Error Concealment in the Network Abstraction Layer for the Scalability Extension of H.264/AVC

Dieu Thanh Nguyen, Miroslav Shaltev and Joern Ostermann
Institut fuer Informationsverarbeitung
Leibniz Universitaet Hannover
Appelstr 9a, 30167 Hannover, Germany
nguyen@tnt.uni-hannover.de

Abstract— **This paper presents an error concealment method applied to the Network Abstraction Layer (NAL) for the scalability extension of H.264/AVC. The method detects loss of NAL units for each group of picture (GOP) and arranges a valid set of NAL units from the available NAL units. In case that there is more than one possibility to arrange a valid set of NAL units, this method uses the information about motion vectors of the preceding pictures to decide if the erroneous GOP will be shown with higher frame rate or higher spatial resolution. This method works without parsing of the NAL unit payload or using of estimation and interpolation to create the lost pictures. Therefore it requires very low computing time and power. Our error concealment method works under the condition that the NAL units of the key pictures, which is the prediction reference picture for other pictures in a GOP, are not lost. The proposed method is the first method suitable for real-time video streaming providing drift-free error concealment at low computational cost.**

Error concealment; scalable video coding;network abstraction layer; video streaming

I. INTRODUCTION

Exchanging video over the Internet with devices differing in screen size, computational power and widely different as well as varying available bandwidth creates a logistic nightmare for each services provider when using conventional video codecs like MPEG-2 or H.264. Scalable video coding is not only a convenient solution to adapt the data rate to varying bandwidth in the Internet but also provide different end devices with appropriate video resolution and data rate. In January 2005, the ISO/IEC Moving Pictures Experts Group (MPEG) and the Video Coding Experts Group (VCEG) of the ITU-T started jointly MPEG's Scalable Video Coding (SVC) project as an Amendment of the H.264/AVC standard. The scalable extension of H.264/AVC was selected as the first Working Draft [1][2]. Furthermore, the Audio/Video Transport (AVT) Working Group of the Internet Engineering Task Force (IETF) started in November 2005 to draft the RTP payload format for the scalable extension of H.264/AVC and the signaling for layered coding structures [3].

The scalable extension of H.264/AVC uses the structure of H.264/AVC that is divided into two parts, so-called the Video Coding Layer (VCL) and the Network Abstraction Layer (NAL) [4]. In the VCL, the input video signal is coded. In the NAL, the output signal of the VCL is fragmented into so-called NAL units. Each NAL unit includes a header and a payload, which can contain a frame, a slice or a partition of a slice. The advantage of this structure is that the slice type or the priority of this NAL unit can be obtained only by parsing of the 8-bit NAL unit header. The NAL is designed based on principle called Application Level Framing where the application defines the fragmentation into meaningful subsets of data such that a receiver can cope with packet loss in a simple manner. It is very important for data transmission over network.

In multimedia communication, transmission errors such as packet loss or bit errors in storage medium causes erroneous bit streams. Therefore, it is necessary to add error control and concealment methods in the decoder. For the scalable extension of H.264/AVC a NAL unit is marked as lost and discarded if the bit error is not remedied by an error correction method. The error concealment methods defined in SVC project attempt to generate missing pictures in the Video Coding Layer by picture copy, up-sampling of motion and residual information from the base layer pictures or motion vector generation [5]. With these methods the decoder can give the output video with maximal available frame rate and resolution. But there will be error drift if the error-concealed picture is used further as a reference picture for other pictures because the error-concealed picture differs from the same reconstructed picture without error. The amount of error drift depends on which spatial layer and temporal level the lost NAL unit belongs to.

In this paper, we present an error concealment method in the Network Abstraction Layer for the scalable extension of H.264/AVC. With the knowledge of the bit stream structure, a simple algorithm will be applied to create a valid bit stream from the erroneous bit stream. The output video will not achieve the maximal resolution or maximal frame rate of the non-erroneous bit stream, but there will be no error drift. This is the first error concealment method for the scalable extension of H.264/AVC that does not require parsing of the NAL unit payload or high computing power. Therefore, it is suitable for real-time video communication.

The rest of the paper is organized as follows. In Section 2 the main scalable techniques in the scalable extension of H.264/AVC and its bit stream structure in the NAL are presented. Furthermore, the effects of error drift by using error concealment methods in the VCL are illustrated. Section 3 describes our proposed error concealment method in the NAL.

