A Comprehensive Review of Text Information Extraction Techniques for Images and Videos

Dr. Kapil Kumar

Assistant Professor, School of Computer Applications & Engineering, Vivek University, Bijnor, Uttar Pradesh, India

DOI: https://doi.org/10.51584/IJRIAS.2026.11060276

Received: 02 July 2026; Accepted: 07 July 2026; Published: 17 July 2026

ABSTRACT

Text information present in pictures and video contain helpful data for programmed explanation, ordering, and organizing of pictures. Extraction of this data includes discovery, restriction, following, extraction, improvement, and acknowledgment of the content from a given picture. Nonetheless, varieties of text because of contrasts in size, style, direction, and arrangement, just as low picture difference and complex foundation make the issue of programmed text extraction very testing. While thorough overviews of related issues, for example, face location, record investigation, and picture and video ordering can be discovered, the issue of text data extraction isn't very much studied. Countless methods have been proposed to address this issue, and the motivation behind this paper is to order and audit these calculations, talk about benchmark information and execution assessment, and to call attention to promising bearings for future exploration.

Keywords: Text data extraction, text location, text restriction, text following, text upgrade, OCR

INTRODUCTION

Text Information Extraction (TIE) from images and videos has become an important research area in computer vision and pattern recognition. It enables the automatic detection, localization, extraction, and recognition of text present in visual content for applications such as document analysis, multimedia retrieval, intelligent transportation, and optical character recognition (OCR). However, variations in text size, font, orientation, illumination, and complex backgrounds make text extraction a challenging task. This paper presents an assessment of different TIE approaches, discusses their strengths and limitations, and highlights their applications and future research directions.

1) Text in pictures

An assortment of ways to deal with text data extraction (TIE) from pictures and video have been proposed for explicit applications including page division [5, 6], address block area [7], tag area [3, 8], and content-based picture/video ordering [2, 9]. Notwithstanding such broad examinations, it is as yet difficult to plan a universally useful TIE framework. This is on the grounds that there are so numerous potential wellsprings of variety while extricating text from a concealed or finished foundation, from low-difference or complex pictures, or from pictures having varieties in text dimension, style, shading, direction, and arrangement. These varieties make the issue of programmed TIE incredibly troublesome.

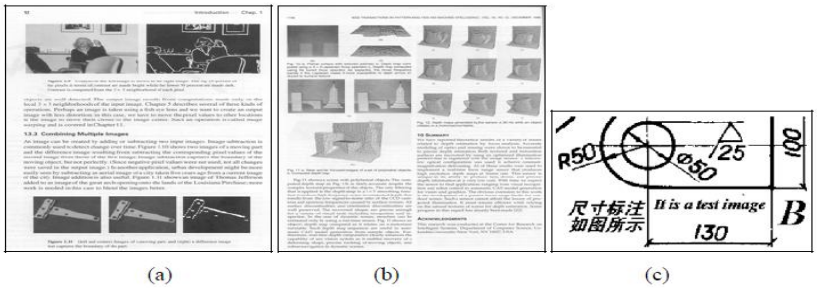

Fig. 1. Grayscale archive pictures: (a) solitary segment text from a book, (b) two-section page from a diary (IEEE Transactions on PAMI), and (c) an electrical drawing (civility of Lu [12]).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Readers may allude to papers on record division/examination [5, 6] for additional instances of report pictures. Despite the fact that pictures obtained by examining book covers, CD fronts, or other multi-shaded records have comparative qualities as the archive pictures (Fig. 2), they can not be straightforwardly managed utilizing a customary archive picture examination method. Likewise, this review recognizes this classification of pictures as multicolor archive pictures from other record pictures.

Fig. 2. Multicolor archive pictures: every content line could conceivably be of a similar shading.

Text in video pictures can be additionally grouped into subtitle text (Fig. 3), which is misleadingly overlaid on the picture, or scene text (Fig. 4), which exists normally in the picture.

