

# EE398B Project Proposal

**Title:**

Intelligent Frame Dropping Algorithm for Rate-Distortion Optimized Video Encoding At Low Bit Rate

**Members:**

Nathan Eagle, [nathan@media.mit.edu](mailto:nathan@media.mit.edu)

Thomas Pun, [thomas.pun@stanford.edu](mailto:thomas.pun@stanford.edu)

Pavithra Srinivasan, [pavisrin@stanford.edu](mailto:pavisrin@stanford.edu)

**Description:**

Despite the complexity built in to modern video codec standards, there is still no ideal method to encode video sequences at an extremely low bit rate. This is a critical problem in many real-world applications because hard constraints on bandwidth are still quite prevalent, particularly for mobile devices. Streaming data at a higher bit rate than the channel can support results in late packets that are simply discarded by the client. Another encoding alternative is to simply drop certain frames. However, randomly dropping frames jeopardizes the user experience because it creates the possibility that the most relevant section of the sequence will be lost. A more intelligent frame-dropping algorithm is needed.

Traditionally, the quality of an encoded bitstream is measured by distortion. Distortion is commonly defined as Mean Squared Error (MSE). Although this measurement provides an objective assessment of the average deviation of the pixels in a frame, it does not account for perceptual quality and user experience. In order to enhance overall experience when the encoder is forced to drop frames, our proposed algorithm intelligently analyzes a buffer of frames and maps them directly to an encoding scheme using a weighted Viterbi algorithm. This algorithm deviates from the standard Viterbi [ 1 ] because it also incorporates a weighting metric - *sequence relevance* - on a short sequences of frames (approximately 2-5 seconds of frames). This metric is determined by a Lagrangian cost function, and is a sum of weighted *'hints'* that come from three different sources:

a) Motion analysis

e.g., scene change detection and “flashing” scene detection. Proper key frame is crucial to motion estimation. Flashing scenes, consecutive scene changes, are hard to code and can be represented by a subsample of scenes.

[ 1 ], [ 1 ]

b) Audio

e.g., pauses within a conversation and introduction of new speakers. Rapid movement in a video phone section may not be relevant if there is no conversation.[ 4 ], [ 5 ], [ 6 ]

c) State of the rate controller in the encoder

e.g., buffer fullness and how likely it will be forced to drop frames. This condition will be an indication on how many frames should be dropped. [ 1 ]

A major difficulty for this project is creating a systematic method of tracking progress. Since our frame-dropping algorithm is designed to enhance user experience, the major criterion will be subjective comparison. Another criterion is the PSNR measurement (between the encoded video and its original) versus the number of frames dropped.

For each test sequence at each bit rate, we will compare three different encoding strategies, all of which utilize the same encoder: uniform subsampling, dynamic frame dropping, and our proposed frame-dropping algorithm. The first one will be an encoding of uniformly temporal subsample of the original sequence. This is a common practice for reducing overall bit rate. However, decreasing temporal correlation between consecutive frames will also decrease compression efficiency. The second candidate for comparison will be an encoding with dynamic frame dropping. Any particular frame will be dropped if it exceeds the buffer constraint. The last candidate will incorporate our proposed frame-dropping algorithm whose output will be fed into the encoder. Our method will replace the dropped frames using the standard frame-repeat method. By doing so, we preserve the frame rate at the decoder. Our performance metrics will consist of both objective and subjective measurements. Objective measurements include calculating the PSNR vs. the number of frames dropped. We will also ask our testers to rank all three encodings according to their preference for best user experience.

In conclusion, we will investigate a frame-dropping algorithm that analyzes segments of the sequence and intelligently determines which frames can be dropped given the current buffer condition. Our method incorporates a weighting metric based on the following hints: motion analysis, audio information, and rate controller's buffer condition, and if time permits, other clues such as structured texture detection. Both objective and subjective comparison will be made against two other common practices: uniform temporal subsampling and non uniform frame dropping.

**Tasks:**

Week 1:

Study more reference papers, especially those related to our *hints*.

Generate test sequences.

Generate encodings with other methods against which our algorithm will be compared.

Week 2:

Design and Implement methods to extract the proposed hints from video sequences.

Link the resulting hints/relevance values with the rate controller of encoder reference software.

Week 3:

Continue improving our frame-dropping algorithm. More debugging, testing.

Compare our algorithm with existing methods.

Week 4:

Conduct user evaluation surveys and obtain subjective feedback.

Prepare for demo and project report.

## References:

- [ 1 ] G. David Forney, Jr. "The Viterbi algorithm", Proc. of the IEEE Vol. 61, No. 3, pp. 268-278. March, 1973.
- [ 2 ] J. S. McVeigh, M. W. Siegel and A. G. Jordan, "Adaptive reference frame selection for generalized video signal coding", Digital Video Compression: Algorithms and Technologies 1996, Vol. 2668, February, 1996, pp. 441 - 449.
- [ 3 ] ISO/IEC JTC1/SG29/WG11, ISO/IEC11172-2, "Information Technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s – Part 2: Video", May 1993
- [ 4 ] Zhu Liu, Jincheng Huang, Yao Wang, Tsuhan Chen, "Audio feature extraction and analysis for scene classification" IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing, Princeton, New Jersey, USA.
- [ 5 ] J. Nam and A. H. Tewfik, "Combined Audio and Visual Streams Analysis for Video Sequence Segmentation," Proc. of ICASSP'97, Vol. 3, pp. 2665-2668, 1997
- [ 6 ] Zhu Liu, Jincheng Huang, Yao Wang, Tsuhan Chen, "Audio feature extraction and analysis for scene classification" IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing, Princeton, New Jersey, USA.
- [ 7 ] ISO/IEC JTC1/SG29/WG11 Test Model Editing Committee, "MPEG-2 Video Test Model 5", ISO/IEC JTC1/SC29/WG11 Doc. N0400, April 1993.