ERROR RESILIENT VIDEO CODING FOR SYSTEMS WITH DELAYED FEEDBACK

Rahul Vanam, Zhifeng Chen, and Yuriy Reznik

InterDigital Communications, LLC, 9710 Scranton Road, San Diego, CA 92121 USA
E-mail: {rahul.vanam, zhifeng.chen, yuriy.reznik}@interdigital.com

ABSTRACT

In systems employing feedback-based error resilience, error propagation can significantly degrade visual quality when feedback delay is in the order of a few seconds. We propose a coding structure based on multiple description coding that mitigates error propagation during feedback delay, and uses feedback to adapt its coding structure to effectively limit error propagation. We demonstrate the effectiveness of our approach at different error rates when compared to conventional coding schemes that use feedback.

***Index Terms*—** Error resilience, error concealment, video coding, RTCP feedback, mobile video telephony.

1. INTRODUCTION

Thanks to the advances in wireless networks and improvements in processing and graphics capabilities of mobile devices, mobile video telephony is now becoming a part of our daily lives [1]. Yet, some technical challenges in the design of mobile video phones still exist. One such a challenge is the *lossy* nature of wireless networks, as well as other communication links connecting one user to the other.

We provide a simple illustrative example of such a system in Figure 1. In this case, video from user A is sent to user B using the RTP transport and RTCP control protocol [2]. Packet loss could occur either at the local link between the phone (UE) and the base station (eNB), in the Internet, or at the remote wireless link. This loss is eventually noticed by user B's application, and information about packet loss can be communicated back to user A by means of an RTCP receiver report (RR) [2, 3]. However, receiver reports are sent only periodically, usually once in every 1-5 second interval, as they should not generate a significant amount of traffic by themselves [2]. Hence, by the time the sender knows that the receiver did not receive some video packets, it is too late to retransmit them. Instead, the sender is usually instructed to send an I- or IDR-frame to stop error propagation caused by lost packets. Additionally, in order to reduce visual artifacts caused by lost packets in periods between receiver reports, the sender must employ video coding techniques that are *resilient* to packet loss.

In this paper we offer a brief review of several existing approaches for error resilient video coding and propose a new approach, which is customarily designed to accommodate long notification delays in RTP/RTCP - based systems.

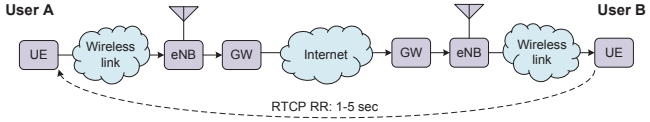

Fig. 1. Mobile video communication system employing RTP transport and RTCP feedback.

1.1. Prior art

The problem of error resilient video coding is well known, and prior research has produced a number of practical techniques for solving it. Recent surveys of such algorithms can be found in [4, 5, 6]. Below we list few general classes of such techniques.

Conventional methods for reducing error propagation. Random intra macroblock insertions, intra slices, and slice interleaving – are among best known practical techniques for error resiliency [4]. Such schemes break the coding dependency of macroblocks or slices in consecutive video frames, thereby limiting error propagation [7]. A recursive optimal per pixel estimate (ROPE) algorithm [8] estimates the overall distortion due to quantization, error propagation, and error concealment, and uses rate-distortion optimization to choose the best intra or inter mode for each macroblock. Stockhammer et al. [9] describe a multidecoder distortion estimation method that improves error resilience. Both methods [8] and [9] show good performance, but require high computational complexity. All these schemes assume no feedback, and offer better resiliency of encoded video at the expense of a moderate increase in the bitrate.

Feedback-based schemes. If feedback is available, it can be used to direct the video encoder to either encode the next frame as an IDR/I-frame, or encode using the most recent correctly transmitted frame as the reference. The former approach is called an *Intra refresh* and latter is called *reference picture selection* (RPS) [11]. Feedback-based methods may also be combined with using hierarchical P-frame coding structures, as in such cases it is sufficient to fix frames that belong to the "base layer" [12]. Most such techniques are only effective when the notification delay is relatively small (on the order of 100s of milliseconds). The longer the no-

tification delay, the longer the part of video sequence that is affected by the error. In practice, video decoders usually employ error concealment techniques, but even with state-of-art concealment, 1-5 seconds of delay before refresh can cause significant and visible artifacts (so-called "ghosting").

