ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Signal Processing: *Image Communication* ∎ (∎∎∎∎) ∎∎∎–∎∎∎

SIGNAL PROCESSING: IMAGE COMMUNICATION

www.elsevier.com/locate/image

H.264 error resilience coding based on multi-hypothesis motion-compensated prediction

Yuh-Chou Tsai[a], Chia-Wen Lin[b,*], Chia-Ming Tsai[c]

[a]*Cyberlink Corp., Taipei 100, Taiwan*

[b]*Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300, Taiwan*

[c]*Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi 621, Taiwan*

Received 11 November 2006; received in revised form 15 March 2007; accepted 22 May 2007

Abstract

In this paper, we propose efficient schemes for enhancing the error robustness of multi-hypothesis motion-compensate predictive (MHMCP) coder without sacrificing the coding efficiency significantly. The proposed schemes utilize the concept of reference picture interleaving and data partitioning to make the MHMCP-coded video more resilient to channel errors, especially for burst channel error. Besides, we also propose a scheme of integrating adaptive intra-refresh into the proposed MHMCP coder to further improve the error recovery speed. Extensive simulation results show that the proposed methods can effectively and quickly mitigate the error propagation and the penalty on coding efficiency for clean channels due to the inserted error resilience features is rather minor.

Keywords: Multi-hypothesis prediction; Video streaming; Error resilience coding; Wireless video; H.264 video coding

1. Introduction

Multimedia applications, such as communications, entertainments, and surveillance, are enabled and getting more and more popular by the evolution of technologies in personal computers and networks. In multimedia applications, compression is a crucially important issue since raw video/audio data demand tremendous storage space and network bandwidth. To reduce the temporal redundancy in raw video data, motion-compensated prediction (MCP) [1] has been widely exploited in the contemporary video coding standards. With MCP, incoming video frames are predicted from previously reconstructed frames stored in frame buffers by spatial displacement vectors (namely, motion vectors). Since only the motion vectors and the quantized prediction errors are coded in the bitstream, the more accurate the prediction, the more compression gain can be achieved. To achieve better prediction accuracy, long-term memory motion compensation (LMMC) was proposed in Ref. [2] to predict the video frames more effectively among multiple reconstructed frames, leading to bit-rate saving of up to 30% using 50 reference frames. The H.264 standard [3] adopts LMMC as a coding tool to achieve high coding efficiency.

*Corresponding author. Tel.: +886 3 573 1152; fax: +886 3 571 5971.

E-mail addresses: michael_tsai@gocyberlink.com (Y.-C. Tsai), cwlin@ee.nthu.edu.tw (C.-W. Lin), tsaicm@cs.ccu.edu.tw (C.-M. Tsai).

0923-5965/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2007.05.002

The concept of multi-hypothesis MCP (MHMCP) was first introduced in Ref. [4], which reported that prediction error reduction between 18% and 40% (0.9–2.2 dB) can be obtained with MHMCP. Subsequent research works [5–7] have been dedicated to the theoretical analysis of the compression improvement and the integration of the MHMCP into a standard video codec. A theoretical analysis in [5] discusses performance bounds for hybrid video coding using MHMCP. In [6], an optimal hypothesis selection and predictor coefficient selection algorithm was proposed. A rate-constrained mode decision algorithm was also proposed in [7]. The combination of LMMC and MHMCP is natural and can achieve better coding efficiency. Simulation results in [7] show that the combination of MHMCP (two hypotheses), variable block-size coding, and long-term memory (10 reference frames) saves up to 30% bit-rate for the same video quality. The performance of generalized B-pictures in H.264, that also adopts MHMCP, was investigated in [8], in which the issues about syntax adjustment, rate-constrained mode decision, entropy coding model adjustment (CABAC), and multi-hypothesis motion estimation were addressed.

Besides the compression issues in multimedia applications, the robustness of multimedia data against transmission errors is also essential, especially for video streaming over unreliable networks. Due to the variable length coding (VLC) adopted in most video coder, a small data error may result in loss of synchronization. Coded data succeeding to the loss will become undecodable even if they are correctly received. Furthermore, due to the MCP exploited in most existing coding standards, the distortion of one video frame which results from transmission errors will propagate to its subsequent frames. These reasons make compressed video data very sensitive to channel errors and cause serious quality degradation. Thus, enhancing the error resiliency of video streams in video coding/transcoding has become a crucial problem [9,10].

The error robustness of MHMCP coders against transmission errors has been investigated recently. In [11], Kim et al. proposed a double-vector motion compensation (DMC) scheme that adopts two-hypothesis MCP (2HMCP) by which each macroblock (MB) is predicted from the weighted superposition of two MBs in two reference frames using two motion vectors. The proposed DMC coder can effectively suppress the error propagation caused by the loss of one reference MB by using the other reference MB for motion compensation, should only one reference MB be corrupted. Lin and Wang [12] presented a decoder distortion model for a given channel condition and a simplified encoder distortion model of a 2HMCP coder. Their proposed algorithm selects the optimal hypothesis weighting factors which minimize the total distortion. In [13], a thorough mathematical analysis was developed to characterize the error propagation effect caused by packet loss in an MHMCP coder. The error resiliency of MHMCP is also analyzed and compared to the well-known random intra-refresh (RIR) scheme. It was concluded in [13] that MHMCP can more effectively suppress the short-term error propagation, whereas IR outperforms the MHMCP for stopping long-term error propagation.

In this work, we propose efficient schemes to further enhance the error robustness of an MHMCP coder with two forward hypotheses involved in the prediction process while retaining good coding efficiency. The contribution of our work is twofold. First, we propose two efficient encoder-side error resilience coding tools that exploit the concept of reference picture interleaving and data partitioning to improve the error robustness of MHMCP-coded video against channel errors, especially for burst error situations. We shall show that an efficient combination of the encoder-side tools and the decoder-side error concealment method proposed in [11] further enhances the error resiliency. Besides, we also propose a scheme of integrating adaptive IR into the proposed MHMCP coder to improve the speed of error recovery.

The rest of this paper is organized as follows: Section 2 gives an overview on MHMCP. The proposed approaches enhancing the resilience of 2HMCP codec with theoretic investigation are presented in Section 3. Simulation results of compression and resilience performance will be given in Section 4. Conclusions are drawn in Section 5.

2. Multi-hypothesis motion-compensated prediction

In a traditional predictive coder, a MB prediction is reconstructed by a given spatial displacement vector (motion vector) and a temporal displacement index (reference frame index). With these two parameters, a decoder can find the prediction of an MB/sub-block among the frame buffer, which is reconstructed during the decoding process of previous frames. In an MHMCP coder, MB/sub-block predictions for the kth frame are formed by linear combinations of multiple hypotheses, each is a block of pixels specified by a spatial displacement vector and a temporal displacement index as formulated in (1). Note that if the weighting factor for each hypothesis varies, the weighting information should also be signaled in the video bitstream

$$\hat{\psi}(k) = \sum_{i=1}^{n} h_i \tilde{\psi}(k, k-i), \quad \text{where } \sum_{i=1}^{n} h_i = 1, \tag{1}$$

where $\hat{\psi}(k)$ is the multi-hypothesis prediction of the kth frame, h_i stands for the weighting factor for each hypothesis, n represents the number of hypotheses, and $\tilde{\psi}(k, k-i)$ represents the motion-compensated prediction for the kth frame using the $(k-i)$th frame as reference.

In this work, we focus on the 2HMCP codec, in which each MB is predicted by a linear combination of two forward hypotheses, which are selected from the $(k-1)$th and $(k-2)$th reference frames, respectively, using two weighting factors, h and $1-h$, as expressed by (2):

$$\hat{\psi}(k) = h\tilde{\psi}(k, k-1) + (1-h)\tilde{\psi}(k, k-2). \tag{2}$$

Fig. 1 illustrates how the 2HMCP coder deals with two situations of frame loss, including a single-frame loss and a burst loss of two consecutive frames. Due to the leaky prediction adopted in the 2HMCP coder, not all the concealment distortion of the corrupted frame will propagate to the next frames since part of the MB prediction (that is, another hypothesis) comes from the previous error-free reference frame. As a result, the distortion would be attenuated and the error propagation is thereby suppressed. A distortion model for the 2HMCP encountering a single-frame loss was presented in [13]. In this model, the error propagation in terms of mean squared error (MSE) to the $(k+i)$th frame, when the kth frame is lost, is derived by the MSE value of the lost frame $(D_d(k))$ as follows [13]:

$$D_d(k+i) = \left(\frac{1-(h-1)^{k+1}}{2-h}\right)^2 \frac{\theta}{1+\gamma_d i} D_d(k), \tag{3}$$

where γ_d and θ are constant model parameters used to take into account the effect of sub-pixel motion compensation.

According to Eq. (3), the 2HMCP coder reacts to a single-frame loss very quickly and mitigates the error propagation more effectively. For transmission in slow-fading wireless channels, burst error, which will usually result in serious video quality degradation, occurs more frequently than transmission in wireline channels. In the additive model presented in [14], the effect of multiple losses is modeled as the superposition of multiple independent losses. This additive model usually cannot characterize a burst loss very well since there typically exist inter-frame/packet dependencies among the lost frames/packets.

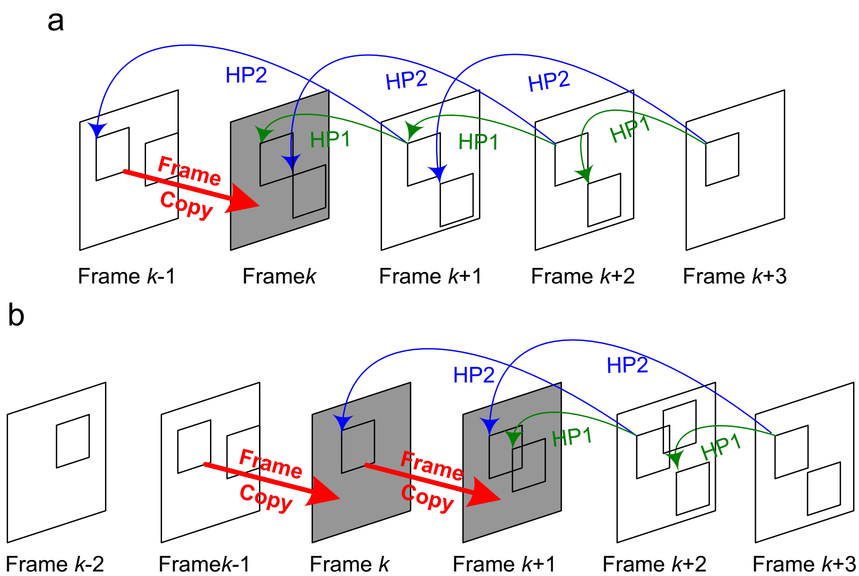


Fig. 1. Illustration of the original 2HMCP scheme encountering loss of frames: (a) a single-frame loss and (b) a loss of two consecutive frames.

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

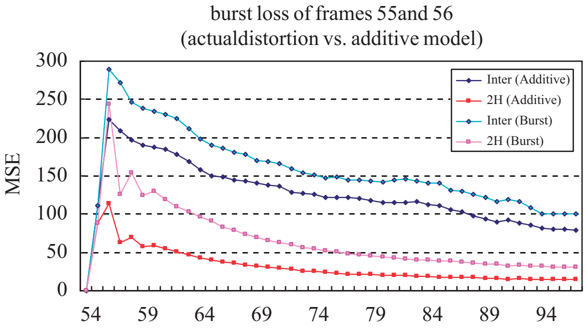


Fig. 2. Comparison of error propagation effects: measured burst-loss distortions versus the additive model approximations.

In [15], a more accurate model about the error propagation effect of a burst loss is presented, which shows that a burst loss generally produces a larger distortion than that produced by summing up the individual distortions of an equal number of isolated single-frame losses using the additive model. Therefore, a burst loss should be treated more carefully than an equal number of individual single-frame losses. Fig. 2 illustrates how worse the burst error can be. The "Inter (Burst)" and "2H (Burst)" curves represent the actual MSE distortion values obtained from the single-hypothesis and two-hypothesis coders, respectively, when a two-frame burst-loss occurs (in this example, both the 55th and 56th frames are lost). "Inter (Additive)" and "2H (Additive)" represent the corresponding estimated MSE distortions obtained by using the additive model that sums up the channel distortions caused by two independent single-frame losses at the 55th and 56th frames together. Evidently, the actual distortions due to a burst-loss (i.e., the "Inter (Burst)" and "2H (Burst)" curves) are significantly higher than the distortions estimated by the additive model (i.e., the "Inter (Additive)" and "2H (Additive)" curves). Therefore, a more effective approach against burst error is desirable for video streaming over lossy channels.

3. Proposed error-resilient H.264 coding scheme for MHMCP

In predictive coding, when a video frame is lost during transmission, the decoder reconstructs the lost frame by using certain error concealment method, which usually results in reconstruction distortion. Such reconstruction distortion will lead to the mismatch between the encoder's and the decoder's reconstructed frames, thereby leading to severe drifting error. The drifting error will propagate to the succeeding frames all the way until reaching an IR point. In the 2HMCP coder, the error propagation still exists since the predictive coding scheme is exploited as well. To mitigate the error propagation due to packet loss, we propose efficient error resilience coding methods at the encoder and decoder sides, respectively. We also provide theoretical analyses to justify the performance of the proposed methods.

3.1. Encoder-side error resilience coding tools

3.1.1. Hypothesis parameters separation

Though the 2HMCP can suppress the error propagation well, the distortion of a lost frame may still be serious. The frame loss cannot be easily recovered by error concealment in the decoder. To reduce the distortion of a lost frame and its error propagation, we propose a hypothesis parameter separation scheme.

In the 2HMCP coder, each hypothesis requires a motion vector and a reference picture index to obtain the predicted block from the corresponding frame buffer. In what follows, the term "hypothesis parameters" is referred to as the motion vectors and the reference block indices required to form the prediction for the two hypotheses.

For the mth MB of the kth frame, there are two sets of hypothesis parameters required for specifying the two hypotheses used for prediction. These parameters are both coded and embedded into the locations between the MB header and the transform coefficients of residues of the mth MB in the bitstream. When a frame is lost during transmission, all the data, including the hypothesis parameters and residue signals, will also get lost since they are usually all encapsulated in the same packet, thereby resulting in significant quality degradation.

The motivation of the scheme is to mitigate the impact caused by a single-frame loss by separating the most important data into two partitions. The hypothesis parameters of the mth MB of the kth frame is separated and embedded into the mth MBs of the kth frame and the $(k-1)$th frame, respectively, as illustrated in Fig. 3(a). As a result, when a single frame is lost, only half of the hypothesis parameters and all the transform coefficient of residues are lost. The decoder can recover the erroneous frame more accurately with the remaining half of the hypothesis parameters rather than reconstructs the lost frame by copying the prior

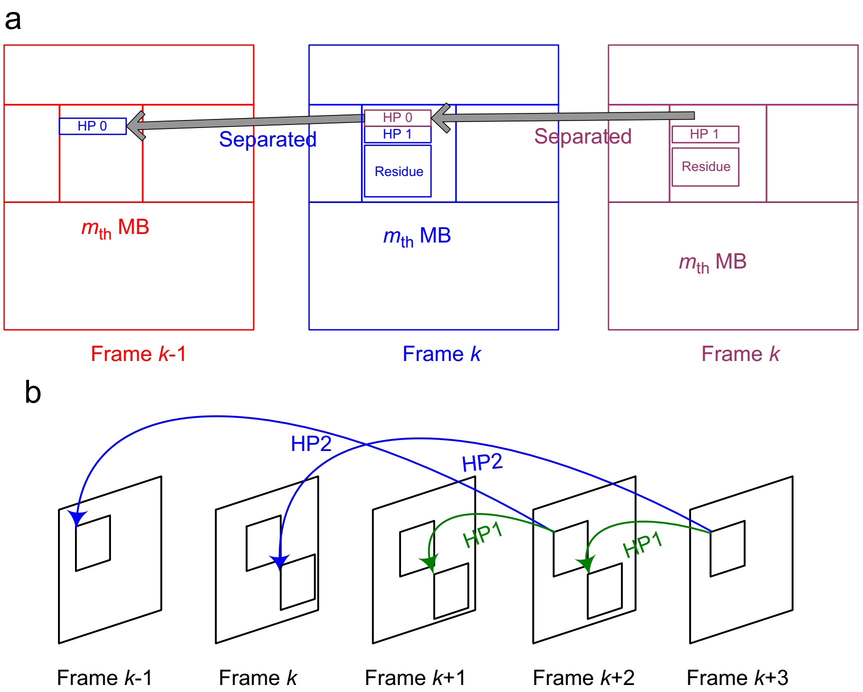


Fig. 3. Encoder-side tools: (a) hypothesis parameter separation and (b) reference interleaving scheme.

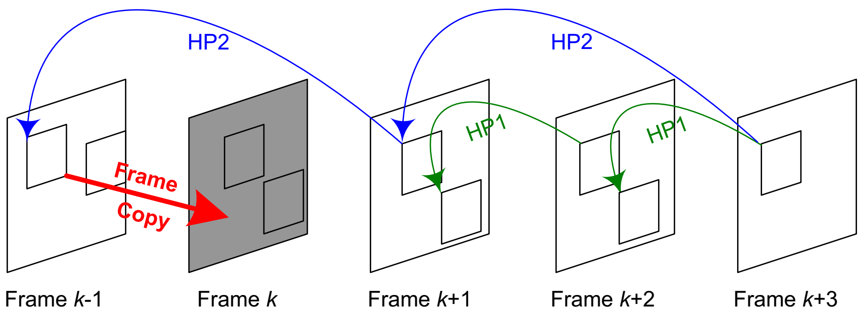


Fig. 4. Decoder side tool: clean hypothesis formation.

reconstructed frame (i.e., the zero-MV error concealment [9]). In this way, the drifting error caused by packet loss would be significantly reduced.

The merit of the hypothesis parameter separation is to mitigate the distortion caused by frame loss without embedding any additional overhead. Besides, this approach does not require a feedback channel between a sender and a receiver. It will, however, introduce extra one-frame encoding latency and storage cost. In order to separate the parameters, the encoder needs to buffer the compressed data of the current encoding frame until the hypothesis parameters for the next frame are available, thereby requiring the extra encoding delay of one frame and the extra memory space for buffering one compressed video frame. Since compressed data of one frame normally will not occupy too much memory space, the extra storage cost should not be a critical issue. However, inserting hypothesis parameters into dynamic locations of a pre-compressed video frame may raise a few complexity issues in system implementations such as flexible data structure and circuit design, which needs to be carefully treated.

