Optimal Slice Size for SVC Error Resilient Video Coding

Mikko Uitto[1], Janne Vehkaperä[1], and Peter Amon[2]

[1]VTT Technical Research Centre of Finland,
Kaitoväylä 1, FI-90571, Oulu, Finland
{mikko.uitto, janne.vehkapera}@vtt.fi
[2]Siemens Corporate Technology,
Otto-Hahn-Ring 6, 81739 Munich, Germany
p.amon@siemens.com

Abstract. Scalable Video Coding (SVC) as an extension to the H.264/AVC standard enables adaptive video transmission, where several types of sub-streams can be decoded from a single encoded stream. The same video content can be streamed to low bit rate mobile phones with low quality as well as high bit rate televisions with extremely high quality. However, strict real-time requirements and unreliable transmission channel can cause packet losses, which means that sufficient error protection and concealment methods are needed. One of the error protection techniques in the encoder enables division of a picture into slices. Introduction of slices will decrease the coding efficiency but at the same time it will improve resilience against transmission errors. In this paper, we try to find a trade-off between the number of slices and coding efficiency for H.264/SVC to be used for video streaming in error-prone networks. In addition to coding efficiency, the selected slice size will affect on error resilience and error propagation inside the video stream. This means that the optimal slice size should not only provide sufficient coding efficiency but it should also provide a good error concealment ratio in relation to packet losses. This paper evaluates the trade-off between the coding efficiency and error resilience of H.264/SVC and the simulation results presented in the paper indicate that a minor increase of the amount of slices per picture greatly improves the error resilience but does not reduce the coding efficiency greatly.

Keywords: Scalable video coding, SVC, error resilience, error concealment

1. Introduction

The demand for fast and location-independent access to multimedia services is increasing day by day. More and more people are willing to access different multimedia services from the Internet, such as watching videos from a streaming server in real-time or have a video chat with a friend using a wireless device. The primary goal for many video compression standards is to provide good coding efficiency. However, in recent use scenarios media adaptation to frequently changing transmission conditions is also one of the challenges to be considered [1].

Scalable video coding has gained interest widely, since it allows flexibility both in transmission as well as in the receiving terminal. The encoding process for complex multi-layered streams can be heavier in SVC compared to H.264/AVC, especially when inter-layer prediction is used. On the other hand, SVC allows video stream adaptation without the need for other adaptation methods, such as transcoding or rate control.

A scalable video stream includes a base layer and one or several enhancement layers. The latter ones cannot be used independently but combined with the base layer they can achieve one of the three scalabilities: spatial, temporal or quality. Spatial scalability enlarges the resolution, temporal scalability increases the frame rate and quality scalability improves the video quality by using smaller quantisation points. The latter one is often referred also as SNR scalability.

The current state-of-the-art video coding standard, H.264/AVC [2], has managed to produce excellent coding efficiency with the help of powerful coding and prediction tools. These tools have been used also in the scalable extension of H.264/AVC, which we will refer to as H.264/SVC throughout the paper, together with some new tools which enable three-dimensional scalability of the stream. Scalability is an anticipated feature for video streaming applications especially when unreliable networks are used for transmission. In order to cope with the unreliability of networks, the H.264/AVC standard has introduced a slice coding tool, which can be used to improve the robustness of video against packet losses. Slices should improve the robustness since in the case of packet loss only a portion of a picture will be lost and the error will not propagate to any other slice in the picture due to independency of slices [3]. It is also commonly known that the error rate in wireless channels will significantly increase with increased packet length [1]. In the literature, many studies of coding efficiency and error resilience of H.264/AVC when slices have been used are published [1, 3, 4] but similar studies have not been made for H.264/SVC. In this paper we will study how the number of slices per picture and thus the slice size will affect both coding efficiency and error resilience. The complexity of H.264/SVC and different intra- and inter-layer prediction tools available will significantly affect especially error propagation in H.264/SVC. Finding a good trade-off between the coding efficiency and error robustness will help when specifying a video streaming system based on H.264/SVC.

This paper is organized in the following way. The second section introduces the scalable video coding extension of H.264/AVC. The third and fourth sections describe existing error protection and concealment strategies focusing on both encoder and decoder techniques used in the simulation studies. In the fifth section we study the effect of the slice size on coding efficiency and also on error resilience when different error concealment techniques are used. The last section draws a conclusion of the results achieved.

