MapReduce based Text Detection using MSER in Big Data Natural Scene Videos

Parul Pathak, Nirmal Gaud
SATI, Vidisha (M.P)

ABSTRACT

Text is one of the most important features in images and videos. It can be used for various analysis purposes. Natural scene text is the text which is automatically present in the scene like sign board etc. In this research work, a method is proposed to detect text in Natural scene videos using MapReduce and MSER (Maximally Stable Extremal Regions). The proposed text detection technique is analyzed on three different classifiers: SVM, LDA and Random forest classifiers. This text detection is very useful in today's era where 'data' is everything.

Keywords

Text detection, MSER, MapReduce, Big Data, Classifiers.

1. INTRODUCTION

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate .It can be defined as high volume, high velocity & high variety information assets which requires cost effective, innovative forms of information processing for faster and better decision making[1].

The process of examining or analyzing big data to discover patterns trends and then find correlations between them to make better decisions. Big data analytics is a link between Information Technology, business users and data scientists.

1.1 Text Analytics:

Text is an important source of information in images and videos. It can provide very important details for analysis purpose. Text analytics is used for information retrieval from data. It involves machine learning, statistical analysis and computational linguistics. E-mails, blogs, online forums, news and call center records are all examples of text data. some of the techniques used in text analytics are Information Extraction, Text Summarization, Question Answering and Sentiment Analysis. Text analytics is a very challenging task due to various problems like diversity of text, noise, low contrast, complexity of the background, illumination etc. But it became popular due to great success of smart phones. In this paper, we have proposed a method for text detection in videos using MSER technique.

1.2 Big Data :

Big data can be defined as large size datasets which requires effective & innovative forms of information processing by using new algorithms & mining techniques. Big data can be explored by its 5 V's which are :volume, variety , velocity, veracity and value.[1]

1.2.1 Hadoop:

Hadoop: Hadoop is an open source, software framework which is used to store & process large data sets in a distributed fashion on large clusters of hardware(commodity hardware). It was created by Doug Cutting in 2005. Apache Hadoop meets the challenges of Big Data by simplifying the implementation of highly parallel distributed applications[2]. It provides a powerful computational platform, providing highly scalable, parallelizable execution that is well suited for the creation of a new generation of powerful data science and enterprise applications.
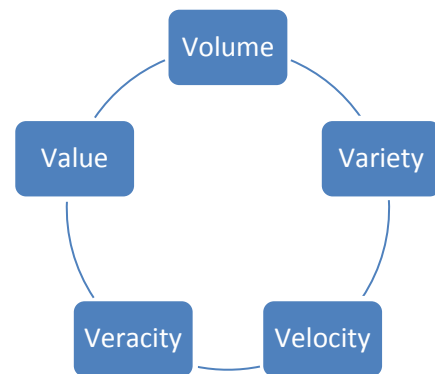

Fig.1 Image showing the 5V's of big data

1.2.1.1 Hadoop Components:

Hadoop has two core components:

1. **HDFS:** A foundational component of the Hadoop ecosystem is the Hadoop Distributed File System (HDFS). It is the storage component of Hadoop. It is natively redundant. HDFS is the mechanism by which a large amount of data can be distributed over a cluster of computers, and data is written once, but read many times for analytics. It provides the foundation for other tools, such as HBase.

2. **MapReduce:** MapReduce is Hadoop's main execution or computational framework. It's a programming model for distributed, parallel data processing, which splitsa task across multiple nodes. It breaks the jobs into *mapping* phases and *reduce* phases (thus the name). Developers write *MapReduce jobs* for Hadoop, using data stored in HDFS for fast data access. Because of the nature of how MapReduce works, Hadoop brings the processing to the data in a parallel fashion, resulting in fast implementation.

1.3 MSER (Maximally Stable Extremal Regions)

MSER was proposed by Matas et al. to find correspondences between image elements from two images with viewpoints. MSER is used in text detection by combining it with canny edges. MSER can be used to track color objects. The term "Extremal Regions" shows connected components which differentiate the pixel intensity (higher or lower) to the outer boundary pixel. MSER performs well on images containing homogeneous regions with distinctive boundaries and it works well for small regions.

