Enhancing Keyframe Extraction in Lecture Videos Using Optical Flow

Sunwoo Baek, Supia Park

Siebel School of Computing and Data Science, University of Illinois Urbana Champaign

INTRODUCTION

Lecture videos have a variety of forms, including a mix of static slides and dynamic annotations, which makes it challenging to extract keyframes that best represent the content. In this research we are addressing the specific problem: given an hour-long lecture video, how do we extract the key frames to create a concise PDF without redundant images from the lecture? Traditional methods like pixel similarity and OCR-based text extraction are not the most effective to capture gradual, non-discrete changes such as handwritten notes, highlights, and incremental annotations.

To address this, we explore a series of methods to complete the stated task. We utilize methods such as optical flow, masking, and subtraction in order to efficiently and accurately select representative frames from a lecture video.

APPROACH

In our approach, a lecture video will go through a series of methods, refining the work from Angrave, Li, and Zhong, 2022, that is currently implemented in ClassTranscribe.

First, we use Farneback dense optical flow to track how pixels move from one frame to the next. This helps us detect scrolling by measuring how much of the screen moves vertically between frames.

Next, we calculate how much of the screen is involved in this movement and how consistent the motion is across different parts of the screen. This allows us to detect whether the video is truly scrolling or just experiencing noise, such as shaky camera footage.

By repeating this process for every pair of frames, we build a set of motion statistics that guide how we group scenes. These motion cues form the foundation for identifying clean content boundaries, improving the accuracy and reliability of our video segmentation pipeline.

DETECTING SCROLLING

Exploring Optical Flow

We analyze motion vectors between consecutive frames; optical flow allows us to track vertical movements. Applying Dense Optical Flow by Gunnar-Farneback effectively detects the gradual evolution of content, and we are working on refining this approach to better differentiate meaningful updates from insignificant motion.

Applying Semantic Scene Segmentation with CNN

Pixel-wise differences for scene changes gave false positives from trivial visual variations (e.g., small text updates) that are not meaningful. To address this, we use CNN segmentation *(Figure 4)*, which captures high-level content rather than low-level noise. This enables us to group semantically similar frames, even when visual differences are subtle.

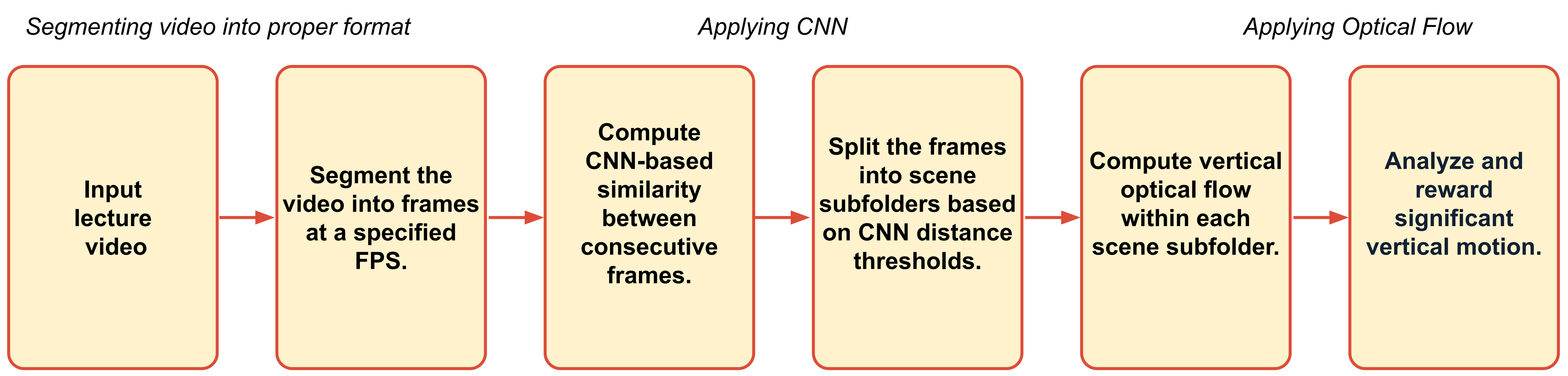
Vertical Motion Estimation via Optical Flow

To accurately measure vertical motion such as scrolling, we first divide the video into subfolders of semantically similar frames using CNN-based scene segmentation *(Figure 5)*. This ensures that motion analysis is performed only within visually coherent segments, avoiding misleading spikes caused by scene changes. Within each subfolder, we apply Farneback optical flow to consecutive frame pairs to quantify vertical motion while excluding irrelevant transitions.

