Available online at www.elixirpublishers.com (Elixir International Journal)

Network Engineering

Elixir Network Engg. 42 (2012) 6310-6312

Elixir ISSN: 2229-712X

An improved rate distortion – optimized rate shaping for Spatio – temporal scalability for applications in heterogeneous communication networks

Aravinda TV[1] and G Mahadevan[2]
[1]PRIST University, Thanjavur
[2]Department of Computer Science Engineering, AMCEC, Bangalore.

ARTICLE INFO

Article history:
Received: 10 November 2011;
Received in revised form:
2 January 2012;
Accepted: 14 January 2012;

Keywords
Rate Distortion,
Scalable Video,
Heterogeneous communication networks.

ABSTRACT

Today, the wide variety of devices in the digital world ranges from desktops to mobile phones. Within the currently available interactive multimedia applications, which are very demanding in terms of video quality and coding efficiency, the cost as well as the limited performances of scalability obtained in the current standards remains unacceptable. That is why; there is a need for intrinsically scalable video coding schemes providing fully progressive bit streams. Scalable Video Coding targets on seamless delivery of digital content and access to the same, enabling optimal user centered multi-channel and cross-platform media services, providing a straightforward solution for universal video delivery to a broad range of applications. Scalable video coding gives a nice way to perform rate shaping for video streams adapting to the available transmission resource. The work in this paper deals with a rate distortion approach in Scalable Video Coding in order to achieve a performance that is comparable to a non-scalable system. The rate distortion curves of the scaled video streams are employed to adapt to both static and dynamic channels. We consider a multiuser scenario which reflects network heterogeneity and propose to perform a joint optimization between these multiple streams which have different performance curves.

Introduction

The proliferation of multimedia on the World Wide Web and the emergence of broadband wireless networks have brought great interest in video communication. With the materialization of Web as a strong competitor for conventional distribution networks, a main challenge relates to the production of easily adaptable video content capable of optimally fitting into various evolving platforms. Network supported multimedia applications like in-home digital networks, video streaming over IP networks, surveillance systems, mobile video, wireless LAN video, multi-party video telephony/conferencing involves many different transmission capabilities.

These applications are used to deliver content to a wide range of terminals and users surrounded by different environments and acting under to tally different circumstances. The challenge now is to make information easily retrievable for a variety of systems.

Improvement in bitrate efficiency between MPEG-2 [ITc] and MPEG-4 [1a] is not significant and new techniques are required to overcome this limitation and to enable a quick and easy access to large multimedia data repositories. To reach this goal, information must be customized in accordance with the various network systems and the features of their devices. Many uncertain parameters exist in the network such as speed, load and bandwidth. Consequently, requirements for bandwidth availability and quick and easy access to large multimedia databases will be more and more stringent. Therefore, meeting bandwidth requirements and maintaining acceptable video quality, simultaneously is a challenge (fig – 1).

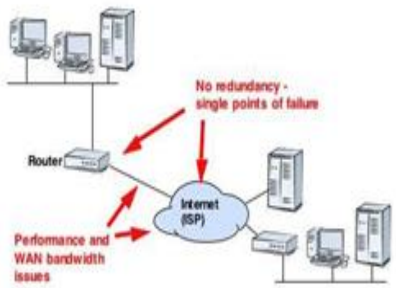

Figure 1: Network model

Continuous rate scalable application can prove valuable in scenarios where the channel is unable to provide a constant bandwidth to the application.

Rather than terminating the session, a decoder can adjust the data rate to use the limited resources, yet produce video of acceptable quality. Such decoders are particularly attractive because of their flexibility.

Scalable Video Coding (SVC) opens the door to some new video coding techniques with the following features:

- Reduced bitrate
- Reduced spatial-temporal resolution
- Coding efficiency comparable to non-scalable video systems

Scalable Video Coding addresses the issue of reliably delivering video to diverse systems over heterogeneous networks using available system resources, particularly in scenarios where the downstream system capabilities, resources, and network conditions are not known in advance. The goal is to provide scalability at bitstream level with good compression

Tele:
E-mail addresses: arvind_cta@yahoo.co.in,
g_mahadevan@yahoo.com

efficiency by allowing free combinations of scalable modes such as spatial, temporal and SNR/fidelity scalability. This new work is presently developed by the Joint Video Team (JVT) as an extension of AVC/H.264 [RSS].

