RD-OPTIMIZED RATE SHAPING FOR MULTIPLE SCALABLE VIDEO STREAMS

Rajyalakshmi Mahalingam, Wei Tu, and Eckehard Steinbach

Media Technology Group, Institute of Communication Networks
Munich University of Technology, Munich, Germany
rajee.mahal@mytum.de, {wei.tu, eckehard.steinbach}@tum.de

ABSTRACT

Delivering streaming video over wired and wireless networks poses many challenges, primarily due to the throughput variations caused by time-varying network conditions. Scalable video coding gives an elegant way to adapt the video streams to the available transmission resource. In this paper, we consider a multi-user scenario wherein multiple scalable video streams compete for a shared transmission link with limited forwarding capacity. We propose a rate-distortion (RD) optimized rate shaping approach to improve the overall video quality. For this, compact RD side information is sent along with the sequences. Our simulation results show that significant improvements are achieved by the proposed RD-optimized rate shaping approach compared to conventional priority-based rate shaping.

1. INTRODUCTION

Within the currently available multimedia platforms, streaming video is evolving as a popular application. Owing to the dynamic heterogeneous network conditions and different user requirements, flexible transmission of streaming video pose a main challenge. H.264/AVC [1] provides excellent coding efficiency, however, this *Single Layer Coding (SLC)* scheme does not fulfill some requirements of the new media applications as it has limited freedom on scalability. On the contrary, Scalable Extension of H.264 [2] (marked as H.264/SVC in this paper) emerges as a new technology enabling *Scalable Video Coding (SVC)* and addressing the issue of reliably delivering video in a heterogeneous environment.

Recently, several studies have been performed on rate control and rate shaping for streaming video over the Internet. RD-optimized packet scheduling has been introduced in [3] and with reduced complexity in [4]. Jointly optimized frame dropping for multiple streaming videos is introduced in [5]. In [6], Hint Tracks are used as side information for RD-optimized frame dropping in a multi-user scenario. None of these works exploit all three types of scalability provided by H.264/SVC, namely, temporal, spatial and SNR scalability.

This paper considers the scenario shown in Fig. 1, where K scalable video streams arrive at a network node and compete for the transmission resource R_{out} on the shared outgoing link. Our RD-optimized rate adaptation approach relies on the RD side information, which is extracted during encoding and is sent along with the scalable video stream. The side information consists of an optimal combination of scalability modes, i.e., spatial, temporal and SNR/fidelity scalability for each of the video streams. The RD approach decides an optimal scaling pattern for every sequence and drops packets that are least important to the reconstruction quality. The objective is to maximize the overall video quality among all the users by forwarding the most important packets to the outgoing link.

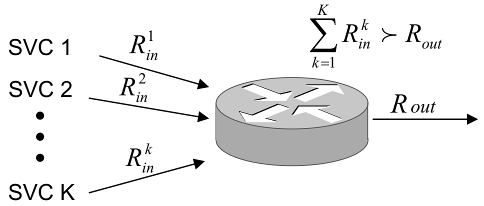

Fig. 1. A network node with K incoming scalable video streams sharing the same outgoing link

The paper continues with an introduction to scalable video coding and the description of RD side information in Section 2. The RD-optimized rate shaping strategy is presented in Section 3. Section 4 gives the performance analysis of the proposed rate shaping approach using H.264/SVC and shows the achievable improvements by our joint RD optimization. Section 5 concludes the paper.

2. SCALABLE VIDEO AND RD SIDE INFORMATION

The scalable extension described in [2] enhances the H.264/AVC coding scheme to support spatial, temporal and quality/SNR scalability for higher compression efficiency and flexibility. In H.264/SVC, the sender encodes a video into a *base layer* that corresponds to the minimally acceptable quality and one or more *enhancement layers* that improve the video quality if received together with the *base layer*. The coded video data is organized into *NAL (Network Abstraction Layer) units*, each of which corresponds to a packet containing an integer number of bytes. SVC can make a flexible combination of certain layers by discarding the corresponding NAL units in order to adapt to different bitrates, frame rates or spatial resolutions of the video content.

