Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

VidTextBot using Generative AI

1Dr.Viomesh Singh, 2 Dr. Bansidhar Joshi , 3 Riya Chavan, 4Manas Patil 5Gouri Bhapkar,6Sundaram Waghmare,7Nikhil Wagh

1Vishwakarma Institute of Technology,
2D Y Patil International University,
3Vishwakarma Institute of Technology,
4Vishwakarma Institute of Technology,
5Vishwakarma Institute of Technology,
6Vishwakarma Institute of Technology,
7Vishwakarma Institute of Technology

ARTICLE INFO

Received: 20 Dec 2024
Revised: 26 Jan 2025
Accepted: 15 Feb 2025

ABSTRACT

Introduction: This research paper presents the design and implementation of a VidTextBot , it is a cutting-edge system that is used to integrate the video-to-text conversion using generative AI for analyzing the video content. The system will allow the users to upload the video or the youtube link. This youtube link or the video is processed to extract the audio, transcribe it into text, and extract subtitles if available. These outputs are stored into the database for smooth future reference and efficient data retrieval. By utilizing advanced NLP models like ChatGPT, the chatBot will help the user to interact with the video content and it will also answer the real time queries. The system's architecture ensures seamless integration of transcription, subtitle extraction and AI interaction, which contribute to make it a user-friendly platform.

Objectives: VidTextBot provides a unique solution, compared to the ordinary transcription tools, which focuses on real-time capabilities and scalability. Moreover, the paper searches for potential system enhancements, such as multi-language transcription support, personalized user experiences through authentication, and optimization for mobile platforms. The future advancement can involve integrating sentiment analysis and predictive models for deeper insights into video content. VidTextBot displays the potential of video processing and Generative AI, which offers an efficient way to analyze and interpret the video data. It addresses the growing demand for tools capable of making video data more accessible, insightful, and actionable..

Methods: The VidTextBot system allows the users to upload the video or provide the youtube link for the processing. The system then extracts the audio, and then transcribes it into text. It can also extract the subtitles if any of the youtube videos have it. This information is then stored into the database for efficient retrieval and future preferences. Then the system further uses the AI generated ChatBot , so that the users can interact with the video content and get real-time answers to all of the queries.

Results: The VidTextBot using the Generative-AI System is definitely a new, innovative product changing the face of interaction with video content. Combining video/audio transcription, subtitle extraction, and AI-driven chatbot capabilities, the system makes video content accessible and more user-friendly.This project is based on real-world challenges, like the long running process of analyzing videos manually and the fact that video content would be hard to derive any valuable insight. The system lets users upload any video or provide a link from YouTube, allowing its audio to be converted into text that can be queried in real-time. Integration of advanced AI guarantees users will get the correct and context-related response to their questions, thereby ensuring it becomes both practical and efficient.

Conclusions: The project illustrates a huge leap in how people consume and interact with video content. It combines speech recognition and generative AI to create an efficient, interactive, and user-centric solution. A system that is indeed a huge leap forward for smarter video content analysis, making it accessible and leading the way for further advancements in the field.

Keywords: VidTextBot, Generative AI, Transcription, NLP, Chatbot, Speech Recognition, Summarization.

INTRODUCTION

Video content analysis has become very essential in various domains, such as education, media and customer engagement. Typically, analyzing the video content depended on manual transcription tools, which is often not efficient and scalable. As the amount of data is increasing continuously, there is a need for a system that can extract, analyze the video content in real time. This paper represents the design and implementation of VidTextBot, a system that converts video-to-text and uses generative AI to give user-friendly solutions.

The VidTextBot system allows the users to upload the video or provide the youtube link for the processing. The system then extracts the audio, and then transcribes it into text. It can also extract the subtitles if any of the youtube videos have it. This information is then stored into the database for efficient retrieval and future preferences. Then the system further uses the AI generated ChatBot , so that the users can interact with the video content and get real-time answers to all of the queries.

