LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

Lecture Video Summarization

UE18CS390B – Capstone Project Phase – 2

Submitted by:

Durjoy Ghosh	**PES1201802124**
Vrushabh Chougale	**PES1201801495**
Shreya Deepak	**PES1201801099**
Surabhi Shivanand	**PES1201800849**

Under the guidance of
Dr. Uma D
Professor
PES University

August - December 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013) 100ft Ring Road, Bengaluru – 560 085, Karnataka, India

PES UNIVERSITY

LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

TABLE OF CONTENTS

1. Introduction

1.1 Overview

This document contains the low-level design of the project, where each module of the design and methods involved in it are explained and depicted with the help of several diagrams like use case diagram, master class diagram and system architecture diagram.

1.2 Purpose

The purpose of this document is to provide the detailed structure of the low-level design and detailed explanation of the modules which are included in the project and detailed details with the help of design diagrams such as use case, class, sequence and package or deployment diagrams.

1.3 Scope

The document tries to provide the maximum understanding regarding the modules present in the suggested project. The document doesn't include the inner depictions of the algorithms used.

2. Design Constraints, Assumptions, and Dependencies

2.1 Design Constraints

1) The video input must be in English.
2) The language assumption of the video is English and no other language.
3) Operating Environment should have the following configuration:

- RAM: 4 GB
- Internet or LAN connection
- Audio system and mic.
- Operating System: Windows/Ubuntu/MAC
- Python with the modules needed for translation such as OpenCV
- APIs for Text to Speech Conversion (Google API)
- Web Browser

2.2 Assumptions:

1) Active internet connection
2) The audio output is clear for better translation of audio.
3) The video is of any length.
4) The video is in the English language.
5) The user wants slides as a summary.

3. Design Description

There are a total of 6 proposed modules.

3.1 Shot Detection: During the lecture, professors generally design their slides in such a way that each point in the slides appears one by one as they speak. From each video split we extract the last image from the split which gives us the summary of that particular split.

3.2 Video Splits: Based on the shots detected, there are 'n' number of video splits generated. The number of video splits will depend on the size of the video and the drastic changes in the frames of the video.

3.3 Video to Audio: It is necessary to know what the professor is saying during the video lecture. Hence Audio is extracted from the video splits.

3.4 Audio to Text: The translation in the user specified language is initially done in the text format and later it is converted into audio format and passed to the user. This method is executed using useful APIs.

3.5 Keyword/Key phrase Extraction: Key Phrases have to be extracted for each text split as each text split can contain a different sub topic. To do so, convert the json file to raw text format. Remove punctuations and special characters etc. Remove stop words. Generate all 1-Grams and 2-Grams and record the frequency of each of them.

3.6 Summary Generation: The slides for the lecture will be an ordered sequence of images extracted from each of the video splits along with the key phrases extracted. This is the final output of the system which is of interest to the user.

