International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 14750-14754
© Research India Publications. http://www.ripublication.com

Text Extraction from Video Images

Nidhin Raju
PG Scholar, Department of Computer Science,
Christ University, Bengaluru, Karnataka -560029, India.
Orcid Id: 0000-0001-5475-9010

Dr. Anita H.B
Associate Professor, Department of Computer Science,
Christ University, Bengaluru, Karnataka-560029, India.
Orcid Id: 0000-0003-1608-0175

Abstract

Video data contains beneficial textual information such as scene text and caption text. The different types of videos like movies, news videos, and TV programs video etc. are created by various video frames based on its purpose. In a country like India, there are only fewer studies has done on text extraction from video data especially in south Indian languages like Malayalam, Telugu, Kannada, and Tamil. The extracted text has many useful applications in video indexing, video key searching and assisting visually challenged people. Malayalam news channel named 'Mathrubhumi News' videos data are considered for the proposed study. It is very beneficial to Kerala people as it is one of the most media-centric regions in the world. In this proposed paper, a new method for text extraction experiments. The anticipated method extracts 13 different features for classifying the image consists of text or not. Both spatial and frequency domain features are extracted to classify. The different types of classification techniques are used to validate the algorithm. Simple Logistic, J48 and Random Forest classification techniques are giving a good result when compared to other methods. Results are encouraging, the average success rate found to be 98%.

Keywords: Discrete Cosine Transform, Fast Fourier Transform, Simple Logistic, J48, Random Forest, and Text Extraction

INTRODUCTION

The availability and efficiency of videos are growing rapidly in recent years. The text extraction can be used for searching important information from the huge amount of video data set. The text extracted from the video is obviously an important component. Using this text anybody can get an idea about the video. Extracted text acts as a key sign to determine the content of the video and it is very easy to categorize videos. Videotext extraction is identified as one of the key components of the video analysis and retrieval system. Videotext extraction can be used in many applications, like multilingual video information access, semantic video summarization and indexing, video security and surveillance, etc. Videotext can be scene text or caption text. Caption text, static or scrolling. Most static texts give the compact and direct information about the content presented in the news video. The procedure of video textual information extraction can be divided into two categories; detection and extraction [13]. Suppose a person who did not know the script of the video then the extracted data can be given to the intelligent systems for converting this to another type of scripts. The extracted text can be converted into a sound signal to assist the visually impaired persons. Most of the research work used edge detection, cc method, region-based method etc. for text extraction from the video [11].

The extraction of text from video frames for South Indian languages like Malayalam has been reported by very fewer researchers. There are various approaches to texture based method, morphological method etc. has been used to extract text from the video images. Most of the researchers worked on English type of script very few have been studied in Indian languages. Proposed work is to extract text from the video frames of a Malayalam news. The considered channel is 'Mathrubhumi News'. The limitations of existing methods are reported less accuracy when the video frame has a complex background. Usually, a video frame of the news has variations in text, font, size, poses, colors and shapes, therefore, feature selection is very challenging [12]. The Matlab tool is considered to implement the anticipated work.

The projected effort is working well for extracting text from 'Mathrubhumi News' channel and produced a promising result. To increase speed and accuracy of the anticipated work, single script means only Malayalam channel's video frames are considered. Many researchers used different techniques for text extraction and which can be classified as edge based, connected component based and texture based methods. Bottom-up technique for gathering littler segments into bigger segments until the point that all regions are recognized in that image is done using in the connected component based method. This is to identify the text components by using a geometric analysis and then to localize text regions to group them. The edge-based method works well on high complexity between text and background. In most of the situation, frequency domain features contain furthermost distinguishing information [2]. So in this situations, frequency domain features give better results when compare to spatial domain features.

The proposed work is based on both spatial and frequency features to extract the text from the complex background. This can be updated for videos based on games, news, television channels etc. Simple Logistic, J48 and Random Forest classifiers are used and conducted an experiment with the Weka tool.

LITERATURE REVIEW

Most of the researchers worked on foreign types of script extraction from video images. Few researchers have worked for Indian type of scripts. As per our knowledge, very few have experimented on South Indian scripts, especially for Malayalam videos. The anticipated work considered only Malayalam news video channels. This helps to extract very significant features which are relevant only for Malayalam scripts and also to improve the accuracy and speed.

