COMSATS University Islamabad, Pakistan

Lecture Video Indexing through OCR

A project presented to
Comsats University Islamabad, Attock Campus

In partial fulfillment

of the requirement of the degree of

Bachelor of Science in Computer Science (2017-2021)

By

Rashid Mehmood CIIT/SP18-BCS-031/ATK

Asad Munir CIIT/FA17-BCS-107/ATK

DECLARATION

We hereby declare that this software, neither whole nor as a part has been copied out from any source. It is further declared that we have developed this software and accompanied the report entirely based on our efforts. If any part of this project is proved to be copied out from any source or found to be a reproduction of some other. We will stand by the consequences. No portion of the work presented has been submitted of any application for any other degree or qualification of this or any other university or institute of learning.

Rashid Mehmood

Asad Munir

CERTIFICATE OF APPROVAL

It is to certify that the final year project of BS (CS) **"Lecture video Indexing through OCR"** was developed by **Rashid Mehmood (CIIT/SP18-BCS-031/ATK)** and **Asad Munir (CIIT/FA17-BCS-107/ATK)** under the supervision of **"Mr. Najam Dar"** and that in his opinion; it is fully adequate, in scope and quality for the degree of Bachelors of Science in Computer Sciences.

Supervisor

External Examiner

Head of Department (Department of Computer Science)

Acknowledgment

All praise is to Almighty Allah who bestowed upon us a minute portion of His boundless knowledge by virtue of which we were able to accomplish this challenging task.

We are greatly indebted to our project supervisor “Mr. Najam Dar”. Without his supervision, advice, and valuable guidance, completion of this project would have been doubtful. We are deeply indebted to them for their encouragement and continual help during this work.

And we are also thankful to our parents and family who have been a constant source of encouragement for us and brought us the values of honesty & hard work.

Rashid Mehmood

Asad Munir

Table of Contents

Table of Figures:

CHAPTER 1

INTRODUCTION

INTRODUCTION:

Our project "**Lecture video indexing through OCR**" is basically a desktop application in which we can search our desired keywords from a recorded video lecture. Like we have a recorded lecture in which presentations slides are also included. In this project user have 1 library which is organized and we have data of our particular user in our database. When user add or import a video the data of that video stores in the database and then whenever user want to access that video it will be very easy for them. This is a very unique kind of project and there is a need for this type of project as well because there are a lot of difficulties and problems in online lectures. You know now day's teachers take online classes due to COVID-19 pandemic and most of the students didn't take their classes at that time so after the class they used to listen the online recorded lectures so you know the lecture is so lengthy and we need enough time to listen the full lecture. So if we want to learn something specific we can just write our keyword in the search box and press enter for our desired result. After that we can get our desired keywords and their time stamps. And if user selects some specific video then users have an option that they can use a keyword to get their desired result.

The system is being designed to use some latest algorithms and tools to cope with the problem of searching for the required information in a lecture. The system will use the OCR to search for the required information inside the video. The proposed system will use various algorithms for various purposes like, for key frames selection, for text detection, for scanning the text etc. The proposed system will get the keyword from the user as input and will provide the time slot if the keyword found in the lecture.

As we know, nowadays a lot of professional share their lecture videos on World Wide Web but the problem is to get something of interest. In this busy world, nobody has time to go through whole videos. As per the study of McKinsey and IDC, in average, 20% of the time wasted by the knowledge workers just in search of particular information they want. This is a time-consuming process, and in the worst case, they have to go through the whole video.

BRIEF:

Lecture video indexing thorough OCR is a desktop based system developed by using the Html, Css, Python and Javascript languages. It has the following goals:

1. First of all to give facility to anyone from which they can find something of their interest.
2. To save time.
3. Get relevant information in short period of time.
4. To avoid wasting with the irrelevant information.

