Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

An error resilient scheme for H.264 video coding based on distortion estimated mode decision and nearest neighbor error concealment

LEE Tien-hsu[1], WANG Jong-tzy[2], CHEN Jhih-bin[3], CHANG Pao-chi[3]
([1]*Department of Electrical Engineering, National Chi Nan University, Taiwan 545, Puli, Nantou County*)
([2]*Department of Electronic Engineering, Jin-Wen Institute of Technology, Taiwan 231, Shindian, Taipei County*)
([3]*Department of Communication Engineering, National Central University, Taiwan 320, Jhongli, Taoyuan County*)
E-mail: thlee@vaplab.ee.ncu.edu.tw; jtwang@jwit.edu.tw; jbchen@vaplab.ee.ncu.edu.tw; pcchang@ce.ncu.edu.tw

Received Dec. 2, 2005; revision accepted Feb. 19, 2006

Abstract: Although H.264 video coding standard provides several error resilience tools, the damage caused by error propagation may still be tremendous. This work is aimed at developing a robust and standard-compliant error resilient coding scheme for H.264 and uses techniques of mode decision, data hiding, and error concealment to reduce the damage from error propagation. This paper proposes a system with two error resilience techniques that can improve the robustness of H.264 in noisy channels. The first technique is Nearest Neighbor motion compensated Error Concealment (NNEC) that chooses the nearest neighbors in the reference frames for error concealment. The second technique is Distortion Estimated Mode Decision (DEMD) that selects an optimal mode based on stochastically distorted frames. Observed simulation results showed that the rate-distortion performances of the proposed algorithms are better than those of the compared algorithms.

Key words: H.264, Error resilience, Error concealment, Data hiding, Mode decision
doi:10.1631/jzus.2006.AS0034 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

With the relatively high error rate of wireless networks, error resilience is extremely important for video coding. H.264 utilizes complicated predictions in temporal and spatial domains to enhance the coding efficiency, such as the directional prediction in intra coding, variable block size, multiple reference picture, and quarter-pixel-accurate motion compensation in inter coding. More efficient variable length code (VLC), in-loop de-blocking filter and no-mismatch integer transform are included in the standard to improve the video quality (Wiegand *et al.*, 2003). However, such predictions may cause serious error propagation effects because some of these techniques are inherently sensitive to transmission errors. Due to the use of intra prediction, errors may spread from neighboring causal macroblocks (MBs) to the inner sub-blocks of the current MB. With the multiple-reference-frame prediction, variable-block-size motion estimation, and sub-pixel generation, the error may be broadcast from preceding frames directly or indirectly and also propagate from the neighboring MBs. In addition, because of the use of VLC, erroneous compressed data usually cannot be correctly decoded until the next resynchronization point. Furthermore, as a result of using de-blocking filter, errors may also diffuse from adjacent MBs (Stockhammer *et al.*, 2003).

To combat these problems, H.264 provides many error resilient tools such as parameter sets, data partition, redundant slices, flexible macroblock ordering (FMO), etc. FMO alters the way how pictures are partitioned into slices and MBs by employing the concept of slice groups. Every slice group consists of one or more slices and a slice is a sequence of MBs. When using FMO, a picture can be split into many

MB scanning patterns such as interleaved slices. Redundant Slice (RS) allows the encoder to place, in addition to the coded MBs of the slice itself, one or more redundant representations of the same MBs into the same bit stream. RS is useful in that the redundant representation can be coded using different coding parameters, such as QP or reference frame index. The parameter set contains information that is expected to rarely change and enable decoding of a large number of VCL NAL units (Wenger, 2003).

