VIDEO RETRIEVAL VIA KEYRAME EXTRACTION IN A CLASSROOM RECORDING ENVIRONMENT

Intelligent video detection and interpretation using distributed robust deep learning

María Fernanda Hernández Montes
B.D. Robotics and Digital Systems

Research Stay 2022

TC3073.603 ESTANCIA DE INVESTIGACIÓN AGOSTO-DICIEMBRE 2022

RESEARCH GROUP: ADVANCED ARTIFICIAL INTELLIGENCE

Instructors:
Dr. Hugo Terashima Marín, Dr. Santiago E. Conant Pablos, Dr. José Carlos Ortíz Bayliss

ABSTRACT

One of video retrieval major downsides is having to work with masive amounts of data, which will vary according to the nature of the videos being processed and the context that determines the information of interest. This is why the solution model is based on a first phase of keyframe extraction, which will permit the simplification of the dataset to be analyzed by creating a summarized set of relevant frames that will describe pivot moments from the original video, hence allowing to reduce processing costs without losing relevant information [1].

Through this work, we aim to propose a machine learning model that allows users to **retrieve information of interest according to the activity shown in a video**, thus through the implementation of a Keyframe Extraction approach that permits improvement in the accuracy of the results obtained **while reducing and simplifying the set of data to be worked with**.

INTRODUCTION

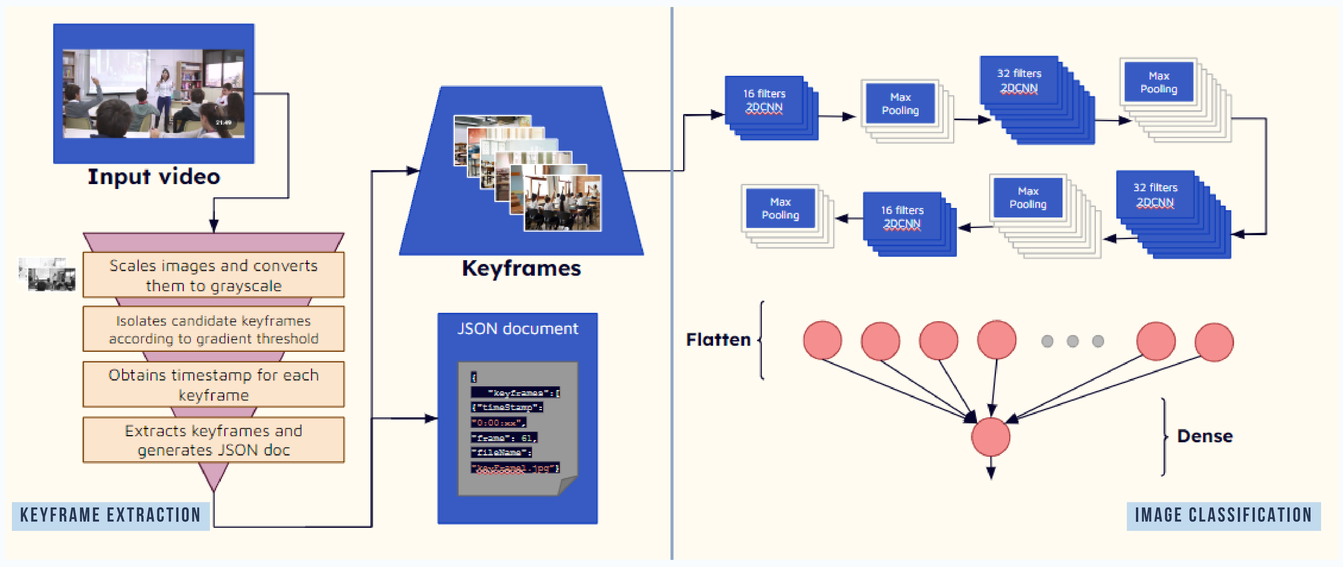
As a result of the pandemic caused by the spreading of SARS-CoV-2 (COVID 19), we have seen the way different barriers to participation were being experienced by students, and how they became particularly evident in collaborative learning tasks. On account of this, many schools have tried to come across with different solutions that could help improve the engagement students show during their online lessons, most of which have stayed even after the sanitary crisis mentioned above [2].

This project works with the idea of the classroom students as a group instead of individual beings, which allows us to establish different scenarios that will be set for analysis, which are: *The classroom shows a high level of movement, the classroom presents an low level of movement*.

