International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1344-1352
 https://dx.doi.org/10.37624/IJERT/13.6.2020.1344-1352

# Recent Advancements in Text Detection Methods from Natural Scene Images

**Shiravale S. S.[1], Sannakki S. S.[2] and Rajpurohit V. S.[2]**

*[1] Assistant Professor, Department of Computer Engineering, MMCOE, Pune, India.*
*[2] Professor, Department of Computer Science and Engineering, GIT, Belagavi, India.*

*ORCID[1] : 0000-0002-9804-3978*

**Abstract**

Effect of digitization and globalization has narrowed down the gap of geographical boundaries. Text/language plays an important role in getting connected with people utilizing oral or written communication. Nowadays text data is easily available in the form of multimedia e.g. audio, videos. A technique is needed to understand and interpret the text present in the videos/images which are rich in contents compared to audio data. Text detection and recognition are two main steps of such text-based applications. Text detection from natural scene images is tedious compared to text detection from document images. Various methods are available for text detection from natural scene images. Text detection methods are generally script specific. The main purpose of this paper is to highlight available text detection methods with pros and cons, challenges in the text detection process, evaluation parameters as well as recent achievements. The paper will act as a roadmap for upcoming researchers to select an appropriate text detection method.

**Keywords:** natural scene, text extraction, text detection, image understanding, text localization

## I. INTRODUCTION

Digitization has completely changed the database architecture that supports storage and retrieval of a rapidly generated huge amount of multimedia data. High availability of computing devices make processing and understanding of content present in multimedia data easier and hence content-based applications are gaining popularity. In this era of globalization, text-based applications can provide ease of communication and connectivity between geographically located different regions. A smartphone-based application can be developed that can capture, process and understand text written in one language and translate it into the target language. Text detection and text recognition are two main steps involved in text processing applications. Text detection is the process of locating and extracting the text present in the image. Text recognition is the process of converting text detected in image format to ready to use text (digital) format. Text detection from camera captured natural scene images is complex due to various challenges [1]. There are several factors affecting text detection process and can be categorized into i) Font perspective: different font size, font style, multi-coloured, multi orientation and multilingual etc. ii) Quality perspective: Complex background, Poor quality due to climatic conditions, deformed image etc. iii) Device perspective: poor resolution, perspective distortion, image with shadow effect and uneven light conditions etc. Few of the text detecting challenges from natural scene images with Devanagari text are mentioned in Fig.1.

The challenges present in the text detection attract many researchers to contribute in this area. Accuracy of any text recognition technique relies on the correctness of the text detection technique. Over the period, remarkable success is achieved in the text detection techniques by many researchers. Paradigm of text detection methods is shifting rapidly from the usage of fundamental features to modern and more intelligent algorithms. The main objective of this paper is to present a brief review of existing state of art methods for text detection and recent advancements in this decade.

## II. EXISTING TEXT DETECTION METHODS

In the literature survey, a lot of work is found on Latin scripts (e.g. English) text detection and recognition. Significant development is observed in Asian languages like Chinese and Indic scripts. Researchers have tried various methods on different language datasets and achieved remarkable success [1, 2]. Conventionally, text detection methods are categorized into sliding window-based and connected component-based methods [3]. Considering the rapid developments and success of state of art methods, in this paper text detection methods are categorized into i) low-level feature-based, ii) high-level feature-based and iii) text region verification methods. Most of the traditional text detection methods are based on basic features like edge, colour, texture etc. These features are used very frequently and easy for the implementation. More semantically rich features like Stoke Width Transform (SWT), Maximally Stable Extremal Regions (MSER), Histogram of Gradient (HoG) etc. can be derived from these basic features and used more efficiently. Non-linear nature of these features limits the scope of intensive improvement in the performance of text detection methods. Thus, more efficient and intelligent machine learning algorithms are adapted by the researchers to handle misdetection of text.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1344-1352
 https://dx.doi.org/10.37624/IJERT/13.6.2020.1344-1352


**Fig. 1.** Challenges in text detection: (a) Poor quality text due to various climatic conditions (b) Perspective distortion (c) Multilingual text (d) Multi-coloured text (e) Curved text (f) Artistic font (g) Images with shadow (h) Uneven light condition(i) Deformed image.

## III. LOW-LEVEL FEATURE-BASED TECHNIQUES

Very fundamental features like edge, colour, texture and connected component are considered as low-level features. These are very popular, frequently used and easy to implement features. The basic concept is to extract these features from an input image and detect text regions using some heuristic rules. For example, in the connected component analysis if height/width is greater than or in between some threshold value then identify that connected component as a text. Basic text detection methods fall into four major categories.