Unfortunately, the error propagation effect still cannot be entirely eliminated by using the above tools. More research works on error resilience are still in progress. The error resilient encoding approach selects intra coded MBs in random or certain update pattern. A typical work (Wiegand and Girod, 2001) considers the use of Lagrangian optimized mode decision when assigning intra MBs with significant improvements in the rate-distortion performance. A rate-distortion optimization approach for H.26L video coding in packet loss environment is analytically investigated by Stockhammer *et al.*(2002). By averaging over several decoders with different channel statistics, the expected MB distortions of decoders are calculated at the encoder and the optimized mode is selected accordingly. An error resilient coding scheme based on data embedding is proposed by Kang and Leou (2005), in which a relatively large volume of important data useful for error concealment are embedded beforehand into video frames at the encoder. The concealment result is satisfactory but 15 error concealment schemes for an MB must be pre-evaluated at the encoder. Even so, the error propagation effect still exists.

Our goal is to develop a robust and standard-compliant error resilient coding scheme for H.264 without a feedback channel. It utilizes techniques of mode decision, data hiding, and error concealment to reduce the damage from error propagation. This paper is organized as follows. The details of the two proposed algorithms, Nearest Neighbor motion compensated Error Concealment (NNEC) and Distortion Estimated Mode Decision (DEMD), are described in the next two sections, respectively. Simulation results and comparisons of the proposed algorithms are given in Section 4, and at last, this research work is concluded in Section 5.

NEAREST NEIGHBOR ERROR CONCEALMENT (NNEC)

The first technique is Nearest Neighbor Error Concealment (NNEC) that performs motion compensated error concealment with a better estimation of motion information. It utilizes a data hiding technique to embed the index code of important information about the most similar neighboring block into the bit stream, presumably the next slice. The hiding information includes the coding mode of current MB, and its corresponding nearest neighboring 8×8 blocks with the most similar motion vectors and reference frames.

The determination of the nearest neighbors for each MB is shown in Fig.1. To simplify the computation, only the modes with block size larger than or equal to 8×8 are considered. Depending on the mode of an MB, the number of motion vectors and the number of candidates for each MB are different. For instance, a 16×16 mode has a set of motion vectors chosen from 8 neighbors that can be represented by 3 bits, while a 16×8 mode has two sets of motion vectors chosen from 4 neighbors each of which can be represented by 2 bits. The detailed bit assignment of the nearest neighbor that utilizes the principle of Huffman coding is shown in Fig.2. No more than 7 bits are needed to represent the nearest neighbors for each MB in terms of the coding mode and motion information.

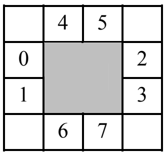


Fig.1 Nearest neighbor determination for each MB (shaded area)

	Coding mode	MV & Ref	Total bits
Skip	2 (00)	0	2
16×16	2 (01)	3 (xxx)	5
16×8	3 (100)	4 (xx xx)	7
8×16	3 (101)	4 (xx xx)	7
8×8	3 (110)	4 (x x x x)	7
Other	3 (111)	0	3
(x) represents the embedded bit, x is 0 or 1			

Fig.2 Huffman coding of NNEC information

The next problem is how to deliver the NNEC information to the receiving end. In this work, a data hiding method is used. We embed each code bit into one quantized residual DCT coefficient by adjusting the coefficient *checksum* in a 4×4 DCT block (Chang *et al*., 2001) to satisfy the rule in Eq.(1).

$$\begin{aligned} &Embeded\ bit = 1 \ \Leftrightarrow\ \sum_{i=0}^{15} LEVEL_i = 2M+1, \\ &Embeded\ bit = 0 \ \Leftrightarrow\ \sum_{i=0}^{15} LEVEL_i = 2M. \end{aligned} \tag{1}$$

$LEVEL_i$ is the quantized value of the ith coefficient and M is an arbitrary integer. By forcing the *checksum* to be 0 or 1, one bit information can be embedded and extracted successfully.

Once the decoder does not receive the correct MB data including its motion vectors, it can use the nearest neighboring blocks with the index codes extracted from the bit steam to perform much more accurate error concealment than the conventional error concealment, like the one used in JM 9.4. By comparing the maximum amount of required hiding bits per MB, NNEC is only 7 bits while that in (Kang and Leou, 2005) is 48 bits. Obviously, the NNEC results in less degradation in the rate-distortion performance from data embedding, and requires relatively low implementation complexity as well.