4. System Architecture

4.1 System Architecture Diagram

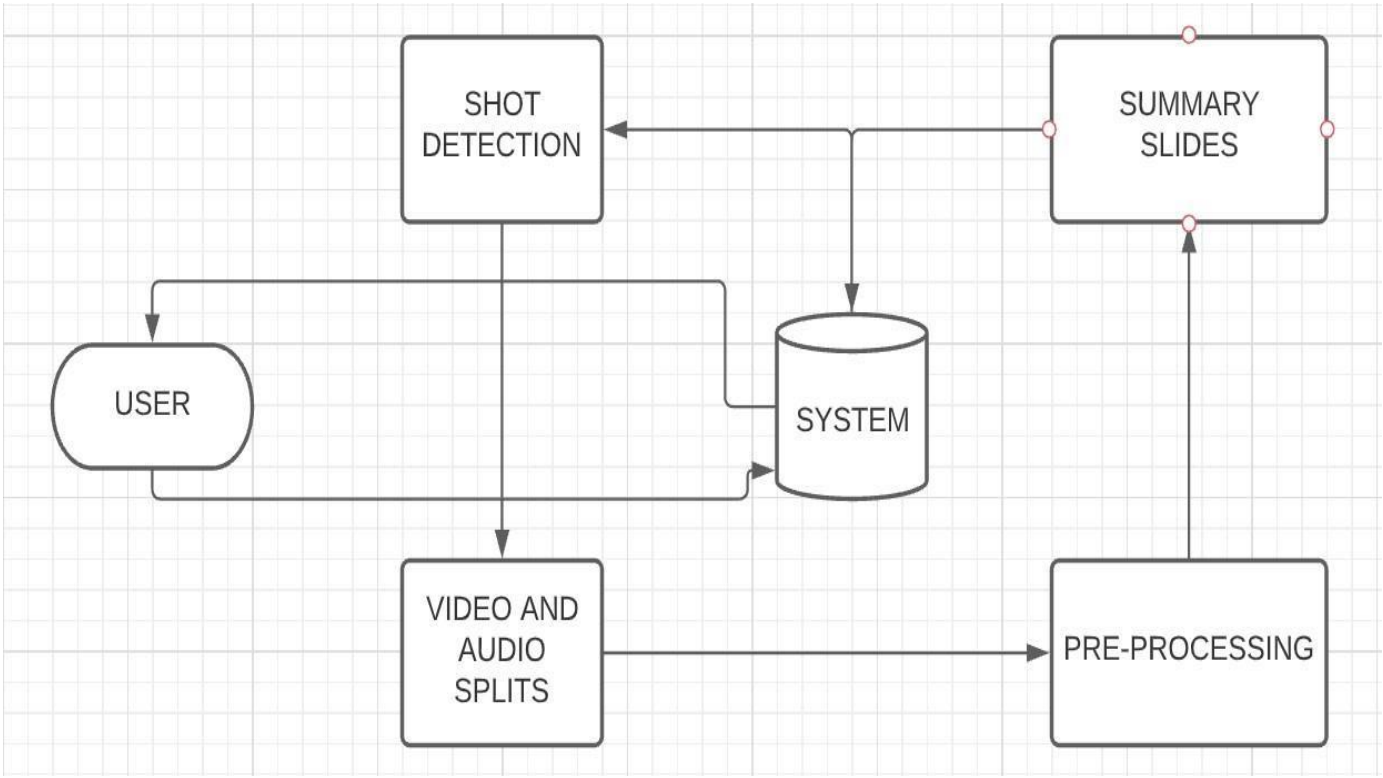

Fig1. System Architecture Diagram

4. 2 Master Class Diagram

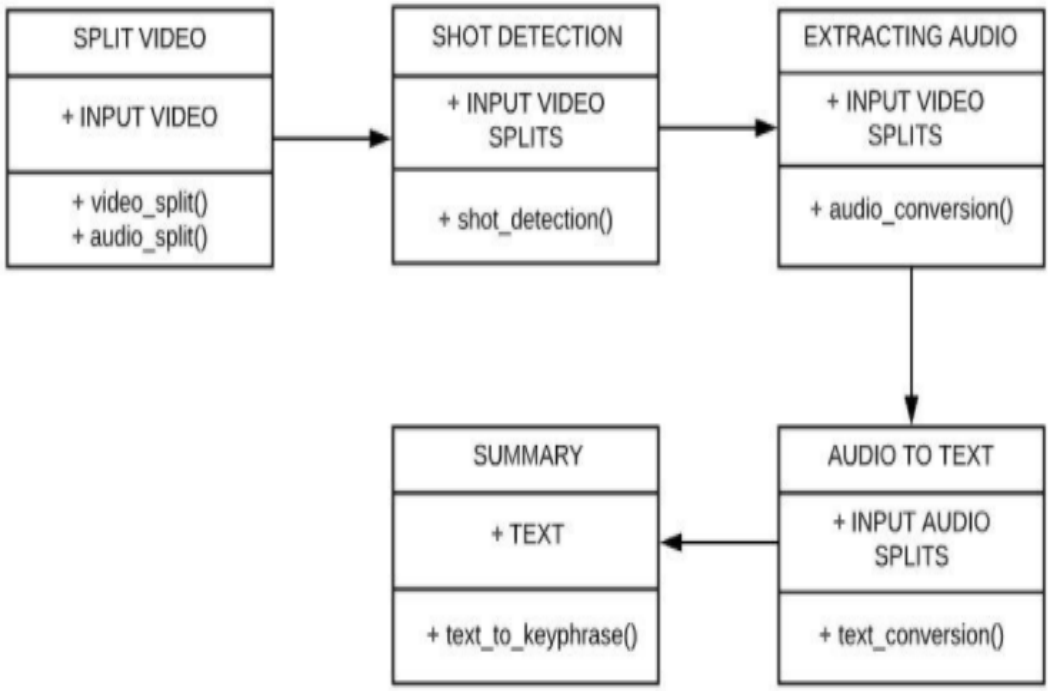

Fig2. Master Class diagram

5. System Description

5.1 Shot Detection : Shot detection is about finding the positions in a video where one scene is replaced by another with a different visual content. Shot transition detection is used to split up a film into basic temporal units called shots.

Shot: A series of interrelated consecutive pictures taken contiguously by a single camera and representing a continuous action in time and space. Each method for shot detection works on a two-phase-principle:

1. **Scoring:** Each pair of consecutive frames of a digital video is given a certain score that represents the similarity/dissimilarity between these two frames. Some common methods include -
    a. **Sum of Absolute Differences (SAD):** The two consecutive frames are compared pixel by pixel, summing up the absolute values of the differences of each two corresponding pixels.