### III.I Edge-based Techniques

Edge-based methods are faster and capable to handle different font size, style, orientation and colour [2]. A character is represented as a combination of lines/edge profiles or curves. Thus usage of edges is advantageous in text detection process. Edge detection operator, mostly 'canny' [3] (suitable to extract horizontal, vertical edges and curve lines) slides over the input image to derive edge map. Morphological operations are applied to the generated edge map to locate text regions [4, 5]. In [6], Sobel filter with adaptive thresholding is applied on the image pyramid for Farsi text localization. The method supports the extraction of text edges from low and high contrast images due to adaptive thresholding. Edge-based methods are simple and produce high recall but they are inefficient to handle complex background and shadow effects. The impact of strong background edges can be minimized by applying pre-processing techniques e.g. Gaussian filtering [7] or eliminating non-significant edges using different features like edge density, strength and orientation variance [8]. Edge-based techniques are script independent and can be efficiently used for multi-script and multi-oriented text localization [9].

### III.II Colour-based techniques

Generally, font colour and background colour holds high contrast whereas colour similarity is found within a word or sentence. This intra-text colour homogeneity and inter-text (font and background) colour contrast properties are beneficial for text detection. The input image can be segmented into text and background based on colour contrast whereas letters with the same colour can be clustered as a word. Camera captured RGB image can be converted into other colour spaces like L*a*b*, HSV, YCbCr etc. for further processing. K-means clustering is a popular choice for grouping similar coloured text, the work based on colour-based clustering in L*a*b colour space is mentioned in [9, 6]. RGB colour-based foreground and background segmentation method is explained in [10]. Performances of text detection in L*a*b*, HSV and YCbCr colour spaces are compared in [11] and experimentally proved that text detection in YCbCr colour space is a better choice for text detection. Problems like shadows or uneven illumination will be tackled by processing input image in each channel separately. For example, in YCbCr model, different colour shades are formed due to the combination of a proportion of 'Y' Luminance component in Chroma red (Cr) or Chroma blue (Cb) component (e.g. High

luminance value with Cr produces pink colour). Shadow effect can be handled by processing only in two channels i.e. Cb and Cr. Colour-based methods are capable to handle most of the text detection challenges like complex background, perspective distortion, shadow effect, uneven light conditioning, poor quality images etc. but multi-coloured and low contrast text handling are the limitations.

### III.III Texture-based Techniques

It is one of the conventional text detection method based on the distinctness of text and background textures. For example, text regions are richer with text lines and strokes. Coarseness texture feature of text line can be a distinctive feature for text and background separation [12]. Some commonly found texture measures are wavelet, entropy, Local Binary Pattern (LBP), Discrete Cosine Transform (DCT) [7] etc. A robust multi–lingual text detection method based on wavelet entropy is proposed in [13]. An input image is decomposed into blocks of variable size and blocks are exposed with several filters to extract texture features based on wavelet entropy. LBP is a spatial structure useful for texture classification, scale-invariant LBP derived from edge profiles is mentioned in [7]. Researchers have contributed in texture-based methods and achieved success to handle multi-oriented text [14], different font size [13] and low-resolution text [15]. These methods are capable to handle poor quality, degraded, complex background and noisy images [16]. Texture-based methods are slow, complex and computationally expensive maybe therefore less advancement is observed in this decade.

### III.IV Connected Components

Text is treated as a collection of connected components (CC). Sometimes script characteristic provides add-on benefits in CC-based text detection process. In Indic scripts like Devanagari and Bangla where alphabets/characters are connected with the header line (‘Shirorekha’) [17, 18] to form a word, CC-based technique provides a complementary role for automatic word formation. Otherwise, connected components having minimum distance can be merged to form a word [10]. Connected components are derived by using edge profiles, colour clustering, morphological operations [19] and Maximally Stable Extremal Region (MSER) techniques instead of deriving directly from the binary image. Generation of the number of CCs varies with the complexity of input images, pre-processing steps applied and features used for extraction. For example, CC derived by applying an edge detector directly on the input image may be in huge numbers compared edge detector applied on the pre-processed image (e.g. after smoothing). MSER based computation produces a comparatively lesser number of CCs and hence became a popular choice of researchers [11, 20]. Though the generation of CC is computationally less expensive some additional computation either based on heuristic rules or classifiers is required for the analysis of CC to decide whether it is true text component or not [21]. A connected component is directly passed as an input for the text recognition module is the major advantage of this technique [22].

Simplicity is the key characteristic of low-level feature-based techniques. Considering the complex nature of scene images, these features are not capable to achieve great success beyond the limit. Text detection methods based on low-level features may produce several irrelevant regions known as non-text regions as an output and thus affects the performance. Features like edge and colour can be used more efficiently by deriving some other stable features from them.

## IV. HIGH-LEVEL FEATURE-BASED TECHNIQUES

Stronger and efficient features can be obtained from low-level features e.g. shapes and stroke width can be derived from edge maps, MSER can be obtained from colour features. Detection based on these features produce more stable results as a relatively lesser number of non-text regions get produced by the techniques. These features are more suitable for training the classifiers for the identification of text and non-text regions. Few frequently used features are mentioned in the following section.