Multiple-description coding (MDC)- based schemes. MDC encoders produce several *descriptions* (subsets of packets), such that reception of any description is sufficient for meaningful reconstruction of video. The more descriptions that are received, the higher the quality of the reconstruction [13]. Simple examples of techniques in this class include temporal-, or spatial sub-sampling of the original video and coding of each sample set as a separate video stream. A survey and classification of MDC-based video coding schemes can be found in [14].

Feedback-based schemes for MDC. Several feedback-based techniques have been proposed for correcting errors in MDC-encoded video. These include: (a) RPS [15, 16], (b) error concealment [16], and (c) retransmission with fast decoding [16]. In the RPS method, the sender on receiving a loss notification predicts the next frame from a correctly transmitted frame [15, 16], and in addition may also use correctly received portions of the corrupted reference frame [16]. In the error concealment method, the encoder on receiving feedback, error conceals the frame in error, and uses it to predict future frames. This approach requires the encoder to know the error concealment used at the decoder (which usually is not the case in practice). The retransmission approach [16] is very similar to [12], except that it uses an MDC structure. All these techniques, however, were proposed for systems with very short notification delay (1-2 frames) [16], and don't seem to be practical in cases when this delay is long.

1.2. Contributions

In this paper, we propose a novel approach, which we call *Inhomogeneous Temporal Multiple Description Coding* (IHTMDC) for video. We consider long feedback delay in the design of our approach, which has not been considered by most prior methods. Our approach lowers error propagation distortion when waiting for the feedback, and on receiving it adapts its coding structure to limit error propagation. We call this adaptation mechanism *Cross-Description RPS* (CDRPS), and show that in the presence of long feedback delay it is more efficient than existing RPS-based methods [15, 16]. In the experimental section, we compare different coding structures at different packet error rates, and show that our approach has better performance over conventional methods at higher error rates.

1.3. Outline

The remainder of this paper is organized as follows. In Section 2, we describe our approach. Details of our experiments and results are provided in Section 3. Conclusions and outlook for future work are given in Section 4.

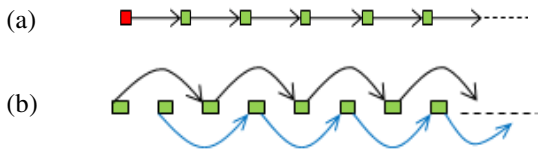

Fig. 2. (a) Conventional "IPPP" coding structure, and (b) Homogeneous temporal MDC.

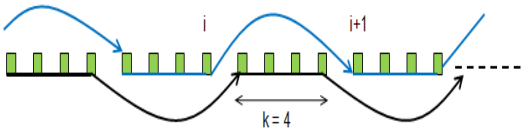

Fig. 3. Inhomogeneous Temporal Multiple Description (IMHDC) coding structure with interleaving factor $k = 4$.

2. DESCRIPTION OF THE PROPOSED SCHEME

In this section, we first describe the coding structure of conventional video codec and its generalization to temporal MDC. We then describe our proposed scheme, and show its relation to conventional (single description) and multiple description schemes. We also describe mechanisms for adaptation of this scheme using delayed feedback, allowing it to limit error propagation.

2.1. Conventional and MDC structures

We show the coding structure employed by the majority of today's real-time video codecs in Figure 2(a). It consists of an Intra- or IDR- frame followed by temporally predicted P-frames. It is commonly referred to as "IPPP" structure. The disadvantage of this scheme is a continuous chain of dependencies between frames and its susceptibility to error propagation.