    b. **Histogram Differences (HD):** HD computes the difference between the histograms of two consecutive frames; the histogram being a table that contains for each color within a frame the number of pixels that are shaded in that color.

2. **Decision:** All scores calculated previously are evaluated and a shot is detected if the score is considered high.
    a. **Fixed Threshold (Threshold-Aware Detection):** The scores are compared to a threshold which was set previously. If the score is higher than the threshold a shot is detected.
    b. **Adaptive Threshold (Content-Aware Detection):** The scores are compared to a threshold which considers various scores in the video to adapt the threshold to the properties of the current video.

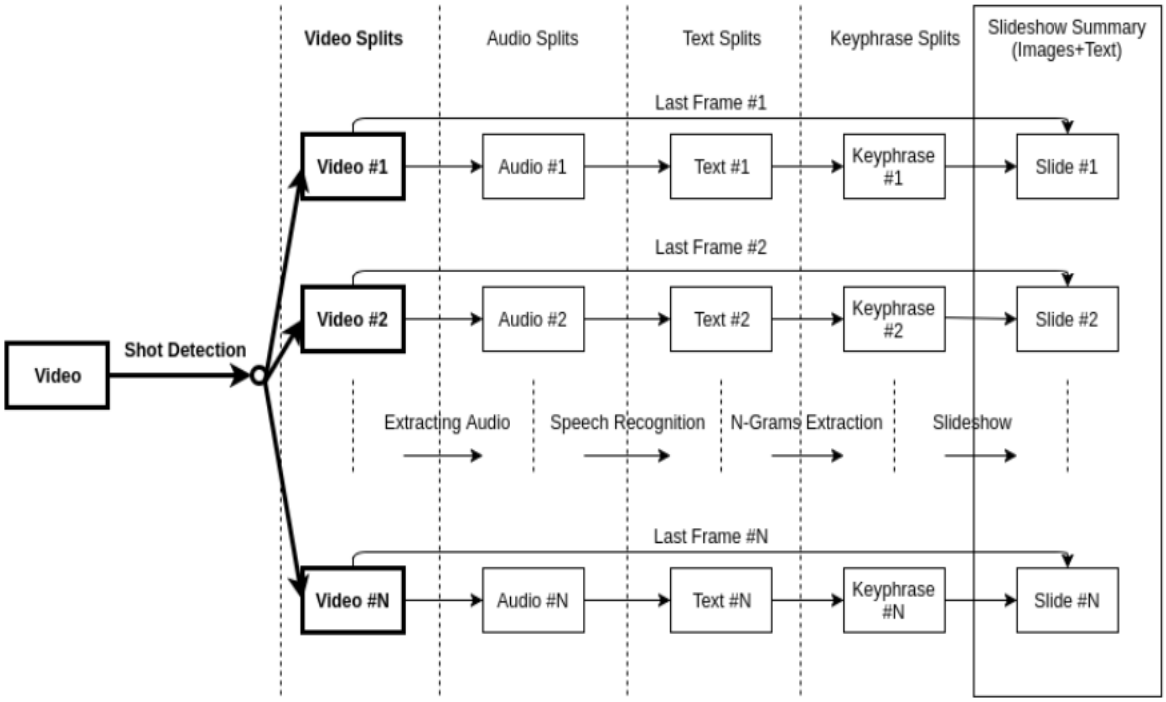

Fig3. Shot Detection

5.2 Video Splits : Speech recognition is the process by which spoken language can be converted to text by computers. Speech recognition systems require training, where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the speaker's voice and uses it to fine-tune the recognition of that person's speech.