Rewarded Vertical Motion (RVM)

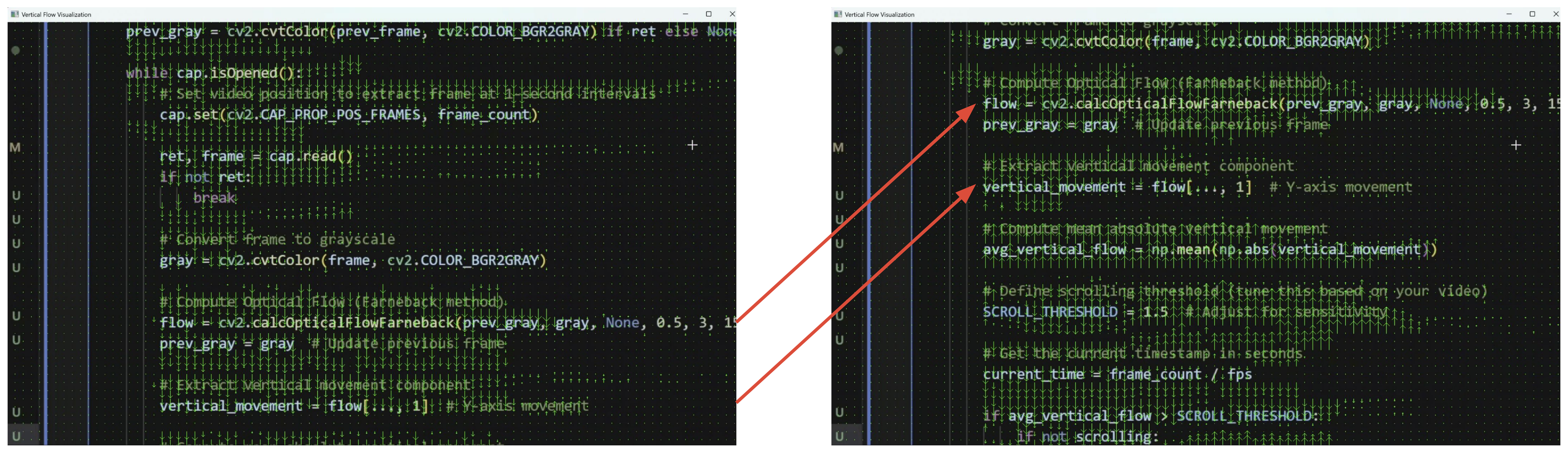
To prioritize widespread vertical movement over localized spikes, we define **Rewarded Vertical Motion (RVM)**, which amplifies motion scores when many pixels exhibit small movements *(Figure 6)*. This is crucial in lecture videos, where subtle scrolling affects a large area, while brief transitions or jitters may involve only a few regions *(Figure 7)*. RVM captures both the magnitude and the coverage of motion, rewarding consistency across the screen

$$\text{RVM} = \left(\frac{\textit{Sum of Significant Motion}}{\textit{Total Pixels}}\right) \times \left(\frac{\textit{Moving Pixels}}{\textit{Total Pixels}}\right)^{\alpha}$$



