adjusted. These variables include the amount of duration considered silence in milliseconds, amplitude of sound in decibels, etc. However, in this project due to the lack of resources required we have restricted these variables to the default constants.

III. Frame Identification: The detected dialogue duration is where the frames will be that contain the subtitles of the video.

$$\text{Timestamp} = \text{frame no.} / \text{frames per second} \quad \text{Eq No. 1}$$

The 'timestamp' in the above equation relates to the timestamp of the frame in which the subtitles could possibly be present. Though this method sounds simple, the idea was really good in our opinion. The process is simple but the requirement for the physical resources for processing is quite large. If we consider the factors with which the execution time and required resources increase, the first on the list would be the size of the input video. The second would be the program itself.

Commented [aw1]:

Commented [aw2]:

## VII. RESULTS

Following are the results of the project. These results contain comparison between the exhaustive approach and the proposed method. The exhaustive method is where there is no support system for extracting and translating the subtitles in the video. As there is no support system, there is no limit on the number of frames that have to be scanned for identifying subtitles in the video. As the execution time for the exhaustive system is a lot longer than the proposed system we are using GPU for the OCR process so as to reduce the time required. The videos that we are using for testing purposes are 60 seconds long. The video is kept shorter because the longer the video, more the resources required will be.

| | Exhaustive Method | Proposed Method |
|---|---|---|
| Execution Time | 25.216 min | 3.836 min |
| Frames Scanned | 1799/1799 | 269/1799 |
| Character Extracted | 370/411 | 273/411 |
| Translation Quality | 7.84/10 | 7.03/10 |

*Table 1: Comparison for Video 1*

| | Exhaustive Method | Proposed Method |
|---|---|---|
| Execution Time | 29.4 min | 1.207 min |
| Frames Scanned | 1799/1799 | 72/1799 |
| Character Extracted | 458/556 | 41/556 |
| Translation Quality | 6.93/10 | 5.46/10 |

*Table 2: Comparison for Video 2*

We will see the results of the two systems on two videos. The execution time for the exhaustive system is approximately the same as it is directly related to the number of frames used by the system. The exhaustive system does its job of extracting characters well enough as the characters that were not extracted successfully belong to the category of the punctuation marks. The translation quality for the exhaustive system is also near average . Now let's see the performance of the proposed system. As mentioned above, the execution time is dependent on the number of frames used for processing. If we compare the number of frames used for video 1 and video 2, we can see that execution time is directly proportional to the number of frames. For video 1 the proposed system extracted a satisfactory amount of characters, but it didn't work that great on video 2. The reason for this is that it is related to the amplitude of the audio. If the amplitude of the audio is not consistent then the system only considers audio that is above the specific threshold. This happened in case of video 2 in the proposed system, in spite of the normalization performed on the audio. The quality of translation for video 2 is just below average.

## VIII. Conclusion

In this paper we proposed a technique to extract subtitles from video and translate the extracted subtitles. The proposed method uses the audio for identifying the frames that may contain the subtitles from the video. There are many parameters that have influence on the experimental results. The results can be improved by adjusting the parameter values and using NLP methods.

## IX. Acknowledgement

We would like to thank our department for giving us enough time and important suggestions at opportune times. We would also like to thank the people who answered our calls for surveys regarding our results of the project.

## References

**[1]** T. Moteetlal, Dr. V. Sreerama Murthy, "Text Detection From a video using frame extraction and text tracking", 2017 International Conference on Intelligent Sustainable Systems (ICISS)

**[2]** Kirti Kaur Sahota, Lalit Kumar Awasthi, Harsh Kumar Verma, "An empirical enhancement using scale invariant feature transform in text extraction from images", 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT)

**[3]** Shuping Liu. Yantuan Xian, Huafeng Li, Zhengtao Yu, "Text detection in natural scene images using morphological component analysis and laplacian dictionary", IEEE/CAA Journal of Automatica Sinica 2020, Volume: 7, Issue: 1

**[4]** Shashank Shetty, Arun S Devadiga, S. Sibi Chakkaravarthy, K.A. Varun Kumar, "Ote-OCR based Text recognition and extraction from video frames", 2014 IEEE 8th International Conference on Intelligent Systems and Control (ISCO)

**[5]** Liu Yongjiu, Li Chungfang, Shi minyong, Shen changxiang, "Video subtitle location and recognition based on edge features", 2019 6th International Conference on Dependable Systems and Their Applications (DSA)

**[6]** Shu Tian, Xu Cheng Yin, Ya su, Hong wei Hao, "A unified framework for tracking based text detection and recognition from videos", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, Volume: 40, Issue: 3

