

Low-Complexity Error-Control Methods for Scalable Video Streaming

Zhijie Zhao and Jörn Ostermann

(Institut für Informationsverarbeitung Leibniz Universität, Hannover D-30167, Germany)

Abstract

In this paper, low-complexity error-resilience and error-concealment methods for the scalable video coding (SVC) extension of H.264/AVC are described. At the encoder, multiple-description coding (MDC) is used as error-resilient coding. Balanced scalable multiple descriptions are generated by mixing the pre-encoded scalable bit streams. Each description is wholly decodable using a standard SVC decoder. A preprocessor can be placed before an SVC decoder to extract the packets from the highest-quality bit stream. At the decoder, error concealment involves using a lightweight decoder preprocessor to generate a valid bit stream from the available network abstraction layer (NAL) units when medium-grain scalability (MGS) layers are used. Modifications are made to the NAL unit header or slice header if some NAL units of MGS layers are lost. The number of additional packets that a decoder discards as a result of a packet loss is minimized. The proposed error-resilience and error-concealment methods require little computation, which makes them suitable for real-time video streaming. Experiment results show that the proposed methods significantly reduce quality degradation caused by packet loss.

Keywords

error resilience; error concealment; SVC; MDC

1 Introduction

Real-time video streaming over packet-switched networks can be impeded by packet loss, which often produces undesirable effects at the decoder. Usually, packet loss can make part of a frame, a whole frame, or even several frames undecodable using a standard decoder. Therefore, error-resilient coding and error-concealment techniques are widely used in video streaming systems to reduce the effect of transmission errors and to minimize end-to-end distortion in error-prone environments.

Error-resilient coding in the encoder produces redundancy, and this limits packet loss. Error-resilient coding tools for scalable video coding (SVC) can be classified as standard or non-standard. SVC supports several standard error-resilient coding tools, including intra-MB/picture refresh, slice coding, parameter sets, flexible MB order, and redundant slices/pictures [1]. Loss-aware rate-distortion-optimized mode decision [1], forward-error correction, and multiple-description coding (MDC) are non-standard error-resilient coding tools. MDC is used to

code a video sequence into two or more bit streams, called descriptions, and these descriptions are transmitted using independent paths. Each description can be decoded independently so that the reproduction of the original source reaches a basic level of quality. A high level of quality can be achieved when all descriptions are reconstructed together. Some early forays into MDC include multiple-description scalar quantizer [2], MDC with pairwise correlating transform [3], multiple-description compensation schemes [4], multiple-state video coding [5], and multiple-description coding based on forward-error correction [6].

Alternatively, error concealment can be used in the decoder to passively reduce transmission errors. In this way, available and correctly decoded information is used without modifying source- and channel-coding schemes. The tools and structure of a special codec are also used to reduce video quality degradation. The temporal and spatial correlation between frames or within a frame is frequently used to conceal the artifacts caused by transmission errors. Motion data is one of the most important types of data for decoding a frame in hybrid video codecs, so motion copy and motion prediction are widely used.

Special Topic

Low-Complexity Error-Control Methods for Scalable Video Streaming

Zhijie Zhao and Jörn Ostermann

With SVC, a video sequence is coded into one or more layers, and data-rate adaptation is allowed. This is an attractive solution for dealing with heterogeneous networks and different terminal capacities. The scalable extension of H.264/AVC is the latest SVC standard [7]. SVC provides temporal, spatial, and quality scalability that can be combined for greater adaptability to different network conditions or terminal capacities. In this paper, we describe low-complexity error-control methods for SVC. In particular, we propose a flexible, standard-compatible MDC method for error-resilient coding [8] and a low-complexity error-concealment method for SVC at the decoder when an MGS layer is used [9]. In section 2, we give an overview of work related to SVC and introduce our proposed methods. In section 3, we present simulation results. Section 4 concludes the paper.

2 Scalable Video Coding and Our Proposed Methods

2.1 Overview of Scalable Video Coding

The SVC extension of H.264/AVC incorporates the key features of H.264/AVC as well as new techniques to improve scalability and coding efficiency. Temporal scalability can be achieved by using hierarchical prediction structures, for example, hierarchical B-pictures or non-dyadic hierarchical prediction structures. Scalable quality can be achieved by using coarse-grain quality-scalable coding (CGS) and medium-grain quality-scalable coding (MGS). Spatial scalability can be achieved by using multilayer coding, and each layer corresponds to a supported spatial resolution. Redundancy between spatial layers can be further exploited by interlayer prediction mechanisms such as interlayer motion prediction, interlayer residual prediction, and interlayer intraprediction.

H.264/AVC has a video coding layer (VCL) and a network abstraction layer (NAL). In the VCL, a coded representation of the input video signal is generated, and in the NAL, this coded representation is fragmented. The NAL provides header information to ease the use of VCL data. Being an extension of H.264/AVC, SVC has the H.264/AVC

MGS supports packet-based quality scalability by distributing the transform coefficients of a slice. When MGS is used to provide quality scalability, an access unit can include several MGS NAL units. With MGS, the enhancement layer transform coefficients can be distributed between a maximum of 16 slices, and each slice corresponds to an NAL unit. MGS layers inside each dependency layer are identified by a quality identifier.

