Error Resiliency of Distributed Video Coding in Wireless Video Communication

Shuiming Ye, Mourad Ouaret, Frederic Dufaux, Michael Ansorge, Touradj Ebrahimi

Multimedia Signal Processing Group
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

Distributed Video Coding (DVC) is a new paradigm in video coding, based on the Slepian-Wolf and Wyner-Ziv theorems. DVC offers a number of potential advantages: flexible partitioning of the complexity between the encoder and decoder, robustness to channel errors due to intrinsic joint source-channel coding, codec independent scalability, and multi-view coding without communications between the cameras. In this paper, we evaluate the performance of DVC in an error-prone wireless communication environment. We also present a hybrid spatial and temporal error concealment approach for DVC. Finally, we perform a comparison with a state-of-the-art AVC/H.264 video coding scheme in the presence of transmission errors.

Keywords: Distributed Video Coding, Error Resiliency, Wireless Communication, Wyner-Ziv coding, H.264/ AVC

1. INTRODUCTION

Nowadays, the most widely used digital video coding solutions are represented by the ISO/IEC MPEG and ITU-T H.26x family of standards, of which the most recent H.264/AVC [1] stands for the state-of-the-art. All these coding schemes rely on a highly complex encoder and a much simpler decoder. However, in some emerging applications, such as wireless low-power video surveillance, multimedia sensor networks, wireless PC cameras, and mobile camera phone, low complexity encoding is required. Distributed Video Coding (DVC) [2], a new paradigm in coding which allows for very low complexity encoding, is well-suited for these applications.

In DVC, the complex task of exploiting the source statistics and motion estimation can be moved from the encoder to the decoder. The Slepian-Wolf theorem on lossless distributed source coding states that the optimal rate of joint encoding and decoding of two statistically dependent discrete signals can be achieved by using two independent encoders and a joint decoder [3]. Wyner-Ziv coding extends this result to lossy coding with Side Information (SI) [4].

DVC generally divides a video sequence into key frames and WZ frames. The major task of exploiting source statistics is carried out in SI generation process to produce an estimation of the WZ frame being decoded. SI has a significant influence on the Rate Distortion (RD) performance of DVC. Indeed, more accurate SI implies that fewer bits are requested from the encoder, so that the bitrate is reduced for the same quality. In common DVC codecs, the SI is obtained by Motion Compensation Temporal Interpolation (MCTI) from the previous and next key frames and utilizes the Block Matching Algorithm (BMA) for motion estimation. However, motion vectors from BMA are often not faithful to true object motions. Unlike classical video compression, it is more important to find true motion vectors for SI generation in DVC. Therefore, it is important to improve the SI generation in

DVC in order to achieve better RD performance. Another appealing property of DVC is its good resilience to transmission errors due to its intrinsic joint source-channel coding framework. A thorough analysis of its performance in the presence of transmission errors has been presented in [5], showing its good error resilience properties. This results from the fact that DVC is based on a statistical framework rather than the closed-loop prediction used in conventional video coding.

Recently, the rapid growth of Internet and wireless communications has led to increased interest for robust transmission of compressed video. However, transmission errors may severely impact video quality as compressed data is very sensitive to these errors [6]. Thus, error control techniques are necessary for efficient video transmission over error prone channels.

H.264/AVC [1] has been developed based on previous MPEG and H.26x standards. Apart from better coding efficiency, the standard has also given strong emphasis to error resiliency and the adaptability to various networks. To give consideration to both coding efficiency and network friendliness, H.264/AVC has adopted a two-layer structure design: a video coding layer (VCL) to obtain highly compressed video data, and a network abstraction layer (NAL) for adaptation to various transportation protocols or storage media [1]. For stream-based protocols such as H.320 and H.324, the NAL delivers compressed video data with start codes such that these transport layers and the decoder can robustly and easily identify the bit stream structure. For packet-based protocols such as RTP/IP and TCP/IP, the NAL delivers the compressed video data in packets without these start codes [7].

In this paper, we perform a comparison of DVC with H.264/AVC error resilience in the presence of transmission errors. We also describe a hybrid temporal and spatial error concealment scheme for WZ frames in DVC, previously introduced in [8]. More specifically, error resilience performance of DVC with error concealment is compared with that of H.264/AVC. The comparisons are done for two different cases. In the first case, no feedback channel is used to inform the encoder when a packet is lost, and therefore no retransmission is requested. In this case, error concealment is implemented as a post process to improve the quality of frames corrupted by transmission errors. In the second case, an Automatic Repeat reQuest (ARQ) retransmission is used with a feedback channel to retransmit lost packets. Experimental results show the good error resilience of DVC compared with H.264/AVC.

