Jurnal
Teknologi

Full Paper

A REVIEW ON TEXT DETECTION TECHNIQUES

Sana Ali[a*], Khalid Iqbal[a] , Saira Khan[a] , Qazi Zohaib Aqil[b], Rehan Tariq[a]

[a]COMSATS Institute Of Information Technology, Attock, Pakistan
[b]Karachi School for Bussiness and Leadership, Karachi, Pakistan

Article history
Received
25 October 2015
Received in revised form
14 December 2015
Accepted
9 February 2016

*Corresponding author
sanazohaib@yahoo.com

Abstract

Text detection in image is an important field. Reading text in image is challenging because of the variations in images. Text detection in images is useful for many navigational purposes e.g. text on google API's and traffic panels etc. This paper analyzes the work done on text detection by many researchers and critically evaluates the techniques designed for text detection and states the limitation of each approach. We have integrated the work of many researchers for getting a brief over view of multiple available techniques and their strengths and limitations are also discussed to give readers a clear picture. The major dataset discussed in all these papers are ICDAR 2003, 2005, 2011, 2013 and SVT(street view text).

Keywords: Text Detection; Images; Videos; ICDAR;

1.0 INTRODUCTION

Nowadays, text detection in images is very important. Many researchers have invented quite a lot of techniques because in our surrounding many types of images billboards, traffic panels, sign boards, building facades images etc. are present and all of these have text embedded in them. But images could be of many variation low illumination, complex back ground, images in big data, images of scan document containing typed or hand written text, images taken in moving environment, images containing text in multi-orientation etc. so different techniques are used for different kind of images. In this paper we have discussed following techniques: Iqbal et al. [1] proposed adaptive classifier threshold method using posterior probabilities. Yin et al.[8] also used posterior probabilities and designed an algorithm for text detection. Gonzalez et al. [14, 15] proposed adaptive classifier threshold method. Jadderberg et al. [2, 17] used neural networks for text detection. Gao et al. [11] invented a visual saliency model for text detection in images. Huang [4] used stroke feature transformation for text localization. Yao et al. [5] and Neuman et al. [7] proposed the use of strokelet for representation of text in image. Ye[6] prosed a combination of appearance and consensus based approach. Minetto et al. [8] proposed snooper text in multi-scale fashion. Ganesh et al.[10] worked ontetx detection in big data. Xiao[13] and Shivakumara[30] used soble edge map for obtaining feature for text detection. Iqbal et al. [16] used Bayes network score and K2 algorithm. Yao et al. [18] proposed a unified framework for text detection in multi oriented text detection. Milevskiy et al. [19] worked on joint energy based detection and classification of text and introduced new hierarchal model MDL. Barlas et al. [20] worked on recognition of hand written or typed text. Yin et al. [22] used Maximally stable extremal region and geometric features for text detection. Hanif et al. [23] used constrained Adaboost and neural networks for text detection. Ye et al. [24] used GLVQ(generalized learning vector quantization) and SVM. Toyamma et al. [27] used eye gaze and the OCR(optical character recognition) technology for reading text. Wang et al. [29] used SLAM(simultaneous localization and mapping) for reading text in videos.

In this paper, we have summarized multiple techniques for text detection in images along with the pros and corns of every technique further we have also discussed the datasets used for each technique. This paper will give a brief summary for the readers and help them to choose the best and also a new better technique could be developed to by looking at all the discussed techniques.

The rest of paper is divided as: section 2 covers the literature review having sub section 2.1 for images and 2.2 for videos. Section 3 covers the critical evaluation and conclusion is done in section 4.

2.0 LITERATURE REVIEW

Natural scene images can be divided into static and moving (video) images. In this section, we reviewed literature on a number of text detection techniques in natural scene images as well as in videos to highlight the strengths and limitations of existing text detection techniques.

2.1 Text Detection in Natural Scene Images

In static natural image, Iqbal et al.[1] suggested an adaptive classifier threshold to detect text in images. Adaptive classifier threshold is based on the geometric mean and standard deviation of posterior probabilities of a MSER-based candidate characters using Bayesian network scores. A candidate character is discarded if posterior probability is below the adaptive classifier threshold value. This method is evaluated on an ICDAR 2013 and achieved a significant competitive performance with a comparison of recently published algorithms. However, the proposed method requires improvements in performance by testing other dataset. In addition, this method only detects on a given dataset rather than any natural image.

Jaderburg et al.[2] does text detection using deep features. They divided the work of text detection into two tasks: spotting the word region and recognizing the words. A Convolutional Neural Network classifier is used for Case sensitive, Case insensitive and bigram classification and generated the saliency maps for these. The convolutional structure of CNN goes through the whole image once. The mining technique goes through the images produces word level and character level annotations. The output is divided as follow: a character/background classifier, a case insensitive classifier, a case insensitive classifier and a bigram classifier. These classifications adds to the efficiency of this work. He used the ICDAR (2003, 2005, 2011, 2013) Robust Reading data set and street view text data set. The system have its limitations due to the saliency maps, because these may sometime give bad resolution and wrong result.

Yin et al.[3] presented a robust method for detecting texts in natural images and designed an effective pruning algorithm using the strategy of minimizing regularized variations to extract Maximally Stable Extremal Regions (MSERs). Further, they proposed distance matric learning algorithm and then used single link clustering algorithm to group candidate character into text region. Finally, non-text candidates were eliminated using Adaboost classifier based on posterior probabilities. The method is evaluated on ICDAR 2011 Robust Reading Competition dataset and also on multilingual (Chinese and English). But the proposed system doesn't work for all natural scene images containing any orientation of text.