2. LITERATURE SURVEY
2.1 Text Detection techniques used earlier:

Julinda Glivata et al.[4] proposed a method which is based on the application of a color reduction technique, which is a method for edge detection and then localization of text regions is done using projection profile analysis and other geometrical properties. This approach allows to detect, localize and extract texts from color images with complex backgrounds. Ankur et al. [5] used a heuristic method of text extraction which can be preferred using 2 different approaches, which are:

i) Connected regions approach and

ii) Edge detection approach.

Connected regions approach implies that a letter can be considered as homogeneous regions and hence it can be used to divide the frame into homogeneous regions. Edge detection approach works on the idea that text contrasts a lot with background.

Chucai et al.[6] proposed an algorithm, which is based on stroke components & descriptive gabor filters to detect text regions in images. Gabor filters are used to analyze the stroke components in text characters. As we know that information in videos and images can be very useful for auxiliary reading, image retrieval etc. This algorithm proposed by [6] is able to handle complex backgrounds & different text patterns. They used image window, as a basic processing cell.

Anna zhu et al.[7] proposed a framework to detect natural scene text by means of scene context information. They used both character feature and context feature for discriminating a connected component into text or non-text.

Q. Ye. et al.[8] proposed a coarse-to-fine text detection framework by applying various text properties in different detection stages. Jain & Yu [9] proposed a classic text detection algorithm which is based on connected component analysis for detecting text in video frames, web images, newspapers etc. In this proposed algorithm the original image is decomposed in to a multi valued map, Then the connected components of same color are selected, and they are classified as text and non text based on the following two strategies:

1. Inter component features (Connection property and geometry)

2. Projection profile features in both direction (Vertical as well as Horizontal)

It is a coarse -to fine detection frame work which provide faster result and also benefit the real time video analysis.

there are three general methods for localizing text[8]:

1: Connected component methods: this method detects the text using property like size form and spatial alignment constraint. It is based on the concept that characters in same text regions exhibit certain similar properties like similar stroke width, proximate pixel value etc.

2: Edge detection method: This method uses edge filters such as sobel and canny for text detection.

3: Texture based classification method: These methods treats text regions as a special type of texture and these methods assumes that text regions have different textual features compared with the background. In texture based method feature design is very important.

Salem et. al.[8] Proposed a method to classify pixels in to text and non text areas using wavelet transformation and neural network. It mainly composed of two steps: First is an off line step for creation of neural network and Second is an online step for text detection. First phase is the creation of neural network which has two parts: Feature vector and Learning Neural network. The second section describes the process to detect text zones. This algorithm converts the normal image in to grey scale image, resize it and then extracts the candidate regions using neural network wavelet transformation and morphological operations. After this, sobel filter is applied and localization of text zones is performed.

Rumin wang et.al.[9] Proposed an approach to detect text in natural scenes using confidence map and context information. They proposed a method to robustly detects text by combining connected component and texture based methods. These two methods are combined the context information.

Too boon chan et.al.[10] proposed a system to extract and recognize text in videos by using and edge detection based text segmentation method. This is performed by combining various techniques, like multiple frame integration, entropy based thresholding, gray scale filtering and line adjacency graph to enhance (improve) the detected text areas. After this, the character recognition is performed using their character side profiles.

Lie Sun et.al.[11] proposed a text detection approach to detect text from natural scene images based on CER(Contrasting extremal region) and neural network.This approach achieved great appreciation on ICDAR 2011 and 2013. This algorithm is based on horizontal alignment assumption vertical and curved lines cannot be treated properly.

Md. Rabiul et al[12].proposed an improved MSER based method for text detection and text recognition. In this approach, MSER is applied to an image to detect the text region.This uses MSER as main connected component. Here, canny edge and MSER are combined to get more reliable result. They have also used SWT (Stroke Width Transformation) to perform filtering & joining the connected components. After the filtering, OCR (Optical Character Recognition) & GMM (Gaussian Mixture Model) is operated on selected text region to identify the actual text information of natural scene images.