2.2 Work Related to Scalable Video Coding

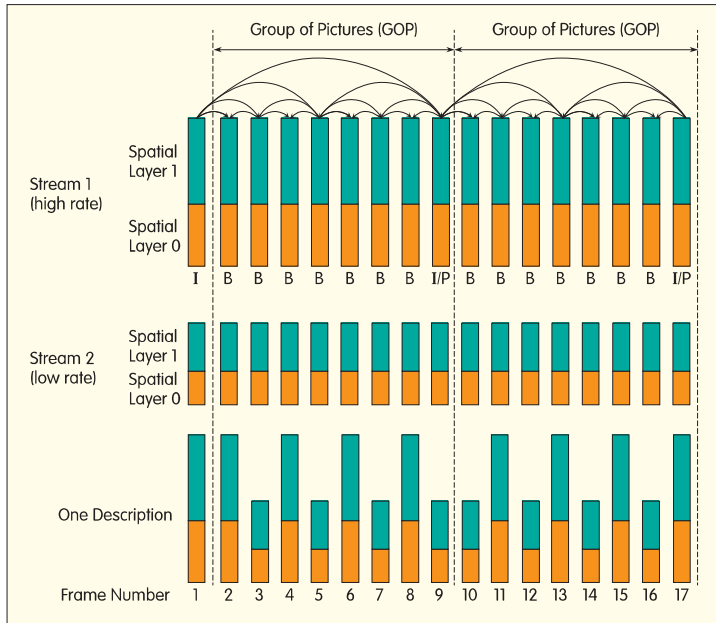
Combined SVC and MDC has attracted the interest of researchers. Multiple-state video coding [5] splits the input video into subsequences in the temporal domain and encodes each subsequence as an independent description. This coding can be used to generate scalable multiple

descriptions with half the temporal resolution. In [10], two complementary descriptions are generated for the high-pass frame of each enhancement layer. These descriptions are generated by assigning only half the motion vectors and texture information of the original coded stream in each description. Alternatively, two descriptions are generated for the base layer of SVC by downsampling the residual data [11]. A general multirate allocation scheme for multiple-description coding is proposed in [12]. This scheme has been used in JPEG 2000 and H.264/AVC to produce multiple descriptions [13]. A fully redundant unbalanced MDC scheme for SVC is proposed in [14]. In [15], an SVC bit stream produces two balanced descriptions by assigning MGS NAL units to one of the descriptions on alternating frames with a period of two group of pictures (GOPs). The spatial scalability of SVC is not taken into account.

Several error-concealment methods have been proposed for SVC [1], [16]. Intralayer error concealment and interlayer error concealment deal with frame loss when there are two spatial layers. Error-concealment algorithms copy a picture from another view, generate a motion vector from the same spatial layer, or upsample motion and residual data from the available base layer to generate a lost picture. With a slice support, the error-concealment scheme in [16] not only uses reference-frame information but also uses correctly received information from the same frame and from higher spatial layers. In this way, packet loss can be concealed on the base layer of the compressed stream [17]. In [18], a frame-loss error-concealment algorithm based on hallucination is proposed for the spatial-enhancement layer. A training database of hallucinations for missing enhancement frames is generated from the two most-recently decoded I or P frames. This algorithm performs better than the motion and residual upsampling proposed in [16]. Another error-concealment method for whole-picture loss in hierarchical B-picture coding is proposed [19]. This method performs better than the error-concealment method in SVC reference software. In contrast, a low-complexity error-concealment algorithm can be used in the network abstraction layer of SVC [20]. The algorithm in [20] recognizes the bit-stream structure and creates a valid sequence of packets from the received packets. In this paper, we call this method NAL unit removal. Unlike other error-concealment algorithms, NAL unit removal does not generate missing frames. In [20], the case of two spatial layers with fine-grain scalability (FGS) is considered. A similar work is [21], which builds on the work in [16] by supporting quality scalability using FGS layer [22].

2.3 Standard-Compatible MDC for SVC

In SVC, each quality base layer and quality-enhancement layer of a spatial layer is usually quantized using different step sizes. The coefficients of a quality refinement picture are quantized with a quantization parameter (QP) and can be distributed over several layers. Each of these layers contains partial refinement coefficients that use MGS. QPs can be cascaded over the temporal levels according to a given pattern, or default QP cascading can be used.



▲ Figure 1. Two descriptions are generated by combining bit streams.

The parameters for quantization step sizes are stored in several places, including the picture parameter set (PPS), slice header (SH), and macroblock layer. The lum quantization parameter is initially QP_v , and this value is used for all macroblocks in the slice until modified by mb_qp_delta in the macroblock layer. QP_v is given as

$$QP_v = 26 + pic_init_qp_minus\ 26 + slice_qp_delta \quad (1)$$

where $pic_init_qp_minus\ 26$ is the initial QP_v of -26 for each slice and is stored in the PPS, and $slice_qp_delta$ is a value that changes the quantizer step size at each slice and is stored in the SH. The slice header contains a codeword that indicates the PPS to be used, and the PPS includes the identifier of the active sequence parameter set (SPS). An active SPS remains unchanged throughout a coded video sequence, and an active PPS remains unchanged within a coded picture.

Two standard-compatible scalable descriptions can be produced by combing streams pre-encoded at different bit rates. To do this, we change only the quantization step size in order to generate low- and high-bit-rate streams. Moreover, in order to combine NAL units of different descriptions at the decoder, both descriptions use the same PPSs and SPSs. Different quantization step sizes are indicated by $slice_qp_delta$ parameter in the SH. Each description generated by combined pre-encoded streams supports spatial, temporal, and quality scalability and can be decoded by a standard SVC decoder.

To generate balanced multiple descriptions, different bit streams are combined so that high- and low-rate NAL units from alternating frames can be assigned to a description over a period of two GOPs. Fig. 1 shows the proposed combination

scheme for generating descriptions.