### IV.I Stroke Width Transform (SWT)

Distance between two parallel edges of a character is considered as a stroke width. Uniformity of text stroke plays an important role in the text detection process. Computational steps for stroke width were proposed by Epshtein et al. in [23]. Stroke width is derived from the target pixel map combined with geometric reasoning called a Stroke width Transform (SWT) map. SWT can be obtained either by gradients derived from the edge map [24] or distance transformation-based approach [25]. In a gradient-based approach, the distance between two pixels that are on the opposite gradient direction of each other is considered as a stroke width. In distance transformation approach, Euclidian distance from object pixel from edge or boundary (i.e. nearest nonzero pixel) is computed. Skeleton like feature [4] or histogram analysis of distance transform [26] can be used to derive stroke width from distance transform. Mean and variance of SWT map are very useful features which are script independent and easy for implementation. On the other side, SWT of natural scene images is sensitive to the complex background and uneven illumination [27]. Additional computation is required to identify whether two parallel edges belong to the same character or not and to check whether the light text is on a dark background or vice versa [28]. In [29], a novel seed-based variant of SWT is proposed to address the issue of false edges or missing edges that affects SWT map. The technique has proven robust to broken/blur edges of the text. Researchers are contributing and trying to enhance the efficiency of SWT.

### IV.II Maximally Stable Extremal Region( MSER)

MSER belongs to the connected component-based method; recently it has become the mainstream method for the text detection due to its wider usage [28]. MSER is a stable connected component of some intensity level set of the image over a range of threshold. MSER components are brighter or darker from its surrounding and hence suitable for separation of font and background in natural scene images. MSER can be obtained from edges [30] or colour components [20]. It is possible to improve the sustainability of MSER for handling low resolution, low contrast images [31] and uneven illumination [32]. MSER is invariant to scale and affine intensity changes. But it needs an additional mechanism for

text candidate construction and pruning of repeated components. Precision of the MSER technique can be improved by using an efficient pruning technique [31], heuristic rules [25] or classifiers [20]. Generation of repeated extremal regions can be avoided by providing some prior information like stroke colour information based MSER extraction as mentioned in [33].

### IV.III Geometric Features

These features are derived from region-based properties. Height, Width, area and aspect ratio are some frequently used features [9, 10, 31, 34, 35] for text detection. These features are computationally simple and commonly used for training the classifiers to classify text and non-text regions. Area: Number of pixels. Aspect ratio: Ratio of width to height. Extent: Ratio of pixels in the region to the total pixels in the bounding box. Eccentricity: Ratio of the distance between the foci of the ellipse and its major axis length. These simple features help in text region identification process. For example, regions with a height greater than width cannot be a text region. Likewise, curvature, smoothness [9] of the boundary can be used for text detection. Due to the complex nature of scene images, text detection methods cannot rely on geometric features alone. Novel symmetric features like Mutual Direction Symmetry (MDS), Mutual Magnitude Symmetry (MMS) and Gradient Vector Symmetry (GVS) had proposed by Anhar et al. for handling curve and multi-oriented font [34]. Other features like gradient features: Multi-script identification [36], shape features: Invariant to scale and rotation transformation and Histogram of Gradient (HoG) descriptor: Useful for extraction of text contours [6, 7, 35, 37] are some other supportive features for text and non-text regions identification.

Stability and efficiency of high-level features made them popular among the researchers. Nowadays, these methods are well accepted as a part of mainstream methods in the field of text detection.

## V. TEXT REGION VERIFICATION TECHNIQUES

Generally, natural scene images have complex backgrounds like trees, building structures, rough surfaces etc. Sometimes text detection method identifies false regions i.e. non-text regions as text regions. Accuracy of text detection degrades if large numbers of false (non-text) regions are detected by the technique. One of the simple solutions is to filter out non-text regions by applying either heuristic rules or classification algorithms. According to the literature survey, machine learning algorithms play an important role in text and non-text region classification and produce more efficient results. These algorithms are capable to handle the non-linear nature of the features. The basic idea is to extract the features from detected regions, train the classifiers with extracted features and classify the regions into text and non-text regions. Few most popular and frequently used classifiers are mentioned in the following sections.

### V.I Support Vector Machine (SVM)

SVM is the topmost choice of the researchers for text and non-text classification due to its experimentally proven classification strength and performance. It is a statistical based classifier, competent to handle nonlinear feature set. Efficiency of the SVM lies in inner product operations used for mapping non-linear space into high dimensional space. Higher dimension space produces a linear perception of non-linear feature space thus provides ease for data partitioning by placing hyperplanes [38]. Different kernel functions available for mapping and their experimental results are discussed in [39]. SVM can be trained with most promising features like SWT, MSER, geometric features, HoG etc. for region classification. Researchers have trained multiple SVM with different features at multiple levels for getting more accurate results. Three-level SVM classification is performed by Jonathan et al. [37]. Each SVM is trained using Fourier descriptor, pseudo-Zernike moments and Polar descriptor respectively for character detection. Final decision of text validation relies on global SVM trained using HoG feature. Anna et al. [35] have proposed simple similarity score filtering as a first layer classifier and HoG feature trained SVM as a second layer classifier. As per the literature survey, SVM is a part of many of the state of art methods.

