Lecture Summarization using Video Processing and Automatic Text Summarization

Janhavi Chadawar
MIT Academy of Engineering,
Alandi (D)
jschadawar@mitaoe.ac.in

Vivek Deshmukh
MIT Academy of Engineering,
Alandi (D)
vvdeshmukh@mitaoe.ac.in

Sahil Kharade
MIT Academy of Engineering,
Alandi (D)
sskharade@mitaoe.ac.in

Tushar Shelar
MIT Academy of Engineering,
Alandi (D)
tlshelar@mitaoe.ac.in,

Prof.Nilesh Bhandare
MIT Academy of Engineering,
Alandi (D)
ndbhandare@mitaoe.ac.in

***Abstract*: Nowadays machine learning has achieved a lot of success in technical fields. One of these is video processing but text is recognized which is in a computer readable format, our paper is focusing on detection of handwritten text and summarizing the document for which we found a better result with some methods which are clearly explained in this paper. This paper is following three modules to achieve the better result as from video selecting keyframes without losing whiteboard data, after selecting keyframes to identify the text which is a difficult task can be achieved. Preparing short notes with more accuracy is required for that this paper shows some ML libraries.These paper will give us a brief idea about which methods are used for modules to achieve a good accuracy.**

Keywords: RNN , Diagonal based feature extraction, CNN, LSA and Lex rank, Fpdf

I.   INTRODUCTION

In today's world, AI is changing the way how things work. One such application of this new age technology is Video Summarization i.e. finding most relevant data from frames of video and converting it to a summarized document. Automatic summarization of text is a method of making a document short with the help of software to create a summary which includes the important points present in the original document.Video summarization helps people to navigate through large videos and retrieve the important information or sentences which are most relevant to the requirements. In a general for processing video summarization system they select the image frames after a particular interval of time and extract that features and then most important frames are selected after analysing accumulated motion of the user and from the voice with comparing to prior frames. Most of this video processing basically relies on some features which are used globally such as color, texture, motion information, etc. Then Character recognition takes place which converts handwritten text into computer readable format after that last module works with the summarization. By using this technique, it will be way easier for students to summarize the entire lecture within a few minutes.

II.   LITERATURE SURVEY

To detect the handwritten text from video we need to follow some methods/modules for analysing the modules which fits our problem perfectly and give more efficient results. We have done surveys on some research papers. Though, basically we identified that we need to distribute project into 3 modules accordingly we surveyed research papers to get deep understanding of the methods for first module i.e. video processing in which we are selecting a key frame from video for text recognition we referred papers for video processing which are [1], [2]. These 3 papers are selecting key frames from video in their own manner with a maximum accuracy, from the video [1] fetches small changes done on whiteboard using the methods divide and conquer, RNN. It is able to identify the changes done on a whiteboard from comparing current frame with the next frame and selects the key frame. Which gives more accurate selection of key frames with the accuracy of 96.28%. Also, we referred to a paper [2] which detects the key frames from video with a simple method, in this paper they have implanted the 2 cameras while shooting video to increase the accuracy but methods used to identify them were not able to identify the key frames exactly.

The second module for our project is text recognition, key frames which we have selected from that we need to identify each character and symbol from frame for that er referred research paper[3], [4], [5], [6]. For detecting the text from frames text may be in standard format i.e computer readable format or may be in handwritten format which varies from person to person to identify the letters. Paper [3] identifies the text from the frame which is in standard format only,which is not suitable for the project to improve the accuracy. Handwritten text is identified with the methods which are stated in the papers [4], [5], [6] but, some of the methods just identify the characters not symbols, numbers as paper [4]. In paper [5] and [6] handwritten text is identified with symbols and numbers with average accuracy of 96% uses NLP. [6] able to identify characters which are present in cursive also. After identification of text from key frames our project is focused on making short notes for students for which we have distributed it as our third module i.e. text summarization for which we referred papers [7], [8], [9]. In this module our main focus is to reduce the data but the data needs to be meaningful which is not present in paper[7] as it converts text to audio and then summarization is time consuming also. Papers [8] and [9] both

give meaningful summarization but paper[7] able to perform summarization once at a time that to paper[8] can summarize multiple documents at a time which is a good approach using methods like LSA and LexRank.

