**International Research Journal of Commerce and Law**
Volume 4 Issue 12, December 2017 ISSN: 2349-705X Impact Factor: 4.616
Journal Homepage: http://ijmr.net.in, Email: irjmss@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal

# A Survey Paper On: Use of OCR Technique for Video Indexing

Prof. Kadam Swati Amol,Prof. Bere Sachin Sukhadeo,Prof. Sodal Jeevankumar Anil ,Prof. Rajpure Amol Subhash, Prof. Amrit Priydarshi

Dattakala Shikshan Sanstha "Dattakala Group of Institution" Swami-Chincholi, Daund, Pune, Maharashtra 413130. India.

***Abstract*—**Video OCR is a method that can extraordinarily find the subjects of enthusiasm for video by means of the programmed extraction and perusing of subtitles and explanations. Content in video can give key ordering data. Perceiving such content for pursuit application is basic. Major troublesome issue for character acknowledgment for recordings is debased and de-designed characters, low goals characters or extremely complex foundation. To handle the issue preprocessing on content picture assumes imperative job. The majority of the OCR motors are dealing with the twofold picture so to locate a superior binarization method for picture to get an ideal outcome is critical. Precise binarization process limits the mistake rate of video OCR.

***Keywords*—Video OCR, binarization, video text, video indexing.**

## I. MOTIVATION

Video OCR is a procedure that can enormously find points of enthusiasm for an extensive advanced video chronicle through the programmed extraction and perusing of inscriptions and explanations [9]. Subtitles by and large give imperative hunt data about the video being introduced - the names of individuals and spots or depiction of items. Understanding the substance of recordings requires the shrewd mix of numerous advances: discourse acknowledgment, characteristic dialect handling, seek systems, picture understanding, and so forth. Separating and perusing inscriptions gives extra data to video understanding. Performing Video OCR on record and joining its outcomes with other video understanding procedures will enhance the general comprehension of the video content. Despite the fact, that there is an incredible requirement for coordinated character acknowledgment in content based video libraries. Programmed character division was performed for titles and credits in movie recordings in be that as it may; papers have lacking thought of character acknowledgment. There are comparable research fields which concern character acknowledgment of recordings. In character extraction from the vehicle tag utilizing video pictures is displayed and characters in scene pictures are fragmented and perceived based versatile thresholding. While these outcomes are connected, character acknowledgment for the video introduces its very own challenges due to various states of title character size and complex foundations. In video inscription goals of character is lower; additionally, the foundation unpredictability is more serious than in other research. The principal issue is low goals of the characters. The extent of a picture is restricted by title number of sweep lines characterized in the NTSC standard; a

character of the video subtitle are little to stay away from impediment of intriguing articles, for example, individuals' countenances. In this manner, the goals of characters in the video subtitle are lacking to actualize steady and powerful Video OCR frameworks. Another issue is the presence of complex foundations. Characters superimposed on recordings frequently have tone and splendor like the foundation, making extraction amazingly troublesome. These issues in video OCR have opened a territory for research. Video OCR is a system that can extraordinarily find points of enthusiasm for a substantial advanced video by means of the programmed extraction and perusing of inscriptions and comments. Video OCR process and all the procedure modules required in video OCR are clarified in segment III. Utilizations of video OCR are clarified in area IV. End dependent on relative work is clarified in part V.

## II. PROBLEM STATEMENT

Performing Video OCR on record and consolidating its outcomes with other video investigation strategies will enhance the general comprehension of the video content. Acknowledgment of videotext is a testing issue because of different factors, for example, the nearness of rich, powerful foundations, low goals, shading, and so on. A system is required to process the video pictures to deliver high-goals binarized content pictures that look like printed message and limit the blunder rate while acknowledgment of debased character.

## III. VIDEO OCR

In this area we survey the development of the distinctive segment innovations that establish a conclusion to-end video content acknowledgment framework. The motivation behind this audit is to recognize pushed zones for research, and capacities that might be prepared for combination into a generation situation. Figure 3.1 contains a square graph that demonstrates the handling arrangement in a regular video content preparing framework. Likewise with a conventional archive handling framework, the initial step is to recognize the content, i.e., discover message in the picture and divide the situation of the content area.

