Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Frame loss error concealment for SVC

CHEN Ying[1], XIE Kai[1], ZHANG Feng[1], PANDIT Purvin[2], BOYCE Jill[2]
([1]*THOMSON Corporate Research, Technology Fortune Center, Beijing 100085, China*)
([2]*THOMSON Corporate Research, Princeton, NJ 08540, USA*)
E-mail: {ying.chen; kai.xie; feng.zhang; purvin.pandit; jill.boyce}@thomson.net
Received Dec. 13, 2005; revision accepted Feb. 20, 2006

Abstract: Scalable video coding (SVC), as the Scalable Extension of H.264/AVC, is an ongoing international video coding standard designed for network adaptive or device adaptive applications and also offers high coding efficiency. However, packet losses often occur over unreliable networks even for base layer of SVC and have severe impact on the playback quality of compressed video. Until now, no literature has discussed error concealment support for standard SVC bit-stream. In this paper, we provide robust and effective error concealment techniques for SVC with spatial scalability. Experimental results showed that the proposed methods provide substantial improvement, both subjectively and objectively, without a significant complexity overhead.

Key words: Scalable video coding (SVC), Error concealment, Spatial scalability
doi:10.1631/jzus.2006.A0677 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Real-time transmission of video data in network environments, such as wireless and Internet, is a challenging task, as it requires coding efficiency, network friendliness and error robustness. Scalability is a possible solution for network adaptive applications. The Scalable Extension of H.264/AVC (SVC) (JVT of ISO/IEC MPEG & ITU-T VCEG, 2005b) aims at achieving both high compression performance and adaptivity for video delivery over heterogeneous networks. SVC is based on H.264/AVC and provides spatial, temporal and quality scalabilities. However, these scalabilities also pose a challenge to design a more efficient and robust decoder to handle packet losses during transmission over different networks.

Packet losses can produce undesirable effects at the decoder, including unpredictable behavior of the decoder and unacceptable playback quality. In order to prevent such effects, error control techniques are highly desirable. A number of algorithms have been proposed in the literature for error concealment. (Wang and Zhu, 1998) contains a good review of error control and other robust encoding techniques. It classifies error concealment into three categories: forward error concealment, post-processing error concealment and interactive error concealment. Some algorithms based on post-processing have been proposed in (Alkachouch and Bellanger, 2000; Atzori *et al*., 2001), which make use of temporal or spatial correlation between the macroblocks (MBs) in damaged area and its adjacent MBs in the same or previous frame. These algorithms assume that if either a single MB or a slice consisting of several consecutive MBs is lost, information from the neighboring available MBs or MBs in the adjacent frames can be used to estimate both motion vectors and texture of the missing MB.

However, the underlying assumptions that the above error concealment algorithms are based on cannot be guaranteed in many applications. During transmission, bit errors could render a whole packet useless. Particularly, when transmitting a low bitrate video sequence over a wireless Local Area Network (LAN) or a Universal Mobile Telecommunications System (UMTS) link, a coded picture typically fits in one packet. Therefore, a transmission error will lead to a loss of a whole slice or frame.

Algorithms handling frame loss have also been proposed. For example, Belfiore *et al.*(2005) addressed a pixel-level algorithm based on the optical flow theory, assuming that the motion between two consecutive pictures does not vary in a dramatic way. It exploited motion information in a few past frames to estimate forward motion of the last received frame.

However, the estimated motion vector (MV) was not very accurate in the optical flow based methods for video signals with fast motion, and usually zero residuals are used. Error concealment in SVC with spatial scalability, can overcome these problems by concealing the enhancement layer if the corresponding base layer packet is correctly received. The presence of the base layer can also improve the accuracy with which MV and residual for the enhancement layer can be predicted.

In this paper, we call the error concealment method "Inter Layer Error Concealment" if it takes advantages of the base layer information. If the error concealment just considers information in the same spatial layer, we call it "Intra Layer Error Concealment".

The error concealment techniques in this paper have been adopted as a non-normative tool at the JVT meeting in Nice, France in October 2005 (Chen *et al.*, 2005).

JSVM 3.0 (JVT of ISO/IEC MPEG & ITU-T VCEG, 2005a) software has no error concealment support. The decoder crashes when there are losses in the bit-stream. In this paper, we propose two inter layer error concealment methods and compare them with two well-known intra layer error concealment methods. We used the packet loss model of VCEG (Wenger, 2002) and the testing conditions of (Wang *et al.*, 2005). In this paper, we focus on two layer spatial scalability and propose methods to conceal entire missing frames of an SVC video sequence while maintaining low complexity and high quality. Our results showed that using base layer information to conceal the enhancement layer can usually outperform methods which do not consider the base layer information. *PSNR* gains of up to 4 dB were observed.

The rest of the paper is organized as follows: Section 2 gives a brief overview of SVC. The proposed algorithms are discussed in Section 3. Experiment results are shown in Section 4 and conclusions are drawn in Section 5.

BACKGROUND FOR SVC

Network abstraction layer of H.264/AVC

SVC (based on H.264/AVC) is organized into two different conceptual layers, the video coding layer (VCL) and the network abstraction layer (NAL): the VCL specifies the syntax elements necessary to give an efficient representation of the coded video signal, whereas the NAL defines the interface between the video codec itself and the outside world. The standard also describes the concept of NAL units (NALUs) which provides support for the packet-based approach of most networking technologies (ITU-T H.264.1, 2005).

SVC spatial scalability

SVC provides spatial scalability mainly by the following three tools: texture prediction, motion prediction and residual prediction.

(1) Texture prediction: Each MB of a base layer can be reconstructed and upsampled as a texture predictor of the enhancement layer MBs. MB in an enhancement layer can choose to be coded as an IntraBL mode by subtracting the upsampled base layer reconstruction.

(2) Motion prediction: If this prediction is enabled, motion information of the base layer MB is upsampled for the enhancement layer MBs. Normally, reference indexes are directly inherited by the enhancement layer. MVs are upsampled by 2 for dyadic case. In addition, quarter pel refinement can be used to further improve the accuracy of the upsampled motion if necessary.

(3) Residual prediction: Base layer residual is simply upsampled and before enhancement layer residual is coded, it subtracts the upsampled base layer residual and the difference is then coded.

Rate-Distortion Optimization (RDO) is used to select which modes will be selected for a particular MB. If base layer information is available, these tools may benefit the error concealment techniques as will be observed in the following sections.

H.264/AVC temporal direct mode

For Intra Layer Error Concealment, the temporal direct mode defined in H.264/AVC specification (ITU-T H.264.1, 2005) can be used to provide efficient MV prediction for lost slices. The MV of MB or MB partition in Temporal Direct Mode is calculated based on the MV of the co-located MB or MB partition. The co-located MB or MB partition is found in the first reference picture of Reference Picture List 1 (RefPicList1). Then, the valid forward MV of co-located MB or MB partition is scaled based on the temporal distance to generate MV for the current MB or MB partition, as shown in Fig.1.

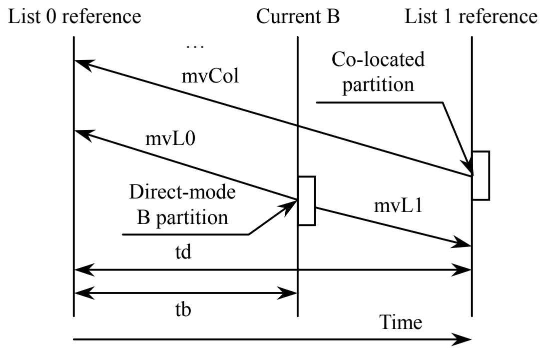

Fig.1 Example for temporal direct-mode motion vector inference

PROPOSED ALGORITHM

We propose error concealment algorithms for bit-streams containing missing coded pictures. Our tests are based on spatial scalability, with two layers as described in (Wang *et al*., 2005) with losses in the enhancement layer. However, base layer only error concealment is also supported by the software and it can be extended to support SNR scalability.

We propose 4 error concealment algorithms (2 intra layer methods and 2 inter layer methods) for cases when the base layer has no loss and when base layer has loss. The error concealment algorithms are: (1) Frame copy (FC); (2) Temporal direct motion vector generation (TD); (3) Motion and residual upsampling (BLSkip); (4) Reconstruction base layer upsampling (RU).

Frame copy (FC)

This is an intra layer error concealment method. In the "frame copy" (FC) algorithm, each pixel value of the concealed frame is copied from the corresponding pixel of the first frame in Reference Picture List 0 (RefPicList0). This algorithm can be invoked for both base layer and enhancement layer.

The base layer FC will only be done when it needs to be rendered. For the two layer spatial scalability case, a lost enhancement layer frame will be concealed using the first frame in the reference RefPicList0. The RefPicList0 is generated from the decoding process of the enhancement layer and is of a higher spatial resolution.

Temporal direct (TD)

This is an intra layer error concealment method. In the "temporal direct motion vector generation" (TD) algorithm, the MVs and reference indices of each subblocks of the missing frame are calculated as if they were coded using the "temporal direct mode". This algorithm can be invoked for both base layer and spatial enhancement layer.

For enhancement layer, if TD is used, each MB in the lost high resolution frame will use TD mode to get and scale the MVs from the collocated MB at the enhancement layer. So the error concealment process does not need base layer information.

Motion and residual upsampling (BLSkip)

This is an inter layer error concealment method. In the "motion and residual upsampling" (BLSkip) algorithm, SVC tools are used and the BLSkip mode is set in the enhancement layer. Residual upsampling is also used to upsample the residual of the base layer for enhancement layer. However, the motion compensation is done at the enhancement layer using the upsampled motion fields. This algorithm can directly be used for the enhancement layer if there is no packet loss in the base layer. If base layer is also lost, it needs to generate the MVs using TD method before BLSkip can use motion field upsampling. We call this method as BLSkip+TD, however for simplicity we will use BLSkip to represent this method throughout this paper.

Reconstruction base layer upsampling (RU)

This is an inter layer error concealment method. In the "reconstruction base layer upsampling" (RU) algorithm, the base layer picture is reconstructed and upsampled using the H.264/AVC 6-tap filter for the lost enhancement layer picture. If base layer packet is also lost, the FC method is used for the enhancement layer, rather than using an upsampled concealed base layer picture.

For packet loss of key frames of the base layer, frame copy error concealment is used. Memory Management Control Operation (MMCO) and Reference Picture List Reordering (RPLR) commands are created during the error concealment process, to emulate the reference picture and non-reference picture patterns at the encoder.

MMCO commands will be generated to delete all the non-key pictures of the previous GOP and the key picture before the previous key picture (if it exists). RPLR commands will be used to guarantee the lost key picture refers to the previous key picture.

This algorithm only works properly if regular GOP coding patterns were used, e.g., adaptive GOP size (AGS) cannot be detected at the decoder and will not be concealed properly.

EXPERIMENT

Experimental conditions

The proposed error concealment algorithms were tested using the packet loss model of VCEG (Wenger, 2002) and the testing conditions of (Wang *et al*., 2005). Our scenario focuses on the two spatial layer cases.

Our simulations only consider the cases when entire frames are lost although the algorithms can be simply extended to support slice loss cases.

FC and TD methods can support error concealment for bit-stream with temporal scalability, such as SVC base layer which is H.264/AVC compliant. Our implementation can be extended to support bit-stream with quality layers by dropping the FGS layer packet if its base layer or lower layer packet is lost. Among all the three scalabilities, error concealment for spatial scalability bit-stream is very critical.

We applied our algorithms to the JVT common conditions video sequences (YUV 4:2:0) as described in (Wang *et al*., 2005), including progressive sequences of News, Football, Foreman, Paris and Stefan with standard frame rate (Chen *et al*., 2005).

