

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with



*The University of Westminster, Coat of Arms*

**Summarizer**

Multimodal Video Summarization System

A Project Proposal by

**Mr. Dilukshan Sathiyamurthi**

Supervised by

**Mr. John Sriskandarajah**

Submitted in partial fulfilment of the requirements for the BEng(Hons) in  
Software Engineering degree at the University of Westminster.

**April 2025**

## Abstract

---

The exponential growth of digital video content has created a critical need for summarization tools that help users quickly extract essential information without watching entire videos. Traditional methods often process visual, audio, and motion streams in isolation, resulting in fragmented, biased, or incomplete summaries.

This study presents a Multimodal Video Summarization System that combines visual frame analysis, audio transcription (OpenAI Whisper), and text summarization (Facebook BART), with motion detection via frame difference analysis. A fallback keyframe mechanism is employed when dynamic actions are minimal, and audio processing is skipped for silent videos. The solution is deployed as a full-stack web application (Flask backend + React frontend), containerized using Docker for scalability.

Experimental results demonstrate the system can reduce video length by over 80–90% while preserving essential content. Evaluation metrics on long-form videos show impressive performance: ROUGE-L: 0.12, Diversity: 0.95, Coverage: 0.68, Temporal Coherence: 1.00. These findings validate the system’s potential for applications in education, surveillance, and content indexing, offering a balanced, fair, and efficient summarization approach.

**Keywords:** Multimodal Video Summarization, Audio Transcription, Action Recognition, OpenAI Whisper, Facebook BART, Frame Difference Analysis, Docker, Flask, React

**Subject Descriptors:**

- Computing methodologies → Artificial intelligence → Computer vision → Video summarization
- Computing methodologies → Machine learning → Machine learning approaches → Neural networks
- Information systems → Information retrieval → Summarization