Keyframe Enrichment and Video Summarization

Duration of the project: 12 weeks

Project Mentors: Meghana Gupta, Tanushree R Ranjanagi

Team Members: Ninad Shetty,Tanay Shekokar,Rian Pinto

Introduction

Project statement: To develop a **video summarization system** which extracts the **key moments** from videos and generates summaries. This provides a time saving solution for users to quickly grasp the essential video content without actually seeing it in its entirety.

Inspiration: The project is inspired by the need to save time. Providing its users with concise summaries which enable them to grasp the information quickly from lengthy videos without having to go over the entire thing is a major tool of convenience especially for those who are expected to watch them but don't want to.

Background info: Such summarization tools exist and they take in audio transcripts to generate their output summaries based on how much in detail the user requires. Very few resources that summarize videos based on frame and images alone exist.

Method

1. *Merging of some datasets that will be utilized for training, like TVSum, SumMe,etc. The datasets have different types of videos.*

2. *Video Processing: The training data set will mostly have videos in which audio doesn't play an important role, and summary can be generated from motion videos. Input video will be processed to prepare it for analysis.*

3. *Keyframe Extraction: These are the frames that represent the most important moments in the video and capture significant events. Keyframes can be extracted by using -*

a. *Threshold-based change detection, where consecutive frames are compared and changes are detected based on pixel intensity or color differences. If the difference exceeds a threshold, then it's a candidate frame.*
b. *Using clustering algorithms on candidate frames to identify keyframes which act as a representative to a group of similar frames.*

4. *Image to text conversion: To generate description for each key frame, image captioning will be done using a pre-built model, based on pipelining.*

---

5. *Post-processing: Once the texts for key frames are ready, it can be further summarized using key-words, and grammar can be fixed.*

Results

The developed video summarization model effectively identifies and extracts the most important moments from videos, creating concise yet informative summaries that provide a comprehensive overview of the content. By presenting the essential elements in a coherent manner, users will be able to quickly grasp the main points and key details without the need to watch the entire video, thus saving valuable time.

It will be a convenient solution for users who are expected to watch lengthy videos but prefer a more efficient way of consuming the information. By automating the process of summarization, users can extract the crucial content from videos without compromising the quality of the information.

Obstacles faced

- *Determination of dynamic threshold for extracting the candidate frames*
- *Image captioning of the Key frames extracted*
- *Implementation of the suitable key frame extraction method*

Conclusion

The objective of this project was to extract the key frames and summarize the video. We have used suitable methods for first extracting the candidate frames, and then the key frames. The threshold method based on pixel intensity difference as well as the OPTICS clustering used for the key frames which takes the DCT coefficients into account has turned out to provide reasonable efficiency with respect to the final output. The captioning algorithm based on pipelining has helped us to generate captions for each key frame that was extracted. Further, the captions were summarized to generate a transcript for the video.

Future work

Further studies can be conducted on finding an efficient method for testing the model which will in turn help to enhance the overall output. Using the videos with transcript as the test data, the transcript can be summarized and the results can be compared with the summary generated by the model using NLP.

References

Advanced Key Frame Extraction from Video | by Varun Mittal | Medium
https://medium.com/@shrutijadon/video-summarization-an-survey-of-existing-algorithms-10b92a53e0a8
https://pictory.ai/pictory-features/auto-summarize-long-videos