2. Overview of scalable video coding

The current state-of-the-art video compression standard H.264/AVC and its scalable extension (Annex G), also known as H.264/SVC, were both jointly developed by the

ISO/IEC MPEG and ITU-T VCEG standardization organizations [2]. The most essential advantages introduced by H.264/SVC compared to previous standards are its layered coding mechanism including spatial, temporal and quality scalabilities. The base layer is compliant to H.264/AVC. The scalability option has been supported also in previous MPEG-2 and MPEG-4 standards, but the scalability schemes combined with the powerful coding and prediction tools in H.264/AVC can improve the flexibility and functionality remarkably.

Fig. 1 shows an example of the variety of SVC streams, where only spatial and temporal scalabilities are present. As can be seen, multiple sub-streams can be decoded from a single encoded stream with varying device capabilities. In this example, the sub-stream with the highest resolution, frame rate, and quality is the HDTV stream. One of the most obvious use scenarios for scalability is to adapt the video according the characteristics of a transmission network especially when the transmission medium is unreliable. Then, with suitable adaptation mechanisms the quality for the receiving terminal could be dropped in order to maintain successful decoding with the price of lower quality and bit rate.

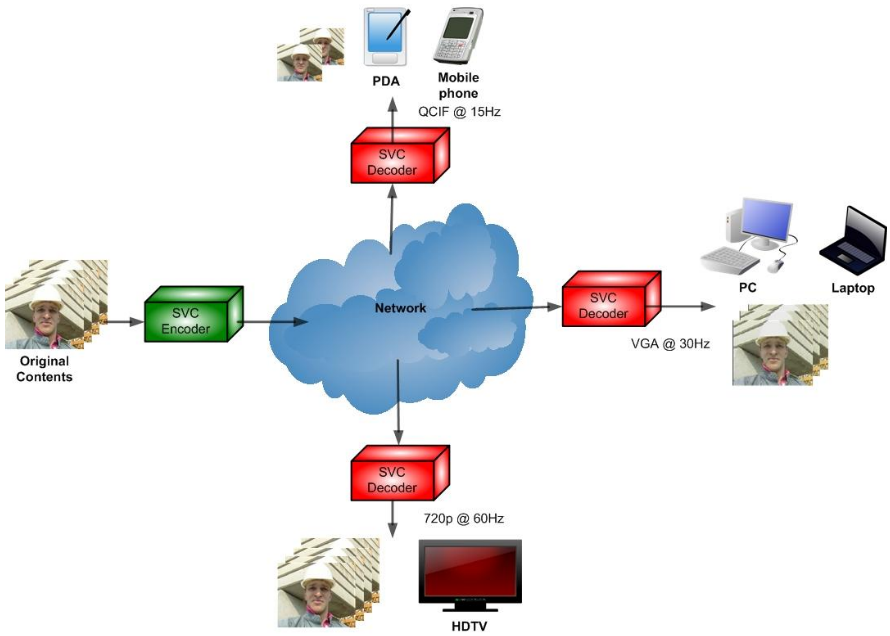


Fig. 1. Scalable video coding and its benefits.

3. Error resilient coding tools for H.264/SVC

Error resilient coding tools of the encoder utilized to prevent errors play an important role in nowadays wireless transmission. The popularity of wireless terminals as a video receiver and limited bandwidth demands to develop new algorithms with higher coding efficiency. However, higher coding efficiency means that more picture information is included in fewer bits, which creates a greater risk that more important information is lost among one packet loss.

This chapter will introduce some of the techniques that have been developed in order to improve error resilience of the encoded stream. Packet re-transmission could be limited or impossible in real-time applications since it causes delays or it is impossible (e.g. for broadcasting). It is known that the longer the RTP packet is in the application layer, the more vulnerable it is to transmission errors especially in a wireless channel. However, the efficient prediction tools, inter-layer prediction structures and adaptation in lossy network environments have increased the popularity of using scalable video coding. Fig. 2 shows the simplified structure of an SVC encoder, depicting the generation of several layers.

