Title

Multiview coding and error correction coding for 3D video over noisy channels

Authors

Vosoughi, Arash
Testoni, Vanessa
Cosman, Pamela C
et al.

Publication Date

2015

DOI

10.1016/j.image.2014.10.006

Peer reviewed

Signal Processing: *Image Communication* 30 (2015) 107–120

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Multiview coding and error correction coding for 3D video over noisy channels ☆

Arash Vosoughi *, Vanessa Testoni, Pamela C. Cosman, Laurence B. Milstein

University of California, San Diego, Department of Electrical and Computer Engineering, 9500 Gilman Drive, Mail Code 0407, La Jolla, CA 92093-0407, USA

ARTICLE INFO

Article history:
Received 10 April 2014
Received in revised form
19 October 2014
Accepted 20 October 2014
Available online 30 October 2014

Keywords:
Joint source–channel coding
Multiview video coding
Unequal error protection
Scalable coding
Asymmetric coding

ABSTRACT

We consider the joint source–channel coding problem of stereo video transmitted over AWGN and flat Rayleigh fading channels. Multiview coding (MVC) is used to encode the source, as well as a type of spatial scalable MVC. Our goal is to minimize the total number of bits, which is the sum of the number of source bits and the number of forward error correction bits, under the constraints that the quality of the left and right views must each be greater than predetermined PSNR thresholds at the receiver. We first consider symmetric coding, for which the quality thresholds are equal. Following binocular suppression theory, we also consider asymmetric coding, for which the quality thresholds are unequal. The optimization problem is solved using both equal error protection (EEP) and a proposed unequal error protection (UEP) scheme. An estimate of the expected end-to-end distortion of the two views is formulated for a packetized MVC bitstream over a noisy channel. The UEP algorithm uses these estimates for packet rate allocation. Results for various scenarios, including non-scalable/scalable MVC, symmetric/asymmetric coding, and UEP/EEP, are provided for both AWGN and flat Rayleigh fading channels. The UEP bit savings compared to EEP are given, and the performances of different scenarios are compared for a set of stereo video sequences.

1. Introduction

A key obstacle for the widespread adoption of 3D video is stereoscopic transmission. The doubled amount of video data to be transmitted can be a challenge for both wired and wireless networks. Therefore, efficiently compressing the 3D source video, while keeping a high quality end-user 3D experience, has been an active research topic.

For noisy channels, the video has to be protected by error-correcting codes, but since these codes increase the number of bits to be transmitted, the error protection should be cleverly applied. This tradeoff between source coding accuracy and channel error protection in error prone channels is a joint source channel coding (JSCC) problem and is a well-studied area for single-view video sequences. The work in [1] presents a comprehensive review on this topic while the work in [2] applies JSCC specifically for video transmission over additive white Gaussian noise (AWGN) channels using rate compatible punctured convolutional (RCPC) codes [3]. The optimal point found by JSCC varies over different AWGN channel signal to noise ratios (SNRs). JSCC for single-view video sequences is also studied for several wireless environments in [4].

One approach to video protection is to employ different channel code rates for each video packet according to its importance, which is a type of unequal error protection (UEP). The importance of each packet can be determined

☆ The material in this paper was presented in part at IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

* Corresponding author. Tel.: +1 607 542 8001.
E-mail addresses: arvosoughi@ucsd.edu (A. Vosoughi), vtestoni@ucsd.edu (V. Testoni), pcosman@ucsd.edu (P.C. Cosman), lmilstein@ucsd.edu (L.B. Milstein).

http://dx.doi.org/10.1016/j.image.2014.10.006
0923-5965/

by the estimation of the distortion in the reconstructed video produced by each packet loss separately. The distortion of the reconstructed video should be reduced when compared to the reconstructed video protected with equal error protection (EEP), where all video packets are coded with the same channel code rate.

The JSCC scheme proposed here differs from prior cited works in the way that it is formulated. For many image/video JSCC problems, a typical optimization approach is to fix a total rate of B bits and then determine the optimal division of B between source and channel coding, where the objective function could be the PSNR to be maximized. An example of this type of optimization for 3D video can be found in [5], where a weighted average of the PSNR of the left view and the PSNR of the right view is used as the quality metric. Formulating the optimization in this way is problematic for 3D video sequences because although the PSNR for the left view is well defined, as is the PSNR for the right view, there is not yet any well-accepted way to quantify the quality of the combined 3D video. A quality metric for 3D video is still an unsolved and difficult problem [6,7]. Maximizing the average PSNR subject to a rate constraint would imply that left/right PSNRs of 20/40 and of 30/30 produce equivalent average PSNRs, although the subjective visual quality might be very different. Indeed, if the rate constraint were such that one of the two views could have distortion driven to near zero, so the PSNR approaches infinity, the distortion of the other view could be arbitrarily bad and yet the average PSNR would be maximized. Our alternative approach to the optimization is to fix the distortion or PSNR of each view to some level, and then attempt to minimize the number of bits required to achieve it. Putting the distortion in the constraint, rather than in the objective function, allows one to choose two separate constraints (one for each view). Therefore, the particular goal of our JSCC scheme is to minimize the total bit-rate, composed of source and error-correction bits, while both reconstructed views achieve predetermined PSNR values. The details of this scheme and of the end-to-end distortion model are presented in Section 3.

Regarding only the source coding, a few aspects of the 3D video can be exploited in order to reduce the amount of coded video and make the sequence more error resilient. Exploiting the redundancy between the left and the right views, the multiview video coding (MVC) [8] extension of H.264/AVC applies disparity compensation for encoding inter-view similarities, and motion compensation for encoding temporal similarities. In our previous work [9], we proposed a JSCC scheme using MVC and rate compatible punctured turbo (RCPT) codes. The 3D video was transmitted over an AWGN channel. In this work, we consider video transmission over both an AWGN channel and a flat Rayleigh fading channel, and exploit additional aspects of the 3D video coding.

The first additional 3D video aspect exploited here is related to binocular suppression theory [10–15]. This theory says that the human visual system (HVS) is insensitive to spatial errors which occur in one view only. This result, determined experimentally, can be explained by the ability of the HVS to compensate for missing information. Because the visual cortex does not always receive perfect information from both eyes, it must infer some information given what is provided. That can mean suppressing errors which occur in a single view, while obtaining the necessary information from the other. Binocular suppression theory has given rise to asymmetric video coding, in which one view is coded with higher quality than the other.

Another 3D video coding possibility exploited here is scalability. Scalability incurs a penalty in rate-distortion performance, but it has the advantage of allowing graceful degradation if necessary to downscale the rate, and it naturally allows the use of unequal error protection. Scalable 3D video coding will be addressed in Section 2, which will also propose a scenario for the combination of multiview, asymmetric and scalable coding for 3D video sequences.

The gains obtained with the JSCC/UEP proposed scheme when compared with the JSCC/EEP approach are shown in Section 4 for both AWGN and flat Rayleigh fading channels. This section also presents bit saving percentages for several scenarios, such as non-scalable MVC against scalable MVC. Finally, Section 5 concludes this work.

2. Combining multiview, scalable and asymmetric coding for 3D video

2.1. Multiview coding

The performance and transmission of MVC bitstreams in error-prone channels have been studied in [16–20]. Some of the works on multiview streaming optimization, as in [16], propose end-to-end distortion models taking into account estimated packet loss probabilities for multiview video packets, but do not include channel error protection schemes. The work in [17] has the same characteristics, but includes a form of UEP by simply setting a smaller packet loss rate for the packets in the base view as well as the packets in the first 20 frames of the other views. Another work [18] that studied the transmission of multiview video sequences over error-prone channels considered UEP through a selective packet discard mechanism. Several error resilience techniques for multiview video sequences are described in [19,20].

In this paper, the MVC base view is called the primary view, and corresponds to the left-eye stereo view, while the right-eye stereo view is called the secondary view, and corresponds to the MVC enhancement view. Due to the inter-view dependencies exploited by MVC, the distortion estimations computed by our UEP scheme for the secondary view must also take into account the distortion generated by primary view packet losses. The formulation is described in Sections 3.1 and 3.2.

2.2. Adding scalable coding

The usual modes of scalability are temporal, spatial, and quality (or SNR) scalability. Temporal scalability is achieved when sub-streams of the scalable bitstream are coded with progressively reduced frame rates. Spatial scalability is achieved when sub-streams are coded with progressively reduced spatial resolutions. With quality scalability, frame rate and resolution are constant for all the sub-streams, but each of them is coded with a progressively lower quality.

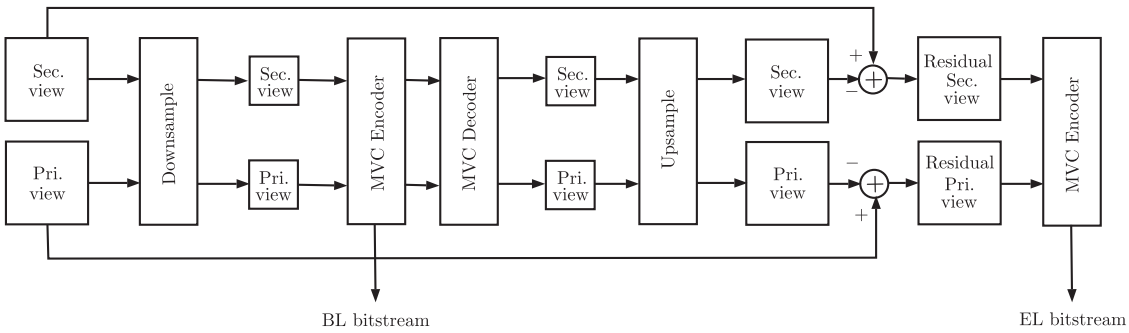

Fig. 1. Proposed scenario for spatial scalable MVC coding per GOP. One MVC bitstream is generated for the base layer and another MVC bitstream is generated for the enhancement layer.

For single-view video sequences, scalable bitstreams can be obtained with the scalable video coding (SVC) [21] extension of H.264/AVC. All types of scalability previously described can be combined with SVC [22] and the final bitstream can be generated with a base layer (BL), which is compatible with H.264/AVC, and several enhancement layers (EL). However, the same flexibility in the bitstream configuration is not supported by the standards for 3D video sequences. Only a temporal scalable MVC bitstream can be achieved with MVC by using the hierarchical prediction structure. However, results such as those presented in [23,24] show that temporal scalability in either just one or both views gives good results for low motion video, but for medium to high motion video, it may be unacceptable due to a visible jumping effect. Although there is no standard-compliant spatial or quality scalable MVC bitstream, several non-standard variants have been proposed [25–28].

Other works proposed for 3D scalable coding video sequences try to define the best mode of scalability for 3D video. Early subjective tests with MPEG-2 in [29,30] show that spatial scalability is preferred over quality scalability. The reason is that in overall stereoscopic perception, especially for low bit rates, blocking artifacts produced by quality scalability implementations are more disturbing than the blurring effect produced by spatial scalability implementations. However, newer results in [31–33], indicate that the perceived quality depends on the 3D display and also that MPEG-2 may cause different artifacts than H.264/AVC on coded video. According to these results, users prefer quality scalability for polarized projection displays and spatial scalability for autostereoscopic parallax barrier displays.

In terms of bit-rate coding efficiency, results in [33] show that, on the average, quality scalability achieves bit streams 15% smaller than spatial scalability for medium to high bit rate scenarios. On the other hand, for low bit rate scenarios and quality around 30 dB (or lower), the achieved bit rate is approximately the same. Results in [33] also show that, if the primary view is encoded at sufficiently high quality and the secondary view is encoded at low quality, users prefer spatial scalability over quality scalability.

2.3. Adding asymmetric coding

Asymmetric coding is another possibility that can be exploited for coding 3D video sequences. Due to binocular suppression theory, asymmetric coding may provide similar perceived 3D quality with a significant decrease in bit rate. Several papers propose asymmetric coding schemes [11,34–37] where one of the views is significantly more coarsely quantized than the other, or is coded with a reduced spatial resolution, generating blurring at the upsampling procedure.

In [38], subjective experiments showed that in the asymmetric coding case, where one view is coded at very high quality (40 dB) and the other view is coded at any level down to a threshold value of approximately 33 dB, the resulting stereo video is indistinguishable from the symmetric high quality case of both views coded at 40 dB. It was found that when both views are coded above their corresponding thresholds, asymmetric coding is preferable to symmetric coding at the same total bit rate, whereas when one or both views are coded below its threshold, symmetric coding is generally preferable. These thresholds are employed in our JSCC problem formulation described in Section 3.3.

2.4. Scalable asymmetric multiview coding

References [39–41] introduced scalability and asymmetry into MVC. For the scalable coding, we use the spatial scalability mode. As shown in Fig. 1, the primary view and secondary view frames of a GOP are each low-pass filtered and downsampled by a factor of 2 in both directions. These are encoded with MVC, and constitute the base layer MVC bitstream. The enhancement layer representations are generated through upsampling and computing the residual views. These residual views are also encoded by MVC. In our JSCC/UEP formulation, for both the spatially scalable and non-spatially scalable coders, the reconstructed views achieve predetermined PSNR thresholds of 40 dB/33 dB for asymmetric, and 40 dB/40 dB for symmetric video coding.

It is important to note that the predetermined PSNR thresholds are aimed at the final display and there is no PSNR threshold specified for the quality of the base layer. Comparisons of packet loss effects in both the base and enhancement layers are shown in Section 4 and the potential advantage of the scalable scheme over the non-scalable one is discussed in terms of the subjective quality of the received video.

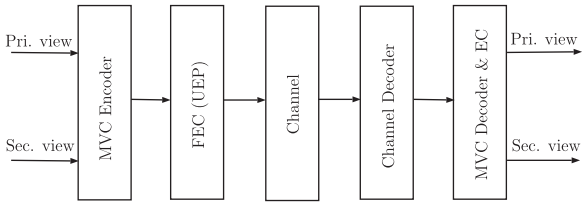

Fig. 2. System block diagram.

3. Problem formulation

The system block diagram is shown in Fig. 2. The primary and secondary views are compressed by an MVC encoder. The encoded bitstream is then protected by FEC and transmitted over an error-prone channel. UEP is utilized, which provides different levels of protection at the packet level through allocating different FEC rates selected from a set of available code rates. At the receiver, channel decoding is applied, where the erroneous packets are detected and not used for display. The primary and secondary views are then decoded by an MVC decoder, where error concealment (EC) is done for the lost packets.

3.1. Modeling the end-to-end distortion

In this section, we model the end-to-end distortion of a GOP of an MVC-encoded 3D video sent over a noisy channel. The MVC extension of the H.264/AVC standard allows encoding the macroblocks into slices for the sake of error resiliency [8], such that each slice can be decoded independently from the others. In this paper, each slice contains a row of macroblocks that includes both the header and payload (i.e. motion vectors, transform coefficients, etc.). We show by simulation that the model is accurate in predicting the actual measured end-to-end distortion for different packet loss ratios. The model is then used in Section 3.2 to derive an estimate of the expected end-to-end distortion. In Section 3.3, we use the estimates for UEP.

We first consider the non-scalable MVC case. Let $f^{(v)}$ represent the original pixel values of view v of a GOP, and $\hat{f}^{(v)}$ be the reconstructed values at the encoder, where $v=1$ represents the primary view and $v=2$ represents the secondary view. We denote the pixel values of view v of the GOP at the decoder as $\tilde{f}^{(v)}$. The distortion of view v is the sum of distortions of all its pixels. It is common in the literature to approximate the source quantization distortion and the channel distortion as being uncorrelated [5,17,42,43]. With this approximation, the expected distortion of view v of the GOP can be written as

$$D^{(v)} = E\{\mathrm{cmse}(f^{(v)}, \hat{f}^{(v)})\} + E\{\mathrm{cmse}(\hat{f}^{(v)}, \tilde{f}^{(v)})\} = D^{(v)}_{Src} + D^{(v)}_{Loss}, \quad (1)$$

where $\mathrm{cmse}(x^{(v)}, y^{(v)})$ is the cumulative mean squared error (CMSE) between the pixels of view v of GOP x and view v of GOP y, $D^{(v)}_{Src}$ represents the source distortion over the entire view v of the GOP, and $D^{(v)}_{Loss}$ denotes the distortion introduced by the channel due to packet losses. In (1), we model the end-to-end distortion such that the source distortion and channel distortion are additive, where the precise value of $D^{(v)}_{Src}$ is computed at the encoder. To compute $D^{(v)}_{Loss}$ for a set of lost packets, we assume that the error signals due to individual losses from either the primary or the secondary view are separate throughout the GOP. For example, if a slice is lost at the top of a frame and another slice is lost at the bottom of that frame (or another frame), the error signals due to the loss of these packets are generally independent. Using this assumption, the CMSE contributions of the individual packets to the CMSE of either the primary view or the secondary view of the GOP are additive. To compute the channel distortion $D^{(v)}_{Loss}$ for a set of lost packets, the model adds up the CMSE values due to the individual lost packets. This additivity assumption is also used for example in [5,44–49]. A CMSE value represents the precise error propagated throughout a view of the GOP, and we assume that it has already been computed offline at the encoder for each packet of the GOP.

Now, we investigate the accuracy of the model in estimating the end-to-end distortion of a GOP. In our experiment, packets of an encoded 3D video are randomly dropped with different PLRs, where 1000 random realizations are done for each PLR. For the non-scalable MVC, for error concealment we implemented linear interpolation for lost I slices, and slice copy for lost P slices, such that a lost P slice is concealed from its reference frame in the same view. Fig. 3(a)–(d) shows histograms of $\Delta\mathrm{PSNR} \triangleq \mathrm{PSNR}_m - \mathrm{PSNR}_{est}$ for PLRs 0.5% and 2% for the sequence 'Oldtimers', where PSNR_m is the actual PSNR measured at the receiver (that is computed between the original uncompressed video and the lossy decoded video) and PSNR_{est} is computed by the model. The model computes the end-to-end distortion using (1) that accounts for both the source distortion and the channel distortion. Histograms similar to the ones depicted in Fig. 3 were obtained for the video sequence 'Race'. We see that the model is accurate in estimating the end-to-end distortion if few packets of a GOP are lost in transmission. If the channel gets bad such that the number of losses after the channel decoder becomes large, the accuracy of the model decreases. However, our JSCC scheme allows us to add as many parity bits as needed to meet the quality constraints for a bad channel condition.

The model adopted for non-scalable MVC can also be used for estimating the end-to-end distortion of scalable MVC. That is, we assume that the source distortion and channel distortion are additive, and that the CMSE contributions of lost packets are additive. This can again be verified by realizing many channel realizations and different PLRs. For scalable MVC, error concealment was implemented such that, when a BL packet is lost, frame copying is used for the BL, and EL information is preserved (linear interpolation is used for lost I slices of the BL); when an EL packet is lost, an upsampled version of the co-located slice of the BL is used for error concealment [50], and if two co-located BL and EL slices are lost simultaneously, frame copying is used for both. Fig. 3(e)–(h) shows histograms of the errors for PLRs 0.5% and 2% for the sequence 'Oldtimers' for the scalable coder.

Table 1 shows the mean absolute value of ΔPSNR, which is defined as $\overline{|\Delta\mathrm{PSNR}|} = (1/N)\sum_{i=1}^{N}|\Delta\mathrm{PSNR}_i|$, where N is the

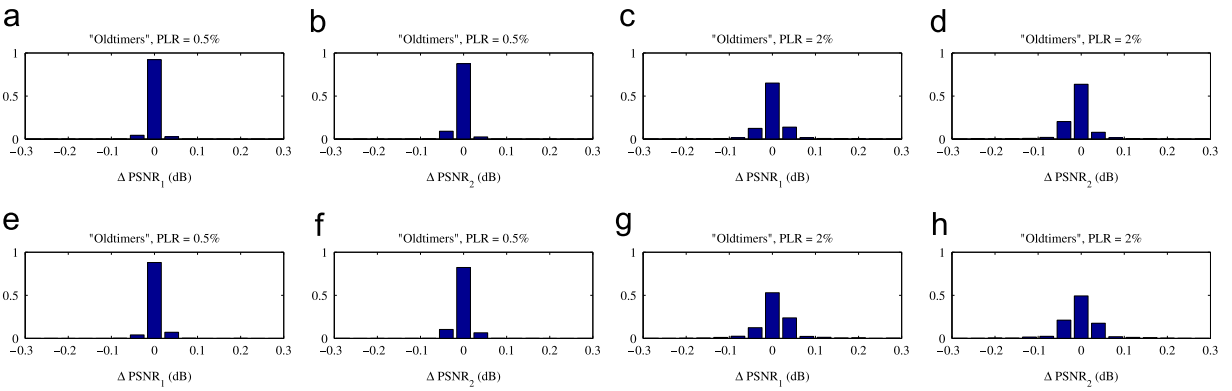

Fig. 3. Histograms of error $\Delta\text{PSNR} \triangleq \text{PSNR}_m - \text{PSNR}_{est}$ for packet loss ratios 0.5% and 2% and video sequence 'Oldtimers'. (a), (b), (c), and (d) Non-scalable MVC and (e), (f), (g), and (h) scalable MVC. ΔPSNR_1 and ΔPSNR_2 correspond to primary and secondary views respectively.

Table 1
Mean absolute value of $\Delta\text{PSNR} \triangleq \text{PSNR}_m - \text{PSNR}_{est}$ in dB for packet loss ratios 0.5%, 1%, 2%, and 5%.

Video name	PLR (%)	Non-scalable		Scalable	
		$\overline{\lvert\Delta\text{PSNR}\rvert}_{\text{Pri}}$	$\overline{\lvert\Delta\text{PSNR}\rvert}_{\text{Sec}}$	$\overline{\lvert\Delta\text{PSNR}\rvert}_{\text{Pri}}$	$\overline{\lvert\Delta\text{PSNR}\rvert}_{\text{Sec}}$
Oldtimers	0.5	0.007	0.01	0.009	0.013
	1	0.016	0.017	0.017	0.023
	2	0.028	0.029	0.034	0.047
	5	0.056	0.067	0.059	0.092
Race	0.5	0.017	0.018	0.029	0.030
	1	0.036	0.045	0.079	0.073
	2	0.089	0.105	0.142	0.150
	5	0.226	0.278	0.237	0.239

number of realizations. The small $\overline{|\Delta\text{PSNR}|}$ values indicate that the model is accurate in estimating the measured PSNR values at the receiver.

We also investigate the accuracy of the additivity approximation for particular packet loss patterns where two packets are lost at the same time. We consider the following packet loss patterns: (1) packets which are in adjacent rows within the same frame, (2) packets which are located in the same row in different frames (spaced apart from 1 to $N_F - 1$ frames, where N_F is the number of frames in a view of a GOP), and (3) all other possible combinations of two packets. Fig. 4(a) shows a histogram of all possible adjacent combinations, which comprise 0.6% of all possible combinations, Fig. 4(b) depicts a histogram of all combinations of two packets located in the same row but in different frames, which comprise 6% of all the possible combinations, and Fig. 4(c) is a histogram of all the other combinations except the ones mentioned above, which comprise 93.4% of all the possible combinations. We observe that the model is not very accurate for some combinations of packet loss patterns (1) and (2). On the other hand, we observe that the model is highly accurate for all other combinations. These observations show that the model is inaccurate only for adjacent losses and losses from the same row (which together comprise a small percentage of all possible combinations) because in such cases, the propagated errors may affect each other.

3.2. Expected end-to-end distortion

In this section, we derive an estimate of the expected end-to-end distortion of a GOP of an MVC-encoded video sent over a noisy channel using the model developed in Section 3.1.

3.2.1. Non-scalable MVC

In the following, $D_{m,v}^{(v')}$ denotes the CMSE contribution of the mth packet of view $v \in \{1,2\}$ to the CMSE of view $v' \in \{1,2\}$, and $\tilde{f}_{m,v}^{(v')}$ represents the reconstructed view v' of GOP at the decoder when the mth packet of view v is lost. Also, $P_i^{(v)}$ is the probability that the ith packet of view v is lost in transmission.

The CMSE contribution of the ith packet of the primary view to the CMSE of the primary view is zero if the packet is not lost, and is equal to $D_{i,1}^{(1)}(q_1)$ if the packet is lost, where q_1 is the quantization parameter used to encode the primary view of the GOP. Thus, following the model assumptions, the average end-to-end distortion of the primary view can be estimated as[1]

$$D^{(1)}(q_1, r_1^{(1)}, \ldots, r_K^{(1)}, \Theta) = D_{Src}^{(1)}(q_1) + \sum_{i=1}^{K} P_i^{(1)}\left(r_i^{(1)}, S_i^{(1)}(q_1), \Theta\right) D_{i,1}^{(1)}(q_1), \tag{2}$$

where K is the number of primary view packets in a GOP (which is the same as the number of secondary view packets in the GOP), and $D_{i,1}^{(1)}(q_1)$ is equal to $\text{cmse}\left(\hat{f}^{(1)}(q_1), \tilde{f}_{i,1}^{(1)}(q_1)\right)$. Packet loss probability $P_i^{(1)}$ depends on the packet size $S_i^{(1)}$ in bits, the code rate $r_i^{(1)}$ by which the packet is protected, and Θ, which represents the channel characteristics; $\Theta = \text{SNR}$ for an AWGN channel and $\Theta = (\text{SNR}, T_c)$ for a flat Rayleigh fading channel, where T_c is the channel coherence time, defined in Section 4. In this work, the quantity $D_{i,1}^{(1)}(q_1)$ is computed at the encoder.

[1] It is assumed that the coded packets are lost independently (a coded packet is composed of the source bits and parity bits). This assumption holds for an AWGN channel. For flat Rayleigh fading channels, independent losses within a GOP are obtained for an archival video by interleaving GOPs such that each interleaved block contains at most one packet from a particular GOP.

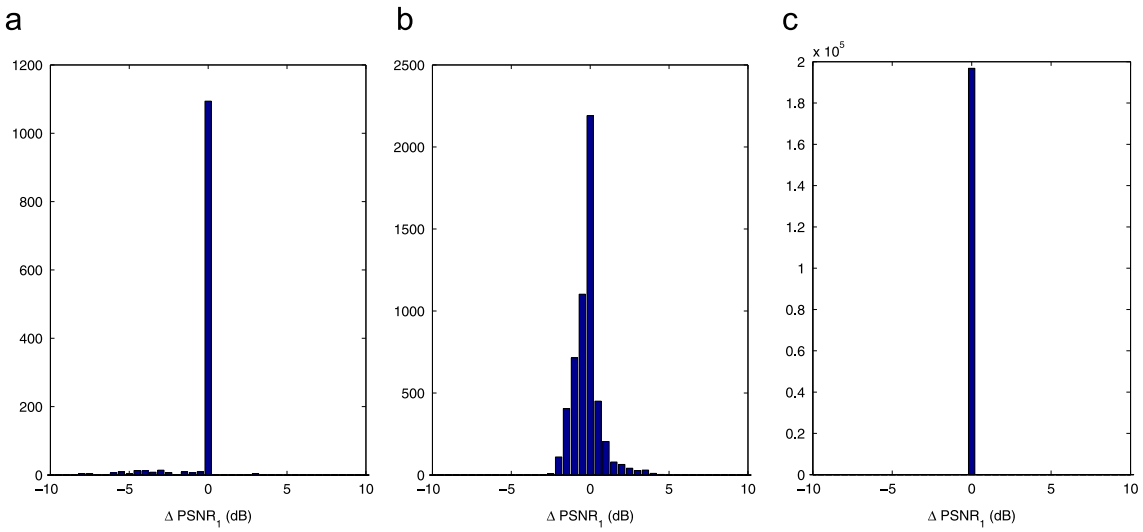

Fig. 4. Histograms of error $\Delta\text{PSNR} \triangleq \text{PSNR}_m - \text{PSNR}_{est}$ for two packets lost in a GOP (see text for description).

In addition, since there is no closed-form expression to compute the packet loss probabilities, $P(r,S,\Theta)$, for the RCPT codes, a lookup table is made by simulation, which yields $P(r,S,\Theta)$ for different ranges of packet sizes. The probability $P(r,S,\Theta)$ is obtained for packet sizes 250, 750, 1500, 2500, 3500, and 5000 in bits, and respectively used for all the packet sizes in the ranges $[0,500)$, $[500,1000)$, $[1000,2000)$, $[2000,3000)$, $[3000,4000)$, and $[4000,\infty)$.

The distortion generated in the secondary view can be formulated in a similar manner. However, since the error due to a lost packet in the primary view propagates in both the primary and secondary views, for the secondary view, the CMSE contribution of lost primary packets should be considered as well as the CMSE contribution of lost secondary packets. Therefore, the average end-to-end distortion of the secondary view can be estimated as

$$\begin{aligned} D^{(2)}(q_1,q_2,r_1^{(1)},\ldots,r_K^{(1)},r_1^{(2)},\ldots,r_K^{(2)},\Theta) &= D_{Src}^{(2)}(q_1,q_2) \\ &+ \sum_{i=1}^{K} P_i^{(1)}\left(r_i^{(1)},S_i^{(1)}(q_1),\Theta\right)D_{i,1}^{(2)}(q_1,q_2) \\ &+ \sum_{j=1}^{K} P_j^{(2)}\left(r_j^{(2)},S_j^{(2)}(q_1,q_2),\Theta\right)D_{j,2}^{(2)}(q_1,q_2), \end{aligned} \tag{3}$$

where $D_{i,1}^{(2)}(q_1,q_2)=\text{cmse}\left(\hat{f}^{(2)}(q_1,q_2),\tilde{f}_{i,1}^{(2)}(q_1,q_2)\right)$ and $D_{j,2}^{(2)}(q_1,q_2)=\text{cmse}\left(\hat{f}^{(2)}(q_1,q_2),\tilde{f}_{j,2}^{(2)}(q_1,q_2)\right)$. The quantities $D_{i,1}^{(2)}(q_1,q_2)$ and $D_{j,2}^{(2)}(q_1,q_2)$ are computed at the encoder and used in the simulations.[2]

3.2.2. Scalable MVC

Similar to the non-scalable MVC, an estimate of the expected end-to-end distortion of the primary view for the scalable MVC case is given by

$$\begin{aligned} D^{(1)}(q_{\text{BL}_1},r_1^{(\text{BL}_1)},\ldots,r_{K/2}^{(\text{BL}_1)},q_{\text{EL}_1},r_1^{(\text{EL}_1)},\ldots,r_K^{(\text{EL}_1)},\Theta) \\ = D_{Src}^{(1)}(q_{\text{BL}_1},q_{\text{EL}_1}) + \sum_{i=1}^{K/2} P_i^{(\text{BL}_1)}\left(r_i^{(\text{BL}_1)},S_i^{(\text{BL}_1)}(q_{\text{BL}_1}),\Theta\right)D_{i,\text{BL}_1}^{(1)}(q_{\text{BL}_1},q_{\text{EL}_1}) \\ + \sum_{j=1}^{K} P_j^{(\text{EL}_1)}\left(r_j^{(\text{EL}_1)},S_j^{(\text{EL}_1)}(q_{\text{BL}_1},q_{\text{EL}_1}),\Theta\right)D_{j,\text{EL}_1}^{(1)}(q_{\text{BL}_1},q_{\text{EL}_1}). \end{aligned} \tag{4}$$

For the secondary view, we have

$$\begin{aligned} D^{(2)}(q_{\text{BL}_1},r_1^{(\text{BL}_1)},\ldots,r_{K/2}^{(\text{BL}_1)},q_{\text{EL}_1},r_1^{(\text{EL}_1)},\ldots,r_K^{(\text{EL}_1)}, \\ q_{\text{BL}_2},r_1^{(\text{BL}_2)},\ldots,r_{K/2}^{(\text{BL}_2)},q_{\text{EL}_2},r_1^{(\text{EL}_2)},\ldots,r_K^{(\text{EL}_2)},\Theta) = D_{Src}^{(2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}) \\ + \sum_{i=1}^{K/2} P_i^{(\text{BL}_1)}\left(r_i^{(\text{BL}_1)},S_i^{(\text{BL}_1)}(q_{\text{BL}_1}),\Theta\right)D_{i,\text{BL}_1}^{(2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}) \\ + \sum_{j=1}^{K} P_j^{(\text{EL}_1)}\left(r_j^{(\text{EL}_1)},S_j^{(\text{EL}_1)}(q_{\text{BL}_1},q_{\text{EL}_1}),\Theta\right)D_{j,\text{EL}_1}^{(2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}) \\ + \sum_{m=1}^{K/2} P_m^{(\text{BL}_2)}\left(r_m^{(\text{BL}_2)},S_m^{(\text{BL}_2)}(q_{\text{BL}_1},q_{\text{BL}_2}),\Theta\right)D_{m,\text{BL}_2}^{(2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}) \\ + \sum_{n=1}^{K} P_n^{(\text{EL}_2)}\left(r_n^{(\text{EL}_2)},S_n^{(\text{EL}_2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}),\Theta\right) \\ D_{n,\text{EL}_2}^{(2)}(q_{\text{BL}_1},q_{\text{EL}_1},q_{\text{BL}_2},q_{\text{EL}_2}), \end{aligned} \tag{5}$$

where in (4) and (5) $D_{t,l}^{(v')}$ denotes the CMSE contribution of the tth packet of layer $l\in\{\text{BL}_1,\text{BL}_2,\text{EL}_1,\text{EL}_2\}$ to the CMSE of view $v'\in\{1,2\}$, and $P_t^{(l)}$ is the probability that the tth packet of layer l is lost in transmission. In (4) and (5), $D_{i,l}^{(v)}=\text{cmse}(\hat{f}^{(v)},\tilde{f}_{i,l}^{(v)})$, where $\tilde{f}_{i,l}^{(v)}$ represents the reconstructed view v of a GOP at the decoder when the ith packet of layer l is lost, and $\hat{f}^{(v)}$ denotes the reconstructed view v of the GOP when there are no packet losses. The quantity $D_{i,l}^{(v)}$ is computed at the encoder.

3.3. JSCC of 3D video for non-scalable and scalable MVC

The objective of our JSCC problem is to minimize the total number of bits, which is the sum of the numbers of

[2] Computing the distortion values at the encoder side requires decoding the whole GOP for each slice of the GOP. The computational complexity of our algorithm at the encoder side is high and it can be done offline.

source bits and FEC bits of both the primary and secondary views. For the non-scalable MVC, the objective function is formulated as

$$\min_{\substack{q_1 \in QP_1 \\ q_2 \in QP_2 \\ r_1^{(1)},\ldots,r_K^{(1)} \in R \\ r_1^{(2)},\ldots,r_K^{(2)} \in R}} \left(\sum_{i=1}^{K} \frac{S_i^{(1)}(q_1)}{r_i^{(1)}} + \sum_{j=1}^{K} \frac{S_j^{(2)}(q_1,q_2)}{r_j^{(2)}} \right), \tag{6}$$

where $R=\{R_1,R_2,\ldots,R_N\}$ is the set of available RCPT code rates, and QP_1 and QP_2 are the sets of quantization parameters. The optimization is done jointly over a primary view and the corresponding secondary view of a GOP. Quantization parameters q_1 and q_2 are applied for all macroblocks of the primary and secondary views of a GOP.

In minimizing the objective function (6), quality constraints must be satisfied. For the symmetric coding case, we require that the expected distortions of the primary and secondary views be less than or equal to a predetermined threshold T_1 at the receiver. However, for the asymmetric coding case, we require that the expected distortion of the primary view and the expected distortion of the secondary view be less than or equal to two different predetermined thresholds, T_1 and T_2. Using (2) and (3), this can be expressed for both the symmetric and asymmetric coding cases as

$$\begin{aligned} &D^{(1)}(q_1,r_1^{(1)},\ldots,r_K^{(1)},\Theta) \le T_1 \\ &D^{(2)}(q_1,q_2,r_1^{(1)},\ldots,r_K^{(1)},r_1^{(2)},\ldots,r_K^{(2)},\Theta) \le T_2, \end{aligned} \tag{7}$$

where for the symmetric coding case $T_1=T_2=10^{(-40\ \mathrm{dB}/10)}$, and for the asymmetric coding case $T_1=10^{(-40\ \mathrm{dB}/10)}$ and $T_2=10^{(-33\ \mathrm{dB}/10)}$. The reason for choosing the particular PSNR values 40 dB and 33 dB is explained in Section 2.

The objective function of the JSCC problem for the scalable MVC can be formulated as was done for the non-scalable case. For scalable MVC, we have

$$\begin{aligned} \min_{\substack{(q_{\mathrm{BL}_1},q_{\mathrm{BL}_2},q_{\mathrm{EL}_1},q_{\mathrm{EL}_2}) \in QP_s \\ r_i^{(\mathrm{BL}_1)},\ r_j^{(\mathrm{BL}_2)},\ r_m^{(\mathrm{EL}_1)},\ r_n^{(\mathrm{EL}_2)} \in R}} \Bigg(&\sum_{i=1}^{K/2} \frac{S_i^{(\mathrm{BL}_1)}(q_{\mathrm{BL}_1})}{r_i^{(\mathrm{BL}_1)}} + \sum_{j=1}^{K} \frac{S_j^{(\mathrm{EL}_1)}(q_{\mathrm{BL}_1},q_{\mathrm{EL}_1})}{r_j^{(\mathrm{EL}_1)}} \\ &+ \sum_{m=1}^{K/2} \frac{S_m^{(\mathrm{BL}_2)}(q_{\mathrm{BL}_1},q_{\mathrm{BL}_2})}{r_m^{(\mathrm{BL}_2)}} + \sum_{n=1}^{K} \frac{S_n^{(\mathrm{EL}_2)}(q_{\mathrm{BL}_1},q_{\mathrm{EL}_1},q_{\mathrm{BL}_2},q_{\mathrm{EL}_2})}{r_n^{(\mathrm{EL}_2)}} \Bigg), \end{aligned} \tag{8}$$

where QP_s is a set of 4-tuple quantization parameters.

Similar to the non-scalable MVC case, two constraints must be satisfied in minimizing the objective function (8). Using (4) and (5), these constraints are written as

$$\begin{aligned} &D^{(1)}(q_{\mathrm{BL}_1},r_1^{(\mathrm{BL}_1)},\ldots,r_{K/2}^{(\mathrm{BL}_1)},q_{\mathrm{EL}_1},r_1^{(\mathrm{EL}_1)},\ldots,r_K^{(\mathrm{EL}_1)},\Theta) \le T_1 \\ &D^{(2)}(q_{\mathrm{BL}_1},r_1^{(\mathrm{BL}_1)},\ldots,r_{K/2}^{(\mathrm{BL}_1)},q_{\mathrm{EL}_1},r_1^{(\mathrm{EL}_1)},\ldots,r_K^{(\mathrm{EL}_1)}, \\ &\quad q_{\mathrm{BL}_2},r_1^{(\mathrm{BL}_2)},\ldots,r_{K/2}^{(\mathrm{BL}_2)},q_{\mathrm{EL}_2},r_1^{(\mathrm{EL}_2)},\ldots,r_K^{(\mathrm{EL}_2)},\Theta) \le T_2, \end{aligned} \tag{9}$$

where for the symmetric coding case $T_1=T_2=10^{(-40\ \mathrm{dB}/10)}$, and for the asymmetric coding case $T_1=10^{(-40\ \mathrm{dB}/10)}$ and $T_2=10^{(-33\ \mathrm{dB}/10)}$.

In the two optimization problems introduced in (6)–(9), different code rates are typically assigned to different packets. The code rate assigned to a particular packet depends on (1) the size of the source packet as determined by the quantization parameters q_1 and q_2 for the non-scalable MVC source encoder, and by q_{BL_1}, q_{BL_2}, q_{EL_1}, and q_{EL_2} for the scalable MVC source encoder, (2) the distortion the packet generates if it is lost in transmission, and (3) the probability that the packet is lost, which depends on channel characteristics specified by Θ. To find the quantization parameters and code rates that minimize the objective functions (6) and (8), we search over a grid of QPs, where for the non-scalable MVC the search is done over a two-dimensional grid specified by vector (q_1,q_2) and for the scalable MVC is done over a four-dimensional grid specified by vector $(q_{\mathrm{BL}_1},q_{\mathrm{BL}_2},q_{\mathrm{EL}_1},q_{\mathrm{EL}_2})$. The solution is obtained as a quantization vector in the grid and a set of code rates, which together produce the smallest total number of bits and, at the same time, meet the quality constraints. The sets QP_1 and QP_2 in (6), and QP_s in (8), are determined for each GOP of a given video sequence. To do that, for the MVC-non-scalable case, we first perform a binary search over q_1 and q_2 to rule out the QPs for which the noise-free encoded video does not meet the quality constraints, and find the largest possible QPs, $q_1^{(max)}$ and $q_2^{(max)}$, that satisfy the constraints. The ruled out QPs are not considered for optimization since they do not meet the quality constraints even in the absence of channel distortion. The sets QP_1 and QP_2 are then defined as the sets whose members are QPs less than or equal to $q_1^{(max)}$ and $q_2^{(max)}$, respectively. For the scalable case, we perform an exhaustive search over the QPs q_{BL_1}, q_{BL_2}, q_{EL_1}, and q_{EL_2}, to rule out the ones for which the noise-free encoded video does not meet the quality constraints.

3.4. Integer optimization

The optimization problems introduced in Section 3.3 are nonlinear integer programming problems, which can be solved by the branch-and-bound (BnB) method [51]. The BnB method is based on binary variables [51]. For the non-scalable MVC case, we transform each variable $r_i^{(1)}$ to N binary variables $x_{i,l}$ $(1\le l\le N)$, and each variable $r_j^{(2)}$ to N binary variables $y_{j,l}$, where x and y take values from the set $\{0,1\}$. We then substitute $r_i^{(1)}$ with $\sum_{l=1}^{N} x_{i,l}R_l$ and $r_j^{(2)}$ with $\sum_{l=1}^{N} y_{j,l}R_l$ in (6) and (7). With these transformations, $2K$ equality constraints are considered along with the inequalities given in (7), which are

$$\begin{aligned} &\sum_{l=1}^{N} x_{i,l} = 1, \quad 1\le i\le K \\ &\sum_{l=1}^{N} y_{j,l} = 1, \quad 1\le j\le K. \end{aligned} \tag{10}$$

Now, we consider the scalable MVC case. We transform each variable $r_i^{(\mathrm{BL}_1)}$ to N binary variables $x_{i,l}$ $(1\le l\le N)$, each variable $r_j^{(\mathrm{BL}_2)}$ to N binary variables $y_{j,l}$, each variable $r_m^{(\mathrm{EL}_1)}$ to N binary variables $z_{m,l}$, and each variable $r_n^{(\mathrm{EL}_2)}$ to N binary variables $t_{n,l}$, where x, y, z, and t take values from the set $\{0,1\}$. We then make the following substitutions in (8) and (9): $r_i^{(\mathrm{BL}_1)}$ is substituted with $\sum_{l=1}^{N} x_{i,l}R_l$, $r_j^{(\mathrm{BL}_2)}$ is substituted with $\sum_{l=1}^{N} y_{j,l}R_l$, $r_m^{(\mathrm{EL}_1)}$ is substituted with $\sum_{l=1}^{N} z_{m,l}R_l$, and $r_n^{(\mathrm{EL}_2)}$ is substituted with $\sum_{l=1}^{N} t_{n,l}R_l$. From these transformations, $3K$ equality constraints must be considered in conjunction with the inequalities in (9),

which are

$$\sum_{l=1}^{N} x_{i,l} = 1, \quad 1 \le i \le \frac{K}{2}$$
$$\sum_{l=1}^{N} y_{j,l} = 1, \quad 1 \le j \le \frac{K}{2}$$
$$\sum_{l=1}^{N} z_{m,l} = 1, \quad 1 \le m \le K$$
$$\sum_{l=1}^{N} t_{n,l} = 1, \quad 1 \le n \le K. \tag{11}$$

4. Simulation results and discussion

Simulation results for AWGN and flat Rayleigh fading channels are given in this section. Binary phase shift keying (BPSK) modulation/demodulation is employed for data transmission over the channel. Samples of the received signal, at a given signal-to-noise ratio E_b/N_0, can be represented by $y = \alpha x + n$, where E_b is energy-per-bit, N_0 is the one-sided power spectral density of the noise, n is a zero-mean Gaussian random variable with standard deviation $\sqrt{N_0/2E_b}$, and $x \in \{-1, 1\}$. For an AWGN channel, α is unity, and for a Rayleigh fading channel, α has Rayleigh distribution with $E\{\alpha^2\} = 1$. The coherence time of a fading channel, T_c, represents the number of symbols affected by the same fade level, and assuming a block-fading channel, each fade is considered to be independent of the others. An interleaver is used to mitigate the effect of error bursts due to the fading channels, and we used a fixed size block interleaver with depth 500 and width 100.

Results are presented for two video sequences, 'Race' and 'Oldtimers', with resolution 640×480, where 'Race' is a high-motion video that contains moving objects and camera panning, and 'Oldtimers' is low-motion. We used the JM 18.2 reference software (stereo profile) for encoding the sequences, where each row of macroblocks of either the primary or secondary view is encoded as a slice. We used the JMVC 8.2 reference software for decoding the MVC bitstream. The primary view frames of a GOP are coded as IPPP…, the secondary view frames are coded as PPPP…, and the GOP size is 20 frames.

We used turbo codes for channel coding. The turbo encoder is composed of two recursive systematic convolutional encoders with constraint length 4, which are concatenated in parallel [52]. The feedforward and feedback generators are 15 and 13, respectively, both in octal. The mother code rate of the RCPT code is $\frac{1}{3}$, the puncturing period $P=8$, and the set of available rates is $\{P/(P+l) | l = 1, 2, \ldots, 2P\}$. An iterative soft-input/soft-output (SISO) decoding algorithm is used for turbo decoding. We considered eight iterations to compute the decoded word error rates.

Fig. 5 shows a scatter plot that depicts how the UEP allocates code rates to different packets of an encoded video. This scatter plot is for 'Race', where the video is encoded by a non-scalable MVC encoder, SNR = 11 dB, and the channel experiences flat Rayleigh fading with $T_c=2000$. Each point on the scatter plot corresponds to a packet that belongs either to the primary or secondary view. The x-axis represents the normalized packet size, and the y-axis represents the inverse of the allocated code rate (higher inverse of code rate corresponds to more protection of the code). For a primary view packet, the z-axis indicates the normalized sum of distortions in the primary and secondary views if that packet is lost, and for a secondary view packet, the z-axis represents the normalized distortion generated in the secondary view if the packet is lost. In this scatter plot, similar levels of distortion are depicted with similar ranges of colors. Packets which generate high distortions are protected with strong codes. These packets are typically the particular slices that generate significant error propagation if they are lost. We also notice that for packets with similar levels of distortion, larger packets are protected less than smaller packets. If two packets generate the same level of distortion, the larger packet might receive less protection than the smaller packet in order to minimize the total number of bits.

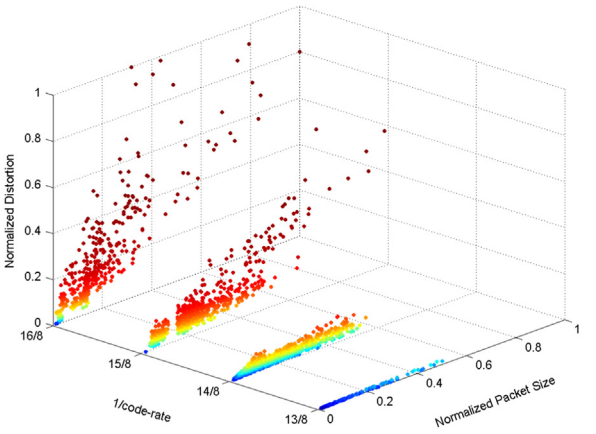

Fig. 5. Scatter plot of the code rates allocated by UEP to different packets of 'Race', where SNR = 10 dB, and $T_c=2000$.

We performed validation tests with many channel realizations to see if the UEP solution obtained using the model and the expected end-to-end distortion estimates (developed in Sections 3.1 and 3.2) meets the quality constraints for realistic channel realizations. Fig. 6 shows the received PSNR histograms of 'Race' and 'Oldtimers', where symmetric coding is considered, and UEP is used for protection of the encoded video over the channel. We see that the average PSNRs meet the specified 40 dB constraints. In Fig. 6(a)–(h), the percentage of received PSNRs which are larger than 40 dB are 87%, 93%, 82%, 82%, 89%, 85%, 88%, 80%.

In the following, we compare the total number of bits required by UEP and EEP for different scenarios. By EEP, we mean that all of the packets are protected by the best single code rate, which is determined by exhaustive simulation over all possible EEP rates. Each scenario is specified by (1) UEP or EEP, (2) channel is AWGN or fading, (3) non-scalable MVC or scalable MVC encoder/decoder is used, and (4) symmetric coding or asymmetric coding is utilized. In all of the comparisons, the percentage of bit savings of scenario A compared to scenario B is defined as

$$e = \frac{\#bits^{(B)} - \#bits^{(A)}}{\#bits^{(B)}} \times 100\%. \tag{12}$$

Fig. 7 shows the results of non-scalable MVC and symmetric coding for 100 frames of video sequences 'Race' and 'Oldtimers', and both AWGN and fading channels. Fig. 7(a), (c), (e),

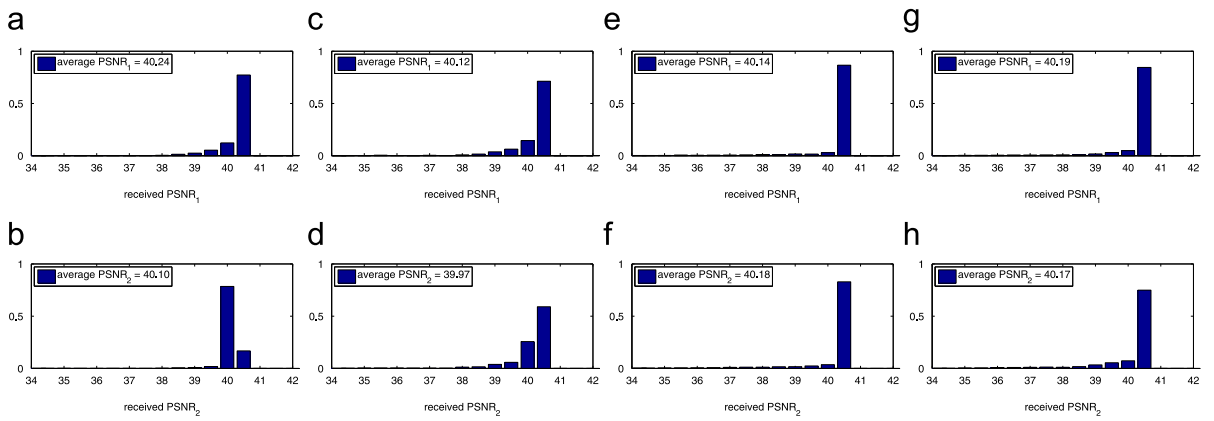

Fig. 6. Received PSNR of the primary view, $PSNR_1$, and the secondary view, $PSNR_2$, for symmetric coding and both non-scalable and scalable MVC. Results are obtained for 2600 channel realizations of the tested SNR values and T_c's. (a) Oldtimers, non-scalable, (b) Oldtimers, non-scalable, (c) Oldtimers, scalable, (d) Oldtimers, scalable, (e) Race, non-scalable, (f) Race, non-scalable, (g) Race, scalable, (h) Race, scalable.

and (g) shows the total number of bits, and Fig. 7(b), (d), (f), and (h) depicts the percentage of UEP bit savings compared to EEP. UEP always requires fewer bits than EEP. We also observe that, as expected, fewer bits are required when the channel SNR increases. For the fading channel, we see that, for a particular SNR, more bits are required for a larger coherence time. This is because when the coherence time gets larger, the diversity order becomes smaller which reduces the capability of a code to protect a packet, and thus, packets need to be protected with a stronger code leading to a larger number of bits. The average gains of UEP over EEP for AWGN and fading channels are 11.6% and 13.4%, respectively, for 'Race', and 13.7% and 16.2% for 'Oldtimers'.

From Fig. 7, we see that the UEP bit savings decrease for higher SNR values, which indicate that the UEP and EEP performances become close for higher SNR values. This is because, as the SNR increases, packets do not need much protection, so both EEP and UEP can use high code rates. For fading channels, we also observe that, for a particular SNR, the UEP bit savings is higher for larger coherence times. As shown in Fig. 7(a) and (e), and discussed above, for a larger coherence time, EEP needs to protect the data with a lower code rate. UEP can flexibly select from many available code rates, leading to a higher bit savings over EEP.

Comparing the 'Race' and 'Oldtimers' results reveals that the number of bits required for 'Race' (high-motion content) is always higher than that of 'Oldtimers' (low-motion content), which is expected. An interesting observation is that the percentage of bit savings of UEP is slightly higher for 'Oldtimers' compared to 'Race'. For low-motion video, there are fewer packets that should be protected using the strong code rates, and these are the ones that contain high motion and their error propagation cannot be concealed efficiently. A larger number of low-motion video packets can be protected with weak codes, which are the ones that belong to the static background or very low-motion regions. These packets generate an insignificant amount of distortion if they are lost in transmission, since their error propagation can be efficiently concealed.

Now we present the results for scalable MVC and symmetric coding. Fig. 8(a) and (b) shows the number of bits required by UEP and EEP, and Fig. 8(c) and (d) illustrate the percentage of bit savings of UEP compared to EEP for fading channels. Comparing the results of Figs. 7 and 8, we see that all the observations made for the non-scalable case are also made for the scalable case. The average gains of UEP over EEP for the fading channels are 17.5% and 19.5% for 'Race' and 'Oldtimers', respectively.

So far, we have presented results for symmetric coding. Asymmetric coding results are presented in Fig. 9 for non-scalable MVC, and in Fig. 8(e) and (f) for scalable MVC. The percentages of bit savings of UEP over EEP are comparable to the symmetric coding case.

Fig. 10 compares the results of symmetric and asymmetric coding for non-scalable video. In this figure, the percentage of bit savings of asymmetric/UEP is compared to both symmetric/UEP and symmetric/EEP. The average gain of asymmetric/UEP over symmetric/UEP and symmetric/EEP for fading channels is 36.4% and 45.2% for 'Race', and 36.8% and 47.1% for 'Oldtimers', respectively. For AWGN, the average gains are 38.3% and 45.4% for 'Race', and 36.1% and 45.0% for 'Oldtimers', respectively. We made similar comparisons between scalable/asymmetric and scalable/symmetric and obtained similar results.

We tested our JSCC/UEP scheme on three more video sequences: 'Ballroom', 'Akko & Kayo', and 'Mobile'. The profiles of the results for these video sequences (not shown) were similar to the ones presented for video sequences 'Race' and 'Oldtimers'. The average percentages of bit savings of UEP compared to EEP for symmetric coding and flat Rayleigh fading were 11.5%, 15.6%, and 14.1% for 'Ballroom', 'Akko & Kayo', and 'Mobile', respectively. The average percentages of bit savings of asymmetric coding/UEP compared to symmetric coding/EEP were 45.6%, 40.3%, and 38.9% for these video sequences, respectively.

It is also interesting to compare the performance of non-scalable and scalable scenarios to see how much overhead (coding inefficiency) is caused by scalability. By comparing the non-scalable and scalable results, we observe that the number of required bits for scalable MVC is always higher than that of non-scalable MVC. Fig. 11 depicts the percentage of overhead of scalable MVC compared to non-scalable MVC for 'Race' and

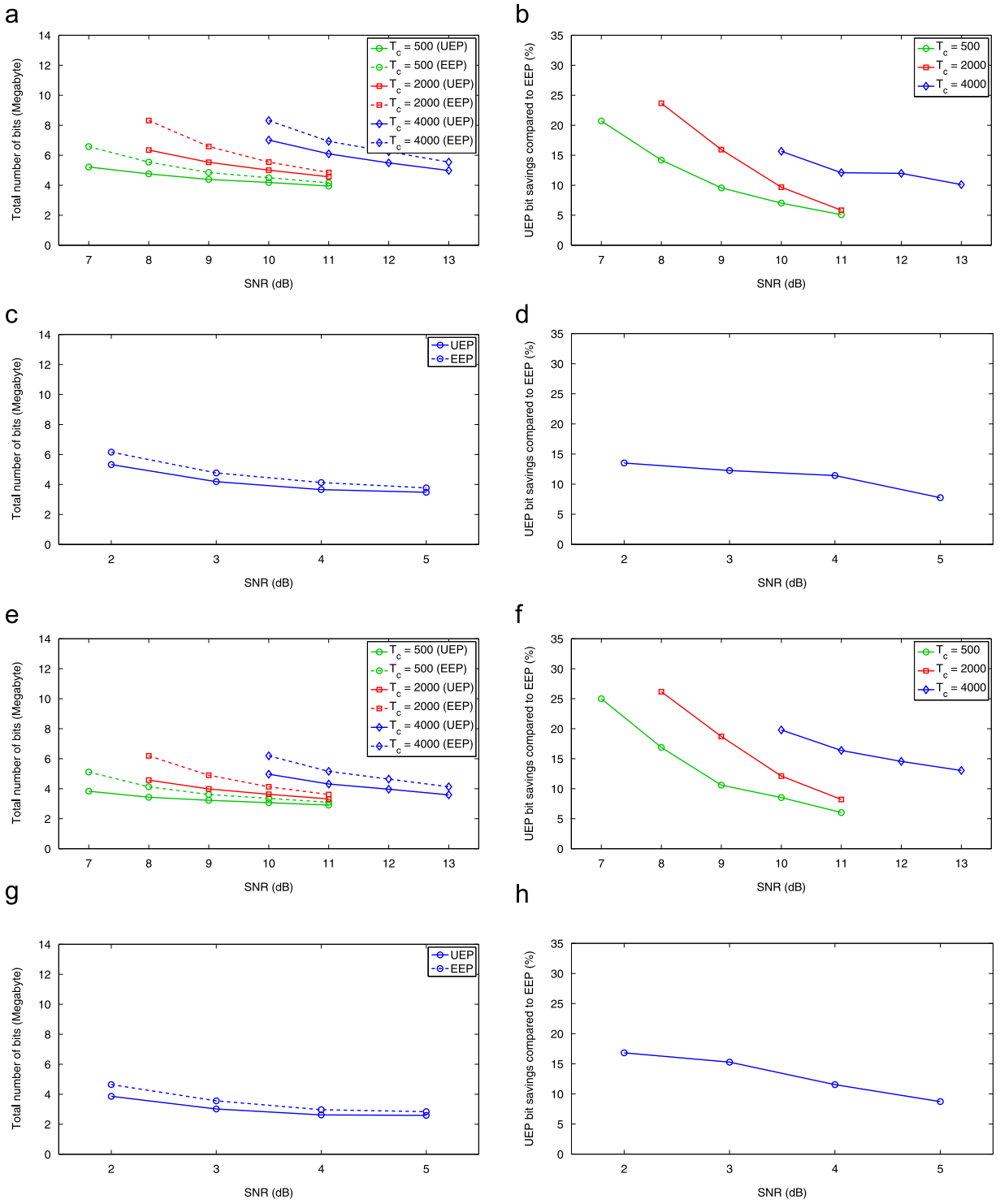

Fig. 7. Results for non-scalable MVC, symmetric coding, and AWGN and fading channels. Total number of bits required by UEP and EEP: (a) Race, fading, (c) Race, AWGN, (e) Oldtimers, fading, (g) Oldtimers, AWGN. Percentage of bit savings of UEP compared to EEP: (b) Race, fading, (d) Race, AWGN, (f) Oldtimers, fading, (h) Oldtimers, AWGN.

'Oldtimers', and for symmetric coding. Comparable results are obtained for asymmetric coding. Although scalable MVC has an overhead penalty, scalability has an advantage if the subjective quality of the lossy decoded bit stream is considered at the receiver. When a BL packet is lost through transmission, frame copying error concealment is used at the decoder, which

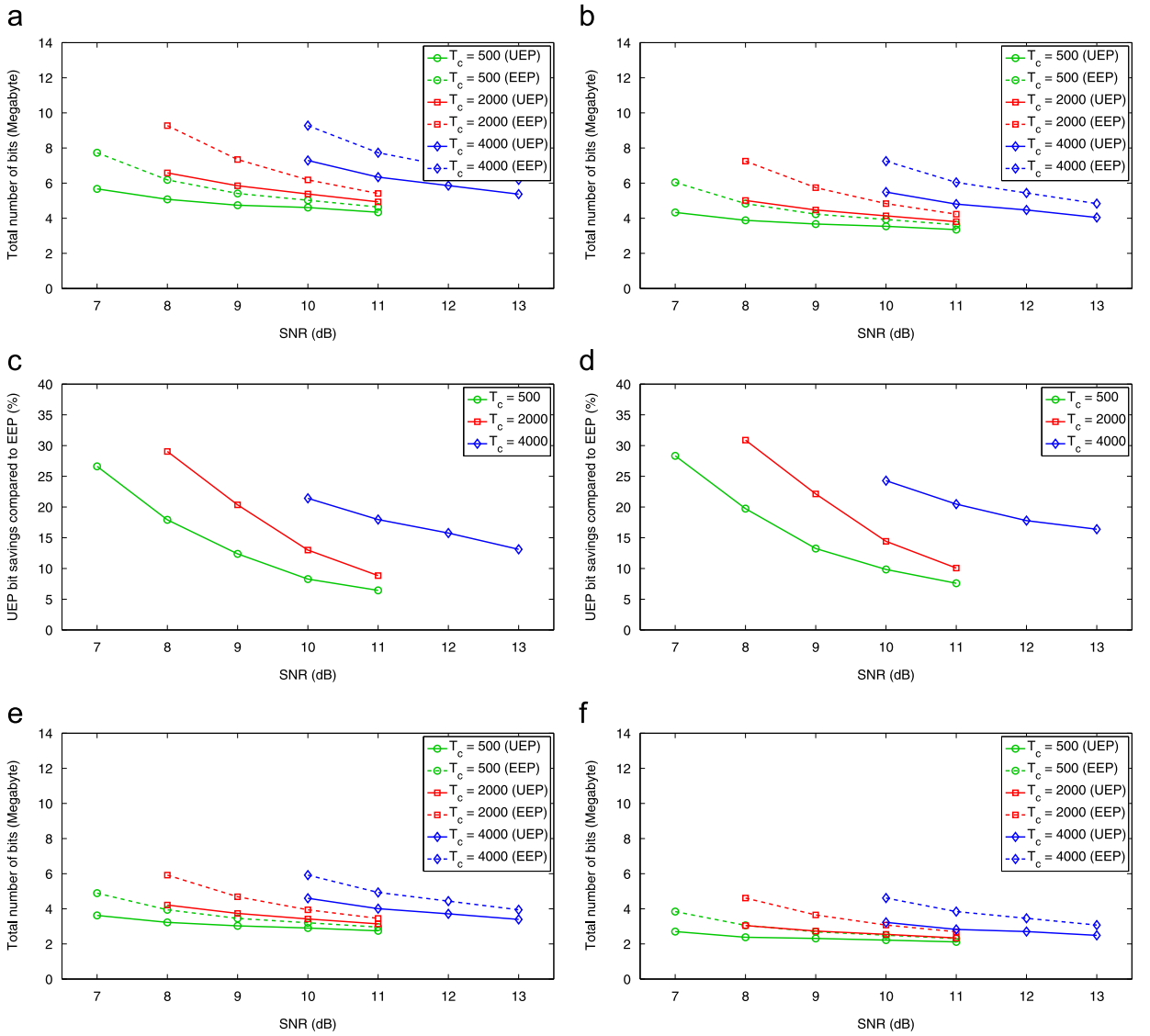

Fig. 8. Results for scalable MVC and fading channels. (a) Race, fading, symmetric, (b) Oldtimers, fading, symmetric, (c) Race, fading, symmetric, (d) Oldtimers, fading, symmetric, (e) Race, fading, asymmetric, (f) Oldtimers, fading, asymmetric.

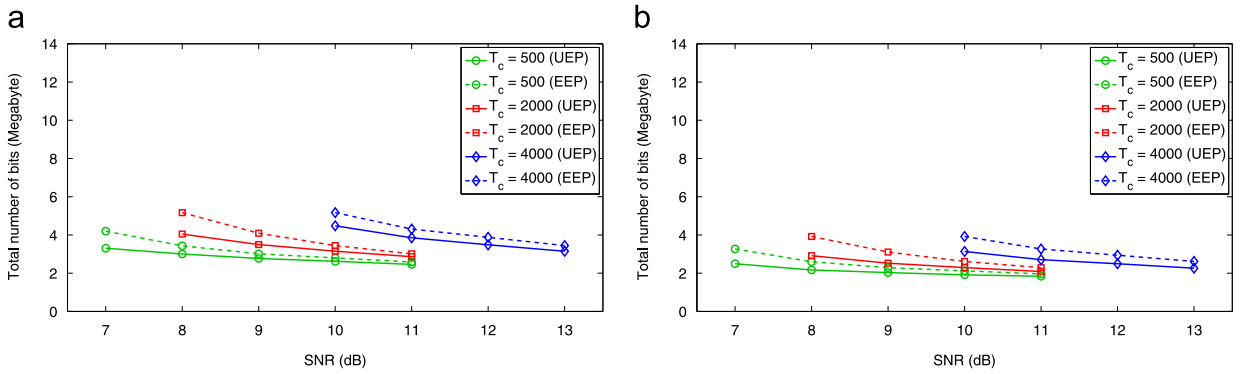

Fig. 9. Results for non-scalable MVC, asymmetric coding, and fading channels. (a) Race, fading, (b) Oldtimers, fading.

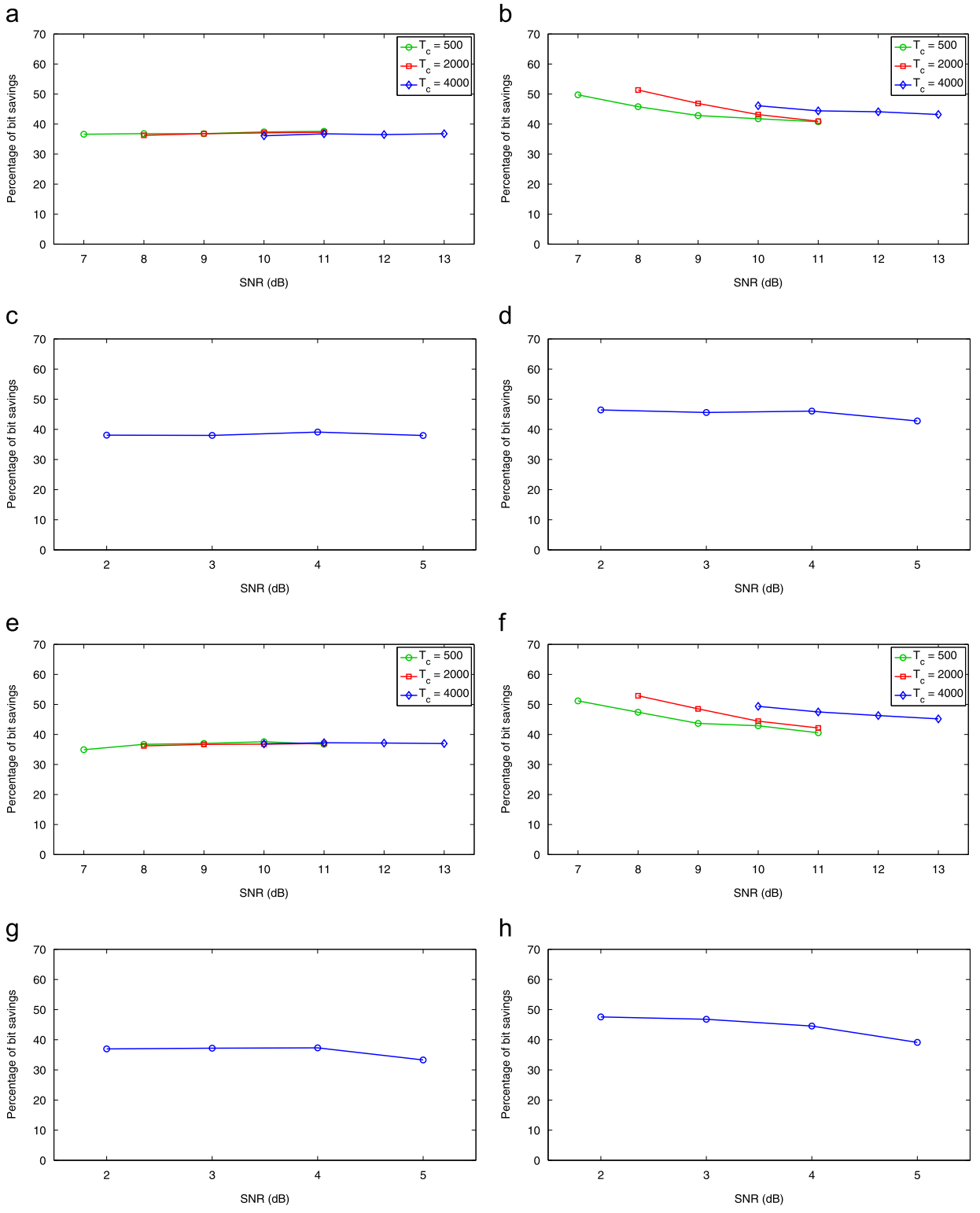

Fig. 10. Percentage of bit savings of asymmetric coding compared to symmetric coding for the non-scalable MVC and both AWGN and fading channels. Bit savings of asymmetric/UEP over symmetric/UEP: (a) Race, fading, (c) Race, AWGN, (e) Oldtimers, fading, (g) Oldtimers, AWGN. Bit savings of asymmetric/UEP over symmetric/EEP: (b) Race, fading, (d) Race, AWGN, (f) Oldtimers, fading, (h) Oldtimers, AWGN.

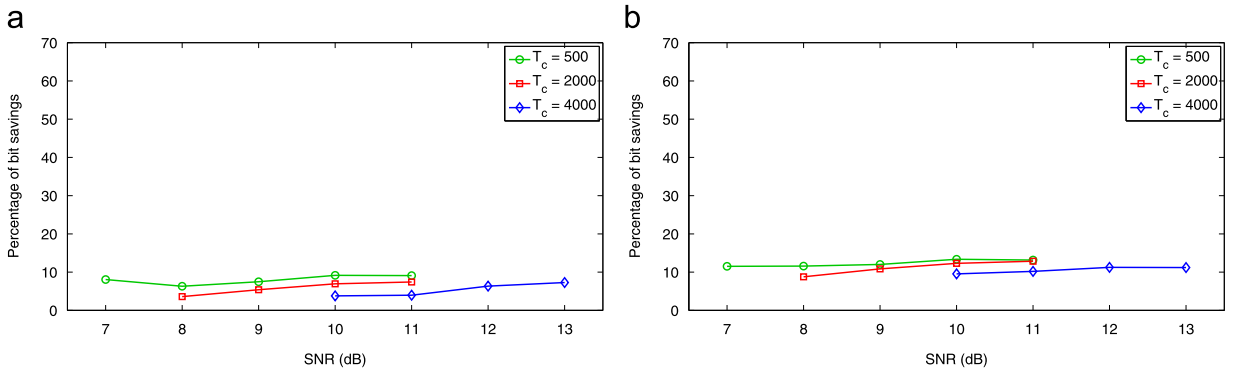

Fig. 11. Percentage of bit savings of the non-scalable MVC compared to the scalable MVC for symmetric/UEP and fading channels. (a) Race, (b) Oldtimers.

generates a noticeable error propagated throughout the GOP [50], specifically for slices possessing high motion content. However, when an EL packet is lost, an upsampled version of the BL is used for error concealment at the decoder, which perhaps causes a less noticeable error [50].

Table 2 shows the percentages of packets that are lost from either the BL, EL, or both layers, for 2600 flat Rayleigh fading channel realizations of all the tested SNR values and coherence times. Results are given for both scalable MVC and non-scalable MVC, where the encoded bit streams are protected using the code rates obtained by the UEP approach. Considering 'Race' for example, we observe that 0.02% and 0.08% of the packets are respectively lost from the BL and the EL. This indicates that the majority of losses occur in the EL, whose errors are concealed more effectively than ones in the BL. In addition, we see that the percentage of BL losses is considerably lower than that of the non-scalable losses. These observations suggest that scalability can perform better than non-scalability in terms of subjective quality.

5. Conclusions

We addressed the joint source–channel coding problem of a 3D video sent over AWGN and fading channels with the goal of minimizing the total number of transmitted bits while subject to video quality constraints. We considered non-scalable MVC and a type of spatial-scalable MVC, and both symmetric and asymmetric coding. The UEP approach proposed here proved to be efficient at achieving this goal when compared to EEP for all the scenarios considered, where the average gains vary from 11.6% to 19.5%. Asymmetric coding was also compared to symmetric coding. Comparable gains were obtained for the non-scalable MVC and scalable MVC, where the asymmetric/UEP gain over symmetric/UEP and symmetric/EEP vary, respectively, from 36.1% to 38.3% and from 45.0% to 47.1%. We also showed that although using scalability leads to an overhead compared to non-scalable MVC, it may have an advantage in terms of the subjective quality of the received video, since most of the lost packets occur in the enhancement layer whose errors are less noticeable to the human visual system compared to the errors due to packets lost in the base layer.

Table 2
Percentage of packet losses of the tested video bitstreams protected by UEP over the flat Rayleigh fading channel.

Video name	Non-scalable (%)	Scalable (%)		
		BL	EL	BL and EL
Race	0.10	0.02	0.08	0.0000
Oldtimers	0.34	0.12	0.24	0.0006

Acknowledgments

This research was supported by the Intel/Cisco Video Aware Wireless Networks (VAWN) program, by InterDigital, Inc., and by the National Science Foundation under Grant number CCF-1160832.

References

[1] R.E. Van Dyck, D.J. Miller, Transport of wireless video using separate, concatenated, and joint source–channel coding, Proc. IEEE 87 (10) (1999) 1734–1750.
[2] M. Bystrom, J.W. Modestino, Combined source–channel coding schemes for video transmission over an additive white Gaussian noise channel, IEEE J. Sel. Areas Commun. 18 (6) (2000) 880–890.
[3] J. Hagenauer, Rate-compatible punctured convolutional codes (RCPC codes) and their applications, IEEE Trans. Commun. 36 (4) (1988) 389–400.
[4] T. Stockhammer, M.M. Hannuksela, T. Wiegand, H.264/AVC in wireless environments, IEEE Trans. Circuits Syst. Video Tech. 13 (7) (2003) 657–673.
[5] A.S. Tan, A. Aksay, G.B. Akar, E. Arikan, Rate-distortion optimization for stereoscopic video streaming with unequal error protection, EURASIP J. Adv. Signal Process. 2009 (2008). 7:1–7:14.
[6] Q. Huynh-Thu, P. Le Callet, M. Barkowsky, Video quality assessment: from 2D to 3D challenges and future trends, in: ICIP, 2010, pp. 4025–4028.
[7] N. Ozbek, A.M. Tekalp, E.T. Tunali, Rate allocation between views in scalable stereo video coding using an objective stereo video quality measure, in: ICASSP, vol. 1, 2007.
[8] A. Vetro, T. Wiegand, G.J. Sullivan, Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard, Proc. IEEE 99 (4) (2011) 626–642.
[9] A. Vosoughi, V. Testoni, P. Cosman, L. Milstein, Joint source–channel coding of 3D video using multiview coding, in: ICASSP, 2013, pp. 2050–2054.
[10] B. Julesz, Foundations of Cyclopean Perception, The University of Chicago Press, Chicago, 1971.
[11] L. Stelmach, W.J. Tam, D. Meegan, A. Vincent, Stereo image quality: effects of mixed spatio-temporal resolution, IEEE Trans. Circuits Syst. Video Tech. 10 (2) (2000) 188–193.

[12] W.A. IJsselsteijn, H. de Ridder, J. Vliegen, Subjective evaluation of stereoscopic images: effects of camera parameters and display duration, IEEE Trans. Circuits Syst. Video Tech. 10 (2) (2000) 225–233.
[13] P. Seuntiens, L. Meesters, W. Ijsselsteijn, Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation, ACM Trans. Appl. Percept. 3 (2) (2006) 95–109.
[14] S. Yasakethu, W. Fernando, B. Kamolrat, A. Kondoz, Analyzing perceptual attributes of 3D video, IEEE Trans. Consum. Electron. 55 (2) (2009) 864–872.
[15] V. De Silva, H. Arachchi, E. Ekmekcioglu, A. Fernando, S. Dogan, A. Kondoz, S. Savas, Psycho-physical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery, in: Packet Video Workshop (PV), 2012 19th International, 2012, pp. 184–189.
[16] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, RD-optimized interactive streaming of multiview video with multiple encodings, J. Vis. Commun. Image Represent. 21 (5) (2010) 523–532.
[17] Y. Zhou, C. Hou, W. Xiang, F. Wu, Channel distortion modeling for multi-view video transmission over packet-switched networks, IEEE Trans. Circuits Syst. Video Tech. 21 (11) (2011) 1679–1692.
[18] D.M. Bento, J.M. Monteiro, A QoS solution for three-dimensional Full-HD H.264/MVC video transmission over IP networks, in: Iberian Conference on Information Systems and Technologies, 2012.
[19] B.W. Micallef, C.J. Debono, An analysis on the effect of transmission errors in real-time H.264-MVC bit-streams, in: IEEE Mediterranean Electrotechnical Conference, 2010, pp. 1215–1220.
[20] Y. Su, A. Vetro, A. Smolic, Common conditions for multiview video coding, JVT-U211, 2006.
[21] H. Schwarz, D. Marpe, T. Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard, IEEE Trans. Circuits Syst. Video Tech. 17 (9) (2007) 1103–1120.
[22] C.A. Segall, G.J. Sullivan, Spatial scalability within the h.264/avc scalable video coding extension, IEEE Trans. Circuits Syst. Video Tech. 17 (9) (2007) 1121–1135.
[23] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G.B. Akar, M.R. Civanlar, A.M. Tekalp, Temporal and spatial scaling for stereoscopic video compression, in: Proceedings EUSIPCO, vol. 6, 2006.
[24] M. Drose, C. Clemens, T. Sikora, Extending single-view scalable video coding to multi-view based on H.264/AVC, in: ICIP, 2006, pp. 2977–2980.
[25] N. Ozbek, A. Murat Tekalp, Quality layers in scalable multi-view video coding, in: IEEE International Conference on Multimedia and Expo, 2009, pp. 185–188.
[26] M.-W. Park, G.-H. Park, Realistic multi-view scalable video coding scheme, IEEE Trans. Consum. Electron. 58 (2) (2012) 535–543.
[27] Y. Chen, R. Zhang, M. Karczewicz, MVC based scalable codec enhancing frame-compatible stereoscopic video, in: IEEE International Conference on Multimedia and Expo, 2011.
[28] Y. Lei, S. Xiaowei, H. Chunping, G. Jichang, L. Sumei, Z. Yuan, An improved multiview stereo videoFGS scalable scheme, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009, 2009.
[29] L.B. Stelmach, W. James Tam, Stereoscopic image coding: effect of disparate image-quality in left-and right-eye views, Signal Process.: Image Commun. 14 (1) (1998) 111–117.
[30] L.B. Stelmach, W.J. Tam, D.V. Meegan, A. Vincent, P. Corriveau, Human perception of mismatched stereoscopic 3D inputs, in: ICIP, vol. 1, 2000, pp. 5–8.
[31] H. Kalva, L. Christodoulou, L.M. Mayron, O. Marques, B. Furht, Design and evaluation of a 3D video system based on H.264 view coding, in: Proceedings of the International Workshop on Network and Operating Systems Support for Digital Audio and Video, 2006, p. 12.
[32] G. Saygili, C. Gurler, A.M. Tekalp, Quality assessment of asymmetric stereo video coding, in: ICIP, 2010, pp. 4009–4012.
[33] G. Saygili, C.G. Gurler, A.M. Tekalp, 3D display dependent quality evaluation and rate allocation using scalable video coding, in: ICIP, 2009, pp. 717–720.
[34] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, J. Kim, Asymmetric coding of stereoscopic video for transmission over T-DMB, in: 3DTV Conference, 2007.
[35] A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G.B. Akar, M.R. Civanlar, A.M. Tekalp, End-to-end stereoscopic video streaming with content-adaptive rate and format control, Signal Process.: Image Commun. 22 (2) (2007) 157–168.
[36] Y. Chen, S. Liu, Y.-K. Wang, M.M. Hannuksela, H. Li, M. Gabbouj, Low-complexity asymmetric multiview video coding, in: IEEE International Conference on Multimedia and Expo, 2008, pp. 773–776.
[37] H. Brust, A. Smolic, K. Mueller, G. Tech, T. Wiegand, Mixed resolution coding of stereoscopic video for mobile devices, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009.
[38] G. Saygili, C.G. Gurler, A.M. Tekalp, Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming, IEEE Trans. Broadcast. 57 (2) (2011) 593–601.
[39] J. Quan, M.M. Hannuksela, H. Li, Asymmetric spatial scalability in stereoscopic video coding, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2011.
[40] N. Ozbek, A.M. Tekalp, E.T. Tunali, A new scalable multi-view video coding configuration for robust selective streaming of free-viewpoint tv, in: IEEE International Conference on Multimedia and Expo, 2007, pp. 1155–1158.
[41] E. Kurutepe, M.R. Civanlar, A.M. Tekalp, Client-driven selective streaming of multiview video for interactive 3dtv, IEEE Trans. Circuits Syst. Video Tech. 17 (11) (2007) 1558–1565.
[42] Y. Wang, Z. Wu, J.M. Boyce, Modeling of transmission-loss-induced distortion in decoded video, IEEE Trans. Circuits Syst. Video Tech. 16 (6) (2006) 716–732.
[43] Z. He, J. Cai, C.W. Chen, Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding, IEEE Trans. Circuits Syst. Video Technol. 12 (6) (2002) 511–523.
[44] Y.-Z. Huang, J. Apostolopoulos, A joint packet selection/omission and FEC system for streaming video, in: ICASSP, vol. 1, 2007, pp. 845–848.
[45] Y.-Z. Huang, J. Apostolopoulos, Making packet erasures to improve quality of FEC-protected video, in: ICIP, 2006, pp. 1685–1688.
[46] W.-Y. Kung, C.-S. Kim, C.-C. Jay Kuo, Packet video transmission over wireless channels with adaptive channel rate allocation, J. Vis. Commun. Image Represent. 16 (2005) 475–498.
[47] K. Stuhlmuller, N. Farber, M. Link, B. Girod, Analysis of video transmission over lossy channels, IEEE J. Sel. Areas Commun. 18 (6) (2000) 1012–1032.
[48] W.-T. Tan, A. Zakhor, Video multicast using layered FEC and scalable compression, IEEE Trans. Circuits Syst. Video Tech. 11 (3) (2001) 373–386.
[49] T.-L. Lin, P.C. Cosman, Efficient optimal RCPC code rate allocation with packet discarding for pre-encoded compressed video, IEEE Signal Process. Lett. 17 (5) (2010) 505–508.
[50] Y. Chen, K. Xie, F. Zhang, P. Pandit, J. Boyce, Frame loss error concealment for SVC, J. Zhejiang Univ. Sci. A 7 (5) (2006) 677–683.
[51] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, Massachusetts, USA, 1999.
[52] European Telecommunications Standards Institute, Universal mobile telecommunications system (UMTS): multiplexing and channel coding (FDD), 3GPP TS 125.212 version 3.4.0, 2000, pp. 14–20.