Temporal Layered vs. Multistate Video Coding

Sila Ekmekci and Thomas Sikora

Technical University Berlin, Einsteinufer 17, Berlin, Germany

ABSTRACT

This document shows the desired format and appearance of a manuscript prepared for the Proceedings of the SPIE. It contains general formatting instructions and hints about how to use LaTeX. The LaTeX source file that produced this document, `article.tex` (Version 2.81), provides a template, which can be used in conjunction with `spie.cls`.

Keywords: Layered coding, Multiple description coding, Error-resilient video communications, rate allocation, Automatic Repeat reQuest (ARQ), Packet path diversity, unbalanced transmission

1. INTRODUCTION

In recent years the demand for communication over wireless networks and Internet is highly increased. But there are still unsolved technical problems such as long delays and packet losses causing quality degradation. Delay sensitive and bandwidth demanding applications still call for transmission solutions encouraging research in this area. Layered Video Coding and Multiple Description Video Coding are both used as source coding techniques for transmission over error-prone networks. In both cases multiple bitstreams are generated. Even if some portion of the bitstream is dropped due to channel problems a reconstruction is still possible with the received rest of the bitstream. While in layered coding the reception of the base layer is mandatory for the decoding of the enhancement layer, multiple description coding enables independent decoding of its descriptions. The enhancement layers in layered coding are mainly used to refine the quality of the video and they are useless if the base layer is lost. Since the reception of the base layer is critical it should be protected against transmission errors via Forward Error Correction Codes (FEC) or Automatic Repeat Request (ARQ). On the other hand the descriptions generated by Multiple Description Coding are mutually refining. If the descriptions are balanced in terms of bitrates they are usually also equally important in terms of quality. Multiple State Video Coding is a special Multiple Description Scheme where the video frames are splitted into threads constituted of even and odd frames. The prediction is performed only between adjacent odd and even frames in each thread. The advantage is the improved state recovery property: If in one of the threads one frame (corresponds to one packet) is lost it is reconstructed by interpolation of its previous and next frames from the other thread. If both channels are error-free the signal can be reconstructed with maximal quality whereas in case of losses a lower but still acceptable quality is enabled. There are several papers reporting comparison results of specific layered and multiple description schemes for special transmission scenarios. A summary of results is given in [], showing that MDC has advantages over LC for networks where feedback is not available due to lack of back channel or due to long round trip times or low delay constraints as for streaming video applications. The results are different when the source coding techniques are combined with FEC or ARQ.

In this work, we will compare temporal layered coding to Multiple State Coding at different packet loss rates. We will assume that one packet contains one frame. Additionally we will assume that the motion vectors are always received. H26L codec (version 9.0) is modified to implement the Multiple State Video Coding. For layered coding the same codec with B frames as enhancement layer are used. According to this the temporal distance between adjacent P-frames in both techniques is the same. The overall frame rate of the reconstructed video is twice the rate of P frames in each thread. In layered coding the rate is doubled due to the temporal enhancement layer (B frames) and in multistate coding due to interleaving of the reconstructed frames after reception and in case of losses after state recovery using interpolations. To allow a fair comparison the total bitrate R is set

Further author information: (Send correspondence to Sila Ekmekci.)
S. Ekmekci: E-mail: ekmekci@nue.tu-berlin.de, Telephone: ++49 30 314-23835

equal in both cases. ARQ is only used for the layered coding as a coding option only if the base layer is lost. Which coding options are allowed for layered and multistate coding is depicted in section x. The resulting coding schemes are compared with each other in terms of average PSNR and average PSNR difference (PSNR difference between adjacent frames). While trying to maximize average PSNR we seek to minimize the average PSNR difference (flickers), since flickers above 0.5 dB are reported as visually disturbing. The performance according to our goal is also dependent on the allocation of the total bitrate R among different sources. For multistate coding R is to be divided between the coding of two threads. For layered coding on the other hand the allocation is performed by setting the quantization step sizes for P and B frames as well as the intra update rates for P frames. The rates of the base and the enhancement layers should add up to the total bitrate R. If ARQ is also allowed for the base layer the additional bitrate resulting from the multiple transmissions of the base layer packets are also considered in calculating the total bitrate. For multistate coding we will assume using two different paths for two threads where each path has a different loss probability.

To sum up, the goal of the work is to compare temporal layered coding with temporal multistate coding at different coding scenarios resulting from different quantization step sizes for different layers or descriptions, usage of ARQ, as well as allowed group of blocks (GOB) or frame updates.

Section II gives the system overview including the general setup for layered and multistate systems. Section III on the other hand presents experimental results showing the performance comparison at different channel conditions, i.e. loss probabilities over one or two channels used. Future work is introduced in Section III and section IV concludes the paper.

Ref. 1, for example.

2. PARTS OF MANUSCRIPT

Subsection a discusses the Multi-state system whereas subsection b the temporal layered system used for the experiments.

2.1. Unbalanced Multi State Video Transmission using Path Diversity

Multiple Description Coding is a forward error correction scheme where two or more descriptions of the source are sent to the receiver over the same or different channels. If only one channel is received the signal can be reconstructed with distortion D_1 or D_2. On the other hand, if both channels are received the combined information is used to achieve a lower distortion D_0. Multiple Description (MD) Coding was already applied to some major coding techniques such as scalar quantization, vector quantization, correlating transforms or quantized frame expansions. Proposed system designs are given in reference [x]. One of the advantages of multiple description codes is their low delay property which makes them attractive for streaming video which requires low latency. Among several MD coding techniques proposed for video .... Multiple State Video Coding (MSVC) is chosen to be used for our comparisons. The main advantage of MSVC is its low complexity.

MSVC is presented in [] and it is composed of two subsystems: multiple state encoding/decoding and a path diversity transmission system. The video is coded into multiple independently decodable streams each with its own prediction process and state. According to this, if one of the streams is lost the other stream is still decodable. Moreover the correctly received stream can enable state recovery for the corrupted using using bidirectional information past and future frames. Figure 1 gives the block diagram of the system and Figure 2 shows the state recovery if a packet is lost. The path diversity system enables the the submission of different subsets of packets over different paths. Since the probability of simultaneous losses on two paths is smaller than the single path loss probabilities an increase in the system performance is expected. But the prediction gain as compared to single description coding of the same sequence decreases due to the increasing distance of adjacent frames coded on the same thread. State recovery ability on the other hand increases the error resilience. In other words high coding gain and error resilience are traded off with each other.

In this work, we concentrate on a two description system. The video is split into the sequences of even and odd numbered frames where both sequences are coded independently. If both streams are received without errors the decoding is performed independently. The final stream is then generated by interleaving the two decoded streams. In case of losses the lost frames are generated by interpolation of past and future frames from the other stream. The balanced operation of MSVC is introduced and discussed in x and y. Unbalanced operation based on frame rate adaptation is given in x. Here the signal is coded into two streams yielding unbalanced frame rates of 2:1. The motivation is network paths with unbalanced bandwidths forcing unbalanced bitrates for transmission. Unbalanced MD video coding based on quantization adaptation is introduced in x. Here Multiple State Video Coding is modified to recover the streams not only from state losses (in case of frame losses) but also from error propagation and more importantly to recover the low quality stream (coarsely quantized) from large quantization errors.

In x two approaches are discussed which are denoted as Approach 1 and Approach 2. In this paper we will consider only the first approach. Surely the system can further be improved by incorporating intra-updates of frames or GOBs which will be investigated in future work.

2.2. Temporal Layered System

H26L codec (version 9) is used to code the video sequence generating a base layer as well an temporal enhancement layer. The temporal distance between frames on the base layer is the same as the temporal distance between frames of each thread of the multi-state video coding. There are mainly four configurations we investigated:

a) Base Layer + Enhancement Layer

The bitrate is allocated between the base layer and the enhancement layer. There no intra-updates of frames or GOB's are allowed. The coding parameters and the total bitrate achieved is given in table x.

b) Base Layer with Frame Updates + Enhancement Layer

An additional coding parameter is the frame update period for the base layer. To keep the bitrate constant, the quantization step sizes of the P and B frames are increased increasing the quantization noise. But through the I frames added periodically the error resilience is increased. If the packet containing the I frame is lost, I frames does not help anymore. The coding parameters are presented in table x.

c) Base Layer with Intra Updates of Group of Blocks + Enhancement Layer

There are 11 GOBs in a QCIF frame. A GOB is a line of macroblocks in a frame where a macroblock is constituted of 16x16 pixels. The additional coding parameter here is the intra update period of GOBs. Again the quantization step sizes of the base and the enhancement layers are adjusted to hold the constant total bitrate. The coding parameters are presented in table x.

d) Base Layer with Frame Updates + ARQ + Enhancement Layer

This is an extension of the second option. If a packet belonging to the base layer gets lost, it is requested at most three times. The additional bitrate needed for the repeated transmissions are also considered in setting the quantization stepsizes of the I, P and B frames. Since the reception of the base layer is crucial for the decoding of the enhancement layer ARQ is expected to increase the general performance. The coding parameters are given in table x.

The total bitrate is equal to the total bitrate of the MSVC system to allow a fair comparison.

2.3. Experimental Results

We coded 200 frames at the overall frame rate of 30 fps from two QCIF sequences Foreman and Akiyo at 140 kbit/s and 19 kbit/s using three techniques: 1- Temporal Layered Coding (LC), 2-Multi-State Video Coding, Approach 1 (MSVC 1) and 3- Multi-State Video Coding, Approach 2 (MSVC 2). We targeted the same bitrate for all coding techniques to allow a fair comparison. We required higher bitrates for the Foreman sequence because of the high motion activity whereas Akiyo requires less coding bitrate because it is a slow motion sequence. If coded in single layer at these bitrates, Foreman sequence has PSNR of 35.9 whereas Akiyo has PSNR of 37.6. We assumed that a packet contains a single and complete frame regardless of the coding method used.

In the next stage the packets are transmitted over a lossy network with a given and constant loss rates. We considered here the loss rates of 0, 5 , 10 and 20 percent. We assume further that the motion vector information is never lost. This assumption has an important effect on the comparison results which will be discussed in section x. In LC, If a packet is lost, it is reconstructed using motion compensated concealment. In MSVC 1 and

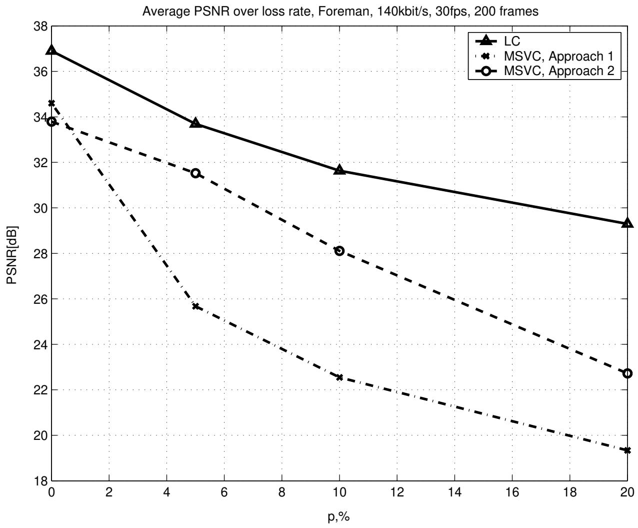

Figure 1. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames

MSVC 2 however, the lost frame is reconstructed using interpolation of the two neighbouring frames from the other stream. MSVC 2 makes use of interpolation based reconstruction even if the current packet received if the achieved frame PSNR is higher this way. Although there is no way to make this objective comparison at the receiver, the decision whether to use the received packet or not, could be performed subjectively.