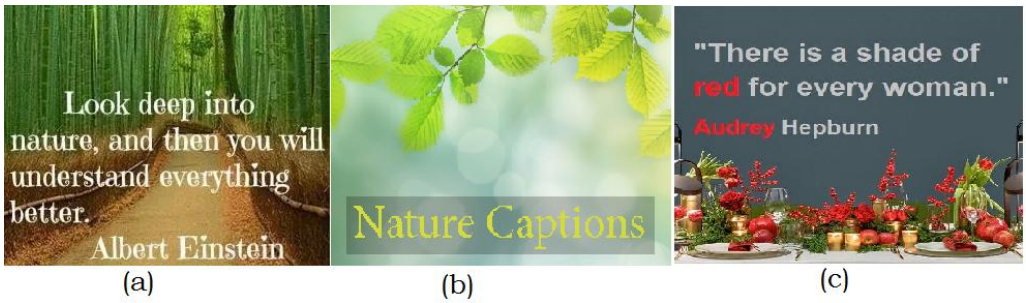

Fig. 3. Pictures with inscription text: (a) shows subtitles overlaid straightforwardly on the foundation. (b) and (c) contain text in casings for better differentiation. (c) Contains a book string that is polychrome.

Fig. 4. Scene text pictures: Images with varieties in slant, viewpoint, obscure, enlightenment, and arrangement.

A few scientists like to utilize the term 'designs text' for scene text, and 'superimposed content' or 'fake content' for inscription text [10, 11]. It is notable that scene text is more hard to identify and next to no work has been done around there. Rather than inscription text, scene text can have any direction and might be misshaped by the viewpoint projection. Besides, it is regularly influenced by varieties in scene and camera boundaries, for example, brightening, center, movement, and so forth before we endeavor to group the different strategies utilized in TIE, it is critical to characterize the ordinarily utilized terms and sum up the attributes of text that can be utilized for TIE calculations. Table 1 shows a rundown of properties that have been used in as of late distributed calculations [13, 15]. Text in pictures can display numerous varieties as for the accompanying properties:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

a. Geometry:

- **Size:** Although the content size can change a ton, presumptions can be made relying upon the application area.
- **Alignment:** The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which can have various perspective distortions. Scene text can be aligned in any direction and can have geometric distortions (Fig. 4).
- **Inter-character distance:** characters in a text line have a uniform distance between them.

b. Color: The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding 'text strings of a single color (monochrome)'. However, video images and other complex color documents can contain 'text strings with more than two colors (polychrome)' for effective visualization, i.e., different colors within one word.

c. Motion: The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.

d. Edge: Most caption and scene texts are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

e. Compression: Many digital images are recorded, transferred, and processed in a compressed format. Thus, a faster TIE system can be achieved if one can extract text without decompression.

Property		**Variants or sub-classes**
Geometry	Size	Regularity in size of text
	Alignment	Horizontal/vertical Straight line with skew (implies vertical direction) Curves 3D perspective distortion
	Inter-character distance	Aggregation of characters with uniform distance
Color		Gray
		Color (monochrome, polychrome)
Motion		Static
		Linear movement
		2D rigid constrained movement
		3D rigid constrained movement
		Free movement
Edge		Strong edges (contrast) at text boundaries
Compression		Un-compressed image
		JPEG, MPEG-compressed image

Table 1: Properties of text in images

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

2) What is Text Information Extraction (TIE)?

The problem of Text Information Extraction needs to be defined more precisely before proceeding further. A TIE system receives an input in the form of a still image or a sequence of images. The images can be in gray scale or color, compressed or un-compressed, and the text in the images may or may not move.

The TIE problem can be divided into the following sub-problems:

(i) Detection,

(ii) Localization,

(iii) Tracking

(iv) Extraction and enhancement

(v) Recognition (OCR) (Fig. 5).

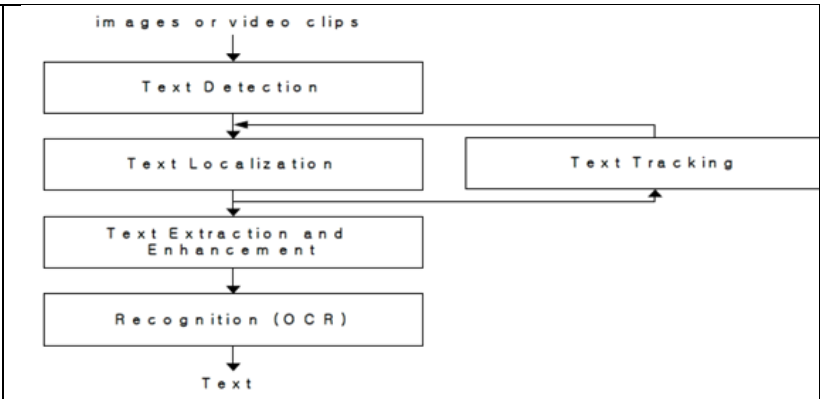

Fig. 5. Architecture of a TIE system.