Houssam Turki et al.[13] proposed a method for text detection based on location of concentration areas of text candidates. This method applies a mask to filter the complex background. After this, they used Otsu technique & edge enhancement. The next step in this method is fine detection of candidate characters by MSER. The non-text components are filtered by various filters like DTW,SIFT and we apply boundry box localization method for detecting text areas.

Abida Tabassum et al[14]. proposed an algorithm for detecting text in images.This algorithm combines MSER and SWT (Stroke Width Transformation). Here, MSER image is obtained and for edge enhancement, canny edge detection is applied. After this, SWT and geometric filtering is used to filter the non text parts.

Lluis Gomez et al.[15] presented a method for tracking and detection of text in natural scenes. In this work, the text detection is combined with MSER tracking, which provides more stability and better performance.

Abdelkarim Ben Ayed et al. [16] proposed a methodfor texture-based text detection in videos using the Apache

Hadoop big data analytics. The proposed system decompose video frames in multiple fixed-size blocks, analyze those blocks using Har wavelet transformation to generate feature vectors and then classify them with a neural network classifier into text blocks and non-text blocks. The proposed system allows detecting texts of various languages and fonts with different colors, size and orientations with less than a half computation time due to the MapReduce parallel programming model.
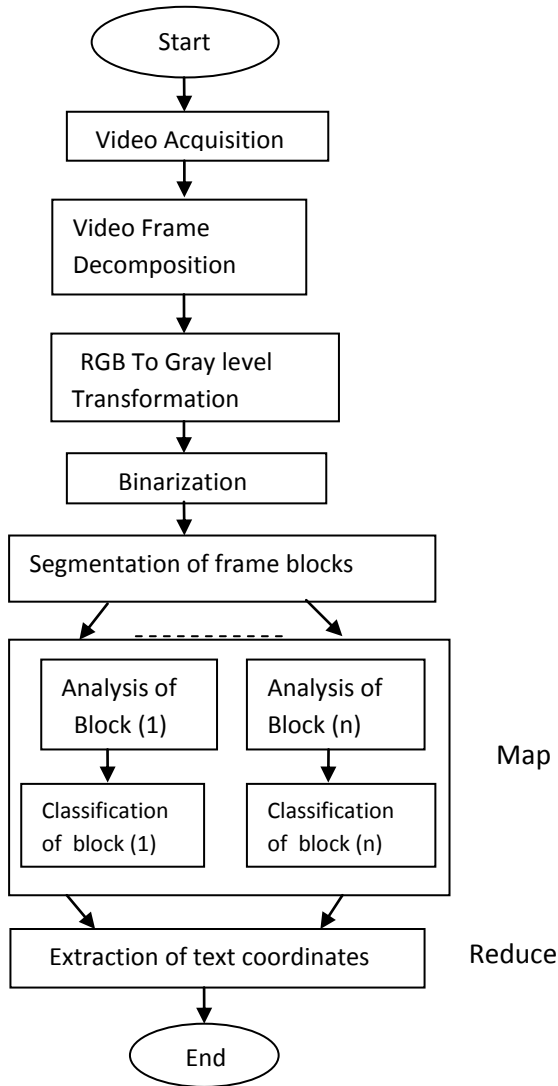
this step, it is required to convert the RGB frames into grey level frames by using the specific function. After converting RGB frames into grey level, binarization is performed for further simplification.
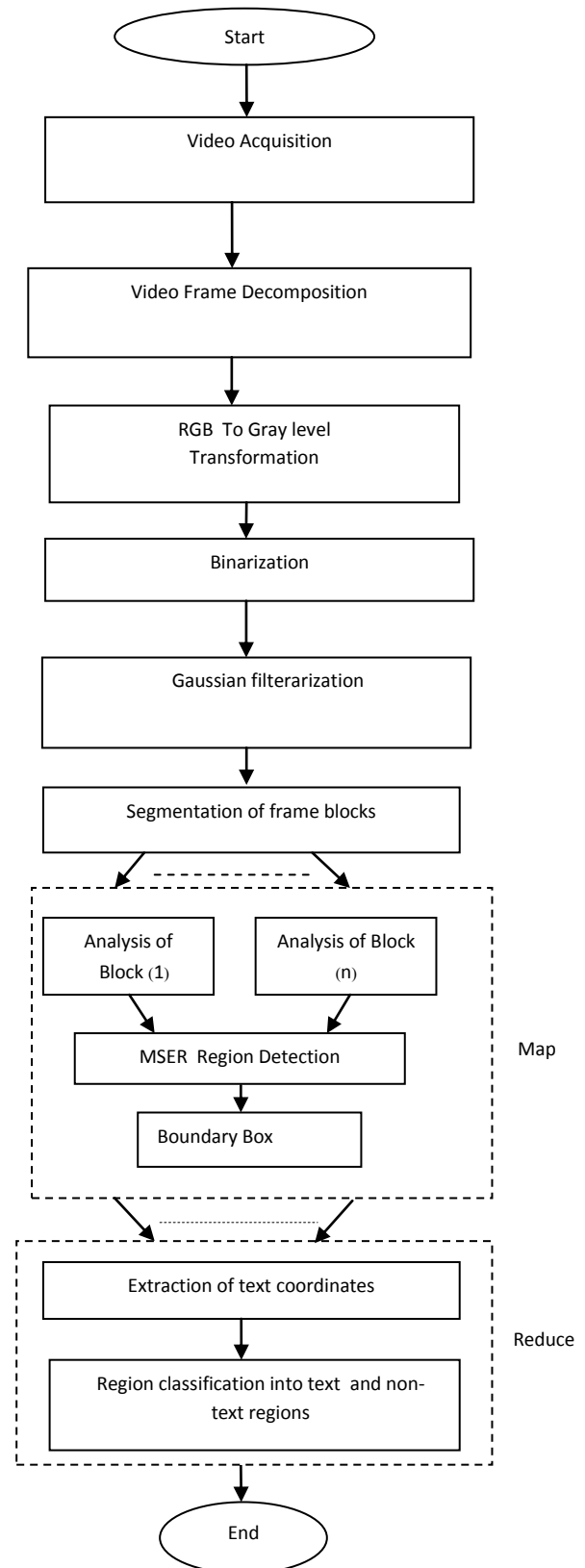

