International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Text Detection and Recognition in Digital Videos

Som Sirurmath [1], Prajwal Gowda B V [2], Sourav Singh [3], Vishal N Kaushik [4], Krupashankari S Sandyal [5]

[1,2,3,4]UG Student, Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India

[5]Assistant Professor, Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India

---

Abstract - *Text dominates video viewing and comprehension because the text contains a wealth of useful knowledge about the video's contents. Humans frequently pay more attention to the text than to other items in a video, according to some studies. Text presented in videos include vital information for content analysis, indexing and retrieval of videos. Finding, verifying, and recognizing video text against complex backgrounds is a key technique for extracting this content. With the increasing availability of low-cost portable digital cameras and video recorders, there are large and growing archives of multimedia data, such as photos and videos hence retrieving videos quickly and efficiently has become a crucial problem. A type of important source that is useful for this task is video text, which contains high-level semantic information. This paper proposes a method of text detection and recognition in digital videos using text detections tools such as the OpenCV East Text Detector to extract the region of interest of text in the image frames and implementing text detection through Optical Character Recognition(OCR) and Pytesseract.*

Key Words: ***Text Detection, Text Recognition, Digital Videos, OCR, OpenCV East Text Detector***

1. INTRODUCTION

Videos have become one of the most popular contents on the Internet. The number of online videos has increased rapidly because of the convenience of uploading and downloading videos. As a result, there is a high demand for retrieval, efficient indexing, and localization of desired content from massive video collections. However, as the size of the database grows significantly, retrieving and indexing the information effectively becomes a major concern. More and more knowledge is now being converted into digital formats. Visual texts can be found in a variety of digital media, including photographs and videos. Unfortunately, these basic tasks do not meet the majority of today's multimedia document analysis users' needs.

This textual information which is important for a variety of reasons and to other people in several ways need to be extracted. The process of information retrieval in this model is done using text segmentation and recognition techniques. This technique helps in extracting the textual information more quickly and effectively. The technique first splits a video into a number of frames and each frame is processed as an image. For each frame text segmentation and recognition techniques are applied and the process is repeated for all the frames in a loop. Finally, the textual information in the videos is extracted.

2. LITERATURE REVIEW

[1] "A Novel Approach for Video Text Detection and Recognition Based on a Corner Response Feature Map and Transferred Deep Convolutional Neural Network" Wei Lu, Hongbo Sun, Jinghui Chu, Xiangdong Huang proposed a method for detecting and recognizing video text by combining a corner response feature map and transferred deep convolutional neural networks. The input video is decoded into frames by using OpenCv library and the text regions in the frames are identified using a corner response feature map. The text is then verified by constructing a transferred deep convolution neural network classifiers from VGG16, ResNet50, and InceptionV3 with a series of layer concatenation and fine-tuning. By using a Fuzzy-c-means clustering-based separation algorithm they were able to extract the text layer from complex backgrounds which then results in OCR ready text lines. To successfully complete this method a test dataset containing 2,000 typical high-resolution video frames collected from various sources, including movies, cartoons, and TV shows were used. The effectiveness of their approach was validated using three public test datasets which resulted in a good performance.

[2] "Text Detection and Recognition: A Review" Chaitanya R. Kulkarni, Ashwini B. Barbadekar. The paper examines and contrasts various stages in the text detection and recognition process, as well as various methods for text extraction from colour images. The methods followed are stepwise and integrated methods. Since the videos are made up of continuous frames the methods can be applied to each individual frame. The two common methods discussed in this paper include further tasks such as text detection and localization, classification, segmentation and text recognition. Stepwise methods have separate

detection and recognition modules and are suitable for the detection of large numbers of words in the image whereas Integrated methods can avoid segmentation or replace it with word recognition and are suitable for identifying specific words from image i.e. small lexicon. Advantages, disadvantages and applications of different approaches have been proposed in the paper.

[3] "Introduction to Video Text Detection'' by Tong Lu, Shivakumara Palaiahnakote, Chew Lim Tan proposed a video text detection model. It first reviews the relevant literature and then discusses characteristics and difficulties in video text detection. Various issues such as the low resolution of video images and captions in the video are examined. It gives us an overview of how video text detection has evolved from the field of document analysis. and also including image processing, pattern recognition, computer vision. While most of them can be applied to video images, detecting and extracting text in video poses specific challenges compared to document images, due to a number of undesirable properties of video that make it difficult to detect and extract text. Fortunately, text in video usually persists for at least several seconds, to give human viewers sufficient time to read it. In this paper, we'll look at the approach that is generally divided into the following stages to solve this video text detection and extraction problem: text detection, text localization, data monitoring, text binarization and text recognition.

[4] "Text Detection and Recognition using Multiple Phase Methods on Various Product Labels for Visual Impaired People" by Rizdanian and Fitri Utaminingrum. This paper examines text detection and identification, which are two important objectives in the field of computer vision and digital image processing science. The study's findings are beneficial to visually impaired people because they may aid them in purchasing and selecting their preferred product from the market. The text detection and recognition processes in this study used the Multiple Phase (MP) methods. The text detection was going through some phases which are the detection of Maximally Stable Extremal Regions (MSER), Canny edge detection, region filtering, and Optical Character Recognition (OCR). For the text recognition method, OCR was used. The experimental outcome for our proposed approach was 80.88 percent, which was better than the previous study, which used a two-stage classifier and only had a score of 69 percent.