### V.II Artificial Neural Network (ANN)

Artificial Neural network is a soft computing approach where data processing is carried out by set of artificial neurons. Various ANN algorithms are available for the text and non-text classification but most popular are multi-layer perceptron (MLP) [11, 40, 41] with back-propagation and recently deep neural networks. Performance of ANN algorithms relies on architectural parameters such as selection of activation function, number of hidden layers, number of neurons in each layer etc. For example, MLP with a single hidden layer with 20 nodes [11] and MLP with the single hidden layer having nodes four times of the number of input features [41] is used by researchers for text region classification. Multilayer perceptron algorithm may have a single hidden layer whereas deep neural networks may have hundreds of layers. Chen et al. [42] have compared performances of MLP and SVM classifiers trained with the same features. Experimental results show SVM performs better compared to MLP. Pan et al. [41] have proposed the novel conditional random field (CRF) model for filtering text and non-text regions. Detailed experimentation is carried out with various combinations of CRF, SVM and MLP. Here, CRF with MLP produces good results compared to SVM. So it can be concluded that the performance of the classifier is dependent on the underlying architecture, measuring attributes and features used for training. ANN and SVM both are efficient for text and non-text classification but ANN is time-consuming whereas SVM is computationally complex.

### V.III Deep Neural Network

Recently deep neural networks are fetching great success in object detection and recognition problems. Deep convolutional neural networks (CNN) with hundreds of hidden layer that are trained using diversified and huge data fetch highly accurate text detection results. Automatic feature extraction and efficiency are key virtues of CNN. Pre-trained NN models and highest accuracy invites researchers to work in this area. CNN outperforms manual feature extracted based shallow neural networks due to automatic adjustment of

**Table 1.** Various methods used for text detection with strengths and issues

| Methods | Strengths | Issues |
|---|---|---|
| Edges-based | Simple, faster, high recall | Sensitive to complex background |
| Colour-based | Handles many detection challenges e.g. Shadow effects | Not suitable for low contrast font |
| Texture-based | Suitable for poor resolution and low contrast text | Time-consuming, Computationally complex |
| CC-based | Simple, directly passed for text recognition | Mechanism for CC analysis |
| SWT | Distinct property of text and efficient | Complexity increases with complex content |
| MSER | Efficient, stable with good precision | Redundant component generation, poor recall |
| Geometric features | Computationally inexpensive and efficient | Mostly act as supplementary features |
| HoG | Strong ability to describe text contour | Sensitive to complex background |
| SVM | Best for text / non-text classification | Computationally complex |
| Naïve Bayesian | Simple, faster | Sensitive to noise |
| NN/Deep NN | Highly accurate results for text detection | Lengthy, not suitable for real-time applications |

tuning parameters at the time of the training network. It is proven that CNN extracted features are more efficient than manually extracted features for text and non-text discrimination. But the performance of CNN can be improved by boosting it with handcrafted features [43, 44]. Cascaded method for text line classification is presented in [45], if the entropy of text line is above some threshold then considered as true text line otherwise passed as an input to CNN model for further decision. Two CNN models are trained separately (one for text extraction and other for text verification) with MSER features and text/ non-text regions respectively by Zhang et al. [33]. Performance of overall text detection is based on the fusion of two CNN results. Many of the researchers have used different flavors of existing CNN algorithms and even contributed by developing their architectures. J. Ma et al. [46] have proposed Rotation Region Proposal Networks (RRPN) to handle multi-oriented text. Orientation features are estimated by adding Rotation Region-of-Interest pooling layer.

CNN performs promisingly in text and non-text discrimination problems [47] and also in text recognition [48, 49]. Availability of task-specific huge dataset for training and high computing processing power is a major hurdle for deep neural networks.

### V.IV Other Classifiers

Researchers have tried other classifiers like Naïve Bayesian classifier and Adaboost. Bayesian classifier is originated from statistical stream and based on probability theory. Probability of pixel belongs to text and non-text classes are computed and pixel gets classified based on the highest probability value [50]. Bayesian classifiers are simple, faster [51] but sensitive with noisy data. Adaboost cascading classifier, a series of weak classifiers trained with single feature to build a strong classifier is proposed in [52]. In [21], Classification and Regression Tree (CART) is used as a weak classifier. Advantage of cascading classifiers is that OCR module for text recognition can be embedded as one of the classifiers in the series. Though Adaboost produces good results they are computationally lengthy. Aneesh Sain et al. [53] have proposed an efficient multi-oriented and curve text detection method based on skeleton feature and HMM classifier. Probability score of HMM determines whether skeleton feature is of text or non-text region.

Classification algorithms are complex but highly efficient for text, non-text region verification compared to heuristic rule-based techniques. Availability of equally proportionate training set of both the classes (text and non-text) is essential to avoid class biasness problem.

Researchers have achieved great success in handing various text detection challenges but artistic font, multi-lingual text etc. are still challenging. Remarkable improvement is observed in other scripts/languages like Farsi, Devanagari, Bangla but work has to be extended for other scripts as well. This paper highlights various and most successful as well as commonly used text detection methods. Different features with their strengths and limitations are summarized in Table 1.