The following encoder parameters were used:

(1) JSVM 3.0, Base layer QCIF, enhancement layer CIF, two-layer spatial scalability with same frame rate;

(2) *GopSize*=16 with one I frame for each 32 frames;

(3) Adaptive inter layer prediction;

(4) *QP* (Quantization Parameter)=28, 32, 36, 40 same *QP* for two layers;

(5) No Slice Groups, FMO, ASO.

Four packet loss patterns with average packet loss rates of 3%, 5%, 10%, and 20% included in ITU-T VCEG were employed. The error patterns were generated from the experiments on the Internet backbone between one sender and three reflector sites. Details of the generation and file format are available in (Wenger, 2002). In our experimentation the packet loss is determined by the location of the character "0" in the error trace file. A "0" indicates a dropped NAL packet and a non zero character indicates successful packet transmission. The base layer packet may have the same or lower packet loss rate compared to the enhancement layer. During our simulations if a base layer packet is lost, we force the corresponding enhancement layer packet to be treated as lost also. We considered the following combinations of base layer and enhancement layer packet loss ratio for our simulations, where 0 means no packet loss for that layer.

All the coded bit-streams are iterated to over 4000 NAL packets and then bit-streams with dropped pictures were created utilizing the packet loss patterns. The lost bit-streams were decoded using the above error concealment algorithms.

Results

In our test suite, we simulated the different loss conditions both in the enhancement layer and the base layer as mentioned in Table 1. Average *PSNR* difference results are reported. Here we compare BLSkip, RU and TD with FC in order to obtain one measure for each sequence. The proposed solution for enhancement layer error concealment significantly depends on the base layer information.

Table 1 Packet loss rate for base layers and enhancement layers

Layer	Loss rate (%)						
B	0	0	3	3	3	5	5
E	3	5	3	5	10	5	10

We report the performance for three different packet loss conditions in Table 2. The results showed that BLSkip, RU, TD obtained significant gains versus the trivial FC (previous frame repetition) concealment. Also, the simplicity of our solution makes it preferable for real-time implementation.

Fig.2 shows comparison of the four schemes with the no error case for the Foreman sequence at different packet loss rates. It is clear that BLSkip and RU proposed schemes have much better performance compared to the other techniques. When the base layer has no loss, the *PSNR* degradation of enhancement layer reduces compared to Intra Layer Error Concealment, since the motion or texture (reconstruction or residual) of base layer can be used for enhancement in BLSkip and RU schemes. Even when the loss of enhancement layer reaches 5%, the *PSNR* value for BLSkip and RU schemes still remains close to the no error case. As expected, the base layer information is very important to BLSkip and RU schemes. More packet loss in base layer would weaken the performance of BLSkip and RU schemes. However, even in this case the BLSkip and RU schemes obtain significant gains over FC and TD schemes. Examples of reconstruction quality of the concealed frames using the four schemes are shown in Fig.3. Fig.4 shows an example to indicate that TD may get texture with more detail than RU in low motion cases.

Table 2 Δ*PSNR* for different base layer packet loss rate (*p*) and enhancement layer packet loss rate (*q*)

Sequence	Δ*PSNR* (dB)								
	p=0%, *q*=5%			*p*=3%, *q*=10%			*p*=5%, *q*=10%		
	BLSkip vs FC	RU vs FC	TD vs FC	BLSkip vs FC	RU vs FC	TD vs FC	BLSkip vs FC	RU vs FC	TD vs FC
News	1.529	0.626	0.175	3.016	1.390	0.069	2.975	1.267	0.405
Foreman	2.240	2.120	0.251	4.017	3.845	0.308	3.921	3.615	0.398
Football	1.469	1.588	0.163	2.738	2.919	0.314	2.750	2.814	0.374
Paris	0.933	−0.118	0.191	1.766	−0.697	0.430	1.715	−0.677	0.508
Stefan	2.119	1.482	0.270	3.760	2.663	0.421	3.673	2.533	0.495
Average	1.658	1.140	0.210	3.059	2.024	0.308	3.007	1.910	0.436

