Improved Redundant Picture Coding Using Polyphase Downsampling for H.264

Jie Jia, Hae-Chul Choi, Jae-Gon Kim, Hae-Kwang Kim, and Yilin Chang

This paper presents an improved redundant picture coding method that efficiently enhances the error resiliency of H.264. The proposed method applies polyphase downsampling to residual blocks obtained from inter prediction and selectively encodes the rearranged residual blocks in the redundant picture coding process. Moreover, a spatial-temporal sample construction method is developed for the redundant coded picture, which further improves the reconstructed picture quality in error prone environments. Simulations based on JM11.0 were run to verify the proposed method on different test sequences in various error prone environments with average packet loss rates of 3%, 5%, 10%, and 20%. Results of the simulations show that the presented method significantly improves the robustness of H.264 to packet loss by 1.6 dB PSNR on average over the conventional redundant picture coding method.

Keywords: H.264/AVC, error resiliency, redundant coded picture, polyphase downsampling.

I. Introduction

H.264/AVC is the latest video coding standard developed by the Joint Video Team (JVT) of ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Expert Group (MPEG) aiming at enhanced coding efficiency and improved network adaptation [1]. The standard was designed to cover a wide range of video applications. Among those applications, video service over unreliable networks is a challenging task in terms of both compression efficiency and error resiliency. However, enhanced coding efficiency somewhat increases the sensitivity of a codec to channel errors in error prone environments. Especially in a packet loss environment, the loss of a whole or partial picture due to packet loss may further result in a crash of the decoder. Therefore, there is a demand for error resiliency tools which make the codec robust to data losses or errors.

To meet this demand, current H.264 has been developed with error resiliency tools (A detailed description may be found in [2]-[6]). Among those tools, slice-based flexible macroblock ordering (FMO) is an effective tool for combating data loss [7]. This tool releases the constraint of macroblock (MB) raster scan order for slice coding; instead of doing that, it flexibly groups MBs into slice groups using an MB allocation map. An example of this is to assign MBs in a checker board mode, as shown in Fig. 1. In this case, when MBs in one slice group are lost, the lost MBs will be concealed by their neighboring MBs which are coded in other slice groups. Apparently, it makes it easier to construct the lost MBs. However, the error concealment uses spatially distant information of neighboring MBs; therefore, the artifacts may degrade the concealed effect. Also, the reconstruction quality greatly depends on the availability of the neighboring MBs.

Manuscript received June 08, 2006; revised Oct. 09, 2006.

Jie Jia (email: joriejiejia@gmail.com) is with the ISN National Key Laboratory, Xidian University, Xi'an, China, and also with the Department of Computer Engineering, Sejong University, Seoul, Korea.

Hae-Chul Choi (email: choihc@etri.re.kr) and Jae-Gon Kim (email: jgkim@etri.re.kr) are with Radio and Broadcasting Research Division, ETRI, Daejeon, Korea.

Hae-Kwang Kim (phone: + 82 2 3408 3757, email: hkkim@sejong.ac.kr) is with the Department of Computer Engineering, Sejong University, Seoul, Korea.

Yilin Chang (email: ylchang@xidian.edu.cn) is with the ISN National Key Laboratory, Xidian University, Xi'an, China.

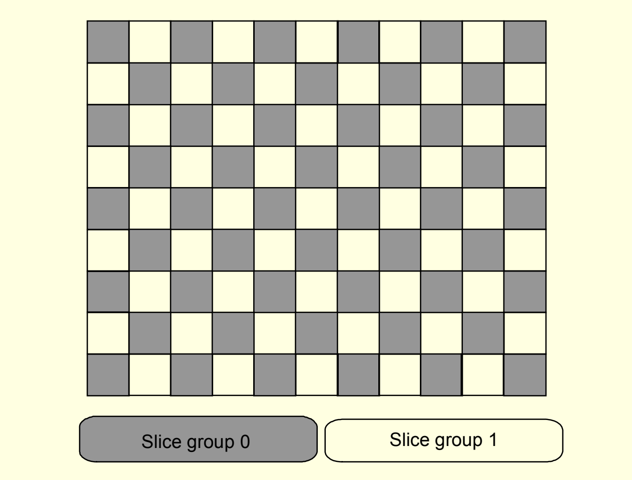


Fig. 1. Checker board mode for FMO.

Another highlighted error resiliency tool is *redundant picture* which has been developed to enhance the robustness of H.264 to data loss. With redundant picture, when the primary coded picture [1] is lost, a redundant one can be used to replace the lost picture in the decoding process, a simple implementation of which is to use a repetition of the primary coded picture as the redundant coded one. However, as the same bits are used to represent the redundant coded picture, the coding efficiency is decreased. Therefore, an alternative strategy was proposed, which uses fewer bits to represent the redundant picture, typically with somewhat degraded quality, for instance, using the greater quantization parameter (QP) in the redundant picture.