Huang et al.[4] introduced a new technique for text localization in natural image using Stroke Feature Transform and Text Covariance Descriptors. At first they used a low level filter "Stroke Feature Transformation (SFT) for removing background pixels. Afterwards, the candidates are obtained using color homogeneity and consistency between stroke widths. And for component detection, two maps are generated using SFT: stroke width map and a stroke color map. Then two Text covariance descriptors (TCD) were used for text region and text line classification. So, by using the SFT and two TCDs, their system's performance improved many folds. The dataset of ICDAR 2005 and ICDAR 2011 are used. The proposed low-level SFT filter, leads to high recall, and the effectiveness of the two-level TCDs, leads to high precision. But this method can only predict text in horizontal manner not vertical.

Yao et al.[5] have proposed a multi-scale representation of text in images using strokelets. Over the conventional approaches strokelets have following four advantages: Usability, robustness, generality and expressivity called the URGE properties. They designed an efficient algorithm for text recognition. First of all the character identification is done by seeking maxima in hough maps. They used the discriminative clustering algorithm designed by Singh et al. [35], certain changes were made according to requirement in this algorithm. The random forest classifier is used for classification of components. The data sets of SVT and ICDAR 2013 are used to evaluate the results. The proposed algorithm consistently outperformed the existing state-of-the art approaches. But the result could show variations over other datasets. This system has limitations due to random forest's behavior.

Ye et al.[6] have invented a new approach on text detection in image by combing both the appearance and consensus component representation into a discriminative model. SVM (support vector machine) is used for dictionary classifiers. This discriminative model performs two tasks: one to differentiate text/non-text and determine component grouping. In text detection, candidate components are built on MSERs. They proposed a definition of "text patterns" using a sequence of classifiers. A multi-class SVM training algorithm is used to train the dictionary classifier. On the given set of samples, they then calculated the classifier responses of the components consensus features and the components. Then hypothesis on the text are made until it become negative. They evaluated the work on following two datasets: ICDAR 2011 scene text dataset and the Street View Text (SVT) dataset. Their proposed approach has improved the precision, recall rate and f-measure than the previous approaches. The problem in the method was, if the distance between candidate characters is as large as text height then the object may be missed to be recognized.

Neumann et al. [7] have worked on text detection in scene image with Oriented Stroke Detection and presented an unconstrained end to end localization and recognition method. This method detects the character as an image region, which contain strokes of specific orientation and relative position. The detected stroke induce the set of rectangles to be classified. The strokes are modelled as responses to oriented filters in the gradient projection scale space and the relative stroke position is modelled by subsampling the responses into a fixed-sized matrix. Character recognition is done by recognizing a known stroke pattern with a trained classifier. Then, to detect words in the image and recognize their content, an optimal sequence was found in each text line. The result was evaluated on the ICDAR 2011 dataset. The advantage of the method is that the no. of rectangles are reduced many fold due classification of set of strokes. And the limitation is due to existence of an ambiguity that a sub-region of a character might be another character.

Minetto et al. [8] have proposed a Snooper text in multi scale fashion. First of all, it locates the candidate characters on the images by image segmentation and shape based binary classification. The segmentation algorithm used, was developed by Fabrizio et al. [9] to define local foreground and back ground. Using the SVM classifier, the character filtering is done. Next, the candidate characters are grouped by simple geometric criteria to form either candidate words or candidate text lines. The grouping module in iTown does not work well when come across wide spaces so to overcome this issue, this module was run twice. One more advantage of the multi-scale approach is that it makes the segmentation algorithm insensitive to character texture like high frequency details. They used four datasets for testing the result: ITW (iTown project's image collection), SVT (Street View Text), EPS (used by Epshtein), ICDAR (half of the 2005). And a comparison was done between Snooper text and TessBack, TesseRact, result showed that TessBack is better one. The advantage of multi-scale fashion is, it makes segmentation algorithm insensitive to character texture like high frequency details. The method has limitations: The grouping module doesn't work well for wide spaces. And it can't detect tilted or vertical text.

Ganesh et al. [10] have worked on text detection in the images of big data. They have taken google API to get many street view images those acted as Big Data in their work. After obtaining the images filters are applied to remove noise. First of all averaging filter is applied, then median filter is applied and finally adaptive filter is applied to eliminate the low frequency regions and noise. Next, image processing is done in two steps: first, color based partition method Second, the classifiers were applied to detect whether the above partitions contain text in them or not. Then by using Hough transformation method, text line grouping method. The strength of the technique is dealing of big data that's very crucial. But it has its limitation that result may vary for other big data rather than dataset used in this technique.

Gao et al. [11] have worked on detection of images in natural scene and invented a Hierarchical Visual Saliency Model. As sometimes the region containing the character is salient, and is detected by saliency based method. So, they have worked on knowing how much these region effect the detection of characters and proposed a new method and compared with Itti et al.'s model [12]. First of all the salient regions are obtained and image is filtered. In second step: evaluation of local saliency region inside global salient region is done. And that is the final map required. The scenery image data base is used which contain 3018 images. The result showed that hierarchal saliency method performed better than Itti et al. [12]. A problem exist in this method is the threshold method and hierarchal clustering method for cropping saliency region, not always give good result.

Xiao et al. [13] have proposed a method for text detection in images of complex back ground. The method uses density-based information and rectangle window in the residual edge image. The input was of low illumination images with complicated background in the RGB format so at first the conversion to Ycbcr is done. Then Tone mapping function is used to enhance the image. After this, the vertical edge of this image is being extracted, soble was used to obtain edge density. In order to detect candidate regions, they estimated edge density across the edge image by applying a Gaussian kernel on it. Next, the complicated background curve and noise is removed by morphological opening, closing. Then text localization is done using, density-based information and rectangle window in the residual edge image. Finally, for text segmentation, they used the work of Nomura et al. [36]. The dataset of ICDAR 2005 is used, result showed that the proposed method worked well even in low illumination. The system's strength is its detection in low illumination while has a drawback that for long curves, the system may scan twice for removing noise.