(Figure 1) Pipeline illustrating how our method detects scrolling behavior

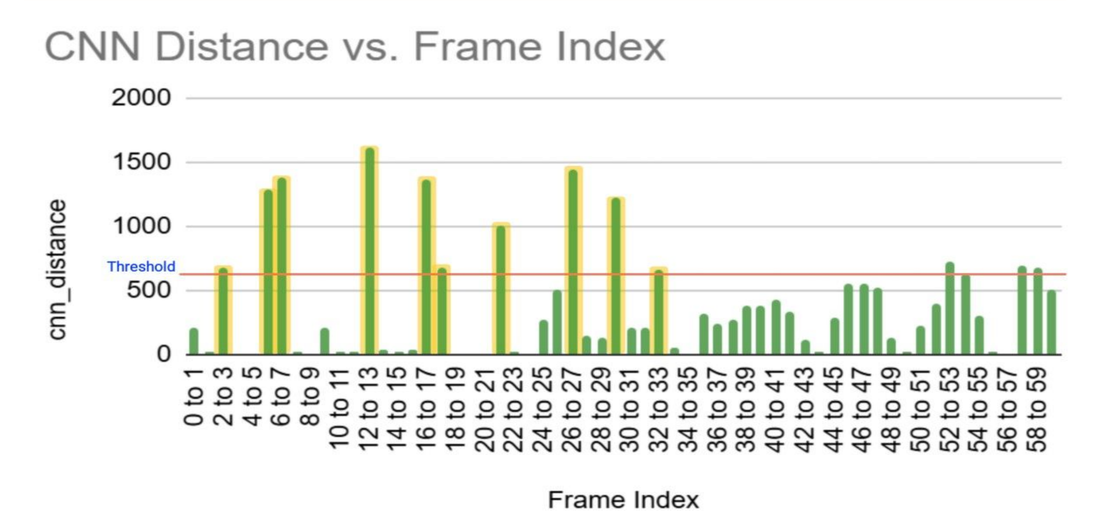
VISUALIZATION OF OPTICAL FLOW



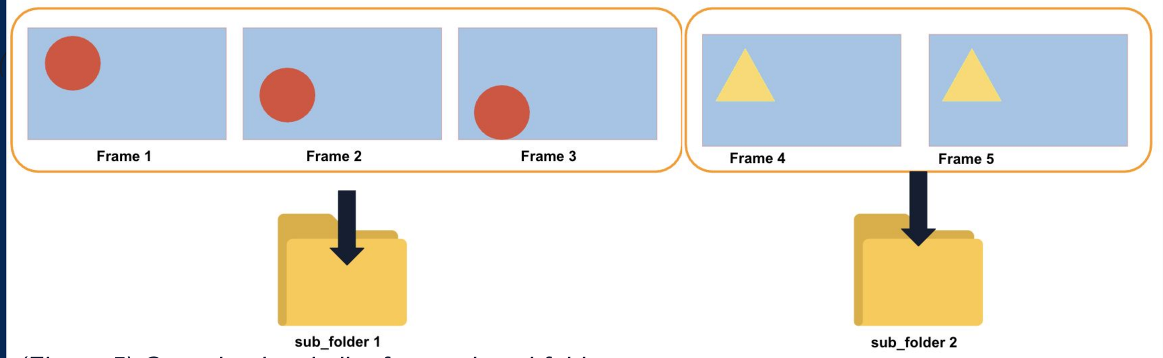
(Figure 2) Optical Flow: Dense Flow Farneback method- Screen scrolling down

(Figure 3) Optical Flow contd.