## VI. PERFORMANCE METRICS

As per the survey [54], precision and recall are the most commonly used accuracy measures for the text detection methods. Precision is the ratio of ‘c’ number of true text regions detected by the technique to the ‘t’ total number of regions identified by the technique. Precision decreases if more number of false (non-text) regions detected by the technique. Recall is the ratio of the number of true text regions detected to the ‘gt’ actual number of text regions present in the ground truth. Precision and recall are illustrated in Fig. 2.

It is not possible to measure the performance of text detection technique by separately considering precision and recall. The overall accuracy of the detection technique is the harmonic mean of precision and recall and represented as f-measure. In simple words, the success of the text detection technique lies in the detection of more number of true text regions and lesser number of non-text regions.

If the classification algorithm is a part of the text detection technique then evaluation protocols may modify accordingly. Accuracy of classification algorithms is measured using the confusion matrix. Thus, accuracy is derived from the

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1344-1352
 https://dx.doi.org/10.37624/IJERT/13.6.2020.1344-1352

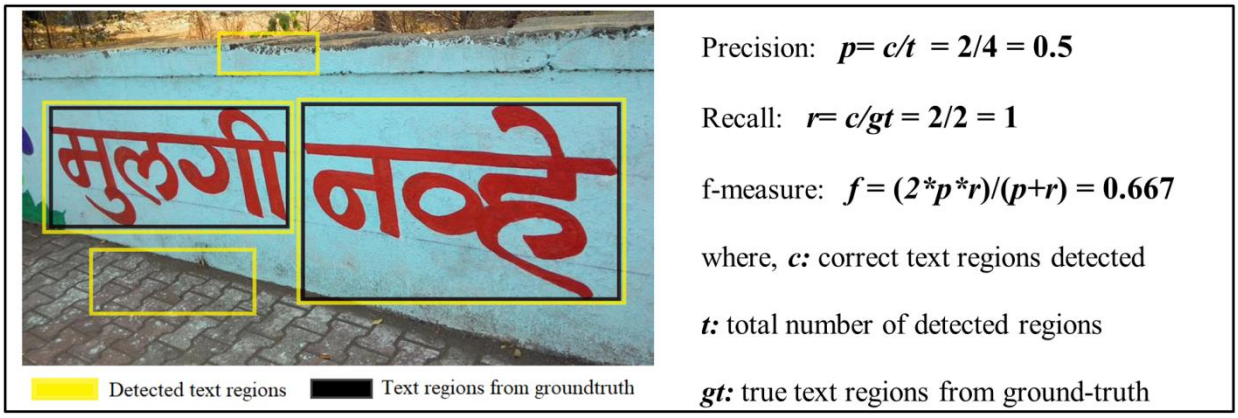


**Fig. 2.** Illustration of precision and recall

confusion matrix in terms of precision and recall. Precision is calculated as p= tp/tp+fp and recall is computed as r=tp/gt. Where, true positive 'tp' is correctly classified text regions, false-positive 'fp' is non-text regions misclassified as text regions.

text detection are considered for the comparisons. Their performances on ICDAR 2011 and ICDAR 2013 datasets are mentioned in Table 2.

It is clear from the Table 2, the success of the text detection algorithm lies in wise usage of many features at a time.

**Table 2.** Performances of scene text detection methods on ICDAR 2011 and ICDAR 2013 datasets

(P: precision, R: recall, F: f-measure).