González et al. [14] have proposed a text reading algorithm in natural image. Their process is based on two steps one: image is analyzed to detect text using geometric feature. Two: the recognition is done. Following are the contribution of their work: 1) a combination of adaptive thresholding method and Maximally Stable Extremal Region, for segmentation, 2) a brief study on different features of text is done to discriminate between text and non-text 3) a restoration stage is proposed for rejected characters, 4) a method for detection of single character using K nearest neighbor and use of DP for misspelled words. The dataset of ICDAR 2003, 2005and 2011 are. Result showed that the proposed system performed better and scored first in precision (mean the no. of false positive was smallest). But the limitation is: method does not work for multi oriented text.

González et al. [15] have worked on text detection in traffic signs boards From Street-Level Imagery Using Visual Appearance. They have used the text detection and recognition technique (with modification) by the same author used in [14].They applied blue and white segmentation on the image. Then the classification is done using Naive Byes and SVM accordingly. The technique is based on color segmentation and BOVM (bag of visual words) approach. And this method is applied to only those areas detected by blue, white masks. Then the feature extraction is done using haris-laplace salient point detector. They have also compared the following descriptors for their work: SIFT, C-SIFT, Hue, Histogram. The strengthening features were: Character recognition was enhanced to detect symbols, invention of a new technique for blue region description. The limitation of the method was: it application only to areas detected blue and white.

Iqbal et al. [16] have worked on text localization in scene images and they have used K2 algorithm and Bayesian network score. At first they have detected the candidate region using MSER algorithm and filtration is done by constraints. As a second step, textual regions are being constructed. Due to intense pixel values every candidate binary region has its own features. So, for each feature, K2 algorithm is used to obtain Bayesian Network score. Finally, classification of true candidates is done by using a text classifier. They have tested their data on ICDAR robust reading competition data set of 2013. So, their method of Bayes Text resulted ranked on number *4th* out of 10 among the recently published results. But the limitation of the method is: dependence on Yin et al [20].

Jaderberg et al. [17] have worked on text detection in natural scenes and using Synthetic Data and Artificial Neural Networks. This method considers the whole image for recognition of word that makes it different. A synthetic engine is used to produce data and generate a word as whole. Then the neural network is use to train that data. They have used three models: dictionary encoding, bag-of-N-grams encoding and sequence encoding. The data set of ICDAR 2003, ICDAR 2013 and Street View Text are used to generate data through synthetic engine. Their method improved greatly over the standard datasets because of simple fast machinery and cost reduction of data acquisition. And the strength of the system is it takes images as whole. There exists a limitation due to the fact that synthetic engine may produce large amount of data that can create difficulties.

Yao et al. [18] have worked on multi-oriented text detection and proposed a new unified framework for it. Text detection and recognition is done all together and same features are utilized. The system works well on multiple orientation of text. A new search based dictionary approach is invented to eliminate errors caused by resembling symbol. Firstly, the candidates are generated via clustering and SWT, then recognition is performed using similar classification schema. For this they used randomized tree and the component level classification, it is done by using Random Forest classifier. Then result is forwarded to a dictionary to correct errors. And finally the detected text is the output of the given framework. The dataset of ICDAR2003, 2005 and 2011 are used. Furthermore dataset like MSRA TD500 is also used because yet there was no state of art dataset for multi-oriented text. Proposed system performed quite well on all these data sets. The strength of the system is multi-oriented text detection and the limitation of system is due to random forest classifier's behavior.

Milevskiy et al. [19] have worked on Joint Energy-based Detection and Classification of Multilingual Text Lines. A new hierarchal model MDL is introduced for classification and detection of the images taken by any hand held device, the energy of the image can be optimized by fusion moves. The method segmented images into multiple classes of language and text line by looking at geometric errors. The original image was detected as text blobs using edge based technique then Adaboost classifier is applied and finally text line is detected using energy based algorithm. The combined text detection and classification was based on energy minimization by BCD (Block Coordinate decent). The dataset was consisted of 500 images taken by mobile cameras. This model combined the geometric error cost which adds to the strength of the method and the narrowness of the method is due to Adaboost classifier.

Barlas at al. [20] have worked on recognizing hand written or typed text in heterogeneous document and developed analysis system for text recognition. Well they have presented LITIS's proposed system for segmentation (into 8 multiple classes). In their work they presented connected component based strategy for identification and segmentation of the text and heterogeneous documents are dealt with it, which make it different and credits to learning based approach. Figure 1 shows the working flow of the system.

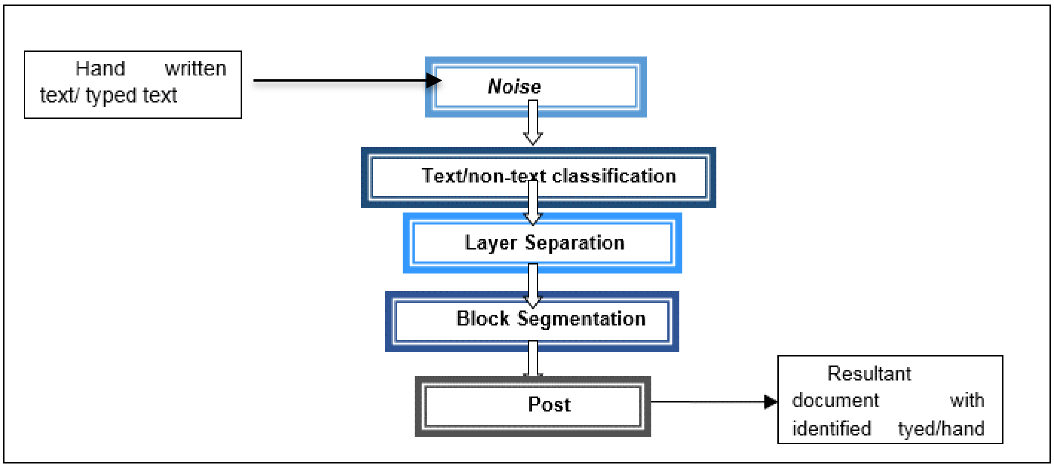
Hand written text/ typed text
Noise
Text/non-text classification
Layer Separation
Block Segmentation
Post
Resultant document with identified tyed/hand

