International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI-Based Video Summarization and Quiz Generation

V. Bhavya Sri [B200484], J. Manasa [B201106], D. Saraswathi [B201234]
Department of Computer Science and Engineering
Rajiv Gandhi University of Knowledge Technologies, Basara (IIIT Basara)
Email: b200484@rgukt.ac.in, b201106@rgukt.ac.in, b201234@rgukt.ac.in
Under the Guidance of Ms. Sindhuja
Assistant Professor
Department of Computer Science and Engineering
Rajiv Gandhi University of Knowledge Technologies, Basara (IIIT Basara)

***Abstract*—The rapid growth of video content on digital platforms has created a need for automatic video summarization systems. This paper presents an AI-based video summarization and quiz generation system that extracts audio from video, converts speech to text, extracts keyframes, generates summaries, and automatically creates quiz questions. The proposed system uses speech recognition, computer vision, and natural language processing techniques to generate meaningful summaries and educational quizzes. This system is useful in e-learning, video indexing, and content understanding applications.**

***Index Terms*—Video Summarization, Artificial Intelligence, Speech Recognition, Keyframe Extraction, Natural Language Processing, Quiz Generation**

I. Introduction

In recent years, there has been a rapid increase in multimedia content, especially video content, on platforms such as YouTube, online learning platforms, and social media. Watching full-length videos to understand the content requires a significant amount of time. Therefore, automatic video summarization systems are needed to extract important information from videos and present it in a short and meaningful format.

Video summarization is the process of creating a short summary of a long video by extracting important frames, scenes, and audio information. The summary helps users understand the main idea of the video without watching the entire video. Video summarization can be classified into two types: static video summarization and dynamic video summarization. Static video summarization generates keyframes, while dynamic video summarization generates a short video skim of the original video.

Artificial Intelligence (AI) techniques such as Machine Learning, Deep Learning, Natural Language Processing (NLP), and Computer Vision are widely used for automatic video summarization. Computer Vision techniques are used to process video frames and extract important keyframes, while Natural Language Processing techniques are used to summarize the text obtained from speech recognition. Speech recognition technology is used to convert audio content into text format, which can then be processed for summarization and quiz generation.

In addition to summarization, quiz generation is also an important feature for educational applications. Automatic quiz generation helps students test their understanding of the video content. The quiz questions are generated from the summarized text using Natural Language Processing techniques. This makes the system useful for e-learning platforms, online courses, and lecture video summarization.

This paper proposes an AI-based video summarization and quiz generation system that integrates speech recognition, computer vision, and natural language processing techniques into a single system. The system takes a video file or YouTube video as input and performs audio extraction, speech-to-text conversion, keyframe extraction, text summarization, and quiz generation. The final output of the system includes transcript, keyframes, summary, and multiple-choice quiz questions.

The main objective of the proposed system is to reduce the time required to understand video content and improve learning efficiency. The system is useful for students, teachers, researchers, and professionals who want to quickly understand video content and test their knowledge using quizzes.

The remainder of the paper is organized as follows. Section II describes the literature review. Section III explains the proposed system. Section IV presents the methodology used in the system. Section V describes the results and discussion. Section VI presents the conclusion and future work.

II. Literature Review

Video summarization has gained significant attention from researchers due to the rapid growth of multimedia and online video content. The main goal of video summarization is to reduce the length of the video while preserving important information and the overall meaning of the video. Many researchers have proposed various techniques for automatic video summarization.

Earlier video summarization methods were based on traditional techniques such as shot boundary detection, frame sampling, and clustering algorithms. These methods used low-level features such as color histogram, texture, edge detection, and motion information to extract important frames. However,

these methods were not able to understand the semantic meaning of the video and often produced inaccurate summaries.

Many researchers have worked on video summarization using deep learning techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM). CNN models are used to extract spatial features from video frames, while RNN and LSTM models are used to capture temporal dependencies between frames. These methods help in identifying important segments in the video more accurately compared to traditional methods.

Previous research shows that video summarization can be classified into two types:

- Static Video Summarization (Keyframe extraction): In this method, important frames are extracted from the video to represent the video content in the form of images.
- Dynamic Video Summarization (Video skimming): In this method, important video segments are selected and combined to create a short summary video.

Recent research focuses on deep learning-based video summarization, which provides better performance compared to traditional methods. Attention mechanisms and transformer models are used in modern video summarization systems to focus on important parts of the video. These models improve the quality of the summary by understanding the context and semantic meaning of the video.

In addition to video summarization, automatic quiz generation is another important research area in Natural Language Processing. Automatic question generation systems generate questions from text using techniques such as keyword extraction, sentence transformation, and semantic analysis. These systems are useful in educational applications and e-learning platforms.