1-4244-0569-6/06/$20.00 ©2006 IEEE

Section 3 provides experimental results of our method. Section 4 concludes the paper.

II. SCALABLE EXTENSION OF H.264/AVC AND ITS BIT STREAM STRUCTURE

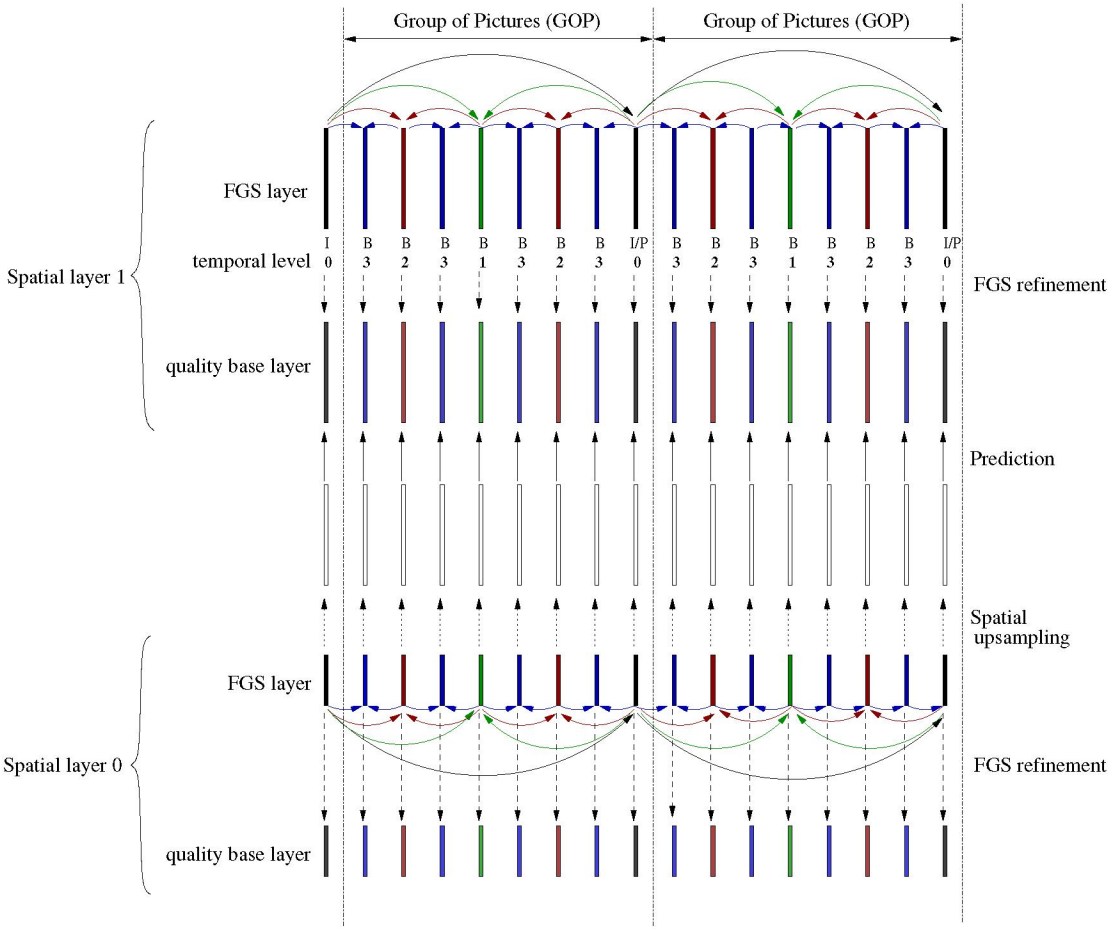
A. *Scalable extension of H.264/AVC*

The scalable video coder employs different techniques to enable spatial, temporal and quality scalability [1][2]. Spatial scalability is achieved by using a down-sampling filter that generates the lower resolution signal for each spatial layer. Either motion compensated temporal filtering (MCTF) or hierarchical B-pictures obtain a temporal decomposition in each spatial layer that enables temporal scalability. Both methods process input pictures at the encoder and the bit stream at the decoder in group of pictures (GOP) mode. A GOP includes at least one key picture and all other pictures between this key picture and the previous key picture, whereas a key picture is intra-coded or inter-coded by using motion compensated prediction from previous key pictures. Figure 1 shows how to generate a scalable video bit stream with 2 spatial layers, 4 temporal levels, a quality base layer and a quality enhancement layer. The input pictures in layer 0 are created by down-sampling of the input pictures in layer 1 by a factor of two. In each spatial layer a group of pictures (GOP) is coded with hierachical B-Picture techniques to obtain 4 temporal levels (i=0,1,2,3). The key picture is coded as I- or P-picture and has temporal level 0. The direction of arrow points from the reference piture to the predicted picture. To remove redundancy within layers, motion and texture information of the temporal level in the lower spatial layer are scaled and up-sampled for prediction of motion and texture information in the current layer.

For each temporal level, the residual signal resulting from texture prediction is transformed. For quality scalability, the transform coefficients are coded by using a progressive spatial refinement mode to create a quality base layer and several quality enhancement layers. This approach is called fine grain scalability (FGS). The advantage of this approach is that the data of a quality enhancement layer (FGS layer) can be truncated at any arbitrary point to limit data rate and quality without impact on the decoding process.

In the Fig. 1, each solid slice corresponds to at least one NAL unit. We see that with the error concealment methods proposed in SVC project the error will affect only one picture if the lost NAL unit belongs to the highest temporal level. The error drift is limited to the current GOP if the lost NAL unit is not in the quality base layer of the key picture. Otherwise, the error drift will expand in following GOPs until a key picture is coded as IDR-picture. An IDR-picture is an intra-coded picture and all of the following pictures are not allowed to use the pictures preceding this IDR picture as a reference.

Figure 1. Example of scalable video coding using the spatial scalibility, hierachical B-pictures and fine grain scalibility (FGS) to create a bit stream containing 2 spatial layers, 4 temporal levels, a quality base layer and a FGS layer. Each solid slice corresponds to at least one NAL unit.



B. *Structure of NAL units in bitstream of scalable extension of H.264/AVC*

Table I shows the NAL units order in a bit stream for a GOP with 2 spatial layers and 4 temporal levels. In the scalable extension of H.264/AVC the NAL header is extended to inform about the spatial layer, temporal level and FGS layer which this NAL unit presents. Because the quality enhancement layer (FGS index greater than 0) only degrades the quality of the corresponding picture and do not affect the decoder process if it is lost, it is not necessary to do error concealment for these NAL units. Therefore, only NAL units of the quality base layer (FGS index equal 0) are shown in Table I for simplification.

The NAL units are serialized in decoding order, but not in picture display order. It begins with the lowest temporal level and the temporal level will be increased after the NAL units of all spatial layers for a temporal level are arranged. The number of NAL units for the quality base layer in each level can be calculated from the GOP size or from the number of temporal level which is found in the parameter sets at the beginning of a bit stream. That means the NAL unit order can be derived from the parameter sets sent at the beginning of a transmission.

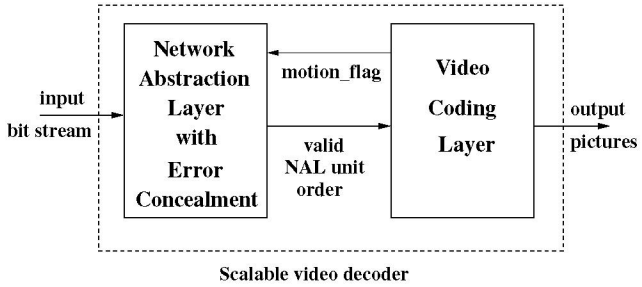
TABLE I. NAL UNIT ORDER IN A BIT STREAM FOR A GOP SIZE OF 8 WITH 2 SPATIAL LAYERS AND 4 TEMPORAL LEVEL S

No.	*Spat. layer*	*Temp. level*	*FGS*
1	**0**	**0**	**0**
2	**1**	**0**	**0**
3	**0**	**1**	**0**
4	**1**	**1**	**0**
5	**0**	**2**	**0**
6	**1**	**2**	**0**
7	**0**	**2**	**0**
8	**1**	**2**	**0**
9	**0**	**3**	**0**
10	**1**	**3**	**0**
11	**0**	**3**	**0**
12	**1**	**3**	**0**
13	**0**	**3**	**0**
14	**1**	**3**	**0**
15	**0**	**3**	**0**
16	**1**	**3**	**0**

III. MOTION-BASED ERROR CONCEALMENT IN NETWORK ABSTRACTION LAYER

Figure 2 shows the block diagram of the proposed scalable video decoder with error concealment in NAL. In our error concealment implementation we assume that the NAL units of a key picture in a GOP are not lost. For those NAL units a regular FEC (Forward Error Correction) method may be used [7]. We define that a lost NAL unit as a NAL unit belongs to a temporal level greater zero. If a NAL unit of a GOP is lost, a valid NAL unit order with a lower spatial resolution and/or lower frame rate is chosen. That means the maximal available spatial layer and/or the maximal available temporal level of this GOP is reduced. For example, if the 9-th NAL unit of a GOP in Table I is lost, our algorithm computes the NAL unit order in Table II to create a valid bit stream with the same resolution and only half of the original frame rate.

Figure 2. Scalable video decoder with motion-based error concealment in Network Abstraction Layer



In case that there are two possible valid NAL unit orders, the order with higher frame rate will be chosen if a lot of motion was observed in the last pictures. Otherwise, the order with the higher spatial resolution will be chosen. The motion flag given by VCL is set, if the average length of motion vectors in the last pictures is above a threshold. For example, if the 6-th or 8-th NAL unit of the GOP in Table I is lost, we can achieve two spatial layer and temporal level combinations. The first has spatial layer 1 and temporal level 1. The second has spatial layer 0 and temporal level 3. If the original bit stream reaches the spatial resolution CIF and a frame rate of 30Hz, than the first valid NAL unit order gives output pictures in (CIF, 7.5Hz) and the second in (QCIF, 30Hz). For the video segment with high motion the resolution (QCIF, 30Hz) makes sense because the human eyes are motion sensible. Furthermore, all of rendering techniques are able to up-sample the picture to a certain spatial resolution using interpolation.

TABLE II. THE VALID NAL UNIT ORDER IF THE 9-TH NAL UNIT OF THE GOP IN TABLE I IS LOST

No.	*Spat. layer*	*Temp. level*	*FGS*
1	**0**	**0**	**0**
2	**1**	**0**	**0**
3	**0**	**1**	**0**
4	**1**	**1**	**0**
5	**0**	**2**	**0**
6	**1**	**2**	**0**
7	**0**	**2**	**0**
8	**1**	**2**	**0**

TABLE III. THE TWO POSSIBLE VALID NAL UNIT ORDERS IF THE 6-TH OR 8-TH NAL UNIT OF THE GOP IN TABLE I IS LOST

No.	*Spat. layer*	*Temp. level*	*FGS*
1	**0**	**0**	**0**
2	**1**	**0**	**0**
3	**0**	**1**	**0**
4	**1**	**1**	**0**

No.	*Spat. layer*	*Temp. level*	*FGS*
1	**0**	**0**	**0**
3	**0**	**1**	**0**
5	**0**	**2**	**0**
7	**0**	**2**	**0**
9	**0**	**3**	**0**
11	**0**	**3**	**0**
13	**0**	**3**	**0**
15	**0**	**3**	**0**

In case that a NAL unit of highest temporal level is lost, for example the 9-te NAL unit of a GOP in Table I, it affects only the corresponding picture. In this case the error concealment algorithm can send a new NAL unit to the VCL to avoid an error drift in this temporal level and send a signal to the VCL or renderer directly requesting a picture repeat.

Moreover, in respect of complexity and error drift our error concealment method is suitable for a scalable video streaming system. In such system, if the packet loss occurs, the congestion control at the server reduces the number of layers and levels to adapt the sending data rate [8]. Therefore, if the client knows the principle of the congestion control at the server, it can predict the layer and level of the next GOP. In case of two possible valid NAL unit order the client can switch the current erroneous GOP in this tendency instead of using the motion flag. So the NAL with error concealment can work independent on the VCL.

IV. EXPERIMENTAL RESULTS

The error concealment in the NAL is implemented in our scalable video decoder [8], which is based on the reference software JSVM 3.0 [6] with the extension of IDR-picture for each GOP to allow the spatial layer switching. For the test a bit stream with 600 frames from the sequences *Mobile & Calendar* and *Foreman* with GOP size of 16 is used. This bit stream has two spatial layers. The lowest spatial layer (layer 0) has QCIF resolution and four temporal levels each at 1.875, 3.75, 7.5 and 15 Hz. The higher spatial layer (layer1) has CIF resolution and five temporal levels that give the additional frame rate of 30 Hz.