For the first description in Fig. 1, the even-numbered frames of the first GOP come from a high-bit-rate stream, and the odd-numbered frames come from a low-bit-rate stream. The even-numbered frames of the second GOP come from a low-bit-rate stream, and the odd-numbered frames come from a high-bit-rate stream. For the second description, the even-numbered frames of the first GOP come from a low-bit-rate stream, and the odd-numbered frames come from a high-bit-rate stream. The even-numbered frames of the second GOP come from a high-bit-rate stream, and the odd-numbered frames come from a low-bit-rate stream. This produces two descriptions that are balanced in terms of bit rate and quality.

At the decoder, a preprocessor is placed before a standard SVC decoder to parse newly arrived packets and extract the packets from the highest-quality bit stream. However, without the preprocessor, each description is still decodable using a SVC decoder. In our proposed scheme, a side decoder is not needed for an MDC description. The received packets from both descriptions are parsed and arranged into a new stream that is passed to an SVC decoder.

2.4 Low-Complexity Error Concealment for SVC

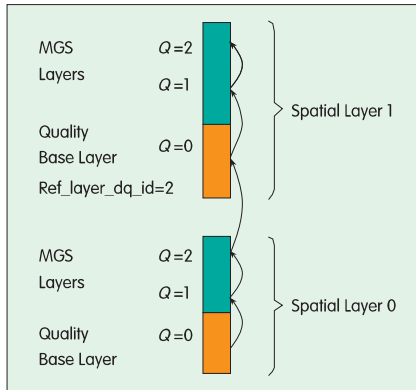
In H.264/AVC and SVC, an NAL unit starts with a single-byte header that signals the type of contained data, for example, NAL unit. In SVC, a three-byte extension header is used to indicate the scalable information for the coded slice in a scalable extension and for the prefix NAL unit. The parameter's dependency identifier (D), temporal identifier (T), and quality identifier (Q) determine which spatial layer, temporal layer, and quality layer an NAL unit belongs to. An access unit corresponds to one picture after decoding and comprises several consecutive NAL units with specific properties. In the SVC design, MGS is used and not FGS; however, no research has been done on MGS quality scalability. We therefore propose extending the NAL-unit-removal algorithm to deal with packet loss in the MGS layer. Our approach is motivated by multilayer adaptation for an MGS-based SVC bit stream [23]. NAL unit headers or slice headers are parsed to produce a valid bit stream from the available NAL units at the receiver. When a frame belonging to the highest temporal level is lost, the handling method of the NAL-unit-removal algorithm is changed.

When MGS is used to provide scalable quality, an access unit can include several MGS NAL units. With the MGS, the enhancement-layer transform coefficients can be distributed between a maximum of 16 slices, and each slice corresponds to an NAL unit. MGS layers inside each dependency layer are identified by a quality identifier. For the quality base layer of a spatial-enhancement layer, a syntax element called $ref_layer_dq_id$ in the slice header is used to signal which MGS layer is used for interlayer prediction (assuming that

Special Topic

Low-Complexity Error-Control Methods for Scalable Video Streaming

Zhijie Zhao and Jörn Ostermann



◀ Figure 2. A layer-prediction structure in an access unit.

interlayer prediction is enabled). For quality-refinement MGS layers with quality identifier $Q > 0$, the preceding quality layer with quality identifier $Q - 1$ is used for interlayer prediction. Fig. 2 shows a layer-prediction structure in an access unit with two spatial layers and two MGS layers. If an MGS layer is employed as reference layer for interlayer prediction and is lost, the received bit stream becomes invalid for a standard decoder. For example, the decoding of MGS layer $Q = 2$ in spatial layer 0 depends on the MGS layer $Q = 1$ in spatial layer 0. If MGS layer $Q = 1$ in spatial layer 0 is lost and the other NAL units are received, the bit stream cannot be decoded by a standard decoder. The packets of MGS layer $Q = 2$ in spatial layer 0 and the whole spatial layer 1 can be discarded [20]. In the following, we discuss how to deal with MGS-layer loss.

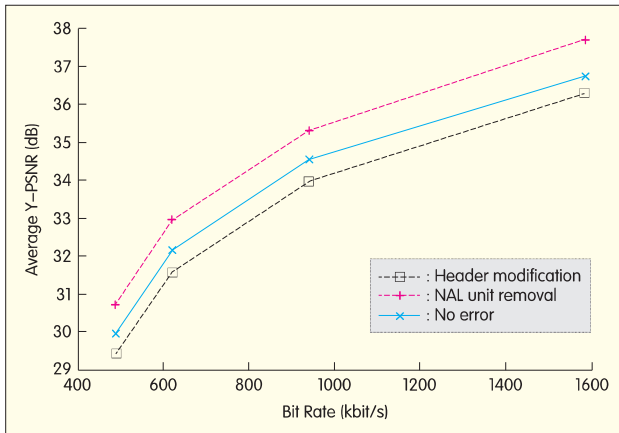
To simplify the description without losing generality, we consider the case where the source video is coded with two spatial layers and two MGS layers within each spatial layer. Table 1 shows the NAL unit order in a bit stream for a group of pictures (GOP) of size four that has three temporal levels. With the NAL unit-removal method, if a NAL unit of a GOP is lost, a valid NAL unit order with lower spatial resolution and/or lower frame rate is chosen. With multiple-quality-layer coding, if an NAL unit not from the highest MGS layer is lost, a valid NAL unit order with a lower-quality layer is chosen. For example, if the 11th NAL unit (MGS layer 1) of a GOP in Table 1 is lost, the 12th NAL unit, which belongs to dependant MGS layer 2, is also discarded to create a valid bit stream, even if the 12th NAL unit is correctly received. Because the slice data of the MGS quality-refinement layers include different distributions of transform coefficients, the 12th NAL unit can still be used to improve the decoded image quality. Therefore, we do not discard the higher MGS layers if one or several lower MGS layers are lost. At the client, we use the same layer-dependent modification described in [23], which is made for data-rate adaptation at the server. If the lost MGS layers belong to spatial layer 1, only the *quality_id* parameter of the NAL headers of the remaining MGS NAL units in spatial layer 1 need to be modified so that continuity of *quality_id* values is maintained. Because an NAL unit header is not compressed, the modification requires very low computing power.

When the 15th NAL unit in Table 1 is lost and the layer-prediction structure in Fig. 2 is used, NAL units 16 to 18 are discarded to create a valid bit stream. To use the received higher MGS layers of a spatial-enhancement layer and to maintain a standard decoder-compliant bit stream, both NAL unit header and slice header are modified. When an MGS NAL unit in spatial layer 0 is lost, the header of the NAL units within spatial layer 0, and the slice header of the quality base layer in spatial layer 1, need to be changed. If the maximum-quality identifier in spatial layer 0 is changed, *ref_layer_dq_id* in the slice header of the quality base layer in spatial layer 1 is updated according to maximum-quality identifier. The slice header is coded using Exp-Golomb codes in SVC, and parsing and modification is not time consuming. Fig. 3 shows the reconstructed video quality of the joint test sequence used in section 0. We discard the first MGS layer of the spatial enhancement layer. In our proposed method, the NAL header of the second MGS layer is modified in order to maintain a valid bit stream. In the NAL unit-removal method, only the quality base layer of the spatial-enhancement layer is kept in order to maintain a valid bit stream. The proposed method improves average luma PSNR by 0.57 dB. Our method may introduce drift, so a

▼ Table 1. NAL unit order in a bit stream for a GOP (four frames) with two spatial layers, three temporal layers and, two MGS layers

No.	D	T	Q	MGS
1	0	0	0	
2	0	0	1	X
3	0	0	2	X
4	1	0	0	
5	1	0	1	X
6	1	0	2	X
7	0	1	0	
8	0	1	1	X
9	0	1	2	X
10	1	1	0	
11	1	1	1	X
12	1	1	2	X
13	0	2	0	
14	0	2	1	X
15	0	2	2	X
16	1	2	0	
17	1	2	1	X
18	1	2	2	X
19	0	2	0	
20	0	2	1	X
21	0	2	2	X
22	1	2	0	
23	1	2	1	X
24	1	2	2	X

D: dependency identifier Q: quality identifier
MGS: medium-grain scalability T: temporal identifier



▲ Figure 3. Average distortion when the first MGS layer in spatial layer 1 is discarded.

two-alternative forced-choice test was performed to assess the subjective quality of our method and the NAL unit-removal method in case the first MGS layer of the spatial-enhancement layer is lost. Two short videos were shown sequentially, and observers had to choose the one they thought was higher quality. For low and medium qualities, the proposed method is preferred, but for higher qualities, the NAL unit-removal method is preferred because of its smoother motion rendition.

With the NAL unit-removal method, if a quality-base-layer NAL unit of the highest temporal layer is lost, an entire temporal layer is removed. For example, if the 13th NAL unit is lost, NAL units 14 to 24 are discarded to arrange a valid bit stream, even if these units are received. However, if hierarchical B picture is used for temporal scalability, the highest temporal layers are B pictures and are not used as reference frames. This means that if one frame of the highest temporal layer is lost, it does not affect the other frames in the temporal layers. In our proposal, the remaining highest temporal layers are retained. The missing frame can be concealed using frame copy or other error-concealment methods for whole-frame loss.

3 Experiment Results

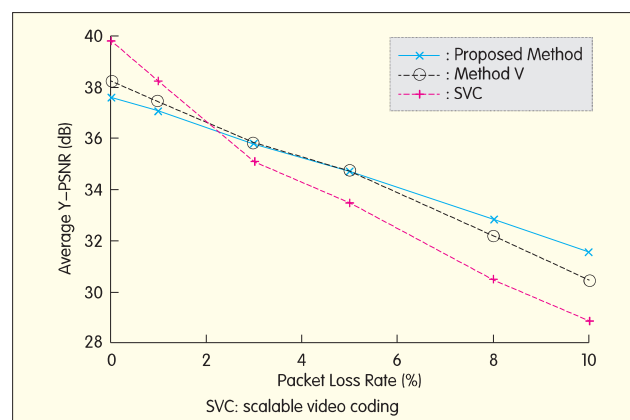
In this section, we present experiment results for the proposed error-resilient and error-concealment methods. JSVM 9.18 SVC reference software was used to encode the input sequences [21]. The tested bits streams had three video sequences—Foreman, Mobile, and Akiyo—combined into a single sequence to produce long test-bit streams. Spatial layer 0 had quarter common intermediate format (QCIF) resolution, and spatial layer 1 had common intermediate format (CIF) resolution. The joint sequence contained 897 frames; the GOP size was 8 frames; and an I frame was used as the key picture. The RTP packet size was limited to 1400 bytes, and packet loss in the transmission channel was simulated by a two-state Markov model—where a good state

means packets are received correctly and promptly, and a bad state means packets are lost.

3.1 Experiment Results for the Proposed MDC Method

In this experiment, we used a streaming scenario with path diversity. The two descriptions are delivered through independent paths. In case of packet loss, all these paths have the same packet-loss probability. Five packet-loss ratios were used: 1%, 3%, 5%, 8% and 10%. We assume that parameter sets are conveyed using a reliable transport mechanism. If spatial scalability is supported, a coded bit stream contains two spatial layers (QCIF and CIF), four temporal levels, and one quality layer. CIF resolution is used if spatial scalability is not considered. For simplicity, spatial base layers of the pre-encoded streams are quantized by the same QP, and only the spatial-enhancement layers or quality-enhancement layers are quantized using two different QPs. Where SVC cannot decode the base layer, or the SMDC receiver lacks both descriptions, one or several frames cannot be decoded. In this case, frame copy is used as error concealment.

First, we compare our proposed SMDC scheme with single-description SVC and method V proposed in [15]. Method V is based on SVC with both descriptions containing the active base layer and every other quality-enhancement layer. Therefore, in method V, the redundancy is only the base layer. In the case of no packet loss, the proposed scheme and method V pay a penalty of reduced coding efficiency compared with single-description SVC. Method V has slightly less coding-efficiency loss than the proposed method. Fig. 4 shows the average luma PSNR as a function of the network packet-loss rate for the joint sequence of Foreman, Mobile, and Akiyo. The proposed scheme outperforms single-description SVC when the packet-loss rate is greater than 3% and outperforms method V when the packet-loss rate is greater than 5% (Fig. 4). At 10% of the packet-loss rate, the gain over single-description SVC is 2.6 dB, and the gain over method V is about 1.1 dB. When the packet-loss rate is less than 2%, the proposed SMDC scheme is inferior to

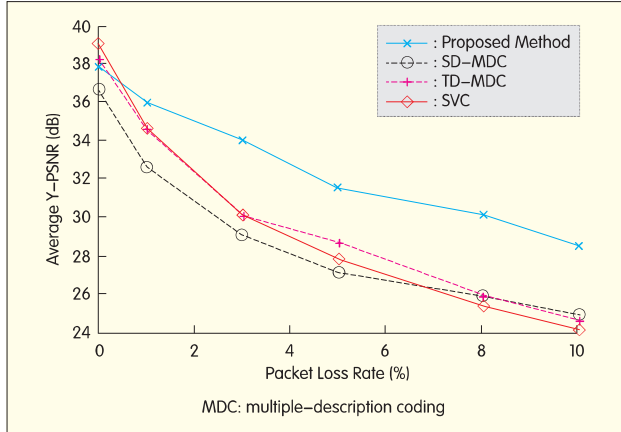


▲ Figure 4. Average Y-PSNR vs. packet loss rate, without spatial scalability.

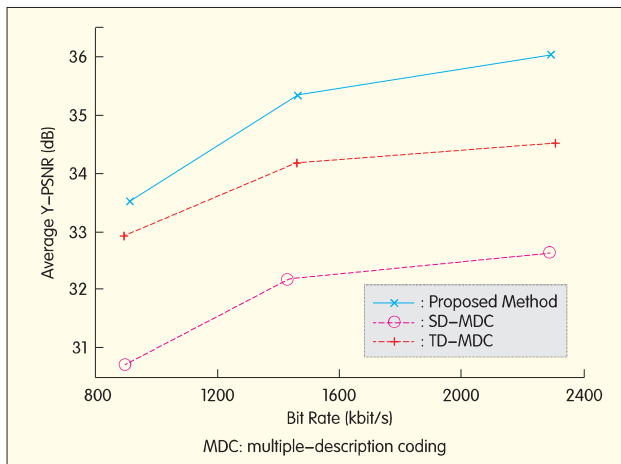
Special Topic

Low-Complexity Error-Control Methods for Scalable Video Streaming

Zhijie Zhao and Jörn Ostermann



▲ Figure 5. Average Y-PSNR vs. packet loss rate, with spatial scalability.



▲ Figure 6. Y-PSNR compared to TD-MDC and SD-MDC with 1% packet loss rate.

single-description SVC, and when the packet-loss rate is less than 3%, the proposed SMDC method is inferior to method V. The additional redundancy introduced in the proposed scheme plays a minor role at low packet-loss rates. However, method V in [15] cannot be extended to support spatial scalability.

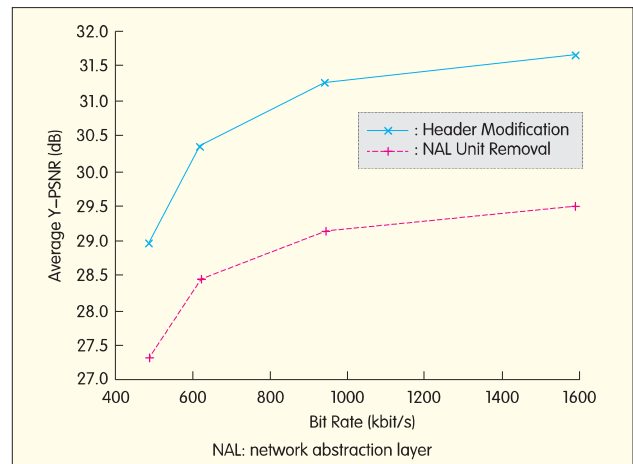
To test performance when supporting spatial scalability, we compare our proposed scheme with single-description SVC, spatial downsampling (SD-MDC), and temporal downsampling MDC (TD-MDC). In SD-MDC and TD-MDC, an original video is first downsampled into two subsequences in the spatial and temporal domain, respectively. Then, the two subsequences are independently encoded by an SVC encoder. Fig. 5 shows how our scheme performs compared with single-description SVC, SD-MDC, and TD-MDC in terms of average Y-PSNR() versus packet-loss rate. The proposed scheme performs best at a packet-loss loss rate of 1–10%. At a 10% loss rate, the proposed scheme outperforms the SD-MDC and TD-MDC by approximately

3.5 dB and 3.9 dB, respectively. Because the descriptions of SD-MDC and TD-MDC are separately encoded, the losses of packets from one description cannot be effectively compensated from the received packets of the other description. However, the proposed scheme can still produce a whole spatial and temporal resolution video when one description is corrupted. Hence, the redundancy introduced in the proposed scheme is more beneficial than SD-MDC and TD-MDC in the case of packet loss. Although single-description SVC has the highest coding efficiency, the proposed scheme has a similar gain over single-description SVC. The gain is 4.3 dB at a 10% packet-loss rate. Fig. 6 shows the average Y-PSNR versus bit rate compared with SD-MDC and TD-MDC at 1% packet-loss rate. The results show that the proposed method is superior to SD-MDC and TD-MDC over the encoding bit rates 912 kbit/s, 1460 kbit/s, and 2294 kbit/s, where the redundancies are 28%, 31%, and 33%, respectively.

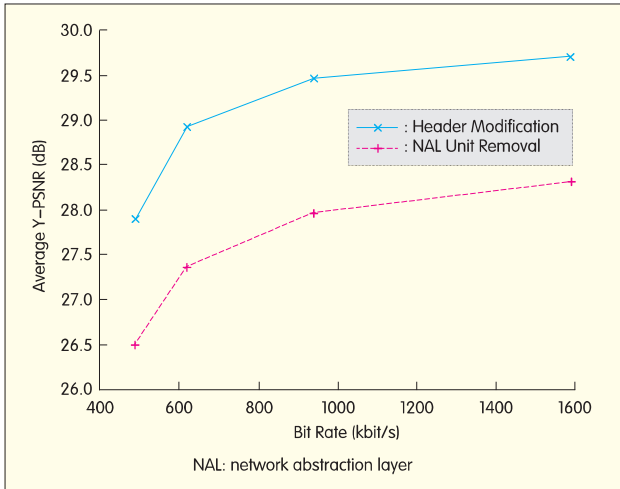
3.2 Experiment Results for the Proposed Error-Concealment Method

The proposed method is implemented as a preprocessing unit before a standard decoder in order to arrange a valid bit stream from the received packets. In these tests, each spatial layer has four temporal layers and two MGS layers. The first MGS layer contains four transform coefficients, and the second MGS layer includes 12 transform coefficients for each spatial layer. The QP difference between the quality base layer and the quality-enhancement layer is set to three. The default cascading of quantization parameters over the temporal levels is used. Hierarchical B picture is also used. In the experiments, we assume that packets of the quality base layer in spatial layer 0 are protected and not lost. Packet-loss ratios of 3%, 5%, and 10% are used. For decoded frames with a spatial resolution of QCIF, we use the upsampling filter in SVC to produce the spatial-resolution CIF.

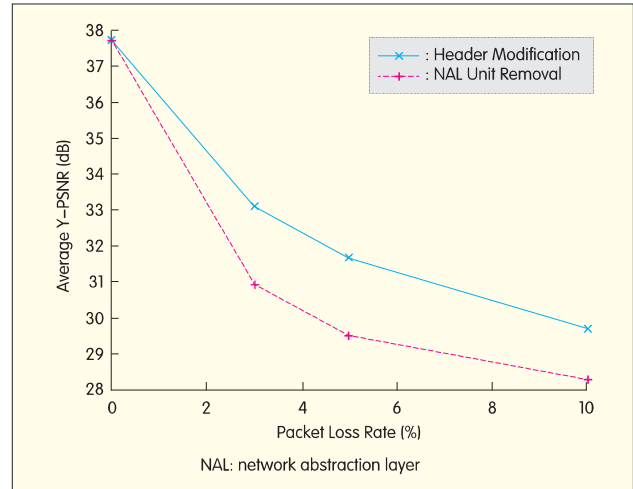
Fig. 7 shows the rate-distortion curves for the



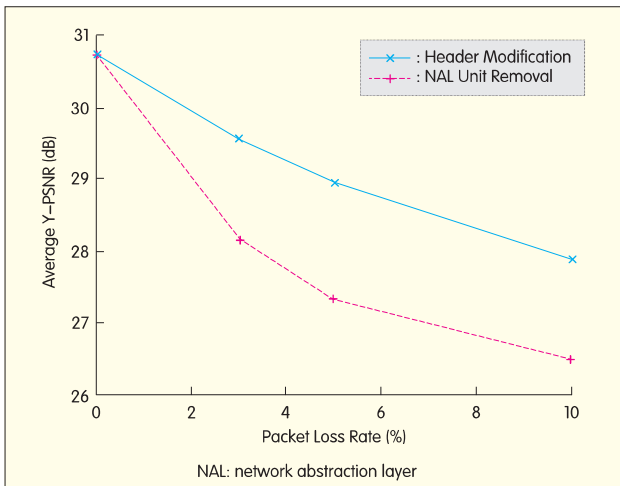
▲ Figure 7. Average distortion, as a function of encoding rate, with 5% packet loss rate (Joint sequence Foreman, Mobile and Akiyo).