Several researchers have developed systems that combine speech recognition, text summarization, and question generation for educational purposes. Speech recognition is used to convert video audio into text, Natural Language Processing is used for summarization, and question generation algorithms are used to generate quizzes from the summarized text.

From the literature review, it is observed that most existing systems focus only on video summarization. Very few systems combine video summarization and quiz generation into a single system. Therefore, the proposed system integrates speech recognition, keyframe extraction, text summarization, and quiz generation into a single system, which makes it more efficient and useful for students and educational applications.

III. System Architecture

The system architecture is shown in Figure 1.

IV. Methodology

The proposed AI-based video summarization and quiz generation system consists of multiple modules such as audio extraction, speech recognition, keyframe extraction, text summarization, and quiz generation. The overall workflow of the system is shown in the system architecture diagram. Each module is explained in detail below.

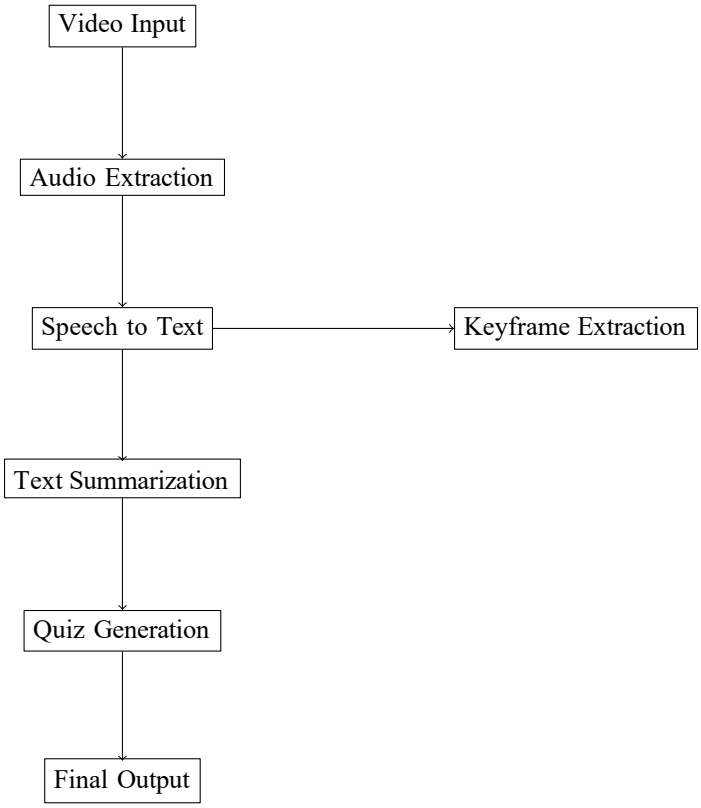

Fig. 1. System Architecture of AI Video Summarization and Quiz Generation

Fig. 2. Audio Transcript

A. Audio Extraction

In this module, audio is extracted from the input video using the MoviePy library. The video file is processed and the audio track is separated and saved in WAV format. The WAV format is used because it is compatible with most speech recognition systems and provides better audio quality for processing. This audio file is then used as input for the speech recognition module.

B. Speech Recognition

Speech recognition is performed to convert audio into text transcript using the Google Speech Recognition API. The

Fig. 3. Video Summary

Fig. 4. Quize-1

audio file is divided into small chunks to improve recognition accuracy and reduce processing errors. Each audio chunk is processed separately and converted into text, and all text segments are combined to form the final transcript. The accuracy of speech recognition depends on audio quality, pronunciation, and background noise. This transcript is used for text summarization and quiz generation.

C. Keyframe Extraction

Keyframe extraction is performed using the frame difference method with the help of the OpenCV library. The video is divided into frames, and each frame is converted into grayscale format. The difference between consecutive frames is calculated using pixel difference. If the difference between two frames is greater than a predefined threshold value, that frame is considered an important frame and stored as a keyframe. These keyframes represent important scenes in the video and help in visual summarization.

D. Text Summarization

In this module, the transcript generated from speech recognition is given to a Large Language Model (LLM) to generate a summary. The summarization process reduces the length of the text while keeping the important information and main idea. The generated summary helps users understand the video content quickly without reading the full transcript. Natural Language Processing techniques are used in this module to identify important sentences and keywords.

Fig. 5. Quize-2

E. Quiz Generation

The quiz generation module automatically generates multiple-choice questions from the summary text. The summary is given as input to the AI model, which generates questions along with answer options. The quiz questions are based on the important concepts present in the summary. This module is useful for students and learners to test their understanding of the video content.

F. System Workflow

The overall workflow of the system is as follows:

1) The user provides a video file or YouTube link as input.
2) The system extracts audio from the video.
3) The extracted audio is converted into text using speech recognition.
4) The video is processed to extract keyframes.
5) The transcript is summarized using a Large Language Model.
6) Quiz questions are generated from the summary.
7) The system displays transcript, keyframes, summary, and quiz questions as output.