| Author | Year | ICDAR 2011 | | | ICDAR 2013 | | | Features used |
|---|---|---|---|---|---|---|---|---|
| | | P | R | F | P | R | F | |
| Soni et al. [51] | 2019 | 83 | 67 | 74 | 84 | 68 | 75 | MSER, TAS, Naïve Bayesian |
| Wang et al.[44] | 2018 | 85 | 70 | 77 | 87 | 68 | 76 | CRF, MSER, CNN |
| Wei et al.[47] | 2018 | 86.9 | 80.9 | 83.7 | 87.3 | 81.1 | 84.3 | CNN |
| J. Ma et al. [46] | 2018 | - | - | - | 90.22 | 71.89 | 80.02 | Rotation Region Proposal Networks |
| Tang and Wu [43] | 2018 | 90.6 | 84.7 | **87.6** | 91.1 | 86.1 | **88.5** | Stroke, Color, Geometric, HOG, CNN |
| Zhang et al. [33] | 2018 | - | - | - | 89.07 | 82.89 | 85.87 | MSER, SWT, CNN |
| Aneeshan et.al [53] | 2018 | - | - | - | 87.4 | 74.2 | 80.26 | CC, Skeleton, HMM |
| Liao et al. [48] | 2017 | 88 | 82 | 85 | 88 | 83 | 85 | Deep neural network |
| Leibin and Chu [28] | 2017 | - | - | - | 82 | 67 | 74 | SWT, MSER |
| Zheng et al.[45] | 2017 | 89.9 | 77.92 | 83.48 | 89.5 | 77.63 | 83.14 | Extremal regions, text-line entropy, CNN |
| H. Cho et al. [3] | 2016 | - | - | - | 86.26 | 78.45 | 82.17 | Canny, ER, SWT, Geometric, Adaboost |
| Zhang et.al [56] | 2015 | 84 | 76 | 80 | 88 | 74 | 80 | Symmetric Features, CNN |
| Neumana and Matas [55] | 2015 | - | - | - | 81.8 | 72.4 | 77.1 | MSER, Stroke, SVM |
| Anna et al. [35] | 2015 | 83 | 68 | 75 | 82 | 71 | 76 | Similarity Score, Stroke, HOG, SVM |
| Su and Xu [29] | 2015 | 78 | 69 | 73 | - | - | - | Seed-based SWT |
| Anhar et al. [34] | 2014 | 83 | 71 | 77 | - | - | - | Edge, Symmetric properties, SIFT |
| Wang et al. [10] | 2013 | 73 | 67 | 70 | - | - | - | RGB colour based CC |
| Yin et al. [31] | 2013 | 86.29 | 68.26 | 76.22 | - | - | - | MSER, Geometric features, SWT |
| Koo and Kim [11], | 2013 | 81.44 | 68.68 | 74.52 | - | - | - | Colour based MSER, MLP |
| Haojin Yang et al. [7] | 2012 | 83.37 | 80.37 | **82** | - | - | - | Entropy, SWT, LBP, HoG, SVM |
| X. Yin et al. [57] | 2012 | 81.53 | 62.22 | 70.58 | - | - | - | MSER, SWT, Geometric, Adaboost |

Benchmarking and publically available dataset of natural scene images are mentioned in [1, 2]. Most of these datasets are in the English language and with own evolution protocols. Researchers may contribute by developing benchmarking datasets in other scripts. An attempt is made here to explore performances of various features that are discussed in the paper. Recent developments proposed in this decade for scene

Recent trend in text detection is deep neural networks due to the highest accurate results. Selection of text detection technique depends on the nature of problem e.g. deep neural networks are not suitable for real-time applications. This highlights the significance of handcrafted features like edge, colour, SWT, MSER etc. Low-level feature-based methods are simple, produce good recall but may have poor precision.

SWT, MSER etc. are very promising features for text localization and produces more accurate results but may produce low recall. Efficiency of text detection algorithms can be improved by eliminating non-text regions. Heuristic rules or classifiers are used for text and non-text verification. Heuristic rule-based methods are simple but require manual adjustment of tuning parameters e.g. threshold which is erroneous and time-consuming. Classification algorithms are powerful and produce highly accurate results. Classification algorithms may restrict the ability of generalization due to training it with task-specific dataset. The success of text detection method lies in the selection of features used for text localization and classifier used for text, non-text verification. As per the literature survey, SVM is highly recommended classifier for text and non-text region classification. Recently, convolutional neural networks are gaining popularity due to highly accurate results and its automatic feature extraction ability for text detection.

## VII. CONCLUSION

Text Detection is greatly rooted domain wherein a good amount of research is being done. There are many languages used across the globe and detection as well as recognition is a critical task. Each script possesses its own unique and distinguishable characteristics making it difficult to fit a single technique to detect a text of different scripts having a structural difference. It is therefore necessary to detect and recognize multi-script text from any language that can be a new avenue opened up for the researchers. This research paper focuses on handing text detection challenges, most promising text detection features, classifiers and evaluation measures. Techniques based on high abstraction features like SWT and MSER are more stable compared to fundamental feature-based techniques. Intelligent machine learning algorithms like SVM and deep neural networks are efficient and used to achieve great success. Upcoming researchers are recommended to consider various factors like script specific characteristics, complexity of input images, availability of dataset and processing power while selecting a method for text detection.

## REFERENCES

[1] Zhu, Y., Yao, C., Bai, X., 2016, "Scene text detection and recognition: Recent advances and future trends", Frontiers of Computer Science, 2016, 10(1), pp. 19–36.

[2] Zhang, H, Zhao, K, Song, K-Y., Guo, J., 2013, "Text extraction from natural scene image: A survey", Neurocomputing, 122, pp.310-323.

[3] Cho, H., Sung, M., Jun, B., 2016 "Canny text detector: Fast and robust scene text localization algorithm", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3566–3573.

[4] Roy, Chowdhury, Bhattacharya, and Parui, 2011, "Text Detection of Two Major Indian Scripts in Natural Scene Images", Springer link. CBDAR, pp. 73-78.

[5] Li, X., Li, J., Gao, Q. and Yu, X., 2019, "Uyghur Text Detection in Natural Scene Images", 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, pp. 1542-1547.

[6] Darab, M., Rahmati, M., 2012, "A Hybrid Approach to Localize Farsi Text in Natural Scene Images", Procedia Computer Science, 13, pp.154-164.

[7] Yang, H., Quehl, B., Sack, H., 2012, "A framework for improved video text detection and recognition", Springer Science+Business Media New York, pp. 217-245.

[8] Liu, X., Samarabandu, J., 2006, "Multiscale edge-based text extraction from complex images" IEEE International Conference on Multimedia and Expo, IEEE, pp. 1721–1724.

[9] Kasar, T, Ramakrishnan, A. G., 2012, "Multi-script and Multi-oriented Text Localization from Scene Images", Camera-Based Document Analysis and Recognition, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 7139:1-14.

[10] Wang, X., Song, Y., Zhang, Y, 2013, "Natural scene text detection with multi-channel connected component segmentation", 12th International Conference on Document Analysis and Recognition, pp. 1375-1379.

[11] Koo, H., and Kim, D. H., 2013, "Scene Text Detection via Connected Component Clustering and Nontext filtering", IEEE Transaction on Image processing, 22(6), pp. 2296-2305.

[12] Huang, X., Ma, H., 2010, "Automatic Detection and Localization of Natural Scene Text in Video", IEEE 2010 International Conference on Pattern Recognition, pp. 3216-3219.

[13] Manjunath, Aradhya, V. N., Pavithra, M. S., and Naveena, C., 2012, "A Robust Multilingual Text Detection Approach Based on Transforms and Wavelet Entropy", ScienceDirect, pp. 232-237.

[14] Shivakumara, P., Dutta, A., Tan, C. L., Pal, U., 2013, "Multi-oriented scene text detection in video based on wavelet and angle projection boundary growing", Multimed Tools Appl @ Springer Science+Business Media New York, pp. 515-539.

[15] Angadi, S. A., Kodabagi, M.M, 2010, "Text Region Extraction from Low Resolution Natural Scene Images using Texture Features" IEEE 2nd International Advance Computing Conference, pp. 121-128.

[16] Kim, K., Jung, K., Kim, J.H., 2003, "Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), pp. 1631-1639.

[17] Jayadevan, R., Kolhe, S.R., Patil, P.M., Pal, U., 2011, "Offline Recognition of Devanagari Script: A Survey", IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, 41(6), pp.782-796.

[18] Bhattacharya, U., Parui, S. K., Mondal, S., 2009, "Devanagari and Bangla Text Extraction from Natural Scene Images" IEEE, Int. Conf. on Document Analysis and Recognition, pp. 171-175.

[19] Uddin, M.S., Sultana, M., Rahman, T., Busra, U.S., 2012, "Extraction of Texts from a Scene Image using Morphology Based Approach", IEEE/OSA/IAPR International Conference on Infonnatics, Electronics & Vision, pp. 876-880.

[20] Sun, L., Huo, Q., Jia, W., Chen, K., 2015, "A robust approach for text detection from natural scene images", Pattern Recognition, 48(9), pp. 2906-2920.

[21] Chang, R.C., 2011, "Intelligent Text Detection and Extraction from Natural Scene Images", Nano, Information Technology and Reliability (NASNIT), 15th North-East Asia Symposium, IEEE, pp. 23 – 28.

[22] Zhang, J., Cheng, R., Wang, K., Zhao, H., 2013, "Research on the text detection and extraction from complex images", Fourth International Conference on Emerging Intelligent Data and Web Technologies, pp. 708-713.

[23] Epshtein, B., Ofek, E., Wexler, Y., 2010, "Detecting text in natural scenes with stroke width transform", in: Proceedings of the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, pp. 2963–2970.

[24] Jameson, J. and Abdullah, S. N. H. S., 2014, "Extraction of arbitrary text in natural scene image based on stroke width transform", 14th International Conference on Intelligent Systems Design and Applications, pp. 124-128.

[25] Gomez, L., and Karatzas, D., 2013, "Multi-script Text Extraction from Natural Scenes", 12th International Conference on Document Analysis and Recognition, pp. 1520-5363.

[26] Wei, L., Neullens, S., Breier, M., Bosling, M., Pretz, T., Merhof, D., 2014, "Text recognition for information retrieval in images of printed circuit boards", Industrial Electronics Society, IECON 2014 - 40th Annual Conference of the IEEE, pp. 3487-3493.

[27] Oh, I., Lee, J.S., 2016, "Smooth Stroke Width Transform for Text Detection", Artificial Intelligence: Methodology, Systems, and Applications. Lecture Notes in Computer Science, Springer, Cham, 9883, pp.183-191.

[28] Guan, L. and Chu, J., 2017, "Natural scene text detection based on SWT, MSER and candidate classification", 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, pp. 26-30.

[29] Su, F. and Xu, H., 2015, "Robust seed-based stroke width transform for text detection in natural images", 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, pp. 916-920.

[30] Chen, H., Tsai, S.S., Schroth, G., Chen, D. M., Grzeszczuk, R., Girod, B., 2011, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions", in: Proceedings of the Eighteenth IEEE International Conference on Im- age Processing, Brussels, Belgium, pp. 2609–2612.

[31] Yin, X.C., Yin, X., Huang, K., Hao, H.W., 2013, "Robust Text Detection in Natural Scene Images", IEEE transactions on pattern analysis and machine intelligence, pp. 970-983.