III. METHODOLOGY

We have distributed our project into 3 modules:

1. Video Processing
2. Text Recognition
3. Text summarization

A. Video Processing- Recurrent Neural Network

RNN uses hidden states to identify the transferred frames from the past frames to the frame which is being processed. Basically it compares the current frame with the next frame data and then selects a key frame from it. Taking key Frames from the video. It is mostly used for data processing in a sequential manner. After selecting frames, we need to do some pre-processing in which we perform noise reduction, binarization, edge detection. Processing is explained in detail in fig. 1
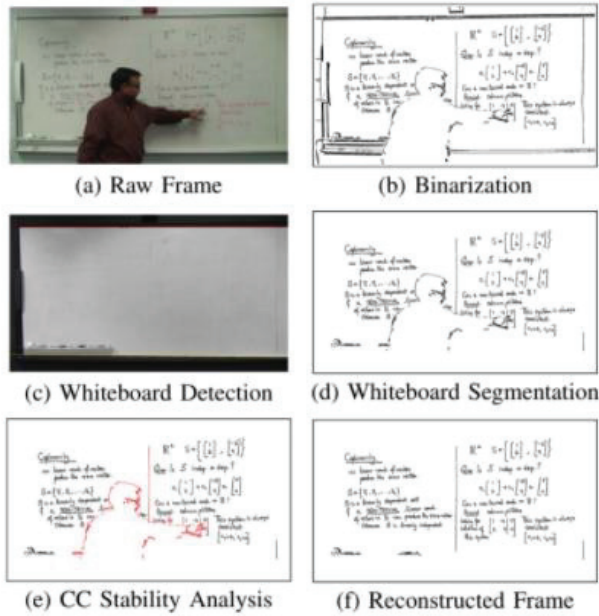

(a) Raw Frame   (b) Binarization

(c) Whiteboard Detection   (d) Whiteboard Segmentation

(e) CC Stability Analysis   (f) Reconstructed Frame

Fig. 1.   Overview of our video processing

B. Text Recognition -Diagonal based feature extraction

It is a segmentation method, an image having text then each character in the word is individually segmented. It is used for character recognition, each segmented character resize it into 100 * 100 pixels every 100*100 pixel is subdivided into 10 * 10 pixels zones. Each zone has 19 diagonals i.e. 1900 diagonals for 1 character. Some diagonals are empty so put their value to zero. Detailed diagonal based feature extraction is explained in fig 2
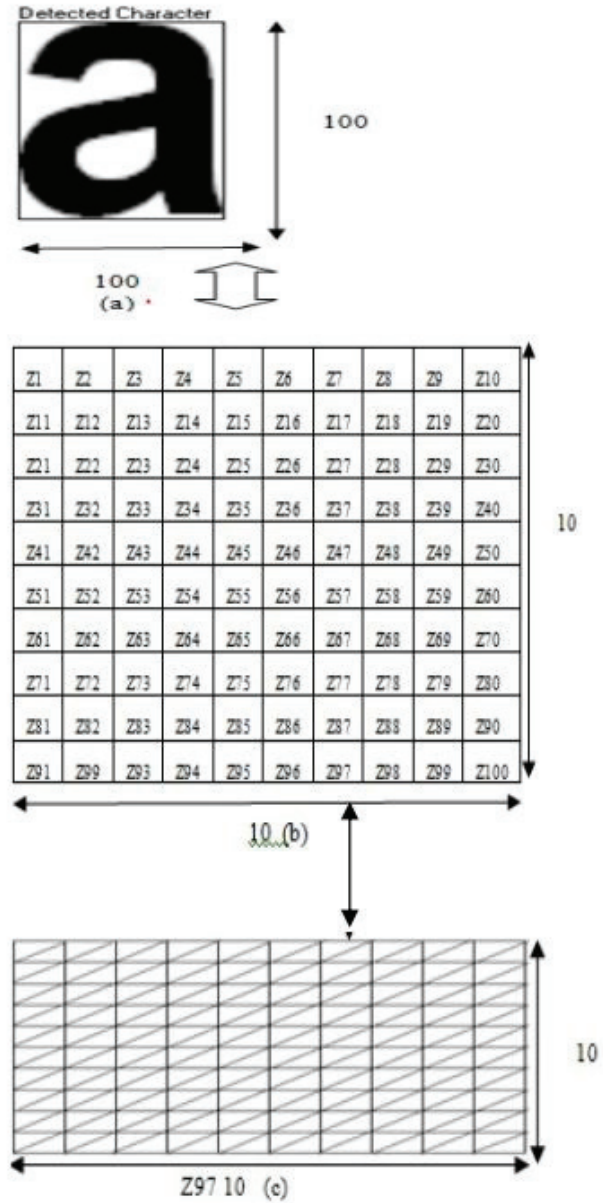