5. To facilitate knowledge workers primarily (can also be helpful for others in special cases)
6. Primary purpose is to work on the lecture videos that also have slide representation.
7. Using modern tools to facilitate users.
8. To create outlines for students through which students can easily get an idea about tutorials.
9. To reduce/minimize human effort
10. Converting complexities to simplicities
11. Enjoying a lot of features without internet
12. To provide accurate results by using the well-known algorithms.
13. Relevance to course modules.

Web Technologies: In the course of Web Technologies, we have studied about the Html and Css languages that will us in the designing of this project.

Report Writing Skills: This course is about learning how to write reports and other formal documentation and in our project we need to write our documentation so this course is helping a lot in this task.

Visual Programming: In the course of Visual Programming, we have studied the C# language that will help us in the development of the project.

Human-Computer Interaction: An interactive system is easy and comfortable for the user to use the system and understand it easily and this course is all about designing an interacting system following standard rules.

Database System: To avoid the scanning of video again and again, database is a crucial part of this project, so the database course that we studied is also some sort of supporter for us.

Project Background:

Due to the advancement in computer vision and image processing, we can see experts are trying to solve all the problems using the great algorithms. For example, we can consider the object tracking, people counting, vehicle counting and video summarization. These are just few problems that computer vision is working, there are so many other problems. Due to this extraordinary performance of computer vision in every field, we have decided to use computer vision to solve a problem. As we know, nowadays a lot of professional share their lecture videos on World Wide Web but the problem is to get something of interest. In this busy world, nobody has time to go through whole videos. As per the study of McKinsey and IDC, in average, 20% of the time wasted by the knowledge workers just in search of particular information they want.

This is a time-consuming process, and in the worst case, they have to go through the whole video. Video search is the most challenging problem of these days. It consumes a lot of time when you start searching for something of your interest in a video. As per the study of McKinsey and IDC, in average, 20% of the time wasted by the knowledge workers just in search of particular information they want. So, this time can only be utilized if we get some solution for this problem. If we are going to design the system so it should be capable enough to show all the text shown on screen. So our proposed system is just to overcome this problem by using the modern tools like OCR (Optical Character Recognition) which provides the facility of recognizing the characters. So hope the proposed system is going to reduce this problem by large fraction. The reason behind choosing OCR is because of its performance these days. The system will perform the segmentation to the video, then removal of elements that can affect the OCR in scanning the text. Text detection and after that text recognition will be performed. Once we get the script, it will be transferred to the database to avoid scanning in the future.

Literature Review:

A literature review surveys books, scholarly articles, and any other sources relevant to a particular issue, area of research, or theory, and by so doing, provides a description, summary, and critical evaluation of these works in relation to the research problem being investigated. In literature review we make a comparison of our project with the existing projects and compare the pros and cons of existing projects with our project.

So we have done a lot of research on Lecture Video Indexing through OCR and it turns out that the Lecture Video Indexing through OCR is only fully operational on web the extra beneficial software, which is called panopto. After understanding panopto, it was discovered that it lacked something's.

First of all, panopto create an organic environment for sharing and expanding knowledge through on-demand video that lets your people learn like they do on YouTube. Another thing that we can compare is resources. Larger servers like panopto requires many resources that are not really easy to afford. It requires internet as well which also makes it time-consuming. Our system uses a local database that is a positive point because it does not require a server to handle a lot of data. Users already have the lecture videos so application will just demand for the path. As panopto requires internet so the network connection can also be the problem. Our application will be a lightweight desktop application that users can install on their personal computers. It doesn't contain any heavy database too that can decrease the performance.

Methodology and Software Lifecycle:

We use the iterative approach to develop this application. The iterative method is a way to breakdown the development process into small patches so that instead of developing the whole software at once which is difficult to develop, we develop in small steps. Following the iterative approach, we will be implementing the small functionalities time to time and will get a system wiith all required functionalities in the end.