Fig. 1 shows the average PSNR over the packet loss rate for the three techniques. MSVC 2 outperforms MSVC 1 at atmost all loss rates. The average PSNR difference is about 5 dB at 5 and 10% loss rates. LC on the other hand achieves the best performance. As the loss rate increases the gap between LC and the other two methods increases further and reaches 6 dB at 20% loss rate between LC and MSVC 2. Approach 1 outperforms Approach 2 only at the lossless case. Although the interpolation could yield a higher PSNR for the current frame, the error propogates to the next frame and the overall average reconstructed frame PSNR decreases.

Fig. 2 shows the result of the similar comparison for the slow motion Akiyo sequence. As for the Foreman sequence the MSVC coding method decreases the average coding PSNR at the same bitrate. As explained in section x, as the temporal distance between adjacent frames increases, the coding gain decreases. The initial PSNR difference between the two coding methods is about 2.5 dB and this difference increases to 7 dB as the loss rate increases. The difference between the two approaches of MSVC is at most 2 dB, smaller than in case of Foreman. The frame splitting costs much higher for Akiyo than for Foreman because of the extra intra-coded frame of the second thread. Since the only 19 kbit/s available the mode decision between intra and inter coding of just one frame among 200 frames makes a big difference. The difference between MSVC Approaches on the other hand are smaller because Approach 2 helps only if after lost frame and an concealment error to the next frame propagates. For the Akiyo however the reconstruction using interpolated concealment works well. As loss rate increases both of the threads are affected, and since there are no intra updates in the coding, the error accumulates on both of the threads. Here instead of a bilinear interpolation using the two neighbouring frames a motion based concealment would yield better results. This effect is also seen in the LC case. As told before this is only the case if the exact motion vector are received. If this is not the case, we have to use the motion vectors from the neighbouring frames and the resulting reconstruction will not be that successful. If no motion vectors are available the lost frame can also be reconstructed by frame repetition. MSVCs strength is its reconstruction cabability even if the motion vectors of the corresponding frames are also lost. Even in this case the motion vectors can be predicted using the former and next frames from other thread. But in the layered case if one frame from the enhancement layer is lost, it can be reconstructed using the same method as in the MSVC case. But if the base layer is lost the reconstruction using the estimated motion vectors from the neighbouring frames is much more difficult. This case is not investigated in this paper and we assume also in the sequel the lossless

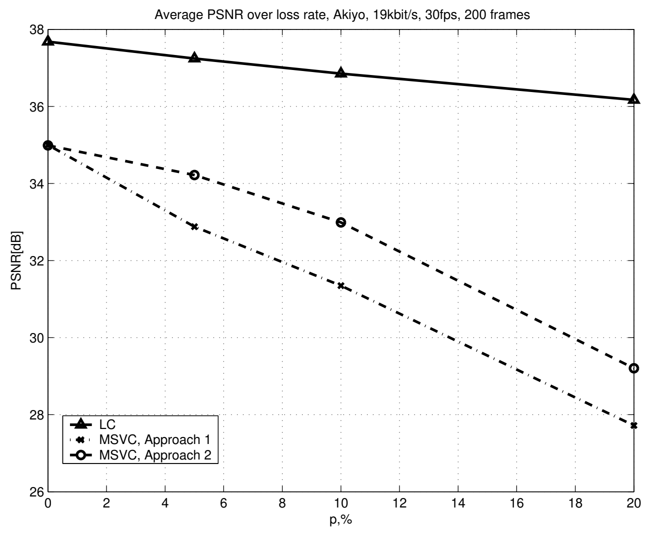

Figure 2. Average PSNR over loss rate, Akiyo, 19 kbit/s, 30 fps, 200 frames

reception of the motion vectors.