Fig. 2.   Diagonal Based Feature Extraction

C. Text Recognition-Convolution neural network

Deep learning base technique that is CNN used for feature extraction and feed forward network used for recognition or classification of text. Which works as shown in the fig 3. First layer performs the pre-processing. After that stretching of the image and then segmentation takes place. Last layer uses a diagonal based feature extraction method to recognize the characters which takes place by a feed forward network.
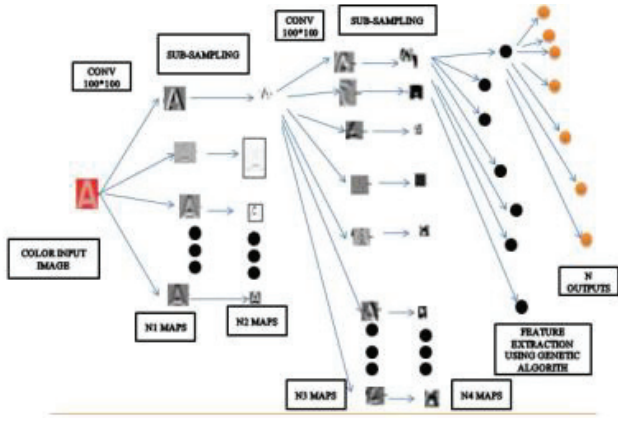
Fig. 3. CNN to identify text

D. Text Summarization- LSA and Lex Rank

LSA and Lex Rank used to convert unstructured to structured text. The stop words removal along with the frequency of words. Extract the important key-phrases in the text. It extracts the sentences according to the rank which has the highest rank extracted first from the input. Sequence of working is shown in detail in fig 4.

E. Convert Summary output into Pdf- Fpdf

Used to convert summarize notes into a proper readable format i.e. using Fpdf which is a python library used for creating pdf. FPDF is simple, small and versatile, with advanced capabilities and easy to learn, extend and maintain.
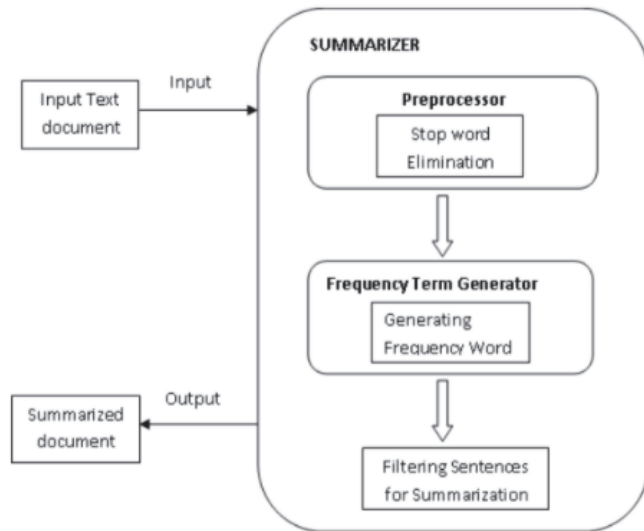

Fig. 4. LSA and LexRank working

IV.   WORK / MODULE DESCRIPTION

As per modules, project is distributed in three modules as follows:

- Video Processing
- Text Recognition
- Text Summarization

These three modules are interconnected and depend on the output of the previous module.

Design of the project is that first we will take input as a video of a whiteboard in a lecture. Then this input of video is taken as input in the first module, i.e. video processing module. In this we will use recurrent neural networks (RNN), to select the key frames from these videos.

Now after we select the key frames, we will perform image processing for each frame. As you can see here, we have used output of the previous module i.e. Video Processing module as input to the Text Recognition module. Here image processing algorithms are applied using python library OpenCV and to recognize the character the CNN is used.

After performing these two modules we will get the output that contains the entire text data from the lecture.

Now, at this stage we need to perform text summarization on the output of the previous model i.e. text recognition module. In the Text Summarization Module, we will use lexrank and lsa algorithms to perform extractive text summarization. We will perform both the algorithms and generate different summaries.

When we get both the summaries, we will perform a set operation which will combine the summary and we will get our final summarized output at this stage.

At last we will be using the Fpdf library to convert the given output to a pdf file and this will be the final output.