The scalability levels of H.264/SVC used in this work are illustrated in Fig. 2, which presents the RD performance for the 'Foreman' sequence. The bitstream provides two spatial resolutions (176x 144, 352x288) and five different temporal resolutions with frame rates of 1.875, 3.75, 7.5, 15, and 30 Hz. As an effective GOP size of 16 pictures is used for both spatial layers, the lowest supported temporal resolution corresponds to the collection of all key pictures. The discrete operational RD points in Fig. 2 are connected by lines for better readability. Additionally, we transmit one *PR (Progressive Refinement) slice* for every picture to provide SNR scalability. The PR slices refine the transform coefficients transmitted in the base and enhancement layers and thus increase the reconstruction quality. Therefore, the generated bitstream consists of 20 different representations (sub-bitstreams), where the PSNR values of the decoded sub-

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under grant STE 1093/3-1

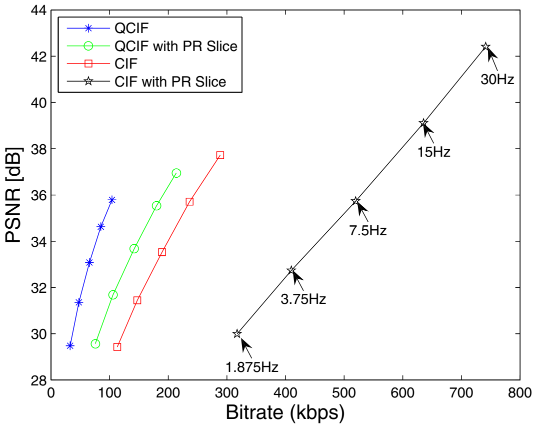

Fig. 2. Operational rate-distortion points for scalable coding of the 'Foreman' test sequence with H.264/SVC.

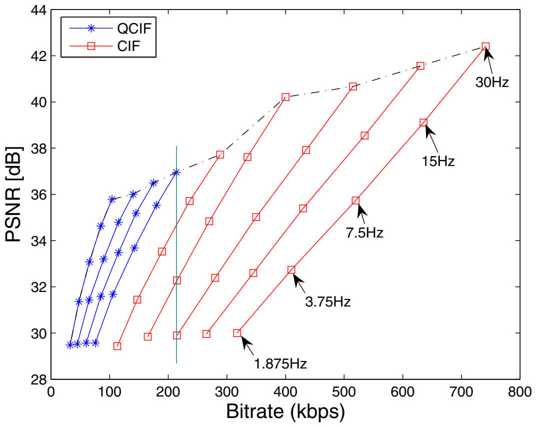

Fig. 3. Extracting additional RD points by dropping PR packets.

bitstreams are measured after reconstructing them to the same temporal and spatial resolution as the original sequence (CIF@30Hz).

2.1. RD Side Information

We can easily generate additional points apart from the operational RD points shown in Fig. 2 by selectively dropping PR slice packets. The coding symbols inside the PR slices are ordered by their importance and thus PR slices can be truncated at any point. A complete (not truncated) PR slice corresponds to a quality improvement that can be obtained by decreasing the QP (Quantization Parameter) with a value of 6 in the H.264 JSVM codec [7]. Fig. 3 gives some achievable rate examples obtained by truncating PR packets for the 'Foreman' sequence. The PSNR values of the decoded substreams are measured similarly as in Fig. 2 by upsampling the scaled video back to CIF@30Hz. As can be seen from Fig. 3, different scalability candidates can be selected for a particular rate. For example, we can have different RD points at a rate of about 210 kbps (shown with the straight line) with significant differences in PSNR: 1) CIF resolution at 1.875 fps encoded with some of the PR slice packets. 2) CIF resolution at 3.75 fps with even fewer PR packets. 3) QCIF at 30 fps encoded with all PR packets. Our RD side information consists of the maximum PSNR points at each rate. In Fig. 3, these points are connected by the dashdotted envelop curve. Please note, for different video sequences the order of the scalability points on the envelop might change. For instance, for 'Akiyo' spatial resolution is more important than temporal resolution.

The RD side information can be used by the network nodes to dynamically decide in a RD-optimized way which packets (NAL units) of which layers of the respective video streams should be dropped in case of network overload. Given all the rate-psnr points of every sequence, the node selects the optimal scalability modes for all incoming video sequences to meet the available transmission resource.

2.2. Subjective Tests

Although PSNR is quite often used to present the video quality, it is also well accepted that perceptual video quality does not always correlate well with PSNR. Human visual characteristics must be considered to provide a more accurate quality measurement. The subjective test described below was conducted to verify the correctness and effectiveness of the RD side information used in our RD optimization, which is based on PSNR.

Table 1. *Foreman* encoded at 140kbps

Test Video	Specification	PSNR [dB]
1	QCIF @ 7.5 Hz with PR slice	33.64
2	QCIF @ 30 Hz	36.00
3	CIF @ 3.75 Hz	31.45

Table 2. *Football* encoded at 400kbps

Test Video	Specification	PSNR [dB]
1	QCIF @ 15 Hz with PR slice	32.83
2	QCIF @ 30 Hz	36.68
3	CIF @ 7.5 Hz	29.52

Table 3. Subjective test results

Sequence	Video 1	Video 2	Video 3
Foreman	2.27	3.65	1.62
Football	3.05	3.67	2.05

Test videos from two different video sequences given in Table 1 and Table 2 were presented to 30 test persons, who were asked to give feedback based on their perceptual feeling of the video quality. The results are on a five pointed scale with five indicating the best quality. The test videos differ in their scalability levels but they are all encoded at the same bitrate. Table 3 presents the survey results, which are averaged over the scores given by all 30 test persons. It can be seen that the results correlate quite well with the PSNR values in our experiments, which justifies that the envelope curve in Fig. 3 can also give the best perceptual quality.

3. RD-OPTIMIZED RATE SHAPING

When the total data rate of all the incoming video streams exceeds the outgoing transmission rate, video packets have to be dropped. We use the RD side information sent along with the scalable video bitstreams to forward meaningful data which achieves highest user satisfaction. Algorithm 1 describes the RD-optimized rate shaping process for every GOP. K is the set of all active users. P_k defines the total number of RD points of user k and p_k is the index of the next available RD point for user k which specifies the achievable quality and needed rate. R_{in} is the total rate assigned to all incoming video streams and should be kept no larger than the given outlink rate R_{out}. The assignment of outlink capacity for a particular user depends on its utility U, defined in PSNR/kbps. It calculates the improvement in PSNR for a particular rate increment from the current RD point to the next RD point. We pick the user with the maximum U and check whether the required rate can be fulfilled. If yes, we assign the resource with the respective scaling pattern to this user and move p of this user to the next RD point on the side information curve. If not, we check the user with the second highest U and see if the required rate can be allocated, then the third and so on. This process is repeated until all the users are checked and all the available resource has been optimally assigned to the video streams. If one user has reached the last RD point, no further quality can be improved and its U equals to zero. We, therefore, exclude this user from the following optimization steps to reduce the complexity of the optimization process. Please note that the key pictures from the base layer of any video stream in most cases have a much higher utility than all other parts of the video.Although the algorithm starts with $R_{in} = 0$, it will not happen that these key pictures of any user are dropped unless the outlink capacity is too small (less than 200 kbps) to hold all of them.

Algorithm 1: RD-optimized Rate Shaping Algorithm

```
begin
    forall $k \in K$ do
        $p_k \longleftarrow 1$;
    $R_{in} \longleftarrow 0$;
    $K' \longleftarrow K$;
    while $R_{in} < R_{out}$ do
        forall $k \in K'$ do
            Calculate $U_k^{p_k} = \Delta D_k^{p_k} / \Delta Rate_k^{p_k}$;
        Choose user $k$ with maximum $U_k^{p_k}$, $(k \in K')$;
        if $R_{in} + \Delta Rate_k^{p_k} \leq R_{out}$ then
            $R_{in} \longleftarrow R_{in} + \Delta Rate_k^{p_k}$;
            $p_k \longleftarrow p_k + 1$;
        else
            repeat
                Choose user $k$ with the next maximum $U_k^{p_k}$;
                if $R_{in} + \Delta Rate_k^{p_k} \leq R_{out}$ then
                    $R_{in} \longleftarrow R_{in} + \Delta Rate_k^{p_k}$;
                    $p_k \longleftarrow p_k + 1$;
            until all candidates have been tried ;
            Break;
        if $p_k == P_k$ then
            $K' \longleftarrow K' - k$;
end
```

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed RD-optimized dynamic rate shaping approach, we compare it to a static priority-based rate shaping strategy for both SVC and H.264/AVC SLC. Both SVC and SLC bitstreams are generated using the JSVM software [7]. In our simulation setup, five different video sequences in CIF resolution are employed, whose characteristics are summarized in Table 4. All the video sequences are encoded at 30 fps and have a GOP size of 16 frames. The results are calculated by averaging the reconstruction quality of all the five video sequences. The simulation runs for 10 seconds and the Football sequence lasts only for the first 3 seconds to simulate a dynamic streaming activity. This means after 3 seconds, one user leaves and only the remaining 4 sequences compete for the outlink capacity.

Table 4. Characteristics of the video sequences

Sequence	SVC		SLC	
	Bitrate Range (kbps)	PSNR Range (dB)	Bitrate (kbps)	PSNR (dB)
Akiyo	9.40-213.31	36.33-45.41	56.96	42.97
Foreman	32.60-741.30	24.49-42.21	270.23	40.13
Coastguard	45.44-1450.91	33.80-43.06	503.38	40.64
Football	49.44-1559.81	25.30-41.35	679.96	38.92
News	22.00-461.59	31.64-43.93	150.33	41.24

As presented in Section 3, the decision which packets to drop is made independently for every GOP. For this, in general individual RD side information for every GOP is required to perform the rate shaping. The downside of this is the additional overhead encountered for sending the RD side information along with every GOP. As an alternative, we can use average RD side information for parts or even the entire video. In our experiments, we have observed a maximum performance degradation of 1 dB when working with average RD side information. Therefore, in the following experiments, the dropping decisions are made individually for all the GOPs using the average RD side information for the entire sequence.

4.1. Rate Shaping for SVC and SLC

We compare the performance of our proposed RD-optimized rate shaping approach with a scheme that uses static priorities for both SVC and SLC videos, whose working principle follows the fixed priority of NAL units. The base layer is the most important layer as all higher layers depend on it for decoding. The respective PR slice has the next level of importance. The enhancement layer comes next in the priority order followed by their PR segments. For static rate adaptation, PR segments of the enhancement layer (lowest priority) are dropped first. The spatial enhancement layer is dropped next in the same spirit. Eventually, the PR slice of the base layer is dropped next followed by temporally dropping the frames in the base layer if congestion persists. Unlike SVC, SLC offers only temporal scalability for rate adaptation.

Fig. 4 illustrates the achievable improvements in average reconstruction quality by the proposed approach. *SVC-optimized* is our proposed scheme using average RD side information. It performs the best among all the schemes at all tested rates. *SLC-optimized* uses RD side information with only temporal scalability and *SLC-unoptimized* uses static priorities. Both have a very close performance to the *SVC-optimized* curve at high transmission rate as very few packets or layers have to be dropped. For low outlink capacity, they perform significantly worse. *SVC-unoptimized* refers to the case with static priority-based dropping. The two SLC approaches outperform the *SVC-unoptimized* method for the middle range of the rates because the *SVC-unoptimized* approach drops all the PR slices of the enhancement layer of all users even when some of them can still be forwarded.

Rate shaping is required especially in cases where the channel is unable to maintain a constant transmission rate. Fig. 5 presents the reconstruction results for a time varying channel. The channel is modeled to vary randomly within 20% around the mean channel rate (X-axis). Rather than terminating the transmission, the network node can adjust the data rate according to the limited channel resource, yet producing gracefully degradation with an acceptable video quality.