DISTORTION ESTIMATED MODE DECISION (DEMD)

The second technique is Distortion Estimated Mode Decision (DEMD). The overall pixel distortion of the reconstruction frame at the decoding side includes the quantization error, concealment error, and propagation error. In DEMD method, the encoder estimates the decoder distortion by calculating and storing expected pixel mean and variance values frame-by-frame for a given packet error rate. It is assumed that each slice will not be packetized into more than one packet, i.e., each data slice can be decoded independently except by referring to previous frames. The notations for the signals are listed as follows:

O_m^i : Original ith pixel in frame m;

D_m^i : Reconstructed ith pixel in frame m at decoder;

C_m^i : Concealed ith pixel in frame m;

E_m^i : Reconstructed ith pixel in frame m at encoder;

p : Packet loss probability;

d_m^i : Expected distortion of ith pixel in frame m.

The expected distortion can be expressed by Eq.(2):

$$d_m^i = E\{(O_m^i - D_m^i)^2\} = (O_m^i)^2 - 2O_m^i E\{D_m^i\} + E\{(D_m^i)^2\}. \tag{2}$$

Furthermore, the NNEC technique described in the previous section can also be incorporated into this mode decision technique. Depending on the used error concealment, the concealed image can be expressed as $\hat{C}_m^i = D_{m-n}^j$, i.e., it is replaced by the jth pixel in the previous n frames, if NNEC can be successfully applied; or $\tilde{C}_m^i = D_{m-o}^k$, i.e., the kth pixel in the previous o frames, if the NNEC embedded information is unfortunately lost, then a conventional error concealment method is employed. Moreover, $E\{D_m^i\}$ and $E\{(D_m^i)^2\}$ are calculated differently in intra and inter coding as follows.

Intra coding

A packet will not be lost with probability $1-p$, then $D_m^i = E_m^i$. Otherwise, a packet will be lost with probability p, then the error concealment will be performed and $D_m^i = C_m^i$. Moreover, NNEC is performed with the probability $1-p$, i.e., no packet loss in the data hiding slice, while conventional error concealment is performed with the probability p if the data hiding slice is lost.

$$\begin{aligned} E\{D_m^i\} &= (1-p)E_m^i + pE\{C_m^i\} \\ &= (1-p)E_m^i + p[(1-p)E\{\hat{C}_m^i\} + pE\{\tilde{C}_m^i\}] \\ &= (1-p)E_m^i + p(1-p)E\{D_{m-n}^j\} \\ &\quad + p^2 E\{D_{m-o}^k\}, \end{aligned} \tag{3}$$

$$\begin{aligned} E\{(D_m^i)^2\} &= (1-p)(E_m^i)^2 + pE\{(C_m^i)^2\} \\ &= (1-p)(E_m^i)^2 + p[(1-p)E\{(\hat{C}_m^i)^2\} \\ &\quad + pE\{(\tilde{C}_m^i)^2\}] \\ &= (1-p)(E_m^i)^2 + p(1-p)E\{(D_{m-n}^j)^2\} \\ &\quad + p^2 E\{(D_{m-o}^k)^2\}. \end{aligned} \tag{4}$$

Inter coding

The motion compensation is performed in the inter-coding. Assume the *j*th pixel in the *n*th previous frame is used to predict the *i*th pixel in the current frame, the residue, i.e., the prediction error, is $e_m^i = E_m^i - E_{m-n}^j$. If the decoding end correctly receives the residue, motion vector, and the reference frame number, then the reconstructed pixel can be obtained by $D_m^i = D_{m-n}^j + e_m^i$. If the packet will be lost with probability p, then the concealed pixel C_m^i will be used to replace this pixel. Therefore, $E\{D_m^i\}$ and $E\{(D_m^i)^2\}$ can be derived as follows:

$$\begin{aligned} E\{D_m^i\} &= (1-p)E\{D_{m-n}^j + e_m^i\} + pE\{C_m^i\} \\ &= (1-p)E\{D_{m-n}^j + e_m^i\} + p[(1-p)E\{\hat{C}_m^i\} \\ &\quad + pE\{\tilde{C}_m^i\}] \qquad (5) \\ &= (1-p)E\{D_{m-n}^j + e_m^i\} + p(1-p)E\{D_{m-n}^j\} \\ &\quad + p^2 E\{D_{m-o}^k\}, \end{aligned}$$

$$\begin{aligned} E\{(D_m^i)^2\} &= (1-p)E\{(D_{m-n}^j + e_m^i)^2\} + pE\{(C_m^i)^2\} \\ &= (1-p)[E\{(D_{m-n}^j)^2\} + 2e_m^i E\{D_{m-n}^j\} + (e_m^i)^2] \\ &\quad + p[(1-p)E\{(\hat{C}_m^i)^2\} + pE\{(\tilde{C}_m^i)^2\}] \\ &= (1-p)E\{(D_{m-n}^j)^2 + 2(1-p)e_m^i E\{D_{m-n}^j\} \\ &\quad + (1-p)(e_m^i)^2 + p(1-p)E\{(D_{m-n}^j)^2\} \\ &\quad + p^2 E\{(D_{m-o}^k)^2\}. \qquad (6) \end{aligned}$$

Mode decision

With the above analysis, the expected distortion d_m^i can be calculated for a given p and then be used as the distortion in the H.264 mode decision procedure. Since d_m^i includes the quantization error, concealment error, and error propagation distortion, the mode decision procedure will choose an optimum mode that minimizes the distortion for a given p.

The encoder and decoder structures of the proposed error resilient video coding system are shown in Figs.3 and 4, respectively. The procedure at the encoding end is as follows:

(1) Calculate the expected distortion based on $E\{D_m^i\}$ and $E\{(D_m^i)^2\}$ in mode decision;

(2) After mode decision, store the best mode, motion vector, and the reference frame number information for skip mode, mode 16×16, mode 16×8, mode 8×16, and mode 8×8;

(3) After a frame is encoded, compare the recorded data with neighboring blocks. Determine the NNEC data for each block;

(4) Encode the NNEC data on a slice and embed them into Huffman coded bit stream of another slice;

(5) Assume each block is lost, perform corresponding error concealment, calculate $\hat{C}_m^i$ and $\tilde{C}_m^i$;

(6) Update $E\{D_m^i\}$ and $E\{(D_m^i)^2\}$ for each frame.

At the decoding end, the DEMD with NNEC procedure is listed as follows:

(1) Decode a slice based on H.264 if the slice is not lost, and extract the NNEC data. If lost, skip this slice, and record the slice position;

(2) Perform error concealment starting from frame boundary to the center after a frame is decoded;

(3) For lost blocks, if the NNEC data can be extracted successfully, perform NNEC. If not, use the conventional error concealment.

The estimates are integrated into a rate-distortion model for optimal switching between various intra and inter coding modes for each MB. The resulting performance is better than the conventional mode decision (without estimating the decoder distortion) provided that the packet error rate can be estimated or fed back by the channel. In addition, unlike the statistical approach introduced in (Stockhammer *et al.*, 2002), the DEMD requires relatively low implementation complexity and is suitable for practical use.

SIMULATION RESULTS

We assess the proposed error resilient scheme in environments with random packet loss rates of 5% and 20%. The illustrated test video sequence is QCIF Foreman and the FMO is enabled with two slice groups mapping as a checker-board type. One slice composed of 11 MBs is packetized into one packet. We adopt the following abbreviations to represent the results with various techniques applied.

(1) JM 9.4 Error Free: error free.

(2) NNEC Error Free: error free with hidden NNEC information codes.

(3) JM 9.4 EC: error concealment (EC) used in JM 9.4.

