DICTA2002: Digital Image Computing Techniques and Applications, 21--22 January 2002, Melbourne, Australia

Bandwidth Adaptation for MPEG-4 Video Streaming over the Internet

K. Ramkishor
Architect – Video
Emuzed India Ltd
Bangalore, India
ramki@emuzed.com

James. P. Mammen
Graduate Student
Standford University
California, USA
jmammen@stanford.edu

Abstract

Video streaming over time varying channels like the Internet involves adapting the source bit rate to the continuous fluctuations in the available bandwidth. Although small bandwidth variations can be managed by changes in the quantisation scale, frame skipping is essential for considerable changes in bandwidth in order to maintain good perceptual quality. This paper presents techniques for adapting the encoded video to the varying bandwidth by having a dynamically varying frame rate algorithm with good spatio-temporal quality tradeoff at the encoder. Strategies for adapting the frame rate in real-time to satisfy the bandwidth constraint and maintain good perceptual quality are presented. Experimental results show that the encoder is able to adapt to a huge range of bandwidth variations.

1. Introduction

The high amount of audio-visual data associated with typical multimedia services call for efficient data compression schemes in order to facilitate transmission and storage applications. Several international standards have been introduced for video compression targeting different application fields. MPEG-1 and MPEG-2 cater to storage media (VCD, DVD) and digital television applications. Communication applications (i.e., video conferencing, video-phone etc.) are covered by the ITU-T standards H.261, H.263 [1], and H.263+. The new MPEG-4 [2] standard offers improved coding efficiency and introduces the concept of content based coding making it better suited to video streaming over the internet and wireless channels.

The compression method in these standards comprises of preprocessing of the video sequence followed by motion estimation, motion compensation, DCT, quantization, and variable length encoding in that order. Motion estimation and motion compensation intend to remove the temporal redundancy presented in natural video sequences while DCT reduces the spatial redundancy.

The Internet offers a constantly varying bandwidth due to its heterogeneity and congestion. The encoder needs to continuously adapt to the changes in available bandwidth while encoding a video stream for transmission over the Internet [3], [4], [5]. It is possible to adapt to small changes in bandwidth by changing the quantisation scale, which in turn affects the visual quality of the video. However, even quantisation at the highest scale may not be sufficient to produce a compressed video stream, which satisfies the available bit rate constraint when the bandwidth drops by a considerable amount. One of the solutions provided by modern video compression standards such as H.263+ and MPEG-4 is scalable or layered video coding. These schemes consist of a video stream with a base layer and one or more enhancement layers so that when the available bandwidth decreases one or more of the enhancement layers can be dropped. Three types of scalability are possible. They are temporal scalability, spatial scalability and SNR scalability. Temporal scalability is commonly achieved by skipping frames. Spatial scalability and SNR scalability increase the complexity of the encoder and the decoder. Moreover there is a significant overhead associated with layering which makes it undesirable for low bit rate video applications. Another solution is called dynamic stream switching [4] where the server switches among multiple streams to serve a client with one that best matches the client's available bandwidth. This solution makes the encoding process more complex as synchronization points will be required in each of the multiple streams where switching can take place. The simplest and most effective approach to accommodate large variations in bit rate is to skip frames, thereby changing the frame rate at which the video is encoded. In order to increase its effectiveness variable frame rate encoding may also be combined with layered coding approach when the overhead due to layering is acceptable.

The solution proposed in this paper involves adaptation at the encoder based on feedback provided by the streaming server. A dynamic frame rate is chosen so that