In this case, when the primary coded picture is lost, a picture with lower quality will be used to replace the lost one. For this reason, the performance is not significantly improved. Besides the methods described above, another redundant picture coding method has been recently proposed in [6], in which the redundant pictures are encoded with different reference pictures from those used for the primary picture coding. This effectively prevents temporal error propagation without relying on feedback information. However, as the redundant pictures are predicted from different reference pictures which are not the optimal ones used for primary picture coding, more bits will be used for the representation of the redundant ones. Coding efficiency still remains a problem.

In order to make a fair tradeoff between the coding efficiency and the error resiliency of H.264 in error prone packet loss environments, a polyphase-downsampling (PD)-based redundant picture coding method is investigated in this paper.

The remainder of this paper is organized as follows. Section II reviews the PD algorithm and section III describes the proposed redundant picture coding method. Simulation results are presented in section IV, and finally, conclusions are drawn in section V.

II. Polyphase Downsampling Algorithm

The PD technique applied to video coding had been investigated in [8] and [9] for improved coding efficiency and in [10] and [11] for multiple description (MD) coding. Moreover, an improved scheme, where PD is applied in the spatial domain, was developed in [12]. A PD-based MD coding combined with data partition in H.264 was proposed in [13] for error resiliency. The PD technique was further extended to improve scalable video coding efficiency in [14]. Until then, the PD technique had been applied to primary picture coding. After that, a PD-based redundant picture coding was proposed in [15]. Further work on this has been encouraged by JVT since the 17th JVT meeting in October 2005. The technique described in [15] became the basis of the current work presented in this paper.

Inspired by FMO, instead of using spatially distant information of neighboring MBs for error concealment, a method utilizing the PD technique is investigated, where the spatial correlation among neighboring samples is exploited. With the proposed method, when errors occur, the lost samples are more effectively concealed with the information of neighboring samples.

For a better understanding of the PD algorithm used in this paper, an exemplification of the PD procedure is illustrated in Figs. 2 and 3 along with the corresponding inverse PD procedure.

Basically, the PD procedure is a sample reorganization process, which rearranges the original sample in the position (r_o,c_o) to the position (r_p,c_p) given by (1) and (2):

$$r_p = (r_o \bmod 2) \times 4 + (r_o / 2)\,, \tag{1}$$

$$c_p = (c_o \bmod 2) \times 4 + (c_o / 2)\,. \tag{2}$$

For the case of inverse PD, it can be represented as (3) and (4). The samples rearranged by the PD process, as shown in Fig. 3(a), will be reorganized back to their original positions, as

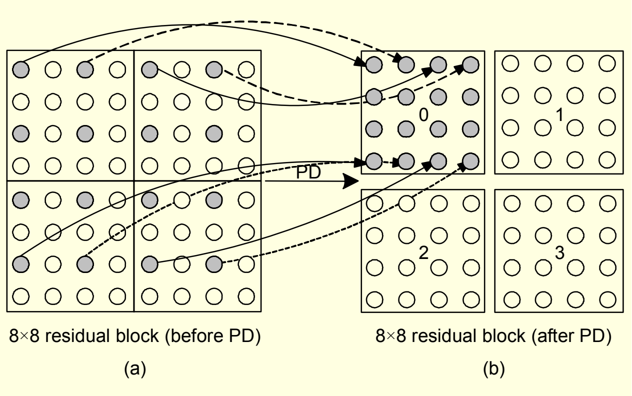


Fig. 2. Exemplification of PD procedure.

shown in Fig. 3(b). Further description of the PD technique applied to redundant picture coding is given in section III.

$$r_o = 2\times[r_a-(r_a/4)\times 4]+(r_a/4)\,, \tag{3}$$

$$c_o = 2\times[c_a-(c_a/4)\times 4]+(c_a/4)\,. \tag{4}$$

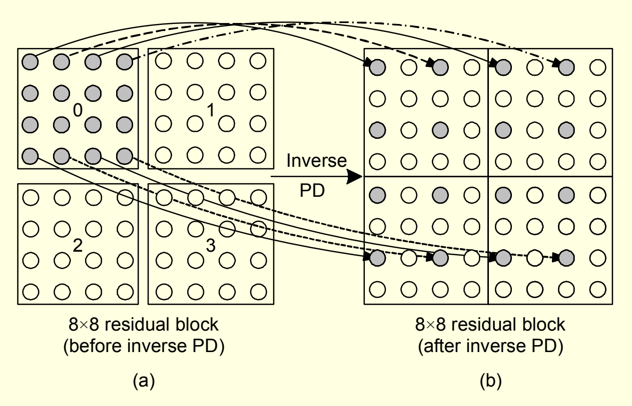


Fig. 3. Exemplification of inverse PD procedure.

III. PD-Based Redundant Picture Coding

This section describes the proposed method in detail, which is investigated in both the encoding and decoding process.

1. PD Process for Residual Blocks

The proposed method is developed for redundant picture coding; therefore, in the following description, the coding process refers to the redundant picture coding process.

Figure 4 gives the block diagram of the PD-based H.264 redundant picture coding structure. Two functional blocks—the PD block and the inverse PD block—are added to the coding process. First, PD is applied to each 8×8 prediction residual block which is the difference between the current block and the reference block obtained from previously-decoded pictures. After that, transform is applied to the rearranged residual blocks, followed by quantization and entropy coding.