Figure 1 A typed and handwritten text block segmentation system for heterogeneous and complex documents

LITIS's segmentation constitutes many detectors. Firstly the noise is being removed, then text and non-text are separated using learning based approach. These characters are then input to MPL classifier (trained on 2000 doc images in MOURDOR's dataset). Then finally layer separation is done to differentiate between textual cc into typed or hand written text using codebook, afterwards the block segmentation is done by using Run length smoothing algorithm and method presented by Thomas Breuel et al. [21]. The method was tested on MOURDOR's dataset. The strength of the method is the detection of hand written text along with typed text in heterogeneous documents. The drawback of system is the text on graphical portion can't be detected.

Yin et al. [22] presented a new approach for localization of text in image using MSER and geometric features and adaboost for classification. The proposed approach works in a way that first the candidate letters are extracted using MSER and based on geometric features non letters are removed. After this candidate groups are classified using AdaBoost classifier, candidate's features are extracted. The dataset of the algorithm was of ICDAR 2011, result were better than the published algorithms till 2011. Contrary to traditional approaches of sliding window and CC, the new method is better. Its detection is limited to only horizontal text.

Hanif at al. [23] have worked on text detection and localization in grey scale image using Constrained AdaBoost Algorithm and presented a text detector that is based on cascade of boost ensemble. The neural networks are utilized to obtain the automatic rules for localization. Feature extraction is done using rectangular text segment. LRT (likelihood ratio test) was picked up as weak classifier. Afterwards, classical Adaboost algorithm is used for removing classification errors ignored by weak classifiers, so a strong classifier is constructed using the hypothesis based on features extracted. Using the attentional cascade many non-object feature are removed at early stages. The cc are being projected to grey level image and an edge map is then computed by Canny filter. Then they used a Multilayer Perceptron (MLP) for validation. Finally, the verified components are clustered to form a single rectangle over the text word. The dataset of ICDAR 2003 is used, the result showed that the false alarm rate is comparatively high for CAdaboost as compared to classical Adaboost but the standard deviation is low. The strength of method is it works well for different font, style and complex background but limited to only grey scale images.

Ye et al. [24] have worked on text detection and text restoration in natural image and presented a robust method for text detection, recognition and restoration.

First of all, GLVQ (generalized learning Vector quantization) algorithm is used to segment pixels to locate characters. Then the differentiation between text and non-text is done and wavelet co efficient is used along with color variance to detect text patterns. Then the SVM classifier is used. Text lined characters are then binarized and input to OCR for reduction of false alarm. After this, the restoration of text is done by a process based on plane to plane homography. It is independent of camera parameters and applied where deformation on the text is detecting. Figure 2 shows the working of system.

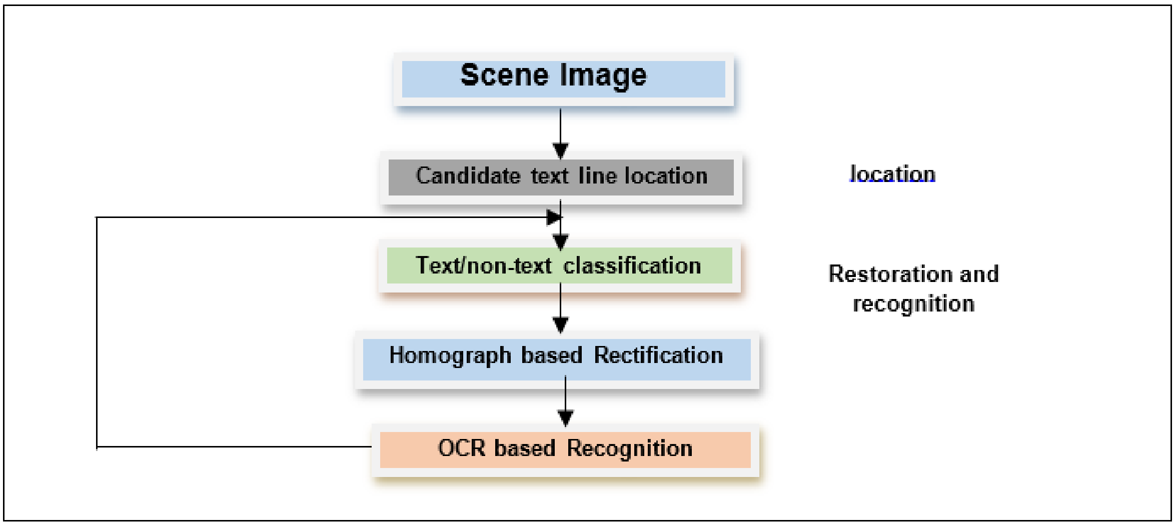
Scene Image
Candidate text line location
Text/non-text classification
Homograph based Rectification
OCR based Recognition
location
Restoration and recognition

Figure 2 Text detection and restoration in natural scene images.

The dataset of 1500 images captured, with different background, illuminations etc. are used. Their method showed robust performance. The OCR reduces the false alarm that adds to strength of the system but the limitation of the system is due to the inconsistent behavior of GLVQ algorithm.