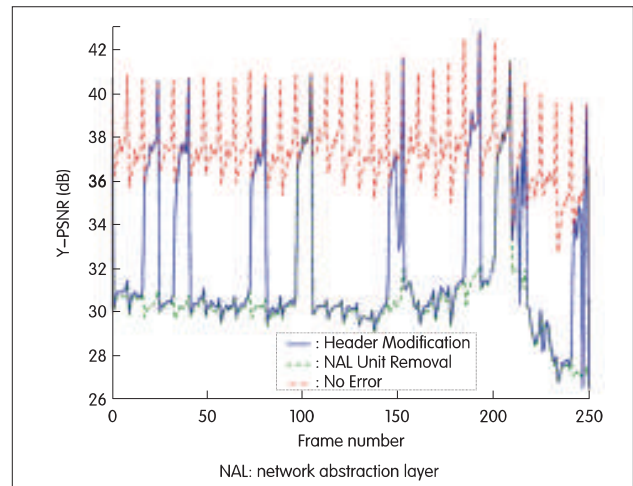
▲ Figure 8. Average distortion vs. encoding rate for 10% packet-loss rate (Joint sequence Foreman, Mobile and Akiyo).



▲ Figure 9. Average distortion vs. packet-loss rate (Joint sequence Foreman, Mobile and Akiyo, 1590 kbit/s).



▲ Figure 10. Average distortion vs. packet-loss rate (Joint sequence Foreman, Mobile and Akiyo, 488 kbit/s).



▲ Figure 11. Y-PSNR between NAL unit-removal and proposed method at 10% packet-loss rate (Foreman, from the joint sequence at 1590 kbit/s).

header-modification and NAL unit-removal methods with 5% packet-loss rate, and Fig. 8 shows the rate-distortion curves for the header-modification and NAL unit-removal methods with 10% packet-loss rate. We compare the proposed method with the NAL unit-removal method only because the proposed method does not substitute methods for concealing a frame loss but only complements them. Fig. 7 and Fig. 8 show that header-modification outperforms NAL unit-removal over the entire considered range of bit rates. For a 5% packet-loss rate, header modification gains 2.16 dB on average, and for a 10% packet-loss rate, NAL unit removal gains 1.55 dB on average for the joint sequence.

To further evaluate the performance of the proposed method, we determine how the average luma PSNR changes in relation to packet-loss rate. Figs. 9 and 10 show that the

proposed method outperforms the NAL unit-removal method at all the three simulated packet-loss rates. The proposed method gains a maximum of 1.64 dB over the NAL unit-removal method when the packet-loss rate for the 488 kbit/s stream is 5%, and this can be as high as 2.16 dB when the packet-loss rate is 5% for the 1590 kbit/s stream. At 10% packet-loss rate, the proposed method improves average luma PSNR by 1.4 dB over the NAL unit-removal method at both bit rates.

Fig. 11 shows luma PSNR against the number of frames. It shows how luma PSNR changes for a Foreman sequence taken from the joint sequence coded at 1590 kbit/s with 10% packet-loss rate. The proposed method still uses the correctly received packets of higher MGS layers in case a lower MGS layer is lost, so the proposed method provides

Special Topic

Low-Complexity Error-Control Methods for Scalable Video Streaming

Zhijie Zhao and Jörn Ostermann

much better video quality than NAL unit removal.

4 Conclusion

In this paper, we have proposed a standard-compatible MDC scheme for SVC based on combined pre-encoded streams. This scheme is designed for video streaming applications in error-prone environments. At the decoder, an error-concealment method in the NAL in case that MGS is used for the scalable extension of H.264/AVC is presented. Experiment results show that the proposed MDC and error-concealment methods can improve video quality in error-prone environments. The proposed methods have low computational complexity and require low computing power. Hence, they are suitable for real-time scalable video streaming.

References

- [1] Y. Guo, Y. Chen, Y. K. Wang and et al., "Error resilient coding and error concealment in scalable video coding," in *IEEE Trans. on Circ. Syst. for Video Tech.*, vol. 19, no. 6, pp. 781–795, 2009.
- [2] V.A. Vaishampayan, "Design of multiple description scalar quantizers," in *IEEE Trans. Info. Theory*, vol. 39, no. 3, pp. 821–834, 1993.
- [3] Y. Wang, M.T. Orchard, V.A. Vaishampayan, A.R. Reibman, "Multiple description coding using pairwise correlating transforms," in *IEEE Trans. Image Processing*, vol. 10, no. 3, pp. 351–366, 2001.
- [4] C.S. Kim, S.U. Lee, "Multiple description motion coding algorithm for robust video transmission," in *Proc. IEEE Int. Symp. Circ. Syst.*, Geneva, Switzerland, Mar. 2000, pp. 717–720.
- [5] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Visual Commun. and Image Processing (VCIP)*, vol. 4310, Jan. 2001, pp. 392–409.
- [6] R. Puri, K. Ramchandran, K.W. Lee, V. Bharghavan, "Forward error correction (FEC) codes based multiple description coding for internet video streaming and multicast," in *Signal Processing: Image Communication*, vol. 16, no. 8, pp. 745–762, May 2001.
- [7] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," in *IEEE Trans. Circ. Syst. for Video Tech.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [8] Z.J. Zhao, J. Ostermann, "Video Streaming Using Standard-Compatible Scalable Multiple Description Coding Based on SVC," in *Proc. 17th IEEE Int. Conf. on Image Processing*, Hong Kong, Sep. 2010, pp. 1293–1296.
- [9] Z.J. Zhao, J. Ostermann, "Error Concealment in the Network Abstraction Layer for Medium Grain Scalability of SVC," in *Visual Commun. and Image Processing 2010*, Proc. SPIE, vol. 7744, pp. 77442P–7744P–8, July 2010.
- [10] H. Mansour, P. Nasiopoulos, V. Leung, "A flexible multi-rate allocation scheme for balanced multiple description coding applications," in *Proc. IEEE Int. Symp. Signal Processing and Info. Tech.*, Athens, Greece, Nov. 2005, pp. 1–4.
- [11] Z.J. Zhao, J. Ostermann and H.X. Chen, "Low Complexity Multiple Description Coding for the Scalable Extension of H.264/AVC," in *Proc. Picture Coding Symposium (PCS'09)*, Chicago, IL, May 2009, pp. 261–264.
- [12] T. Tillo, E. Baccaglioni and G. Olmo, "A flexible multi-rate allocation scheme for balanced multiple description coding applications," in *Proc. 7th IEEE Workshop on Multimedia Signal Processing*, Nov. 2005.
- [13] T. Tillo, E. Baccaglioni and G. Olmo, "Multiple descriptions based on multirate coding for JPEG 2000 and H.264/AVC," in *IEEE Trans. on Image Processing*, vol. 19, no. 7, pp. 1756–1767, Jul. 2010.
- [14] P. Schelkens, A. Gavrilescu, A. Munteanu and et al., "Error-Resilient Transmission of H.264 SVC Streams over DVB-T/H and WIMAX Channels with Multiple Description Coding Techniques," in *Proc. 15th European Signal*

