MULTI-STATE VIDEO CODING WITH SIDE INFORMATION

Sila Ekmekci Flierl, Thomas Sikora

Technical University Berlin
Institute for Telecommunications
D-10587 Berlin / Germany

ABSTRACT

Multi-State Video Coding (MSVC) is a multiple description scheme where the video is splitted into two or more subsequences. Each subsequence is encoded and transmitted separately and can be decoded independently. The prediction gain decreases due to sequence splitting but error resilience of the system increases since reconstruction capabilities improve. The lost frames in one subsequence are reconstructed by using state recovery, i.e., interpolation of the past and/ future frames from the other subsequence. Unbalanced Quantized MSVC is realized by using the same scheme but coding the subsequences with different quantization stepsizes yielding different bitrates. The advantage of unbalanced operation is the increased system performance in case of unbalanced transmission channel characteristics. In our previous work, we proposed an advanced reconstruction algorithm to support the unbalanced coding of the subsequences: State recovery is not only used for the lost frames but also for received frames when state recovery yields a higher frame PSNR than using the received packet and applying motion compensation. But to figure out which reconstruction method gives a higher frame PSNR a comparison with the original sequence is necessary. Therefore the method is applicable at the decoder only if a feedback mechanism between the encoder and decoder is present. In this work, we present an alternative way, MSVC with side information (MSVCSI), for guiding the optimized reconstruction stategy by estimating the reliabilies of several possible reconstruction alternatives. The reliabilty values are calculated recursively for each frame using the loss history of the frames and the side information representing the specific sequence characteristics. We show that under unbalanced transmission conditions, MSVCSI outperforms the original MSVC method (Approach 1) and the advanced MSVC (Approach 2) upto 1 dB depending on the loss rates of the transmission channels. The gain increases as the loss rates and the unbalance rate increase.

keywords: multiple description coding, optimal rate allocation, unbalanced quantization, path diversity.

1. INTRODUCTION

Multimedia communication over Internet has conflicting requirements on high compression and high error resilience. Multiple Description Coding (MDC) is an error resilient source coding method, where two or more descriptions of the source are sent to the receiver over different channels [1]. If only one description i is received, the signal is reconstructed with distortion D_i. If all descriptions are available, we achieve a lower distortion D_0. Multi-State Video Coding (MSVC) is a special multiple description scheme where the video sequence is splitted into the subsequences of even and odd numbered frames [2]. A MSVC system has two main components: multiple state encoding/decoding (Figure 1) and a path diversity transmission system. The generated subsequences are coded into multiple independently decodable streams each with its own prediction process and state. The advantages are that the streams are independently decodable and that the correctly received stream can enable state recovery for the corrupted stream using bidirectional information from past and future frames.

With increasing heterogeneity in network infrastructures, it becomes interesting to build descriptions with different coding rates adaptable to the streaming conditions. Unfortunately, unbalanced multiple description video coding has not been widely explored. Unbalanced descriptions can be generated based on adaptation of the quantization, temporal [3] or the spatial resolution of the frame-wise splitted video signal. In [4], we investigated Unbalanced Quantized Multi-State Video Coding where the subsequences are quantized with different quantization stepsizes yielding different bitrates. We also proposed to use the state recovery property, not only to recover from errors [2] but also to substitute the coarsely quantized frames by interpolation of the received past and future frames whenever it is possible to achieve a higher frame PSNR [4], [5]. In the sequel, the original MSVC scheme will be referred to as Approach 1 and the MSVC with extended state recovery as Approach 2.

In this work, we investigate Multi-State Video Coding with Side Information (MSVCSI) where encoder transmits off-line data calculated using the video sequence to the de-

coder. The side information reflects the PSNR change due to applying a specific reconstruction method instead of the motion compensation. Therefore it is highly related to the scene activity of the sequence. The decoder calculates a reliability value for each frame and each reconstruction option and then chooses the reconstruction option yielding the highest frame PSNR for this frame. We use also a larger set of reconstruction methods to choose from depending on whether the adjacent past and future frames from the same and other thread are already coded or not. In addition to the interpolation of the previous and the next frame from the other thread, we consider also copying the previous frame (from the same or other thread) as well as the next frame from the other thread as possible reconstruction methods. We compare MSVCSI to Approach 1 and Approach 2 of the MSVC scheme at the same loss rate of the transmission channels. We assume that the unbalanced quantized streams are sent over channels with different loss rates and measure the average reconstructed frame PSNR for each loss rate combination.

The algorithm for MSVCSI is given in Section 2. Section 3 describes the experiments and presents the experimental results. Conclusions can be found in Section 4.

2. MSVC WITH SIDE INFORMATION

In MSVCSI, as in Approach 2, we use state recovery not only in case of losses but also when the packet is received but we can achieve a better reconstruction by using other reconstruction options through the past and future frames from the other stream. In Approach 2, reconstruction options available for a received frame are: 1- using the data from the received packet and applying motion compensated interpolation, 2-motion compensated interpolation using the previous and next adjacent frames from the other stream. The optimal reconstruction method is chosen by comparing the frame PSNR's achieved by both methods. But this comparison can be only directly performed at the encoder since the original sequence is required. Therefore Approach 2 is to be used if there is a feedback mechanism between the decoder and encoder, sending information about the current status, e.g. loss history. The results presented in section 3 are calculated as if the decoder can perform the PSNR comparison exactly on the frame base. The reconstruction options for a received frame in MSVCSI that are different than in Approach 2 are: 3-copying the previous frame from the same thread, 4-copying the previous frame from the other thread, 5-copying the next frame from the other thread. For the lost frames the same reconstructions options are available except the first one. Another difference from Approach 2 is that the optimal reconstruction method is chosen by the calculation and comparison of reliability terms associated with each reconstruction option. The reliability terms on the other hand are dependent on the following:

1. loss history of the frames

2. $PSNR_q$: PSNR of each frame due to quantization

3. $dPSNR_{interp}$: PSNR change of each frame due to interpolation

4. $dPSNR_{left}$: PSNR change of each frame due to copying of the previous frame from the other thread

5. $dPSNR_{left_2}$: PSNR change of each frame due to copying of the previous frame from the same thread

6. $dPSNR_{right}$: PSNR change of each frame due to copying of the next frame from the other thread.

Whereas the loss history is available at the decoder the other five terms are to be calculated offline at the encoder and transmitted to the decoder as side information. The side information represent in this case characteristics of the video sequence such as the variation of the quantization distortion on the frame base, and the PSNR change when a specific reconstruction method is applied instead of motion compensation using the motion compensated frame difference. Our goal is to calculate a reliability term recursively for each frame to guide the overall optimal reconstruction stategy. The reconstruction method is chosen on the frame base to maximize the reliability, i.e. reconstructed frame PSNR. Section 2.1 gives the calculation of the reliability terms and Section 2.2 the system setup.

2.1. Calculation of Reliability Values

Reconstruction by using the packet received:

$$
\begin{aligned}
R(1) &= PSNR_q(1) \\
R(2) &= PSNR_q(2) \\
R(n) &= R(n-2)
\end{aligned}
$$

where $R(n)$ is the reliability value for frame n. Motion compensation can be applied if the packet containing the motion compensated frame difference is received. The reliability of the first frame in each thread is set to $PSNR_q$; the PSNR value due to quantization only.

Reconstruction by interpolating the previous and next frame from the other thread:

$$
R(n) = \frac{R(n-1) + R(n+1)}{2} - dPSNR_{interp}
$$

where n is the index of the current frame and $n-1$ and $n+1$ the indices of the previous and next frames.

Reconstruction by copying the previous frame from the same thread

$$R(n) = R(n-2) - dPSNR_{left2}$$

Reconstruction by copying the previous frame from the other thread

$$R(n) = R(n-1) - dPSNR_{left}$$

Reconstruction by copying the next frame from the other thread

$$R(n) = R(n+1) - dPSNR_{right}$$

The first frame of the first thread must be reconstructed by motion compensation and its reliability is therefore $R_1 = PSNR_q(1)$. For the remaining frames all other reconstruction options are available if for the corresponding reconstruction necessary frames are already reconstructed. For example, the previous and the next frames from the other thread should already habe been reconstructed to perform the interpolation based reconstruction of the current frame.

2.2. System Setup

We modified the H.264 codec (version 9.0) to support the MSVC structure. Two parallel decoders are implemented which help each other to recover from losses as explained in Section 2. The optimal reconstruction method depends on both the loss history and the scene activity. Side information as listed in 2.1 is sent by the encoder to help choosing the best reconstruction strategy. We assume that each frame (I or P) is transmitted in a single packet [2]. Moreover we assume that the very first frame in each sequence is never lost (e.g. retransmission). If the packet is lost (I or P), all information is lost for the corresponding frame including the motion vectors for P frames. The reconstruction methods for the lost I and P frames are the same. The block diagram of the MSVC system is given in Figure 1 and the frame interpolation in Figure 2.

3. EXPERIMENTS

The H.264 codec (TML, version 9.0) is used to implement and test the MSVC algorithm, Approach 1, Approach 2 and
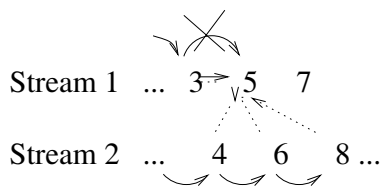

Fig. 2. Frame Interpolation in MSVC.

MSVCSI. Video sequences are in QCIF format and coded into two streams consisting of even and odd streams respectively, each at 15 fps. The state recovery is performed through motion controlled interpolation. Each frame of any type (I or P) is transmitted in a single packet. The packets in each thread are lost with given loss rates of p_1 and p_2 respectively. Lossy channel is simulated with a random loss generator and 100 different loss patterns are used for each loss rate. 200 frames from each sequence are used for the experiments. Due to lack of space the results are presented only for the sequence Foreman. The total bitrate is 188.77 kbit/s and the rates of the subsequences of odd and even frames are $R_1 = 142.77kbit/s$ and $R_2 = 46kbit/s$ corresponding to $PSNR_{avg,1} = 37.84dB$ and $PSNR_{avg,2} = 32.84dB$ at lossless reception. We compared MSVCSI to Approach 1 (original MSVC) and to Approach 2 (extended state recovery) in terms of the average reconstructed frame PSNR at different channel loss rates. We differentiate between three cases: 1- the first stream is lossless and the loss rate of the second channel varies. 2-both of the channels are lossy, and the loss rates are equal to each other. 3-both of the channels are lossy but the loss rates are different from each other. The experimental results for these three cases are given in Figures 3, 4 and 5 respectively. In Figure 3 we see that the MSVCSI outperforms Approach 1 by about 1dB and gives almost the same performance as Approach 2. If the two channels have balanced loss rates MSVCSI outperforms both Approach 1 and Approach 2, by about 0.3 dB and 0.6 dB respectively when the loss rates increase beyond 10%. The advantage of MSVCSI is especially visible when the loss rates are unbalanced. In Figure 5, on the x axis, $p_1 + p_2 = 15\%$ corresponds to $p_1 = 5, p_2 = 10\%$, $p_1 + p_2 = 25\%$ to $p_1 = 10, p_2 = 15\%$ and $p_1 + p_2$ to $p_1 = 10\%, p_2 = 20\%$ respectively. We observe that the advantage increases as the loss rates increase. When $p_1 = 10\%, p_2 = 20\%$, difference in $PSNR_{avg}$ between MSVCSI and Approach 2 is about $1dB$. The difference to Approach 1 is at least $1dB$ at all loss rates.

4. CONCLUSIONS

Unbalanced Descriptions are particularly interesting for video streaming applications over heterogeneous networks where transmission channels have varying transmission characteristics such as loss rate and bandwidth. By using flexible and adaptive rate allocation over available transmission paths the reconstructed signal quality at the receiver can be improved. In this work, unbalanced descriptions of the video signal are generated using the Multi-State Video Coding technique where the video sequence is divided into the subsequences of odd and even frames which coded independently. The subsequences are quantized with different step sizes yielding different bitrates adaptive to the sequence as
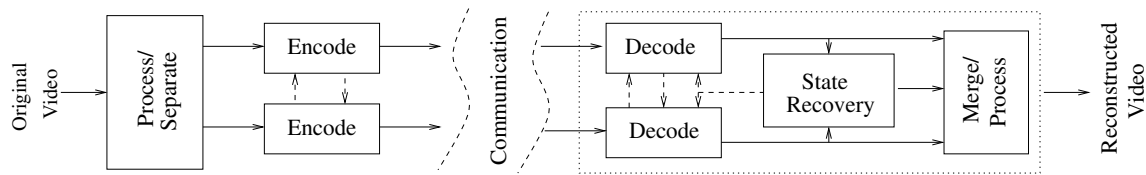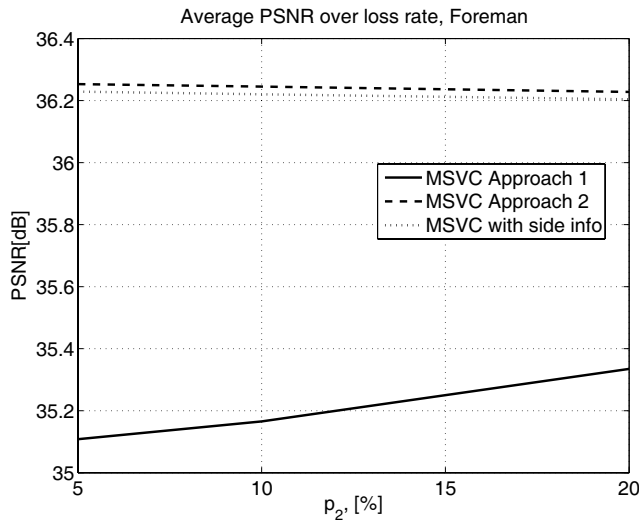
Fig. 1. Block Diagram of the MSVC System.

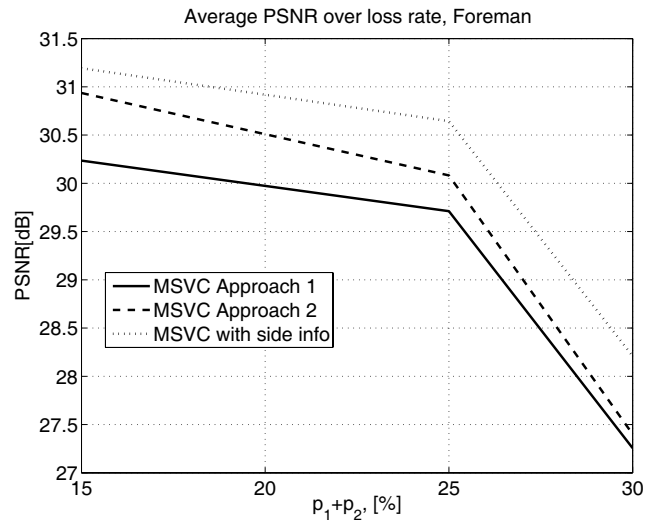**Fig. 3**. $PSNR_{avg}$ over p_2, $p_1 = 0$; Foreman.

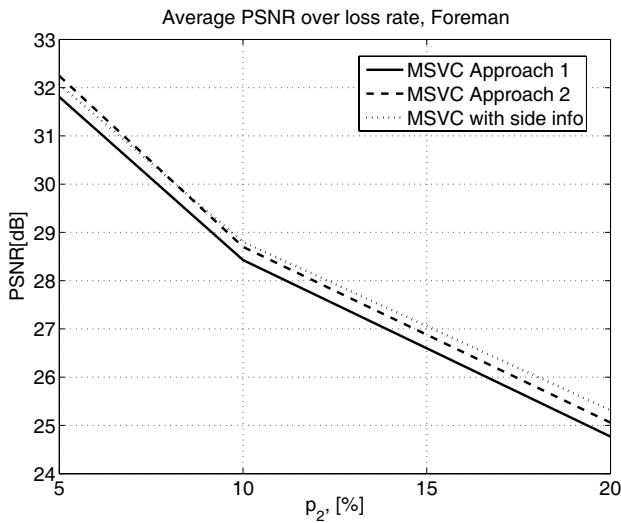**Fig. 5**. $PSNR_{avg}$ over $p_1 + p_2$, $p_1 \neq p_2$; Foreman.

Fig. 4. $PSNR_{avg}$ over p_2, $p_1 = p_2$; Foreman.

well as to the loss rates of the transmission channels. In this paper, we investigated the Multi-State Video Coding with Side Information, where side information reflecting the characteristics of the sequence calculated at the encoder are sent to the decoder to guide the optimal frame by frame reconstruction strategy. For each frame a number of different reconstruction methods are available depending on whether its corresponding packet is received and the adjacent frames on the same and the other thread are received or not. The side information reflects mainly the scene activity and gives a measure about how the frame PSNR is affected by using a specific reconstruction method with respect to motion compensation. Using the side information, reliability terms are calculated recursively for each frame and for each reconstruction option. For each frame the reconstruction option with the highest reliability is applied. We presented experimental results showing that Multi-State Video Coding with Side Information outperforms the original Multi-State Video Coding by upto 1dB depending on the loss rates of the transmission channels. The gain increases as the loss rates and the unbalance in loss rates increase.

5. REFERENCES

[1] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag., vol. 18, no. 5, pp. 74-93*, Sept. 2001.

[2] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *VCIP*, January 2001.

[3] J. Apostolopoulos and S. Wee, "Unbalanced multiple description video communication using path diversity," *ICIP*, October 2001.

[4] S. Ekmekci and T. Sikora, "Unbalanced quantized multiple description video transmission using path diversity," *Electronic Imaging, 2003, SPIE*, January 2003.

[5] ——, "Unbalanced quantized multi-state video coding: Potentials," *Picture Coding Symposium (PCS'04)*, December 2004.