Text extraction is where the content segments are divided from the foundation. Upgrade of the removed content parts is required in light of the fact that the content district generally has low-goal and is inclined to commotion. From that point, the extricated text pictures can be changed into plain content utilizing OCR innovation.

Text Detection

Text recognition alludes to the assurance of the presence of text in a given casing (typically text identification is utilized for a grouping of pictures). In this stage, since there is no earlier data on whether the info picture contains any content, the presence or non-presence of text in the picture must be resolved. A few methodologies expect that particular sorts of video casing or picture contain text. This is a typical suspicion for filtered pictures (e.g., smaller circle cases or book covers). In any case, on account of video, the quantity of edges containing text is a lot more modest than the quantity of casings without text. The content recognition stage looks to identify the presence of text in a given picture.

Kim [1] chose an edge from shots identified by a scene-change discovery strategy as a competitor containing text. In spite of the fact that Kim's scene-change discovery strategy isn't depicted in detail in his paper, he specifies that low limit esteems are required for scene-change recognition on the grounds that the segment involved by a book locale comparative with the entire picture is generally little. This methodology is exceptionally delicate to scene-change recognition. Text-outline determination is performed at a timespan seconds for inscription text in the distinguished scene outlines. This can be a straightforward and effective answer for video ordering applications that lone need catch phrases from video cuts, as opposed to the whole content.

Text Localization

Text limitation is the way toward deciding the area of text in the picture and creating jumping boxes around the content. As per the highlights used, text confinement techniques can be classified into two sorts: Region-based and surface based.

A) Region-based strategies

District based techniques utilize the properties of the shading or dark scale in a content locale or their disparities with the comparing properties of the foundation. These techniques can be additionally partitioned

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

into two sub-approaches: associated segment (CC)- based and edge-based. These two methodologies work in a base up style; by recognizing sub-structures, for example, CCs or edges, and afterward blending these sub-structures to check jumping boxes for text. Note that a few methodologies utilize a blend of both CC-based and edge-based techniques.

a) CC-based Methods

CC-based strategies go through a base methodology by gathering little segments into progressively bigger parts until all districts are distinguished in the picture. A mathematical examination is expected to consolidate the content parts utilizing the spatial game plan of the segments in order to sift through non-text segments and imprint the limits of the content locales.

b) Edge-based strategies

Among the few literary properties in a picture, edge-put together techniques center with respect to the 'high difference between the content and the foundation'. The edges of the content limit are distinguished and blended, and afterward a few heuristics are utilized to sift through the non-text locales. Normally, an edge channel (e.g., a Canny administrator) is utilized for the edge location, and a smoothing activity or a morphological administrator is utilized for the blending stage.

B) Texture-based methods

Surface based strategies utilize the perception that text in pictures have particular textural properties that recognize them from the foundation. The strategies dependent on Gabor channels, Wavelet, FFT, spatial difference, and so forth can be utilized to distinguish the textural properties of a book district in a picture.

Text Tracking is performed to diminish the preparing time for text restriction and to keep up the respectability of position across neighboring edges. Despite the fact that the exact area of text in a picture can be shown by jumping boxes, the content actually should be sectioned from the foundation to encourage its acknowledgment. This implies that the separated content picture must be changed over to a paired picture and upgraded before it is taken care of into an OCR motor.

Text Extraction in Compressed Domain

In view of the possibility that most advanced pictures and recordings are typically put away, prepared, and communicated in a compacted structure, TIE techniques that straightforwardly work on pictures in MPEG or JPEG packed configurations have as of late been introduced. These techniques just require a limited quantity of disentangling, consequently bringing about a quicker calculation. Additionally, the DCT coefficients and movement vectors in a MPEG video are likewise helpful in content location [14].

Yeo and Liu [20] proposed a strategy for subtitle text restriction in a decreased goal picture that can be effortlessly recreated from packed video. The diminished goal pictures are remade utilizing the DC and AC segments from MPEG recordings. In any case, the picture goal is decreased by a factor of 16 or 64, which brings about a lower limitation rate. Since the content districts are restricted dependent on a huge entomb outline distinction, just unexpected casing succession changes can be recognized. These outcomes in missed looking over titles. The technique is likewise restricted by the suspicion that text just shows up in pre-characterized areas of the edge.