**[7]** Too Kipyego Boaz, Prabhakar C. J., "A novel approach for detection and localization of caption in video based on pixel pairs", National Conference on Challenges in Research & Technology in the Coming Decades (CRT 2013)

**[8]** Zhanzhan cheng, Jing lu, Baorui zou, Liang Qiao, Yunlu xu, Shiliang pu, Yi niu, Fei wu, Shuigeng zhou, "FREE: A fast and robust end to end video text spotter", IEEE Transactions on Image Processing, 2021, Volume 30

**[9]** Mrs. G. Gayatri Devi, Dr. C. P. Sumathi, "Text Extraction from images using gamma correction method and different text extraction methods - A comparative analysis", International Conference on Information Communication and Embedded Systems (ICICES2014)

**[10]** Hui Fu, xiabi liu, Yunde jia, Hingbin deng, "Gaussian mixture modelling of neighbor characters for multilingual text extraction in images", 2006 International Conference on Image Processing

**[11]** Tianyi Gui, Jun Sun, Satoshi Naoi, Yutaka Katsuyama, Akihiro Minagawa, "A fast caption detection method for low quality video images", 2012 10th IAPR International Workshop on Document Analysis Systems

**[12]** Xiaoqian Liu, Weiqiang Wang, "Robustly extracting captions in videos based on stroke-like edges and spatio-temporal analysis", IEEE Transactions on Multimedia, 2012, Volume: 14, Issue: 2

**[13]** Pannag Sanketi, Huiying She, James M. Coughlan, "Localizing blurry and low resolution text in natural images", 2011 IEEE Workshop on Applications of Computer Vision (WACV)

**[14]** Palaiahnakote Shivkumara, Rushi Padhuman Sreedhar, Trung Quy Phan, Shijian Liu, "Multioriented video scene text detection through bayesian classification and boundary growing", IEEE Transactions on Circuits and Systems for Video Technology, 2012, Volume: 22, Issue: 8

**[15]** Kamrul Hasan Talukder, Tania Mallick, "Connected component based approach for text extraction from color image", 2014 17th International Conference on Computer and Information Technology (ICCIT)

**[16]** A. Thilagavathy, K. Aarthi, A. Chilambuchelvan, "A hybrid approach to extract scene text from videos", 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)

**[17]** Yuming Wang, Naoki Tanaka, "Text string extraction from scene image based on edge feature and morphology", 2008 The Eighth IAPR International Workshop on Document Analysis Systems

**[18]** Liang Wu, Palaiahnakote Shivkumara, Tong Lu, Chew Lim Tan, "A new technique for multi-oriented scene text line detection and tracking in video", IEEE Transactions on Multimedia, 2015, Volume: 17, Issue: 8

**[19]** Xu Zhao, Kai-Hsiang Lin, Yun Fu, Yuxiao Hu, Yuncai Liu, Thomas s. Huang, "Text from corners: a novel approach to detect text and caption in videos", IEEE Transactions on Image Processing, 2011, Volume: 20, Issue: 3

**[20]** Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu, "A hybrid approach to detect and localize texts in natural scene images", IEEE Transactions on Image Processing, 2011, Volume: 20, Issue: 3

**[21]**Pooja, Renu Dhir, "Video text extraction and recognition: a survey", 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)

**[22]** Zhujun Wang, Lei Yang, Xiaoyu Wu, Ying Zhang, "A survey on video caption extraction technology", 2012 Fourth International Conference on Multimedia Information Networking and Security

**[23]** Hongliang Bai, Jun Sun, Satoshi Naoi, Yutaka Katsuyama, Yoshinobu Hotta, Katsuhito Fujimoto, "Video caption duration extraction", 2008 19th International Conference on Pattern Recognition

**[24]** B. Bazeer Ahamed, D. Yuvraj, S. Shanmuga Priya, "Image Denoising WIth Linear and Non-linear Flters", 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)

**[25]** Ashutosh Dehuri, Rupeli Rupanita Dash, Siba Sanyena, Mihir Narayan Mohanty, "A Comparative Analysis of Filtering Techniques on Application in Image Denoising", 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)

**[26]** Shashidhar Ram Joshi; Roshan Koju, "Study and Comparison of Edge Detection Algorithms", 2012 Third Asian Himalayas International Conference on Internet

**[27]** Chinu; Amit Chhabra, "A hybrid approach for color based image edge detection", 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)

**[28]** Ganesan P., G. Sajiv, "A Comprehensive study of Edge Detection for Image Processing Applications", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

**[29]** Pooja, Renu Dhir,"Video text extraction and recognition: a survey", 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNE)