

A Cut Above The Rest

Analyzing the Impact of Shot Detection Algorithms on the Quality of Chunk Encoded Videos.

Sike Ogieva, Amherst College '25.

When we upload a video to YouTube or Twitch, the raw bytes of our file are encoded into optimized formats that are more conducive to storage and transfer. When we play it later, a partial reversal occurs as our devices decode the file.

Advances in video encoding seek to minimize video file sizes while retaining quality. However, for the algorithms that work the best, we accept considerably increased complexities and computational overhead. One way to mitigate these costs is chunk-based encoding. Breaking videos up and encoding them in parallel is nearly infinitely scalable.

When partitioning video chunks, grouping qualitatively similar frames helps to ensure that the same encoding parameters influence all its frames in uniform ways. Finding these natural qualitative boundaries within a video is called shot transition detection. A shot is technically a series of frames representing the same, continuous action– all usually captured by the same camera.

Methods of shot detection include the sum of absolute differences (SAD), the histogram differences (HD), and the edge change ratio (ECR). The ECR is considerably the most accurate but also the most computationally intensive. While they all work by measuring differences between frames, the ECR is more sensitive to differences that correlate more closely with actual shot transitions.

A question that comes up now is, how important is it to detect accurate shot transitions when splitting videos into chunks for encoding? When an algorithm overlooks a shot transition because the changing frames still appear "similar," how much do these "wrong" decisions affect the resultant quality of the video?

This study examines whether suboptimal shot detection, such as failing to capture a hard shift with SAD or HD, significantly affects the final encoded video's quality, where video quality is measured by Video Multi-method Assessment Fusion (VMAF) scores.

VMAF was introduced in 2016 by researchers from Netflix and the University of Southern California to replace the mean squared error as a video quality measure. The MSE is often a poor reflection of the opinions of human viewers about a video's quality, and the VMAF was designed to capture the subjective experience of human viewers more closely.

This project uses simulated video files processed through SAD and ECR shot detection algorithms and encoded with Google's VP9 codec. The final quality will be evaluated using VMAF scores. By comparing the performance of these shot detection methods, the study aims to determine whether more sophisticated techniques like ECR significantly improve video quality over more straightforward approaches such as SAD and HD.