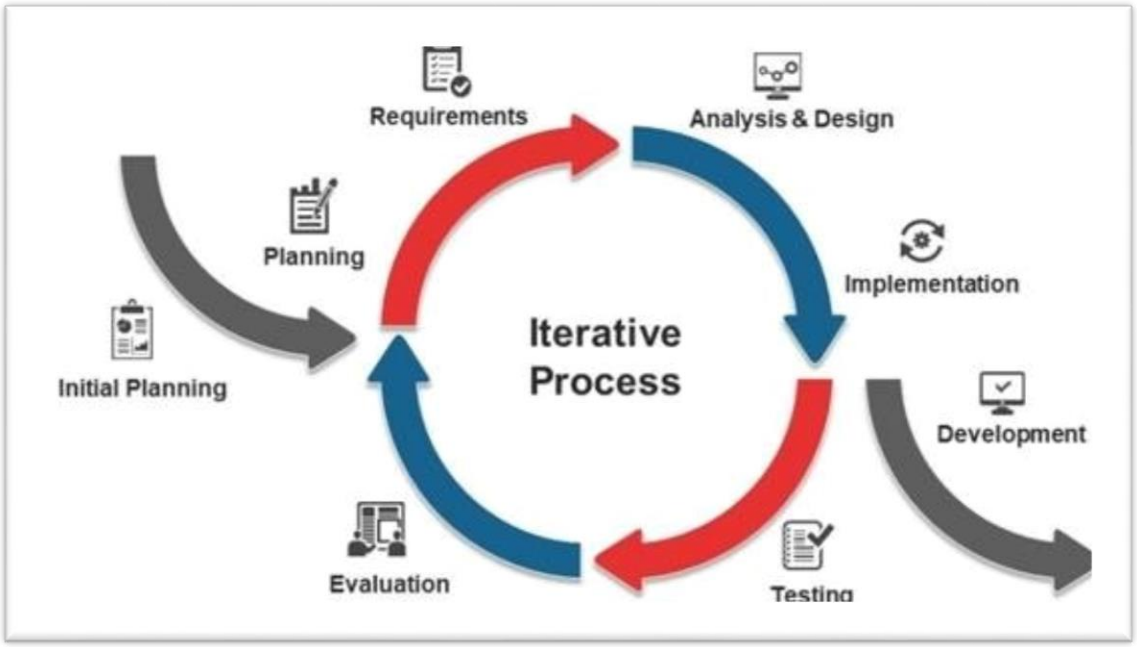

Figure 1.1 1 Iterative Model

The rationale behind the selected methodology:

The iterative process is the way to breakdown the whole process into smaller chunks in order to develop the complete software in one go. The reason behind this approach is that it is easily measurable and small chunks are easy to test and debug as compare to developing the whole software and then testing it, at this point debugging becomes difficult. Another good feature of iterative process is that all of its sub processes show some kind of functionality that keeps a developer motivated.

CHAPTER 2

Problem Statement:

As we know, nowadays a lot of professional share their lecture videos on World Wide Web but the problem is to get something of interest.

Problem definition

In this busy world, nobody has time to go through whole videos. As per the study of McKinsey and IDC, in average, 20% of the time wasted by the knowledge workers just in search of particular information they want. This is a time-consuming process, and in the worst case, they have to go through the whole video.

Video search is the most challenging problem of these days. As per the study of McKinsey and IDC, in average, 20% of the time wasted by the knowledge workers just in search of particular information they want. So, this time can only be utilized if we get some solution for this problem. If we are going to design the system so it should be capable enough to show all the text shown on screen. So our proposed system is just to overcome this problem by using the modern tools like OCR (Optical Character Recognition) which provides the facility of recognizing the characters. So hope the proposed system is going to reduce this problem by large fraction. The system will perform the segmentation to the video, then removal of elements that can affect the OCR in scanning the text. Text detection and after that text recognition will be performed.