One way to break this dependency is to create two or more sub-sequences of frames, which are not cross-referencing each other. We illustrate this approach in Figure 2(b), where we use two uniformly sampled sub-sequences to produce two independent encodings or *descriptions* of video. This is a very simple example of an MDC scheme for video, which we will call *homogeneous temporal MDC* (HMDC).

2.2. Inhomogeneous temporal MDC

We now propose a modification of temporal MDC method, where the temporal distances between adjacent frames in each description are not equal. We call this approach *Inhomogeneous Temporal MDC* (IHTMDC), and we illustrate it with an example in Figure 3. In this figure, frames i and (i+1) are

set five frames apart, while frames (i+1) and (i+2) are set one frame apart.

Our motivation for using this scheme is to maintain the correlation between frames to a large extent while generating descriptions, which results in a hybrid coding structure shown in Figure 3.

2.2.1. *Connection to a single description and HMDC*

We characterize IHTMDC by an interleaving interval k. In our example in Figure 3, this factor is set to $k = 4$. Different coding structures can be derived from the IHTMDC by varying k. For example, when $k = 1$, IHTMDC turns into a homogeneous temporal MDC scheme, shown in Figure 2(b). Similarly, if we set $k = \infty$, IHTMDC effectively becomes a single description IPPP coding structure as shown in Figure 2(a).

2.2.2. *Effects of packet loss*

In the IPPP coding structure, a packet loss would corrupt all successive frames. On the other hand, in HTMDC, the error propagates through one of the descriptions as illustrated in Figure 4 (a). Moreover, HTMDC structure allows the decoder to better conceal successive frames of a corrupted description by using neighboring frames belonging to uncorrupted description, thereby limiting error propagation drift to at most k consecutive frames.

2.2.3. *Effects of interleave factor k on overall distortion*

When considering transmission over a lossy channel, the overall (end-to-end) distortion of received video can be approximately expressed as:

$$D_{\mathrm{ETE}}(k) \approx D_{\mathrm{Q}}(k) + D_{\mathrm{T}}(k), \tag{1}$$

where D_{ETE}, D_{Q}, and D_{T} denote the end-to-end-, source coding-, and transmission- induced distortions, respectively. Specific conditions under which (1) holds true and related discussion can be found in [10].

Assuming that (1) holds true, we may conjecture that for a given source and a given channel there may exist an optimal choice of parameter k for our proposed coding scheme:

$$k^* = \arg \min_{k \in \mathbb{Z}^+} D_{\mathrm{ETE}}(k). \tag{2}$$

Intuitively, with no transmission errors, single description ($k = \infty$) is most desirable since it yields least coding distortion. However, in the presence of packet loss, $k = \infty$ may not be a good choice since error propagates through the length of the video yielding larger D_{ETE}. Using smaller k would increase D_{Q}, but it would also make the bitstream less sensitive to transmission errors, as errors propagates through one of the descriptions, thereby resulting in smaller D_{ETE}.

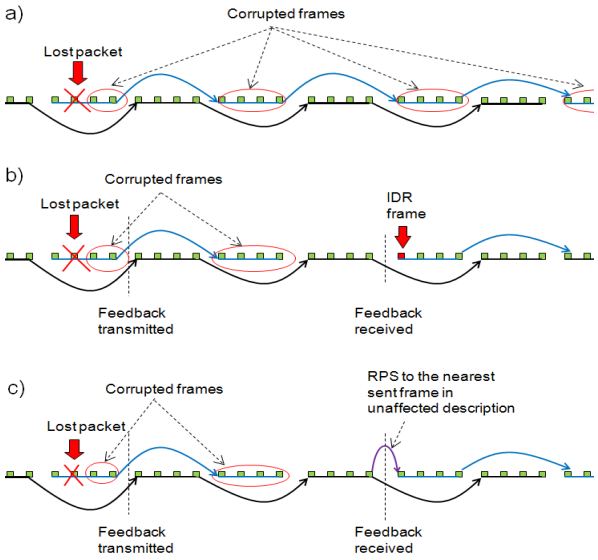

Fig. 4. IHTMDC subjected to errors. (a) Error propagation in IHTMDC without feedback. IHTMDC with feedback when using (b) intra refresh, and (c) cross description reference picture selection.

Therefore, the optimal choice of k must depend on the packet error rate.

In order to test this theory, in Section 3, we will perform experiments to study the effect of k on the rate-distortion performance under different packet error rates.

2.3. Adapting IHTMDC in response to feedback

We will now discuss uses of RTCP feedback for limiting error propagation in IHTMDC-coded video. There are at least two possible solutions:

- Intra refresh: Encode the next frame belonging to the corrupted description as an IDR/I-frame as illustrated in Figure 4 (b).
- Cross description RPS (CDRPS): In this approach, the encoder based on rate-distortion optimization decides whether to encode the next frame belonging to the corrupted description as an intra/IDR frame, or encode it using the *nearest frame from the uncorrupted description* as the reference. The latter approach is illustrated in Figure 4 (c).

Performing an intra refresh or CDRPS on the next corrupted frame limits error propagation of the corrupted description. When $k = \infty$, the above two methods turn into conventional single description intra refresh and RPS, respectively. However, when k is finite, the CDRPS method is different compared to traditional RPS techniques [11]. In traditional RPS schemes, the reference is always set to last frame that was confirmed as delivered. With long 1-5 second feedback, this means that such reference would have to be 25-100 frames back. On the other hand, with IHTMDC and one

surviving description - such a reference can always be found within the last k frames. This makes this scheme much more suitable for systems with delayed feedback.

In practical implementations, the CDRPS and intra-refresh techniques can be used in a complementary fashion. For example, when the encoder knows that both the descriptions have been lost since the last feedback, it may insert a new IDR frame in one description, and use cross-description reference in another description to restart the encoding process.

3. EXPERIMENTAL RESULTS

In this section, we describe our experimental setup and results.

3.1. Experiment setup

In our tests we have utilized standard CIF and high-definition test sequences [18], and looped them back and forth to generate 1000 frames for each test. We used "Foreman", "Soccer", and "News" for CIF sequences (352×288, 30 fps), and "Pedestrian" for HD sequence (1080p, 25 fps). Since we generate IHTMDC bitstream using a modification of the x264 encoder [19][17], the computational complexity of our scheme is comparable to the single description coding scheme. Constant QP rate control option and one reference frame was used in all experiments. We used the H.264 JM decoder with frame-copy error concealment method enabled.

For CIF sequences, we set $QP = 26, 28, 30, 32$, and 34, and use a frame as a slice. Here a lost packet corresponds to a lost frame. For the 1080p sequence, we set $QP = 30, 34, 38$, and 42, and encode using 14 slices per frame, as this was necessary to keep the NAL unit size within 1400 bytes for our operating bitrates. In order to understand the effectiveness of the proposed methods we have setup an experiment in which we have simulated a channel with no errors, 10^{-2} and 3×10^{-2} packet error rates (PER), which are typical for conversational services over LTE. We have also implemented RTCP notification with a one second delay. We have tested IHTMDC with interleaving factors $k = 1, 2, 4$, as well as conventional H.264 single-description coding scheme ($k = \infty$), and a single-description coding scheme that uses random intra macroblock refresh ($k = \infty$ + RIM). The RPS technique (CDRPS in case of IHTMDC) is used to correct errors upon RTCP notification. For the random intra macroblock refresh scheme, 5% and 9% intra macroblocks (MBs) are used for CIF and HD videos, respectively.

3.2. Results

Table 1 illustrates visual quality achievable with single description coding ($k = \infty$) vs. IHTMDC with $k = 4$. Sequence "Pedestrian.yuv" is used in this experiment. The error starts at frame number 166 for both schemes. As expected, single description scheme ($k = \infty$) propagates error into frame 186, while in the case of IHTMDC ($k = 4$) the error is not noticeable in frames 176 and 186.