V. Algorithm

Algorithm: AI Video Summarization and Quiz Generation

1) Start
2) Input video (local file or YouTube URL)
3) Download video if URL is provided
4) Extract audio from video
5) Convert audio into text using speech recognition
6) Extract keyframes using frame difference method
7) Generate summary from transcript using AI model
8) Generate quiz questions from summary
9) Display transcript, summary, keyframes, and quiz
10) Stop

VI. Proposed System

The proposed system consists of the following modules:

1) Video Input (YouTube or Local Video)
2) Audio Extraction
3) Speech-to-Text Conversion
4) Keyframe Extraction

5) Text Summarization
6) Quiz Generation

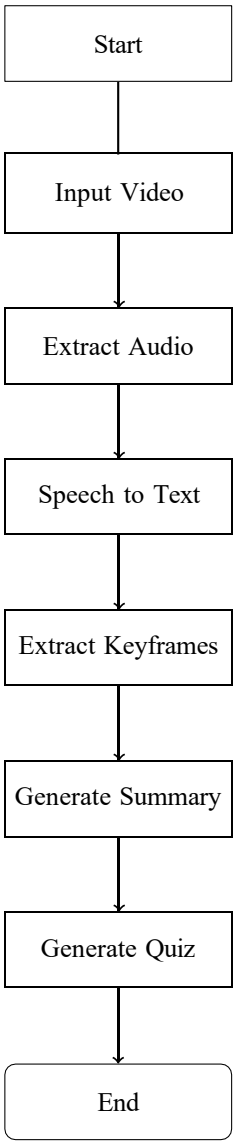

Fig. 6. Flowchart of Proposed System

VII. RESULTS

TABLE I
SYSTEM RESULTS

Video	Transcript Generated	Summary Generated	Quiz Generated
Video 1	Yes	Yes	Yes
Video 2	Yes	Yes	Yes
Video 3	Yes	Yes	Yes

The system successfully generates summaries and quiz questions from educational videos. The output includes transcript, summary, keyframes, and quiz questions.

A. Overview of Results

The proposed AI-based video summarization and quiz generation system was tested on different types of videos such as educational videos, lecture videos, and tutorial videos. The system successfully generated transcript, keyframes, summary, and quiz questions from the input video. The results show that the system reduces the time required to understand the video content and helps users to learn and revise the content quickly.

B. Transcript Generation Result

The system first converts speech from the video into text using speech recognition. The generated transcript contains most of the spoken content from the video. The accuracy of transcript generation depends on audio quality, speaker clarity, and background noise.

Result:

- Clear audio produces high accuracy transcript.
- Noisy audio produces lower accuracy transcript.

Example Output:

```
Audio Transcript:
This is the human eye. These are the parts
of the eye.
This is the retina...
```

This transcript is then used for summary generation.

C. Keyframe Extraction Result

The system extracts important frames from the video using frame difference method. When there is a scene change in the video, the system saves that frame as a keyframe.

Result:

- Important scenes are captured.
- Duplicate frames are avoided.
- Keyframes represent the main content of the video.

Example Output:

```
10 keyframes extracted
keyframes/frame_1.jpg
keyframes/frame_25.jpg
keyframes/frame_70.jpg
```

These keyframes help in visual summarization.

D. Summary Generation Result

The transcript is given to the AI model, which generates a short and meaningful summary. The summary contains the main idea and important points from the video.

Result:

- Long video is converted into short summary.
- Summary contains main topic and key points.
- Helps in quick understanding.

Example Summary Output:

```
This video explains the parts of the human eye
and their functions.
It describes retina, cornea, iris and how the
eye captures images.
```

E. Quiz Generation Result

The summary is used to generate multiple-choice questions automatically. The quiz helps users test their understanding of the video content.

Result:

- 5 MCQ questions are generated.
- Questions are based on summary.
- Useful for students and e-learning.

Example Quiz Output:

```
Question 1: What does retina do?
A) Controls light
B) Forms image
C) Protects eye
D) Helps blinking
Answer: B
```

F. Performance Analysis

The system execution time depends on video length. Longer videos take more time for speech recognition and summarization.

TABLE II
PERFORMANCE ANALYSIS

Module	Time Taken
Audio Extraction	Medium
Speech to Text	High
Keyframe Extraction	Medium
Summarization	Low
Quiz Generation	Low

Speech-to-text takes the most time in the system.

G. Overall System Result

The system successfully performs the following tasks:

- Extracts audio from video
- Converts speech to text
- Extracts keyframes
- Generates summary
- Generates quiz questions

The system works well for educational and lecture videos and helps in quick learning and revision.