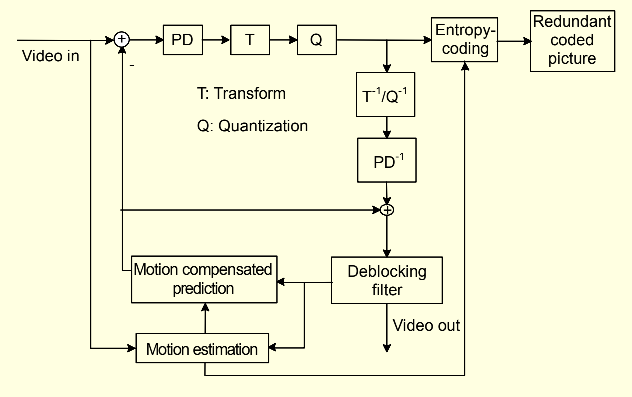


Fig. 4. Block diagram of the redundant picture coding.

Inverse PD, denoted by PD^{-1} in Fig. 4, is implemented in both the encoding and decoding process. Hence, an inverse PD module would be added for a normal H.264 decoder. Inverse PD is applied to the decoded 8×8 residual blocks obtained from the inverse transform in both the encoder and the decoder, followed by sample construction and a deblocking process.

2. Encoding Process for Rearranged Residual Blocks

In order to maintain coding efficiency while providing error resiliency, the redundant picture is typically coded with the higher QP, which results in more zero coefficient levels in the quantized transform block. This inspired the idea not to encode all the rearranged residual blocks, but to selectively encode only some of them. Figure 5 gives the block group patterns for the selective block coding. The rearranged residual blocks are classified into two block groups: one includes blocks to be coded; the other comprises blocks not to be coded.

Since some of the residual blocks are not to be coded, information related to those residual blocks is not considered for reference in the coding process of blocks being coded. Therefore, regarding the coding process of transform coefficient levels, when context-based adaptive variable length coding (CAVLC) is used, the following modifications are needed for the prediction of the number of non-zero transform coefficient levels:

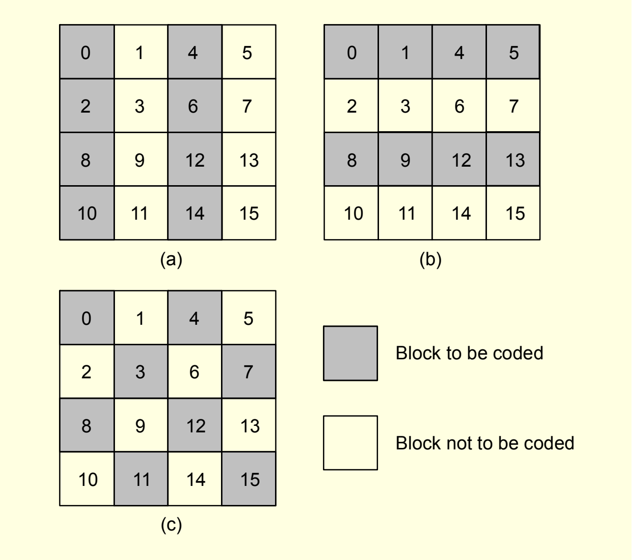


Fig. 5. Block group patterns for selective block coding.

a) If the blocks to be coded and the blocks not to be coded are distributed as in Fig. 5(a), then the available neighboring block is limited to the top neighboring block. In this case, the number of non-zero transform coefficient levels is predicted from that of the top neighboring block.

b) If the blocks to be coded and the blocks not to be coded are distributed as in Fig. 5(b), then the number of non-zero transform coefficient levels is predicted from that of the left neighboring block.

c) If the blocks to be coded and the blocks not to be coded are distributed like a checker board as shown in Fig. 5(c), then the number of non-zero transform coefficient levels is predicted from that of the top-left neighboring block.

3. Sample Construction Process for Redundant Pictures

Only some of the prediction residual blocks have been coded; therefore, for those samples whose prediction residual have not been coded, a spatial-temporal sample construction method is developed for the corresponding picture construction process as illustrated in Fig. 6.

Using the PD technique in the coding process, each sample whose prediction residual has not been encoded is surrounded by four samples whose prediction residual have been encoded. Therefore, in the redundant picture decoding process, those samples whose residual signals have not been encoded will be constructed by referring to not only the inter predicted samples obtained from the previously decoded picture, but also the previously decoded neighboring samples within the current picture.

Two steps are included in this construction process. First, a zero prediction residual is assumed to have been decoded for the samples whose prediction residual have not been encoded. In this way, the temporally predicted sample value is primitively used as the constructed sample value. Then the average of the spatial neighboring samples is used to adjust the primitive sample value. For simplicity, in this paper, an experimental threshold of 10 is used in the simulations. When the difference between the primitive value and the average value is larger than the threshold, the average value is used to replace the primitive sample value. Further study is needed on threshold selection, which will further improve the reconstructed picture quality.