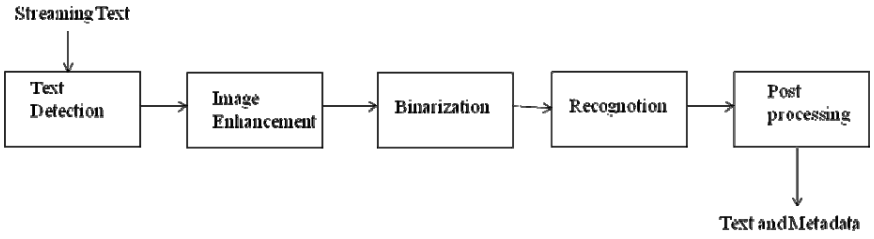


**Figure 3.1 Steps in video OCR**

The essential test being that we have to initially follow the grouping of cases of a video content item over progressive edges of video, even within the sight of components, for example, unpredictable, convulsive movement of a handheld or vehicle mounted camera, nonlinearity of content plane movement, impediments, shadows and other light changes, etc. Along these lines, following the underlying discovery step, the subsequent stage is to improvement. Furthermore, as a major aspect of the picture upgrade step, video content pictures that show viewpoint bending (regularly of content that is inside the scene and not straightforwardly confronting the camera); should be amended with the goal that their appearance is like content that specifically faces the camera. Content amendment includes the estimation of the position and introduction of the content plane in respect to the camera. As the greater part of the OCR's is chipping away at the binarized picture, binarization of improved picture is required. At last, binarized content is perceived by standard OCR and post-preparing is connected on perceived content.

The endeavor of recognizing, dividing and seeing substance apparently appearing in complex pictures and video is apparently particularly described. Regardless, many structure decisions must be taken subject to the general target.

Run of the mill structure decisions are:

- What sort of content events ought to be considered?

In view of its beginning there exist two various types of content in recordings and pictures. Scene content is content that was recorded as a feature of scene, for example, road names, shop names, and content on T-shirts. It for the most part shows up unintentionally and is only here and there proposed. Conversely; the presence of overlay content is cautiously coordinated. Usually an imperative bearer of data and herewith appropriate for ordering and recovery.

- With what text style properties?

Content events can contrast essentially in text dimension, type, style, and shading. Some exploration work has been customized to unmistakable areas with restricted varieties in these traits.

- In what sort of media information?

Should the basic content location, division and acknowledgment approach be picture based (i.e., regarding a video as a lot of autonomous pictures) or should it misuse the way that a similar content line happens in recordings for quite a while and that, thusly, the different examples of a similar content line can be used to accomplish better identification, division and acknowledgment execution.

- How will the yield of the Video OCR framework be utilized? Diverse utilizations have distinctive dimensions of resilience against blunders. For example, if the Video OCR yield is utilized for picture/video ordering dependent on the translated content, pixel mistakes in the confinement and division ventures and additionally acknowledgment blunders can be endured and redressed. Assuming, be that as it may, the yield is utilized for item based video encoding, the framework must limit the mistakes in pixel characterization. The framework in, for precedent, was unequivocally intended to mark every pixel in a video as whether it has a place with content or not.

### *A. Text Detection*

The vast majority of the current video content location techniques have been proposed based on shading, edge, and surface based highlights [2].

1. Color-based approach

Shading based methodologies expect that the video content is made out of a uniform shading. In the methodology by the red shading part is utilized to acquire high differentiation edges among content and foundation. In (Hua et al, 2004) the "uniform shading" hinders inside the high difference video outlines are chosen to accurately extricate content locales.

2. Edge-based approach
3. Edge-based methodologies are additionally viewed as helpful for overlay content location since content districts contain rich edge data. The generally embraced strategy is to apply an edge locator to the video edge and afterward distinguish locales with high edge thickness and quality.
4. Texture-based approach

Surface based methodologies, for example, the remarkable point identification and the wavelet change, have likewise been utilized to identify the content districts.

