Multiple Description Video Coding with H.264/AVC Redundant Pictures

Ivana Radulovic, Pascal Frossard, Ye-Kui Wang, Miska M. Hannuksela, and Antti Hallapuro

***Abstract*—Multiple description coding offers interesting solutions for error resilient multimedia communications as well as for distributed streaming applications. In this letter, we propose a scheme based on H.264/AVC for encoding of image sequences into multiple descriptions. The pictures are split into multiple coding threads. Redundant pictures are inserted periodically in order to increase the resilience to loss and to reduce the error propagation. They are produced with different reference frames than the corresponding primary pictures. We show, given the channel conditions, how to optimally allocate the rates to primary and redundant pictures, such that the total distortion at the receiver is minimized. Extensive experiments demonstrate that the proposed scheme outperforms baseline solutions based on loss and content-adaptive intra coding. Finally, we show how to further reduce the distortion by efficient combination of primary and redundant pictures, if both are available at the decoder.**

***Index Terms*—H.264/AVC, multiple description video coding, redundant pictures.**

I. Introduction

THERE has been recently a rapid development of multimedia services and applications such as video conferencing, mobile video, or Internet Protocol TV (IPTV). These applications are often subject to packet loss and bandwidth variations on current packet networks. Error resilience techniques have been shown to provide elegant solutions that offer a sustained quality to the users in the absence of guarantee from the transmission channels. Among these, multiple description coding (MDC) [1] has recently emerged as a promising solution, especially in low-latency applications. It offers improved performance compared to schemes based on forward error correction, especially when the channel conditions are not accurately estimated [2].

The most popular MDC schemes for video, such as video redundancy coding (VRC) [3] or multiple state video coding (MSVC) [4], split the input video sequence into subsequences of frames that are independently coded, with their own prediction process and state. With this solution, even if one description is completely lost, another one can be independently decoded and reconstructed at half of the frame rate. Moreover, the frames lost in one description can be reconstructed by interpolation from the neighboring frames in another description. Other examples of multiple description video coding schemes based on information splitting in the temporal domain include multiple description motion compensation schemes [5], [6], rate-distortion optimized unbalanced MDC [7], and the optimal selection of different MDC coding modes, investigated in [8].

In this letter, we propose a standard compatible MDC video scheme for low-delay applications over lossy networks. We build on our previous work [9] and use H.264/AVC redundant pictures to provide robustness to transmission errors. The video information is split into several encoding threads, and redundant pictures are inserted to reduce the error drift in case of packet loss. In contrary to the classical construction, redundant pictures are coded in a different thread than the corresponding primary pictures. Given the channel conditions, we show how to allocate the coding rate to primary and redundant pictures, such that the total distortion experienced at the receiver is minimized. We finally show how the decoding quality can be further improved by a proper handling of the different versions of received pictures available at the decoder. Extensive simulations demonstrate that our MDC algorithm outperforms state-of-the-art single and two-description video coding schemes in terms of average quality, as well as quality variation and resiliency to incorrect estimation of the channel state. It is worth noting that a parallel work of Tillo *et al.* [10] also proposes an MDC video coding scheme based on redundant pictures. The descriptions are however not completely independent, and the decoding process does not exploit all the information available at the decoder.

This rest of this letter is organized as follows. Section II describes the proposed scheme in detail, while in Section III we compare its performance with state-of-the-art techniques. We discuss decoder improvements in Section IV. Finally, Section V summarizes the letter.

Manuscript received January 25, 2008; revised December 20, 2008. First version published July 7, 2009; current version published January 7, 2010. This work was supported in part by the Swiss National Science Foundation grant PP-002-68737. This paper was recommended by Associate Editor G. Wen.

I. Radulovic was with the Signal Processing Laboratory (LTS4), Ecole Polytechnique Federale de Lausanne, Lausanne 1015, Switzerland. She is now with Ericsson Research, Stockholm 16480, Sweden (e-mail: ivana.radulovic@gmail.com).

P. Frossard is with the Signal Processing Laboratory (LTS4), Ecole Polytechnique Federale de Lausanne, Lausanne 1015, Switzerland (e-mail: pascal.frossard@epfl.ch).

Y.-K. Wang was with Nokia Research Center, Tampere 33720, Finland. He is now with Huawei Technologies, Bridgewater, NJ 08807 USA (e-mail: yekuiwang@huawei.com).

M. M. Hannuksela and A. Hallapuro are with Nokia Research Center, Tampere 33720, Finland (e-mail: antti.hallapuro@nokia.com; miska.hannuksela@nokia.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2009.2026815

II. MDC with Redundant Pictures

We extend the MSVC scheme [4] and increase the resiliency to temporal propagation of errors by the addition of redundant pictures. Redundant pictures (RP) are one of tools included in H.264/AVC that can be used efficiently for error resilient video coding [11]–[13]. Typically, each (primary) picture in the encoded video sequence may be associated with one or more RPs. The decoder can reconstruct the redundant picture in case a primary picture (or parts thereof) is missing. On the other hand, RPs are usually discarded by the decoder if the corresponding primary picture is correctly received.

The proposed coding scheme (MSVC-RP) is illustrated in Fig. 1. We consider a simple I-P-P-... scenario with a single reference frame, since we mostly target low-delay applications. It can be noted, however, that MSVC-RP can be extended to B-pictures or multiple reference frames.

The input video sequence is split into sequences of odd and even source pictures. When encoding, each primary picture in the even/odd description is predicted only from other pictures of the same description, typically the previous picture. In addition, redundant pictures are included in the bitstream of each description, thus carrying the information from the alternate description. In the time domain, they are positioned such that they can replace a lost primary picture. Unlike the primary pictures, which use the previous primary frames from the *same thread* as a reference, redundant pictures are predicted from the previous frame in the input sequence. Redundant pictures are coded as P pictures (except the first two, which are intra coded) and each primary frame has its redundant version. Redundant pictures are not used as a reference for any subsequent picture. The descriptions are typically placed in different transmission packets and sent to the network. They could be transmitted over two different lossy channels, if such an arrangement is supported by the network. Sending the descriptions over a single link consists of sending primary pictures followed by their corresponding redundant pictures, which is the normal decoding order of H.264/AVC bitstreams. If the descriptions are sent over independent paths, pictures within a description are sent in their decoding order (from left to right in Fig. 1).

The redundant pictures typically use the same coding modes as the corresponding primary pictures, but they are more coarsely quantized in order to save on the overall coding rate. Naturally, the quality of redundant pictures should be chosen by taking the network loss rate into account. If the loss rate is very low, the probability that a primary picture is lost and has to be replaced by the corresponding RP is also low, and this is why RPs should be quantized coarsely. On the other hand, as the loss rate increases, better quality of RPs becomes more advisable. Clearly, this comes at a price of reducing the quality of primary pictures when the total rate is fixed. On average, having better quality RPs is however beneficial, since there is a higher probability that a primary picture is lost and replaced by a RP.

The receiver can face three different situations, depending on if the primary and redundant pictures are lost or not. First, if primary pictures are received error-free, the standard suggests that the RPs should be discarded. In our letter, we will first follow this approach, thus keeping the decoding process as simple as possible, which can be of great importance for delay-sensitive applications. In Section IV, we will eventually improve the decoding process with more efficient handling of all the received pictures. Second, if a primary picture (or parts thereof) has been lost, the corresponding redundant picture is decoded and used to replace its missing parts. Since the quantization is generally coarser in RPs, this operation typically leads to artifacts in the decoded sequence. However, the degradation is generally smaller than the error generated by simple concealment with the information from the neighboring macroblocks from the same and/or subsequent frames. Third, if both primary and redundant parts of a picture are lost, the missing information is reconstructed using an error concealment algorithm, e.g., by copying the closest available previous frame from either description. After the necessary discarding/replacement/concealment, the two descriptions are subsequently interleaved to produce the final reconstruction.

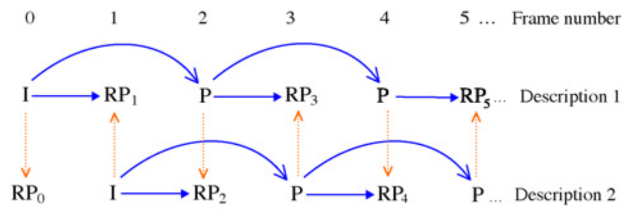

Fig. 1. Proposed scheme for MDC video.

III. Performance Evaluation

In this section, we compare our scheme with three solutions proposed in the literature that represent viable solutions for low-delay applications. We start by describing the testing conditions. Then, we compare the average quality, as well as quality variation and resiliency to incorrect estimation of the channel state for all the schemes. For the detailed analysis of MSVC-RP, including redundancy analysis, source and channel distortion models, and optimal rate allocation between primary and redundant pictures, please refer to [14].

A. Testbed

Our testbed corresponds to the common error resilience testing conditions specified in JVT-P206 [15], which specifies the required testing sequences, together with the corresponding bitrates and frame rates, as well as the bitstream packetization. The NAL unit size is limited to 1400 bytes, and the maximal size of each slice is chosen such that it fits in one NAL unit. Therefore, depending on the bitrate and the sequence format, there may be several slices per frame. Finally, an overhead of 40 bytes for the RTP/UDP/IPv4 headers is also taken into account when calculating the total bitrates.

We compare our MSVC-RP with three state-of-the-art schemes.

1) MSVC scheme: the video is encoded into two independent coding threads, without redundant pictures. Note that the author in [4] considers several error concealment strategies when an entire frame is lost. In our work we

only consider the simple scheme, where a lost frame is replaced with the closest possible received frame from either description, similarly to [8].

2) *Adaptive intra refresh* (AIR) scheme [16], which takes into account both the source distortion and the expected channel distortion (due to losses) and chooses an optimal mode for each macroblock based on Lagrange optimization. Therefore, it is likely to place intra macroblocks in more "active" areas.

3) *Random intra refresh* (RIR) scheme, which increases the robustness to losses by randomly inserting macroblocks whose number is proportional to a packet loss rate.

To have a fair comparison, we fix the total bit rates for all the schemes to be equal. In case of loss, parts or entirely lost pictures are replaced with their redundant versions taken from the alternate description in our MSVC-RP implementation. If both primary and redundant pictures are lost, we copy the temporally closest decoded picture from either description. For the other schemes, in case of partial frame losses, the missing pieces are copied from the corresponding places in the previous pictures. If an entire picture is lost, we copy the entire previous picture, as it is implemented in the MSVC-RP scheme. In addition, only the first frames in all the video sequences are encoded as I pictures.

We have tested all the sequences specified in JVT-P206 and at several loss rates [15]. To obtain statistically meaningful results, all the bitstreams are concatenated and tested with the entire loss patterns containing 10 000 binary characters, for all the packet loss rates. We show here the results for the three sequences: *News QCIF*, *Foreman QCIF*, and *Stefan CIF*, while similar results for other test sequences can be found in [14].

B. Selecting Q_p and Q_r

The optimal values Q_p and Q_r are obtained by full search over all combinations of quantization parameters that satisfy the total bitrate constraint. Table I shows these optimal parameters for the *Foreman QCIF* sequence, encoded at 7.5 fps and 144 kbits/s and *Stefan CIF* sequence at 30 fps and 512 kbits/s. We can observe that the value of Q_r decreases when the loss rate increases, as expected. When the losses are very high (20%), the primary and redundant pictures are coded with very similar quantization parameters. The increase in quality of redundant pictures comes clearly at the expense of decreasing the quality of primary pictures when the overall bit rate is constrained. This however improves the average distortion, since the probability of using the redundant pictures becomes significant. On the other hand, when the loss rate is low, the optimal allocation tends to give as much rate as possible to primary pictures, while the redundant pictures are made very coarse. In this case, the system avoids wasting bits on the redundant pictures that are unlikely to be used in the decoding process.

C. End-to-End Distortion Analysis

TABLE I

OPTIMAL QUANTIZATION PARAMETERS THAT MINIMIZE THE AVERAGE DISTORTION, AS A FUNCTION OF p. SEQUENCES *Foreman QCIF* AT 7.5 FPS AND 144 KBITS/S AND *Stefan CIF* AT 30 FPS AND 512 KBITS/S

	Foreman QCIF		*Stefan CIF*	
p	Q_p^{opt}	Q_r^{opt}	Q_p^{opt}	Q_r^{opt}
3%	25	42	41	49
5%	26	34	41	49
10%	28	29	42	44
20%	28	29	42	44

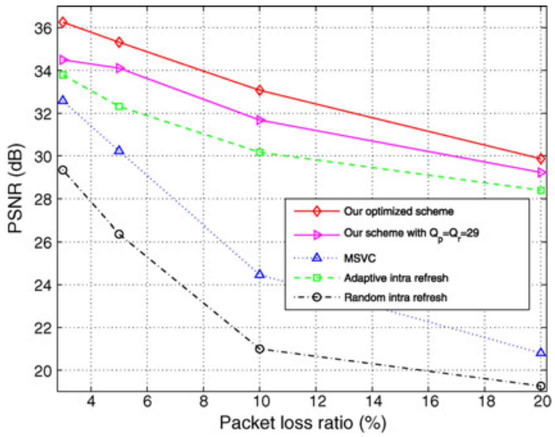

Fig. 2. Average PSNR versus loss probability. Sequence: *Foreman QCIF*, 7.5 fps, 144 kbits/s.

We analyze here the performance of the different error resilient coding solutions in terms of average distortion, for different loss ratios. The average PSNR is illustrated in Fig. 2 for the *Foreman QCIF* test sequence. We compare our optimal MSVC-RP solution with the MSVC, RIR, and AIR schemes, as well as with MSVC-RP with maximal redundancy (i.e., $Q_p = Q_r$). It can be seen that the MSVC-RP scheme performs generally the best at all packet loss rates, and that the AIR scheme also provides an efficient solution at either low or high packet loss rate, depending on the activity in the vide sequence. At 10% loss probability, the MSVC-RP scheme outperforms the AIR and MSVC schemes by approximately 3 and 8 dB for the *Foreman* sequence. The quality gain due to MSVC-RP generally increases with the loss rate, since the redundancy offered by the design of two descriptions is really beneficial in this case, compared to joint coding with only one coding thread. For the complex sequences like *Stefan* encoded at medium bitrate, we can see that the performance of the MSVC-RP stays close to the AIR scheme, due to the limitations of the simple error concealment method that is unable to provide a sustainable quality when the loss of one description becomes frequent. On the contrary, the coding of intra blocks in areas of high activity helps to improve the quality for the AIR scheme at high loss rates.

We further study the performance of the proposed scheme in a wider range of rates, and we compare it to the AIR scheme. Fig. 3 shows the average PSNR as a function of the rate constraint R, for the *Foreman* sequence, when the loss rate is equal to $p = 5\%$. We can see that our approach gives the best performance in the whole range of bitrates, from 0.6 dB at 32 kbits/s up to 2.7 dB at 192 kbits/s. Moreover, the gain increases as the bitrate increases.

Finally, Fig. 4 presents the temporal evolution of the PSNR for the different encoding schemes, for the same loss trace. The

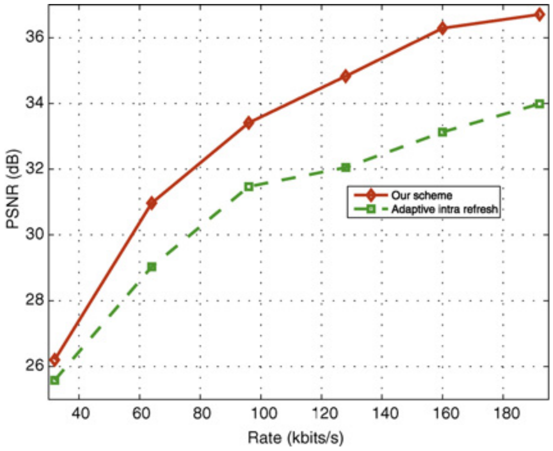

Fig. 3. Average PSNR, as a function of encoding rate, when PLR = 5%. Sequence: *Foreman QCIF*, 7.5 fps.

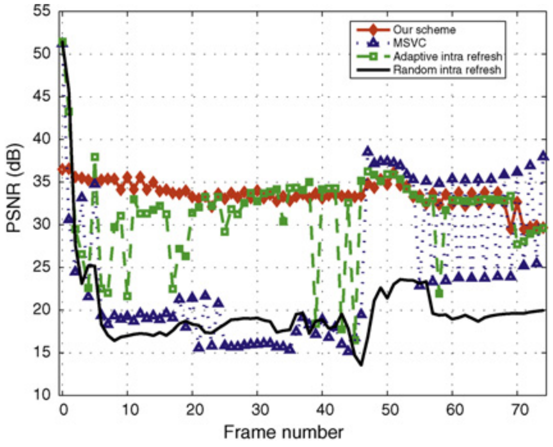

Fig. 4. Reconstructed video quality, on a frame basis, when PLR = 10%. Sequence: *Foreman QCIF*, 7.5 fps, 144 kbits/s.

error pattern is taken from a random entry in the error pattern file. The MSVC-RP scheme generally gives the best decoding quality. We can also notice that the AIR succeeds in catching up with our scheme, but with big variations in quality and with the performance similar to MSVC-RP in short intervals. The MSVC scheme performs very bad before the scene change around the frame 45. Then it recovers, thanks to inserted intra macroblocks after the scene change, but the frame-by-frame quality varies in significant amounts, up to 12 dB between two consecutive frames. Overall, it can be observed that the variations of quality for the MSVC-RP scheme are much smaller than for the other schemes. This illustrates the benefits of the design of two descriptions that can be decoded independently. Similar results have been observed for other loss rates, other sequences, and other video formats [14].

D. Robustness to Inexact Loss Rate Estimation

We finally discuss the robustness of the encoding schemes to incorrect loss rate estimation, which is likely to happen in practical scenarios. We compare the MSVC-RP and the AIR approaches that are optimized for a given loss ratio p, but where the actual loss rate is different from the expected one. This is actually a common situation in practical scenarios. Fig. 5 presents the end-to-end quality for the *Foreman* sequence, when all the schemes are optimized for $p = 5\%$, but when the actual loss ratio varies from 3% to 20%. For the sake of completeness, we also plot the best performance of MSVC-RP and AIR at each loss ratio. The differences between the optimized and actual performance for both schemes are 0.39 dB and 0.14 dB respectively, when $p = 3\%$. Not surprisingly, the gap between the optimized and actual performance increases as the actual loss ratio moves away from 5%. At $p = 10\%$, these gaps for both schemes are 0.9 dB and 1.33 dB respectively, while at $p = 20\%$ the corresponding gaps are 1.32 and 2.78 dB. Therefore, we can conclude that MSVC-RP is more robust to unknown network conditions. This can be a very desirable property, especially if the sender cannot change the encoding parameters as fast as the network conditions change. A similar behavior is observed for the other test sequences.

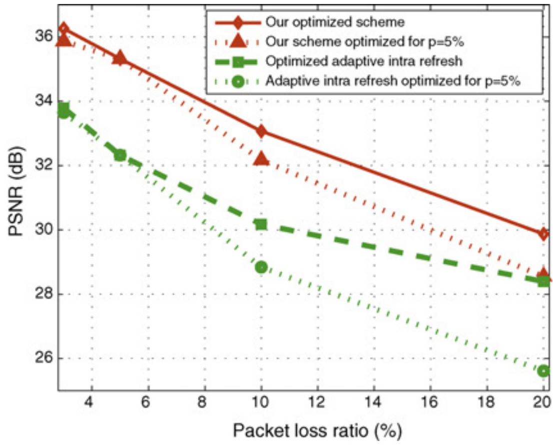

Fig. 5. Actual and minimal distortion versus the actual PLR, when all the schemes are optimized for PLR = 5%. Sequence: *Foreman QCIF*, 7.5 fps, 144 kbits/s.

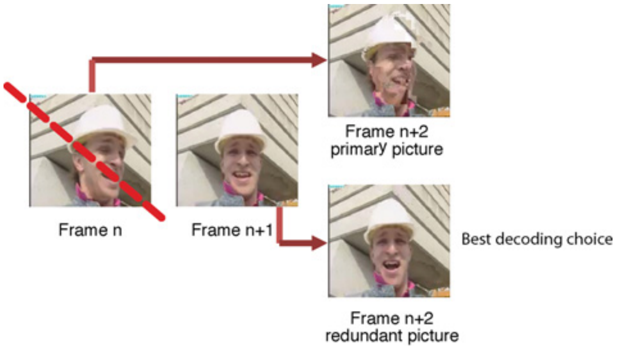

Fig. 6. Reconstruction of the $(n+2)$th frame from the *Foreman QCIF* sequence ($Q_p = 28$, $Q_r = 29$) when its both primary and redundant pictures are received. Here the nth frame, used as a reference for the $(n+2)$th primary frame, is entirely lost, while the $(n+1)$th frame, used as a reference for the $(n+2)$th redundant frame is correctly received.

IV. Improved Frame Reconstruction

A. Combination of Pictures

In the first part of this letter, we discarded redundant pictures if the corresponding primary pictures were available at the

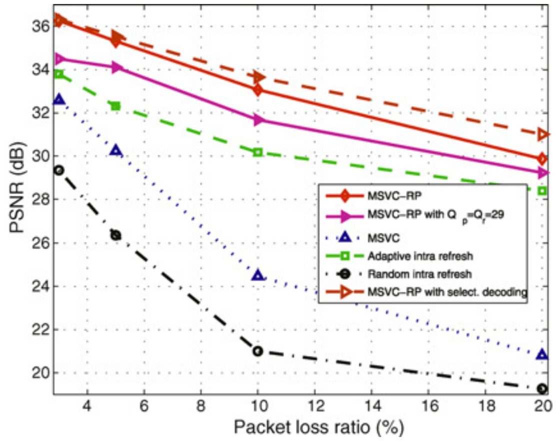

Fig. 7. Minimal achievable average distortion, as a function of a probability of loss, p (sequence: *Foreman QCIF*, 7.5 fps, 144 kbits/s).

decoder. This solution has an advantage of simplicity, but it is clearly suboptimal. Although a primary picture is correctly received, it may happen that its reference frames are affected by losses, which causes error propagation that also affects the primary picture. At the same time, it may happen that the thread from which a redundant picture is decoded is error-free or less affected by transmission errors. In these scenarios, choosing a redundant instead of a primary picture may be beneficial. This especially makes sense if the quantization parameters for primary and redundant pictures of the same original picture are very similar, which further induces similar visual qualities for both frames. Since primary and redundant pictures are decoded from different threads, the transmission error propagates only in one thread or description. We can therefore choose to use the best possible frame in case of loss in the reference frames, as depicted in Fig. 6. A model that addresses rate-distortion optimal macroblock selection between a primary coded picture and the respective redundant coded picture is detailed in [14]. However, it can be noted that the improved solution is not standard-compatible anymore, since both primary and redundant pictures need to be decoded in this case.

We report in Fig. 7 the benefits of the improved decoding process in terms of average distortion for *Foreman QCIF* sequence, while similar results can be found in [14]. We can see that the PSNR quality improvement ranges from 0.07 dB when $p = 3\%$ to 1.14 dB when $p = 20\%$. In general, the improvement at low loss rates is rather small, and gets more important at high loss rates. As the loss rate gets higher, it becomes very likely that an entire frame can be lost, in which case a serious quality degradation can be seen in subsequent frames. At the same time, the probability that both threads are simultaneously affected stays small, so that the possibility of choosing the frame to decode becomes beneficial. We conclude that discarding the redundant pictures by default is not optimal, as the additional information provided by these pictures can be very helpful against temporal error propagation.

V. Conclusion

In this letter, we have proposed a simple and H.264/AVC compatible Multiple Description Video Coding scheme based on redundant pictures. Compared to the state-of-the-art error resilient coding for low-latency applications, the proposed scheme offers significant gains in terms of average PSNR, fewer temporal fluctuations in the picture quality, and improved robustness to bad estimation of the loss probability in the network. We finally propose an improved decoding process that exploits the best information available at the decoder in primary or redundant picture. We plan to further study the efficient and adaptive allocation of redundant pictures in the video descriptions, based on the scene content.

References

[1] V. K. Goyal, “Multiple description coding: Compression meets the network,” *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

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

[3] S. Wenger, “Video redundancy coding in H.263+,” in *Proc. Workshop Audio-Visual Services Packet Netw.*, 1997, pp. 23–28.

[4] J. Apostolopoulos, “Reliable video communication over lossy packet networks using multiple state encoding and path diversity,” in *Proc. Vis. Commun. Image Process.*, Jan. 2001, pp. 392–409.

[5] Y. Wang and S. Lin, “Error-resilient video coding using multiple description motion compensation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 438–452, Jun. 2002.

[6] C.-S. Kim and S.-U. Lee, “Multiple description motion coding algorithm for robust video transmission,” in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4. Mar. 2000, pp. 717–720.

[7] D. Comas, R. Singh, A. Ortega, and F. Marques, “Unbalanced multiple-description video coding with rate-distortion optimization,” *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 81–90, Jan. 2003.

[8] B. Heng, J. Apostolopoulos, and J. S. Lim, “End-to-end rate-distortion optimized MD mode selection for multiple description video coding,” *EURASIP J. Appl. Signal Process.*, vol. 2006, no. Article ID 32592, p. 12.

[9] I. Radulovic, Y.-K. Wang, S. Wenger, A. Hallapuro, M. M. Hannuksela, and P. Frossard, “Multiple description H.264 video coding with redundant pictures,” in *Proc. ACM Multimedia 2007*, Sep. 2007, pp. 37–42.

[10] T. Tillo, M. Grangetto, and G. Olmo, “Redundant slice optimal allocation for H.264 multiple description coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 59–70, Jan. 2008.

[11] Y.-K. Wang, M. M. Hannuksela, and M. Gabbouj, “Error resilient video coding using unequally protected key pictures,” in *Proc. Very Low Bit Rate Video*, Sep. 2003, pp. 290–297.

[12] P. Baccichet, S. Rane, and B. Girod, “Systematic lossy error protection based on H.264/AVC redundant slices and fexible macroblock ordering,” in *Proc. Packet Video Workshop*, Hangzhou, China, Apr. 2006.

[13] C. Zhu, Y.-K. Wang, M. Hannuksela, and H. Li, “Error resilient video coding using redundant pictures,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 1, pp. 3–14, Jan. 2009.

[14] I. Radulovic, “Balanced multiple description coding in image communications,” Ph.D. thesis, Ecol. Polytech. Federale Lausanne, Lausanne, Switzerland, Dec. 2007.

[15] Y.-K. Wang, S. Wenger, and M. M. Hannuksela, “Common conditions for SVC error resilience testing,” *JVT Output Document, JVT-P206*, Aug. 2005.

[16] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, “Error resilience video coding in H.264 encoder with potential distortion tracking,” in *Proc. Int. Conf. Image Process.*, vol. 1, Oct. 2004, pp. 163–166.