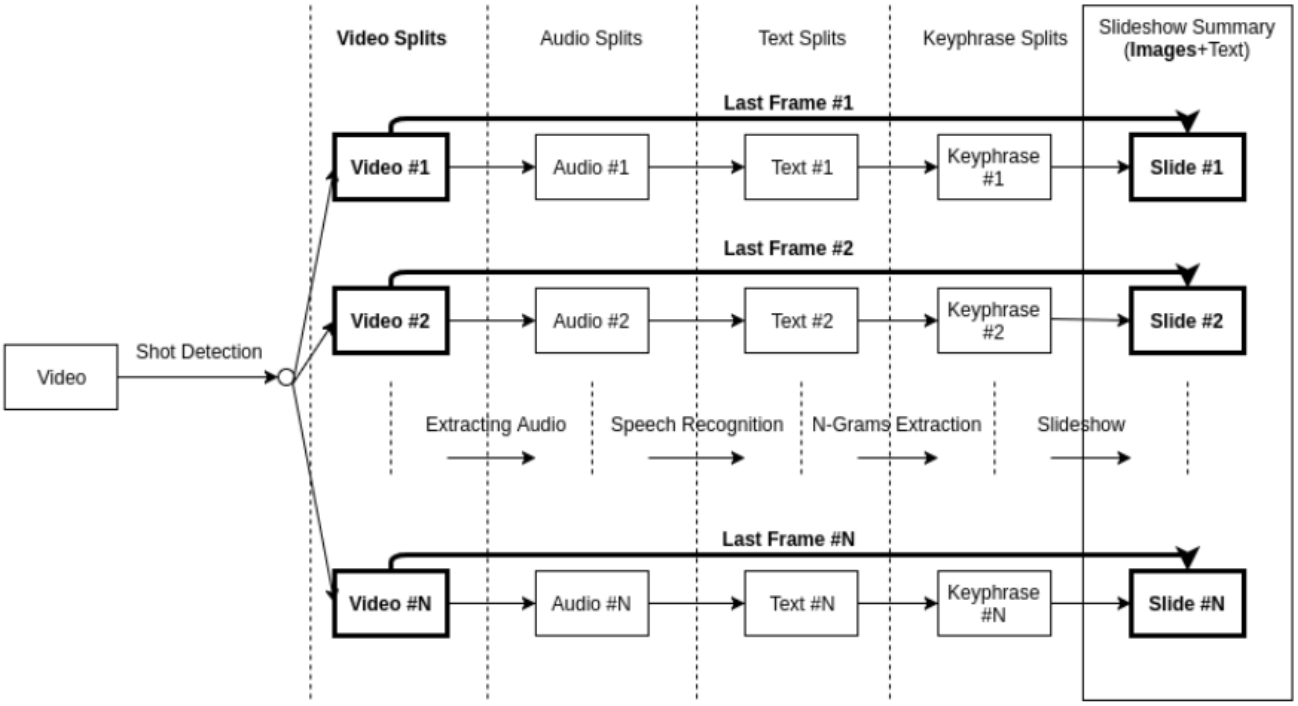

Fig4. Video Splits

5.3 Video to Audio : It is necessary to know what the professor is saying during the video lecture. Hence Audio is extracted from the video splits.