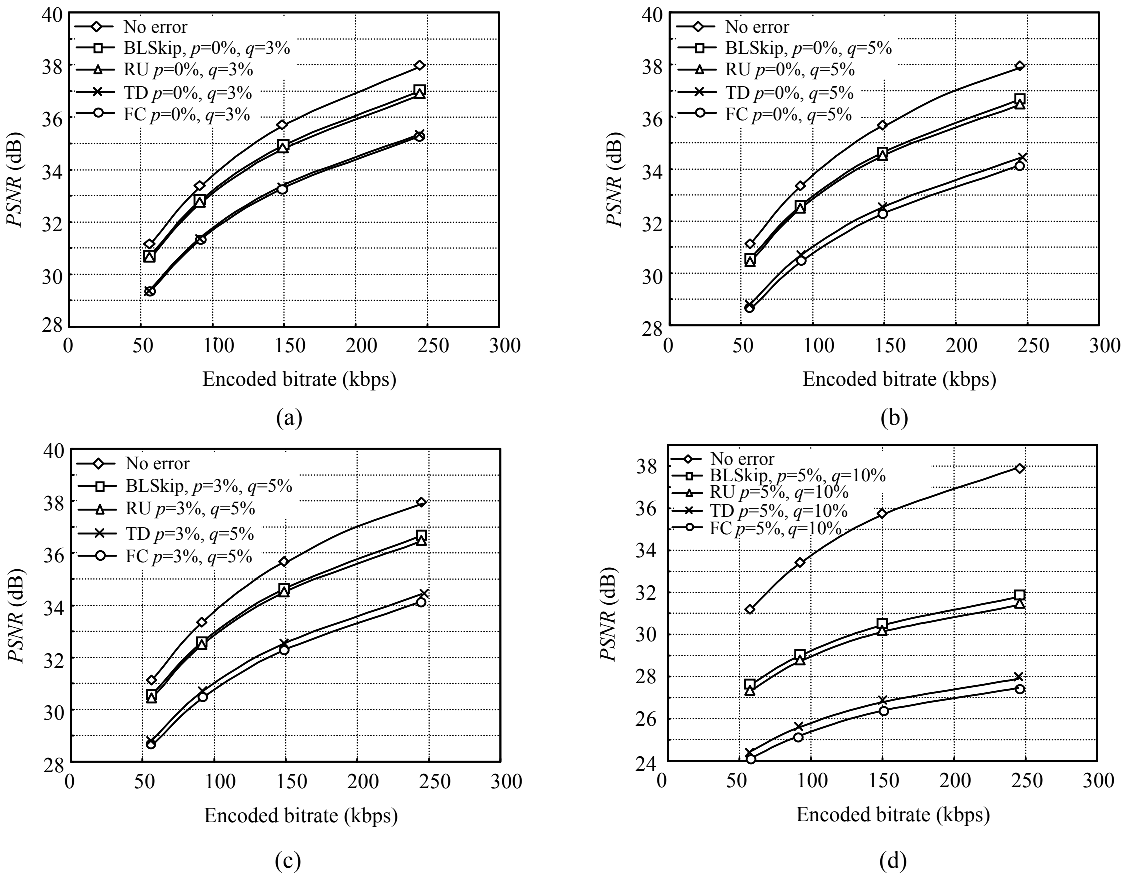

Fig.2 Comparison BLSkip, RU, TD, FC and No error cases with "Foreman" sequence for different base layer packet loss (*p*) and enhancement layer packet loss (*q*). (a) *p*=0%, *q*=3%; (b) *p*=0%, *q*=5%; (c) *p*=3%, *q*=5%; (d) *p*=5%, *q*=10%


(a)


(b)


(c)


(d)