(4) NNEC: nearest neighbor motion compen-

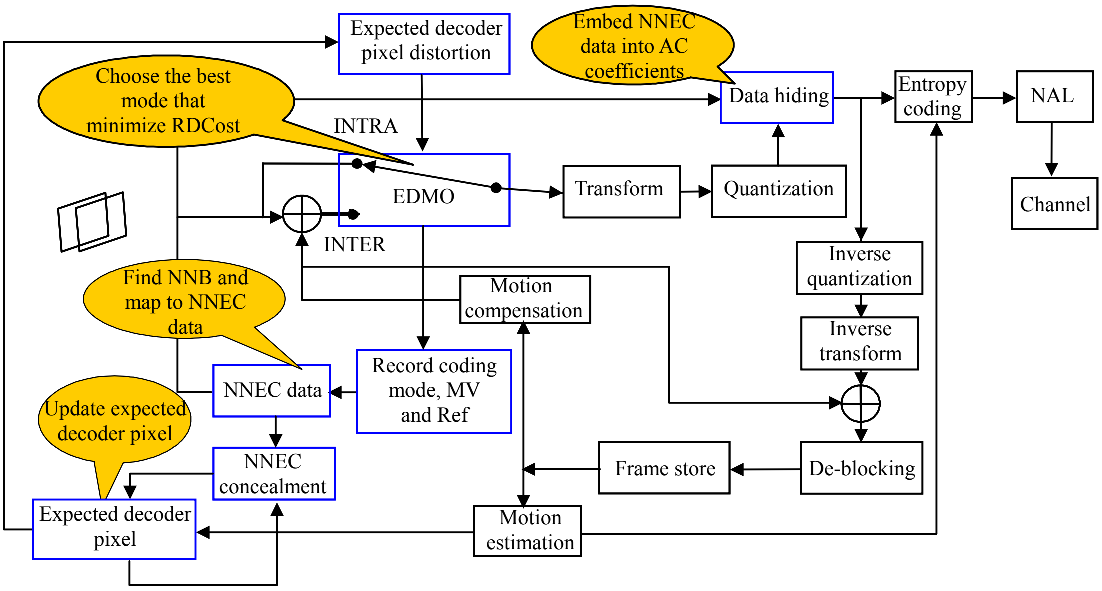


Fig.3 Encoder structure of the proposed error resilience system

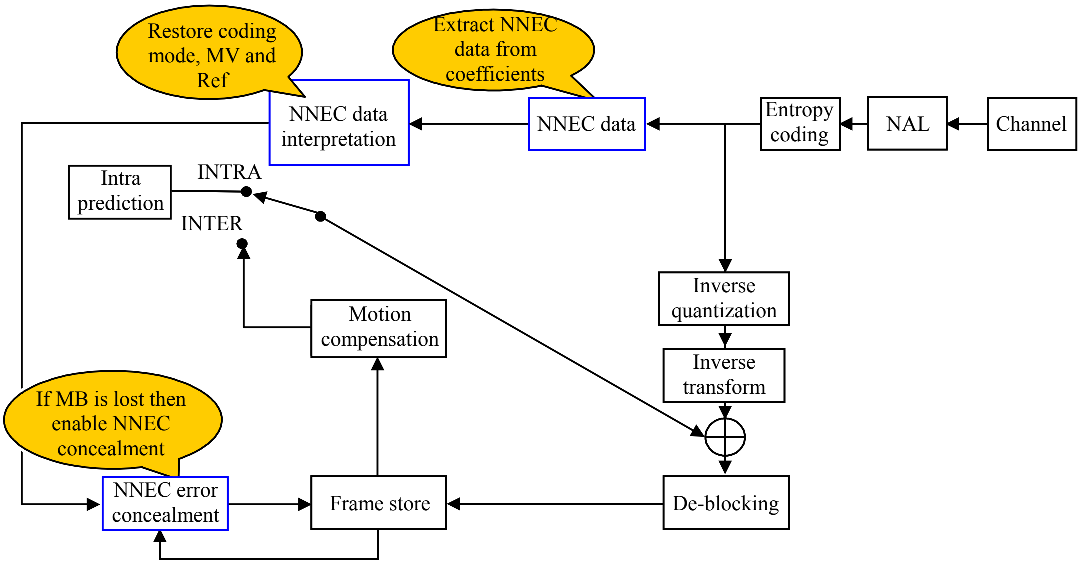


Fig.4 Decoder structure of the proposed error resilience system

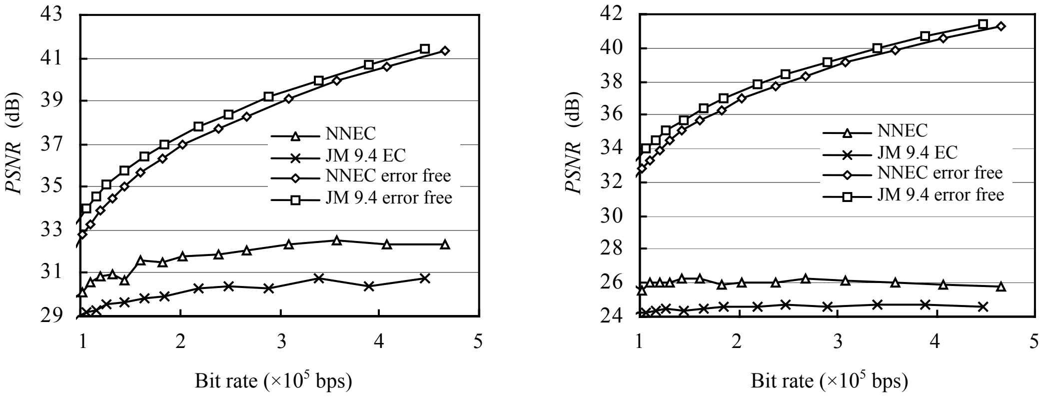
sated EC.

(5) DEMD: distortion estimated mode decision.

(6) RIR: random intra refresh.

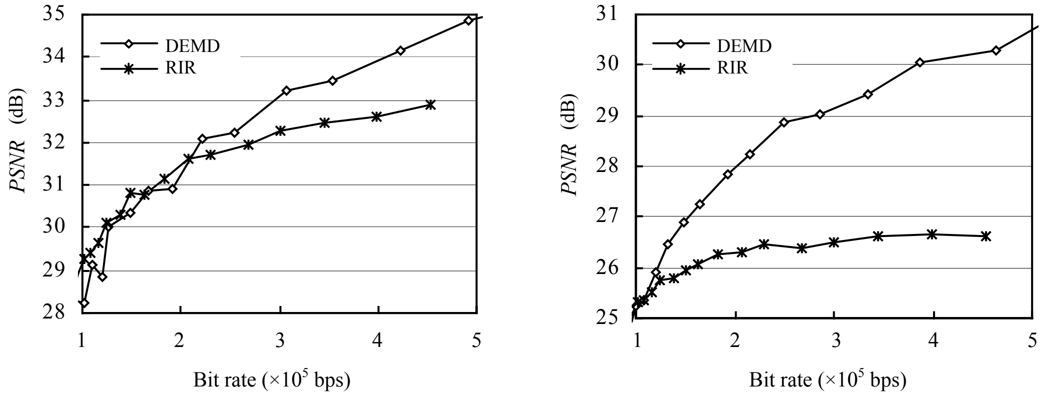
By investigating the simulation results, the rate-distortion (R-D) performances of the proposed techniques are better than those of the compared algorithms in all experiments. For example, Fig.5 shows that the proposed NNEC only degrades the PSNR by about 0.5 dB in the error free case. However, NNEC can additionally enhance the error concealment approach suggested in JM 9.4 by more than 1.5 dB in noisy environment. Obviously, the enhancement cannot be further increased especially in a high bit rate situation due to the inherent characteristics that error propagation effects cannot be avoided by simply applying error concealment. In contrast to Fig.6, the proposed DEMD technique can significantly further improve the video quality by more than

4 dB compared with the error concealment or random intra refresh approach especially at the high bit rate of 500 kbps and packet loss rate of 20%. Of course, by combining DEMD with NNEC, the proposed error resilient scheme achieves the best rate-distortion performance as shown in Fig.7.