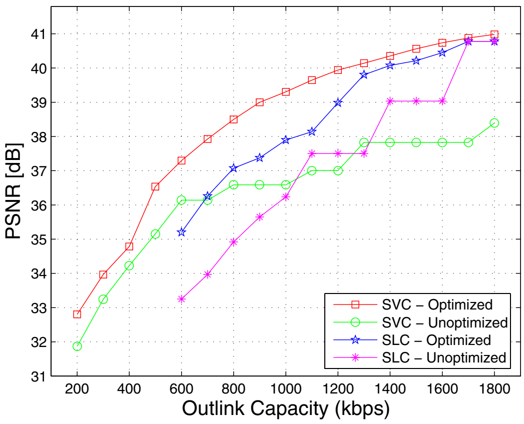

Fig. 4. Performance evaluation in a static channel

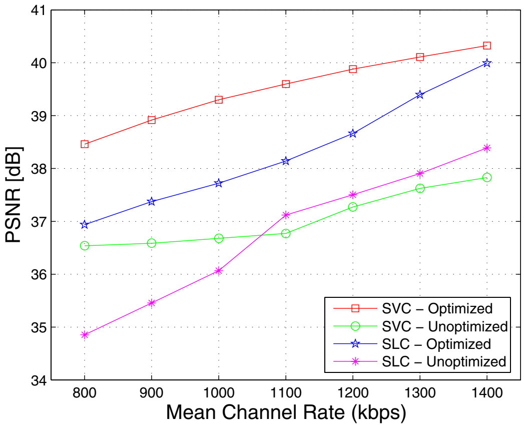

Fig. 5. Performance evaluation in a dynamic channel

4.2. Fair RD-optimization

For a given outgoing rate, the RD approach searches for the set of rate-distortion points which maximizes the sum of the PSNR values for all users. However, it might lead to unfairness among different video sequences because of their different characteristics. Furthermore, different rate shaping decisions for every GOP might also result in an annoying time-varying quality fluctuation. To overcome this problem, we maximize the overall PSNR based on the precondition that a minimum reasonable quality should be achieved for all the video streams. In order to achieve that required quality, we determine the appropriate operational RD point and assign corresponding resources to all users. If the system can not provide this needed resource, the quality has to be reduced. On the other hand, if there is still some transmission rate left, we assign that rate to users in the similar way as described in Algorithm 1. Fig. 6 shows the result of the approach. *Required Quality [85%]* means for every user, 85% (in PSNR) of the original quality (PSNR of all layers) should be guaranteed. The solid line presents results which fulfill the basic required quality for all the users. As can be seen, the fair approach has some slight performance loss compared with the *Max PSNR* principle, in return some basic QoS can be achieved. The higher the percentage of achievable quality, the higher the requirement in fairness and as expected the bigger the gap from the *Max PSNR*. Please note that, different minimum quality constraints for resolution and frame rate can easily be added to this framework.

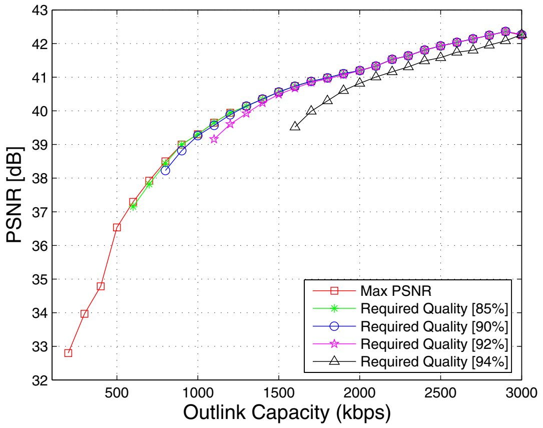

Fig. 6. Fair RD-optimization

5. CONCLUSION

This paper presents a RD-optimized rate shaping approach for scalable video streams. RD side information extracted during the encoding is used as input to the joint optimization over all incoming video sequences. Our subjective test results verify the correctness of the side information and the dropping strategy. The simulation results indicate that our proposed approach improves the performance of streaming video applications over static and dynamic channels and is more robust to bandwidth fluctuations in comparison to the static priority controlled dropping of video data. The advantages of deploying such an adaptive rate scalable framework are that it can achieve suitable and fair QoS for video over wired and wireless networks, as well as an efficient and fair sharing of the transmission resource.

6. REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. on CSTV*, vol. 13, no. 7, pp. 560–576, July 2003.

[2] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable H.264/MPEG4-AVC extension,” in *ICIP 2006*, Atlanta, GA, Oct. 2006.

[3] Philip A. Chou and Z. Miao, “Rate-distortion optimized streaming of packetized media,” Tech. Rep. MSR-TR-2001-35, Microsoft Research, CA, 2001.

[4] J. Chakareski, J. Apostolopoulos, and B. Girod, “Low-complexity rate-distortion optimized video streaming,” in *ICIP 2004*, Singapore, Oct. 2004.

[5] W. Tu, W. Kellerer, and E. Steinbach, “Rate-distortion optimized video frame dropping on active network nodes,” in *Packet Video Workshop 2004*, Irvine, California, Dec. 2004.

[6] J. Chakareski and P. Frossard, “Rate-distortion optimized packet scheduling over bottleneck links,” in *ICME 2005*, Amsterdam, Netherlands, July 2005.

[7] HHI, “JSVM reference software,” *http : //ip.hhi.de/imagecom_G1/savce/index.htm.*