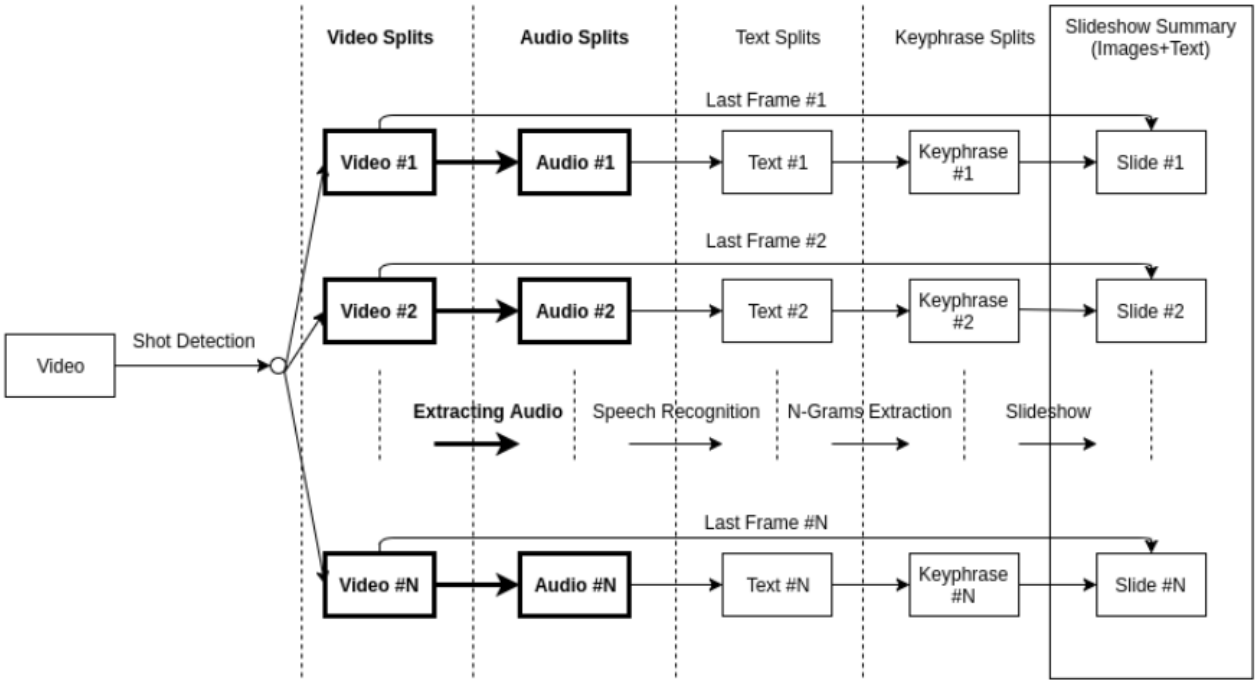

Fig5. Video to Audio

5.4 Audio to Text : It is important to convert the audio to text, so that we can extract the important keywords spoken for each of the sub-topics.

curl -s -H 'Content-Type: application/json' -H "Authorization: Bearer `gcloud auth print-access-token`" https://speech.googleapis.com/v1/speech:longrunningrecognize -d "{'config': {'encoding':'FLAC','sampleRateHertz': 44100,'languageCode': 'en-US','enableWordTimeOffsets': false},'audio': {'uri':'gs://bamboo-foundation-6245/AUDIO.flac'}}" > OP_ID
17

curl - H "Authorization: Bearer `./google-cloud-sdk/bin/gcloud auth print-access-token`" "https://speech.googleapis.com/v1beta1/operations/OP_ID" > TEXT . txt