VISUALIZATION OF SEGMENTATION



(Figure 4) CNN score between each consecutive frames

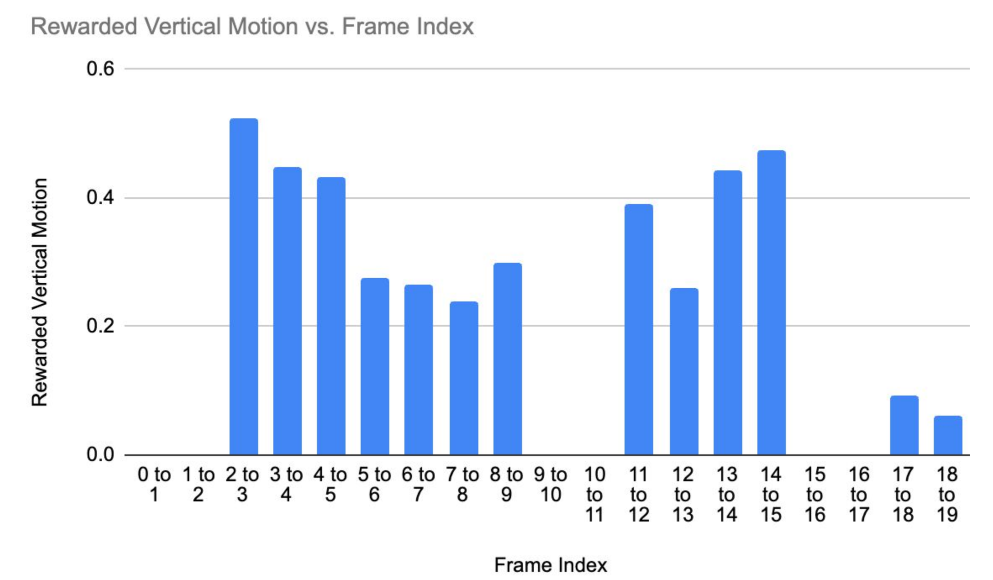


(Figure 5) Grouping by similar frames in subfolder

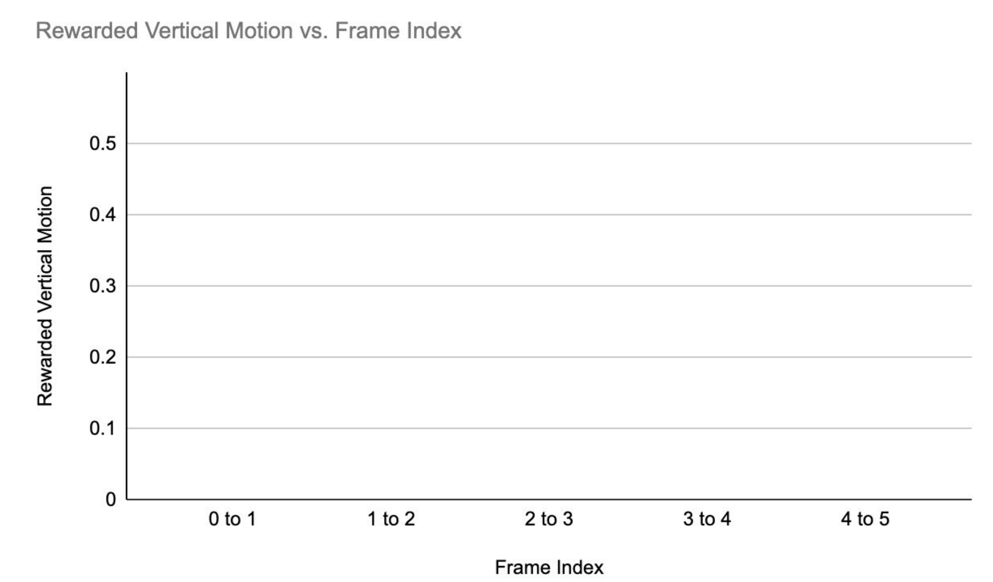
FUTURE WORK

We plan to enhance all key components of our system: scroll detection, masking and subtraction, and mean shift. Particularly, we aim to improve the scroll detection module to handle a wide range of scrolling speeds, as extremely slow or fast scrolling can prevent accurate frame detection. Additionally, we plan to explore improved thresholding strategies to better balance false positives and false negatives, ensuring robust scene segmentation across diverse types of videos. Once these improvements are in place, our next step is to integrate these methods to develop a robust scene detection pipeline. Ultimately, we aim to apply this system to **ClassTranscribe**, our university's lecture video platform, to enable more accurate and efficient segmentation of lecture content.

RESULTS



(Figure 6) Rewarded Vertical Motion of Scrolling scene



(Figure 7) Rewarded Vertical Motion of Non-Scrolling scene

REFERENCES

[1] Angrave, L., Li, J., and Zhong, N. 2022. Creating TikToks, Memes, Accessible Content, and Books from Engineering Videos? First Solve the Scene Detection Problem. In Proceedings of the 2022 ASEE Annual Conference & Exposition, Minneapolis, MN.

[2] OpenCV. Optical Flow. OpenCV Documentation. https://docs.opencv.org/4.x/d4/dee/tutorial_optical_flow.html

ILLINOIS