MODEL ARCHITECTURE

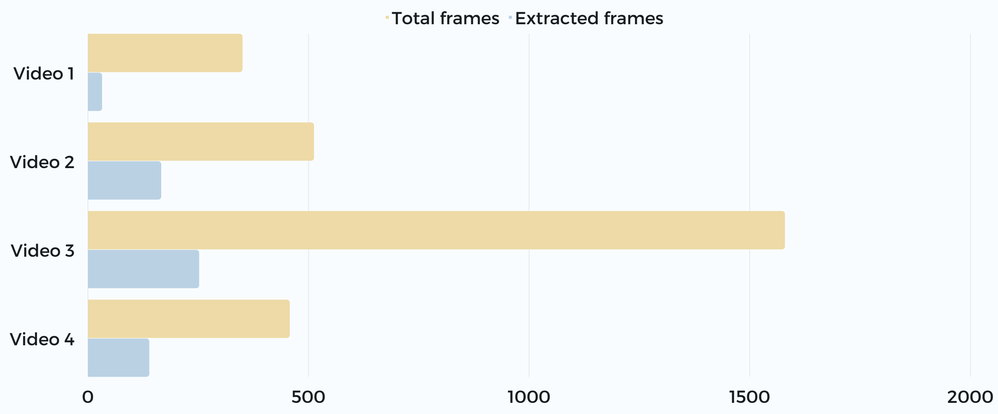
Although there are different approaches that have been explored when it comes to Text-Based Video Retrieval and data-set pre-training methods, for the purposes of the present project two of these have been taken into account as a ground for the desired model in which our research is based upon: Video summarization through Keyframe extraction and Image Classification to classify said keyframes into status of interest [3].

During the first phase we implemented a Keyframe Extraction model to summarize the video of interest, in order to reduce processing costs of analyzing a full-length video. After this, we used an Image Classification model which will generate the analysis of the video according to the activity being visually shown.



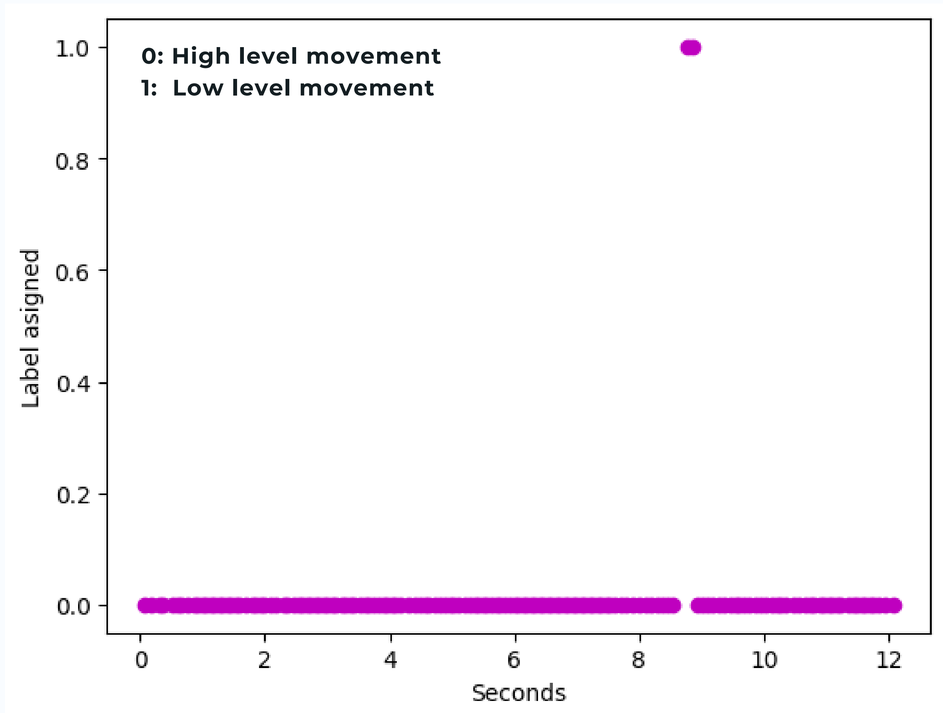
RESULTS

Since our model proposes an approach that aims to reduce the costs of video processing while simplifying the process of retrieval according to the time it takes to analyze the frames, in table 1 we can observe our results on these considerations.



Graphic 1. As we can observe, the first phase of our model allows us to summarize the video from 9% to 31% of the total length, thus reducing both processing times and costs.

Now, after we have summarized these test videos, we classify their correspondant keyframes according to different states. In the next graphic, we can observe an analysis made on one of the videos tested above.



Graphic 2. The results of the present model show us the moments throughtout the time where we can observe high and low level movement in the classroom recorded.

CONCLUSIONS

As we can observe, after processing a video with the model presented, interesting results were obtained, since costs related to time were reduced while an analysis according to the activity visually shown was generated. This way, the present work represents a useful tool for both teachers and educational institutions to evaluate the efficacy of the class dynamics adaptation in a pospandemic world.

REFERENCES

[1] Bashir Olaniyi Sadiq, Bilyamin Muhammad, Muhammad Nasir Abdullahi, Gabriel Onuh, Ali Abdulha-keem Muhammed, and Adeogun Emmanuel Babatunde. Keyframe extraction techniques: A review. Journal of Electrical Engineering, 19(3):54–60, 2020.

[2] Jenna Gillett-Swan. The challenges of online learning: Supporting and engaging the isolated learner. Journal of Learning Design, 10(1):20–30, 2017.

[3] Matthew Bihis and Sohini Roychowdhury. A generalized flow for multi-class and binary classification tasks: An azure ml approach. In 2015 IEEE International Conference on Big Data (Big Data), pages 1728–1737, 2015.