Fig. 2 Flow Chart of Map Reduce Based Text Detection Process.

3. PROPOSED METHODOLOGY

Based on the method given by [16], we have proposed a methodology which is used to detect text regions in natural scene videos (even if they are of low quality) by adding a Gaussian filter before segmentation and we have also applied MSER **(Maximally Stable Extremal Regions)** for detecting possible text regions and then applied various classifiers to differentiate between text and non-text regions. As shown in the proposed flow chart, it starts from getting/ acquiring the video from any source like CCTV footage, movie clip etc. After the acquisition of video, the video is decomposed into frames. Every video has a large number of frames depending up on the quality and length of the video. So, the process becomes simpler if we break the videos into frames. After

Fig. 3 Flow chart of proposed text detection process using MSER (Maximally Stable Extremal Regions)

Segmentation is done to identify the regions based on their texture. The texture based detection methods analyzes blocks of pixels using different filters. Analyzing of each frame is done by using various features like energy, entropy etc. We have used standard filter function to calculate the array of the pixels contained in the input image. (Another function we have applied is medfilt(), which is used to provide the corresponding value of the neighboring pixels of the text in the input image).

After getting analyzed blocks of pixels, MSER(Maximally Stable Extremal Region) region detection is applied which detects the areas of interest which contains text and form bounding box around them for text detection. MSER is very efficient in finding text regions because the high contrast and consistent color of the text leads to stable intensity profiles. Sometimes, there are some stable regions are also detected by MSER, which are non-text. These regions can be eliminated by some geometric properties like aspect ratio, eccentricity, extent etc. to get better results, we also have applied a machine learning approach i.e. a classifier. We have applied LDA(Linear Discriminant Analysis), SVM(Support Vector Machine) and Random forest classifier & compared them on the basis of various parameters to find out which is the best classifier for our algorithm among these three classifiers. These factors are recall, precision and accuracy. After applying the classifier, we finally get the text regions present in the input video.

