Math from paper [20] –

In our cognitive adaptation system, video is adapted when the available network bandwidth changes. As in \cite{DASH}, we estimate the link capacity and alter segments based on the bandwidth. The multidimensional adaptation operation that we perform is dependent on a class label predicted at the time of adaptation. In order to predict this label, we use two support vector machines which learn the user’s perception of the importance of frame rate, frame size, and SNR resolution under different contexts. The context is determined based on features of the video and the streaming environment, such as network bandwidth and video content type. The predicted label takes on values from 0 to 3, conveying the following information:

* Label 0: User weighs both frame rate and frame size of equal importance, while quality is of less importance
* Label 1: User weighs frame rate as highest importance, with quality of secondary importance and size of least importance
* Label 2: User weighs frame size as highest importance, with quality of secondary importance and frame rate of least importance
* Label 3: User weighs quality of highest importance, with frame size and rate of less importance

The motivation for using this model of labeling is that we only desire to know the order of preference the user has for each video dimension. With this information, we perform a simple calculation that will selectively alter the encoding parameters based on the user’s preference. This will be highlighted in the discussion on multidimensional adaptation.

Our system determines the correct labels based on the quality, spatial, and temporal resolution of the video when a training sample is recorded. For the purposes of this study, where the primary motivation was to show the utility of support vector machines in multidimensional video adaptation, we employ a heuristic in which these resolutions are inspected and the ground truth is found by comparing them to a threshold value. For example, suppose the maximum frame size is CIF (352x288) and the maximum frame rate is 30fps. We can define the threshold to be half the max resolution value. In this, if the user is viewing video at QCIF frame size and 30fps, they are given label 1 as defined above. In the same way, if they are viewing video at CIF frame size and at 15fps, they will be given label 2. The remaining labels are determined in the same fashion. This method serves well for our purposes, and can be replaced in future studies by using an intelligent labeling scheme such as clustering.

The perceived QoE of a video stream is highly dependent on how the video adapts to changes in network capacity. Multidimensional video adaptation, in which the temporal, spatial, and quality resolution are subject to alteration, is essential for serving various users on disparate devices with individual preferences. The multidimensional adaptation problem in this study is characterized in \cite{wang2005classification} by the following:

where is a multidimensional adaptation function defined in the space **A**. This model states that the correct adaptation function to choose should maximize the utility , which can be any metric that represents the quality of the user experience. The constraint is simply the available resource, in our case being the bandwidth of the network, and , the new resource requirement due to the adaptation, should be strictly less than this. We can equally represent the constraint as . In addition, we assume that , the total bitrate of the video after the adaptation operation. We model quality resolution in bits per pixel as follows:

where is the spatial resolution of the video after adaptation and is the temporal resolution after adaptation. In the adaptation operation, we select the spatial resolution and temporal resolution from a discrete set of values, then calculate the quality resolution as a result of these selections. We can call these sets **C** and **S**, where is spatial resolution , and is temporal resolution , . We assume that can be determined within a reasonable degree of accuracy. We then select a value for K, representing the percentage of available bandwidth that is acceptable to fill and providing a bit budget that we use to optimize the video encoding parameters. For this study we also define a value , being the minimum target bits per pixel. This value is selected heuristically, but it can theoretically be configured or learned, representing the weight given to quality resolution. To perform the correct adaptation operation, we use the class label predicted by our support vector machines. The operations are defined below.

* : Select and , is calculated using equation (above)
* : Select and ; maximize such that . If none exists, select and calculate using equation (above)
* : Select and ; maximize such that . If none exists, select and calculate using equation (above)
* : Minimize and , calculate such that

These operations provide an effective way to maximize the QoE based on user classification. Each video parameter is maximized according to the preference of the user, giving precedence to the parameter found most important. In our performance evaluation, we will compare the above method against single dimensional SNR resolution adaptation.