Spatial Scalability

Spatial scalability is a type of scalability in which a higher layer uses predictions from data derived from a lower layer without using motion vectors. The layers can have different frame sizes, frame rates or formats. Typically, spatially scalable video is encoded in an efficient way with representation of the same video in different spatial resolutions or sizes. Fig. 2 shows a block diagram of a two layer spatially scalable coding system. For the base (lower) layer, the raw video is first spatially down-sampled, then DCT transformed, quantized and VLC coded. For the enhancement (higher) layer, the base layer image is reconstructed by inverse quantization and inverse DCT. The base layer image is spatially up-sampled and the up-sampled base layer image is subtracted from the original image. The residual is then DCT-transformed and quantized by a quantization parameter smaller than that of the base layer and finally coded by VLC.

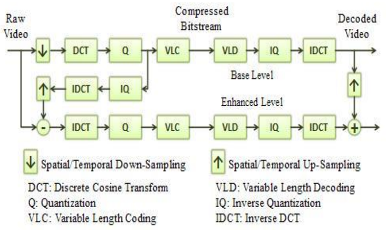

Figure 2: Spatial/Temporal scalability

At the decoder end, the spatially scalable decoder operates exactly as the non-scalable video encoder for the base layer. For the enhanced layer, both layers must be received, decoded by VLD, inversely quantized and inversely DCT transformed. Then the base layer image is spatially up-sampled and is combined with the enhanced layer refinements to form enhanced layer decoded video.

Temporal Scalability

The temporal scalability allows diffrent picture rates. It is defined as representing the same video in different temporal resolutions or frame rates. The block diagram of temporally scalable codec is the same as that of spatially scalable codec, shown in Fig. 2. The only difference is that the spatially scalable codec uses spatial down-sampling and spatial up-sampling while the temporally scalable codec uses temporal down-sampling and temporal up-sampling.

SNR Scalability

The SNR scalability allows the enhancement of the video quality by using different quantization parameters. It represents the same video in different SNR or perceptual quality. A SNR scalable coder with two level scalability is depicted in Fig. 3. For the base level, the SNR scalable encoder operates in the same manner as that of a non-scalable video encoder. However, for the enhancement level, the base layer DCT coeffcients are reconstructed by inverse quantization and subtracted from the original DCT coeffcients. The residual is quantized by a quantization parameter, which is smaller than that of the base level and the quantized bits are VLC coded.

The decoder performs in the reverse fashion. In general, spatial, temporal and SNR scalability provides multiple video representations in different spatial, temporal and quality resolutions. Each video representation has different significance and and width requirement. The base layer is more important while the enhanced layer is less important. The base layer needs less transmission bandwidth due to its coarser quality while the enhanced layer requires more transmission bandwidth due to its finer quality.

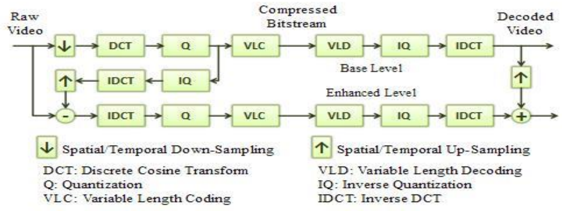

Figure 3: SNR scalability

Approach Proposed

RD Optimization

A widely investigated approach to deal with varying importance of video frames is rate distortion (RD) optimization. RD-optimized frame handling has been successfully used in many situations. Consider again the scenario in Equation - 1, where K video streams arrive at an active network node and leave the node at the same outgoing link. The outgoing link has a transmission rate R_{out}. The outgoing link has a link buffer size of size B_{max} (Bytes). When incoming streams have different bitrates, the decision which packets (NAL units) to discard has to be jointly made for all the streams to adapt to the available bandwidth. The rate shaping strategy proposed here, relies on the scalable vector that is sent as side information along with the video streams. The scalable vector consists of the reconstruction distortion observed for different scalability levels and the required bitrate for the particular scalability. The side information can be used by the network nodes to dynamically decide in a RD-optimized way which packets of which layers of the respective video stream should be dropped in case of node overload. Basically, the nodes have to choose an optimal scalable functionality for each of the video sequences to meet network resource limitations. We propose a Lagrangian Cost Function that uses the scalable vectors of all the incoming streams together with the outgoing link capacity to find the optimum dropping pattern at the respective network node. The dropping strategy is based on maximizing the Lagrangian Cost Function J.

$$J(n) = \sum_{(k=1)}^{k} \left[P(n)ks - \lambda(n) \sum_{(k=1)} R(n)ks \right] \qquad (1)$$