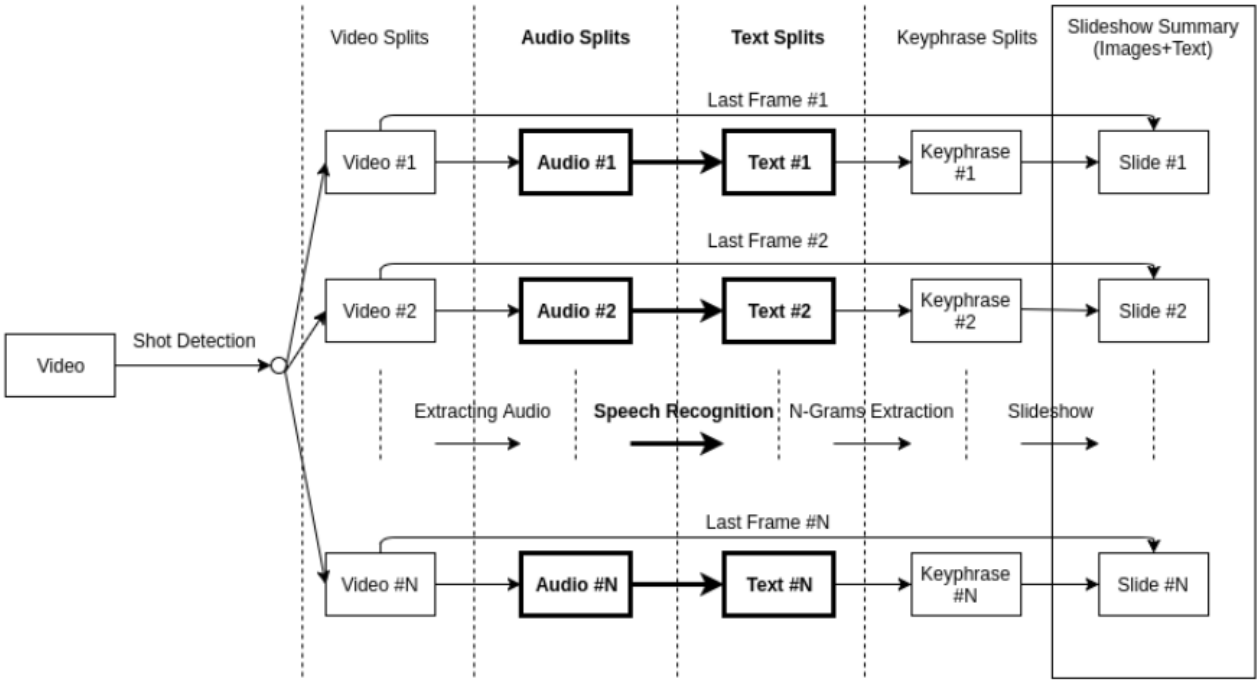

Fig6. Audio to Text

5.5 Keyword Extraction : To extract the keyphrases of each sub-topic, we need to do so for each text split individually.

1. Convert Json to Raw Text.
2. Remove special characters, punctations etc.
3. Remove stopwords.
4. Generate all possible 1-Grams and 2-Grams, and record the frequency of each of them.