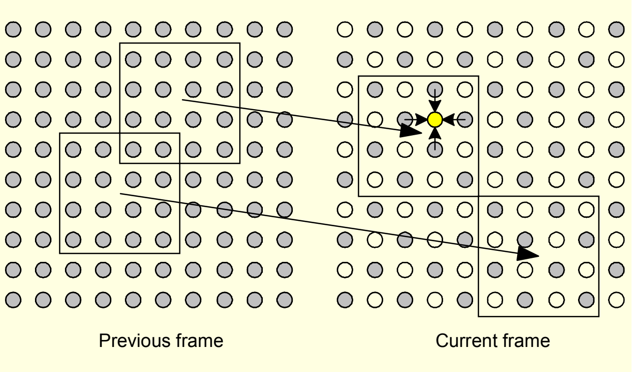


Fig. 6. Spatial-temporal sample construction.

IV. Simulation Results

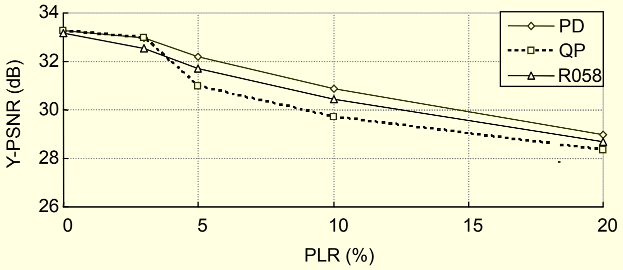
To evaluate the performance of the proposed method, simulations were run based on the reference software JM11.0 [16]. Two methods were used as anchors in this study. In the first method, the redundant picture is coded with higher QP (denoted by QP). The other method is described in [6], and is denoted by R058 in the following descriptions. Both the proposed method and the anchor methods were tested on several standard video test sequences (YUV 4:2:0): Foreman.qcif (400 frames), News.qcif (300 frames), and Stefan.cif (300 frames). The coding parameters were set as follows:

- sequence coding type: IPPP, with one redundant picture coded for every other frame for the PD method and the higher QP method
- intra period: 1 second
- number of reference frames: 5, in the R058 method, when the redundant slice is enabled, the parameter *NumberReferenceFrames* set to be no less than the parameter *PrimaryGOPLength*
- entropy coding method: CAVLC
- RD optimization: on
- allocation of redundant pictures in the R058 method: GOP size = 10 and *NumRedundantHierarchy* = 2 used for News and Stefan sequences; GOP size = 12 and NumRedundantHierarchy = 2 used for Foreman sequence
- multi-slice not used, each coded slice assumed to be encapsulated in one RTP packet
- forced intra MB update not used
- *RestrictRefFrames* = 1 and *UseConstrainedIntraPred* = 1.

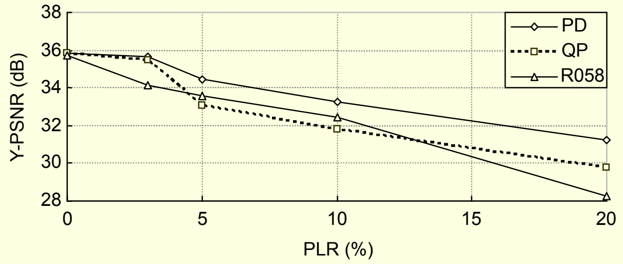
For the higher QP anchor, the value of the QP used for the redundant picture was 3 larger than that used for the primary picture. For the proposed method, the block group pattern for selective block coding as illustrated in Fig. 5(c) was used.

Four packet loss patterns, as specified in ITU-T VCEG and [17], [18], with the average PLR of 3%, 5%, 10%, and 20%, were employed in the simulations. A 0 in the packet loss pattern indicates a packet loss, while a non-zero numeral indicates a successful packet transmission. Note that, regarding residual blocks obtained from intra prediction, the conventional coding process is applied.

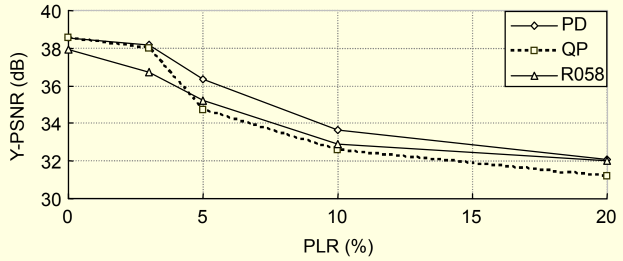
Figures 7 through 9 compare the performance of the PD-based



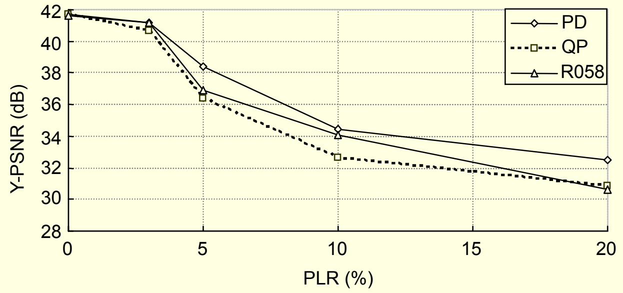
(a) 64 kbps



(b) 128 kbps



(c) 192 kbps



(d) 320 kbps

Fig. 7. Performance comparison between the proposed method (PD), the higher QP method (QP), and the method described in JVT-R058 (R058) in error prone environments—Foreman, QCIF, and 10 fps.

redundant picture coding method with that of the higher QP method and the R058 method in various packet loss environments. Figure 7 presents the simulation results for the Foreman sequence based on different bit rates, where the target bit rates of 64 kbps, 128 kbps, 192 kbps, and 320 kbps were tested. All four of the simulation results show that the proposed method achieves better reconstruction quality in terms of PSNR than the anchor methods. This is due to the fact that the proposed method uses fewer bits to represent the redundant coded picture while keeping as many as possible of the reconstructed sample values in the redundant coded picture the same as those in the primary coded picture. In this way, with the same bit rate, more redundant coded pictures can be used for the protection of the primary coded pictures. Furthermore, the experimental results demonstrate that, for each case, the higher the PLR, the more the proposed method outperforms the anchor. Up to 2 dB improvement can be observed when the proposed method is compared with the higher QP method and up to 3 dB enhancement can be seen when it is compared with the R058 method.

It should be noted that at lower PLR, the performance improvement of the proposed method over other methods is relatively minor (around 0.2 dB improvement). This is because at lower PLR, fewer redundant coded pictures are really used in the decoding process, resulting in similar performance for all three of the methods.

In general, similar performance improvements can also be observed from the simulation results for the Stefan sequence and the News sequence, as shown in Figs. 8 and 9. Also, it is clear that when the transmission environment becomes worse, the proposed method gives more a gradual degradation of the reconstructed picture quality than the higher QP method. There is one counterfactual point in Fig. 8. When the PLR is 3%, the anchor method in which the redundant pictures are coded with higher QP outperforms the proposed method by approximately 0.1 dB. This is due to the very infrequent use of redundant coded pictures in the decoding process when PLR is low.

As the simulation results demonstrate, the error resilience of H.264 has been significantly enhanced with the proposed method by an average 1.6 dB PSNR improvement over the higher QP redundant picture coding method and around 0.8 dB PSNR enhancement over the R058 method in various packet loss environments. Moreover, a further comparison between

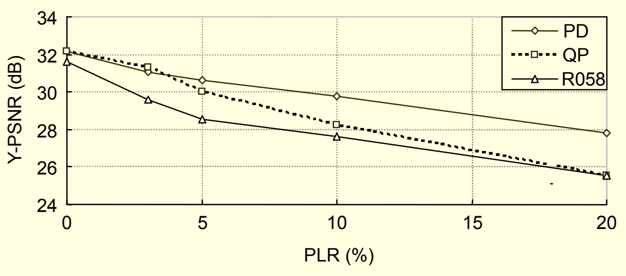


Fig. 8. Performance comparison between the proposed method (PD), the higher QP method (QP), and the method described in JVT-R058 (R058) in error prone environment—Stefan, CIF, 30 fps, and 1.0 Mbps.

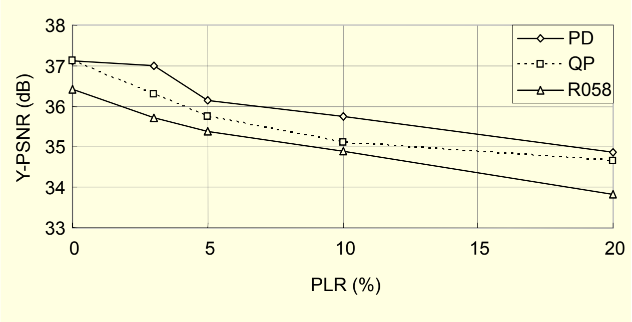
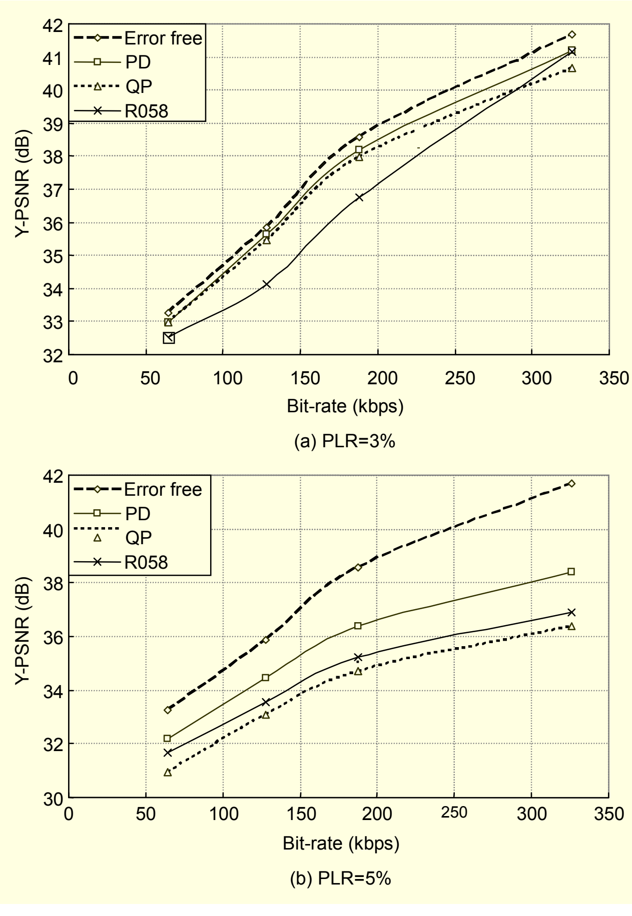


Fig. 9. Performance comparison between the proposed method (PD), the higher QP method (QP), and the method described in JVT-R058 (R058) in error prone environment—News, QCIF, 10 fps, and 64 kbps.

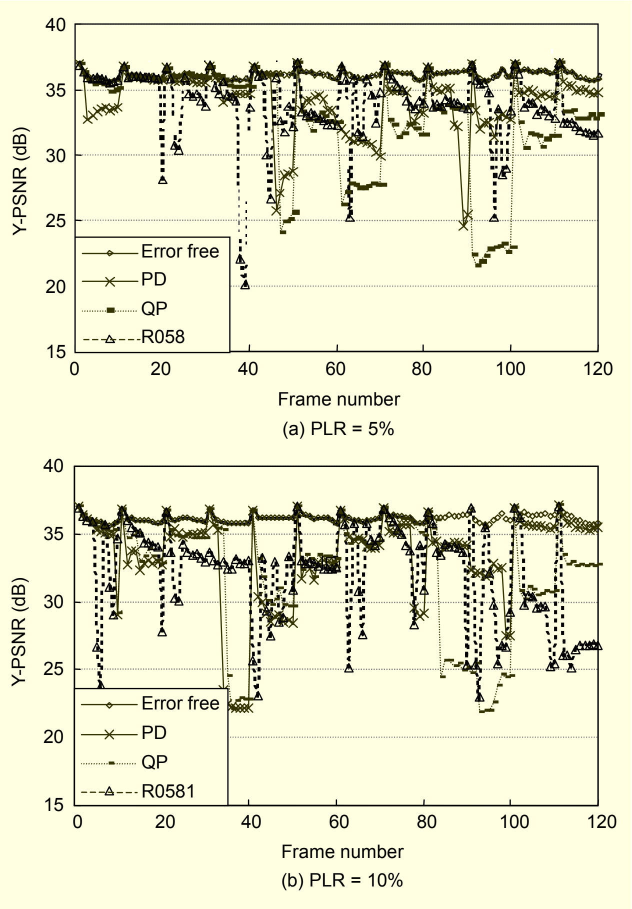


(a) PLR=3%

(b) PLR=5%

Fig. 10. Comparison of R-D performance between the proposed method (PD), the higher QP method (QP), and the method described in JVT-R058 (R058) in error prone environment—Foreman, 10 fps.

different test sequences shows that the proposed method achieves higher PSNR enhancement on high motion sequences like Foreman than on low motion sequences like News. This is due to the fact that as motion in video sequences increases, the loss of one picture will have an increasingly serious affect on the reconstruction quality of the following pictures. Therefore, the more the pictures are protected, the better the reconstruction quality will be.



(a) PLR = 5%

(b) PLR = 10%

Fig. 11. PSNR performance comparison against frame numbers between the proposed method (PD), the higher QP method (QP), and the method described in JVT-R058 (R058) in error prone environment—Foreman, 10 fps, and 128 kbps.

Figure 10 shows a comparison of R-D performance between the proposed method and the anchor methods in error prone environments for the Foreman sequence. It can be seen that at a lower PLR, a slightly better performance is obtained by the proposed method. At a higher PLR, for instance, when the PLR is 5%, the proposed method gives obvious consistent PSNR improvement from 1.1 dB to 2 dB over the higher QP method and from 0.5 dB to 1.5 dB over the R058 method. This is in accordance with the conclusions drawn from the previous analysis.

To further illustrate the improved error resiliency of H.264 with the proposed method, Fig. 11 gives the frame-by-frame objective PSNR performance of the three methods in detail.

The experimental results demonstrate that the proposed method outperforms the anchor methods most of the time. Quality degradation has been observably restricted with the


(a) (b)

(c) (d)