4. RESULT ANALYSIS:

We have gathered YouTube videos to create datasets The dataset has 15 videos to analyze the algorithm which is approx. 2000 frames per video. The result shows that proposed algorithm takes less time & provides better and fast results and enhanced performance.(The proposed algorithm is analyzed on the basis of precision recall and accuracy). We have analyzed our proposed techniques on 3 different classifiers: SVM, LDA and Random Forest and we found that SVM has enhanced performance over LDA and Random Forest.
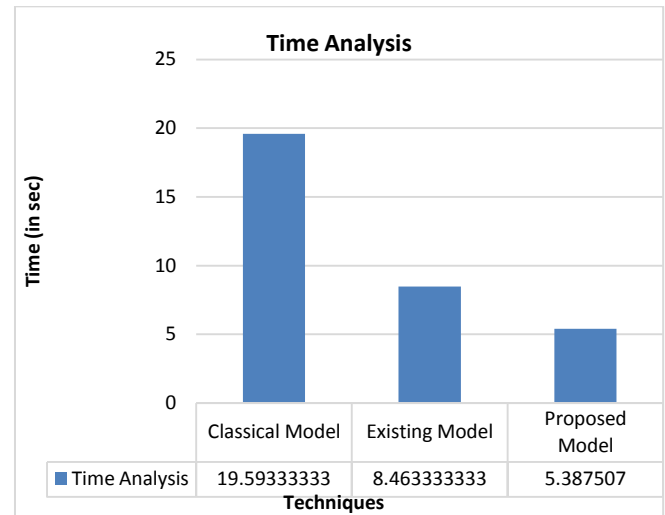

Fig. 4 Graph showing The Time taken by previous methodology and proposed methodology

	Classical Model	Existing Model	Proposed Model
Time Analysis	19.59333333	8.463333333	5.387507

Table 1: Table Representing the Time Taken in Detecting Text from Videos

S. No.	Input Image	Output Image	Time Elapsed (in Sec)
1			8.94371
2			4.910006
3			2.153514

4			3.305994
5			2.317037
6			12.784148
7			5.564572
8			3.121083
	AVERAGE TIME ELAPSED		**5.387507**

Performance Evaluation Parameters:

Accuracy = (TN+TP)/(TN+TP+FP+FN)
Precision = (TP)/TP+FP)
Recall = (TP)/(TP+FN)
(**TP-** True Positive
TN- True Negative
FP- False Positive
FN- False Negative)

Table 2: Table showing the Accuracy, precision and recall of the three classifier.

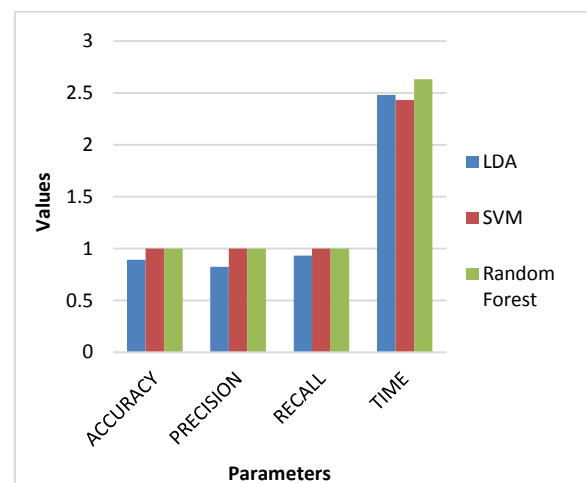| PARAMETERS	TECHNIQUES		
	LDA	**SVM**	**Random Forest**
ACCURACY	0.8919	1	1
PRECISION	0.8235	1	1
RECALL	0.9333	1	1
TIME	2.479826	2.431897	2.631818

Fig. 5 Graph showing the performance of the classifiers

The graph shown above depicts the performance of the classifiers based on the parameters like Accuracy, Precision, Recall and Time. And it clearly shows that SVM (Support Vector Machine) is the best classifier among these three classifiers.