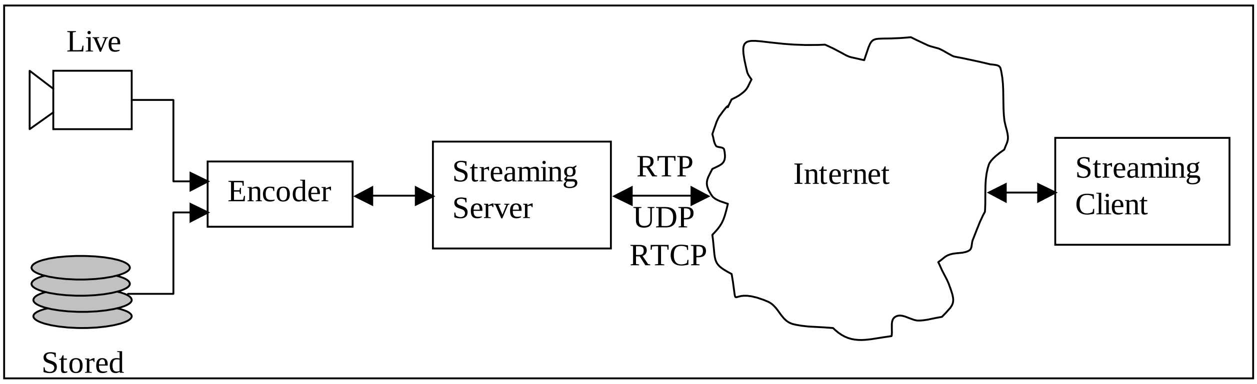

Figure 1. Model for Video Streaming Over the Internet

the quality of frames is in tolerable range under sudden motion change and time varying communication channel environments without obvious degradation in the perceptual motion smoothness. Since, it is difficult to support good quality in both spatial and temporal resolution, the frame rate is adapted for a tradeoff of spatial/temporal quality based on the motion in the video and the available channel bandwidth. Section 2 presents the model used for video streaming over the Internet and explains the two conditions in which frames are skipped in order to provide good quality video while being constrained by the available bit rate. Section 3 explains the algorithms for deciding the skipping of frames under both conditions. This is followed by experimental results of encoding various video sequences at various bit rates using the MPEG-4 simple profile encoder.

2. Streaming Model and Frame Skipping Conditions

The model used for video transmission over the Internet is described in Figure 1. The encoder compresses the raw video available from some source and passes it to a streaming server, which streams it over the Internet to the streaming client. The streaming server uses the Real Time Streaming Protocol (RTSP), which employs RTP/UDP/IP for data and Real Time Control Protocol (RTCP) for feedback and control. The packet loss ratio available from the RTCP channel is used to estimate the available bandwidth in various ways as described in [6], [7], [8], etc. The client communicates the packet loss observed by it through RTCP. Depending on the packet loss, the server either increases the bandwidth or decreases the bandwidth. This information regarding the available bandwidth is given by the streaming server to the encoder, which encodes video at this bit rate. Until the bit rate changes again the encoder assumes a Constant Bit Rate (CBR) channel at this bit rate. Thus the encoder uses a time varying CBR model to characterize the channel behavior.

The work presented in this paper deals with the MPEG-4 standard for video streaming over the Internet. However it can be applied to other video compression standards like MPEG-2, H.263+, etc with suitable modifications. Moreover it can be used for bandwidth adaptation over any time varying channel, where feedback regarding the available bandwidth is available.

When the available bandwidth decreases, frames can be encoded using a coarser quantisation scale in order to meet the bandwidth constraint. However, when the drop in bandwidth is high it may not be possible to encode at the specified bit rate even with the highest quantisation scale. In this case it is necessary to skip frames in order to meet the bandwidth constraint. This is referred to as frame skipping for satisfying the bandwidth constraint. However in some cases, depending on the bit rate and the complexity of the video to be encoded, it may be possible to encode using a very high quantisation scale. It may sometimes be visually more pleasing to view video at a lower frame rate with better spatial quality in each frame than a higher frame rate video with very coarse quantisation resulting in very low spatial quality in each frame. In such cases, it is desirable to skip frames in order to increase the spatial quality of the video by making available the extra bits saved by skipping frames to those that are not skipped. This is referred to as frame skipping for quality enhancement. Depending on the available bit rate and the encoding complexity of the video, frames may be skipped for enhancing the quality or for meeting the bit rate constraint using the algorithms described in the following section. Thus there is a dynamically varying frame rate throughout the video sequence, which depends on the available bandwidth and the complexity of the video content to be encoded.

