Demonstration of Improved Multimedia Streaming by Using Content-Aware Video Scaling

Avanish Tripathi and Mark Claypool
Department of Computer Science
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA-01609, USA
{avanish,claypool}@cs.wpi.edu

1. INTRODUCTION

The Internet disseminates enormous amounts of information for a wide variety of applications all over the world. As the number of active users on the Internet has increased so has the tremendous volume of data that is being exchanged between them, resulting in periods of congestion in the network. To overcome this congestion various congestion control strategies have been built into the Transmission Control Protocol (TCP), the de facto transport protocol on the Internet. For multimedia traffic however, TCP is not the protocol of choice because unlike traditional data flows, multimedia flows do not necessarily require a completely reliable transport protocol since they can absorb a limited amount of loss without significant reduction in perceptual quality. Also, multimedia flows have fairly strict delay and delay jitter requirements that are difficult to meet when using TCP [2]. Multimedia flows generally use the User Datagram Protocol (UDP). This is significant since UDP does not have a congestion control mechanism built in, therefore most multimedia flows are unable to respond to network congestion and adversely effect the performance of the network as a whole.

Media scaling is one technique to reduce bandwidth of video streams. Media scaling techniques for video can be broadly categorized as follows [1]:

- *Spatial scaling:* In spatial scaling, the size of the frames is reduced by transmitting fewer pixels and increasing the pixel size, thereby reducing the level of detail in the frame.
- *Temporal scaling:* In temporal scaling, the application drops frames. The order in which the frames are dropped depends upon the relative importance of the different frame types.
- *Quality scaling:* In quality scaling, the quantization levels are changed, chrominance is dropped or DCT and DWT coefficients are dropped. The resulting frames are of a lower quality and may have fewer colors and details.

It has been shown that the content of the stream can be an important factor in influencing the choice of the preferred scaling technique (i.e. temporal, spatial or quality) [1]. For instance, if a movie scene has high motion and had to be scaled then it would look better if all the frames were played out albeit with lower quality. That would imply the use of either quality or spatial scaling mechanisms. On the other hand, if a movie scene has little motion and needed to be scaled it would look better if a few frames were dropped but the frames that were shown were of high quality. This work demonstrates quantitatively the effects of the different scaling techniques on perceptual quality of video, which to the best of our knowledge, has not been done before.

2. APPROACH

In order to successfully develop a system that makes scaling decisions based upon the amount of motion in the video stream, we develop an automated means of measuring the amount of motion in the stream in real-time and then integrate this with a media filtering system [5]. The whole system is then capable of making content-aware decisions for the choice of the scaling mechanism to use for a particular sequence of frames.

2.1 Motion Measurement

In our system, we have used an MPEG video stream to explore our approach. A typical MPEG stream contains three types of frames: Intra-encoded frames (I), Predicted frames (P) and Interpolated frames (B-for Bidirectional prediction). Each frame is further decomposed into 16x16 blocks called macroblocks, the basic motion-compensation unit. All macroblocks in the I-frames are encoded without prediction and the I-frame is thus independent of any other frames. The macroblocks in the P-frame are encoded with forward prediction with references made from previous I-frames and P-frames or may be intra-coded. Macroblocks in B-frames may be coded with forward or backward prediction from past or future I-frames or P-frames, with interpolated prediction from past and future I-frames or P-frames or they may be intra-coded.

Our system uses the percentage of interpolated macroblocks in the B-frames as a measure of motion. A high number of interpolated macroblocks implies that a greater portion of the frame is similar to frames that are already existing in the stream (i.e. less motion) and a low number of interpolated macroblocks implies that there are a greater number of changes between frames (i.e. more motion). Our pilot studies have shown this methodology to work well when making decisions regarding scaling policies.

For our system, we need to categorize the sequence of frames into two categories, low motion or high motion. Sequences having greater than 45% interpolated macroblocks generally fall in the low motion category while those having less than 45% interpolated macroblocks could be classified as having high motion. Also, the 45% cut-off appears to statistically place half the selected clips in the high motion category and half in the low motion category. This classification may be made more fine grained as the need arises.