Performance Evaluations

There are a few challenges identified with execution assessment in virtually all exploration zones in PC vision and example acknowledgment (CVPR). The observational assessment of CVPR calculations is a significant undertaking as methods for estimating the capacity of calculations to meet a given arrangement of necessities. Albeit different examinations in CVPR have explored the issue of target execution assessment, there has been next to no emphasis on the issue of TIE in pictures and video. This segment surveys the current assessment strategies utilized for TIE and features a few issues in these assessment techniques.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

The exhibition measure utilized for text location, which is simpler to characterize than for restriction and extraction, is the identification rate, characterized as the proportion between the quantity of distinguished content casings and all the given edges containing text. Estimating the exhibition of text extraction is very troublesome and up to this point there has been no correlation of the diverse extraction techniques. All things being equal, the exhibition is simply construed from the OCR results, as the content extraction execution is firmly identified with the OCR yield.

Execution assessment of text confinement isn't basic. A portion of the issues identified with the assessment of text confinement strategies have been summed up by Antani et al. [4]

(i) **Ground truth information:** Unlike assessing the programmed identification of other occasions, for example, video shot changes, vehicle recognition, or face discovery, the level of accuracy of TIE is hard to characterize. This issue is identified with the development of the ground truth information. The ground truth information for text limitation is typically set apart by limited square shapes that incorporate holes between characters, words, and text lines. Notwithstanding, if a calculation is exceptionally precise and recognizes text at the character level, it wo exclude the above holes and in this way won't have a decent review rate [10].

(ii) **Performance measure:** After deciding the ground truth information, a choice must be made on which measures to use in the coordinating cycle between restricted outcomes and ground truth information. Typically, the review and exactness rates are utilized. Furthermore, a technique is additionally required for contrasting the ground truth information and the calculation yield: pixel-by-pixel, character-by-character, or square shape by-square shape examination.

(iii) **Application reliance:** The point of every content limitation framework can contrast. A few applications necessitate that all the content in the info picture must be found, while others just spotlight on removing significant content. Likewise, the presentation additionally relies upon the loads alloted to bogus alert or bogus excusal.

(iv) **Public information base:** Although numerous scientists look to contrast their techniques and others, there are no area explicit or general far reaching data sets of pictures or recordings containing text. In this way, specialists utilize their own information bases for assessing the presentation of the calculations. Further, since numerous calculations incorporate explicit suppositions and are typically enhanced on a specific information base, it is difficult to lead a complete target correlation.

(v) **Output design:** The outcomes from various restriction calculations might be unique, which additionally makes it hard to think about their exhibitions.

4 Applications

There are various uses of a book data extraction framework, including archive investigation, vehicle tag extraction, specialized paper examination, and article situated information pressure. In the accompanying, we quickly depict a portion of these applications.

(i) **Wearable or versatile PCs:** with the quick improvement of PC equipment innovation, wearable PCs are presently a reality. A TIE framework including a hand-held gadget and camera was introduced as a utilization of a wearable vision framework. Watanabe's [23] interpretation camera can recognize text in a scene picture and make an interpretation of Japanese content into English in the wake of performing character acknowledgment. Haritaoglu [24] likewise exhibited his TIE framework on a hand-held gadget.

(ii) **Content-based video coding or report coding:** The MPEG-4 standard backings object-based encoding. At the point when text areas are portioned from different districts in a picture, this can give higher pressure rates and better picture quality. Feng et al. [25] and Cheng et al. [26] apply versatile

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

vacillating in the wake of portioning a report into a few unique classes. Accordingly, they can accomplish a more excellent delivering of reports containing text, pictures, and designs.

(iii) License/holder plate acknowledgment: There has just been a ton of work done on vehicle tag and compartment plate acknowledgment. In spite of the fact that holder and vehicle tags share numerous qualities with scene text, numerous suppositions have been made in regards to the picture procurement measure (camera and vehicle position and course, brightening, character types, and shading) and mathematical properties of the content. Cui and Huang [9] model the extraction of characters in tags utilizing Markov arbitrary field. Then, Park et al. [44] utilize a learning-based methodology for tag extraction, which is like a surface based content recognition strategy [18, 19]. Kim et al. [30] use angle data to extricate tags. Lee and Kankanhalli [16] apply an associated segment based technique for freight holder confirmation.

(iv) Text-based picture ordering: This includes programmed text-based video organizing techniques utilizing inscription information [4, 14].

(v) Texts in WWW pictures: The extraction of text from WWW pictures can give significant data on the Internet. Zhou and Lopresti [9, 10] utilize a CC-based strategy after shading quantization.

(vi) Video content examination: Extracted text districts or the yield of character acknowledgment can be helpful in sort acknowledgment [1, 28]. The size, position, recurrence, text arrangement, and OCR-ed results would all be able to be utilized for this.