$P(n)^k s$ is the PSNR quality achieved for video k for dropping strategy s and $R(n)^k s$ is the required bitrate for dropping strategy s. We replaced the continuous time t by the GOP index n where n is the GOP duration since the dropping decision is made only at multiples of GOP duration. In case the streams have different GOP durations, the dropping decision is made synchronized to the stream with the highest GOP duration. λ(n) is set to a value equal to 1 for all n since it is assumed that all the network sources have sufficient buffer to store and forward all the incoming video packets. In other words, rate shaping of the video sequences to the available rate is independent pent of the buffer limitations in our proposed scenario. λ(n) can be modeled to different values to reflect further stringent buffer conditions. We, therefore, determine λ(n) as a function of buffer fullness B(n). If we denote the number of possible dropping strategies at time n for video k as $S^k(n)$ then for K videos, we have

$$C(n) = Y^{k}_{K=1} S^{k}(n) \qquad (2)$$

different ways to rate shape the video sequences. One of these dropping patterns will minimize the cost function (Equation 3) that represents the optimum strategy based on the current outgoing transmission rate. The main approach is realized using the following target functions;

max (mean(P SN Rk)) for all k=1..K video streams (3)

max(min(P SN Rk) for all k=1….K video streams(4)

The proposed target functions aim to maximize the average sequence quality and guarantee a minimum acceptable quality for the video streams.

Experimental Results

Results provided in this section correspond to qcif scenario1 (4:2:0). They are obtained by encoded soccer video at the various bit – rates, frame rates and resolutions described in table 1.

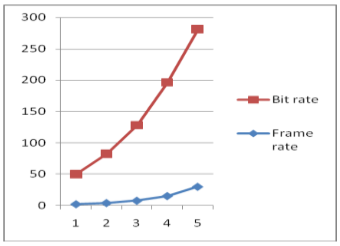

Figure 4: Performance evaluation of scalable coding for resolution 176x144

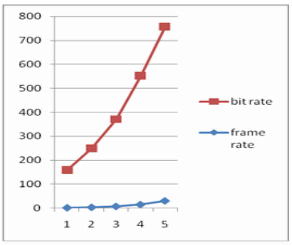

Figure 5: Performance evaluation of scalable coding for resolution 352x288.

Conclusion

We proposed a rate-distortion optimized video rate shaping strategy in case of heavy traffic load. QoS labeling of the video packets together with priority mechanisms, support importance controlled dropping of data. This technique allows to cover a very wide range of bit – rate and improves significantly video quality at lower spatial resolutions, with out noticeable penalty at high bit rate and full resolution.

References

[1] B. Haskell, A Puri, AN Netravali, Digital video: An introduction to MPEG - 2, New York: Chapman % Hall, 1997.

[2] T. Tsunashima, J. Stampleman, V. Bove, A scalable motion compensated sub band image coder, IEEE Trans. On communication, vol,42, pp. 1894 – 1901, 1994

[3] L. Yang, F.C.M. Martins, and T.R. Gardos, “Improving H.264+scalabilty performance for very low bit rate applications,” in Proc. SPIE-VCIP’99, San Jose, CA, Jan.1999, Vol.3653, pp.768-779.

[4]. ISO/IEC JTC 1. Coding of audio–visual objects – part2: visual. ISO/IEC 14496-2(MPEG-4 part 2), Jan.1999 (with several subsequent amendments and corrigenda).

[5] ISO/IEC JTC 1. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s - part 2: Video. ISO/IEC 11172 (MPEG-1), Nov. 1993.

[6] Philip A. Chou and Zhourong Miao. Rate- distortion optimized streaming of packetized media. In IEEE Transactions on Multimedia, Feb 2001.

[7] Pankaj Topiwala Gary J. Sullivan and Ajay Luthra. The h.264/avc advanced video coding standard: Overview and introduction to the fidelity range extensions. In SPIE Conference on Applications of Digital Image Processing XXVII, Aug 2004.

[8] K. R. Ramakrishnan G. Ravindra, N. Balakrishnan. Active router approach for selective packet discard of streamed mpeg video under low bandwidth conditions. In IEEE International Conference on Multimedia and Expo (ICME 06), Jul 2000.

Table 1: performance of video sequences for different resolutions

Layer	Resolution	Frame rate(frames/s)	Bit rate(Kbits/s)
0	176x144	1.8750	48.00
1	176x144	3.7500	78.53
2	176x144	7.5000	120.53
3	176x144	15.0000	181.13
4	176x144	30.0000	251.13
5	352x288	1.8750	157.00
6	352x288	3.7500	246.26
7	352x288	7.5000	364.26
8	352x288	15.0000	538.12
9	352x288	30.0000	728.12