3. Bandwidth Adaptation Algorithm

The dynamic frame rate at any instant is denoted by f and is related to the actual frame rate of the video content denoted by F as follows.

$$f = s*F$$

Where s is the dynamic frame rate scaler. For example, assuming the actual frame rate F = 20 frames per second and a dynamic frame rate scaler s = 0.5, the dynamic frame rate would be f = 10 frames per second which corresponds to skipping 1 in every 2 frames. Based on the algorithms proposed below, a decision is taken to skip m out of n frames leading to a dynamic frame rate scalar of $s = (n-m)/n$. The bits saved by skipping the m frames are distributed among the remaining $n-m$ frames. However, this distribution is not even as explained below.

In video compression algorithms, for achieving high compression only the motion compensated residual is coded and transmitted. When one or more frames are skipped before a frame to be coded, its temporal correlation with the last frame that was not skipped is less compared to that of consecutive frames due to the increased temporal distance. As a result, in order to maintain the same spatial quality it requires more bits. Hence out of the $n-m$ frames, the first frame needs to be given a greater proportion of the bits saved by skipping the m frames.

The decision making for the dynamic frame rate is made based on whether there is a need to skip frames in order to either meet the bandwidth constraint or to enhance the spatial quality. Since producing a compressed video stream at the stipulated bit rate is essential, the decision to skip frames for maintaining the target bit rate is taken first. The decision to skip frames for enhancing the spatial quality is taken only if no frames are required to be skipped for meeting the bit rate constraint. It must be noted that both decisions can affect each other as the effect of both kinds of frame skips is to increase the number of bits available to the frames that are not skipped. Thus by skipping frames for enhancing the quality, there will be a loosening of the bandwidth constraint and by skipping frames for satisfying the bit rate constraint the additional bits contribute to increased spatial quality. The algorithms proposed for making these decisions are described below.

3.1. Skipping Frames for Satisfying Bandwidth Constraint

The decision to skip frames in order to satisfy the bandwidth constraint needs to be taken when the current frame rate is not able to meet the stipulated bit rate. Rate control modules in video compression algorithms work by assigning a number of bits called the target bits to a frame based on its encoding complexity and the available bit rate. Thereafter a quantisation scale is selected such that after encoding the number of bits produced is close to the target bits. For a tighter rate control incorporating perceptual measures, a macroblock layer rate control scheme is also applied, which selects the quantisation scale for each macroblock based on its encoding complexity. If the actual number of bits required for encoding the frame turn out to be much in excess of the target bits in spite of a high quantiser, it indicates that the frames are unable to satisfy the bit rate constraint.

Hence after encoding a frame, based on the number of bits that are consumed in excess of the estimated target (*ExcessBits*), a decision is taken to skip either 1 out of every 3 frames (s = 0.667) or 1 out of every 2 frames (s = 0.5) or 2 out of every 3 frames (s = 0.333) or 3 out of every 4 (s = 0.25) frames. The pseudo code for the proposed scheme follows: where Q is the average quantisation scale over the whole frame and $AvgBits = BitRate/F$ i.e. the average number of bits per frame.

Pseudo code:

```
if(Q > Qthresh && ExcessBits > 100)
{
  if(ExcessBits > 0.5*AvgBits)
      s = 0.250;  /* Skip 3 in 4 */
  else if(ExcessBits > 0.25*AvgBits)
      s = 0.333;  /* Skip 2 in 3 */
  else if(ExcessBits > 0.1*AvgBits)
      s = 0.500;  /* Skip 1 in 2 */
  else if(ExcessBits > 0.05*AvgBits)
      s = 0.667;  /* Skip 1 in 3 */
  else
      s = 1.000;
}
else
{
      s = 1.000;
}
```

For the case of MPEG-4 video, where the quantisation scale is restricted to the range of 1 to 31, Qthresh is selected as 25. The value is selected after doing extensive experimentation over wide range of sequences.