Deliverables and Development Requirements:

Following is the list of deliverables of our project:

- Video Gallery
- Video Searcher
- Video Player
- Keyword timeslot provider
- Grid and List view for the video gallery

Development requirements include the following software requirements:

- **IDE :** Visual Studio , Electron, Node.js
- **Database:** Sqlite
- **Programming Languages:** Html, Css, Javascript, Python

CHAPTER 3

REQUIREMENT ANALYSIS

Requirement Analysis:

Software Requirement Analysis (SRS) is not an ordinary document because it provides a basic understanding of functional as well as non-functional requirements. We can consider it as the starting point of the project because it searches for the well-known and algorithms with the best performance. With the help of SRS, the users will get their benefits.

- Study existing systems like this if any.
- Search for the well-known and algorithms with the best performance.
- Understand the working of algorithms.
- Choosing the best algorithms among all.
- Designing a prototype to understand the environment of a system.
- Machine Learning Implementation.
- Testing the model.
- Building the interface.
- Linking the interface and the model.

Functional Requirements:

Functional requirements are the description of the service that the software must offer. It describes a software system or its component. A function is nothing but inputs to the software system, its behavior, and outputs. It can be a calculation, data manipulation, business process, user interaction, or any other specific functionality which defines what function a system is likely to perform.

Functional requirements of our project are as follows:

Video Segmentation:
In video segmentation, the frames are decomposed from the video and key frames are identified. Key frames show the different content than the previous frames. When proper segmentation is done the exact key frames are selected without missing any important texts. Generally for video lectures the text contained in frame can be used for key frame identification.

Key frame selection:

Select the frames from the segmentation process. Now, Select a key frame for further processing, number of key frames can be similar. From similar key frames select only one and discard others.

Feature extraction:
There are two ways to handle the feature first is spatial and second is temporal. Color, shapes, edges are spatial feature. Spatial feature is nothing but low level feature .Motion and audio is temporal feature.

Text recognition:
Following are the steps required for the text recognition using OCR.

- Get input path for processing
- Load images from input path
- Convert input image into low scale 312 by 256
- Extract characters
- Set upper/lower case
- Save character set

Non Functional Requirements:

The non-functional requirements describe how the system should work. It also specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system.

Following are the non-functional requirements of our project:

Performance:
Performance describes the efficiency of the system. Our project will be very good in performance like it will help users to find their thing of interest in just seconds and they will get their desired result.

Portability:
Portability means that the user must not be restricted to use the application in a specific area/place. Our desktop application is very portable because it allows the users to use it everywhere.

Usability:
Usability means that the application should be easy to use. The user should not feel any difficulty in using this system. In our project of "Lecture Video Indexing through OCR", it is very easy for the user to search their desired keyword. They just have to import video in player so there is nothing difficult in it.

Compatibility:
Compatibility is described as the user can use the application in all the devices and it must not be restricted to some specific devices. Our desktop project is very compatible with all the devices and it can run on every windows, Mac and Linux devices.

Security:
Security is one of the most important factors in the project that must not be compromised at any cost. In our project all the data is safe and secured. All the information that a user gives is completely secured.

CHAPTER 4

DESIGN ARCHITECTURE

Project Design:

Project design is an early phase of the project where a project's key features, structure, criteria for success, and major deliverables are all planned out. The point is to develop one or more designs which can be used to achieve the desired project goals. The project design phase might generate a variety of different outputs, including sketches, flowcharts etc.

- Activity Diagram
- Use Case Diagram
- Workflow Diagram

Use Case Diagram:

A use case diagram is a way to summarize details of a system and the users within that system. It is generally shown as a graphic depiction of interactions among different elements in a system. Use case diagrams will specify the events in a system that how those events flow.