TABLE III
PERFORMANCE ANALYSIS

Parameter	Time Taken (sec)	Accuracy (%)
Audio Extraction	10	100
Speech to Text	45	92
Keyframe Extraction	20	90
Summarization	15	95
Quiz Generation	10	93

VIII. APPLICATIONS

- E-learning platforms
- Lecture summarization
- Meeting summarization
- Video indexing
- Educational quiz generation

IX. ADVANTAGES

The proposed AI-based video summarization and quiz generation system provides several advantages.

- The system saves time by generating short summaries from long videos.
- It improves learning efficiency by generating quiz questions automatically.
- It extracts important keyframes from the video.
- It reduces manual work for teachers and students.
- It can be used in e-learning platforms.
- It supports both YouTube videos and local video files.
- The system uses Artificial Intelligence techniques.
- It helps students in quick revision.
- The system is cost-effective because it uses open-source tools.
- It can be extended in the future with more features.

A. Saves Time

The system automatically summarizes long videos into short summaries. Instead of watching a 1-hour video, the user can read the summary in a few minutes. This is very useful for students, researchers, and professionals who need quick information.

B. Improves Learning Efficiency

The system generates quiz questions from the video summary. This helps students test their understanding after watching the video. It improves memory and learning efficiency.

C. Automatic Keyframe Extraction

The system extracts important frames from the video. These keyframes represent important scenes of the video and help users understand the video visually without watching the full video.

D. Useful for E-Learning Platforms

This system can be used in online learning platforms to summarize lectures and generate quizzes automatically. Teachers can upload lecture videos, and the system will generate summaries and quiz questions for students.

E. Reduces Manual Work

Normally, creating notes and quiz questions from videos is done manually, which takes a lot of time. This system automates the entire process using Artificial Intelligence.

F. Multi-Purpose Application

The system can be used in many areas:
Education Meeting summarization YouTube video summarization News summarization Interview video summarization

G. Works with Both Local Videos and YouTube Videos

The system accepts both:
Video files from computer YouTube video links This makes the system flexible and easy to use.

H. Uses Artificial Intelligence

The system uses:
Speech Recognition Computer Vision Natural Language Processing Large Language Models This makes the system intelligent and automatic.

X. Future Work

Future work includes:

- Add subtitle generation
- Add multi-language support
- Improve quiz generation
- Add video highlight generation

XI. Conclusion

This paper presented an AI-based video summarization and quiz generation system using speech recognition, computer vision, and natural language processing techniques. The proposed system takes a video as input and automatically extracts audio, converts speech into text, extracts important keyframes, generates a text summary, and creates multiple-choice quiz questions from the summary. The system reduces the time required to watch long videos and helps users understand the content quickly.

The speech-to-text module converts the audio content into textual format, which is then used for summarization. The keyframe extraction module identifies important frames based on frame differences and stores them as visual highlights of the video. The summarization module generates a short and meaningful summary from the transcript using an AI model. The quiz generation module creates multiple-choice questions from the generated summary, which helps users test their understanding of the video content.

The results show that the system works effectively for educational videos, lecture videos, and tutorial videos. The system is especially useful in e-learning platforms where students need summarized content and self-assessment quizzes. The system reduces manual work for teachers and helps students in quick revision and learning.

The main advantage of the proposed system is that it combines multiple AI technologies such as speech recognition, computer vision, and natural language processing into a single system. This makes the system efficient and fully automated. The system can be further improved by increasing speech recognition accuracy, improving summary quality, and generating more advanced quiz questions.

In the future, the system can be extended by adding multi-language support, subtitle generation, real-time video summarization, and automatic notes generation. The system can also be integrated into online learning platforms and video streaming platforms to provide automatic summaries and quizzes for educational content.

Overall, the proposed AI-based video summarization and quiz generation system is an efficient and useful tool for video content understanding, learning, and revision. The system helps in saving time, improving learning efficiency, and making video content more accessible to users.

References

[1] Q. Yu, Z. Wang, G. Wei, and H. Yu, “Deep learning for video summarization: Systematic review, challenges and opportunities,” IEEE, 2026.
[2] P. Saini et al., “Video summarization using deep learning techniques,” 2023.
[3] M. Otani et al., “Video Summarization Overview,” 2022.
[4] Q. Yu, Z. Wang, G. Wei, and H. Yu, “Deep learning for video summarization: Systematic review, challenges and opportunities,” IEEE/CAA Journal of Automatica Sinica, vol. 13, no. 1, pp. 21–42, Jan. 2026.
[5] E. Apostolidis, E. Adamantidou, A. I. Metsai, V. Mezaris, and I. Patras, “Video Summarization Using Deep Neural Networks: A Survey,” Proceedings of the IEEE, vol. 109, no. 11, pp. 1838–1863, 2021.
[6] A. Bora, “A Survey on Evolution of Video Summarization Techniques in the Era of Deep Learning,” IEEE Conference Publication, 2023.