V.   PROPOSED PROJECT DESIGN

We have divided our project in three modules and the output of the previous module works as input for the next module.
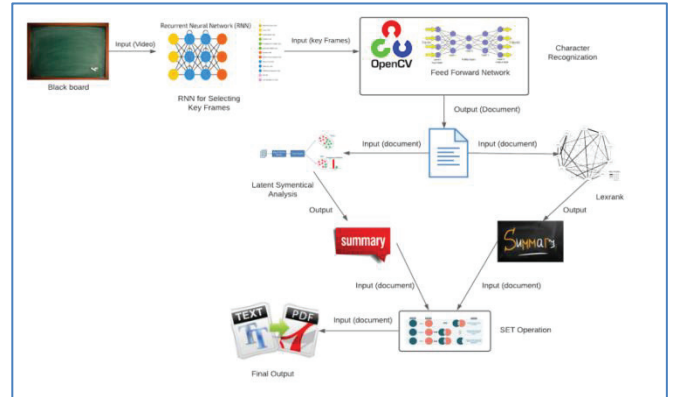

Fig. 5. System architecture

VI.   EXPECTED OUTPUT (FROM DESIGN)

Expected output for our project from the 3 modules are as follow:

1. Video processing
   - Selection of keyframes without losing a small amount of data.
   - Selected keyframes are with a good clarity to recognize text clearly.

2. Text Recognition
   - Handwritten text should be identified.
   - All kinds of text (italic, bold, handwritten, cursive) should be identified with accuracy.
   - Characters, numbers, symbols should be recognized.
3. Text Summarization
   - Meaningful extraction of sentences can be done.
   - While summarizing important data can not be lossed.
   - Summarized documents may cover all parts of the text with less no of data.
   - Students should be able to recall particular topics by reading this summarized document.

VII. CONCLUSION

There are many ways to identify the computer readable text with more accuracy but handwritten text recognition is a bit difficult task also, selecting keyframes from video with no loss of data is a laborious task too. To get a better module solution to achieve more accuracy we have found some methodology by surveying some research papers. RNN will be a more efficient method to select keyframes as it compares data between current and next frame to reduce the loss of data.For handwritten text recognition diagonal based feature extraction and CNN will give an accuracy upto 94% with this we can able to identify characters, numbers and symbols. LSA and LexRank can be used for text summarization. Hybridization of this may give more human-like summary i.e meaningful sentence formation can be done.

REFERENCES

[1] Kenny Davila, Richard Zanibbi. "Whiteboard Video Summarization via Spatio-Temporal Conflict Minimization". *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR).* Published: 29 January 2018

[2] Greg C. Lee, Fu-Hao Yeh, Ying-Ju Chen1, Tao-Ku Chang. "Robust handwriting extraction and lecture video summarization". *2016 Springer Science+Business Media New York.* Published: 9 February 2016

[3] Neha Agrawal, Arashdeep Kaur. "An Algorithmic Approach for Text Recognition from Printed/Typed Text Images".*2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence).* Published: 23 August 2018

[4] Pratik Madhukar Manwatkar, Shashank H. Yadav. "Text Recognition from Images". *2015 IEEE Sponsored 2nd International Conference on Innovations in Information,Embedded and Communication systems (ICIIECS).* Published: 24 March 2015

[5] Mohit Agarwal,Baijnath Kaushik. "Text recognition from images using Artificial Neural Network and Genetic Algorithms". 2*018 IEEE Sponsored 2nd International Conference on Innovations in Communication systems.* Published: 04 November 2013

[6] Polaiah Bojja , Naga Sai Satya Teja Velpuri, Gautham Kumar Pandala,S D Lalitha Rao Sharma. "Handwritten Text Recognition using Machine Learning Techniques in Application of NLP". *2019 International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. Published: 02 December 2019

[7] J.N.Madhuri, Ganesh Kumar.R. "Extractive Text Summarization Using Sentence Ranking". *2012 10th IAPR International Conference on Document Analysis and Recognition (ICDAR).* Published: 15 April 2013

[8] Nikhil S. Shirwadkar, Samidha Kulkarni. "Extractive Text Summarization using Deep Learning". *2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA).* Published: 15 June 2016

[9] Anusha Bagalkotkar, Ashesh Kandelwal, Shivam Pandey, S. Sowmya Kamath. "A Novel Technique for Efficient Text Document Summarization as a Service". *2013 Third International Conference on Advances in Computing and Communications.* Published: 19 December 2013