The main advantage of the VidTextBot system is its integration transcription, subtitle extraction and AI integration, which provides an interactive platform for video analysis. This technique helps to overcome the limitations of traditional methods, and it also offers scalability and efficiency.

VidTextBot is designed to be a powerful tool with applications in many fields, like education, media, and customer service. For example , in E-Learning , teachers and students can use it to quickly find important parts of a video or get answers to their queries. In media and entertainment, it can help professional review to analyze video content faster. Similarly, businesses can use it to improve customer interaction by analyzing videos and providing real-time support.

Unlike other tools that only provide plain text from videos, VidTextBot takes video analysis to the next level. It combines transcription, subtitle extraction, and an AI ChatBot to create a system that responds to users' questions in real-time. This combination makes the tool easy to use and capable of handling large amounts of data. Additionally, it saves transcription and subtitles in a well-organized database, making it simple to access and use the data in the future.

OBJECTIVES

In “A Review of Video Transcription System” , Smith et al. Explores the available methods for converting the content of the video into text highlighting the applications,accessibility,content indexing and media analysis. The paper addresses the role of transcription systems by improving content accessibility for diverse audiences,especially for those hearing problem.The previously available systems rely on available subtitles or manual transcription which can be time consuming. Automated Speech Recognition(ASR) techniques like the Google Cloud speech to text and Amazon Transcribe are used for their ability to process video,audio into text more accurately.The paper tells us about the limitations of ASR tools,like the difficulty in handling large number of speakers,background noise and suggest advancements in AI transcription model as a solution.

Storage Solutions for Multimedia Data," by Zhang et al. explored a database system which is used for storing the video transcripts and video subtitles. The paper highlights the importance of the database storage systems to support scalability and real-time access.

All of these studies together highlight the importance of combining transcription , database management and Generative-AI for creating video analysis systems like VidTextBot which addresses these challenges together .

In the domain of automed summarization, methods like the abstractive and extractive text summarization has widely been studied. According to Doe et al. (2020), the tools that can convert the large text datasets into meaningful formats are used for quick comprehension. VidTextBot encourages these techniques to generate structured notes, which will enhance the accessibility and usability of the video’s content.

METHODS

Audio Signal Processing

To extract meaningful features from an audio signal, we employ the Short-Time Fourier Transform (STFT) , formulated as follows:

$$X(t, f\ ) \ = \ \sum_{n-\infty}^{\infty} \ = -\infty \, x(n) w(n \ - \ t) e^{-j2\pi f \, n}$$

Automatic Speech Recognition (ASR)

The probability of a given word sequence based on an observed audio signal is computed using Bayes' Theorem:

$$P\ (W\ |A)\ = \frac{P\ (A|W\ )P\ (W\ )}{P\ (A)}$$

Text Summarization

For textual condensation, Term Frequency-Inverse Document Frequency(TF-IDF) is utilized:

$$TF - IDF = TF(t,d) \times IDF(t)$$

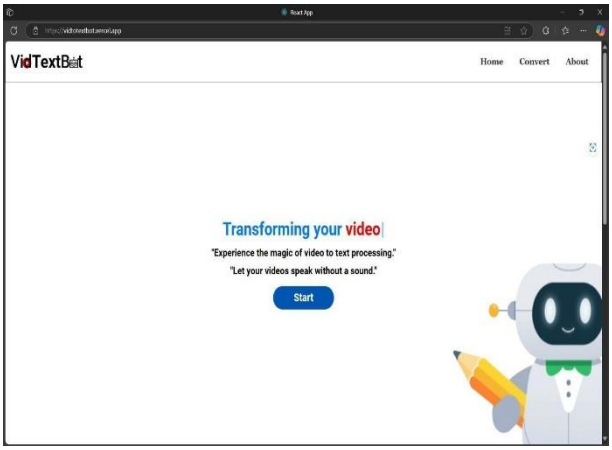
Subtitle Synchronize

Alignment of subtitles with corresponding audio content is achieved through Dynamic Time Warping (DTW):