[32] Sun, L., Huo, Q., Jia, W., Chen, K., 2014, "Robust text detection in natural scene images by generalized color-enhanced contrasting extremal region and neural networks" in: Proceedings of the ICPR.

[33] Zhang, X., Gao, X., Tian, C., 2018, "Text detection in natural scene images based on color prior guided MSER", Neurocomputing, 307, pp.61-71.

[34] Risnumawan, A., Shivakumara, P., Chan, C.S., Tan, C.L., 2014, "A robust arbitrary text detection system for natural scene images", Expert Systems with Applications, 41(18), pp. 8027-8048.

[35] Zhu, A., Wang, G., Dong, Y., 2015, "Detecting natural scenes text via auto image partition, two-stage grouping and two-layer classification", Pattern Recognition Letters, 67(2), pp. 153-162.

[36] Mandal, R., Roy, P.P., Pal, U., Blumenstein, M., 2015, "Multi-lingual date field extraction for automatic document retrieval by machine", Information Sciences, 314, pp.277-292.

[37] Fabrizio, J., Marcotegui, B., Cord, M., 2013, "Text detection in street level images", Pattern Anal Applic, Springer-Verlag London, pp. 519-533.

[38] Jiawei, H., Kamber, M., "Data Mining: Concepts and Techniques", Elsevier Publishers

[39] Chen, G.Y., Bhattacharya, P., 2006, "Function Dot Product Kernels for Support Vector Machine", 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, pp. 614-617.

[40] Sun, L., Huo Q., Jia, W., Chen, K., 2015, "A robust approach for text detection from natural scene images", Pattern Recognition, 48(9), pp. 2906-2920.

[41] Pan, Y.F., Hou, X., Liu, C.L., 2011, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE Transaction on Image Processing, 20(3), pp. 800-813.

[42] Chen, D., Odobez, J.M., Thiran, J.P., 2004, "A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods", Signal Processing: Image Communication, 19(3), pp. 205-217.

[43] Tang, Y., Wu, X., 2018, "Scene Text Detection Using Super pixel-Based Stroke Feature Transform and Deep Learning Based Region Classification, IEEE Transactions on Multimedia, 20(9), pp. 2276-2288.

[44] Wang, Y., Shi, C., Xiao, B., Wang, C., Qi, C., 2018, "CRF based text detection for natural scene images using convolutional neural network and context information", Neurocomputing, 295, pp. 46-58.

[45] Zheng, Y., Li, Q., Liu, J., Liu, H., Li, G., Zhang, S., 2017, "A cascaded method for text detection in natural

scene images", Neurocomputing, 238, pp. 307-315.

[46] Ma, J., 2018, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals", IEEE Transactions on Multimedia, 20(11), pp. 3111-3122.

[47] Wei, Y., Shen, W., Zeng, D., Ye, L., Zhang, Z., 2018, "Multi-oriented text detection from natural scene images based on a CNN and pruning non-adjacent graph edges", Signal Processing: Image Communication, 64, pp. 89-98.

[48] Liao, M., Shi, B., Bai, X., Wang, X, Liu, W., 2017, "TextBoxes: A fast text detector with a single deep neural network", in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, pp. 4161–4167.

[49] Ansari, G. J., Shah, J.H., Yasmin, M., Sharif, M., Fernandes, S.L., 2018, "A novel machine learning approach for scene text extraction", Future Generation Computer Systems, 87, pp.328-340.

[50] Shivakumara, P., Sreedhar, R., Phan, T.Q., Lu, S., Tan, C.L., 2012, "Multioriented Video Scene Text Detection Through Bayesian Classification and Boundary Growing" IEEE Transaction on Circuits and System for video Technology, 22(8), pp.1227-1235.

[51] Soni, R., Kumar, B., Chand, S., 2019, "Text detection and localization in natural scene images based on text awareness score", Appl Intell , 49, pp.1376–1405.

[52] Yan, J., Gao, X., 2014, "Detection and recognition of text superimposed in images base on layered method", Neurocomputing, 134, pp. 3-14.

[53] Sain, A., Bhunia, A.K., Roy, P.P., Pal, U., 2018, "Multi-oriented text detection and verification in video frames and scene images", Neurocomputing, 275, pp.1531-1549.

[54] Lucas, S. M., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R., 2003, Icdar2003 robust reading competitions, in Proceedings of the Seventh International Conference on Document Analysis and Recognition, 2, pp. 682–687.

[55] Neumann, L., Matas, J., 2015, "Efficient scene text localization and recognition with local character refinement", in: International Conference on Document Analysis and Recognition, pp. 746–750.

[56] Zhang, Z., Shen, W., Yao, C., Bai, X., 2015, "Symmetry-based text line detection in natural scenes", in: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 2558–2567.

[57] Yin, X., Yin, X.C., Hao, H.W., Iqbal, K., 2012, "Effective text localization in natural scene images with mser, geometry based grouping and adaboost", International Conference on Pattern Recognition (ICPR), pp.725- 728.