3.2 Skipping Frames for Enhancing Quality

The decision to skip a frame for enhancing the quality has to be made if the extra bits saved by skipping a frame can be used to enhance the spatial quality of other frames thus offsetting the jerkiness introduced due to skipping the frame. The spatial quality of a frame is dependent on how finely or coarsely it is quantised, i.e. the average quantisation scale (*Q*) over the frame. Moreover, in general, when the same quantisation scale is used for two different frames, the frame with higher encoding complexity appears to be at a higher visual quality. The encoding complexity is measured using the Mean of Absolute Differences (MAD) over the whole frame from its prediction after doing motion compensation.

Hence the decision for skipping a frame is based on the adaptively changing threshold and the threshold is dependant on the average quantisation scale (Q) and average MAD of the frame. This kind of adaptively selecting the threshold takes care of sequences with different complexities. Here, frame skips of 1/4, 1/3 and 1/2 are supported, where 1/*n* indicates that 1 out of *n* frames is skipped. The corresponding dynamic frame rate scalers are s = 0.75, 0.667 and 0.5, respectively. When there is significant motion in the video content, the increase in spatial quality cannot be perceived due to the motion, whereas the jerkiness due to loss of frames is still visible. As a result, frame skipping for enhancing the quality is not attempted if there is high motion in the video. The amount of motion in the video can be measured using the average of the absolute values of the motion vectors used for motion compensated prediction.

Pseudo code:

```
if(motion is not high)
{
   if(Q > Q1)
       s = 0.500;  /* Skip 1 in 2 */
  else if(Q > Q2 + MAD)
       s = 0.667;  /* Skip 1 in 3 */
   else if(Q > Q2 + MAD/2)
       s = 0.750;  /* Skip 1 in 4 */
  else
       s = 1.000;
}
else
{
       s = 1.000;
}
```

For the case of MPEG-4 video, *Q1* and *Q2* are selected as 25 and 13, respectively. The motion is considered to be not high when the mean of the absolute values of motion in both vertical and horizontal directions are less than 5 pixels. The values are chosen after doing experimentation over wide range of sequences and evaluating the subjective quality. The sequences used in the evaluation are Foreman, Carphone, Akiyo, Silent, CoastGuard, Bus, Tempete, Paris etc.

4. Results

Since, the encoder has a time varying CBR view of the channel, it is sufficient to test the algorithm for bandwidth adaptation at various constant bit rates which cover the entire range over which it would be expected to vary with time. The proposed scheme is tested for several sequences at various spatial resolution, frame rates and bit rates. The results for a few representative sequences are tabulated in 0. The video is encoded using the MPEG-4 standard (simple profile). The details of the sequences are given in the first column followed by the bit rates at which they are required to be encoded. The number of frames skipped by the proposed algorithm for meeting the bandwidth constraint and enhancing the quality are listed next. The PSNRs and encoded bit rates with and without the bandwidth adaptation scheme form the last four columns of the table.

It can be observed that very few or no frames are dropped when the bit rates are high enough for the complexity of the video sequence to be encoded. At higher bit rates most of the skips are for enhancing the quality whereas at very low bit rates, most of the frame skips are for satisfying the bandwidth constraint. It can also be seen that for the bus sequence where there is high motion no frames are skipped for enhancing the spatial quality since the enhancement would not be perceivable because of high motion.

The table shows that when the bandwidth adaptation scheme is not used, the encoder is not able to produce a compressed video at the stipulated bit rate. For the QCIF sequences like ‘carphone’, ‘coastguard’ and ‘foreman’, the encoded bit rate exceed the bit rate constraint by more than 50% when the bit rate is as low as 16kbps. Whereas the bandwidth adaptation algorithm skips sufficient number of frames in order to maintain the bit rate and at the same time provide better spatial quality as seen by the increase in PSNR.