Shahab at al. [25] done the analysis of multiple techniques for text detection in natural scene images based on ICDAR 2011's challenge 2. According to them there three sub part in text reading in an image: 1) text location, 2) character detection and 3) word detection. In 2011 a competition was held to test the algorithm developed to test these two parts text localization and word recognition. So they enhanced the dataset of 2003 by adding images in it. For text localization the method of Wolf et al. [26] was used. Following 9 methods participated 1) Yi's Method, 2) Kim's Method, 3) Text Hunter, 4) KAIST AIPR System, 6) LIP6-Retin, 7) "TDM IACAS, 8) TH-TextLoc, TH-OCR System, 9) ECNU-CCG Method. Kim's method attained the first position in text localization and in text detection there were three participant among them TH-OCR performed best. The result showed that still there was a lot of room for improvement. Only few methods took part in word recognition but still there is room for lot of improvement.

Toyama et al. [27] have worked on text recognition using eye gaze. And develop a system to translate the text from French to English. The OCR (optical-character-recognition) technology is used and human eye gaze was taken as input to get efficient result. They have extended the work of [28]. The feedback is given through Head mount display. Not only the translation but the navigational help is also provided. They utilized SMIs Eye Tracking Glasses (ETG)2 for gaze input. These gaze algorithms are used: Gaze Repetitive Leap (GRL), Gaze Scan. Then the OCR is activated, the text region is extracted from the, end and start fixation point of gesture. Then the navigation is provided on HDM screen. The advantage of the method is its navigational help for foreigners but limitation is: the result showed only the basic effectiveness of their algorithm.

2.2 Text Detection in Videos

Wang et al. [29] focused on videos to detect text by using simultaneous localization and mapping (SLAM) to extract planar scene surfaces "tiles". And all the sensor data is input using LCM (Lightweight Communications and Marshaling) then the system maintain the sensor ring's motion using incremental LIDAR scan-matching and a local map is built consisting of line segment. Tiles are then projected onto the camera which generated the observation and then a fronto-parallel view was observed through a homography transformation. These observations are then considered for text detection and decoding. They used DCT and MSER to produce detection regions for character which then provided to Tesseract. Then a clustering process grouped down the characters word candidate. The output came up to be as sequence of characters with each comprising a small number of candidates. They themselves designed a matrix to evaluate the test result of their work. The strength of the method is text detection in images captured in motion but there is a limitation as well: When the distance is 1.5m while taking observation then the decoder may not work.

Shivakumara et al. [30] has worked basically on arbitrary oriented text detection in videos using Gradient Vector Flow and Grouping based Methods. Soble edge map of the image is used to identify dominant text pixel. Further edge components in Soble edge map corresponding to dominant pixels are extracted and they called them Text Candidates (TC). Then to overcome arbitrary orientation they proposed two stages grouping for TC, at first stage grows the perimeter of each TC to identify the nearest neighbor, which gave the text component. Then in next stage, the tails of CTC were used to identify the direction of text to find nearest neighbor, the objective of this step was to form a word. Next, word patches are combined to detect text lines. The datasets used are as follow: Hua's dataset, ICDAR 2003 dataset, arbitrarily-oriented data, non-horizontal data and horizontal data. The proposed method outperforms the existing methods. The strength of the system lies in the fact that it detects text in arbitrary orientation in videos, but the problem in the method is it may not give good result for horizontal text line with less space. Table 1 provides a detailed analysis of all the above technique

Table 1 Shows a cross comparison of all the techniques discussed

Author	Method	Advantages	Limitations	Dataset	Performance		
					Precision	Recall	F-measure
Iqbal et al.[1]	K2 algorithm based text detection using classified threshold	Better than ICDAR 2011's reading competition.	K2 algorithm is used here which have a drawback that is its high dependence on ordering of nodes in a structure.	ICDAR 2013	84.97%	62.37%	71.94%
Jaderb-urg et al.[2]	Deep features for text spotting	CNN goes through the whole image once, which simplifies the task, automatic production of word level and character level annotations. The output was classified that added to the efficiency of method. Better than Wu[31], Alsharif [32].	The saliency maps generated, may sometime give bad resolution and wrong result.	ICDAR 2003, ICDAR 2005, ICDAR 2011, ICDAR 2013, Street View Text dataset	X	X	X
Yin et al.[3]	Robust Text Detection in Natural Image	Minimized regularized variation is better than linear smoothing or median filtering which reduce noise but smoothens the edges to some degree.	In single link clustering could lead to chaining phenomenon which may lead to impractically heterogeneous clusters and difficulties.	ICDAR 2011, multilingual dataset(Chinese and English	86.29%	68.26%	76.22%
Huang et al[4]	Text Localization in Natural Images using SFT and Text Covariance Descriptors	SFT filter and two TCD classifiers enhanced the system performance many folds and leads to high recall and high precision respectively. Better than Neumann[33].	Using SFT, the direction of stroke is not determined so can only detect horizontal text line.	ICDAR 2005, ICDAR 2011	81% and 82%	74% and 75%	72% and 73% respectively
Yao et al.[5]	Strokelets: A Learned Multi-Scale Representation for Scene Text Recognition	Strokelets have URGE properties (usability, robustness, generality and expressivity) which make it better than previous traditional approaches. Proposed algorithm achieved better result than present algorithms. RF classifier is used instead of SVM because, performs better than later.	Random forest is used for classification but in this classification, Regression can't predict beyond a particular range in the training data.	ICDAR 2003(full) with accuracy 80.33%, ICDAR 2003(half) with accuracy 88.48%, SVT with accuracy 72.89%, ICDAR 2013	X	X	X
Ye at al[6]	Scene Text Detection via Integrated Discrimination of Component Appearance and Consensus	Gemma co relation is used to extract low contrast text components. Using all hypothesis generated in method could be time consuming so they used loose	If the distance between candidate character is as large as text height then the object may be missed to be recognized.	ICDAR 2011 and SVT	43.89%	67.52%	53.20%