Lajish V.L and Anoop K. proposed a morphology-based approach for extract and detect text from Malayalam news videos. The proposed method includes the selection of key frame images (which contain textual information) from the input videos and a robust text extraction from the keyframe images. Experiments are conducted over the keyframes extracted from MPEG-4 compressed news video streams of DD Malayalam. It is a Malayalam satellite channel supported by Doordarshan. They finalized the future research work as to be extended towards the recognition of the extracted text information so as to effectively use it for video indexing and related applications. The average precision and recall rate obtained for the overall process was 88.52% and 94.13% respectively [3]. Thika Ali H. Subber and Abbas H. AL-Asadi suggested a method to localize or detect and segment Arabic static artificial scripts located in the video. The dataset was created by taking video frames of five different videos. The different type of Acquisition used are broadcast and DVD. Around 500 number of textboxes without multiframe integration are taking in each video. From each video more than 5000 Number of characters without multiframe integration obtained. The amount of the correctly segmented characters recognized correctly was 84.43% [4]. Anubhav Kumar used Gabor Filter method for text extraction from video images. To localize the text region, heuristic and morphological filtering process methods were used. The main features of this method are, by using Gabor Filter, it is possible to text extraction from complex video images and the information retrieval applications in video frames and images. In the dataset, the various kind of data like News, Commercial and Video Frames collected [5]. V.Vijayakumar and R.Nedunchezhian provided an innovative approach for distinguishing video text regions containing score and information about players in the various sports videos. Based on five minute's video of ESPN sports channel for the superimposed text extraction they built their own video database [6]. The different frames extracted from the videos. Then the frame converted to Gray scale and cropped it. The text detection was done by using the canny edge detection algorithm. The proposed experiment could identify the videotext in the image's boundary only [4]. Anil K. Jain, Hongjiang Zhang, and Yu Zhong proposed a technique to automatically localize captions in the I-frames of MPEG compressed videos and JPEG compressed images. The database made by taking video frames from TV videos. The symbols and text that are existing at exact locations in a video image used to identify program associated with the video and the TV station [7]. Jianzhuang Liu, Hongjiang Zhang, Xiaoou Tang, and Xinbo Gao, offered a video caption recognition and detection a fuzzy-clustering neural network (FCNN) classifier based system. The dataset build by the Asian TV and TVB news programs in Hong Kong. Video data are programmed in an MPEG-1 format with a resolution of 288*352 and for the classification, Self-organizing neural network classifier used [8].

B.H.Shekara and Smitha M.L.b proposed an approach based on the gradient difference to localize text in the videos and scene images [9]. Marios Anthimopoulos proposed various methods for text extraction. The author created two datasets. The first one consists of 214 frames from athletic videos while the second contains 172 video frames from news broadcasts [10]. Bo Lilo, Xaoou Tang, Zhan Hongiiang and Jianzhuang Li proposed a method to extract text data from video image by completely applying the sequential information enclosed in the video image. The dataset made upon different videos by entirely using the temporal data limited in the video. Efforts had been made to recognize and extract the characters repeatedly to permit access to video data with high-level content [11].

IMPLEMENTATION/METHODOLOGY

Data collection and Pre-processing

The news videos were collected from YouTube. A Malayalam channel named 'Mathrubhumi News' videos are used to evaluate our algorithm. The downloaded videos have the format of 640*352. Dataset is created by extracting frames from videos. For every 3 seconds, we considered one frame from the video. The size of the text in videos has varied in size, color, texture and complex background [2]. So the frame image is divided into blocks of size 102*26.

The proposed algorithm is working in Matlab and will extract each and every 3 seconds frames from the video. The extracted frames will store as jpeg format images in a new folder on the local disk of the system. The next step is that to take a frame manually from the dataset and crop it into another 50 images and store it in jpeg format in another folder. From the cropped images, 13 different features are taking. The well-known Ostu's global thresholding method is used for converting the images into binary format [1]. The binary and grayscale representation of some sample cropped images are given in Table 1. The Table 2 picturizes the DCT and FFT conversion of a sample cropped images in both binary and grayscale format.