5. Conclusions

A bandwidth adaptation strategy for video streaming over time varying channels like the Internet is presented. It is implemented at the encoder, which in turn obtains feedback from the streaming server regarding the available bandwidth. The encoder skips frames based on the available bit rate and the encoding complexity of the sequence in order to meet the stipulated bit rate and to produce acceptable quality video. The results show that this scheme is able to adapt to a wide range of bit rates for a variety of sequences. Although the schemes presented in this paper are for the specific case of the

MPEG-4 standard, they can be easily used for other video coding standards with corresponding changes.

6. References

[1] "Video Coding for Low Bit Rate Communications", *H.263 Standard*, ITU-T Recommendation H.263, Feb. 1998.

[2] "Information Technology – Generic Coding for Audio-Visual Objects – Part 2: Visual," *MPEG-4 Standard, 14496-2, ISO/IEC Standard*.

[3] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, and H. Kimata, "RTP Payload Format for MPEG-4 Audio/Visual Streams," *RFC3016, Internet Engineering Task Force (IETF)*, November 2000.

[4] Jian Lu, "Signal Processing for Internet Video Streaming: A Review", *Proc. of SPIE Image and Video communications and Processing*, Jan 2000.

[5] Dapeng Wu, Yiwei Thomas and Ya-Qin Zhang, "Transporting Real-time Video over the Internet: Challenges and Approaches," *Proceedings of IEEE*, Vol. 88, December 2000.

[6] D. Wu, Y. T. Hou, W. Zhu, Ya-Qin Zhang and H. J. Chao, "MPEG-4 Compressed Video over the Internet," *ISCAS*, 1999.

[7] D. Wu, Y. T. Hou, J. Yao and H. J. Chao, "Transmission of Real-Time Video over the Internet: A Big Picture", *IEEE NetWorld+Interop*, May 2000.

[8] J. Kim, Y. Kim, H. Song, T.-Y. Kuo, Y. J. Chung and C.-C. Jay Kuo, "TCP-friendly Internet Video Streaming employing Variable Frame-rate Encoding and Interpolation," *PCS*'99.

[9] Hwangjun Song, Jongwon Kim, C. C. Jay Kuo, "Real-time Encoding Frame Rate Control for H.263+ Video over the Internet," *Signal Processing: Image Communication*, September 1999.

Table 1. Results of the Proposed Bandwidth Adaptation Scheme for Some Representative Sequences

Sequence Size @*F*(fps) Length	Bit Rate (kbps)	Bandwidth Skips	Quality Skips	Total Skips	With Bandwidth Adaptation		Without Bandwidth Adaptation	
					PSNR (db)	Encoded Bit Rate (kbps)	PSNR (db)	Encoded Bit Rate (kbps)
Akiyo QCIF @15fps 300	32	0	0	0	38.29	31.95	38.29	31.95
	24	0	1	1	36.37	23.76	36.40	23.76
	16	7	17	24	34.01	15.98	33.78	15.97
Carphone QCIF @15fps 382	64	0	9	9	33.15	64.01	33.02	64.01
	32	34	62	96	31.13	31.85	29.83	32.81
	24	99	55	154	30.40	23.80	29.05	28.63
	16	208	17	225	29.60	15.76	28.75	27.02
Coastguard QCIF @15fps 300	64	0	0	0	30.54	63.91	30.54	63.91
	32	2	69	71	28.23	31.95	27.47	31.95
	24	22	91	113	27.46	23.76	26.69	24.58
	16	154	20	174	26.97	15.97	26.64	24.17
Foreman QCIF @15fps 400	64	1	0	1	31.84	63.90	31.84	63.90
	32	50	61	111	29.69	31.95	28.61	31.95
	24	118	58	176	29.08	23.96	28.13	28.88
	16	229	25	254	28.22	15.97	28.07	28.57
Bus CIF @30 150	384	9	0	9	26.02	383.39	25.86	383.39
	256	86	0	86	26.29	255.59	25.68	362.09
	192	100	0	100	25.78	193.33	25.66	360.45
Tempete CIF @30 150	384	0	37	37	28.05	383.61	27.57	383.61
	256	1	82	83	27.30	254.4	26.04	256.79
	192	23	82	105	26.66	190.62	25.69	232.37