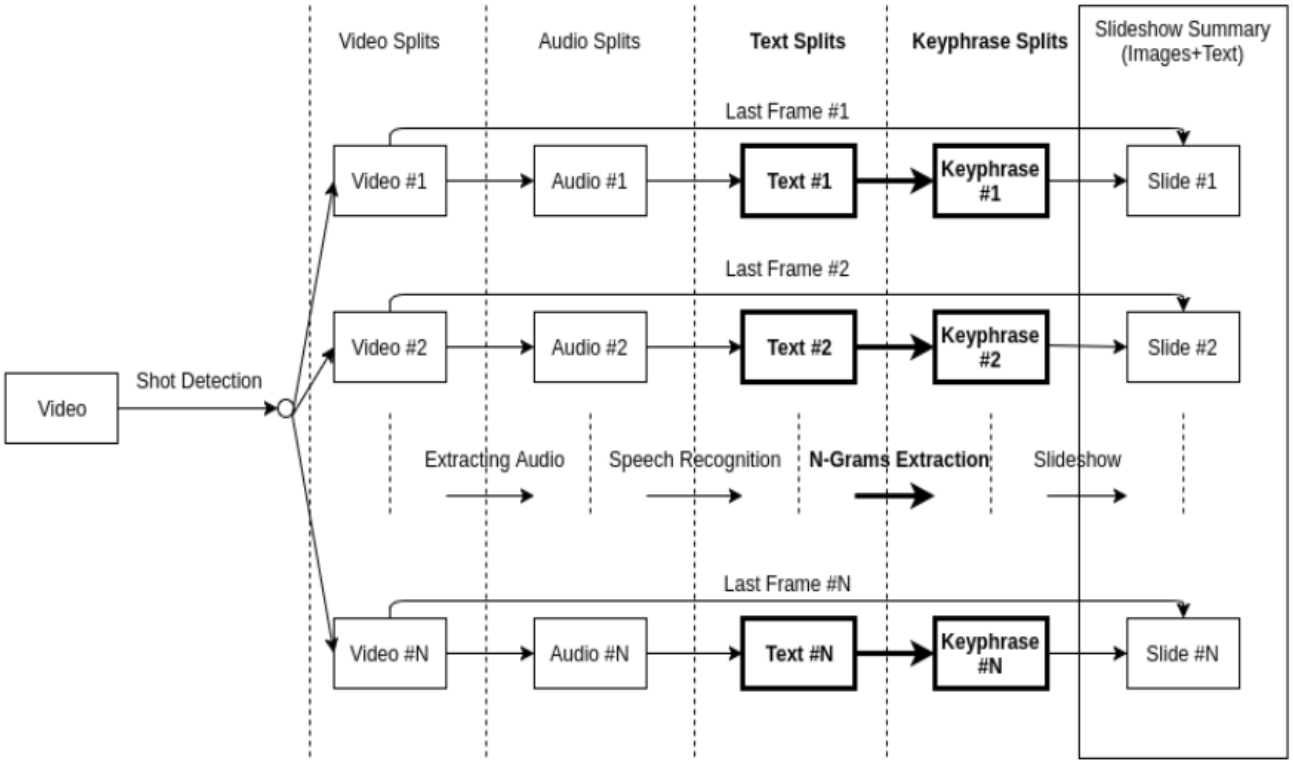

Fig7. Keyword Extraction

5.6 Summary Slides : The slides for the lecture will be an ordered sequence of images extracted from each of the video splits along with the keyphrases extracted from the text corresponding to the audio of each of the video splits.

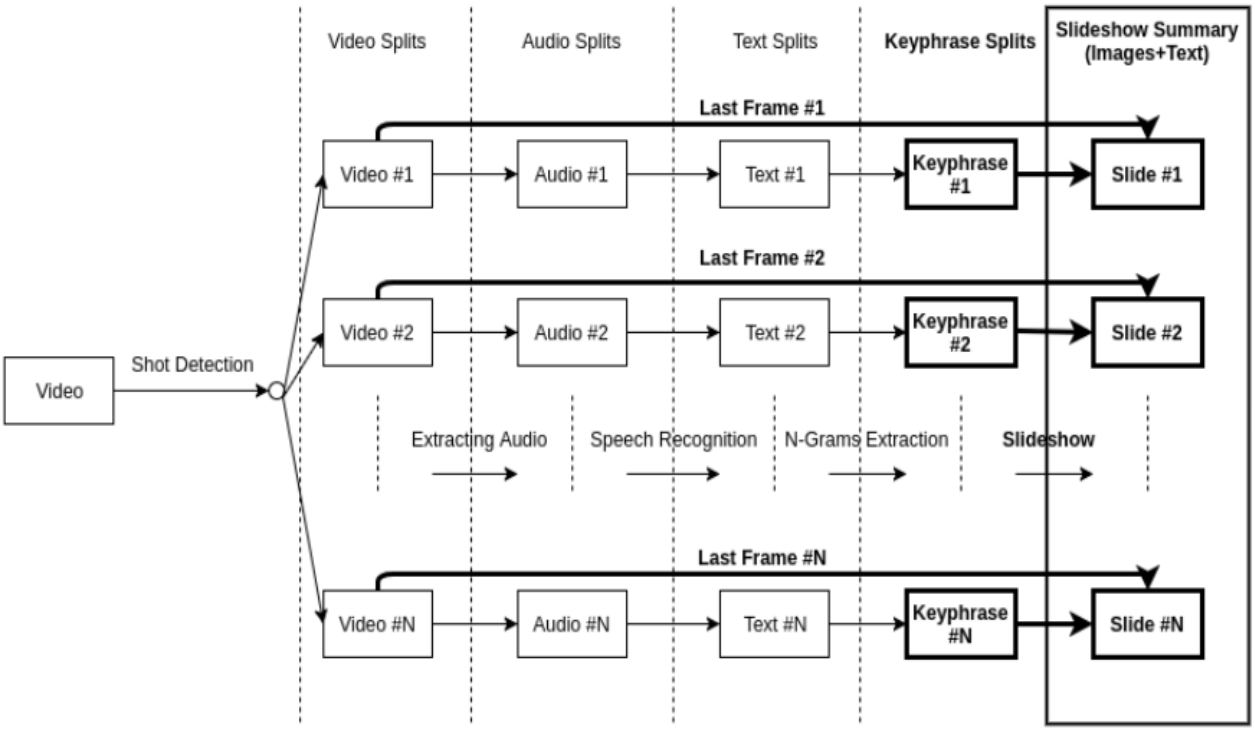

Fig8. Summary Slides

standard language (ie. English) in our case . By doing this , the language identification part can be restrained. This standard language is considered as the base for the translation into the user-specified language as this will be the input for the translating module.