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

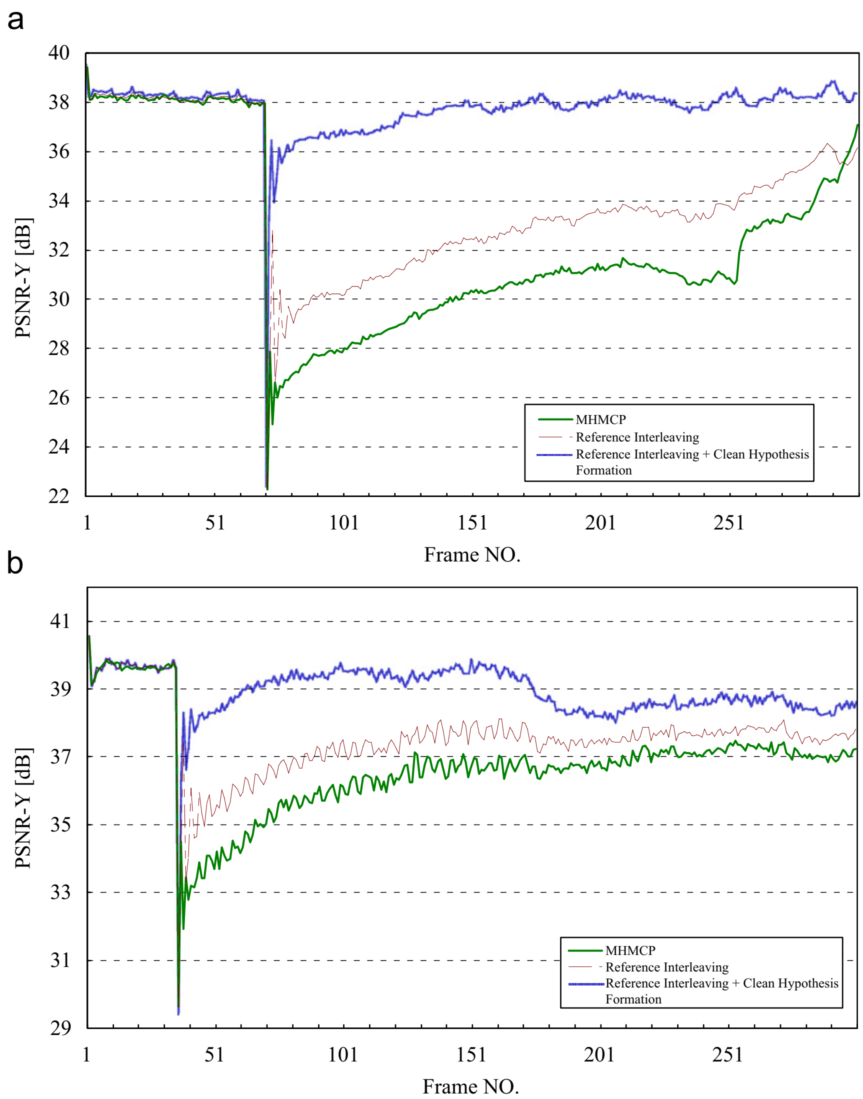


Fig. 5. Illustration of the proposed 2HMCP scheme encountering loss of frames: (a) a single-frame loss and (b) a loss of two consecutive frames.

3.1.2. *Reference Interleaving*

The original design of 2HMCP may suffer from burst packet loss a lot, since both hypotheses will be lost in the original 2HMCP when a burst loss of length larger than two happens. As a result, the next frame after the lost frames may suffer from serious error propagation which is difficult to recover.

We propose a "Reference Interleaving" scheme to mitigate the impact of a burst-loss with a length of two or more. In our method, we consider only the cases with a burst-length of two or less because it was shown in [16] that the average burst length caused by random packet loss is around 1.5–2. For a burst loss with a long burst-length caused by slow signal fading, packet interleaving can be used to effectively spread out the long burst loss into short individual packet losses. Fig. 3(b) illustrates the reference interleaving method with an interleaving distance of two to combat against packet losses with a length of two or less, in which the two hypotheses are selected from the $(k-1)$th and $(k-3)$th reference frames, respectively. This reference selection can effectively mitigate the error propagation caused by a burst loss, since a clean hypothesis still exists when two consecutive frames are lost.

3.2. Decoder-side error concealment

At the decoder, we adopt the error concealment scheme proposed in [11], namely "clean hypothesis formation" as illustrated in Fig. 4, to make best use of the proposed encoder-side tools so as to further mitigate the error propagation effect caused by packet loss. Assume that frame k is lost during transmission. At the decoder side, the predictions of MBs within subsequent frames (frame k and frame $k+1$) that contain the hypothesis predicted from frame $k-1$ will only be estimated with the clean

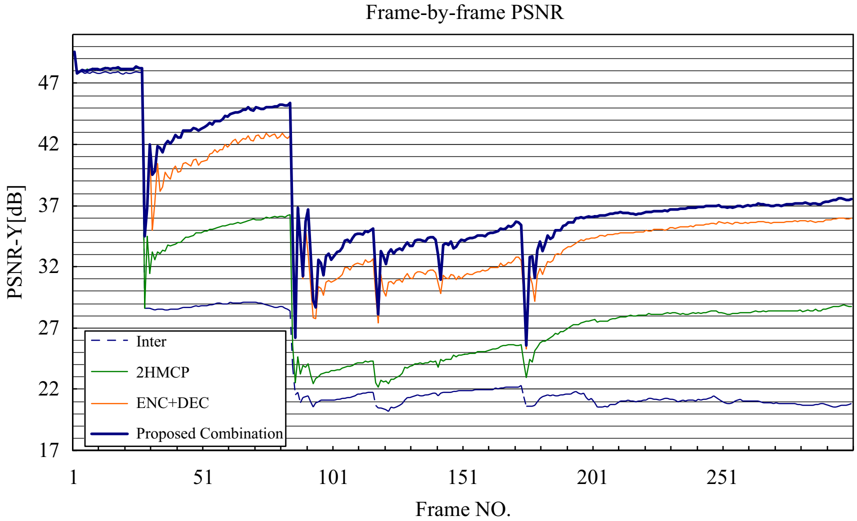


Fig. 6. Frame-by-frame PSNR performance comparisons of the conventional 2HMCP, Reference interleaving, reference interleaving + clean hypothesis formation under a single-frame loss with QP = 24 for two test sequences: (a) *Foreman* (frame #70 lost) and (b) *Carphone* (frame #35 lost).

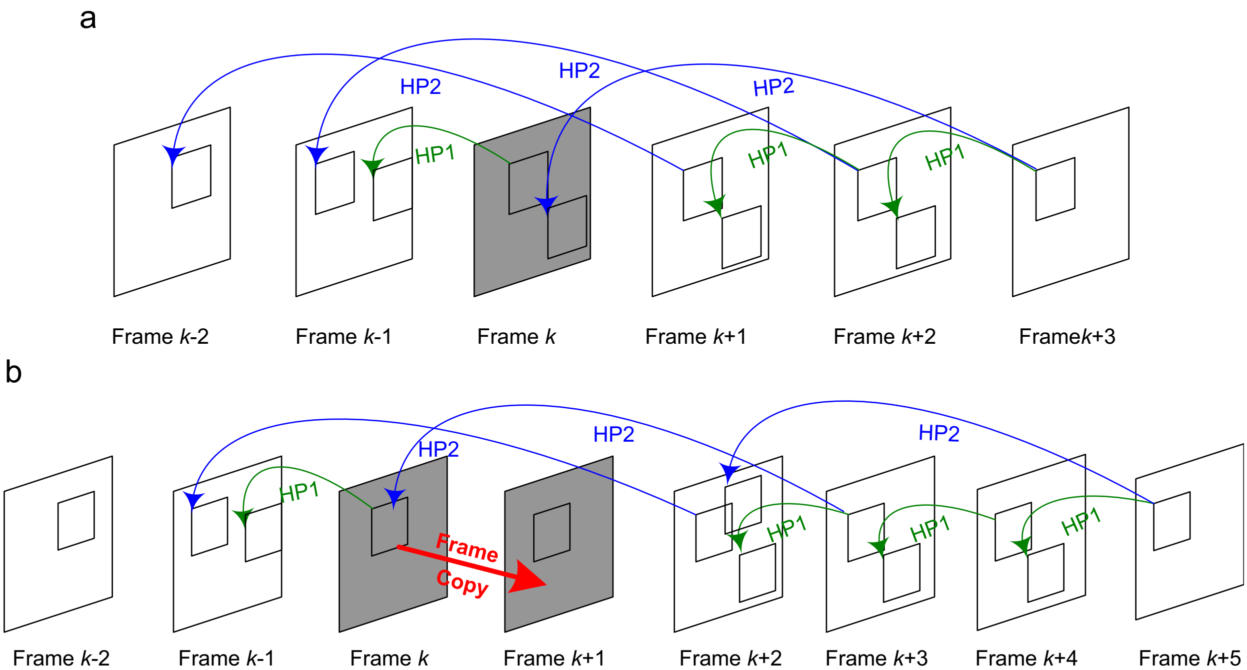


Fig. 7. Rate–PSNR performance comparisons of four methods for the *Carphone* sequence (QP = 12, PLR = 5%).

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

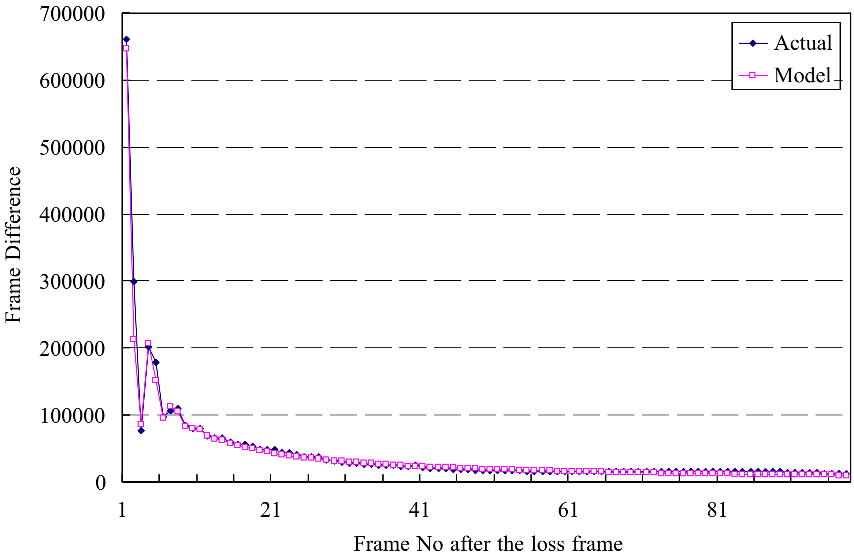


Fig. 8. Comparison of the derived model and the actual distortion with a single-frame (frame #50) loss for the *Foreman* sequence.

Table 1
Encoding configuration for 2HMCP

QP	12, 15, 18, 21, …, 36
GOP	IPPMMMM…
Frame rate	30 fps
No. of reference frames	2 (*n*-1 and *n*-3 are used as references)
Block mode	8 × 8
Rate-control	N/A (fixed QP)
ME search range	16
ME accuracy	Quarter-pel
Entropy coder	CAVLC

hypothesis. That is, the corrupted hypotheses are discarded when constructing the predictions of the subsequent frame at the decoder side.

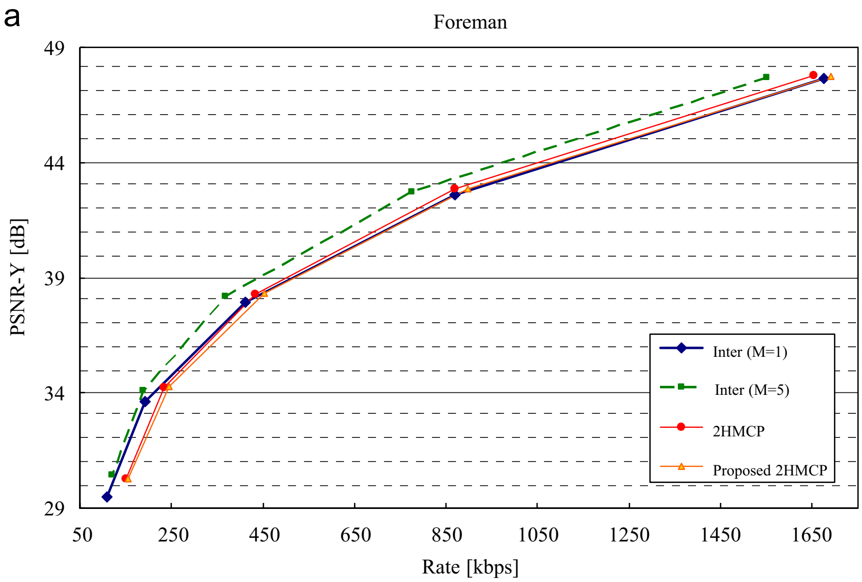
With this method, the distortion of the frames after the erroneous one can be reduced significantly compared to the conventional 2HMCP method [11]. Although the drifting error due to the reconstruction distortion is not avoidable, the overall resilience against frame loss is significantly improved, especially for single-frame loss situations.

3.3. Combination of the encoder-side and decoder-side tools

Since the proposed encoder-side error resilience coding tools are orthogonal to each other, they can be combined together to further benefit from each other. Note that, the decoder-side error concealment scheme would not be useful without separating the hypothesis parameters or reasonably estimating the parameters. Our simulation results show that the combination of the "hypothesis parameters separation + reference interleaving" tools (denoted as ENC) and the "clean hypothesis formation" (denoted as DEC) tool works well for burst-loss situations, whereas the ENC tools themselves performs better for single-loss situations. The reason is, using Fig. 5(a) as an example, if only frame k is lost, it may be reconstructed with reasonably good quality from the available hypothesis using the combination of tools. Therefore, when reconstructing frame $k+1$, it is possible to obtain a better reconstruction from frame $k-2$ (error-free) and frame $k-1$ (concealed) with two motion vectors by turning off "cleaning hypothesis formation" than from only a single hypothesis (frame $k-2$) by turning on the decoder tool. We therefore propose to combine the ENC and DEC tools with a little modification. If the frame before the lost frame that is the reference of the second hypothesis of the current frame is not lost during transmission, then the DEC tool is disabled.

As shown in Fig. 6, we compare the PSNR performances of the conventional 2HMCP, reference interleaving, and reference interleaving + clean hypothesis formulation under a single-frame loss with $QP = 24$ for the *Foreman* and *Carphone* sequences. The Reference Interleaving scheme achieves significant PSNR performance improvements over the conventional 2HMCP coder. The combination of Reference Interleaving and Clean Hypothesis Formation (i.e., the method in [11]) achieves further substantial PSNR

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002



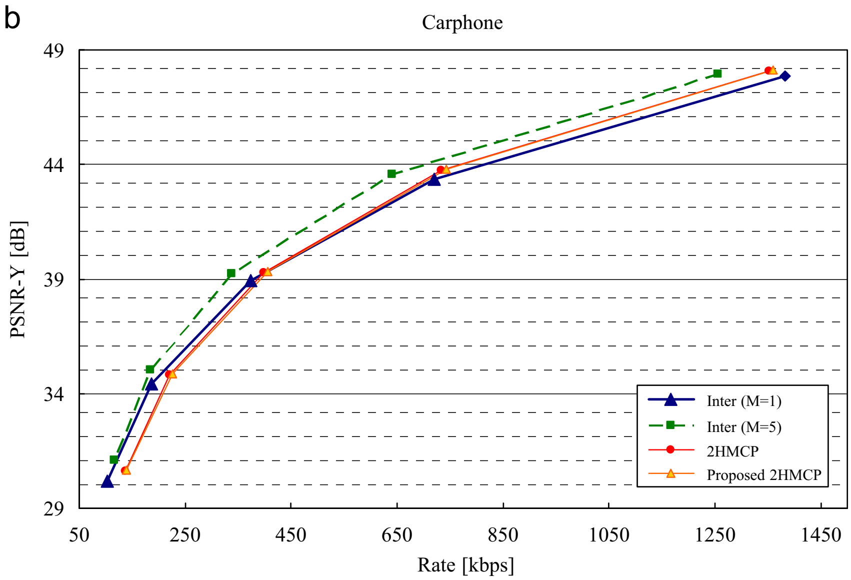


Fig. 9. Comparisons of average rate–PSNR performances of single-hypothesis MCP with one and five reference frames, the original 2HMCP, and the proposed 2HMCP in an error-free channel for: (a) *Foreman* and (b) *Carphone*.

performance gains. Our experimental results also show that the adaptive combination always achieves better resilience except for only a few special cases. Fig. 7 depicts the error resilience performance comparison of the single-hypothesis MCP, the conventional 2HMCP, 2HMCP with a direct combination of ENC + DEC, and the proposed combination method for the *Carphone* sequence with QP = 12 and PLR = 5%. We can observe that the proposed combination of encoder and decoder error resilience tools always yields significantly better results than the original 2HMCP.

3.4. Analysis of the joint error resilience coding method

The error propagation of the proposed method is analyzed and compared to the conventional 2HMCP in this section. Assuming that the kth frame is lost during transmission, with the proposed HP Separation, as shown in Fig. 7(a), the kth frame can be reconstructed from the $(k-1)$th frame using the first-half hypothesis parameters. The mismatch between the reconstructed values of the kth frame at

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

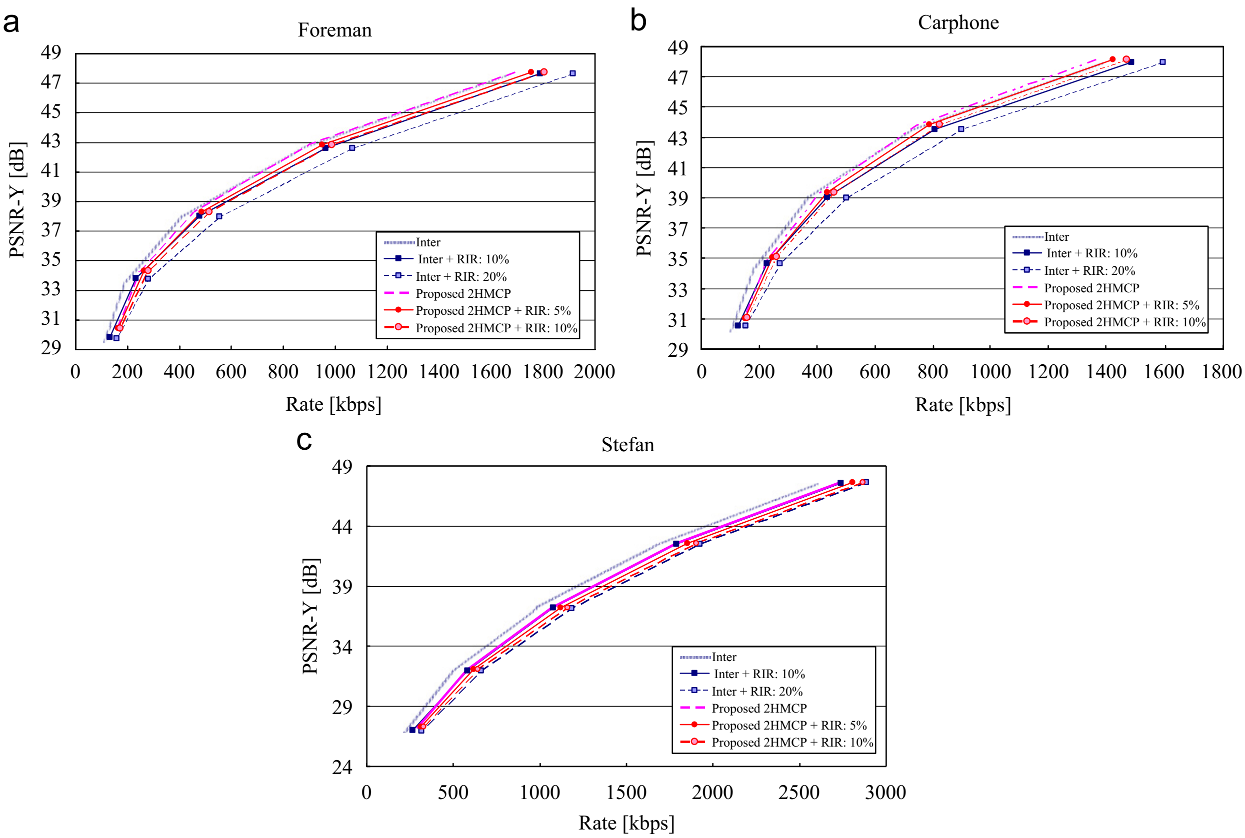


Fig. 10. Comparisons of average rate–PSNR performances of MCP and the proposed 2HMCP with different IR rates in an error-free channel for: (a) *Foreman*, (b) *Carphone*, and (c) *Stefan*.

the encoder and the decoder is as follows:

$$\begin{aligned}\varepsilon[k] &= \tilde{f}_{\mathrm{e}}[k] - \tilde{f}_{\mathrm{d}}[k]\\ &= \underbrace{h\tilde{\psi}(k,k-1) + (1-h)\tilde{\psi}(k,k-3) + Q(r[k])}_{\tilde{f}_{\mathrm{e}}[k]}\\ &\quad - \underbrace{\tilde{\psi}(k,k-1)}_{\tilde{f}_{\mathrm{d}}[k]}\\ &= (1-h)(\tilde{\psi}(k,k-3) - \tilde{\psi}(k,k-1))\\ &\quad + Q(r[k]),\end{aligned} \tag{4}$$

where $\tilde{f}_{\mathrm{e}}[k]$ and $\tilde{f}_{\mathrm{d}}[k]$, respectively, represent the reconstructions of the kth frame in the encoder (i.e., the error-free reconstruction) and the decoder (i.e., the reconstruction by error concealment), $\tilde{\psi}(k,k-i)$ represents the prediction of the kth frame from the $(k-i)$th frame, $r[k]$ stands for the prediction residual of the kth frame, and $Q(\cdot)$ represents the quantization function.

Similarly, the reconstruction errors of the subsequent frames due to error propagation can be expressed as the following recursive equations.

$$\begin{aligned}\varepsilon[k+1] &= (\tilde{f}_{\mathrm{e}}[k+1] - \tilde{f}_{\mathrm{d}}[k+1])\\ &= (h\tilde{\psi}(k+1,k) + (1-h)\tilde{\psi}(k+1,k-2)\\ &\quad - \tilde{\psi}(k+1,k-2))\\ &= (h(\tilde{\psi}(k+1,k) - \tilde{\psi}(k+1,k-2))),\end{aligned} \tag{5}$$

$$\varepsilon[k+2] = h\varepsilon[k+1], \tag{6}$$

$$\varepsilon[k+i] = h\varepsilon[k+i-1] + (1-h)\varepsilon[k+i-2], \quad i \geqslant 2. \tag{7}$$

Similar to (3), the distortion propagation from the loss of the kth frame to the $(k+i)$th frame can be obtained by

$$D_{\mathrm{d}}(k+i) = E[\varepsilon^2(k+i)]\frac{\theta}{1+\gamma_{\mathrm{d}}i}, \tag{8}$$

where the factor $\theta/(1+\gamma_{\mathrm{d}}i)$ is incorporated to take into account the attenuation effect of sub-pixel motion compensation on the distortion [13,14]. Fig. 8 shows that the proposed theoretic model is rather accurate,

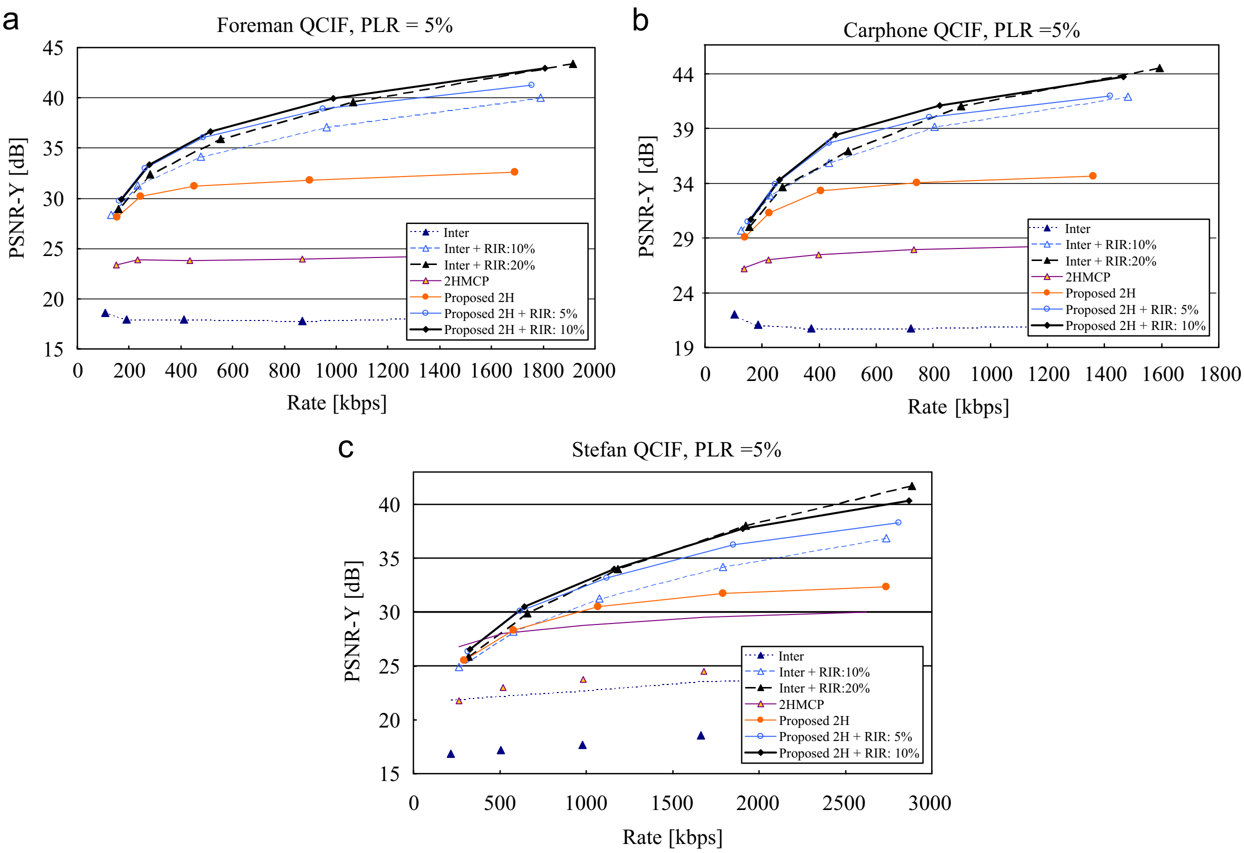


Fig. 11. Comparisons of average rate–PSNR performances of MCP, 2HMCP, and the proposed 2HMCP with different intra-refresh rates (PLR = 5%, 10 test patterns) for: (a) *Foreman*, (b) *Carphone*, and (c) *Stefan*.

and that the proposed 2HMCP coder does effectively mitigate the error propagation due to a single-frame loss. In these simulations, θ and γ are fixed parameters that are determined empirically. For the *Foreman* sequence, they are set to 1 and 0.3, respectively.

Compared to the conventional MHMCP coder, the proposed scheme results in much less distortions on frame $k+1$ and the subsequent frames if frame k gets lost. This is because, as derived in (5), $\varepsilon[k+1]$ with the proposed scheme is proportional to the difference between the two hypotheses, that is $\varepsilon(k+1) \propto \tilde{\psi}(k+1,k) - \tilde{\psi}(k+1,k-2)$. On the other hand, in the conventional 2HMCP, $\varepsilon[k] = h\varepsilon[k] = h(\tilde{f}_{\mathrm{e}}[k] - \tilde{f}_{\mathrm{d}}[k]) = h(\tilde{f}_{\mathrm{e}}[k] - \tilde{f}_{\mathrm{e}}[k-1])$, $\varepsilon[k+1]$ is proportional to the frame difference. The difference of two hypotheses (i.e., motion-compensated predictions) of the same block is usually much smaller than the direct difference of two consecutive frames. Therefore, the proposed scheme will usually provide better reconstruction for a single-frame loss.

For burst-loss situations, as shown in Fig. 5(b), the difference between error-free reconstructed frames and the decoder reconstructed frames (with transmission error), if the kth frame is corrupted, is formulated as

$$\begin{aligned}\varepsilon[k+1] &= (\tilde{f}_{\mathrm{e}}[k+1] - \tilde{f}_{\mathrm{d}}[k+1]) \\ &= \underbrace{(h\tilde{\psi}(k+1,k) + (1-h)\tilde{\psi}(k+1,k+2) + Q(r[k+1]))}_{\tilde{f}_{\mathrm{e}}[k+1]} \\ &\quad - \underbrace{\tilde{\psi}(k+1)}_{\tilde{f}_{\mathrm{d}}[k+1]},\end{aligned} \tag{9}$$

$$\begin{aligned}\varepsilon[k+2] &= (h\tilde{\psi}(k+2,k+1) + (1-h)\tilde{\psi}(k+2,k-1) \\ &\quad - \tilde{\psi}(k+2,k-1)) \\ &= (h(\tilde{\psi}(k+2,k+1) \\ &\quad - \tilde{\psi}(k+2,k-1))),\end{aligned} \tag{10}$$

$$\varepsilon[k+3] = (h\varepsilon[k+2] + (1-h)\varepsilon[k]), \tag{11}$$

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

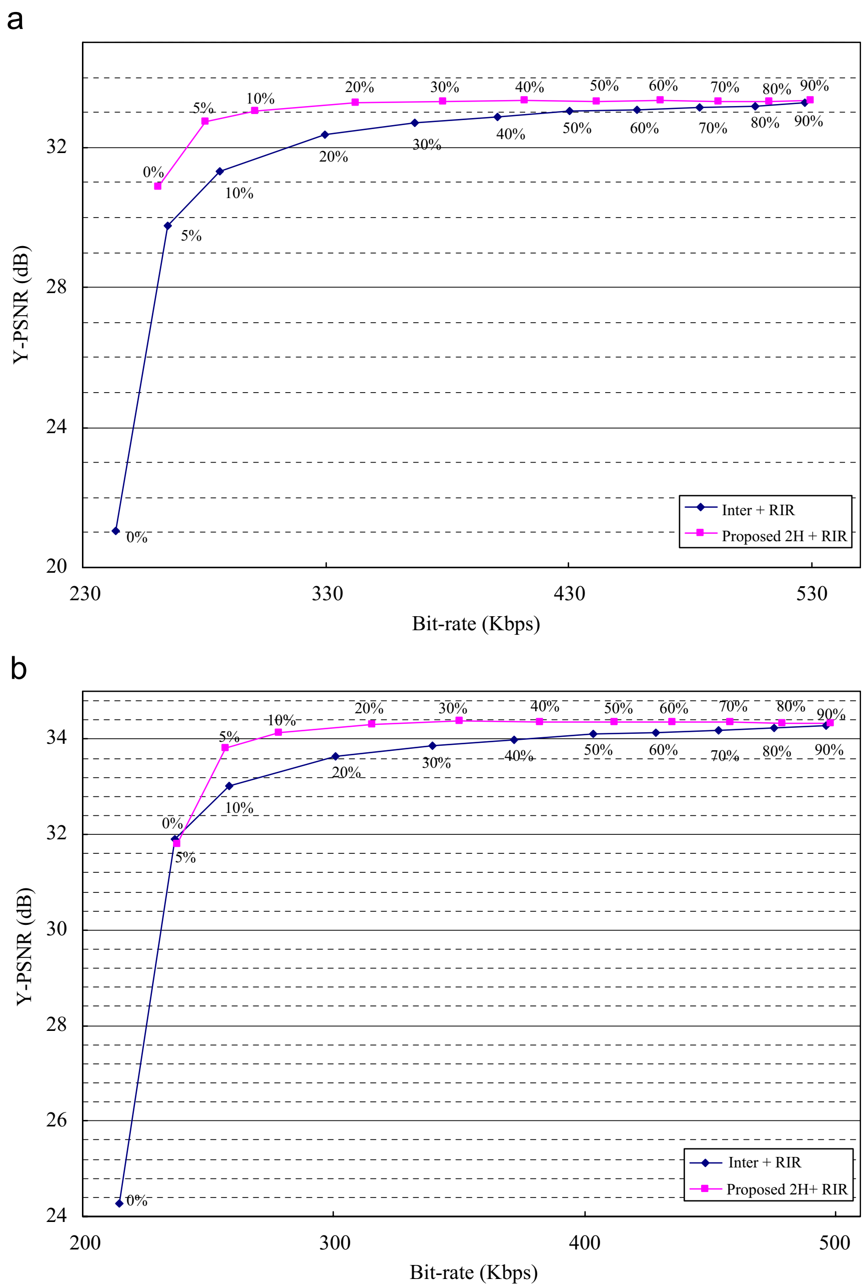


Fig. 12. Comparisons of average rate–PSNR performances versus the IR rate with the single-hypothesis and the proposed 2HMCP (QP = 30, PLR = 5%, 10 test patterns) for: (a) *Foreman* and (b) *Carphone*.

$$\begin{aligned}\varepsilon[k+4] &= (h\varepsilon[k+3] + h\tilde{\psi}(k+4,k+3) \\ &\quad + (1-h)\tilde{\psi}(k+4,k+1) - \tilde{\psi}(k+4,k+3)) \\ &= (h\varepsilon[k+3] + (1-h)\tilde{\psi}(k+4,k+1) \\ &\quad - \tilde{\psi}(k+4,k+3)), \qquad (12)\end{aligned}$$

3.5. *Adaptive intra-refresh*

The 2HMCP reacts to a packet loss swiftly, but it cannot completely recover from the packet loss. As for IR methods, it recovers from packet loss slowly

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

but in long term will completely eliminate the error propagations [13]. Due to this reason, appropriately combining the IR tool with the proposed 2HMCP will do a better job. Our simulation results justify that the combined method is able to compensate for the drawback of the 2HMCP coder, while benefiting from the merits of 2HMCP a lot.

In order to have better error resilience performance, we also propose an adaptive IR scheme based on our 2HMCP coder. Our scheme first computes for each MB the distortion caused by its corrupted hypotheses using the additive model. A predetermined number of MBs with top-ranking distortions are then intra-coded. The priority decision is summarized as follows:

Step 1: Encode the video with the GOP (group of pictures) structure: IPPMMM…IPPMMM…, where M-frames are 2HMCP-coded frames.

Step 2: For each M-frame, collect distortion_cost for each MB in frame k. Calculate the concealment error of frame k using (4) and (8). Calculate the error propagation from frame k to frame $k+1$ using (5) and (8). Use the additive model to estimate the distortion_cost $= \varepsilon[k]+\varepsilon[k+1]$.

Step 3: When encoding frame $k+1$ with the proposed 2HMCP, sort the distortion_cost for MBs

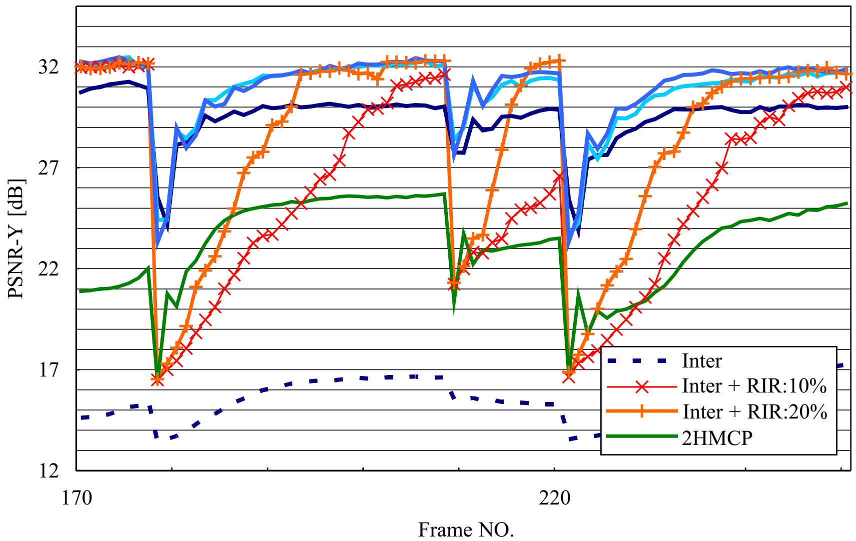


Fig. 13. Comparisons of frame-by-frame PSNR performances of MCP, 2HMCP and the proposed 2HMCP with various IR rates for *Stefan* (QP = 30, error pattern #2, frames 170–250).

Table 2
The average PSNR over 10 error patterns for *Foreman* with QP = 30 and PLR = 5%

QP = 30	M1	M5	2H	M1_IR10	M1_IR20	Proposed 2H	Proposed 2H + IR5	Proposed 2H + IR10
Error pattern 1	21.74	24.20	26.09	31.60	32.59	32.66	33.15	33.45
Error pattern 2	16.65	19.43	21.79	31.50	32.53	31.05	33.46	33.61
Error pattern 3	17.84	21.11	24.42	31.53	32.50	29.04	32.84	33.31
Error pattern 4	20.25	23.27	26.04	31.61	32.58	32.37	33.70	33.76
Error pattern 5	20.29	23.08	25.65	31.60	32.60	31.06	33.13	33.42
Error pattern 6	18.20	21.26	24.11	31.12	31.96	28.40	32.52	33.04
Error pattern 7	15.49	19.03	22.37	30.51	32.02	28.80	32.30	32.86
Error pattern 8	16.27	19.87	22.14	30.95	32.03	26.11	32.25	32.84
Error pattern 9	15.38	18.59	22.34	31.07	32.41	30.45	33.27	33.51
Error pattern 10	17.57	21.24	23.47	31.62	32.62	31.54	33.34	33.54
Average	17.97	21.11	23.84	31.31	32.38	30.15	33.00	33.33

in frame k and a number of MBs with top ranking distortion_cost will be intra-refreshed. The number of intra-coded MBs is calculated according to a predetermined IR rate.

Step 4: Perform Steps 2 and 3 iteratively to encode the rest M-frames in the current GOP until all the M-frames are processed. Go to Step 1 to encode the next GOP.

Note that, more sophisticated methods, such as ROPE [17] and the method in [18], can also be integrated with the proposed 2HMCP by modifying the error model to further improve the performance of adaptive IR at the cost of significantly higher computational complexity.

4. Experimental results

In our experiments, we modify the H.264 JM7.3 reference codec to implement 2HMCP. Three 300-frame QCIF test sequences: *Foreman*, *Carphone*, and *Stefan* are used for performance evaluation. The first frame of each GOP is coded as an I-frame, while the second and third frames are coded as P-frames. The I- and P-frames are the same as those coded with an H.264 coder. The remaining frames in the GOP are then all coded in the 2HMCP mode (M-frames). Each frame is encoded with various fixed quantization parameters, ranging from 12 to 36. Since the error resilience performance is the central focus of this work, for the sake of simplicity but without loss of generality, all MBs are coded in fixed 8×8 block size with quarter-pel precision ME. Furthermore, the mode decision for block size or inter /intra- mode is disabled, that is, all the MBs within one frame are coded in the same block mode (except for those MBs chosen to be intra-refreshed). Table 1 lists the encoding configuration of our experiments.

Table 3
The average PSNR over 10 error patterns for *Carphone* with QP = 24 and PLR = 5%

QP = 24	M1	M5	2H	M1 + IR10	M1 + IR20	Proposed 2H	Proposed 2H + IR5	Proposed 2H + IR10
Error pattern 1	22.74	24.58	27.72	35.52	36.77	36.47	41.50	43.09
Error pattern 2	19.70	21.77	25.30	35.81	37.01	31.07	36.84	37.78
Error pattern 3	19.17	21.97	27.01	36.22	37.51	32.60	37.59	37.98
Error pattern 4	21.09	23.18	27.20	36.31	37.34	34.79	37.85	38.32
Error pattern 5	23.93	26.32	29.43	36.54	37.63	34.91	37.17	37.88
Error pattern 6	21.62	25.48	29.40	36.44	37.33	31.89	37.15	37.77
Error pattern 7	19.56	23.63	27.81	36.26	37.46	34.92	37.43	38.07
Error pattern 8	20.65	23.19	28.34	36.24	37.49	29.69	36.46	37.34
Error pattern 9	19.10	21.38	25.26	32.90	33.74	32.83	37.49	37.92
Error pattern 10	20.21	22.81	27.46	36.13	37.34	33.71	37.23	37.90
Average	20.78	23.43	27.49	35.84	36.96	33.29	37.67	38.40

Table 4
The average PSNR over 10 error patterns for *Stefan* with QP = 24 and PLR = 5%

QP = 24	M1	M5	2H	M1_IR10	M1_IR20	Proposed 2H	Proposed 2H + IR5	Proposed 2H + IR10
Error pattern 1	19.83	20.44	24.81	31.10	33.83	30.86	33.34	34.36
Error pattern 2	17.23	18.58	23.71	30.49	33.37	31.27	30.60	30.85
Error pattern 3	15.19	17.53	23.97	31.43	34.13	30.19	33.29	34.26
Error pattern 4	20.02	20.74	25.18	32.00	34.14	30.52	34.10	34.71
Error pattern 5	19.62	20.58	24.21	31.20	33.97	32.08	33.99	34.54
Error pattern 6	17.45	18.91	22.86	30.84	34.08	30.00	33.27	34.07
Error pattern 7	16.02	17.61	22.59	30.34	34.02	30.18	33.12	34.07
Error pattern 8	17.05	18.90	23.21	32.25	34.38	27.64	31.88	33.51
Error pattern 9	14.92	16.10	22.19	31.36	33.75	30.31	33.66	34.45
Error pattern 10	19.44	20.71	24.6	31.45	34.20	31.89	34.24	34.87
Average	17.68	19.01	23.73	31.25	33.99	30.49	33.15	33.97

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

In our experiments, each frame is encapsulated into a single packet. For channel condition simulations, a two-state Markov model [19], which adopts a simplified Gilbert channel [20,21], is used to generate packet loss patterns. The channel packet loss rate (PLR) is set to 5% and the mean burst error length is set to two packets. Ten different patterns are generated for each channel PLR under this configuration. For performance comparison, we also implemented the RIR, which is a widely adopted error resilience coding tool, with refresh rates of 10% and 20%, respectively. The proposed 2HMCP method is also integrated with the RIR.

The average rate-PSNR performances of the proposed 2HMCP, the conventional 2HMCP, the inter-MCP with one and five reference frames are compared in Fig. 9. Fig. 10 compares the average rate–PSNR performances of the proposed 2HMCP and the single-hypothesis MCP with RIR for a wide range of bit-rates. In the two figures, 'Inter ($M = 1$)' and 'Inter ($M = 5$)' stand for Inter-MCP with one and five reference frames, respectively. '2HMCP' stands for the conventional 2HMCP, and 'Proposed 2H' represents the proposed method. The IR scheme used in Figs. 10–12 is RIR, where 'Inter + RIR: X%' and 'Proposed + RIR: Y%' stand for Inter-MCP with X% and the proposed 2HMCP with Y% out of total MBs of each frame being randomly selected to be intra-coded, respectively. The simulation results show that, with quarter-pixel ME, the coding efficiencies of single-hypothesis MCP and 2HMCP are very close, since the coding gain obtained from quarter-pel ME for the single-hypothesis MCP is comparable with the coding gain from adding an additional temporal hypothesis from another reference frame for 2HMCP. Overall, the 2HMCP coder performs better than (or equally with) the single-hypothesis MCP at medium to high rates, but poorer at low rates due to the extra overhead for sending the hypothesis parameters. The proposed 2HMCP coder lowers down the compression efficiency slightly compared to the conventional 2HMCP coder since the proposed Reference Interleaving lengthens the reference distance.

The error resilience performances of MCP, 2HMCP, and the proposed 2HMCP are compared in Figs. 11–13 and Tables 2–4. Fig. 11 illustrates the rate–PSNR curves for the three test sequences, respectively, where the average PSNR values for each test sequence are computed for 300 frames and 10 error patterns with PLR = 5%. The results show

Fig. 14. Subjective performance comparison of different methods for frame #80 of *Foreman* (PLR = 5%, error pattern #2, QP = 36): (a) Inter–MCP (14.33 dB), (b) 2HMCP (19.91 dB), (c) Inter-MCP with 20% IR rate (26.06 dB), and (d) the proposed 2HMCP with 10% IR rate (30.18 dB).

Fig. 15. Subjective performance comparison of different methods for frame #185 of *Stefan* (PLR = 5%, error pattern #2, QP = 36): (a) Inter-MCP (17.00 dB), (b) 2HMCP (23.42 dB), (c) Inter–MCP with 20% IR rate (23.47 dB), and (d) the proposed 2HMCP with 10% IR rate (26.86 dB).

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

that for the three test sequences, the proposed 2HMCP with 5% IR rate outperforms the single-hypothesis MCP with 10% IR rate by about 1.9 dB on average. It can even beat the Inter–MCP with 20% IR rate at low to medium bit-rates. As shown in Fig. 10, the coding efficiency of the proposed 2HMCP with 5% IR is better than those of the single-hypothesis MCP with 10% and 20% IR rates. Therefore, the proposed 2HMCP with a lower-rate IR significantly outperforms Inter MCP with the same or higher IR rate for most types of sequences in terms of both coding efficiency for clean channels and error robustness for lossy channels. Among these error patterns and QP settings, the proposed method evidently yields a significantly better error resiliency over the original 2HMCP coder. The average PSNR improvement ranges from 3 to 6 dB for the same bit-rates.

Increasing the IR rate will increase the error resiliency of a compressed video, while also reducing the coding efficiency. Since the impact of IR rate on the PSNR performance and output bit-rate is

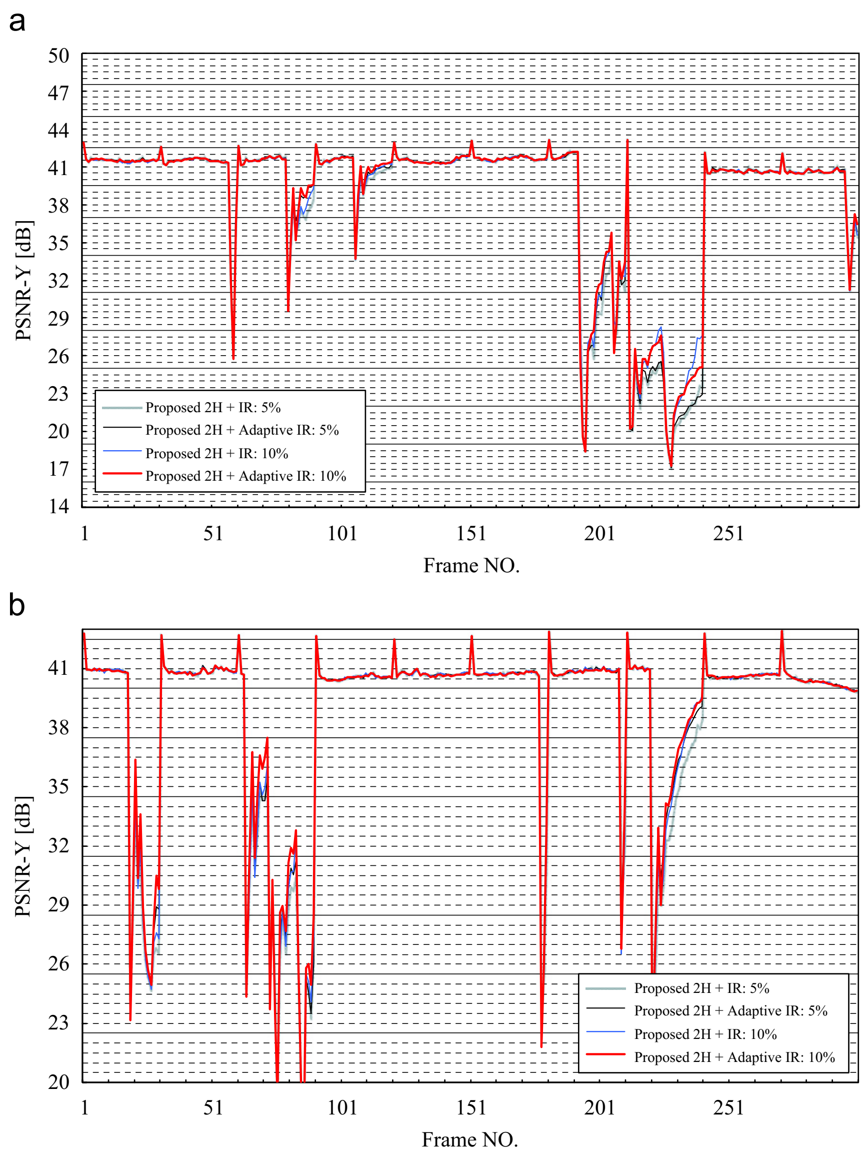


Fig. 16. Frame-by-frame PSNR performances of the proposed MHMCP with fixed and adaptive IR for: (a) *Foreman* (QP = 12; error pattern #1) and (b) *Stefan* (QP = 12; error pattern #2).

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

Table 5
Average PSNR performance comparisons for corrupted frames using proposed 2HMCP with random IR and adaptive IR

	Proposed 2HMCP + random IR (dB)		Proposed 2HMCP + adaptive IR (dB)	
	5% IR	10% IR	5% IR	10% IR
Foreman	34.36	34.89	34.52	35.02
Stefan	30.80	31.48	31.14	31.78

dependent on the PLR and the quantization parameters, we only show in Fig. 12 two examples about the rate–PSNR performances versus the IR rate under PLR = 5% and with QP = 30 for *Foreman* and *Carphone*, respectively. The numbers around the data points indicate the IR rates used. We can observe that the PSNR performance of the proposed 2HMCP saturates at the IR rate of about 20–30% in both cases. Applying an intra-fresh rate higher than the saturation rate only slightly improves the PSNR performance while significantly increasing the output bit-rate. For the single-hypothesis MCP coder, the saturation point of IR rate is about 50%. These comparisons justifies that the proposed coder has much stronger error resilience than the single-hypothesis coder. Others simulations with different quantization step sizes for different test sequences also show similar results.

Fig. 13 illustrates a frame-by-frame PSNR performance comparison of the three coders with various IR rates for *Stefan*. We can observe from Fig. 13 that the differences between the characteristics of IR and 2HMCP in terms of error resiliency and coding efficiency. The 2HMCP reacts to error quickly but is usually not able to terminate the error propagation entirely, whereas the IR scheme reacts to channel relatively slowly but can recover from error propagation almost entirely in the long run. Combining the proposed method with IR can do a better job in further enhancing the error robustness of the 2HMCP coder. Figs. 14 and 15 illustrate the snap-shots of some corrupted frames coded with different coders for subjective performance evaluation.

The performances of the proposed 2HMCP with adaptive IR and RIR are compared in Fig. 16 for the *Foreman* and *Stefan* sequences. From Fig. 16, we can see that the proposed adaptive IR scheme reacts to channel error more swiftly than the RIR, leading to about 0.2–0.3 dB average PSNR improvements for the corrupted frames as shown in Table 5.

5. Conclusion

In this paper, we proposed efficient schemes for enhancing the error resiliency of 2HMCP, especially for burst error conditions. We have shown that combining the proposed 2HMCP with RIR can achieve faster error recovery speed. We have also proposed an adaptive IR scheme to further improve the error recovery capability of the proposed MHMCP coder. Extensive simulation results show that the proposed methods effectively and quickly mitigates the error propagation and the penalty on coding efficiency for clean channels due to the inserted error resilience features is rather minor. The proposed 2HMCP with a lower-rate IR can achieve better or comparable error resilience compared to Inter MCP with a much higher-rate IR, while retaining significantly better coding performance for clean channels.

References

[1] B. Girod, The efficiency of motion-compensating prediction for hybrid coding of video sequence, IEEE J. Select. Areas Commun. 5 (August 1987) 1140–1154.
[2] T. Wiegand, X. Zhang, B. Girod, Long-term memory motion-compensated prediction, IEEE Trans. Circuits Syst. Video Technol. 9 (1) (February 1999) 70–84.
[3] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H.264/AVC video coding standard, IEEE Trans. Circuits Syst. Video Technol. 13 (7) (July 2003) 560–576.
[4] G.J. Sullivan, Multi-hypothesis motion compensation for low bit-rate video coding, in: Proceedings of the IEEE International Conference Acoustics, Speech and Signal Processing, vol. 5, April 1993, pp. 437–440.
[5] B. Girod, Efficiency analysis of multihypothesis motion-compensated prediction for video coding, IEEE Trans. Image Process. 9 (2) (February 2002) 173–183.
[6] M. Flierl, T. Wiegand, B. Girod, A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction, in: Proceedings of the Data Compression Conference, April 1998, pp. 239–248.
[7] M. Flierl, T. Wiegand, B. Girod, Rate-constrained multi-hypothesis prediction for motion-compensated video com-

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002

pression, IEEE Trans. Circuits Syst. Video Technol. 12 (11) (November 2002) 957–969.

[8] M. Flierl, B. Girod, Generalized B pictures and the draft H.264/AVC video-compression standard, IEEE Trans. Circuits Syst. Video Technol. 13 (7) (July 2003) 587–597.

[9] Y. Wang, Q.F. Zhu, Error control and concealment for video communication: a review, Proc. IEEE 86 (5) (May 1998) 974–997.

[10] J. Xin, C.-W. Lin, M.-T. Sun, Digital video transcoding, Proc. IEEE 93 (1) (January 2005) 84–97.

[11] C.-S. Kim, R.-C. Kim, S.-U. Lee, Robust transmission of video sequence using double-vector motion compensation, IEEE Trans. Circuits Syst. Video Technol. 11 (9) (September 2001) 1011–1021.

[12] S. Lin, Y. Wang, 2002. Error resilience property of multi-hypothesis motion-compensated prediction, in: Proceedings of the IEEE International Conference Image Processing, vol. 3, June 2002, pp. 545–548.

[13] W.-Y. Kung, C.-S. Kim, C.-C. Jay Kuo, Analysis of multihypothesis motion compensated prediction (MHMCP) for robust visual communication, IEEE Trans. Circuits Syst. Video Technol. 16 (1) (January 2006) 146–153.

[14] K. Stuhlmüller, N. Färber, M. Link, B. Girod, Analysis of video transmission over lossy channels, IEEE J. Select. Areas Commun. 18 (6) (June 2006) 1012–1032.

[15] Y.J. Liang, J.G. Apostolopoulos, B. Girod, Analysis of packet loss for compressed video: does burst-length matter, in: Proceedings of the IEEE International Conference Acoustics, Speech and Signal Processing, vol. 5, April 2003, pp. 684–687.

[16] S. Gnavi, M. Grangetto, E. Magli, G. Olmo, Comparison of rate allocation strategies for H.264 video transmission over wireless lossy correlated networks, in: Proceedings of the IEEE International Conference Multimedia and Expo., vol. 2, July 2003, pp. 517–520.

[17] R. Zhang, S.L. Regunathan, K. Rose, Video coding with optimal inter/intra-mode switching for packet loss resilience, IEEE J. Select. Areas Commun. 18 (6) (June 2006) 966–976.

[18] H.-J. Chiou, Y.-R. Lee, C.-W. Lin, Content-aware error resilient transcoding using prioritized intra-refresh for video streaming, J. Visual Commun. Image Represent. 16 (4–5) (August-October 2005) 563–588.

[19] S. Aramvith, C.-W. Lin, S. Roy, M.-T. Sun, Wireless video transport using conditional retransmission and low-delay interleaving, IEEE Trans. Circuits Syst. Video Technol. 12 (6) (June 2006) 558–565.

[20] J.R. Yee, E.J. Weldon. Jr., Evaluation of the performance of error-correcting codes on a Gilbert channel, IEEE Trans. Commun. 43 (8) (August 1995) 2316–2323.

[21] E.N. Gilbert, Capacity of a burst-noise channel, Bell Syst. Tech. J. (September 1960) 1253.

Please cite this article as: Y.-C. Tsai, et al., H.264 error resilience coding based on multi-hypothesis motion-compensated prediction, Signal Process. Image Commun. (2007), doi:10.1016/j.image.2007.05.002