### *B. Enhancing the Text Image*

Picture Enhancement is used to improve the general idea of an image, with the objective that the result is more fitting than the primary picture for express application. A typical for substance in video is that a given substance region persists over several housings of video feed in the midst of

which the establishment may change. Honestly, when in doubt the establishment vacillates while the substance remains static. The enhanced picture is handled by modifying the differing instances of an explicit substance locale transversely over housings and, for each pixel, picking the shading that identifies with the base power a motivation across over edges. We can endeavor other demand estimations, for instance, the mean, center, and the best anyway the base demand estimation yielded the best picture similar to visual acknowledgment.

### C. *Binarization*

Most OCR motors are dealing with binarized content picture for acknowledgment. There are different diverse systems accessible for it. Finding the better one among them is essential to get precise outcome is imperative. Target of this report is to think about the accessible techniques and consolidate them to get a required outcome.

*Correlation-Based Technique*

A relationship method for binarizing videotext pictures is a standout amongst the most prevalent regular procedures [3]. Four separate channels to display flat strokes, vertical strokes and two corner to corner strokes are utilized to discover different strokes. The channels are prepared by stamping appropriate districts on test preparing information. Instances of stamped preparing districts for flat and vertical channels and relating prepared channels are appeared in figure 3.2 [5]. Each videotext picture is corresponded independently with every one of the four channels, and the connection yields are thresholded to yield four middle paired pictures. The last binarized picture is the association of the four middle of the road binarized pictures.

Problem with this systems are –

i. The most prevalent problem is that background components occasionally exhibit text-stroke-like characteristics and are routinely picked up by the filters. Furthermore, such background components have high correlation scores and cannot be completely eliminated by adjusting the threshold.
ii. Constraining the binarization procedure to use the grayscale image alone, valuable information in the color image is summarily discarded.
iii. The binarized images also lack the smoothness that characterizes the curves and loops in characters such as **c, d, *0,*** etc.

In spite of most of the issues recorded over, the relationship framework attempts to flawlessness of finding the circumstance of substance pixels. In this way, while the last binarized picture may be morphologically insufficient with respect to, it contains most of the substance pixels in the principal picture [4]. In perspective of this discernment another binarization plan can be made that uses the relationship procedure as the underlying advance and after that profits to the shading picture for upgraded execution.

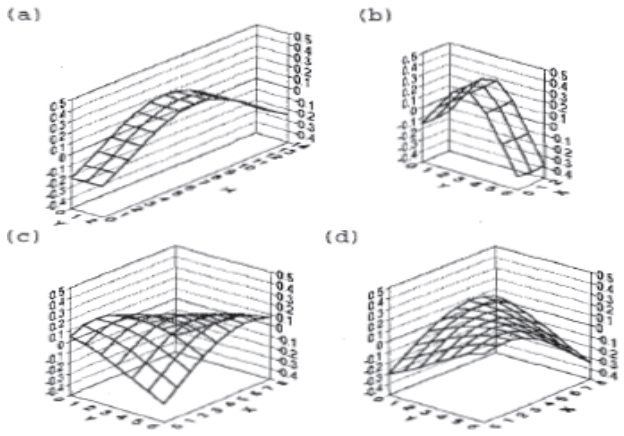


**Figure 3.2 Character extraction filters (a) Vertical. (b) Horizontal. (c) Left diagonal. (d) Right diagonal [8].**

*D. Recognition*

Most OCR engines at last work on binarized pictures of content. At the end of the day, regardless of whether the info picture is grayscale or shading, the acknowledgment framework first believers the picture into a bi-level (dark white) picture before handling the picture through the acknowledgment motor.

*E. Post-processing*

To procure content data for substance based access of video databases, high word acknowledgment rates for video OCR are required. When the content for a casing has been remembered, it is put away to be contrasted with the content removed from neighboring edges for ordering. We apply post-handling, which assesses contrasts between acknowledgment results with words in the lexicon, and chooses a word having the minimum contrasts. Diverse post-handling strategies are utilized for ordering. Video ordering can be utilized in different applications like computerized libraries, advanced News.

## IV. APPLICATIONS OF VIDEO OCR

Performing Video OCR on record and uniting its results with other video understanding frameworks will upgrade the general cognizance of the video content. Content in video gives rich information to content based interest applications. There are diverse applications in which video OCR is used. Figure 4.1 shows utilization of video OCR where message in video is recognized and criticalness of that substance can be resolved which can be used in various applications.