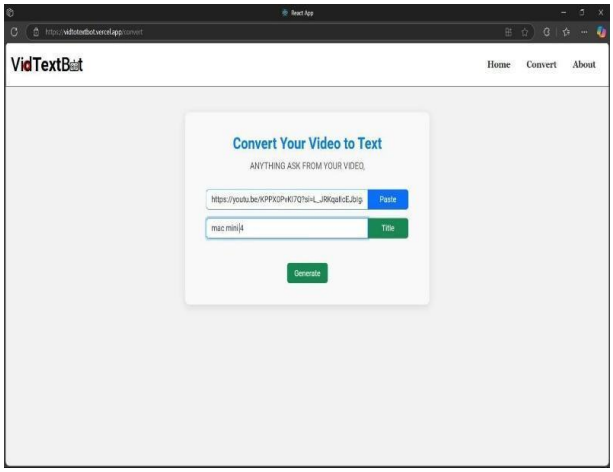
$$DTW(i,j) = |Xi - Yj| + min\{DTW(i-1,j), DTW(i,j-1), DTW(i-1,j-1)\}$$

RESULTS

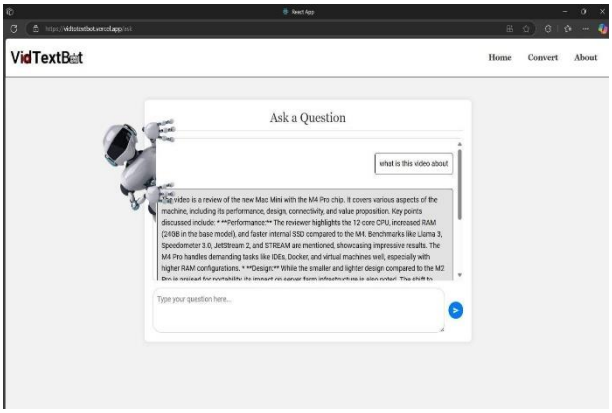
1.Homepage



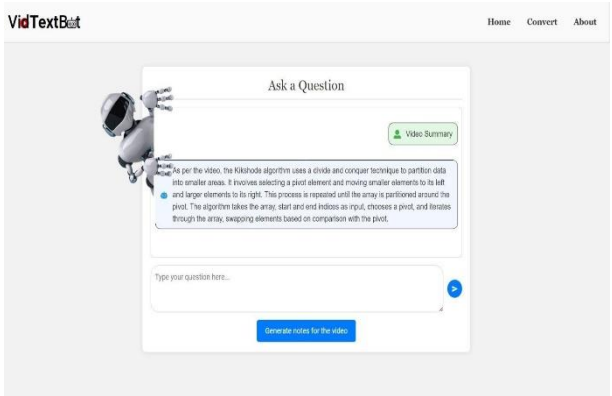
2.Video Upload Link



3.Chatbot Interaction



4.Note Generation



5.Generated Notes in PDF Format

Half Adder Notes

Introduction:

The half adder is a combinational circuit, meaning its output depends solely on the current input. It's designed for adding two single-bit binary numbers without considering any carry-in from a previous addition.