The next experiment shows the case intra update of a frame in each thread is allowed once in a second. There are two intra frame updates for MSVC scheme, one for each thread. The bitrate is still kept constant for the two sequences. The average PSNR of the coded sequence is about the same as for the intra frame update option for the Foreman sequence. For the low motion sequence Akiyo however, the penalty of using intra updates is high, because of the low total bitrate available the average coded sequence PSNR decreases around 3 dB. Figure 3 and Figure 4 show the average PSNR of the reconstructed sequence over the loss rate. The common observation for both cases is that the drop of PSNR is much smaller as the loss rate increases when using intra frame updates because of the state refreshment every second. If an intra coded frame is lost, it is concealed like a P Frame (using motion compensated interpolation) in MSVC. In LC however we assumed that the motion vectors are sent additionally for intra coded frames to allow motion compensated concealment. It is important to code that the additional bitrate for coding the extra motion vectors for I frames are not included in the total bitrate.

Moreover comparing Figures 3 and 4, we see that the overall drop in PSNR as the loss rate increases is smaller in Akiyo compared to Foreman. The reason is the better reconstruction results due to slow motion of the sequence.

The third experiment involves the usage of periodic GOB (Group of Block) updates instead of the intra frame updates. A GOB consist of a horizontal line of macroblocks where there are 9 lines of macriblocks in a QCIF frame. We coded 9 lines of macroblocks in a second so that we had a complete frame update every second if there were no losses and also no spatial error propagation in this time interval. The average PSNR drop in encoded sequences are the same as in the case of frame updates when the total bitrate is kept constant as before. The same is applied to LC method. What we observe in Figures 5 and 6 is that as the loss rate increases the performance goes much faster down compared to the case with frame updates. This effect is mainly due to the spatial error propagation between GOBs. Another observation is that the gap between Approach 1 and Approach 2 of MSVC increases and Approach 2 yields better results than LC in the middle range loss rates. This is due to the statistically independent losses and independent intra coded GOBs on the two streams. Moreover the gap between the two MSVC approaches is larger because the probability that the state of one stream at a given point is better than the other one, is higher as compared to complete update case.

Similar observations can also be done for the Akiyo sequence with the difference that MSVC method never reaches the performance of LC. This could be generally changed if we allow the two threads to share the same intra frame as reference at the beginning which is not explored in this work.