1. Automatic broadcast annotation: creates a structured, searchable view of archives of the broadcast content.
2. Digital media asset management: archives digital media files for efficient media management.
3. Video editing and cataloguing: catalogs video databases on basis of content relevance.
4. Library digitizing: digitizes cover of journals, magazines and various videos using advanced image and video optical character recognition (OCR).
5. Mobile visual sign translation: extracts and translates visual signs or foreign languages for tourist usage, for example, a handhold translator that recognizes and translates Asia signs into English or French.
6. Named matching with face: Name and title information is valuable in matching the people's face with their name.

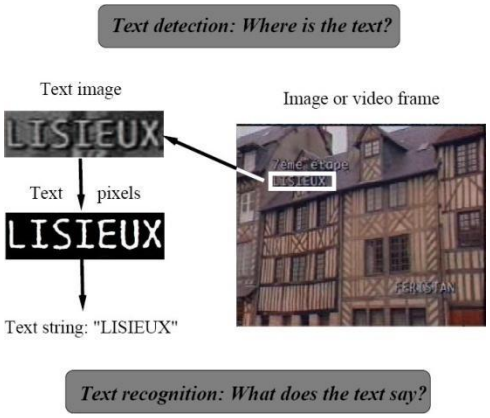


**Figure 4.1 Application of video OCR.**

## V. CONCLUSIONS

Most OCR motors, including at last work on binarized pictures of content. At the end of the day, regardless of whether the info picture is grayscale or shading, the acknowledgment framework first proselytes the picture into a bi-level (dark white) picture before handling the picture through the acknowledgment motor. On account of video content pictures, the traits recorded prior (e.g., low goals, point of view twists, pressure ancient rarities) make the

binarization stage a testing one. Truth be told, test results show that a solitary binarization approach may not be sufficient for managing various types of content in video, and a half breed method that consolidates different methodologies offers generally guarantee.

## REFERANCES

[1] Lyu M., J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Trans. PAMI*, Feb 2005, vol. 15, no. 2, pp. 243–255.

[2] Li h., D. Doermann, 0. Kia, "Automatic Text Detection and Tracking in Digital Video.", *IEEE Trans on Image Processing*, 2000, vol. 9, no.1, pp. 147 -156.

[3] Mori, M. Sawaki, M. Yamato, J.," Robust Character Recognition Using Adaptive Feature Extraction", *Image and Vision Computing New Zealand*, IVCNZ 2008. 26-28 Nov. 2008, pp.1-6.

[4] Natarajan P., B. Elmieh, R. Schwartz, and J. Makhoul, "Videotext OCR using Hidden Markov Models," *Proceedings Sixth International Conference on Document Analysis and Recognition*, 2001, pp. 947 – 951.

[5] Natarajan P., R. Schwartz, and J. Makhoul, "OCR of Low-Resolution Images from Diverse Sources", *SDIUT 2001, Columbia, Maryland*, April 2001, pp. 253-267.

[6] Peng T., P. H. Tian, B. C. Li, "A robust video text extraction method based on text travering line and stroke connectivity," in *Proc.ICSP*, 2008, vol.14, pp. 123-127.

[7] Rui MaI2, Wei Hu2, Qiao Huang3, Jiaxin Wang1, Tao Wang2, Yimin Zhang2, "Robust text stroke extraction from video"**,** *In IEEE International Conference on Multimedia and Expo*, July 2007, pp.1391 – 1394.

[8] Sato T., et al., "Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Caption", *ACM Multimedia Systems Special Issue on Video Libraries*, 1999, pp. 385-395.

[9] Sato T., T. Kanade, E. K. Hughes, and M. A. Smith, "Video OCR for digital news archive," *in Proc. IEEE Workshop Content-Based Access Image Video Database*, 1998, pp. 52-60.

[10] Tsai, Y. C. Chen, C. L. Fang, "A Comprehensive Motion Videotext Detection localization and Extraction Method," *In Proc.CCSP*, 2006, pp.515-519.