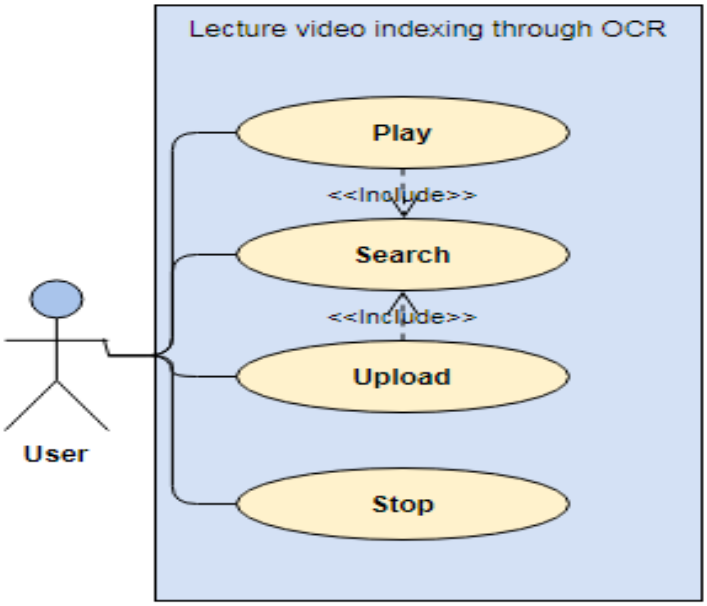

Figure 4. 1 UseCase Diagram

Activity Diagrams:

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc. So, our activity diagram tells about the whole activity of importing a video, extraction of the text and taking the user to the keyword searched.