The paper is organized as follows. First, the DVC architecture and other related work are introduced in section 2. Section 3 introduces a new hybrid spatio-temporal error concealment based the improved SI generation techniques. The simulation results are presented in section 4. Finally, section 5 concludes the paper.

2. DISTRIBUTED VIDEO CODING

Without loss of generality, in this paper, we consider the Transform Domain Wyner-Ziv (TDWZ) DVC architecture from [9], as shown in Figure 1. A video sequence is divided into key frames (Y_k) and WZ frames (X_k). Hereafter, we consider a Group of Pictures (GOP) size of 2, namely the odd and even frames are key frames and WZ frames, respectively. Key frames Y_k are conventionally encoded using H.264/AVC Intra coding [1]. Conversely, for WZ frames X_k, a DCT transform is firstly applied to the input video frames, and the resulting transform coefficients undergo quantization. The quantized coefficients are then split into bitplanes which are turbo encoded.

At the decoder, SI approximating the WZ frames is generated by MCTI of the decoded key frames. The SI is used in the turbo decoder, along with WZ parity bits requested from feedback channel, in order to reconstruct the decoded WZ frames X'_k. This new video coding paradigm enables to explore the signal statistics, partially or totally, at the decoder; in other words, DVC enables to shift complexity from the encoder to the decoder. The Slepian-Wolf decoding is performed using an iterative turbo decoding procedure. The iterative decoding procedure stops when a given convergence criteria is satisfied. The WZ parity bits requested at the decoder can not only correct the errors between SI and it original frame, but also the transmission errors. Due to this intrinsic joint source-channel coding framework, DVC is robust to channel errors.

Furthermore, since in DVC signal statistics are explored at the decoder, no prediction loop is needed at the encoder side, hence avoiding the inter-frame error propagation typical of traditional video coding systems. Therefore, in DVC schemes improved error resilience can be achieved in a more natural way, i.e. without sending additional information to increase the bitstream error robustness. Conversely, in traditional video coding schemes, channel coding techniques are typically employed to make the source encoded bitstream more robust to channel errors.

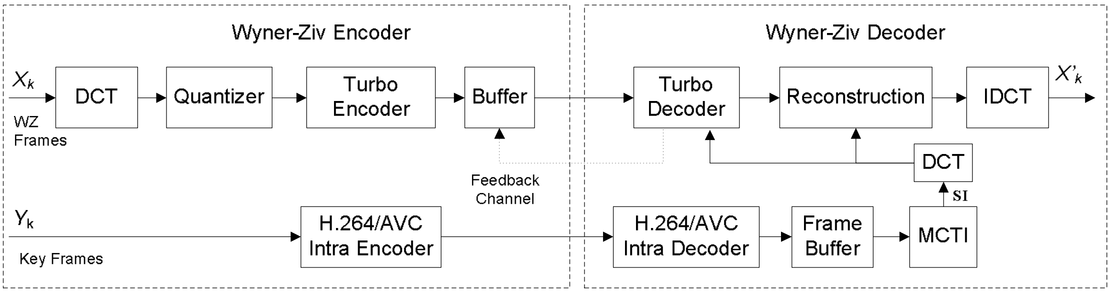


Figure 1: DVC architecture.

3. HYBRID ERROR CONCEALMENT FOR DVC

To conceal errors in WZ frames of DVC, we adopt the hybrid spatial-temporal error concealment scheme proposed in our previous work [8] (Figure 2). It uses the error-concealed results from spatial Error Concealment (EC) to improve the performance of the temporal EC, instead of simply switching between spatial and temporal EC. Spatial EC based on the edge directed filter [10] is firstly applied to the corrupted blocks, and the spatially concealed blocks are used as partially decoded WZ frames to improve the performance of temporal EC. An enhanced temporal EC is then applied, which includes motion vector refinement and smoothing, optimal compensation mode selection based on the spatial error-concealed results, and a new matching criterion for motion estimation. In other words, the temporal EC is not only based on the key frames, but also on the WZ bits already decoded. Experimental results show that the proposed scheme significantly improves the quality of the SI and RD performance of DVC, and the performance of the proposed hybrid scheme is superior to spatial EC and temporal EC alone.

The locations of corrupted blocks are firstly detected. In this paper, we assume that the error locations are detected at the decoder, as often presumed in error concealment literature. It can be done at transport level, or based on syntax and watermarking [6]. Spatial EC is then applied to obtain a partially error-concealed frame. This frame is much closer to the error free frame than the corrupted one. The partially error-concealed frame is used for motion vector refinement, smoothing, and optimal compensation mode selection, to obtain an estimate of the motion vector of the corrupted block. Motion compensation is finally used to obtain the reconstructed block.

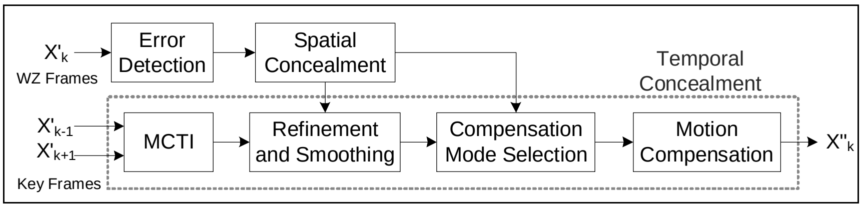


Figure 2: Proposed spatio-temporal error concealment.

In DVC, the decoded WZ frames are based on the SI generated by MCTI of the key frames. SI is then used by the turbo decoder, along with the WZ bits, to obtain the decoded WZ frame. The transmission errors in WZ bits tend to cause noises around edges in the corrupted WZ frames [8]. It is different from traditional video coding schemes such as H.264/AVC, where errors result in loss of whole blocks. Therefore, an edge directed filter is constructed to remove the noises around edges caused by errors in the WZ bits. Anisotropic diffusion techniques have been widely used in image processing for their efficiency at smoothing noisy images while preserving sharp edges. We adopt the anisotropic diffusion as a direction diffusion operation, and use the diffusion method for spatial error concealment as in [10].

An enhanced temporal EC is then used to exploit spatio-temporal correlations. The approach is based on MCTI and motion vector filtering as proposed in [11]. One of the key novelties is that the partially error-concealed frame is used to improve the temporal EC, unlike [11] where MCTI is based on the previous and next key frames. Indeed, the reconstructed frame by spatial concealment contains additional information about the current frame carried by the correctly received WZ bits. Therefore, by using the partially error-concealed frame resulting from spatial EC, the spatio-temporal correlations between this frame and the reference key frames can be better exploited. Hence the performance of the temporal EC is improved.

4. ERROR RESILIENCE PERFORMANCE EVALUATION

In our simulations, only luminance data is coded, and a communication channel, characterized by the error pattern files provided in [12] with different packet loss ratios (PLR), is used. Test sequences in the QCIF format at 15 fps are corrupted with packet loss rates of 5%, 10%, and 20%. For the various test conditions in terms of PLR and quantization parameters (QP), the average PSNR is measured. For every testing condition, results are obtained by averaging over ten runs using different error patterns.

In this paper, if the bit error probability of the decoded bitplane is higher than 10^{-3}, the decoder uses the corresponding bitplane of the side information, which improves the quality of the decoded frame. The header of the WZ bitstream, which contains critical information such as frame size, quantization parameters, and intra period, is assumed to be correctly received. The turbo decoder stops requesting more bits if the bitplane bit-error rate is below a given threshold equal to 10^{-3}.

In order to evaluate the error resilience of the DVC, its RD performance is compared to that of the H.264/AVC, which represents the state-of-the-art in video coding. The three H.264/AVC modes used for performance comparisons are:

- H.264/AVC Intra: Coding with H.264/AVC without exploiting temporal redundancy is used.
- H.264/AVC Inter: Coding with H.264/AVC in IPP mode with GOP size 15 is used in simulations.

- H.264/AVC Inter No Motion: Coding with H.264/AVC in IPP mode but without performing any motion estimation, which is the most computationally expensive encoding task. In this paper, GOP size 15 in IPP mode is used in simulations.

The H.264/AVC codec used in this paper is the H.264/AVC JM 11.0 reference software. We have enabled the Flexible Macroblock Ordering (FMO) feature in dispersed mode at the encoder to improve its error resilience. The performance evaluation are carried out considering two different scenarios: (1) Without feedback channel, the proposed hybrid error concealment is adopted for DVC, and the error concealment defined in H.264/AVC JM software is enabled; and (2) With feedback channel, ARQ is adopted to re-transmit the lost packets.

4.1 Error Resilience Perform without Feedback Channel

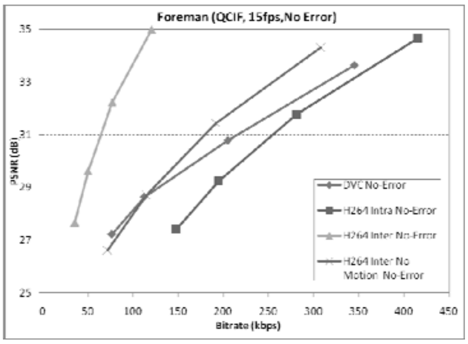
In video streaming, due to the tight delay constraint, a feedback channel is not preferred since it will cause additional delay, and the received retransmitted bits beyond time constraint will be useless. In this case, no feedback channel is used when errors occurred during the transmission.

The error concealment algorithm implemented in the H.264/AVC JM11.0 Reference Software is adopted in this paper for DVC when there are transmission errors in key frames, as well as for H.254/AVC Intra, Inter, and Inter No Motion decoding [13]. For Intra coding, the H.264/AVC software Intra error concealment is adopted using spatial interpolation based on weighted average of boundary pixels of the missing block. On the other hand, for Inter coding, the H.264/AVC software Inter error concealment is used based on frame copy method.

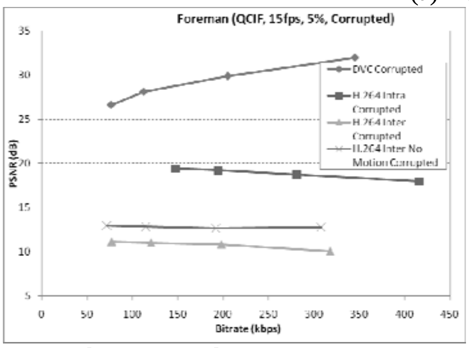
Figure 3 shows the rate-distortion performance for the *Foreman* sequence at Packet Loss Rate (PLR) of 5%, 10%, and 20%. These results show that, without error concealment, the qualities of the corrupted sequences of H.264/AVC coding are very low, especially for Inter coding with/without motion. On the other hand, the quality loss of DVC without error concealment is much smaller.

We can also notice that when error concealment is enable for DVC and H.264/AVC, at a low PLR (Figure 3b), H.264 (Intra, Inter, and Inter No Motion) can compensate the channel errors, and for higher PLR values (Figure 3c and Figure 3d), H.264 Inter coding gives the worst results, especially at high bitrates. On the contrary, the error concealment for DVC can significantly improve the quality for all PLRs at high bitrates. Furthermore, it is worth noting that the key frames of DVC in our simulations are coded with H.264/AVC Intra mode, which means the error resilience performance of DVC are affected by that of H.264/AVC Intra coding. If we only evaluate the WZ frames in DVC, as shown in [8], the error resilience performance of DVC would be much better.

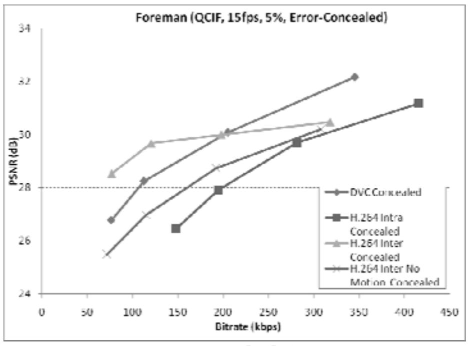
Similar results have been obtained for sequence with low motion such as *Hallmonitor*, as shown in Figure 4. The performances of H.264/AVC Inter and H.264/AVC Inter No Motion for such low motion sequence are very close. These results support similar conclusions from *Foreman* that the distortions caused by transmission errors in DVC are much smaller than those in H.264/AVC.



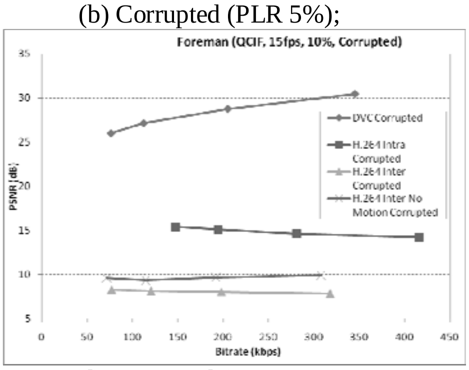
(a) No-error



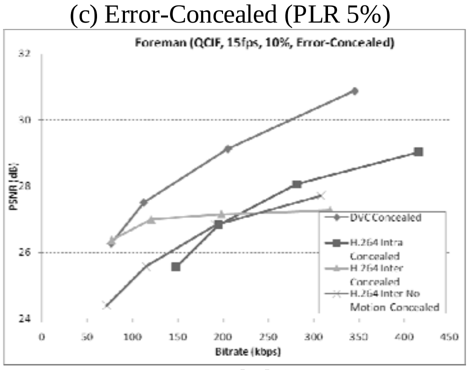
(b) Corrupted (PLR 5%);



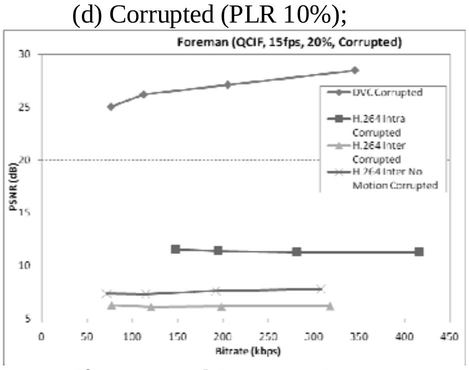
(c) Error-Concealed (PLR 5%)



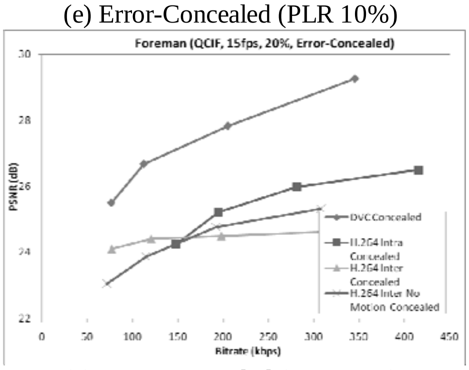
(d) Corrupted (PLR 10%);



(e) Error-Concealed (PLR 10%)

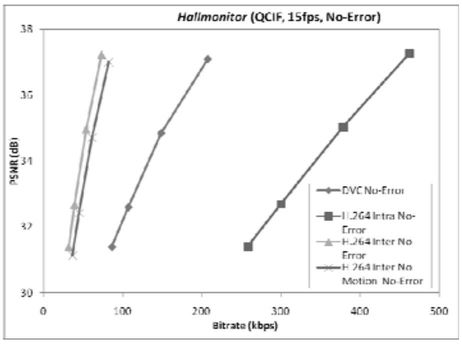


(f) Corrupted (PLR 20%);

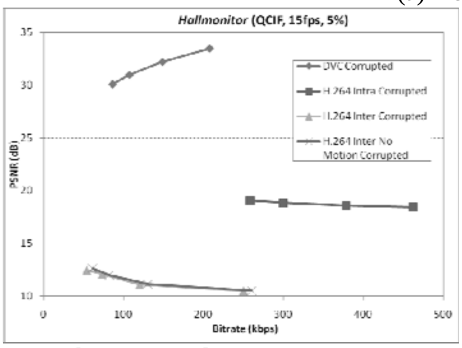


(g) Error-Concealed (PLR 20%)

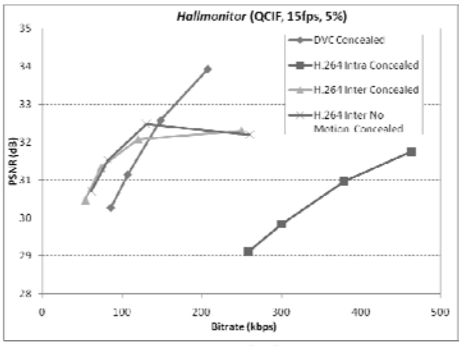
Figure 3: Error Resilience Performance of *Foreman* (without feedback channel)



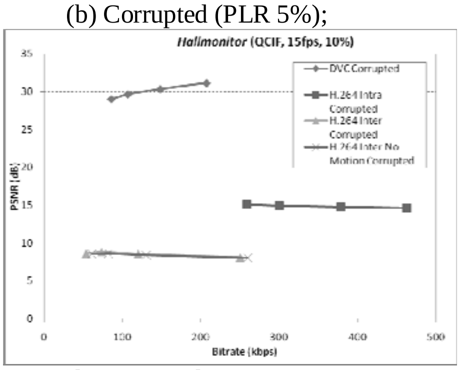
(a) No-Error



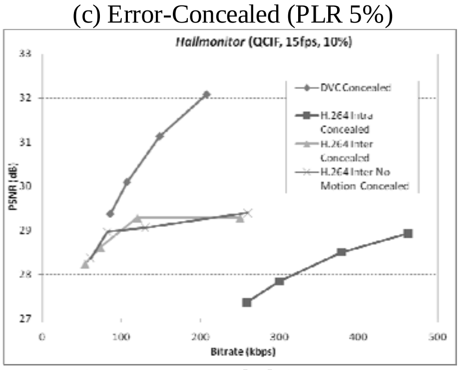
(b) Corrupted (PLR 5%);



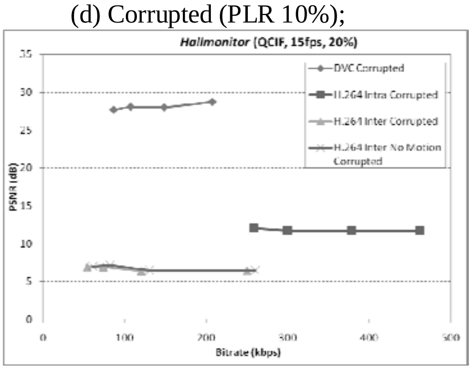
(c) Error-Concealed (PLR 5%)



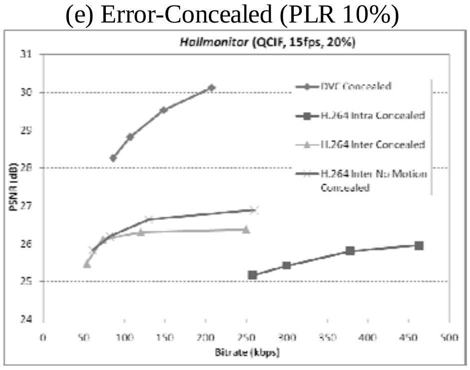
(d) Corrupted (PLR 10%);



(e) Error-Concealed (PLR 10%)



(f) Corrupted (PLR 20%);



(g) Error-Concealed (PLR 20%)

Figure 4: Error Resilience Performance of *Hallmonitor* (without feedback channel)

4.2 Error Resilience Perform with Feedback Channel

Feedback channel can be efficiently used for packets retransmission. The receiver can send ACK messages or NAK messages to the sender to report whether the RTP packets have been received or not. We also compare the error resilience performance in case of feedback channel. Namely, an ideal automatic retransmission is adopted when errors are detected and the information is sent back to the encoder by the feedback channel. However, packet retransmission may bring time delay to the decoder, and it is restricted by conversational video services requirement. ARQ is usually used to guarantee error-free delivery when there is a feedback channel to inform the encoder which packets are lost [7].

With the feedback channel available, ideal ARQ retransmission is adopted in our simulations. For this case, it is assumed that, whatever the packet size, the network protocol may ask for the retransmission of lost packets until they are correctly received. The results for *Foreman* and *Hallmonitor* are shown in Figure 5 and Figure 6, respectively. The final quality is the same for each PLR because the lost packets are retransmitted by the encoder, meaning that all the packets are received correctly. The bitrate for a specific PLR (R_{PLR}) is determined based on the rate of original no-error rate ($R_{No\text{-}error}$) as:

$$R_{PLR} = \frac{R_{No-error}}{1-PLR} \qquad (1)$$

In this case, an increase of bitrate is observed which depends only on the original rate-distortion performance.

5. CONCLUSIONS

In this paper, the error resilience of DVC has been evaluated, compared to that of H.264/AVC. In particular, we have first compare their error resilience performances when there is no feedback channel. In this case, error concealment methods are adopted at the decoder to improve the quality of the corrupted frames. We have then presented a comparison with a feedback channel and ARQ to recover the lost packets. All these results show that the distortions caused by transmission errors in DVC are much smaller than those in H.264/AVC. At high PLR and high bitrates, the DVC codec can outperform the H.264/AVC JM codec. The results confirm the intrinsic error resilience capability of DVC. The reason is the intrinsic robustness of the DVC decoding paradigm. There are a lot of techniques adopted in H.264/AVC to improve the error resilience, but few have been dedicated to DVC. Further work could also be done to further improve the error resilience of DVC. For example, improved error concealment proposed for H.264/AVC can be adopted to improve the quality of key frames corrupted by errors in DVC.

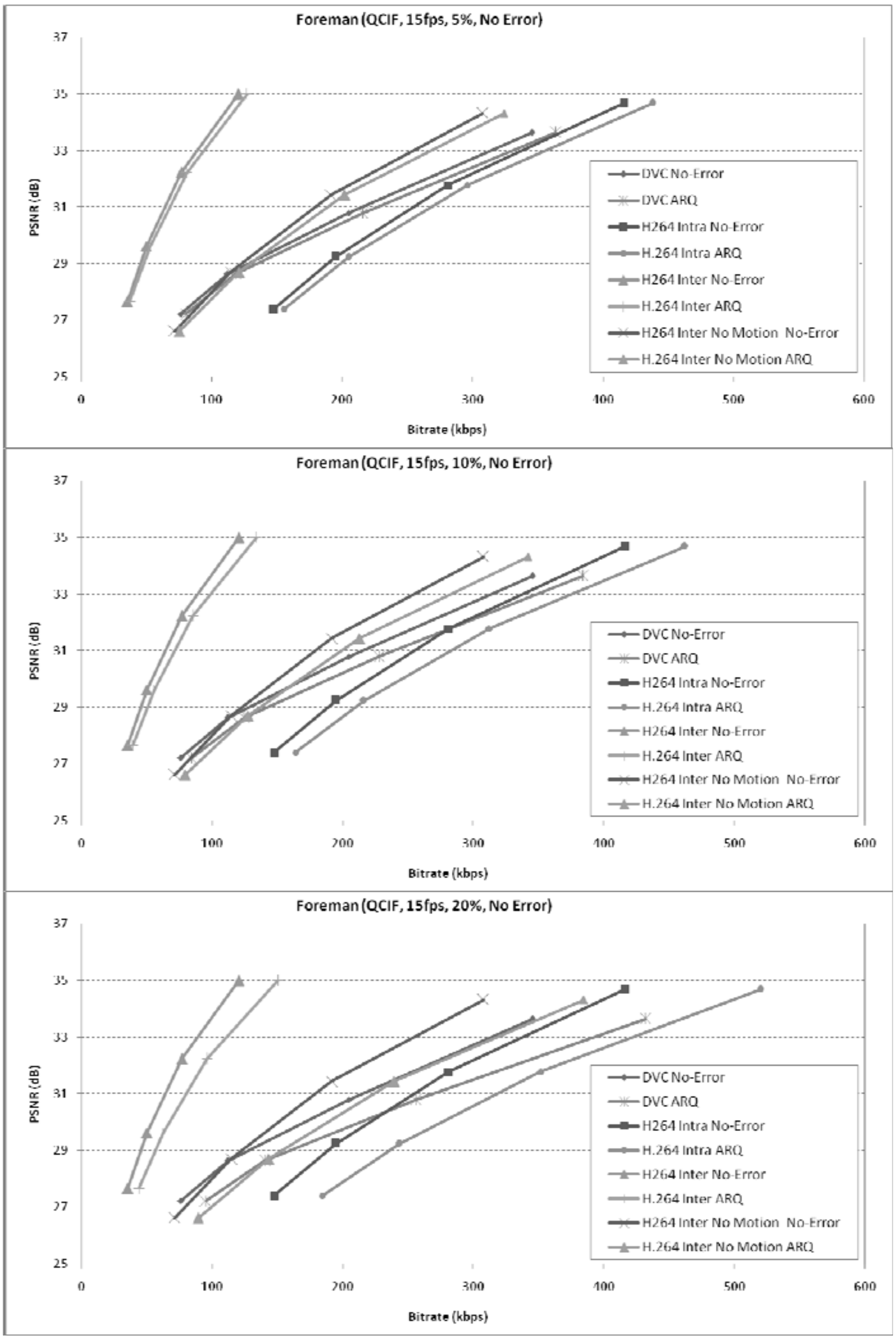


Figure 5: Error resilience performance of *Foreman* with feedback channel

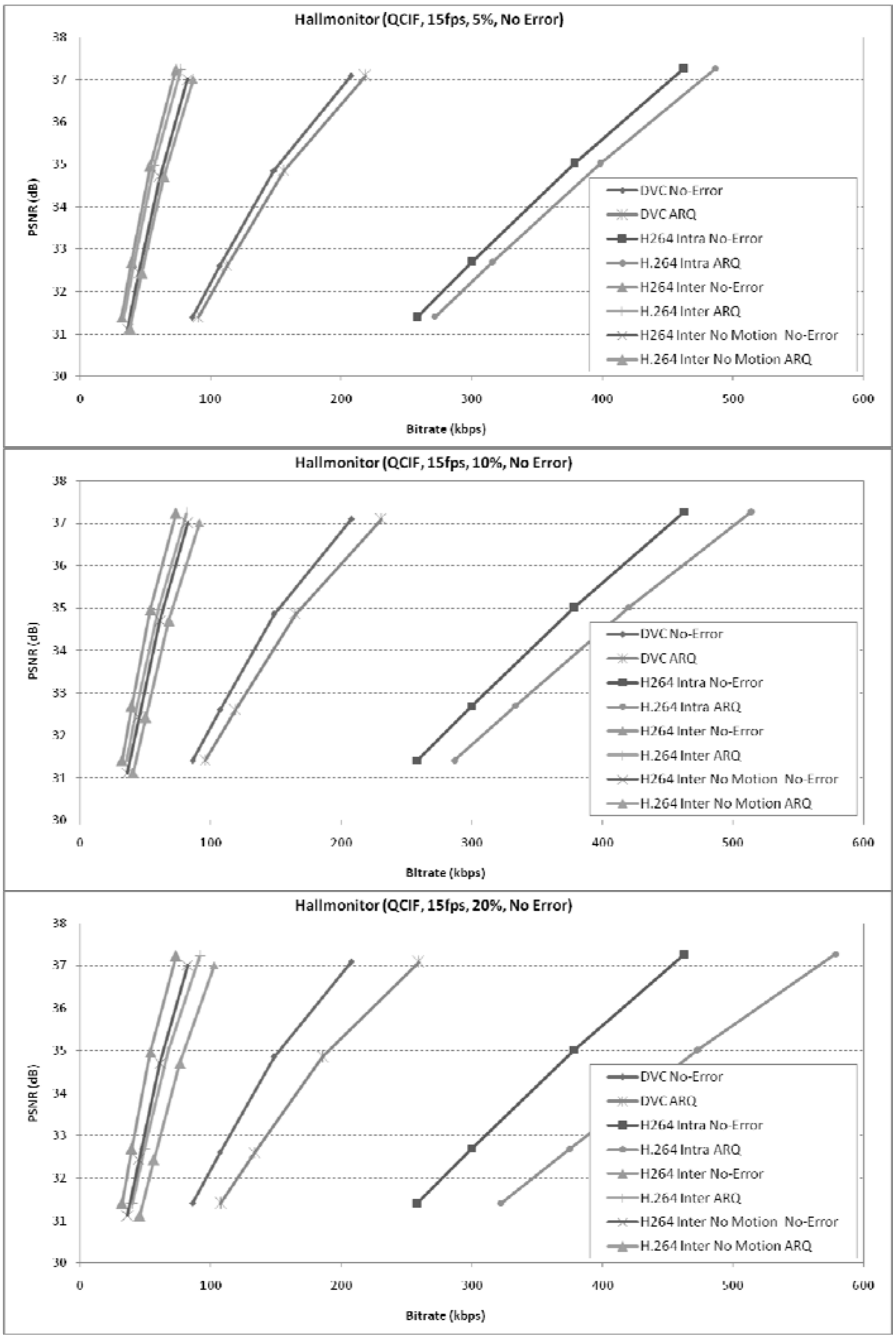


Figure 6: Error resilience performance of *Hallmonitor* with feedback channel

ACKNOWLEDGEMENT

This work was partially supported by the European Network of Excellence VISNET II (http://www.visnet-noe.org) IST Contract 1-038398, funded under the European Commission IST 6th Framework Program.

REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.13, No.7, pp. 560-576, July 2003.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed Video Coding", *Procedings of the IEEE*, Vol. 93, No. 1, Jan. 2005, pp. 71-83.

[3] J. D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Transactions on Information Theory*, Vol. IT-19, pp. 471–480, Jul. 1973.

[4] A. Wyner and J. Ziv, "The Rate-distortion Function for Source Coding with Side Information at the Decoder", *IEEE Transactions on Information Theory*, Vol. IT-22, No. 1, pp. 1–10, Jan. 1976.

[5] J. Pedro, L. Soares, C. Brites, J. Ascenso, F. Pereira, C. Bandeirinha, S. Ye, F. Dufaux, and T. Ebrahimi, "Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec", *Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.

[6] Y. Wang and Q. Zhu, "Error Control and Concealment for Video Communication: A Review", *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998, pp.974-997.

[7] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments", *IEEE Transaction on Circuits and Systems for Video Technology*, vol.13, pp. 657–673, July 2003.

[8] S. Ye, M. Ouaret, F. Dufaux and T. Ebrahimi, "Hybrid Spatial and Temporal Error Concealment for Distributed Video Coding", *IEEE International Conference on Multimedia and Expo*, Hannover, Germany, June 2008.

[9] C. Brites, J. Ascenso, and F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance", *International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[10] S. Ye, Q. Sun, and E.C. Chang, "Edge Directed Filter based Error Concealment for Wavelet-based Images", *IEEE International Conference on Image Processing*, Singapore, 2004.

[11] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak, July 2005.

[12] S. Wenger, "Proposed Error Patterns for Internet Experiments", Doc. VCEG Q15-I-16R1, VCEG Meeting, Red Bank, NJ, USA, Oct. 1999.

[13] K.-P. Lim, G. Sullivan, T. Wiegand "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods", JVT-K049, Munich, Germany, March 2004.