- Processing Conference*, Poznan, Poland, Sep. 2007, pp. 1995–1999.
- [15] T. Berkin Abanoz and A. Murat Tekalp, "SVC-based scalable multiple description video coding and optimization of encoding configuration," in *Signal Processing: Image Communication*, vol. 24, no. 9, pp. 691–701, Oct. 2009.
- [16] Y. Chen, K. Xia, F. Zhang and et al., "Frame loss error concealment for svc," in *J. Zhejiang University Science A*, vol. 7, no. 5, pp. 677–683, 2006.
- [17] T. Keränen, J. Vehkaperä, and J. Peltola, "Error concealment for svc utilizing spatial enhancement information," in *Proc. 4th Int. Mobile Multimedia Commun. Conf. (MobilMedia '08)*, Oulu, Finland, July 2008, article 10.
- [18] Q.R. Ma, F. Wu and M.T. Sun, "Error concealment for spatial scalable video coding using allucination," in *Proc. IEEE Int. Symp. Circ. Syst. (ISCAS'09)*, Valencia, Spain, pp. 129–132, 2009.
- [19] X.Y. Ji, D.B. Zhao and W. Gao, "Concealment of whole-picture loss in hierarchical b-picture scalable video coding," in *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 11–22, 2009.
- [20] D.T. Nguyen, M. Shaltev, M and J. Ostermann, "Error concealment in the network abstraction layer for the scalability extension of H.264/AVC," in *Proc. Int. Conf. Commun. Electronics (ICCE06)*, Beijing, pp. 274–278, 2006.
- [21] M. Stoufs, A. Munteanu, J. Cornelis and P. Schelkens, "Error concealment for the scalable extension of H.264/mpeg-4 avc," in *Proc. Picture Coding Symp. (PCS2007)*, Lisbon, Portugal, Nov. 2007.
- [22] J. Reichel, H. Schwarz and M. Wien, Joint scalable video model 11 (jsvm 11)," Joint Video Team, doc.JVT-X202, July 2007.
- [23] T. C. Thang, J. W. Kang, J. J. Yong and J. G. Lee, "Multilayer adaptation for MGS-based SVC bitstream," in *Proc. 16th ACM Int. Conf. on Multimedia*, Vancouver, BC, pp. 689–692, 2008

Manuscript received: January 17, 2012

Biographies

Zhijie Zhao(zhao@tnt.uni-hannover.de) received his MSc degree in communications and information systems from Jilin University. He is currently working towards his PhD degree at the Institut für Informationsverarbeitung, Leibniz University, Hannover, Germany. His research interests include video streaming and video coding.

Jörn Ostermann has studied electrical engineering and communications engineering at the University of Hannover and Imperial College London. He received his Dipl.-Ing. and Dr.-Ing. degrees from the University of Hannover in 1988 and 1994. From 1988 to 1994, he was also a research assistant at the Institut für Theoretische Nachrichtentechnik and conducted research on low bit-rate, object-based analysis-synthesis video coding. From 1993 to 1994, he chaired the European COST 211 sim group coordinating research in low-bitrate video coding. From 1994 to 1995 he worked on video coding in the Visual Communications Research Department at AT&T Bell Labs. From 1996 to 2003, he was a member of Image Processing and Technology Research Team within AT&T Labs-Research. In 1998, he received the AT&T Standards Recognition Award and the ISO award. Since 2003, he has been a full professor and head of the Institut für Informationsverarbeitung at Leibniz Universität, Hannover, Germany. In 2007, he became head of the Laboratory for Information Technology at the same university. Since 2008, he has been the chairperson of the MPEG Requirements Group (ISO/IEC JTC1 SC29 WG11). Jörn was a scholar of the German National Foundation.

Dr. Ostermann has organized the evaluation of video tools to start defining the MPEG-4 standard. He chaired the Adhoc Group on the coding of arbitrarily-shaped objects in MPEG-4 video. He is a fellow of the IEEE and member of the IEEE Technical Committee on Multimedia Signal Processing. He has been the chair of the IEEE CAS Visual Signal Processing and Communications (VSPC) Technical Committee and has also been a Distinguished Lecturer of the IEEE CAS Society. He has published more than 100 research papers and book chapters. He is coauthor of a graduate-level text book on video communications and holds more than 30 patents.



Back Cover:
ZTE Corporation