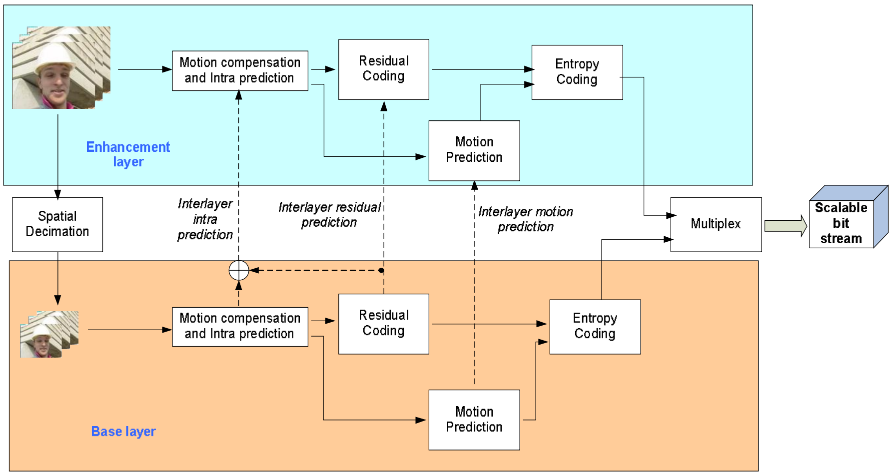


Fig. 2. Simplified SVC encoder .

The H.264 video coding standard supports several different tools which can be used to improve error robustness. In this paper, we will focus on slice coding and how this affects on coding and error concealment efficiency. Other important error resilience tools supported by the H.264 video coding standard are intra MB/picture refresh, reference picture identification and selection, data partitioning, spare picture signaling, redundant slices/pictures, parameter sets, flexible macroblock ordering (FMO), gradual decoding refresh (GDR), scene information signaling, SP/SI pictures, constrained intra prediction and reference picture marking repetition (RPMR) [5]. All tools except SI/SP pictures and data partitioning are included in also in H.264/SVC profiles and H.264/SVC introduces also three new error resilient tools, namely quality layer integrity check signaling, redundant picture property signaling, and temporal level zero index signaling. In the following, slice and FMO coding tools are introduced.

3.1. NAL packetisation and slice support

Network abstraction layer (NAL) packetisation for H.264/AVC, and naturally also for SVC, is designed to support packet-based transmission and have been found to perform well in lossy network environments [6]. In the most traditional case, the base layer picture contains only one NAL unit (NALU) and the enhancement layer picture forms a second NALU, correspondingly. Each of the enhancement layer pictures is packetised into its own NALU. The advantage is that the total amount of NAL packets, which will be sent over the network, is low. However, the effect of one missing packet is then more significant since more video data will be lost and some fragmentation might be needed due to limitation of data unit size at lower system layers, e.g., maximum IP packet size or maximum transmission unit (MTU) size in general.

The pictures in all the layers can be divided into slices as presented on the left in Fig. 3. The total number of NALUs will be increased since each slice is packetised into its own NALU. Naturally, the advantage of using slices is that one missing packet has a smaller influence for the video quality. The disadvantage of using slices is that it will decrease the coding efficiency as presented in [3] for H.264/AVC due to the restriction of prediction from another slices. In Section 5, we will study how the slice size will affect the coding efficiency in H.264/SVC.

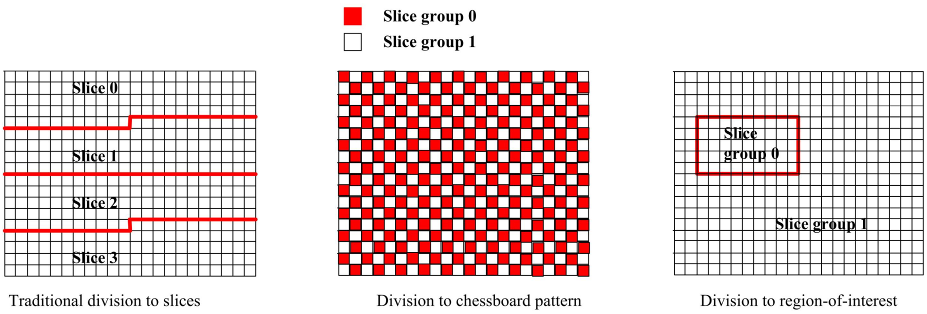


Fig. 3. Traditional division to slices and slice groups (FMO).

3.2. Flexible macroblock ordering

Flexible macroblock ordering (FMO) is a very efficient error-resilience tool for H.264/AVC. The main benefit using FMO is that it is not restricted to slices consisting of neighboring macroblocks. With the help of this, each macroblock can be assigned freely to certain slice group. The benefit of using a scattered macroblock order is that reconstruction of missing blocks is easier since it is possible to use the information of the surrounding macroblocks. As a result, the errors are scattered more equally to the whole picture than if they were limited to one certain area. FMO has a lot of potential especially in error resilience. Both subjective and objective quality is improved and its use is recommended especially in environments, where severe packet losses are expected [7].

One of the FMO patterns is the chessboard pattern: It is widely used in different applications, but the negative aspect is its great number of header data since all macroblock addresses are needed. A better coding efficiency can be reached with FMO interleaving, although the recovery from errors is then more problematic [4]. Fig. 3 presents two possible FMO patterns, the chessboard pattern and region of interest.

4. Error concealment for H.264/SVC

Error concealment is an important tool especially when forward error correction fails. The goal of error concealment is to exploit correctly received video data to recover the missing data or minimizing the deterioration caused by the loss. Some of the techniques in H.264/SVC are applicable also to H.264/AVC, while some of them take advantage of the SVC coding structure and coding tools as presented in [5]. Novel error concealment tools for H.264/SVC can utilize the correlation between different layers which is not possible in the traditional single-layer coding. Error concealment tools utilizing this inter-layer correlation can improve the performance of error concealment. Data for the scalable enhancement layer can be partially predicted from the base layer, which emphasizes the importance of the base layer. The enhancement layer data is usually useless if the corresponding base layer slice is missing. Furthermore, the non-existing macroblock information inside a frame can lead to wide-ranging error propagation to other frames as well. The adaptive inter-layer prediction can decrease the effects of the missing picture data, but error concealment is still needed in order to provide better quality of experience.
This section presents four different error concealment strategies which we have implemented in the H.264/SVC reference decoder: pixel-value interpolation, frame copy, spatial enhancement layer utilization, and upsampling. All these methods are designed to function in spatially scalable video and they rely on the macroblock map, which tells the missing macroblock locations inside a frame. The first two error concealment techniques work both for spatial base and enhancement layer slices whereas the spatial enhancement layer utilization can be used only for concealing the enhancement layer slice when base layer slice is missing. Furthermore, the upsampling procedure only functions for the missing spatial enhancement layer slice (base layer needs to be correctly received).

4.1. Error concealment for the base and enhancement layer

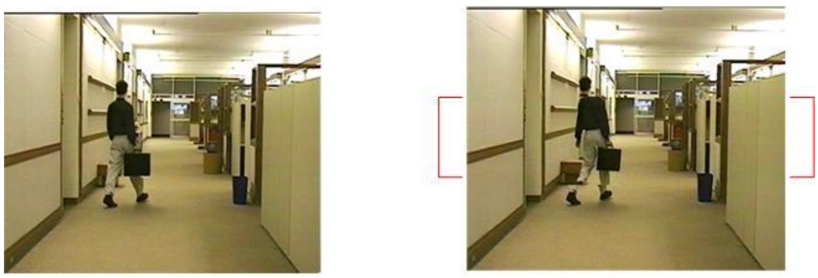
The implemented pixel-value interpolation function uses the correctly received and decoded macroblock and pixel areas from the same frame where the missing slice is located. This concealment method uses above and below macroblocks as sources to conceal the intermediate macroblock. If only one of the two sources is available, the interpolation is done using only one reference. The worst case is that both sources are unavailable, which means that macroblock lines are skipped as long as a correctly received macroblock is found. The reconstruction is done in raster scan order. Fig. 4 shows an example of interpolation.

a) Original, error-free frame. PSNR 42,5 dB

b) Concealed frame. PSNR 21,5 dB

Fig. 4. An example of interpolation.

Similarly as interpolation, frame copy (FC) is an intra layer error concealment method, which functions for both, base and enhancement layers. Its principle is simply to copy the missing pixel values to the erroneous frame from the corresponding pixel of the first frame in the reference picture list. Two separate reference frame lists can be applied where the missing pixels can be copied. One of the lists consists of temporally previous frames in the video stream of the enhancement layer and it is a higher spatial resolution. The second one works only to B-frames (list 1) and it consists of temporally latter frames in the sequence. When the key frames are missing from the beginning of a GOP, the FC method is always used [9]. Since motion between frames is not considered at all, this method is not very effective to use in video sequences that contain a lot of motion. Fig. 5 shows an example of frame copy.



a) Original, error-free frame. PSNR 41,0 dB

b) Concealed frame. PSNR 27,6 dB

Fig. 5. An example of frame copy.

4.2. Error concealment for the base layer

As was mentioned earlier, the missing base layer slice usually leads to rejecting also the dependent enhancement layer slice. In many cases, this is true if the higher layer needs the base layer data for the picture formation. The macroblocks in a frame can be coded differently. The basic assumption is that the blocks with a lot of motion are intra-coded while the static areas are inter-coded. The inter-coded macroblocks are predicted from other pictures from the same layer and also from the lower layer, since the hierarchical prediction structure is usually used in SVC.

The spatial enhancement layer utilization method was presented in [10]. In this method, the texture elements of the correctly received enhancement layer data are gathered and used in the reconstruction. Since some of the base layer data is still needed, the macroblock copy method is applied to all the blocks that are inter-coded. Fig. 6 shows an illustrative example of enhancement layer utilization. As can be seen, the visual results are very good. The particular reference picture contains many intra-coded macroblocks since it has motion from one frame to another. The majority of the frame is nicely reconstructed and the blocking phenomenon is small. The PSNR value is quite poor, which confirms the assumption that the PSNR value is not the best indicator for the assessment of artifacts caused by a packet loss.

a) Original, error-free frame. PSNR 42,2 dB

b) Concealed frame. PSNR 28,3 dB

Fig. 6. An example of spatial enhancement layer utilisation.

4.3. Error concealment for the spatial enhancement layer

The scalable video layers are usually sent to different RTP ports as well as signed with different priority in order to protect the video data better. In spatially scalable video, the base layer is usually better protected with higher priority. Since the spatial enhancement layer can be more vulnerable to packet losses, the error concealment techniques for the higher layers must be taken care of even better than for the base layer in some cases. The two concealment method presented in the previous section, interpolation and frame copy, can be also used in the concealment of spatial enhancement layers. This section presents the upsampling approach, which was also implemented into H.264/SVC decoder.

The upsampling technique is an effective solution for concealing the missing spatial enhancement layer picture if the corresponding base layer picture is correctly received. The H.264/SVC decoder usually needs the target layer resolution to be decoded. If a spatial enhancement layer is missing and no error concealment is used, the current frame is then simply discarded by the decoder since the resolution is smaller. This affects other frames as well, especially if the missing frame is a reference to these frames. The usage of the upsampling method is appealing because the change in video resolution in the middle of the playback is a very annoying quality phenomenon. With this error concealment method, the lower resolution data will be upsampled to the target resolution from the lower layer (e.g., base layer) at the expense of quality.

The upsampling requires that the base layer (or lower layer) picture data is correctly received. This picture is then upsampled to the target resolution using normative integer-based 4-tap filters. Since the base layer has often lower quality, also the upsampled enhancement layer has then lower quality which can be seen as a blurred picture as illustrated in Fig. 7. Basically, upsampling assumes that the enhancement layer frame is totally missing. For slice-oriented frames it is not always sensible to discard the whole enhancement frame, if only few slices are missing.

(a) Base layer picture (lower layer) (b) Spatial enhancement layer picture (lower layer) (c) Upsampled picture (from the base layer)

Fig. 7. An example of upsampling.

5. Simulations

The simulations section comprises two parts. Section one evaluates the effect of slice partitioning on coding efficiency whereas the section two evaluates the influence of slice partitions to error concealment efficiency. Three different sequences were used in the simulations: ‘City’, ‘Harbour’ and ‘Soccer’. The “City’ sequence has a slightly moving camera and background, whereas the ‘Harbour’ has a static camera and only couple of moving targets. The ‘Soccer’ contains a lot of motion, both from the camera as well as the background. The encoding parameters are presented in Table 1. All the sequences were encoded with JSVM 9.15 reference encoder [11].

Table 1. Encoding parameters.

Sequence	*City*	*Harbour*	*Soccer*
Number of frames	300	300	300
Frame rate(Hz)	30	30	30
IDR period	only first	only first	only first
Intra period	32	32	32

GOP length	8	8	8
Sequence structure	IPP…PPI	IPP…PPI	IPP…PPI
Resolution (BL)	QCIF/CIF	QCIF/CIF	QCIF/CIF
Resolution (EL)	CIF/4CIF	CIF/4CIF	CIF/4CIF
Slices (BL)	1,2,3,5,9	1,2,3,5,9	1,2,3,5,9
Slices (EL)	1,2,3,5,9	1,2,3,5,9	1,2,3,5,9

5.1. The effect of slice partitioning to coding efficiency

Fig. 8 and Fig. 9 show the effect of using slice partitioning on coding efficiency for CIF and 4CIF resolution, respectively. As we can see from Fig.8, the loss in coding efficiency for smaller resolutions like CIF is quite significant, i.e., about 0.5 dB for low data rates around 400 kbps in the case of 5 slices per picture. A loss of almost 1 dB can be observed for 9 slices per picture. Most of this loss results from slice partitioning of the QCIF base layer, as further analysis has revealed Note that enhancement layer as well as base layer are partitioned into slices, i.e., in the case of 5 slices per picture we have 10 slices per access unit, i.e., frame.

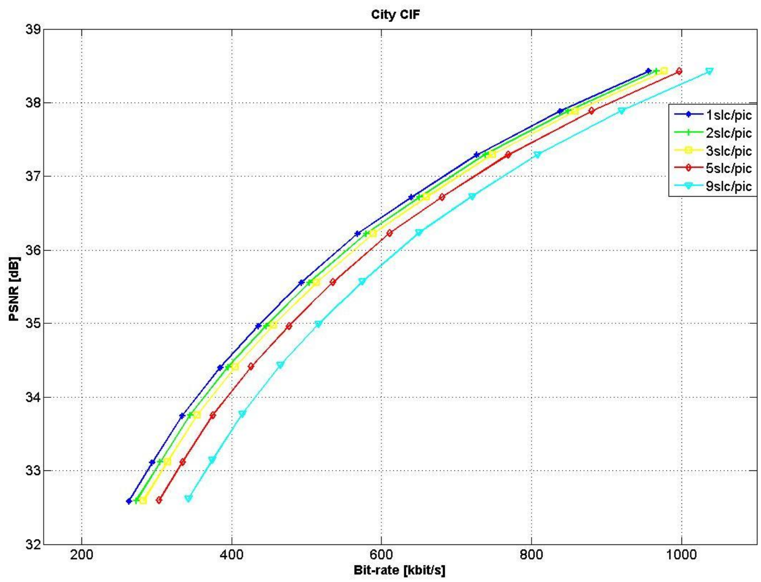


Fig. 8. Effect of slice number on coding efficiency: CIF resolution with QCIF base layer (‘City’ sequence).

Only a small loss can be observed for 4CIF resolution with a CIF base as depicted in Fig. 9. This is due to the effect that the number of macroblocks in a slice is four times higher compared to the previous case, thus making intra prediction and entropy coding within a slice more effective.

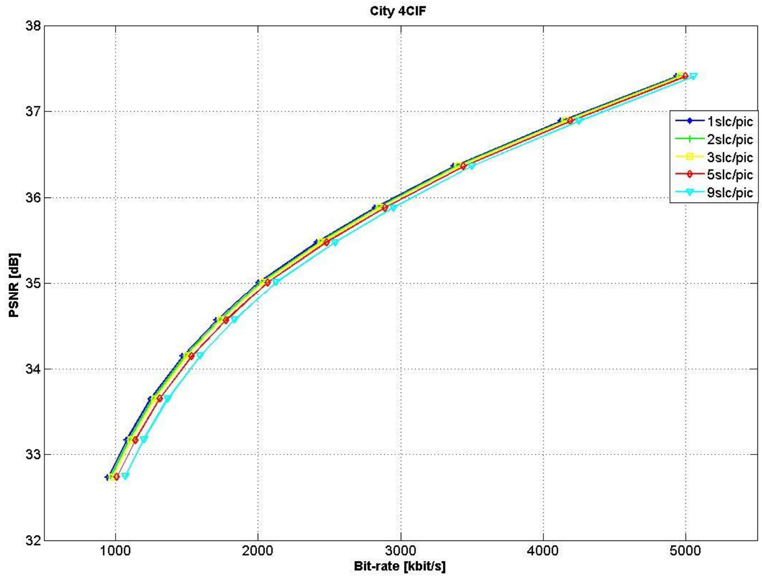


Fig. 9. Effect of slice number on coding efficiency: 4CIF resolution with CIF base layer ('City' sequence).