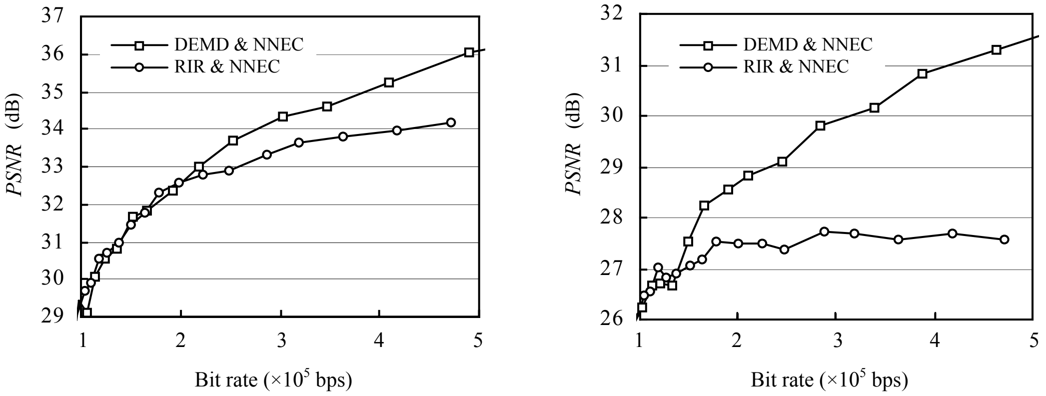
(a) (b)

Fig.5 R-D comparison of NNEC and JM 9.4 error concealment at packet loss rates of (a) 5% and (b) 20%



(a) (b)

Fig.6 R-D comparison of DEMD and RIR at packet loss rates of (a) 5% and (b) 20%



(a) (b)

Fig.7 R-D comparison of DEMD & NNEC and RIR & NNEC at packet loss rates of (a) 5% and (b) 20%

CONCLUSION

The proposed error-resilient NNEC and DEMD techniques are not mutually exclusive, i.e., they can be applied together in a system without degrading each individual performance. The combined system performs better than the compared methods without requiring a feedback channel.

ACKNOWLEDGEMENT

This work was supported in part by grants from the National Science Council of Taiwan under contracts NSC 94-2213-E-008-021 and NSC 94-2219-E-260-006.

References

Chang, P.C., Wang ,T.H., Lee, T.H., 2001. An Efficient Data Embedding Algorithm for H.263 Compatible Video Coding. Proc. Data Compression Conference, p.489.

Kang, L.W., Leou, J.J., 2005. An error resilient coding scheme for H.264/AVC video transmission based on data embedding. *J. Visual Communication and Image Representation*, **16**(1):93-114. [doi:doi:10.1016/j.jvcir.2004.04.003]

Stockhammer, T., Kontopodis, D., Wiegand, T., 2002. Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment. Proc. Int. Packet Video Workshop. Pittsburgh, USA.

Stockhammer, T., Hannuksela, M., Wiegand, T., 2003. H.264/AVC in wireless environments. *IEEE Trans. Circuits Syst. Video Technol.*, **13**(7):657-673. [doi:10.1109/TCSVT.2003.815167]

Wenger, S., 2003. H.264/AVC over IP. *IEEE Trans. Circuit Syst. Video Technol.*, **13**(7):645-656. [doi:10.1109/TCSVT.2003.814966]

Wiegand, T., Girod, B., 2001. Lagrange Multiplier Selection in Hybrid Video Coder Control. Proc. IEEE International Conference on Image Processing, **3**:542-545.

Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A., 2003. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, **13**(7):560-576. [doi:10.1109/TCSVT.2003.815165]