Table 1: Binary and Grayscale representation

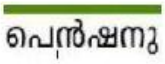
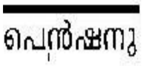
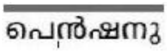
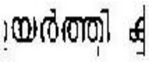
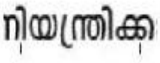
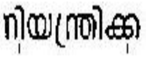
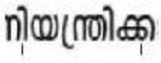
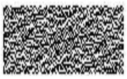
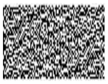
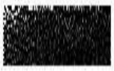
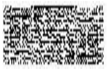
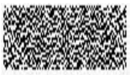
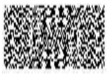
Image Name	Original Image	Black and White	Grayscale
Img1			
Img2			
Img3			
Img4			
Img5			

Table 2: DCT and FFT of divided video frame images

Image Name	Black and White DCT	Gray scale DCT	Black and White FFT	Gray scale FFT
Img1				
Img2				
Img3				
Img4				
Img5				

Feature Extraction

Thirteen different features are extracted from both binary and gray scale images. Feature extraction includes the deriving the meaningful information from the video frames. The features can be classified into Global features and Local features. When extracting features from the whole image then it is called global features. The features that extract from the blocks are known as local features. A brief description of each feature used is mentioned in the algorithm.

Algorithm

Input: Divided blocks of images from the extracted Frames

Output: Total number of 13 Features

Method:

1. Sum of 1's and 0's in binary images. This forms 4 features.
2. Standard deviation of row, column and diagonal elements from gray scale images. This forms 3 features.
3. Mean and standard deviation of gray scale and binary images. This forms 4 features.
4. Applied DCT and FFT on images and calculated standard deviation of gray scale and binary images. This forms 4 features

CLASSIFICATION

The classifiers discover best suitable class for input and distinguish the input feature with the stored pattern. There are many approaches used for classification such as Simple Logistic, Random Forest, J48 etc. Classifiers used in this experiment are given below,

Simple Logistic

Simple Logistic fits a multinomial strategic relapse show utilizing the Logit Boost calculation. In every emphasis, it includes one Simple Linear Regression show for each class into the logistic regression model. Logit Boost utilizes a symmetric model, yet in the event that you run Simple Logistic with an adequately huge number of iterations, you get an indistinguishable predictor from utilizing Logistic, modulo the distinction in portrayal [15]. It's significantly more effective all things considered. Note that you can likewise join Logistic with Greedy Stepwise characteristic choice, alongside the Wrapper evaluator.

Random Forest

Random forest algorithm is an example for the supervised classification algorithms. As the term recommend, this calculation makes the forest with various trees. When all is said in done, the further trees in the forest are stronger the forest resembles [16]. Similarly in the Random forest classifier, the higher the quantity of trees in the forest provides the great precision comes about. To receive the prediction utilizing the trained irregular forest algorithm we have to breeze through the test includes the standards of each randomly made trees [17].

J48

A decision tree is an example of the analytical machine-learning model. Based on various attribute values of the accessible data, it chooses the target value of another sample. To classify a new item with a specific end goal, the J48 Decision tree classifier initially desires to make a decision tree in view of the attribute values of the accessible training [14]. In this way, at of any kind point, it comes across a training set it distinguishes the property that separates the different instances most unmistakably.

EXPERIMENTAL RESULTS

Simple Logistic, Random Forest and J48 algorithms are used to evaluate our work. Simple Logistic is a neural network based classification gave the best accuracy when compared to simple logistics and Random Forest. the Random Forest and J48 algorithms are tree-based classifiers. Table-3 gives the details regarding the result analysis of all the three classifiers. The cross-validation is done with 10 fold.