In Fig. 3 the blue curve shows the PSNR of output pictures from the erroneous bit stream with 5% loss of NAL units by using the proposed error concealment method and the red curve gives the PSNR of output pictures from the non-erroneous bit stream for the first 97 pictures. The PSNR calculation is based on the maximal spatial and temporal resolution, namely (CIF, 30Hz). If a GOP has lower frame rate, the output pictures are repeated to achieve 30Hz. For GOPs with a spatial resolution of QCIF we use the up-sampling filter in SVC with the following coefficients to obtain higher the spatial resolution CIF.

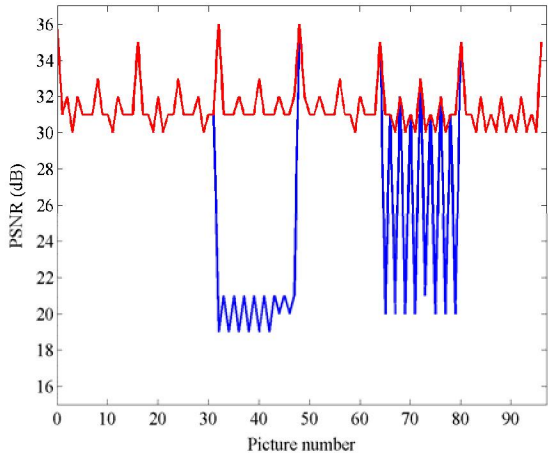


Figure 3. Red curve: PSNR of scalable video at high spatial temporal resolution. Blue curve: PSNR of scalable video with lower spatial resolution (Picture 33-49) or lower temporal resolution (Picture 65-81) due to 5% loss of NAL units.

$$h[i] = \{1,0,-5,0,20,32,20,0,-5,0,1\}$$

In Fig. 3 the pictures from 33 to 49 belongs to a GOP with an erroneous NAL unit order. The error concealment method chooses the new order to give the spatial resolution QCIF and a frame rate of 15 Hz. This gives soft images with relative smooth motion. For the GOP with the pictures from 65 to 81 the spatial resolution CIF and a frame rate of 15Hz are chosen resulting in sharp images with jerky motion.

Effectively the performance of this error concealment method is determined by the selected NAL unit order, which is based on the lost packet. This NAL unit order is an order that the server might choose to select based on network condition. Essentially our algorithm selects packets to be ignored based on actually lost packets in a computationally very efficient and pre-computed manner.

V. CONCLUSION

In this paper an error concealment method in the Network Abstraction Layer for the scalable extension of H.264/AVC is presented. The method can detect the NAL unit loss in a group of picture based on the knowledge of NAL unit order that can be derived from the parameter sets at the beginning of a bit stream. If a NAL unit loss is detected, a valid NAL unit order is arranged from this erroneous NAL unit order. In case of two possible valid NAL unit orders the order providing higher frame rate is chosen, if a lot motion was observed in the previous pictures. Otherwise, the valid NAL unit order providing the higher spatial resolution is selected. The error concealment method works under the condition that the NAL units of the key pictures are not lost. Our proposed method requires low computing power and does not produce error drift. Therefore, it is suitable for real-time video streaming.

REFERENCES

[1] J. Reichel, H. Schwarz and M. Wien, "Scalable Video Coding - Working Draft I," Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Doc. JVT-N020, January 2005.

[2] R. Schaefer, H. Schwarz, D. Marpe, T. Schierl and T. Wiegand, "MCTF and Scalability Extension of H.264/AVC and its Application to Video Transmission, Storage, and Surveillance," Proc. VCIP 2005, Bejing, China, July 2005.

[3] S. Wenger, Y.K. Wang and M. Hannuksela, "RTP payload format for H.264/SVC scalable video coding," 15th International Packet Video Workshop, Hangzhou, China, April 2006.

[4] "H.264: Advanced video coding for generic audiovisual services," International Standard ISO/IEC 14496-10:2005.

[5] J. Reichel, H. Schwarz, M. Wien, "Joint Scalable Video Model JSVM-6," Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Doc. JVT-S202, April 2006.

[6] J. Reichel, H. Schwarz, M. Wien, "Joint Scalable Video Model JSVM-3," Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Doc. JVT-P202, July 2005.

[7] S. Lin and D.J. Costello, "Error Control Coding:Fundamentals and Application," Englewood Cliffs, NJ: Prentice-Hall, 1983.

1-4244-0569-6/06/$20.00 ©2006 IEEE

[8] D.T. Nguyen and J. Ostermann, "Streaming and Congestion Control using Scalable Video Coding based on H.264/AVC," 15th International Packet Video Workshop, Hangzhou, China, April 2006.