

# Unbalanced Quantized Multiple State Video Coding

Sila Ekmekci Flierl,<sup>1</sup> Thomas Sikora,<sup>2</sup> and Pascal Frossard<sup>1</sup>

<sup>1</sup>École Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Institute, 1015 Lausanne, Switzerland

<sup>2</sup>Institute for Telecommunications, Technical University Berlin, 10587 Berlin, Germany

Received 18 March 2005; Revised 24 August 2005; Accepted 12 September 2005

Multiple state video coding (MSVC) is a multiple description scheme based on frame-wise splitting of the video sequence into two or more subsequences. Each subsequence is encoded separately to generate descriptions which can be decoded independently. Due to subsequence splitting, the prediction gain decreases. But since reconstruction capabilities improve, error resilience of the system increases. Our focus is on multiple state video coding with unbalanced quantized descriptions, which is particularly interesting for video streaming applications over heterogeneous networks where path diversity is used and transmission channels have varying transmission characteristics. The total bitrate is kept constant, while the subsequences are quantized with different stepsizes depending on the sequence as well as on the transmission conditions. Our goal is to figure out under which transmission conditions unbalanced bitstreams lead to good system performance in terms of the average reconstructed PSNR. Besides, we investigate the effects of intra-coding on the error resilience of the system and show that the sequence characteristics, and in particular the degree of motion in the sequence, have an important impact on the decoding performance. Finally, we propose a distortion model that is the core of an optimized rate allocation strategy, which is dependent on the network characteristics and status as well as on the video sequence characteristics.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

## 1. INTRODUCTION

Multimedia communication over the 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. 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  $D_0$ . Multiple 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 [1]. An MSVC system has two main components: multiple state encoding/decoding 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 decodable independently and that the correctly received stream can enable state recovery for the corrupted stream using bidirectional information from past and future frames.

With the 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 the adaptation of the quantization temporal or the spatial resolution of the frame-wise splitted video signal. In [2], we investigated the unbalanced quantized multiple state video coding. We also proposed to use the state recovery property, not only to recover from errors [1], 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 [2]. We apply the substitution by interpolation only if the error propagation on the next frame due to interpolation is below a given threshold.

Our goal in this paper is to build on our previous work, in order to figure out under which conditions unbalanced bitstreams lead to efficient system performance in terms of the average reconstructed frame PSNR. We differentiate between low and high motion sequences. The sequences are reconstructed with the extended MSVC approach at different loss probabilities and unbalance rates. Besides, we investigate the effects of extra intra-coding on the error resilience of the system. Next, we discuss the effect/penalty of the unbalanced descriptions on the reconstruction performance if unbalancing is inevitable (e.g., because of bandwidth limitations). Finally,

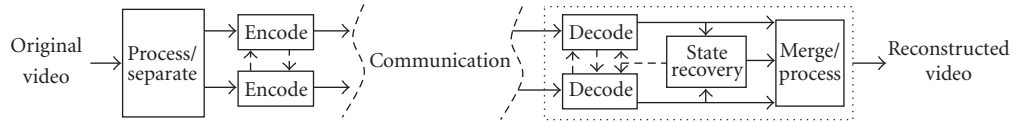


FIGURE 1: Block diagram of the MSVC system.

we propose a distortion model that is the core of an optimized rate allocation strategy, dependent on the network characteristics and status.

Section 2 gives background information for MDC and MSVC. Section 3 analyzes the system performance of a one-dimensional AR(1) source as a simplified model for video to understand the effect of unbalanced rates on the average distortion. In Section 4, an end-to-end distortion model is proposed for MSVC dependent on the network status, and sequence characteristics. Next an optimized rate allocation scheme for unbalanced quantized MSVC is discussed based on this model. Section 5 provides a performance analysis at different streaming conditions, whereas Section 6 concludes the paper.

## 2. BACKGROUND

Multiple description coding is applied to some major coding techniques such as scalar quantization, vector quantization, correlating transforms, or quantized frame expansions. A summary of the state-of-the-art system designs can be found in [3]. MDC techniques are also investigated and applied for video coding. Some of them are MD protection of the most significant DCT coefficients [4], MDC of motion vectors [5], altering of prediction loops [6], scalar quantizers [7], matching pursuit [8], or forward error correction [9], respectively. High rates, low latency requirements, and error drift are however the main problems encountered in MDC schemes for video streaming due to possible desynchronization of encoders and decoders.

The idea of channel splitting has a longer history. It became first popular with speech coders and information theorists in Bell Laboratories in 1978 and 1979. Gersho proposed the use of modulo PCM encoding for channel splitting [10], followed by Jayant who proposed the separation of odd and even samples for speech coding [11, 12]. A more recent technique combines prediction with simple separation for speech coding [13]. Reudink dealt also with scalar quantization for channel splitting. He was the first to propose channel splitting techniques that do not significantly increase the total rate and do not rely entirely on the inherent redundancy in the source sequence [14]. Multiple state video coding (MSVC) approach [1], which is the MDC technique subject to this work, is inspired by the idea of frame-wise splitting of the video signal as in video redundancy coding [15]. The block diagram of MSVC is given in Figure 1 and state recovery method in Figure 2. Figure 2 shows that the packet carrying the coded data for frame #5 is lost and interpolated using the reconstructed frames #4 and #6 whose data were coded in the second stream. Coding gain (bitrate reduction)

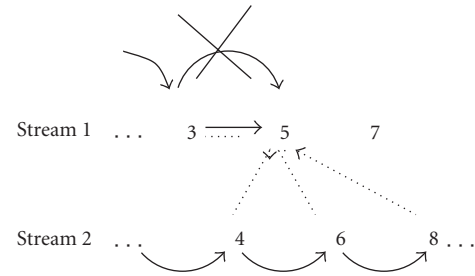


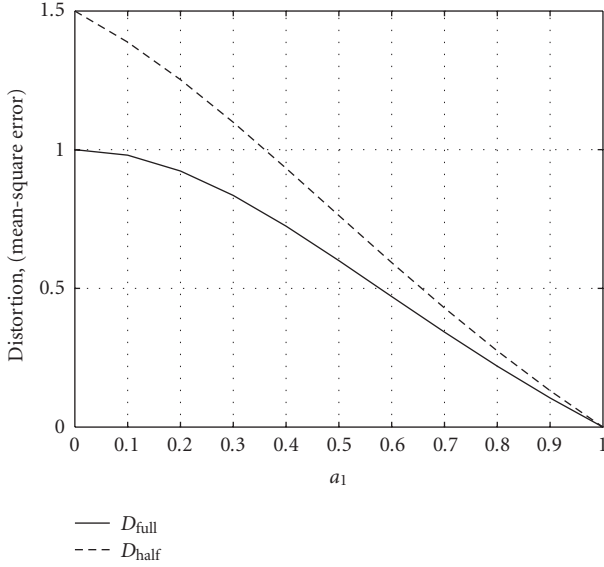
FIGURE 2: Error concealment in MSVC.

in MSVC is smaller than that for single description coding due to the larger temporal distance between adjacent frames in each subsequence. In other words, coding gain is traded off with the resilience to transmission errors.

Unbalanced multiple description video coding, where the descriptions are coded with different bitrates, has not yet been widely explored. One of the works investigating unbalanced MD video coding is [16], where the system produces two descriptions with different resolutions of transform coefficients. The video data is encoded into a high-resolution video stream using an encoder that produces an H.263 compliant stream. In addition, a low-resolution video stream is also generated by duplicating the most important information from the high-resolution video stream such as headers, motion vectors, and some of the discrete cosine transform (DCT) coefficients. The scheme is especially designed for packet loss rates below 10%. The number of duplicated coefficients are determined for a given packet loss probability and rate budget by minimization of the expected distortion. The main disadvantage of the work is that error propagation is not considered in the expected distortion formulation. Another work on unbalanced video coding is given in [17], where MSVC is extended to an unbalanced system based on frame rate adaptation. The video is coded into two distinct streams producing unbalanced frame rates of 2 : 1. But for high motion sequences, due to the increased temporal distance, the recovery of the high-rate stream does not work well causing a decrease in the average reconstruction quality. Unbalanced quantized MSVC is the topic of this paper which investigates the unbalanced extension of MSVC based on quantization adaptation.

## 3. ANALYSIS: ODD/EVEN SEPARATION FOR AN AR(1) SOURCE

A video signal can be roughly approximated as a collection of AR(1) sequences of corresponding (motion compensated)

FIGURE 3: Comparison of  $D_{\text{full}}(R)$  to  $D_{\text{half}}(R)$  at  $R = \log_2(1 + a^2)$ .

pixels along time [18]. In this section, we will analyze the effects of channel splitting on one-dimensional autoregressive sources of first order, AR(1). The insights we gain from the analysis of one-dimensional sources will give us some hints about the error-prone transmission of video signals. Odd/even sample separation makes use of the correlation between consecutive samples. Consider a discrete time first-order autoregressive, AR(1), source model given as

$$x[k] = ax[k-1] + z[k], \quad (1)$$

where  $k \in \mathbb{Z}$  and  $z[k]$  is a sequence of independent, zero-mean Gaussian random variables and  $a$  is the correlation coefficient between consecutive samples.  $x[k]$  is normalized to have unit power if the variance of  $z[k]$  is set to  $1 - a^2$ . The distortion rate function for an AR(1) source is given as [19]

$$D(R) = (1 - a^2)2^{-2R}, \quad \text{for } R \geq \log_2(1 + a). \quad (2)$$

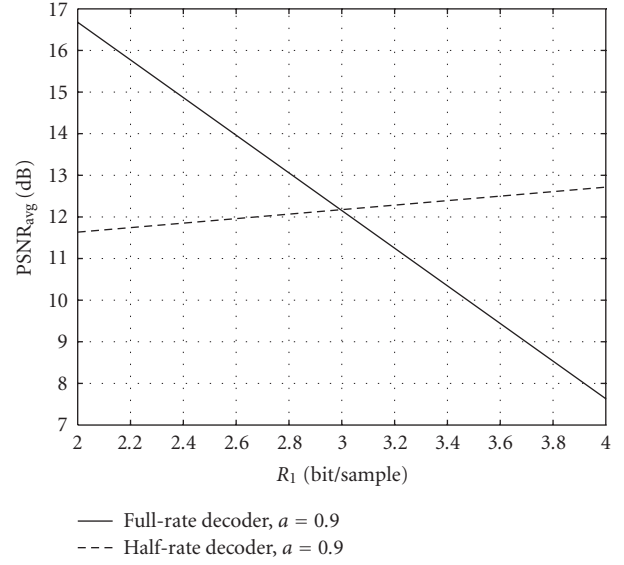
The subsequences of odd and even samples are AR(1) sequences each with the correlation coefficient  $a^2$  and have the distortion rate expression

$$D_{\text{full}}(R) = (1 - a^4)2^{-2R}, \quad \text{for } R \geq \log_2(1 + a^2). \quad (3)$$

If we assume that the even samples are received and the odd samples are reconstructed by interpolation (half-rate decoder), the average distortion over all samples is given as

$$D_{\text{half}}(R) = \frac{1}{2} \left[ (1 - a)^2 + \frac{1}{2}(1 - a^2) \right] + \frac{3}{4}D_{\text{full}}(R) \quad (4)$$

if the quantization error is uncorrelated. The distortion of the even samples does not change, whereas the distortion term on the interpolated samples depends on two factors: (1) interpolation distortion and (2) quantization distortion of the odd samples. At high rates, the component

FIGURE 4:  $\text{PSNR}_{\text{avg}}$  over  $R_1$ , comparison of full-rate to the half-rate decoder in the high-rate region.

of  $D_{\text{half}}(R)$  dependent on quantization becomes almost zero and  $D_{\text{half}}(R)$  approaches asymptotically to interpolation distortion  $1/2[(1 - a)^2 + 1/2(1 - a^2)]$ . At high rate,  $D_{\text{half}}(R)$  exceeds  $D_{\text{full}}(R)$ . At rate  $R = \log_2(1 + a^2)$ , we have

$$D_{\text{full}}(R) = \frac{1 - a^2}{1 + a^2}, \quad (5)$$

$$D_{\text{half}}(R) = \frac{a^4 - 4a^3 + a^2 - 4a + 6}{4(1 + a^2)}.$$

Figure 3 depicts the performance comparison of both decoders for  $1 > a > 0$  stating that  $D_{\text{full}}(R) \leq D_{\text{half}}(R)$ .

However in the low-rate region, where  $R < \log_2(1 + a^2)$ , half-rate decoder yields competitive and at times smaller average distortion than the full-rate decoder [3]. The above equations for  $D_{\text{full}}$  and  $D_{\text{half}}$  are however not valid in the low-rate region [19].

Next, we test the unbalanced quantized operation in the high-rate region, where  $R > \log_2(1 + a^2)$ ,  $R_i > \log_2(1 + a^2)$ , for all  $i \in [1, 2]$ ,  $R = R_{\text{avg}} = (R_1 + R_2)/2 = 2$  bit/sample, and  $a = 0.9$ . Figure 4 shows the comparison of the full-rate decoder to the half-rate decoder. We see that at balanced operation, full-rate coder performs best, but when  $R_1$  becomes larger, (i.e., the unbalance increases), half-rate decoder performs better.

Based on these insights, we investigated what happens when the transmission is lossy, that is, samples are lost with given probabilities. The results are obtained via simulations, where 100 samples are transmitted and 20 000 iterations are used. Increasing the number of iterations beyond did not change the experimental results. The transmission rate is 0.2 bit/sample averaged over both subsequences, that is, lower than the threshold rate,  $\log_2(1 + a^2)$ . The loss probabilities vary between 0 and 0.1. We consider also the unbalanced rate allocation between the subsequences.

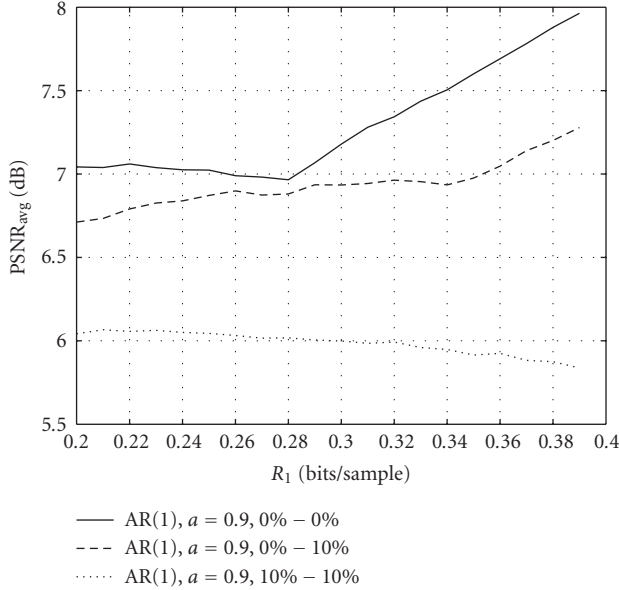


FIGURE 5: PSNR<sub>avg</sub> over  $R_1$ ,  $a = 0.9$ ,  $R_{\text{avg}} = 0.2$  bit/sample.

Average PSNR over the rate of the first subsequence  $R_1$  is plotted for  $a = 0.9$  in Figure 5. Since the total average rate is constant, as  $R_1$  increases  $R_2$  decreases. The curves depicted are generated by comparing the full rate and the half-rate decoders at each unbalance rate and picking the larger value. As  $R_1$  increases and  $R_2$  decreases, after some point, half-rate decoder exceeds the full-rate decoder since the distortion due to error propagation and coarse quantization exceed the interpolation distortion.

In Figure 5, when no loss occurs and  $R_1 > 0.27$  bit/sample, half-rate decoder performs better than the full-rate one. When the first channel is lossless but the second channel is lossy, the threshold rate where half-rate coder exceeds the full-rate coder is smaller. If the loss probabilities are moderate and rate unbalance is high, using half-rate coder increases average PSNR.

The same comparisons are depicted for  $a = 0.5$  in Figure 6. Due to the decreased correlation between consecutive samples, half-rate coder cannot perform as good as the full-rate coder. Only when both channels are lossless and  $R_1 > 0.37$  bit/sample, half-rate decoder outperforms the full-rate decoder. Moreover, we see that the highest average PSNR is achieved at balanced rate allocations. The optimal rate allocation is reached at unbalance only when the second loss probability is larger than 10%. When the loss probabilities are balanced, full-rate decoder at balanced operation is preferred.

To sum up, correlation coefficient matters. This corresponds to the correlation of corresponding blocks in video sequences as discussed in Section 5. If the correlation is low, the error due to interpolation is large. In this case, when the quantization error is not large enough (i.e., high coding rate), the half-rate decoder cannot outperform full-rate coder. Similarly, for low motion sequences, since the

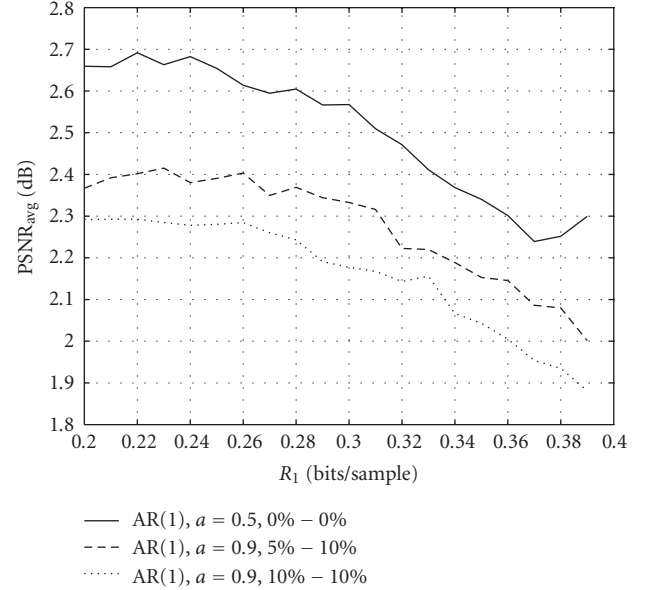


FIGURE 6: PSNR<sub>avg</sub> over  $R_1$ ,  $a = 0.5$ ,  $R_{\text{avg}} = 0.2$  bit/sample.

correlation between corresponding blocks is high, we expect the half-rate to yield good results. In contrary, for high motion sequences, the correlation is low, and therefore there is a penalty associated with the half-rate decoder. Basically, as we will see in the next section, activity in the sequence directly drives the interpolation efficiency and thus the coding strategy.

#### 4. RATE ALLOCATION FOR UNBALANCED QUANTIZED MSVC

Based on a distortion estimation model for unbalanced quantized multiple state video coding system [20], we will present a simplified, nearly optimal solution to the rate allocation problem to guide our optimized coding strategy. Summary of the distortion model is given in Section 4.1 and the model-based rate allocation in Section 4.2.

##### 4.1. Distortion model

For the optimized coding strategy, we will use a recursive block-based distortion estimation model. The model gives a comprehensive estimation of the average distortion which depends on the channel conditions and scene activity. The detailed analysis of the estimation method is provided in [20]. According to the model, the distortion on the current block is dependent on the distortion of its corresponding blocks on the previous frames of the video sequence. The corresponding blocks on the previous frames are located using the motion vector field extracted in the encoding stage. For each block, four cases are considered depending on the reception of the current and corresponding blocks on the previous adjacent frame on the same thread and the next adjacent frame on the other thread.

(i) Blocks  $x(n)$  and  $x(n-2)$  are both received:

$$D_1(n) = \alpha(D(n-2) - \sigma_{q_2}^2) + \sigma_{q_2}^2. \quad (6)$$

(ii) Block  $x(n)$  is received but block  $x(n-2)$  is lost:

$$D_2(n) = \alpha(D(n-2) + \sigma_{q_2}^2) + \sigma_{q_2}^2. \quad (7)$$

(iii) Block  $x(n)$  is lost but block  $x(n+1)$  is received:

$$\begin{aligned} D_3(n) = & \frac{(1 + \sqrt{\alpha})^2}{4} D(n-1) + \frac{1 + \alpha}{4} \sigma_{q_1}^2 \\ & - \frac{(1 + \sqrt{\alpha})\sqrt{\alpha}}{2} \sigma_{q_1}^2 (1 - p_1) \\ & + \sigma_{\text{interp}}^2. \end{aligned} \quad (8)$$

(iv) Blocks  $x(n)$  and  $x(n+1)$  are both lost:

$$D_4(n) = D(n-1) + \sigma_{\text{rep}}^2, \quad (9)$$

where  $x(n)$  is the current block on the first thread,  $x(n-2)$  the corresponding block on the previous frame on the same thread,  $x(n+1)$  the corresponding block in the next frame, and  $x(n-1)$  on the previous frame of the other thread; whereas  $D(n)$ ,  $D(n-2)$ ,  $D(n+1)$ , and  $D(n-1)$  are their corresponding distortion terms, respectively.  $\sigma_{\text{interp}}^2$  denotes the interpolation distortion and  $\sigma_{\text{rep}}^2$  the repetition distortion if interpolation is impossible and the lost block is to be replaced by its corresponding block on the previous frame.  $\sigma_{q_1}^2$  and  $\sigma_{q_2}^2$  are the average quantization distortions for both streams corresponding to  $R_1$  and  $R_2$ , respectively. The model is applicable to inter-coded as well as to intra-coded blocks so that intra-updates can also be considered.  $\alpha$  is a variable which depends on the scene activity of the video sequence and should be determined adaptively. All parameters used in the model are estimated at the encoder and are obviously sequence dependent. To predict  $\alpha$ , we can match the simulation results to the real distortion values on a frame basis and minimize the gap for a given loss rate. The predicted  $\alpha$  value can then be used for the distortion estimation at other loss rates. The formulas for a block on the other thread is symmetric to the given one, where the subscripts “1” and “2” are to be exchanged.

The block distortion values for each case are weighted by their respective probabilities to yield the overall block distortion:

$$\begin{aligned} D(n) = & (1 - p_2)(1 - p_2)D_1(n) + (1 - p_2)p_2D_2(n) \\ & + p_2(1 - p_1)D_3(n) + p_2p_1D_4(n). \end{aligned} \quad (10)$$

The individual frame distortions are calculated from the block distortions, whereas the frame PSNR values from the frame distortions. The average PSNR values for specific rate allocations are calculated by averaging the frame PSNR estimations over the whole sequence. To get the optimal coding rates, the distortion should be minimized while the rate constraints of the channels are respected.

## 4.2. Optimized rate allocation

In this section, a nearly optimal rate allocation will be described based on the average distortion estimation. For our model, we will use an  $\alpha$  value averaged over all frames, blocks and streams as an approximation. To be exact,  $\alpha$  is slightly dependent on the coding rate, therefore different  $\alpha$  values are associated with differently quantized streams. The average quantization distortions of the two substreams are  $\sigma_{q_1}^2$  and  $\sigma_{q_2}^2$ , respectively. Similarly, the sum of the average distortion terms of the corresponding blocks along the odd and even video subsequence are denoted by  $D_O$  and  $D_E$ . Our goal is minimizing the total distortion  $\min\{D_E + D_O\}$  subject to  $f(\sigma_{q_1}^2, \sigma_{q_2}^2) \leq C$ , that is,  $R_1 + R_2 \leq R_T$ , for a given constant  $C$ .  $D_O$  can be expressed as

$$\begin{aligned} D_O = & \sum_{n=1:2}^{N-1} D(n) \\ = & \sigma_{q_1}^2 + \sum_{n=3:2}^{N-1} [(1 - p_1)^2 D_1(n) + (1 - p_1)p_1 D_2(n) \\ & + p_1(1 - p_2)D_3(n) + p_1p_2D_4(n)], \end{aligned} \quad (11)$$

using the distortion model.  $N$  is the number of frames considered. A symmetric formula can also be written for  $D_E$  by summing only the even numbered frame distortions by exchanging the subscripts. After expanding the summation terms we obtain for  $D_O$

$$\begin{aligned} D_O = & \frac{1}{1 - (1 - p_1)\alpha} \\ & \times \left[ \sigma_{q_1} \left[ 1 + (1 - p_1)^2 \frac{N-2}{2} (1 - \alpha) \right. \right. \\ & \left. \left. + (1 - p_1)p_1 \frac{N-2}{2} (1 + \alpha) \right] \right. \\ & + \sigma_{q_2} \left[ (1 - p_2)p_1 \frac{N-2}{2} \frac{(1 + \alpha)}{4} \right. \\ & \left. - (1 - p_2)^2 p_1 \frac{N-2}{2} \frac{N-2}{2} \frac{(1 + \sqrt{\alpha})}{2} \sqrt{\alpha} \right] \\ & + D(N-1)[2(p_1 - 1)] \\ & + D(N) \left[ - (1 - p_2)p_1 \frac{(1 + \sqrt{\alpha})^2}{4} - p_1p_2 \right] \\ & + \sigma_{\text{interp}}^2 \left[ (1 - p_2)p_1 \frac{N-2}{2} \right] \\ & + \sigma_{\text{rep}}^2 \left[ p_1p_2 \frac{N-2}{2} \right] \\ & + D_E \left[ (1 - p_2)p_1 \frac{(1 + \sqrt{\alpha})}{4} + p_1p_2 \right], \end{aligned} \quad (12)$$

where the averaged values for  $\alpha$ ,  $\sigma_{\text{interp}}^2$ , and  $\sigma_{\text{rep}}^2$  over all blocks of the whole sequence are used to simplify the equation. A similar equation can also be written for  $D_E$  by exchanging the subscripts. Next, the equations for  $D_E$  and  $D_O$  can be combined and formulated in the following way:

$$\begin{aligned} D_E = & K_1\sigma_{q_1}^2 + K_2\sigma_{q_2}^2 + K_3D(N-1) + K_4D(N) + K_5, \\ D_O = & K_6\sigma_{q_1}^2 + K_7\sigma_{q_2}^2 + K_8D(N-1) + K_9D(N) + K_{10}, \end{aligned} \quad (13)$$

where  $K_i$ ,  $i \in [1, 10]$ , are given by the system and dependent on  $p_1$ ,  $p_2$ ,  $\alpha$ ,  $N$ ,  $\sigma_{\text{interp}}$ , and  $\sigma_{\text{rep}}$ . For large  $N$ ,  $D(N-1)$  and  $D(N)$  can be neglected in comparison to the remaining terms. Adding the equations for  $D_E$  and  $D_O$  together, we obtain

$$D_E + D_O = (K_1 + K_6)\sigma_1^2 + (K_2 + K_7)\sigma_2^2 + (K_5 + K_{10}), \quad (14)$$

which is an equation with two variables  $\sigma_{q_1}$  and  $\sigma_{q_2}$  which are dependent on the quantizers  $Q_1$  and  $Q_2$ . The constraint for the optimization is  $R_1(Q_1) + R_2(Q_2) \leq R$ .

## 5. PERFORMANCE ANALYSIS FOR UNBALANCED QUANTIZED MSVC

### 5.1. Preliminaries

In balanced MSVC, each subsequence is quantized with the same step size after splitting the original sequence into odd and even frames. The resulting descriptions require the same bitrate, have the same frame rate, and yield the same average PSNR if no loss occurs. On the other hand, unbalanced descriptions could be advantageous if we have transmission paths with different characteristics, such as different bandwidths and loss probabilities. We investigate here unbalanced quantized MSVC [2], where the subsequences are quantized with different stepsizes depending on the sequence as well as on the transmission conditions. In our work, MSVC is modified to recover the streams not only from state losses, but also from error propagation, and more importantly to improve the low quality stream which is coarsely quantized to increase the average reconstructed PSNR. The extended recovery method chooses the best frame PSNR reconstruction method from all available alternatives at every stage. This information can be sent to the decoder as side information. Using this, the decoder reconstructs the best possible sequence depending on the received video packets.

If the current packet is lost, there is only one reconstruction method available: interpolating the previous and next frames from the other description. If it is received, on the other hand, two alternatives are present: (1) using the received packet and (2) interpolating as if the packet was lost. If the interpolation using the finely quantized frames yields a higher frame PSNR than the coarsely quantized and the reconstructed one, the second alternative should be chosen. An important point is that due to prediction in hybrid coding, the interpolation error on the reconstructed frame propagates to the following frames in the same subsequence. The severity of error propagation depends on the scene activity and the interpolation method. To take this effect into account, we employ the second alternative only if the PSNR degradation of the future frame (after the current one) is below a certain threshold. The threshold can be preset dependent on the content as well as on the application. The comparison based on PSNR cannot be performed directly at the decoder since the original sequence is not present. But the encoder can predetermine the rate allocation between given paths so that the average reconstructed frame

PSNR,  $\text{PSNR}_{\text{avg}}$ , is maximized. Our goal is to allocate a given total bitrate  $R_T$  optimally between two streams by setting their quantization stepsizes, as well as intra-group of blocks (GOB) or frame coding periods based on a comprehensive end-to-end distortion model which takes the scene content, loss probabilities, and coding rates into account. Alternatively, the encoder can detect the motion activity of the sequence or part of the sequence and adapt the rate unbalance according to the precalculated and prestored tables.

### 5.2. System setup

We modified the H.264 codec (version 9.0) to support the MSVC structure, and two parallel decoders are implemented which help each other to recover from losses as explained here above. The reliability values can be determined recursively. The optimal reconstruction method depends on both the loss history and the scene activity. Side information is sent by the encoder to help choosing the best reconstruction strategy. The side information can be some hint specifying the motion activity for each particular frame (or frame block), that is, representing the difficulty of interpolating it using the adjacent frames. The hint track can for example be generated offline at the encoder. An alternative scenario that we do not consider here consists in estimating sequence activity directly at the receiver to choose the best reconstruction method.

Finally, we assume that each frame (I or P) is transmitted in a single packet [1]. 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 method for the lost I and P frames is the same.

### 5.3. Performance analysis

We now analyze the performance of unbalanced MSVC for different sequences and different streaming conditions. A careful analysis will provide us with important insights for the design of efficient coding strategies. The results are given for two sequences here: Foreman and Akiyo representing the high motion and low motion sequences, respectively. For each coding option (non-intra, with intra-GOB updates, and with intra-frame updates), four operating points (P.1, P.2, P.3, and P.4) are considered. Operating points correspond to different settings of quantization stepsizes of both streams. The first operating point denotes the balanced rate allocation, whereas the fourth one represents the most unbalanced one. Tables 1, 2, and 3 list the bitrates and the corresponding QP's for the two streams chosen for coding the sequences Foreman and Akiyo without and with intra-coding of GOB's and frames. The total rate  $R_T$  is set to 140 kbit/s for Foreman and 19 kbit/s for Akiyo which is to be split between  $R_1$  and  $R_2$ . For Foreman, every 9th frame and for Akiyo, every 36th one is intra-coded for the coding option with intra frame updates. Similarly, one GOB in every frame for Foreman and in every 4th frame for Akiyo is intra-coded for the option with

TABLE 1: Unbalanced quantized MSVC, quantization stepsizes, and the corresponding bitrates for the two subsequences: Foreman and Akiyo.

	$QP_1$	$R_1$ (kbits/s)	$QP_2$	$R_2$ (kbits/s)
P.1	14/18	111.88/13.92	26/27	27.48/5.03
P.2	15/19	95.94/12.16	21/24	43.01/6.75
P.3	16/20	83.16/10.64	19/22	54.91/8.41
P.4	17/21	71.38/9.41	17/21	71.04/9.41

TABLE 2: Unbalanced quantized MSVC, quantization stepsizes, and the corresponding bitrates for the two subsequences: Foreman and Akiyo with GOB-intra-updates.

	$QP_1$	$R_1$ (kbits/s)	$QP_2$	$R_2$ (kbits/s)
P.1	17/22	106.38/13.22	27/30	34.02/5.97
P.2	18/23	94.55/11.67	24/28	46.98/7.14
P.3	19/24	83.27/10.59	22/26	59.14/8.72
P.4	20/25	73.83/9.64	21/25	65.68/9.64

TABLE 3: Unbalanced quantized MSVC, quantization stepsizes, and the corresponding bitrates for the two subsequences: Foreman and Akiyo with frame-intra-updates.

	$QP_1$	$R_1$ (kbits/s)	$QP_2$	$R_2$ (kbits/s)
P.1	20/23	83.90/11.26	29/30	55.68/7.97
P.2	21/24	78.23/10.56	26/28	61.36/8.61
P.3	22/25	74.04/9.95	24/27	66.73/8.97
P.4	23/26	70.24/9.45	23/26	70.24/9.45

intra-GOB updates. The vertical location of the intra-coded GOB is shifted downwards periodically every frame. A total of 200 frames are considered from each sequence. To investigate the system performance, 100 different random loss patterns for each loss probability pair are generated with a uniformly distributed independent loss model. Although some of the loss patterns generated contain bursty errors (especially if the loss probability is high), bursty loss models will be investigated in the future work. Encoded subsequences are decoded and reconstructed using the extended recovery approach. Block-based motion controlled interpolation is employed for extended state recovery.

Figures 7, 8, and 9 show for Foreman the average reconstructed frame PSNR over the rate allocated to the first stream for balanced and unbalanced loss probability pairs in case the subsequences are coded without intra-coding, with GOB-intra-updates, and with frame updates, respectively. The same comparison is depicted for Akiyo in Figures 10, 11, and 12.  $PSNR_{avg}$  is given on the  $y$ -axis, whereas  $R_1$ , the rate allocated to the first stream, is given on the  $x$ -axis. The left corner of the  $x$ -axis corresponds to balanced rate allocation, whereas the right corner represents the most unbalanced one. Observations for the three categories of loss probability pairs are listed below.

(a) First stream is lossless and second stream is lossy.

(i) For Foreman, whether the periodic intra-GOB coding is applied or not, the unbalanced operation is preferred. The optimal rate  $R_1$  is smaller with

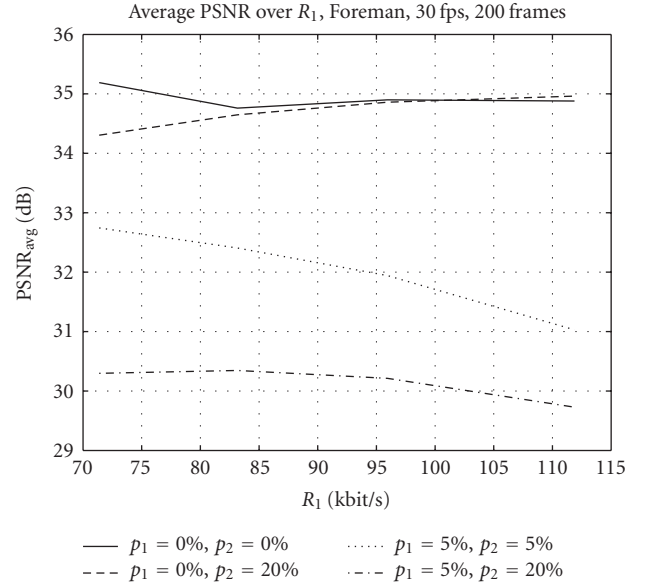


FIGURE 7:  $PSNR_{avg}$  over  $R_1$ , Foreman.

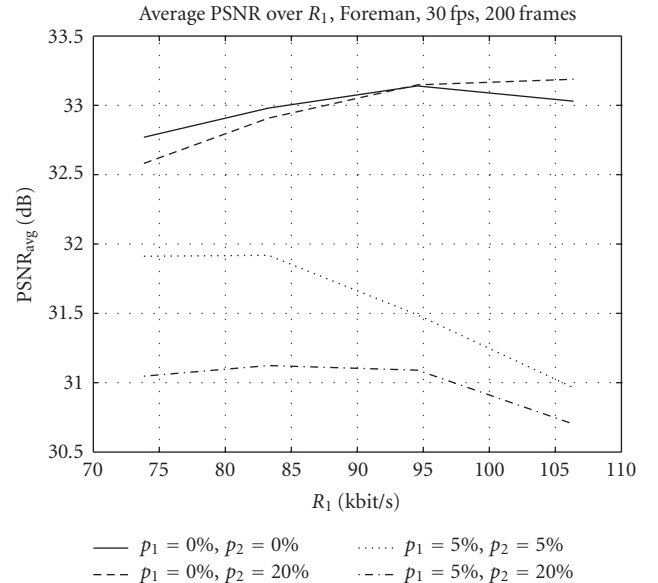
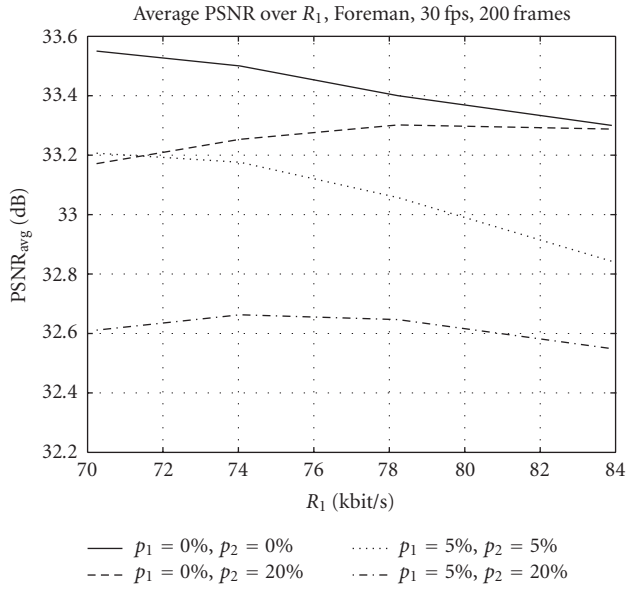
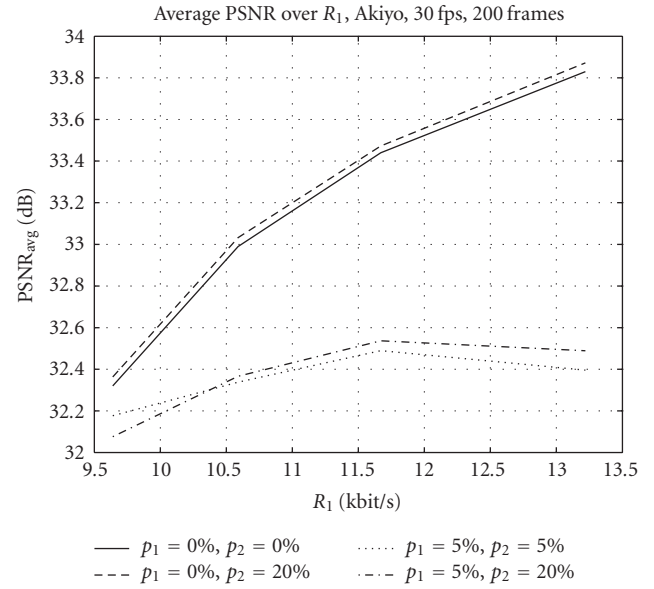
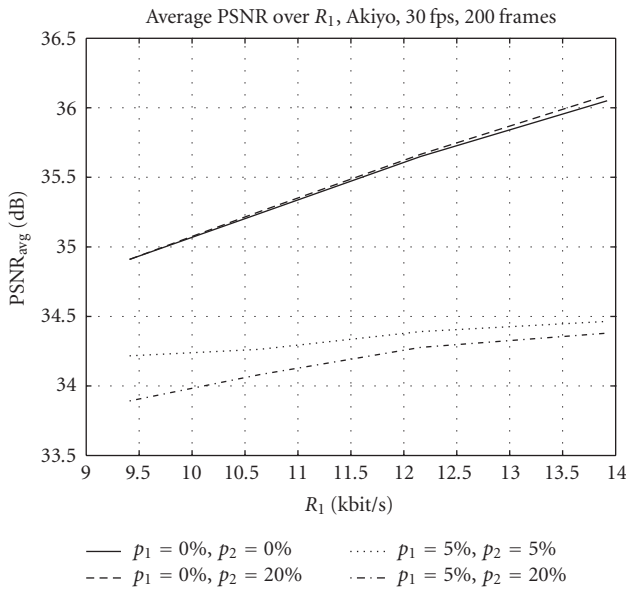
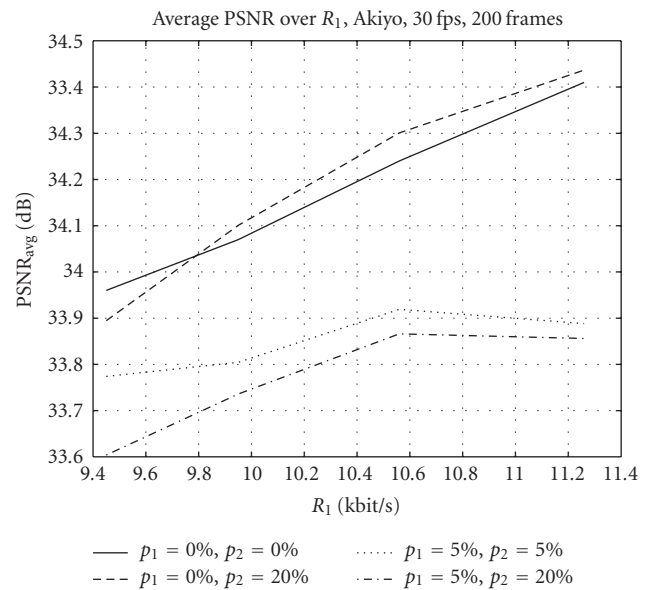


FIGURE 8:  $PSNR_{avg}$  over  $R_1$ , Foreman with GOB-intra-updates.

the intra-GOB coding than with no-intra-coding. With intra-frame coding, unbalanced rates are favorable only when  $p_2$  increases beyond 5%.

- (ii) For Akiyo, however, interpolation yields very good results due to low motion. The optimal operating point for Akiyo is therefore the most unbalanced one whatever intra-coding option is used.
- (iii) Note that, the packets which are received but not used (since interpolation yields better results) are not employed to enhance the reconstruction results in the experiments (subject of future work). If the received but discarded packets, that is, coarsely

FIGURE 9: PSNR<sub>avg</sub> over  $R_1$ , Foreman with frame-intra-updates.FIGURE 11: PSNR<sub>avg</sub> over  $R_1$ , Akiyo with GOB-intra-updates.FIGURE 10: PSNR<sub>avg</sub> over  $R_1$ , Akiyo.FIGURE 12: PSNR<sub>avg</sub> over  $R_1$ , Akiyo with frame-intra-updates.

quantized images, can be incorporated to enhance the reconstruction further, we expect the optimal rate  $R_1$  to increase, favoring unbalanced operation even more. Using coarsely quantized images as side information is discussed in [21] in the context of the Wyner-Ziv coding.

- (iv) For Foreman, if both streams are lossless, PSNR<sub>avg</sub> decreases slightly, first as  $R_1$  increases, followed by a slight increase. The slight decrease in performance is due to the conservative (suboptimal) threshold setting in the simulations.

(b) Both streams are lossy.

- (i) Balanced loss probabilities call for balanced rate allocations whether periodic intra-coding is applied or not and whether the sequence has high motion content or not.
- (ii) For Foreman at unbalanced loss probabilities, balanced operation is still slightly favored. For Akiyo at unbalanced loss probabilities and at loss rates that are smaller than 20%, unbalanced operation is preferred.



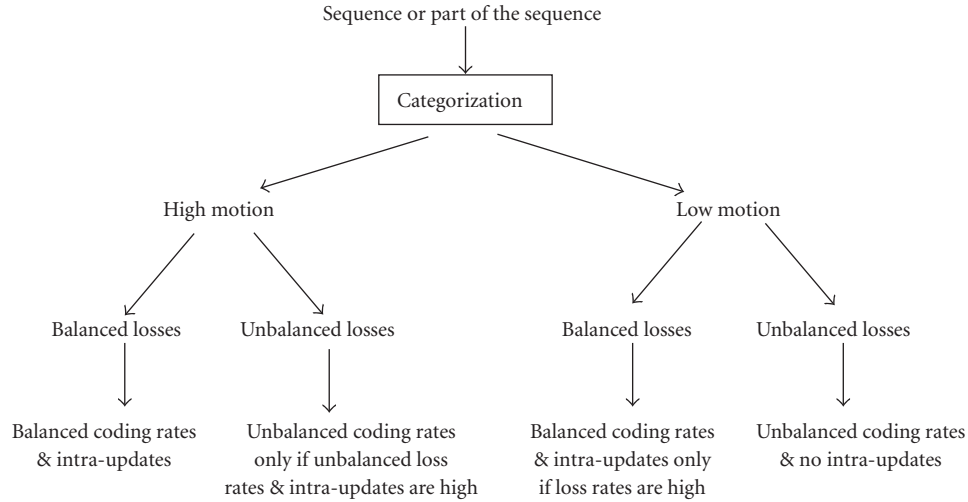


FIGURE 13: Heuristics for coding decisions.

- (iii) The low-rate stream is not fully used in the current work. Since the higher-rate stream is also lossy, the low-rate stream is to be used to recover the high-rate stream. If the sequence length until the next refreshment is long, the overall quality can decrease due to the coarse quantization of the second description.
- (iv) As noted before, the results can be improved in favor of unbalanced operation by incorporating the received but discarded packets into the reconstruction.
- (v) Performance differences between operating points are smaller when intra-updates, especially frame updates are used.

Heuristics and results we obtained from the here-above analysis and experiments are summarized in Figure 13. First, we have to distinguish between high and low motion sequences. Second, we should check whether the loss probabilities of the channels are balanced or not. For high motion sequences, if we have balanced loss probabilities, balanced coding rates with intra-updates are optimal. But if the loss probabilities are highly unbalanced, unbalanced rates in combination with intra-updates are to be preferred. For low motion sequences, on the other hand, at balanced loss rates, balanced rates are favorable and the frame updates are to be used only if the loss rates are high (more than 10%). For unbalanced loss probabilities, however, no intra-updates are necessary and unbalanced rates are the best strategy.

## 6. CONCLUSIONS

Unbalanced descriptions are particularly interesting for video streaming applications over heterogeneous networks using path diversity, where transmission channels have varying transmission characteristics. 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 multiple state video coding approach. The total bitrate is kept constant while the subsequences are quantized with different stepsizes adaptive to the sequence as well as to the transmission conditions. Besides, the state recovery approach of MSVC is extended to enhance the quality of the coarsely quantized stream. The system performance in terms of average PSNR is investigated for different loss rates, rate allocations, coding options, and sequences. To determine the optimal coding parameters, we distinguish between high and low motion sequences as well as between balanced and unbalanced transmission conditions (bandwidth or loss probabilities). For high motion sequences and for balanced loss probabilities, balanced coding rates with intra-updates are optimal. But if the loss probabilities are highly unbalanced, unbalanced rates in combination with intra-updates are to be preferred. For low motion sequences and balanced loss rates, balanced rates are favorable and the frame updates are to be used only if the loss rates are high (more than 10%). If the loss probabilities are unbalanced, no intra-updates are necessary and unbalanced rates are optimal. Another result is that extended state recovery reduces the penalty in the average system performance due to unbalancing. At last, we introduce a distortion model to guide the optimized coding strategy. Additionally, we present a nearly optimal rate allocation method based on this distortion model. Future work will focus on the improvement of the results in unbalanced operation using the coarsely quantized images as side information at the interpolation process.

## ACKNOWLEDGMENTS

This work was supported in part by the Technical University Berlin and the Swiss National Science Foundation.

## REFERENCES

- [1] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path

- diversity,” in *Visual Communications and Image Processing (VCIP '01)*, vol. 4310 of *Proceedings of SPIE*, pp. 392–409, San Jose, Calif, USA, January 2001.
- [2] S. Ekmekci and T. Sikora, “Unbalanced quantized multiple description video transmission using path diversity,” in *Image and Video Communications and Processing (IVCP '03)*, vol. 5022 of *Proceedings of SPIE*, pp. 301–311, Santa Clara, Calif, USA, January 2003.
- [3] V. K. Goyal, “Multiple description coding: compression meets the network,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, 2001.
- [4] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, “A robust codec for transmission of very low bit-rate video over channels with bursty errors,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1403–1412, 2000.
- [5] C.-S. Kim and S.-U. Lee, “Multiple description motion coding algorithm for robust video transmission,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 4, pp. 717–720, Geneva, Switzerland, May 2000.
- [6] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, “Multiple-description video coding using motion-compensated temporal prediction,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 193–204, 2002.
- [7] V. A. Vaishampayan and J. Domaszewicz, “Design of entropy-constrained multiple-description scalar quantizers,” *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 245–250, 1994.
- [8] X. Tang and A. Zakhor, “Matching pursuits multiple description coding for wireless video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 566–575, 2002.
- [9] R. Puri, K. Ramchandran, K. W. Lee, and V. Bharghavan, “Forward error correction (FEC) codes based multiple description coding for Internet video streaming and multicast,” *Signal Processing: Image Communication*, vol. 16, no. 8, pp. 745–762, 2001.
- [10] A. Gersho, “The channel splitting problem and Modulo-PCM coding,” *Bell Labs Memo for Record (not archived)*, October 1979.
- [11] N. S. Jayant, “Subsampling of a DPCM speech channel to provide two self-contained half-rate channels,” *Bell System Technical Journal*, vol. 60, no. 4, pp. 501–509, 1981.
- [12] N. S. Jayant and S. W. Christensen, “Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure,” *IEEE Transactions on Communications*, vol. 29, no. 2, pp. 101–109, 1981.
- [13] A. Ingle and V. A. Vaishampayan, “DPCM system design for diversity systems with applications to packetized speech,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 48–58, 1995.
- [14] D. O. Reudink, “The channel splitting problem with interpolative coders,” Tech. Rep. TM80-134-1, Bell Labs, Murray Hill, NJ, USA, October 1980.
- [15] S. Wenger, G. D. Knorr, J. Ott, and F. Kossentini, “Error resilience support in H.263+,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 867–877, 1998.
- [16] D. Comas, R. Singh, A. Ortega, and F. Marques, “Unbalanced multiple-description video coding with rate-distortion optimization,” *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 81–90, 2003.
- [17] J. G. Apostolopoulos and S. J. Wee, “Unbalanced multiple description video communication using path diversity,” in *Proceedings of IEEE International Conference on Image Processing (ICIP '01)*, vol. 1, pp. 966–969, Thessaloniki, Greece, October 2001.
- [18] C. Chen, *Motion-compensated hybrid coders in video communications*, Ph.D. thesis, Monash University, Victoria, Australia, 1992.
- [19] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1971.
- [20] S. Ekmekci and T. Sikora, “Model-based unbalanced multiple description video transmission using path diversity,” in *Visual Communications and Image Processing (VCIP '03)*, vol. 5150 of *Proceedings of SPIE*, pp. 1023–1034, Lugano, Switzerland, July 2003.
- [21] A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video,” in *Conference Record of the 36th Asilomar Conference on Signals, Systems and Computers (ACSSC '02)*, vol. 1, pp. 240–244, Pacific Grove, Calif, USA, November 2002.

**Sila Ekmekci Flierl** received the B.S. degrees in electrical engineering and in computer engineering from Bogazici University, Turkey, in 1996 and the M.S. and the Ph.D. degrees from the Technical University Berlin, Germany, in 1998 and 2004, respectively. She joined the Signal Processing Institute of the Swiss Institute of Technology in 2004 as a Senior Researcher.



**Thomas Sikora** is a Professor and the Director of the Communication Systems Department at Technical University Berlin, Germany, since 2002. He received the Dipl.-Ing. degree and Dr.-Ing. degree in electrical engineering from Bremen University, Germany, in 1985 and 1989, respectively.



**Pascal Frossard** received the M.S. and Ph.D. degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. From 1998 to 2000, he worked with the Signal Processing Laboratory, EPFL, as a Research and Teaching Assistant under a grant from Hewlett-Packard. From 2001 to 2003, he was a Member of the research staff at the IBM TJ Watson Research Center, Yorktown Heights, NY. Since April 2003, he has been an Assistant Professor at EPFL. His research interests include image representation and coding, nonlinear representations, visual information analysis, joint source and channel coding, multimedia communications, and multimedia content distribution.