Table 3: Result Analysis

	Simple Logistic	**Random Forest**	**J48**
Correctly classified instances	98.333 %	96.667 %	95.667 %
Incorrectly Classified Instances	1.667 %	3.333 %	4.333 %
Total Number of Instances	1200	1200	1200

CONCLUSION

In this paper, text extraction from video frames of a news channel is presented. The anticipated work considered only Malayalam news video channels. Experiments are performed on scripts from a single Malayalam news channel named ‘Mathrubhumi News’. This helps to extract very significant features which are relevant only for Malayalam scripts and also to improve the accuracy and speed. The average success rate of 98% is gained using features extracted from Algorithm-1. This shows that the proposed experiment is robust. In future, we extend this to convert extracted text into a sound signal to assist the visually impaired persons.

REFRENCES

[1] G. G. Rajput and Anita H.B., “Handwritten Script Recognition using DCT and wavelet Features at Block Level”, IJCA, Special Issue on RTIPPR (3):158-163, 2010.

[2] G. G. Rajput and Anita H.B., “Kannada, English, and Hindi Handwritten Script Recognition using multiple features”, Proc. Of National Seminar on Recent trends in Image Processing and Pattern Recognition, ISBN: 93-80043-74-0, pp 149-152, 2010.

[3] Lajish V.L and Anoop K, “Mathematical Morphology and Region Clustering Based Text Information Extraction from Malayalam News Videos”, Springer International publishing, pp.431-442, 2015.

[4] Thika Ali H. Subber and Abbas H. AL-Asadi, “Arabic Text Extraction from Video Images”, Journal of Basrah Researches ((Sciences)), Vol. 39 , No. 4, pp. 120-136, 2013.

[5] Anubav Kumar, “An Efficient Approach for Text Extraction in Images and Video Frames Using Gabor Filter”, International Journal of Computer and Electrical Engineering, Vol. 6, No. 4, pp. 316-320, 2014.

[6] V.Vijayakumar and R.Nedunchezhian, “A Novel Method for Super Imposed Text Extraction in a Sports Video”, International Journal of Computer Applications, vol.15– no.1, pp.1-6, 2011.

[7] Hongjiang Zhang, Anil K. Jain, and Yu Zhong, “Automatic Caption Localization in Compressed Video”, IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 22, no. 4, pp.385-392, 2000.

[8] Jianzhuang Liu, Hongjiang Zhang, Xiaoou Tang, and Xinbo Gao, “A Spatial-Temporal Approach for Video Caption Detection and Recognitio”, IEEE Transactions On Neural Networks, vol. 13, no. 4, pp.961-971, 2002.

[9] B.H.Shekara , Smitha M.L., “Gradient Difference Based Approach for Text Localization in Compressed Domain”, International Conference on Emerging Research in Computing, Information, Communication and Applications (ERCICA-14), pp.1-11, 2014.

[10] Marios Anthimopoulos, “Text Detection in Images and Video”, International Conference on Document Analysis and Recognition, pp. 57–63, 2012.

[11] Jianzhuang Liu, Bo Lilo, Hongiiang Zhan, and Xaoou Tang, “Video Caption Detection And Extraction Using Temporal Information”, IEEE, pp.297-300.

[12] G.Gayathri Devi, T. Santhanam and C.P. Sumathi, “A Survey On Various Approaches Of Text Extraction In Images”, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, pp. 27-42, 2012.

[13] Dr. Prof. M.S. Panse, Shrugal Varde, and Akhilesh Panchal, “Comparative study of Image processing techniques used for Scene text detection and extraction”, International Journal of Engineering Research and General Science Volume 4, no.2, pp.183-188, 2016.

[14] P Nagabhushan1, Vimuktha Evangeleen Jathanna, "A Hybrid Method for Text Extraction from Mosaiced Image of Text Dominant Video", International Journal of Computer Science and Information Technologies, Vol. 7, no.3, pp.1061-1068, 2016.

[15] Divya gera and Neelu Jain, "Comparison of Text Extraction Techniques- A Review", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2, pp.621-626, 2015.

[16] Ankur Srivastava , Dhananjay Kumar,Om Prakash Gupta , Amit Maurya,and Mr.sanjay kumar Srivastava, "Text Extraction in Video", International Journal of Computational Engineering Research,Vol.03,Issue 5,pp.48-53, 2013.

[17] Vivek Dhanapal Sapate, "A Survey: Text Extraction from Images and Video", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, pp.393-400, 2016.