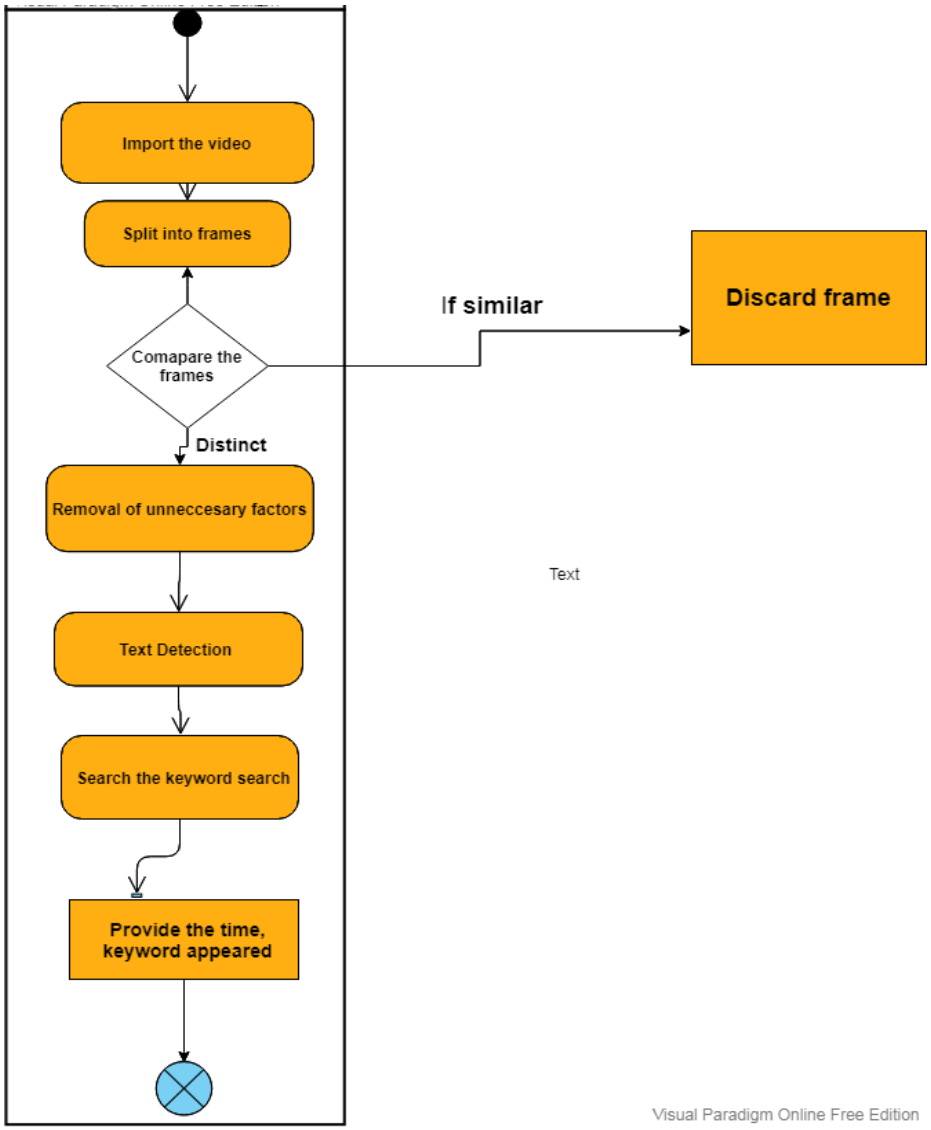

Figure 4. 2 Activity Diagram

Workflow Diagram

A workflow diagram defines the proper working of the system. It gives information about the steps involved in the operation as per their position. As by looking at our workflow diagram, we can understand the system well, like we will get so many frames out of the video, second step is to extract the key frames out of it and discarding the redundant ones, first 2 steps are the part of lecture segmentation. After that, our task is text detection, once the text is detected, some image enhancement techniques can be applied in order to improve the text recognition. In the end, we can store the script per frame in our database to make it more meaningful.

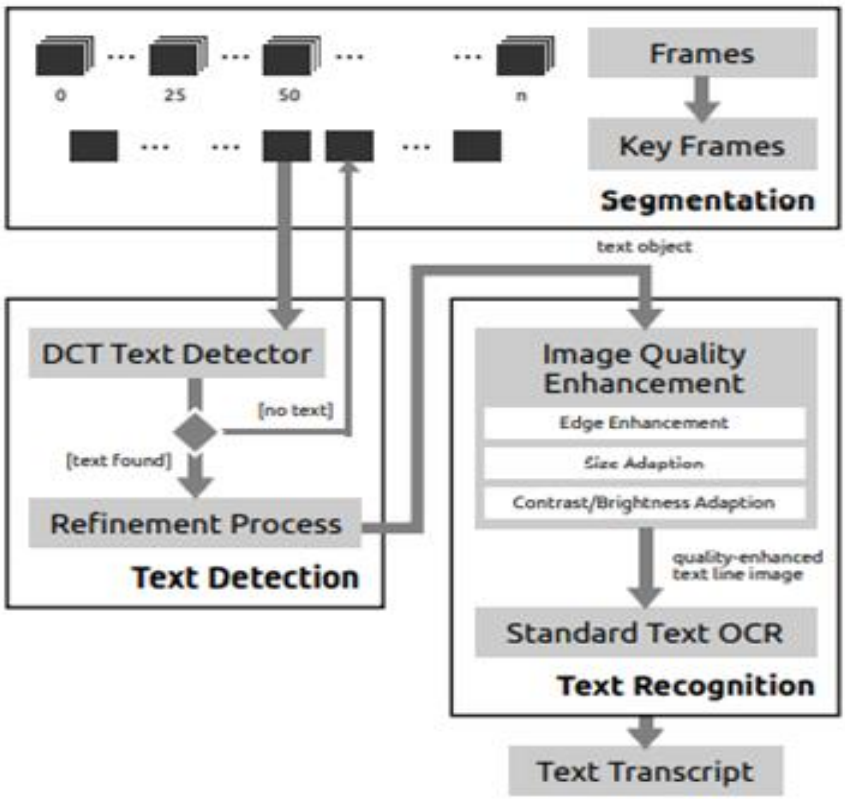

Figure 4. 3 Workflow Diagram