[5] "Segmentation and Recognition of Text from Image Using Pattern Matching" by N. Anandhi and R. Avudaiammal. This paper proposes a method for segmenting and recognition of text in an image using pattern matching techniques. The approach taken for extracting text involves two phases: Data Creation Phase and Searching Phase. In the first phase images with text are collected and then the features such as character extraction, feature extraction and pattern matching are extracted for each image and stored in the database as a vector. In the second phase, the text extracted from the input image is identified using the following modules: Image Pre Processing, Mask Image, Character Extraction, Feature Extraction and Pattern Matching. This paper takes on a time and frequency domain analysis for segmentation and recognition of text from an image.

[6] "OntoSeg: a Novel Approach To Text Segmentation using Ontological Similarity" by Mostafa Bayomi, Killian Levacher, M.Rami Ghorab, Séamus Lawless. This paper proposes a novel approach to text segmentation known as OntoSeg. OntoSeg is based on the ontological similarity between text blocks and is used to explore the conceptual relations between text segments using a Hierarchical Agglomerative Clustering (HAC) algorithm to represent the text in a tree-like hierarchy that is conceptually structured. This rich tree-like structure further allows the segmentation of text in a linear fashion at various levels of granularity. The results of the experiments performed shows that using ontological similarity performs successful segmentation with low error rates. Another method is proposed which combines ontological similarity with lexical similarity and the results show an enhancement of the segmentation quality.

3. IMPLEMENTATION

To perform text segmentation and recognition in digital videos, the first step involves decoding the given video into frames. Each frame is processed as an image in a suitable format (jpg, jpeg, png, etc..). Once the video is split into frames, the process is carried out in a loop for all the frames and finally the segmentation and recognition techniques are applied to each frame and the textual information is extracted.

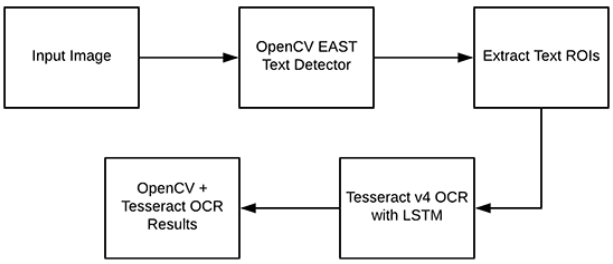

Fig -1: System Architecture

To make sure that the image processing is done only for specific regions, the region of interest should be obtained. This is done to reduce the processing time and increase the accuracy of OCR recognition for the textual information present in the video streams.

In order to obtain the region of interest (ROI), cv2.findcontours() technique is applied to each video frame, in order to ensure the image processing is done specifically and only on the region of interest(ROI). This helps in faster and effective processing of video frames and results in more accuracy for OCR recognition. Once the ROIs are obtained, the frames are resized in order to ensure the memory and operational costs of segmentation and recognition are reduced. After the frames are resized and ROIs are finalized, the tesseract tool is used for the recognition phase. The resized frames are given as input in the recognition phase, the cv2.findcontours() technique is applied to each frame and finally the textual information is extracted in this phase using the EAST text detector. After the textual information is extracted, a suitable bounding box is drawn around the region and then this is displayed as output in a suitable format, such that even end-users or people without any technology background can use the product implemented. The project is centered mainly on image processing and computer vision.

3.1 TECH STACK AND RESULTS

Python is the programming language used for the implementation phase, to build all the modules and functionality. To perform the image processing tasks and provide an infrastructure for the computer vision application, OpenCV is used. To perform the recognition of OCR characters, tesseract is used, which is an OCR engine trained to recognize OCR characters. The NumPy library is also used, which provides support to large and multi-dimensional arrays and matrices.

To perform the segmentation and recognition, the videos are initially decoded into frames, each frame is processed as an image and the necessary segmentation and recognition is done in a loop for all the frames. The product developed using the tools, technologies and methodologies mentioned above, is able to perform OCR recognition for the textual information in digital videos with greater accuracy and lesser processing time.

4. CONCLUSIONS

An image processing model is developed to detect and recognize textual information in digital videos. The video streams are decoded into frames and each frame is processed as an image. For each frame the developed model uses the EAST Text Detector for text detection and the Pytesseract for Optical Character Recognition(OCR) which recognizes the text and generates the results with better accuracy. The Proposed model overcomes the disadvantages of the existing model with faster processing time and better accuracy. The Proposed model can also recognise text in live video streams which was not possible in the existing model. The product implemented can be used in the various fields such as game design and development and animation studios.

REFERENCES

[1] W. Lu, H. Sun, J. Chu, X. Huang and J. Yu, "A Novel Approach for Video Text Detection and Recognition Based on a Corner Response Feature Map and Transferred Deep Convolutional Neural Network," in IEEE Access, vol. 6, pp. 40198-40211, 2018.

[2] Text Detection and Recognition: A Review by Chaitanya R. Kulkarni, Ashwini B. Barbadeka, 2017. International Research Journal of Engineering and Technology, IRJET.

[3] Lu T., Palaiahnakote S., Tan C.L., Liu W. (2014) Introduction to Video Text Detection. In: Video Text Detection. Advances in Computer Vision and Pattern Recognition. Springer, London.

[4] Rizdania and F. Utaminingrum, "Text detection and recognition using multiple phase methods on various product labels for visual impaired people," 2017 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, 2017, pp. 398-404.

[5] N. Anandhi and R. Avudaiammal, "Segmentation and recognition of text from image using pattern matching," 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2017, pp. 0066-0069.

[6] M. Bayomi, K. Levacher, M. R. Ghorab and S. Lawless, "OntoSeg: A Novel Approach to Text Segmentation Using Ontological Similarity," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 2015, pp. 1274-1283.