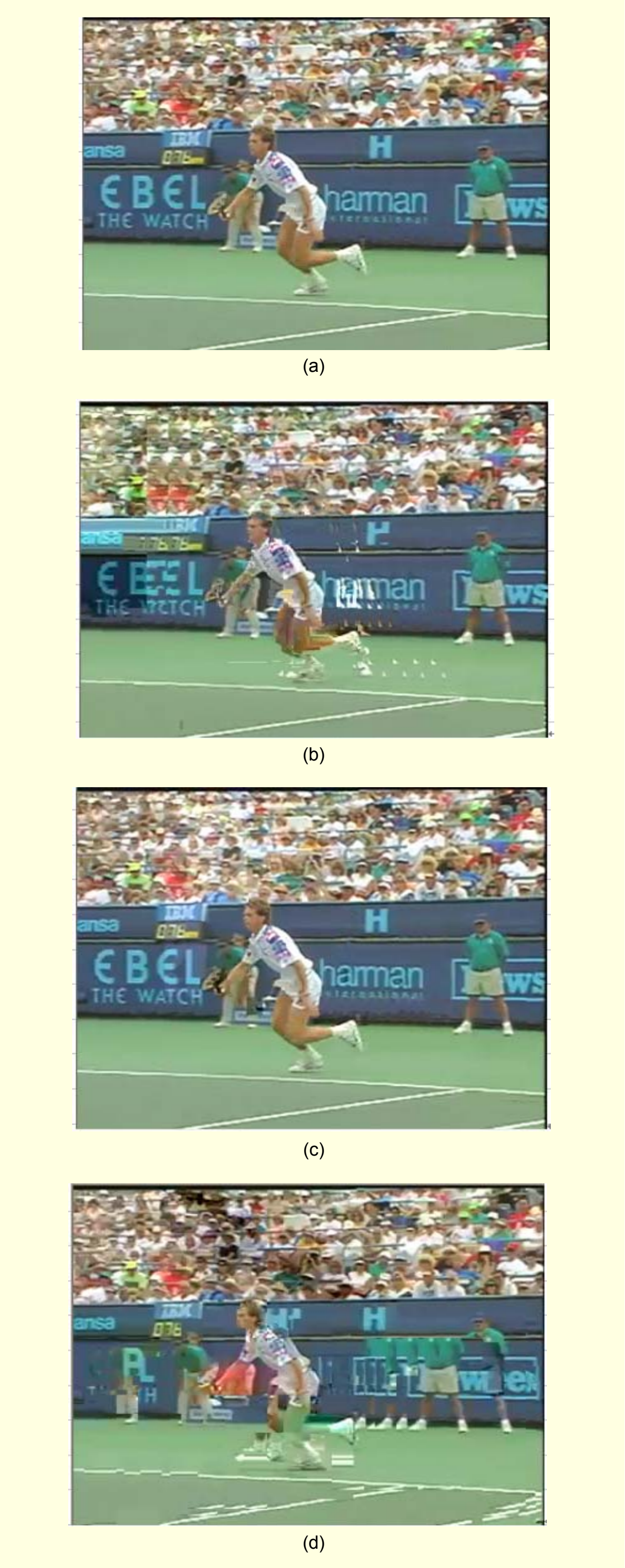
Fig. 12. Frame #92 of the reconstructed Foreman sequence coded at 128 kbps, with frame rate = 10 fps and PLR=10%: (a) error free, (b) redundant slice coded with higher QP, (c) JVT-R058, and (d) redundant slice coded with PD based method.

proposed redundant picture coding method.

For a subjective quality comparison, Fig. 12 presents reconstructed pictures from the Foreman sequence under error free and PLR=10% conditions. Figure 13 shows a comparison from the reconstructed Stefan sequence of frame #226 with PLR=20%. In both of the figures, (a) gives the reconstructed picture in error free conditions, while (b), (c), and (d) give the reconstructed picture in an error prone environment, where the redundant pictures are coded with the higher QP method, the method described in JVT-R058, and the proposed PD method, respectively. From the decoded pictures, it can be seen that the proposed method gives better reconstructed subjective quality than either of the anchor methods.

V. Conclusion

In this paper, an improved redundant picture coding method considering both the coding efficiency and the error resiliency of H.264 has been presented. The proposed method applies a polyphase downsampling technique to 8×8-sized inter prediction residual blocks, followed by selectively encoding part of the rearranged residual blocks. The spatial-temporal sample construction method proposed in this paper further improves the reconstructed picture quality in error prone environments. The results of simulations show that the proposed method efficiently enhances the error resiliency of H.264 in error prone environments.


(a)

(b)

(c)

(d)

Fig. 13. Frame #226 of the reconstructed Stefan sequence coded at 1.0 Mbps, with frame rate = 30 fps and PLR=20%: (a) error free, (b) redundant slice coded with higher QP, (c) JVT-R058, and (d) redundant slice coded with PD based method.

Acknowledgment

The authors would like to thank Dr. G.J. Sullivan, Dr. T. Wiegand, and Dr. Y.K. Wang for their insightful discussions and comments on this work. The authors would also like to thank Chunbo Zhu for making the anchor software available for this work.

References

[1] ITU-T, *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec.H.264, Mar. 2005.

[2] T. Wiegand, G.J. Sullivan, G. Bjǿntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. on CSVT*, vol. 13, no. 7, Jul. 2003, pp. 560-576.

[3] T. Stockhammer, M.M. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments," *IEEE Trans. on CSVT*, vol. 13, no. 7, Jul. 2003, pp. 657-673.

[4] S. Wenger, "H.264/AVC Over IP," *IEEE Trans. on CSVT*, vol. 13, no. 7, Jul. 2003, pp. 645-656.