5.2. The effect of slice partitioning to the efficiency of error concealment

The packet losses for the error concealment simulations were generated with a separate software, which created random packet losses for the base and enhancement layers with ratios 1,2,5 10 and 15. The losses could occur for all types of pictures, excluding the SEI, SPS and PPS packets at the beginning of each sequence. A modified JSVM 9.15 decoder with the implemented error concealment techniques was used in the decoding process. The PSNR values were measured with the program included in the JSVM reference software [11]. The simulations were repeated 50 times for each packet loss ratio, after which the average PSNR values were calculated.

Fig. 10, Fig. 11 and Fig. 12 present the simulation results for the error concealment evaluation. The 'Harbour' sequence has the smallest variation between each slice partition, whereas the 'Soccer' holds the greatest variation, mainly caused by the motion in the video. As can be seen, 1 slice per picture gives the smallest average PSNR values in all the tests. The reason is that the whole picture is then lost and concealment is more difficult.

When the packet loss ratio exceeds 10% the best results are achieved by using 9 slices per picture. This partition seems to be better in a transmission environment where high losses are expected. On the other hand when the loss rate exceeds 10% it is worth thinking whether it is reasonable to transmit the video at all due to the poor quality-of-experience. As can be seen, the 3-slice partition defeats the competitors in all the test cases, when the loss ratio stays under 10%. The difference between 1 slice and 3 slices per picture with 10% packet loss ratio for the ‘Soccer’ sequence is approximately 4 dB, which is significant.

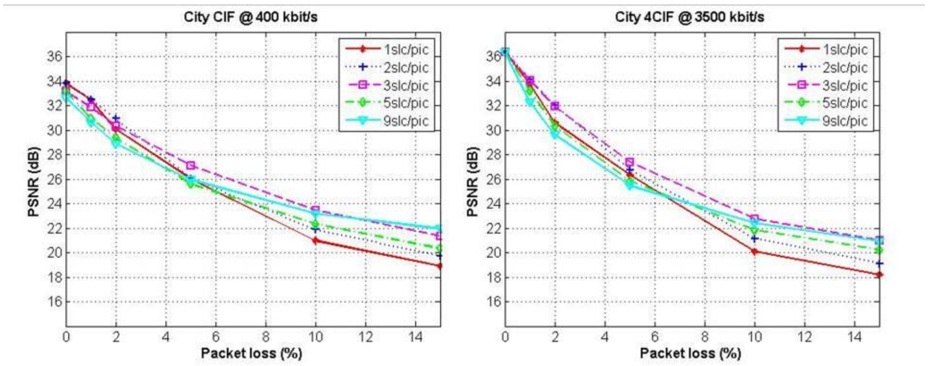


Fig. 10. Simulation results for ‘City’ sequence.

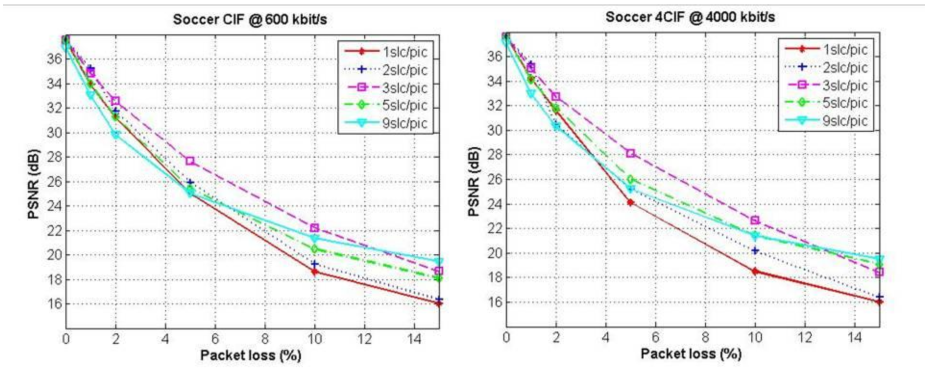


Fig. 11. Simulation results for ‘Soccer’ sequence.