(vii) Industrial robotization: Part distinguishing proof can be cultivated by utilizing the content data on each part [29].

CONCLUSIONS

We have given a complete review of text data extraction in pictures and video. Despite the fact that an enormous number of calculations have been proposed in the writing, no single technique can give good execution in all the applications because of the huge varieties in character text style, size, surface, shading, and so on

There are a few data hotspots for text data extraction in pictures (e.g., shading, surface, movement, shape, math, and so on) It is favorable to combine different data sources to upgrade the exhibition of a book data extraction framework. It is, nonetheless, not satisfactory with regards to how to coordinate the yields of a few methodologies. There is a reasonable requirement for a public area and agent test information base for objective benchmarking. The absence of a public test set makes it hard to analyze the exhibitions of contending calculations, and makes troubles when consolidating a few methodologies.

For inscription text, critical advancement has been made and a few applications, for example, a programmed video ordering framework, have just been introduced. Nonetheless, their content extraction results are wrong for general OCR programming: text upgrade is required for inferior quality video pictures and greater versatility is needed for general cases (e.g., opposite characters, 2D or 3D disfigured characters, polychrome characters, etc). Next to no work has been done on scene text. Scene text can have various attributes from inscription text. For instance, part of a scene text can be impeded or it can have complex development, change in size, textual style, shading, direction, style, arrangement, lighting, and change.

Although many scientists have just explored text limitation, text discovery and following for video pictures is needed for usage in genuine applications (e.g., portable handheld gadgets with a camera and constant ordering frameworks). A book picture structure-investigation, similar to an archive structure examination, is expected to empower a book data extraction framework to be utilized for a picture, including both filtered report pictures and genuine scene pictures through a camcorder. Regardless of the numerous troubles in utilizing TIE frameworks in true applications, the significance and convenience of this field keeps on pulling in much consideration.

REFERENCES

1. H. K. Kim, Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database, Journal of Visual Communication and Image Representation 7 (4) (1996) 336-344.
2. H. J. Zhang, Y. Gong, S. W. Smoliar, and S. Y. Tan, Automatic Parsing of News Video, Proc. of IEEE Conference on Multimedia Computing and Systems, 1994,
3. Y. Cui and Q. Huang, Character Extraction of License Plates from Video, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 502 –507.
4. T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith, Video OCR for Digital News Archive, Proc. of IEEE Workshop on Content based Access of Image and Video Databases, 1998, pp. 52-60.
5. A. K. Jain, and Y. Zhong, Page Segmentation using Texture Analysis, Pattern Recognition, 29 (5) (1996) 743-770.
6. Y. Y. Tang, S. W. Lee, and C. Y. Suen, Automatic Document Processing: A Survey, Pattern Recognition, 29 (12) (1996) 1931-1952.
7. B. Yu, A. K. Jain, and M. Mohiuddin, Address Block Location on Complex Mail Pieces, Proc. of International Conference on Document Analysis and Recognition, 1997, pp. 897-901.
8. D. S. Kim and S. I. Chien, Automatic Car License Plate Extraction using Modified Generalized Symmetry Transform and Image Warping, Proc. of International Symposium on Industrial Electronics, 2001, Vol. 3, pp. 2022-2027.
9. J. C. Shim, C. Dorai, and R. Bolle, Automatic Text Extraction from Video for Content-based Annotation and Retrieval, Proc. of International Conference on Pattern Recognition, Vol. 1, 1998, pp. 618-620.
10. S. Antani, D. Crandall, A. Narasimhamurthy, V. Y. Mariano, and R. Kasturi, Evaluation of Methods for Detection and Localization of Text in Video, Proc. of the IAPR workshop on Document Analysis Systems, Rio de Janeiro, December 2000, pp. 506-514.
11. S. Antani, Reliable Extraction of Text From Video, PhD thesis, Pennsylvania State University, August 2001.
12. Z. Lu, Detection of Text Region from Digital Engineering Drawings, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (1998) 431-439.
13. D. Crandall, S. Antani, and R. Kasturi, Robust Detection of Stylized Text Events in Digital Video, Proceedings of International Conference on Document Analysis and Recognition, 2001, pp. 865-869.
14. Yu Zhong, Hongjiang Zhang, and Anil K. Jain, Automatic Caption Localization in Compressed Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, (4) (2000) 385-392.
15. R. Lienhart and F. Stuber, Automatic Text Recognition In Digital Videos, Proc. of SPIE, 1996, pp. 180-188.
16. J. Ohya, A. Shio, and S. Akamatsu, Recognizing Characters in Scene Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (2) (1994) 214-224.
17. S. H. Park, K. I. Kim, K. Jung, and H. J. Kim, Locating Car License Plates using Neural Networks, IEE Electronics Letters, 35 (17) (1999) 1475-1477.
18. A. K. Jain, and K. Karu, Learning Texture Discrimination Masks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (2) (1996) 195-205.
19. K. Jung, Neural network-based Text Location in Color Images, Pattern Recognition Letters, 22 (14) December (2001) 1503-1515.
20. B.L. Yeo, and B. Liu, Visual Content Highlighting via Automatic Extraction of Embedded Captions on MPEG Compressed Video, Proc. of SPIE, 1996, pp.142-149.
21. J. Zhou, D. Lopresti, and Z. Lei, OCR for World Wide Web Images, Proc. of SPIE on Document Recognition IV, Vol. 3027, 1997, pp. 58-66.
22. J. Zhou, D. Lopresti, and T. Tasdizen, Finding Text in Color Images, Proc. of SPIE on Document Recognition V, 1998, pp. 130-140.
23. Y. Watanabe, Y. Okada, Y. B. Kim, and T. Takeda, Translation Camera, Proc. of International Conference on Pattern Recognition, 1998, Vol. 1, pp. 613-617.
24. Haritaoglu, Scene Text Extraction and Translation for Handheld Devices, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2001, Vol. 2, pp.408-413.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

25. G. Feng, H. Cheng, and C. Bouman, High Quality MRC Document Coding," Proc. of International Conference on Image Processing, Image Quality, Image Capture Systems Conference (PICS), Montreal Canada, 2001.
26. H. Cheng, C. A. Bouman, and J. P. Allebach, Multiscale Document Segmentation, Proc. of IS&T 50th Annual Conference, 1997, pp. 417-425, Cambridge, MA.
27. B. Shahraray and D. C. Gibbon, Automatic Generation of Pictorial Transcripts of Video Programs, Proc. of SPIE, 1995, Vol. 2417.
28. S. Fisher, R. Lienhart, and W. Effelsberg, Automatic Recognition of Film Genres, Proc. of ACM Multimedia'95, 1995, pp. 295-304, San Francisco.
29. Y. K. Ham, M. S. Kang, H. K. Chung, and R. H. Park, Recognition of Raised Characters for Automatic Classification of Rubber Tires, Opt. Eng., 1995, Vol. 34, pp.102-108.
30. S. Kim, D. Kim, Y. Ryu, and G. Kim, A Robust License-Plate Extraction Method under Complex Image Conditions, Proc. of International Conference on Pattern Recognition, 2002, Vol. 3, pp. 216-219.
31. S. Long, X. He, and C. Yao, "Scene Text Detection and Recognition: The Deep Learning Era," International Journal of Computer Vision, vol. 129, no. 1, pp. 161–184, 2021.
32. F. Borisyuk, A. Gordo, and V. Sivakumar, "Rosetta: Large Scale System for Text Detection and Recognition in Images," Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.
33. J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review," IEEE Access, vol. 8, pp. 142642–142668, 2020.
34. A. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What Is Wrong with Scene Text Recognition Model Comparisons? Dataset and Model Analysis," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
35. A. Halder, U. Pal, P. Shivakumara, and M. Blumenstein, "A Comprehensive Review on Text Detection and Recognition in Scene Images," Artificial Intelligence and Applications, vol. 2, 2024.
36. U. Maria, A. Singh, and R. Verma, "Exploring Text Recognition, Segmentation and Detection in Natural Scene Images: A Review," Journal of Digital Systems, vol. 19, no. 2, pp. 66–82, 2024.
37. E. Eli, "A Comprehensive Review of Non-Latin Natural Scene Text Detection and Recognition," Engineering Applications of Artificial Intelligence, vol. 145, 2025.
38. M. B. Gohil and A. A. Desai, "A Review on Text Detection and Recognition in Images and Videos Using Machine Learning and Deep Learning Techniques," International Journal of Engineering Research & Technology (IJERT), vol. 15, no. 1, 2026.
39. A. K. Kashyap, M. Upadhya, V. S. Panwar, et al., "Integrated Framework Utilizing Scene Text Detection and Recognition Techniques for Enhancing Point of Interest Extraction from Name Boards in All Indic Languages," Scientific Reports, vol. 16, 2026.
40. V. Kadha and R. Kumar, "A Deep Learning Survey of Scene Text Detection and Recognition," Computer Vision and Image Understanding, 2026.