Figures 5 (a)-(i) show the rate-distortion performance of IHTMDC with CDRPS for different packet error rates and values of k for CIF sequences. As expected, the RD performance of the single description scheme ($k = \infty$) performs the best for the no-error case as shown in Figures 5 (a), (d), and (g). With packet loss, IHTMDC and HMDC show better performance over the single description scheme for the "Foreman" and "Soccer" sequences as shown in Figures 5 (b), (c), (e), and (f). For the "News" sequence at PER = 10^{-2}, single description ($k = \infty$) performs the best at bitrates less than 250 kb/s, and $k = 4$ performs the best at higher bitrates as shown in Figure 5 (h). This clearly indicates that the choice of k is also dependent on the video content and operating bitrate. For PER = 3×10^{-2}, $k = 4$ shows the best R-D performance. With packet loss, HMDC ($k = 1$) shows poor performance over the single description scheme as shown in Figures 5 (h) and (i). Under packet loss, the random intra MB scheme has comparable RD performance to the single description scheme as shown in Figures 5 (b), (c), (e), and (f). For the "News" sequence, the single description scheme exploits the slow video content by using a large number of skip MBs. Therefore, inserting random intra MBs affects the overall RD performance even under packet loss as shown in Figures 5 (h) and (i).

For the HD "Pedestrian" sequence, we only test for PER = 10^{-3} and 2×10^{-3}, since our IHTMDC scheme at $k = 4$ demonstrates good performance at such low packet error rates. Specifically, for PER = 10^{-3}, $k = 4$ has similar performance to single description ($k = \infty$) for bitrates less than 1.4 Mb/s, and has best performance for higher bitrates, yielding up to 0.7 dB gain over single description as shown in Figure 6 (b). For PER = 2×10^{-2}, $k = 4$ has the best performance yielding up to 1.5 dB gain over single description as shown in Figure 6 (c).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an inhomogeneous multiple description video coding technique that provides excellent error resilience properties, and is suitable for systems with long feedback delay. Our scheme effectively uses feedback to reset the coding structure, thereby limiting error propagation. Our scheme can be used to derive different coding structures by varying the interleaving factor. We compare our scheme with single description coding, random intra macroblock refresh, and homogeneous temporal multiple description coding, which all use feedback, and find that our scheme provides better visual quality and RD performance at higher error rates. In our current work, we studied our approach using a fixed interleaving factor. In our future work, we plan to adapt the interleaving factor based on the observed packet error rates.