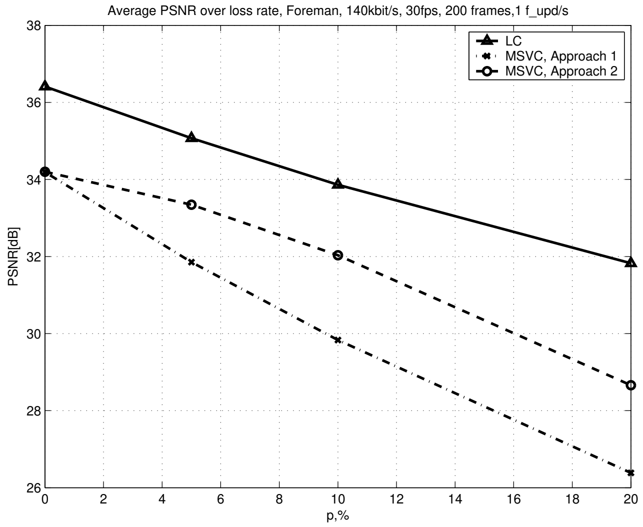

Figure 3. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

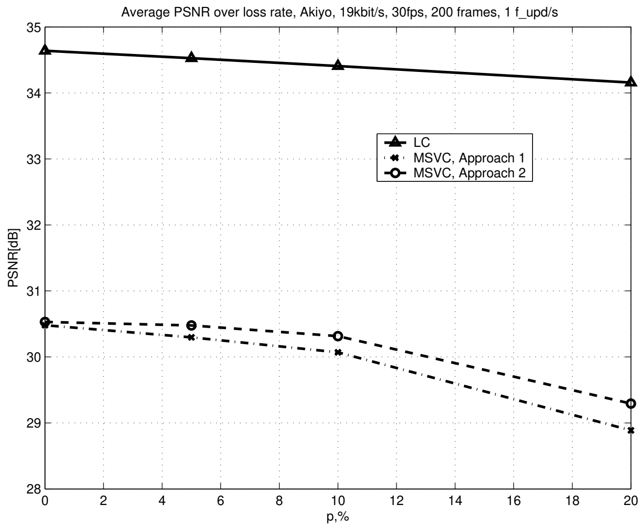

Figure 4. Average PSNR over loss rate, Akiyo, 19 kbit/s, 30 fps, 200 frames, 1f_upd/s

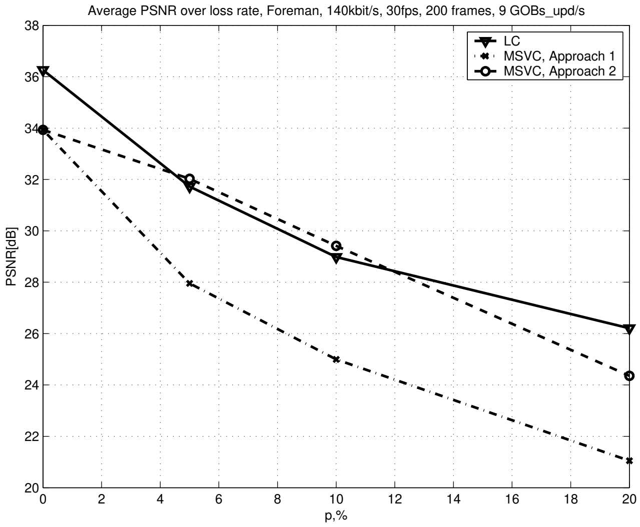

Figure 5. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

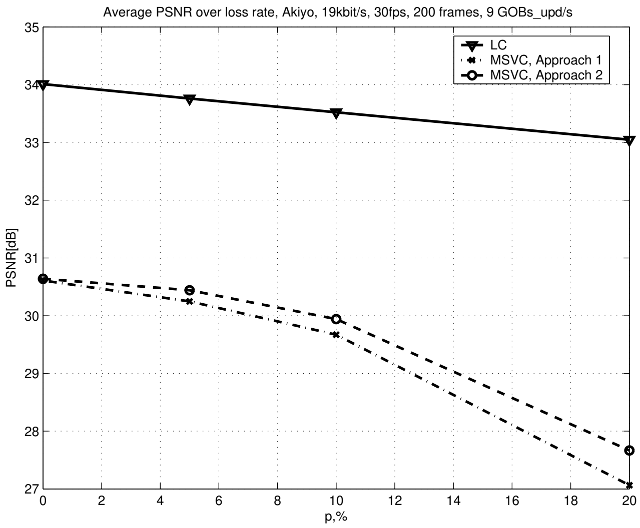

Figure 6. Average PSNR over loss rate, Akiyo, 19 kbit/s, 30 fps, 200 frames, 1f_upd/s

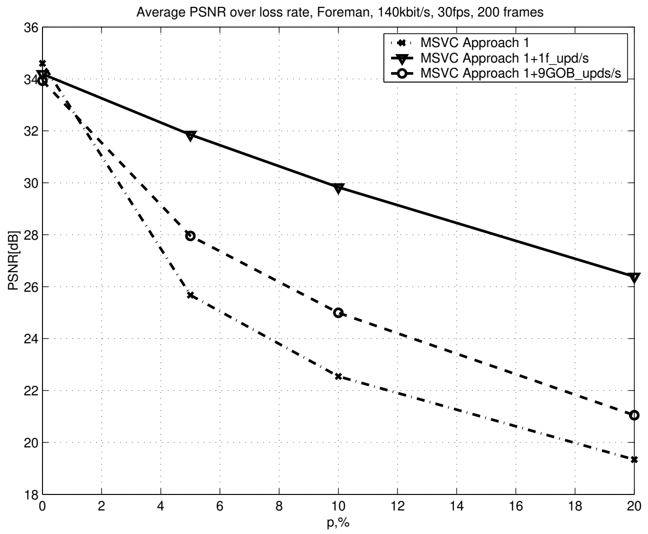

Figure 7. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

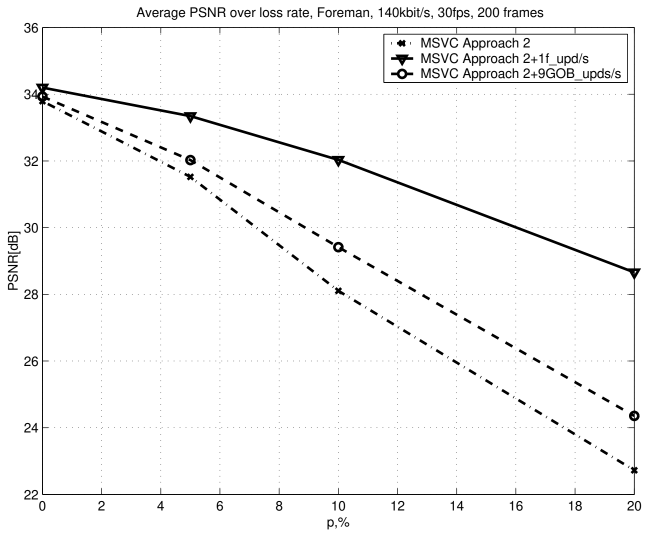

Figure 8. Average PSNR over loss rate, Akiyo, 19 kbit/s, 30 fps, 200 frames, 1f_upd/s

Figures 7 and 8 compares the three coding options (without intra updates, intra update of frames and intra updates of GOBs) for the Foreman sequence. As discussed above intra updates of frames yields the best results followed by the intra updates of GOBs. The gap between different methods for Approach 1 is bigger than for Approach 2 since Approach 2 involves more interaction between the two independently streams.

Figures 9 and 10 on the other hand, shows the comparison of the three coding modi for two Approaches. The important observation here is that coding without any intra updates yields better results, especially at lower bitrates. Since the total bitrate chosen for Akiyo is small, allocating bitrate for intra updates is ineffective. Only for Approach 1 and as the loss rate increases over 10% the modi using intra frame updates performs better.

Figures 11 and 12 depicts the comparison results for the Layered Coding at 4 different coding modi: 1- Layered Coding without any intra updates, 2- with frame intra updates, 3- with GOB intra updates, 4- frame intra updates and additionally at most 3 transmission repetitions in case of losses of base layer packets for

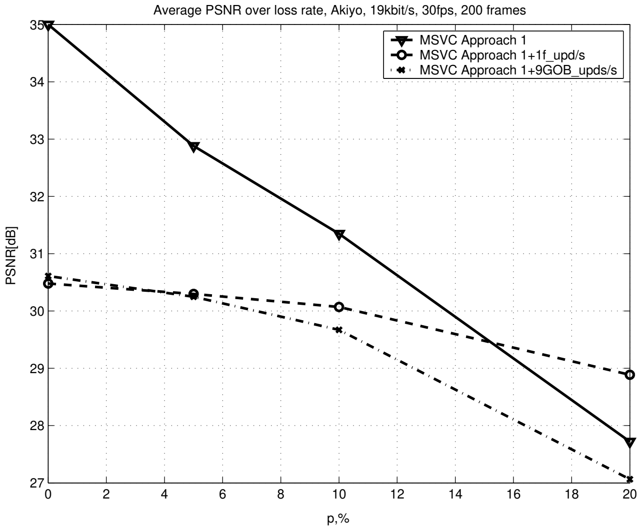

Figure 9. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

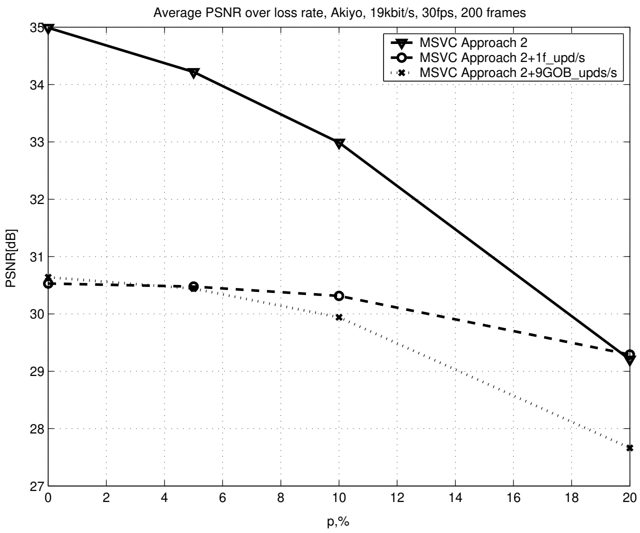

Figure 10. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

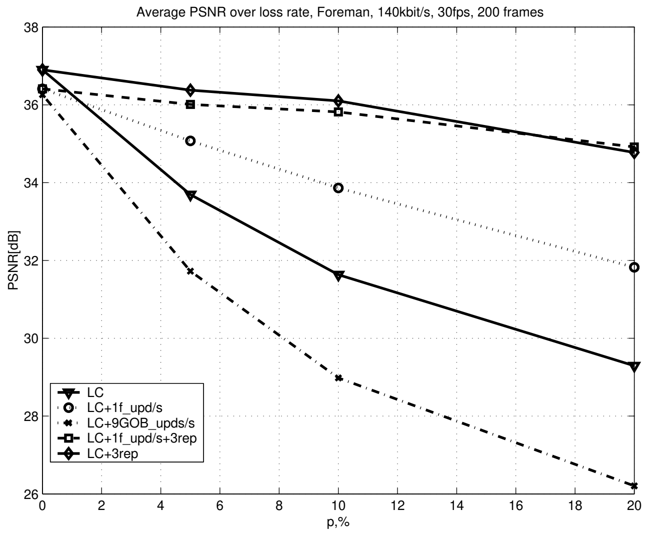

Figure 11. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

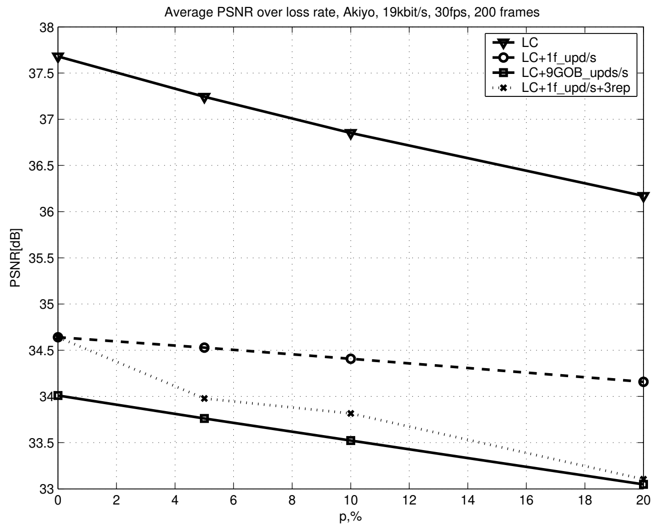

Figure 12. Average PSNR over loss rate, Foreman, 140 kbit/s, 30 fps, 200 frames, 1f_upd/s

sequences Foreman and Akiyo respectively. The additional bitrate needed for retransmissions are also accounted for in the total bitrate depending on the corresponding loss rate of the base layer packets.

2.4. Conclusion

REFERENCES

1. A. Eisenberg, *Guide to Technical Editing*, Oxford University, New York, 1992.