5. CONCLUSION & FUTURE WORK

In this research work, a method is proposed to detect text in natural scene videos using MapReduce and MSER(Maximally Stable Extremal Regions). And we have concluded that after using MSER in the previous approach, the processing time is reduced. Three classifiers used classifiers to differentiate text and non-text regions. We have analyzed our proposed technique on 3 different classifiers: SVM, LDA and Random Forest classifier and we found out that SVM has enhanced performance over LDA and Random forest on the basis of the performance evaluation parameters like Accuracy, Precision and Recall. In future work, we can reduce the processing time and improve the efficiency of the method by using different techniques.

6. REFERENCES

[1] K.Arun, Dr.L.Jabasheela, 'Big Data: Classification and Analysis Survey' International Journal of Innovative Research in Information Security (IJIRIS), Vol.1 Issue 3(September 2014)

[2] 'Big Data and Analytics' Seema Acharya, Subhashini Chellappan by Wiley India pvt. Ltd.

[3] PROFESSIONAL Hadoop® Solutions Boris Lublinsky Kevin T. Smith Alexey Yakubovich Copyright c 2013 by John Wiley & Sons, Inc., Indianapolis, Indiana.

[4] Julinda Gllavata, Ralph ewerth and Bernd Freisleben ' Arobust algorithm for text detection in images'

[5] Ankur Shrivastava,Dhananjay Kumar, Om shrivastav : Text extraction in video International Journal of Computational Engineering Research, Vol. 3 issue 5.

[6] Chucai Yi, Yingli Tian ' Text detection in natural scene images by Stroke Gabor Words' 2011 International conference on document analysis and recognition.

[7] Anna Zhu, Renwu Gao, seiichi Uchida, 'could scene context be beneficial for scene text detection' Pattern Recognization,58 2016, 204-215

[8] Salem Sayahi, Mohamed Ben Halima,' An intelligent and robust multi oriented image scene text detection'. Internatioanl conference on Soft computing and pattern recognition

[9] Qixiang Ye, Qingming Huang, Wen Gao, Debin Zhao ' Fast and Robust detection in images and video frames' Image and vision computing 23, 2008.

[10] Teo Boon Chen, D. Ghosh, S.Ranganath, 'Video text extraction and recognition..

[11] Lei Sun, Qiang Huo, Wei jia, Kai Chen ' A robust approach for text detection from natural scene imges' Pattern Recognition, 45, 2016, pg 2906- 2920.

[12] Md.Rabiul Islam, Chayan Mondal, md.Kawsar Azam and Abu Syed, Md. Jannat Ul Islam 'Text detection and recognition using enhanced MSER Detection and a novel OCR technique. 2015 fifth international conference on Informatics, Electronics and vision.

[13] Houssem Turki, Mohamed Ben Halima, Adel M.Alimi, Scene Text Detection Images With Pyramid Images and MSER enhanced. 2018

[14] Adiba Tabassum, Shweta A. Dhondse : 'Text detection using MSER and stoke Width Transform'. 2015,fifth international conference on communication systems and network technologies.

[15] Lluis Gomez, Dimosthenis Karatzas: 'MSER- based Real-Time Text detection and Tracking'22nd International conference on Pattern Recognition, 2014.

[16] AbdelKarim Ben Ayed,Mohamed Ben Halima and Adel M. Alimi, MapReduce based text detection in big data natural scene videos procedia computer science, INNS conference on BigData.vol.53 2015, pages 215- 223.

[17] Ashwin Ittoo,Le Minh Nguyen, Antal Van Den Bosch, Text Analytics in industry: Challenges desiderata and trends.Computers in industry, 78, 2016 pg 99-107.