		constraints on component spatial distance and alignment.					
Neumann et al.[7]	Scene Text Localization and Recognition with Oriented Stroke Detection	No. of rectangles are reduced many fold due classification of set of strokes. Better than Neumann and Matas[33].	There exist an ambiguity that a sub-region of a character might be another character, failure to detect letter on word boundaries which consist of single stroke(e.g. 'l' , '1')	ICDAR 2011	79.3%	66.4%	72.3%
Ganesh et al.[10]	Extraction of Text from Images of Big Data	This method worked really good, as it is v crucial to deal big data.	This method took images of google API as big data, the result may differ for other big data.	Google API is used to obtain images of street view to get dataset.	X	X	X
Minetto et al[8]	SnooperText: A text detection system for automatic indexing of urban scenes	The multi-scale fashion make segmentation algorithm insensitive to character texture like high frequency details.	The grouping module doesn't work well for wide spaces. So need to run twice. Due to the fact that text that is near to low legibility, the character that can't be separated, excessively distorted fonts and isolated character, those were eliminated by grouping modules. And can't detect tilted or vertical text.	ICDAR 2003, 2005	74%	63%	68%
Gao et al.[11]	A Hierarchical Visual Saliency Model for	Better than conventional Itti's method[12]	A problem is the threshold method and hierarchal clustering method for	3018 scenery images are used as dataset	X	X	X
Xiao et al[13]	An Efficient Method of Text Localization in Complicated Background Scenes	Ycbcr is used which is a complete model used conventionally. Work well in low illumination.	For long curves of the background and noise the image is scanned twice for removing noise.	ICDAR 2005, 1000 images of low illumination text in them are used as dataset.	X	X	X
Gonzalez et al.[14]	A text reading algorithm for natural images	The no of false positive was so small in this method's evaluation.	This method does not work for multi oriented text.	ICDAR 2003, and 2011	81% and 72.76%	70% and 56.00 %	69% and 63.25%.
Gonzalez et al.[15]	Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance	Character recognition was enhanced to detect symbols.	This method was only applied to the area where panel was detected and only on blue and white region of traffic panel.	Images taken from google street view.	X	X	X
Iqbal et al.[16]	Bayesian Network Scores Based Text Localization in Scene Images	Ranked 4th among 10 top published methods.	Dependence on yin et al.[20].	ICDAR 2013	84.30%	63.54 %	72.44%
Jaderburg et al.[17]	Synthetic Data and Artificial Neural Networks	Whole image is used to detect word. This method improved over standard method because of its	Synthetic engine sometimes provide a large amount of data that may complicates the problem.	ICDAR 2003, 2013, SVT	X	X	X

	for Natural Scene Text Recognition	simple fast machinery and cost reduction of data acquisition. Better than Wang[34] and Wang and Wu[31].					
Yao et al.[18]	A Unified Framework for Multi-Oriented Text Detection and Recognition	This method work for multi orientation of text which makes it diff. from traditional approaches. Resembling symbols errors are eliminated using dictionary search method.	In random forest Regression can't predict beyond a particular range in the training data.	ICDAR 2011, MSRA TD500	82% and 64%	65% and 62%	73% and 61%
Milevskiy et al.[19]	Joint Energy-based Detection and Classification of Multilingual Text Lines	This model combined the geometric error cost.	The narrowness of our algorithm was due to Adaboost classifier but can be enhanced.	Dataset consisted of 500 images taken from handheld devices.	84%	71%	76%
Barlas et al.[20]	A typed and handwritten text block segmentation system for heterogeneous and complex documents	Heterogeneous documents are dealt.	Drawback of system is that can't perform well in text present on image and graphical part of any document.	MOURDOR's dataset	82.9%	82.7%	X
Yin et al.[22]	Effective Text Localization in Natural Scene Images with MSER, Geometry-based Grouping and AdaBoost	Contrary to traditional approaches of sliding window and CC, a new better way is proposed. Better than method published till 2011.	Effective for only horizontal text.	ICDAR 2011	81.53%	62.2%	70.58%
Hanif et al.[23]	Text Detection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm	The method work well for different font, style and complex background.	Works only for grey scale images.	ICDAR 2003	56%	64%	X
Ye et al.[24]	Text detection and restoration in natural scene images	The SVM classifier have good generalization capability as compared to neural network and decision tree. OCR removes the false candidates.	The drawback lies due to the inconsistent behavior of GLVQ algorithm.	1500 images captured by handheld devices,	X	92.5%	X
Shahab et al.[25]	ICDAR 2011 Robust Reading Competition Challenge 2:	Provide a good comparison of multiple methods and points out the best one.	Only a limited no. of methods participated in word recognition.	ICDAR 2011	X	X	X
Toyama et al.[27]	A Mixed Reality Head-Mounted Text Translation System Using Eye Gaze Input	Provide help to foreigners.	System just check the effectiveness of algorithm.	X	X	X	X
Wang et al.[29]	Spatially Prioritized and Persistent Text Detection and Decoding	As the images captured during motion so they were blurry and	When the distance is 1.5m while taking observation then the decoder may not work well but perform better	Self-designed dataset	X	X	X

		proposed method does the detection well.	on wrapped observation than the original one.				
Shivakumara et al.[30]	Scene Text Localization and Recognition with Oriented Stroke Detection	Detect multi oriented text in videos .	May not give good result for horizontal text line with less spaces. API(average processing time) is larger.	ICDAR 2003	36%	42%	35%