Fig.3 Subjective quality comparison of concealed pictures for "Foreman", with a base layer packet loss rate of 3% and an enhancement layer packet loss rate of 3%, *QP*=28. (a) BLSkip; (b) RU; (c) TD; (d) FC


(a)


(b)

Fig.4 Frame 101 of the "Paris" sequence with base layer 3% loss and enhancement layer 3% packet loss, *QP*=28. (a) TD; (b) RU

Analysis

From the comparisons above, BLSkip method, which utilizes upsampled base layer motion, in most cases is better than the other error concealment methods. Although, there are cases where it might also cause more artifacts compared to RU method, which uses the upsampled reconstruction of the base layer. While RU requires multiple-loop decoding, BLSkip is well suited for single loop decoding.

TD method can also occasionally give good concealed picture for low motion case. As shown in Fig.4, since the whole background is almost still, TD method gives better objective and subjective quality than RU method.

FC method turns out to be the worst method, especially for cases where there is error prorogation. However, it can also be used as a tool for low complexity applications by forcing the decoder to drop pictures with high temporal level to use the temporal scalability property of SVC.

Comparing the subjective quality of BLSkip and RU methods, we can find that BLSkip is usually better than RU. The essential reason is that BLSkip processes the motion compensation at the enhancement layer, so temporally the quality variance is low even if there are artifacts in some independent pictures. Since the motion compensation is done at base layer and upsampling is addressed to get the high resolution pictures for RU method, the concealed pictures are blurred. The effect caused by RU method looks like the video is out of focus from time to time, to which human eyes are very sensitive.

CONCLUSION

Scalable video coding (SVC), as the Scalable Extension of H.264/AVC, is an ongoing standard and apart from better coding efficiency it provides more adaptivity to heterogeneous network. Efficient and fast error concealment thus is highly desired for the robustness and flexibility of SVC based applications. In this paper, we present error concealment algorithms for SVC, focusing on spatial scalability. Simulation results showed that using base layer information to conceal the enhancement layer usually gives better reconstruction results with *PSNR* gains of up to 4 dB. Our proposed methods can conceal entire missing frames of an SVC video sequence with low complexity and high quality.

References

Alkachouch, Z., Bellanger, M.G., 2000. DCT-based spatial domain interpolation of blocks in images. *IEEE Transactions on Image Processing*, **9**(4):729-732. [doi:10.1109/83.841948]

Atzori, L., DeNatale, F.G.B., Perra, C., 2001. A spatio-temporal concealment technique using boundary matching algorithm and meshbased warping (BMA-MBW). *IEEE Trans. on Multimedia*, **3**(3):326-338. [doi:10.1109/6046.944476]

Belfiore, S., Grangetto, M., Magli, E., Olmo, G., 2005. Concealment of whole-frame losses for wireless low bit-rate video based on multi frame optical flow estimation. *IEEE transactions on multimedia*, **7**(2):316-329. [doi:10.1109/TMM.2005.843347]

Chen, Y., Boyce, J., Xie, K., 2005. Frame Loss Error Concealment for SVC. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Document JVT-Q046. Nice, FR.

ITU-T H.264.1, 2005. Telecommunication Standardization Sector of ITU, Series H: Audio-Visual and Multimedia Systems, Infrastructure of Audiovisual Services—Coding of Moving Video. Conformances Specification for H.264 Advanced Video Coding.

Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 2005a. Draft Output Document from JVT JVT-P202-JSVM3, Draft of Joint Scalable Video Model JSVM-3. 16th Meeting. Poznan, Poland.

Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 2005b. Draft Output Document from JVT JVT-Q201, Draft of Scalable Video Coding—Working Draft 4. 17th Meeting. Nice, France.

Wang, Y., Zhu, Q.F., 1998. Error control and concealment for video communication: a review. *Proc. IEEE*, **86**:974-997.

Wang, Y.K., Wenger, S., Hannuksela, M.M., 2005. Common conditions for SVC error resilience testing. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Document JVT-P206d1.

Wenger, S., 2002. Error Patterns for Internet Experiments. VCEG Q15-I-16r1.