4.1.1. Use Case Diagram

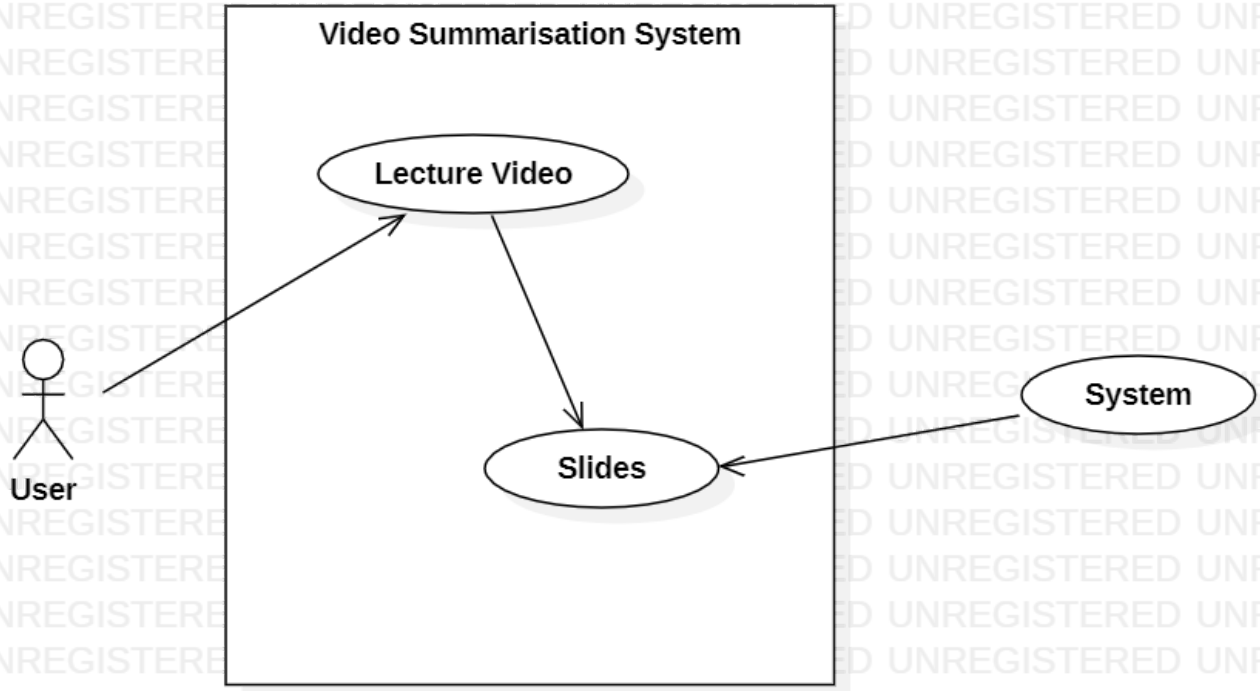

Fig9. Use Case Diagram

Use Case Item	Description
Video to Slides	The video is given to the system by the user and slides are generated as a summary of the video.

Appendix A: Definitions, Acronyms and Abbreviations

Keywords : Important words of interest
Frames : The stagnant capture of the video
User : It refers to the customers
API : Application Programming Interface

Appendix B: References

For the diagrams creation we have used the following softwares:

- Star UML

Appendix C: Record of Change History

#	Date	Document Version No.	Change Description	Reason for Change
1.	22-09-2021	1.0	Shot detection and Split Video	Initial Commit
2.	24-09-2021	1.1	Changes to diagrams	Deployment diagrams are wrong
3.	25-09-2021	1.2	Final Commit	Change in description of modules