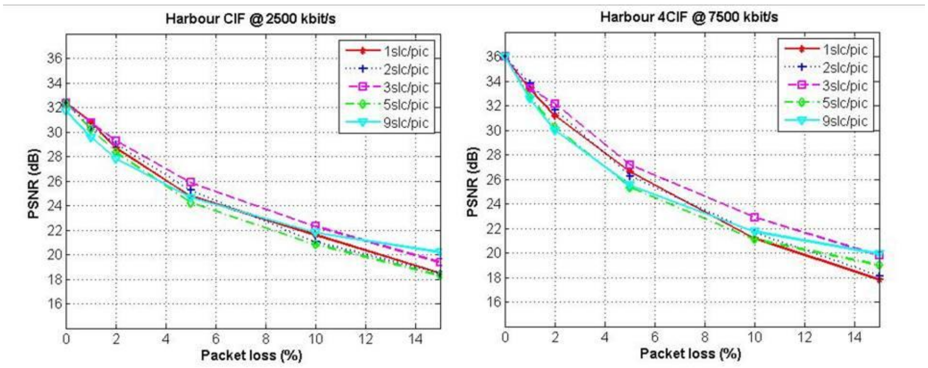


Fig. 12. Simulation results for 'Harbour' sequence.

6. Conclusion

This paper studied the scalable video coding in wireless packet-switched networks. The paper aimed at finding out the best trade-off between slice partitioning, coding efficiency, and error concealment efficiency. The paper also introduced the error resilient coding tools as well as the implemented error concealment tools in H.264/SVC reference codec.

The evaluation was done using three different test sequences. The coding efficiency of H.264/SVC was first evaluated by encoding several sequences with different slice partitions and target bit rates. After this, the error concealment efficiency for each slice partition with different packet loss ratios was evaluated.

Obviously the best coding efficiency was reached with the minimum number of slices. On the other hand, it was shown that error concealment was then harder and it gives poor results. The best tradeoff was achieved with three slices, especially regarding the quality after error concealment.

Acknowledgments

This work was carried out in the ICT-OPTIMIX project (see http://www.ict-optimix.eu), which was partially funded by the European Union in the 7th Framework Program under grant agreement n∘INFSO-ICT-214625. The authors would like to thank for the support.

7. References

1. Stockhammer, T., Hannuksela, M.M., Wiegand, T.: H.264/AVC in Wireless Environments. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, p. 657-673. (2003)
2. ITU-T (2007) Series H: Audiovisual and Multimedia Systems, Recommendation H.264: Advanced video coding for audiovisual services, 564p.
3. Sood, A., Chilamkurti, N.K., Soh, B.: Analysis of Error Resilience in H.264 Video Using Slice Interleaving Technique. Book Chapter, Information Networking. Advances in Data Communications and Wireless Networks, Springer, pp. 912-921 (2006)
4. Kumar, S., Xu, L., Mandal, M.K., Panchanathan, S.: Error Resiliency Schemes in H.264/AVC Standard. In: Elservier J. of Visual Communications & Image Representation, vol.17, pp. 1-26. (2006)
5. Guo, Y., Chen, Y., Wang, Y.-K., Li, H., Hannuksela, M.M., Gabbouj, M.: Error Resilient Coding and Error Concealment in Scalable Video Coding. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, pp. 781-795. (2009)
6. Kung, W.-Y., Kim, C.-S., Jay Kuo, C.-C.: Spatial and Temporal Error Concealment Techniques for Video Transmission over Noisy Channels. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, pp. 789-803. (2006)
7. Dhondt, Y., Lambert, P.: Flexible Macroblock Ordering – an error resilience tool in H.264/AVC. In: Fifth FTW PhD Symposium, 1st December, Faculty of Engineering, Ghent University, Belgium, paper nr. 106, pp. 1-2. (2004)
8. Mys, S., Lambert, P., De Neve, W., Verhoeve, P., Van de Walle, R.: SNR scalability in H.264/AVC Using Data Partitioning. In: Advances in Multimedia Information Processing, vol 4261, pp. 329-338. (2006)
9. Chen, Y., Xie, K., Zhang, F., Pandit, P., Boyce, J.: Frame loss error concealment for SVC. In: Journal of Zhejiang University – Science A 7, pp. 677-683. (2006)
10. Uitto, M., Vehkaperä, J.: Spatial Enhancement Layer Utilisation for SVC in Base Layer Error Concealment. In: Proceedings of the 5th International ICST Mobile Multimedia Communications Conference, pp. 1-8, London, United Kingdom. (2009)
11. SVC Reference Software (JSVM software). [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm