I. Initial Ideas, System Outline

In this section we propose the outline of our framework for adaptive streaming of stored videos.

A channel estimator model is used to determine the available bandwidth of the network based on feedback information such as packet loss. The target number of key frames N_{KF} is determined from the available bandwidth. To extract the key frames, two approaches can be taken:

- **video-level optimization:** Appropriate values for SVD parameters are determined (e.g. iteratively) to output N_{KF} frames. Now we need to show these frames are the most representative frames. This can be done with subjective tests or we should find a metric that serves the purpose. Image quality metrics such as MSE, SSIM, or L_1 distance of color histograms do not capture the essence of video and seem unsuitable for our application. For example a trivial temporal sampling beats our results in terms of MSE since it distributes the key frames more uniformly. However I need to explore this further to see if restricting the MSE calculation to individual shots helps. This requires the computation of shot boundaries which is not yet clear to me since different values for parameters result in different peaks in frame significance plots. Another issue with such metrics (except color histograms) is their sensitivity to camera motions. For example MSE and SSIM perform very poor while comparing images that are translated or rotated copies. What we need is a metric that could compare two sequences of frames that are of the same length, yet do not have the same frames. Image quality metrics are not suitable for this purpose as they are designed for evaluating the quality of images after some processing (e.g. JPEG compression). It seems to me that any pixel-based content insensitive metric does not help us as a trivial temporal sampling would always beat us. Am I right?
- **Shot-level optimization:** After detecting the shot boundaries, the N_{KF} key frames are distributed among different shots. Key frames can be allocated based on shot length, motion, significance variations, etc. Now, for each shot s_i, we know the number of key frames N_i that need to be selected. These frames can be selected to minimize some distance metric. Another option is to run the SVD algorithm to find the most significant frames in each shot.