3.0 CRITICAL EVALUATION

We have critically analyzed the methods being discussed in this paper, Yao et al. [5] and Neuman et al. [7] proposed the use of strokelet for representation of text in image but Yao et al. [5]'s method worked not only for single scale but also multi-scale representation. Huang[4] also used stroke feature transformation along with two TCD(text Covariance descriptor) that really enhanced the performance. Ye et al. [6] prosed a combination of appearance and consensus based approach, so this combination approach was better than separate one's. Iqbal et al. [1] proposed adaptive classifier threshold method using posterior probabilities whereas Yin et al. [8] also used posterior probabilities and designed an algorithm for text detection but Yin et al. [8]'s method was multi-lingual. Gonzalez et al. [14],[15] proposed adaptive classifier threshold method, but [15] was only designed for traffic panels. Gao et al. [11] invented a visual saliency model for text detection in images and it is better than conventional Itti et al. 's method[12]. Minetto et al. [8] proposed snooper text in multi-scale fashion and snooper text was proved to be best than TessBack, TesseRact. Ganesh et al. [10] worked on text detection in big data as compared to rest of the methods discussed in the paper. Yao et al. [18] also proposed a unified framework for detection of multi oriented text. Iqbal et al. [16] used Bayes network score and K2 algorithm and result showed that this technique performed well on among the top 10 methods of its time. Milevskiy et al. [19] worked on joint energy based detection and classification of text and introduced new hierarchal model MDL. Barlas et al. [20] worked on recognition of hand written or typed text that was quite challenging as the hand writing varies person to person. Yin et al. [22] used Maximally stable extremal region and geometric features for text detection whereas Hanif et al. [23] used constrained Adaboost and neural networks for text detection whereas Jadderberg et al. [2], [17] also used neural networks for text detection, ref[17] was better because it used synthetic engine and worked on synthetic data also its better than Wang et al. [34] and Wang et al. [31]. Ye et al. [24] used GLVQ(generalized learning vector quantization) and SVM. Toyamma et al. [27] used eye gaze as an input and the OCR(optical character recognition) technology for reading text, this method was also multi-lingual. Wang [29] used SLAM(simultaneous localization and mapping) for reading text in videos, though this method worked was not tested on some standard dataset but it performed quite well. Xiao et al. [13] and Shivakumara et al. [30] used soble edge map for obtaining feature for text detection among these shivakumara at al.[30]'s method was better as it works for multi-orientation of text in videos whereas the other only does detection in images.

4.0 CONCLUSION

In this paper we have mentioned many methods for detection and recognition of text present in images and videos, every method have strengths and limitations. The text detection is done using many techniques and method like MSER for candidate extraction, saliency maps, SFT, density based method, geometric features based methods, color homogeneity based method. Different techniques are used for different kind of images so we can conclude that there could not be a technique like "one size fits all" due to variation in the images but we can improve or transform every technique. The dataset used by most of the methods are ICDAR 2003, 2005, 2011, 2013 and result is measured in term of precision, recall and f-measure. The field is still in adolescence so better technique could be developed.

References

[1] Iqbal, K., Yin, X. C., Hao, H. W., Asghar, S., and Ali, H. 2014. K2 Algorithm-based Text Detection with An Adaptive Classifier Threshold. *International Journal of Image Processing (IJIP)*, 8(3): 87-94.

[2] Jaderberg, M., A. Vedaldi, and A. Zisserman 2014. Deep Features For Text Spotting. *In Computer Vision–ECCV*. 6 September 2014. 512-528.

[3] HU, Q., L. R., HONG, and MA, L. L. 2013. Text Detection in Natural Scene Images. *Computer Knowledge and Technology*. 22: 45-47.

[4] Huang, W., Lin, Z., Yang, J., & Wang, J. 2013. Text Localization In Natural Images Using Stroke Feature Transform And Text Covariance Descriptors. In *Computer Vision (ICCV), 2013 IEEE International Conference*. 1(2): 1241-1248.

[5] Yao, C., X., Bai, B., Shi and W., Liu. 2014. Strokelets: A Learned Multi-Scale Representation For Scene Text Recognition. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference*. 9 February 2014. 4042-4049.

[6] Ye, Q., and D. Doermann. 2014. Scene Text Detection Via Integrated Discrimination Of Component Appearance And Consensus. In *Camera-Based Document Analysis and Recognition* 5th International Workshop, CBDAR 2013. Washington, DC, USA. 23 August 2013. 47-59.

[7] Neumann, L., and J., Matas. 2013. Scene Text Localization And Recognition With Oriented Stroke Detection. In *Computer Vision (ICCV), 2013 IEEE International Conference*. 1(2):97-104.

[8] Minetto, R., N., Thome, M., Cord, N. J., Leite, and J. Stolfi. 2014. Snoopertext: A Text Detection System For Automatic Indexing Of Urban Scenes. *Computer Vision and Image Understanding*. 122: 92-104.

[9] Fabrizio, J., M., Cord, and B. Marcotegui. 2009. Text Extraction From Street Level Images. *CMRT09-CityModels, Roads and Traffic*. 38(3): 199-204.

[10] Ganesh, V., and L. G. Malik. 2014. Extraction of Text from Images of Big Data. *International Journal*. 2(3): 40-46.

[11] Gao, R., F., Shafait, S., Uchida, and Y., Feng. 2014. A Hierarchical Visual Saliency Model for Character Detection in Natural Scenes. In *Camera-Based Document Analysis and Recognition*. 23 August 2014. 18-29.

[12] Walther, D., L., Itti, M., Riesenhuber, T., Poggio, and C. Koch. 2002. Attentional Selection For Object Recognition A Gentle Way. *Biologically Motivated Computer Vision*. 22 November 2002. 472-479.

[13] Xiao, H., and Y. Rao. 2014. An Efficient Method of Text Localization in Complicated Background Scenes. *Journal of Software*. 9(6): 1538-1544.