2.2 System Architecture

The system consists of 4 distinct modules: *server*, *filter*, *network feedback generator* and *client*.

- *Server:* The server in the system takes as input an MPEG file, parses and packetizes it and streams it over the network to the client. The server is also capable of quantifying the amount of motion in the video stream by using the motion measurement sub-module.
- *Filter Module:* The filter module has two kinds of filters: *Temporal Filter* and *Quality Filter*. The temporal filter is a frame dropping filter and does scaling in the temporal domain. The quality filter is a re-quantization filter and scales the video stream in the quality domain. These modules are an extension of the work done by [5].
- *Network Feedback Generator:* The network feedback generator module resides on the client side and monitors the congestion in the network by keeping track of the sequence numbers of the packets received. Dropped packets (i.e. packet loss) are taken to be a measure of network congestion. In the event of congestion the feedback module will send control messages to the server indicating a reduction in available bandwidth. At predefined intervals the feedback generator can have the server increase the bandwidth to probe for the optimal level. Our system is compatible with most responsive, rate based multimedia protocols, such as [2, 3]
- *Client:* The client module is a regular MPEG decoder that is capable of playing out frames that are received over network sockets.

3. RESULTS

We conducted a user study with 22 users in order to evaluate the effectiveness of our content-aware scaling system. Figure 1 shows the graph of the user perceived quality against the different scaling levels for a high motion clip (a car commercial). This clip has an average of 20% interpolated macroblocks over its entire duration. We observe that quality scaling is consistently better to users than temporal scaling for this high motion clip. For low motion clips we found that

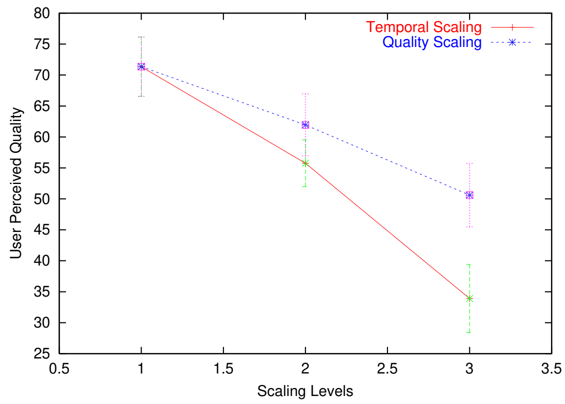

Figure 1: High Motion Clip (20% Interpolated Macroblocks)

temporal scaling is consistently better than quality scaling. Further studies [4] with longer clips having multiple scene changes have shown that although frequent scene changes do have a detrimental effect on perceptual quality, the overall quality is still better than if the clip were not scaled at all in the presence of congestion.

4. DEMONSTRATION

We demonstrate the effectiveness of our system by showing four different clips (2 high motion and 2 low motion) at full quality, quality scaled and temporally scaled. The system scales the video stream temporally if the stream has low motion and it does quality scaling if the video stream has high motion. We demonstrate that the perceptual quality of the high motion clips is better when they are quality scaled whereas the perceptual quality of the low motion clips are better when they are temporally scaled. Further details on this work, including additional results, can be found in [4].

5. REFERENCES

[1] P. Bocheck, A. Campbell, S.-F. Chang, and R. Lio. Utility-based Network Adaptation for MPEG-4 Systems. In *Proceedings of Ninth International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, June 1999.

[2] J. Chung and M. Claypool. Better-Behaved, Better-Performing Multimedia Networking. In *Proceedings of SCS Euromedia Conference (COMTEC)*, May 2000.

[3] S. Floyd, M. Handley, J. Padhye, and J. Widmer. Equation-Based Congestion Control for Unicast Applications. In *Proceedings of ACM SIGCOMM 2000*, August-September 2000.

[4] A. Tripathi and M. Claypool. Improving multimedia streaming with content-aware video scaling. Technical Report WPI-CS-TR-01-02, Worcester Polytechnic Institute, March 2001.

[5] N. Yeadon, F. Garcia, D. Hutchinson, and D. Shepherd. Continuous Media Filters for Heterogeneous Internetworking. In *Proceedings of SPIE Multimedia Computing and Networking (MMCN'96)*, January 1996.