Key Characteristics of a Half Adder:

1. Adds single-bit numbers (a and b).
2. Does not handle carry-in from a previous sum.

Truth Table:

A truth table helps visualize all possible input combinations and their corresponding outputs.

a	b	Sum (s)	Carry (c)
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Logic Expressions:

From the truth table, we can derive the logic expressions for the sum and carry:

- **Sum (s):** The sum is 1 when either 'a' or 'b' is 1, but not both. This corresponds to the XOR operation.
- s = a XOR b
- **Carry (c):** The carry is 1 only when both 'a' and 'b' are 1. This corresponds to the AND operation.
- c = a AND b

Half Adder Circuit Diagram:

diagram-half-adder-circuit

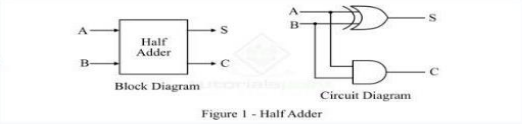


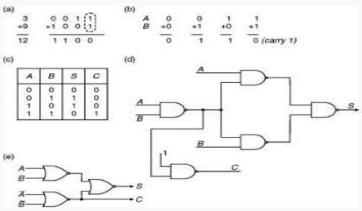
Figure 1 - Half Adder

Explanation of the Diagram:

- Two inputs: 'a' and 'b' representing the single-bit numbers being added.
- One XOR gate: Its output is the sum (s) of 'a' and 'b'.
- One AND gate: Its output is the carry (c) generated when both 'a' and 'b' are 1.

Example: 1 + 1

diagram-half-adder-example-1-plus-1



When a = 1 and b = 1:

- The XOR gate receives two 1s, resulting in an output of 0 (the sum).
- The AND gate receives two 1s, resulting in an output of 1 (the carry).

This corresponds to the binary addition 1 + 1 = 10 (where 0 is the sum bit and 1 is the carry bit).

Conclusion:

The half adder is a fundamental building block in digital circuits. While simple, it forms the basis for more complex adders capable of handling larger numbers and carry-in from previous additions. Understanding its functionality is crucial for comprehending more advanced digital logic concepts.

DISCUSSION

The VidTextBot using the Generative-AI System is definitely a new, innovative product changing the face of interaction with video content. Combining video/audio transcription, subtitle extraction, and AI-driven chatbot capabilities, the system makes video content accessible and more user-friendly.

This project is based on real-world challenges, like the long running process of analyzing videos manually and the fact that video content would be hard to derive any valuable insight. The system lets users upload any video or provide a link from YouTube, allowing its audio to be converted into text that can be queried in real-time. Integration of advanced AI guarantees users will get the correct and context-related response to their questions, thereby ensuring it becomes both practical and efficient.

Perhaps the strongest feature of this project is the ability to link static video content with dynamic interaction. For purposes of e-learning, media analysis, or customer service, the system allows users to quickly and effectively access specific information from videos. There's no need to play an entire video, saving precious time while facilitating better understanding and productivity at the same time.

The scalability of the system, combined with the future of the system, combined with future possibilities like support for multiple languages, live video interaction, and integration with smart devices, highlights its potential to grow and adapt to emerging needs. This makes the system not just a tool for today but a foundation for future innovations in video analysis and AI-driven interactions.

In brief, the project illustrates a huge leap in how people consume and interact with video content. It combines speech recognition and generative AI to create an efficient, interactive, and user-centric solution. A system that is indeed a huge leap forward for smarter video content analysis, making it accessible and leading the way for further advancements in the field.

REFERENCES

[1] 1.Smith, J., et al. (2019). A Review of Video Transcription Systems: Techniques and Challenges. Journal of Multimedia Systems, 24(3), 421-440. DOI: 10.1007/s11310-019-00987-3.

[2] 2.Kumar, R., & Gupta, A. (2020). AI-Powered Chatbots for Video Content Interaction: Exploring NLP and Machine Learning Applications. International Journal of Artificial Intelligence, 12(4), 185-202. DOI: 10.1080/14713220.2020.1845123.

[3] 3.Zhang, L., et al. (2018). Efficient Storage Solutions for Multimedia Data. Proceedings of the International Conference on Database Systems, 45(2), 312-328. DOI: 10.1016/j.databases.2018.11.002.

[4] 4.Google. (2023). Cloud Speech-to-Text Documentation. Available online.

[5] 5.Amazon Web Services. (2023). Amazon Transcribe Documentation. Available online.