[14] Gonzalez, A., and L. M. Bergasa. 2013. A Text Reading Algorithm For Natural Images. *Image and Vision Computing*. 31(3): 255-274.

[15] Gonzalez, A., L. M., Bergasa, and J. J. Yebes. 2014. Text Detection And Recognition On Traffic Panels From Street-Level Imagery Using Visual Appearance. *Intelligent Transportation Systems, IEEE Transactions.* Beijing, China. 6-*11 July 2014.* 15(1): 228-238.

[16] Iqbal, K., X. C., Yin, H. W., Hao, S., Asghar, and H., Ali. 2014. Bayesian Network Scores Based Text Localization In Scene Images. In *Neural Networks (IJCNN), 2014 International Joint Conference.* Beijing, China. 6-*11 July 2014.* 15(1): 2218-2225.

[17] Jaderberg, M., K., Simonyan, A., Vedaldi, and A., Zisserman. 2014. Synthetic Data And Artificial Neural Networks For Natural Scene Text Recognition. *Arxiv Preprint Arxiv*. 9 June 2014. 1406. 2227.

[18] Yao, C., X., Bai, and W., Liu. 2014. A Unified Framework For Multi-Oriented Text Detection And Recognition. *Image Processing, IEEE Transactions*. 23(11): 4737-4749.

[19] Milevskiy, I., and Y., Boykov. 2014. Joint Energy-based Detection and Classificationon of Multilingual Text Lines. *arXiv preprint arXiv*. 23 July 2014*. 1407.6082*.

[20] Barlas, P., S., Adam, C., Chatelain and T. Paquet. 2014. A Typed And Handwritten Text Block Segmentation System for *Analysis Systems (DAS). 11th IAPR International Workshop on* heterogeneous and complex documents. Tours, France. 7-10 April 2014*. 46 – 50.*

[21] Breuel, T. M. 2002. Two Geometric Algorithms For Layout Analysis. *Document Analysis Systems v*. 19 August 2002. 188-199.

[22] Yin, X., X. C., Yin, H. W., Hao, and Iqbal, K. 2012. Effective Text Localization In Natural Scene Images With MSER, Geometry-Based Grouping And AdaBoost. In *Pattern Recognition (ICPR), 2012 21st International Conference. Tsukuba, Japan. 11-15 November 2012*. 725-728.

[23] Hanif, S. M., and L. Prevost. 2009. Text Detection And Localization In Complex Scene Images Using Constrained Adaboost Algorithm. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference. Barcelona, Spain. 26-29 July 2009. 1-5*

[24] Ye, Q., J., Jiao, J., Huang and H. Yu. 2007. Text Detection And Restoration In Natural Scene Images. *Journal of Visual Communication and Image Representation*. 18(6): 504-513.

[25] Shahab, A., F., Shafait, and A. Dengel. 2011. ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text In Scene Images. In *Document Analysis and Recognition (ICDAR), 2011 International Conference*. Beijing, China. 18-21 Sept. 2011. 1491-1496

[26] Wolf, C., and J., M., Jolion. 2006. Object Count/Area Graphs For The Evaluation Of Object Detection And Segmentation Algorithms. *International Journal of Document Analysis and Recognition (IJDAR)*. *8*(4): 280-296.

[27] Toyama, T., D., Sonntag, A., Dengel, T., Matsuda, M., Iwamura and K. Kise. 2014. A Mixed Reality Head-Mounted Text Translation System Using Eye Gaze Input. In *Proceedings Of The 19th International Conference on Intelligent User Interfaces*. Haifa, Israel. 24-27 February 2014. 329-334.

[28] Kobayashi, T., T., Toyamaya, F., Shafait, M., Iwamura, K., Kise, And A., Dengel. 2012. Recognizing Words In Scenes With A Head-Mounted Eye-Tracker. In *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop*. Gold Cost, QLD 27-29 March 2012. 333-338.

[29] Wang, H. C., Y., Landa, M., Fallon, and S., Teller. 2014. Spatially Prioritized And Persistent Text Detection And Decoding. In *Camera-Based Document Analysis and Recognition*. 23 August 2014. 3-17

[30] Shivakumara, P., Phan, T. Q., Lu, S., and Tan, C. L. 2013. Gradient Vector Flow And Grouping-Based Method For Arbitrarily Oriented Scene Text Detection In Video Images. *Circuits and Systems for Video Technology, IEEE Transactions. 23*(10): 1729-1739.

[31] Wang, T., D. J., Wu, A., Coates, and A., Y., Ng. 2012. End-To-End Text Recognition With Convolutional Neural Networks. In *Pattern Recognition (ICPR), 2012 21st International Conference. Tsukuba, Japan. 11-15 November 2012*. 3304-3308.

[32] Alsharif, O., and J. Pineau. 2013. End-To-End Text Recognition With Hybrid Hmm Maxout Models. *arXiv preprint arXiv:* 7 October 2013. *1310.1811*.

[33] Neumann, L., and J., Matas. 2012. Real-Time Scene Text Localization And Recognition. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference. Providence, RI. 16-21 June 2012.* 3538-3545.

[34] Wang, K., B., Babenko, and S. Belongie. 2011. End-to-end scene text recognition. In *Computer Vision (ICCV), 2011 IEEE International Conference. Barcelona, Spain. 6-13 Nov. 2011.* 1457-1464.

[35] Singh, S., A., Gupta, and A. A. Efros. 2012. Unsupervised Discovery Of Mid-Level Discriminative Patches. *Computer Vision–ECCV 2012*. 3(4): 73-86.

[36] Nomura, S., K., Yamanaka, O., Katai, H., Kawakami and T., Shiose. 2005. A Novel Adaptive Morphological Approach For Degraded Character Image Segmentation. *Pattern Recognition*. *38*(11): 1961-1975