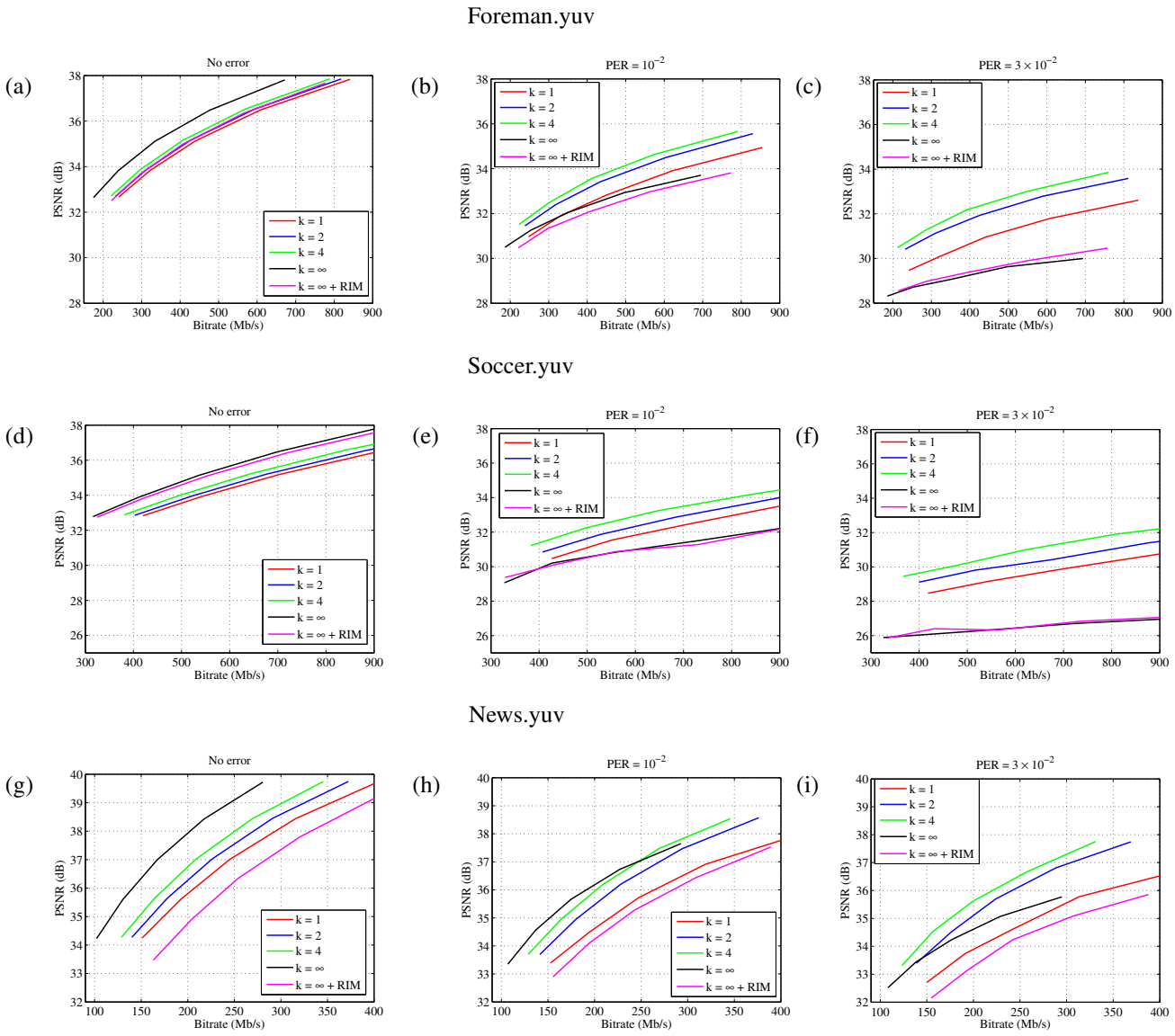

Fig. 5. Rate-distortion performance of IHTMDC for different interleaving factors k, packet error rates, and frame resolutions. Plots for "Foreman.yuv" : (a) no errors, (b) PER = 10^{-2}, and (c) PER = 3×10^{-2}. Plots for "Soccer.yuv": (d) no error, (e) PER = 10^{-2}, and (f) PER = 3×10^{-2}. Plots for "News.yuv": (d) no error, (e) PER = 10^{-2}, and (f) PER = 3×10^{-2}. Cases when $k = \infty$, $k = \infty$ + RIM, and $k = 1$ correspond to the single description scheme, random intra MB refresh scheme, and homogeneous temporal MDC, respectively.

5. REFERENCES

[1] T. Weigand and G. J. Sullivan, "The picturephone is here. Really," *IEEE Spectrum*, vol. 48, no. 9, pp. 50–54, Sept. 2011.

[2] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC 3550: RTP: A transport protocol for real-time applications," July 2003.

[3] J. Ott, S. Wenger, N. Sato, C. Burmeister, and J. Ray, "IETF RFC 4585: Extended RTP profile for real-time transport control protocol (RTCP)-based feedback (RTP/AVPF)," 2006.

[4] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Review of error resilient coding techniques for real-time video communications," *IEEE Signal Proc. Magazine*, vol. 17, pp. 61–82, 2000.

[5] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication – a review," in *Proc. IEEE*, 1998, pp. 974–997.

[6] S. Kumar, L. Xu, M. K. Mandal, and S. Panchanathan, "Error resiliency schemes in H.264/AVC standard," *J. Visual Comm. Image Rep.*, vol. 17, no. 2, pp. 425–450, 2006.

[7] T. Stockhammer, "Error robust macroblock mode and reference frame selection," in *VCEG JVT-B102*, Jan 2002.

[8] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with

Table 1. Illustration of error propagation in single-description coding vs. IHTMDC using CDRPS with one second feedback delay using "Pedestrian.yuv" sequence. Error occurs at frame number 166. The red ellipses highlight errors. For single description ($k = \infty$), error propagates all the way until frame number 186, while in the IHTMDC case ($k = 4$) error propagation is not noticeable in frames 176 and 186.

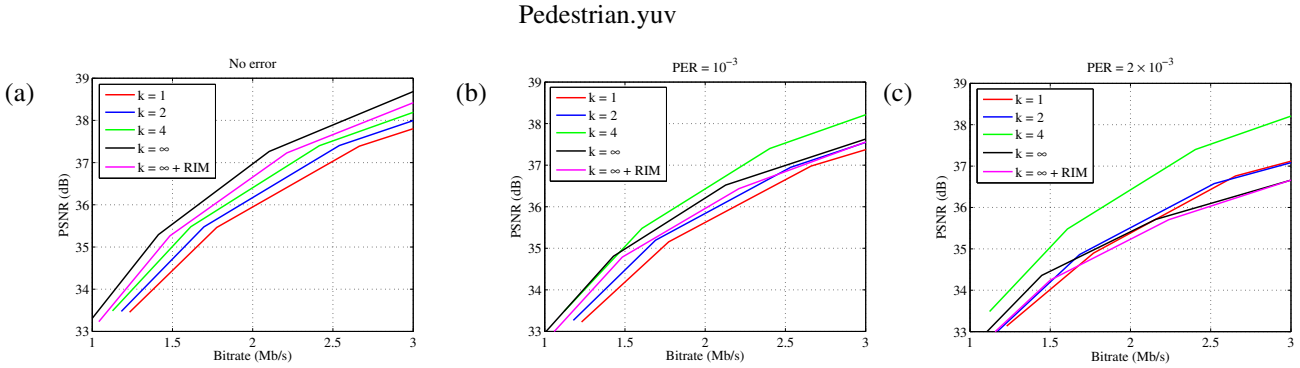

Fig. 6. Rate-distortion performance of IHTMDC for different interleaving factors k and packet error rates for "Pedestrian.yuv" sequence: (a) no error, (b) PER = 10^{-3}, and (c) PER = 2×10^{-3}.

optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Area Comm.*, vol. 18, no. 6, pp. 966–976, 2000.

[9] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Cir. Sys. Video Tech.*, vol. 13, pp. 657–673, 2003.

[10] Z. Chen and D. Wu, "Rate-Distortion Optimized Cross-Layer Rate Control in Wireless Video Communication," *IEEE Trans. Cir. Sys. Video Tech.*, vol. 22, no. 3, pp. 352–365, March 2012.

[11] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," in *Proc. IEEE*, 1999, pp. 1707–1723.

[12] I. Rhee and S. R. Joshi, "Error recovery for interactive video transmission over the internet," *IEEE J. Sel. Area Comm.*, vol. 18, pp. 1033–1049, 2000.

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

[14] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *in Proc. IEEE*, vol. 93, no. 1, pp. 57–70, 2005.

[15] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. GLOBECOM 1996*, 1996, vol. 3, pp. 1503 – 1508.

[16] W. Tu and E. G. Steinbach, "Proxy-based reference picture selection for error resilient conversational video in mobile networks," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 19, no. 2, pp. 151–164, Feb 2009.

[17] L. Merritt and R. Vanam, "Improved Rate Control and Motion Estimation for H.264 Encoder," *in Proc. ICIP*, vol. 5, pp. 309–312, Sept. 2007.

[18] "Raw video sequences," ftp.ldv.e-technik.tu-muenchen.de.

[19] "x264 encoder," http://www.videolan.org/developers/x264.html.