[5] Y. Wang and Q.F. Zhu, "Error Control and Concealment for Video Communication: A Review," *Proc. of the IEEE*, vol. 86, no. 5, May 1998, pp. 974-997.

[6] C. Zhu, H. Li, Y.K. Wang, and M.M. Hannuksela, "Coding of Redundant Pictures for Improved Error Resilience," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-R058, Bangkok, Thailand, Jan. 2006.

[7] Y. Dhondt and P. Lambert, "Flexible Macroblock Ordering: an Error Resilience Tool in H.264/AVC," *Fifth FTW PhD Symposium*, Faculty of Engineering, Ghent University, Dec. 1, 2004, paper no. 106.

[8] H.C. Huang and J.L. Wu, "Real-Time Software-Based Moving Picture Coding (SBMPC) System," *Signal Processing: Image Comm*., vol. 6, no. 2, 1994, pp. 173-187.

[9] Y.K. Wang and G.F. Tu, "Study of Decimation and Interpolation and its Application to DCT Image Coding," *Proc. 16th IFIP World Computer Congress/2000 Int'l Conf. on Communications Technology* (WCC'2000/ICCT'2000), Beijing, China, Aug. 2000, pp. 1.981-1.984.

[10] M. Caramma, M. Fumagalli, and R. Lancini, "Polyphase Down-Sampling Multiple Description Coding for IP Transmission," *Proc. SPIE 2001—Visual Communications and Image Processing*, San Jose, CA, 2001, pp. 545-552.

[11] N. Franchi, M. Fumagalli, and R. Lancini, "Flexible Redundancy Insertion in a Polyphase Down-Sampling Multiple Description Image Coding," *Proc. ICME 2002*, Lausanne, Switzerland, Aug. 2002, pp. 605-608.

[12] N. Franchi, M. Gumagalli, R. Lancini, and S. Tubaro, "Multiple Description Video Coding for Scalable and Robust Transmission over IP," *IEEE Trans. on CSVT*, vol. 15, no. 3, Mar. 2005, pp. 321-334.

[13] J. Jia and H.K. Kim, "Error Resilience for SVC Base and Enhanced Layers," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-O048, Busan, Korea, Apr. 2005.

[14] J. Jia, H.K. Kim, H.C. Choi, and J.G. Kim, "An Improvement on Coding Efficiency for SVC Enhancement Layer by Polyphase Down Sampling," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-P051, Jul. 2005.

[15] J. Jia, H.K. Kim, H.C. Choi, and J.G. Kim, "Polyphase Downsampling Based Redundant Slice Coding for SVC Error Resilience," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-P050, Jul. 2005.

[16] JM11.0.zip, http://iphome.hhi.de/suehring/tml/download/, Aug. 2006.

[17] Y. Guo, H.Q. Li, and Y.K. Wang, "SVC/AVC Loss Simulator," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-Q069, Oct. 2005.

[18] Y.K. Wang, S. Wenger, and M.M. Hannuksela, "Common Conditions of SVC Error Resilience Testing," *ISO/IEC JTC 1/SC 29/WG 11*, JVT-P206d1, Jul. 2005.

Jie Jia received the BS and MEng degrees in communication engineering from Xidian University, Xi'an, China, in 2000 and 2003, respectively. Since then, she has been working toward the PhD degree in communication engineering at Xidian University, Xi'an, China. She has been a researcher at Sejong University, Seoul, Korea, since 2005. Her current research interests include video compression, scalable video coding, and error- and packet-loss resilient video coding.

Hae-Chul Choi received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1997, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1999 and 2004, respectively. Since then, he has been a Senior Member of Research Staff, in the Broadcasting Media Research Group of Electronics and Telecommunications Research Institute (ETRI), Korea. His current research interests include video coding and scalable video coding. He is the co-chair of JVT SVC error resilience AHG.

Jae-Gon Kim received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1990, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1992 and 2005, respectively. Since 1992, he has been a Senior Member of Research Staff, in the Broadcasting Media Research Group of Electronics and Telecommunications Research Institute (ETRI), Korea. He is currently the Team Leader of the Convergence Media Research Team. From 2001 to 2002, he was a staff associate at Columbia University, New York. His research interests include video processing, networked video, multimedia applications, MPEG-7, and MPEG-21. He is a member of the IEEE.

Hae-Kwang Kim received the BS degree in electronic engineering from Hanyang University, Korea in 1986, and the D.E.A. and PhD in computer science from Paul-Sabatier University, France in 1994 and 1997, respectively. He worked on the development of multimedia systems and applications at Samsung Electronics from 1986 to 1992. He worked on MPEG-4 and MPEG-7 multimedia standard technology domains at Hyundai Electronics from 1997 to 2000. He has been working on MPEG-4, MPEG-7, and MPEG-21 technology at Sejong University since 2000. His current research interests include video coding, multimedia retrieval, digital content protection, digital contents, electronic commerce, and the adaptation of computer graphics contents.

Yilin Chang was born in Jiangsu, China in 1944. He received the BS degree in Xi'an Jiaotong University and the MS degree in Xidian University. His major interests are in the fields of networked multimedia and network management. He has published over 100 papers.