

Robust MMSE Video Decoding: Theory and Practical Implementations

Chang-Su Kim, *Member, IEEE*, JongWon Kim, *Senior Member, IEEE*, Ioannis Katsavounidis, *Member, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

Abstract—A novel video decoding algorithm based on the minimum mean square error (MMSE) criterion is investigated in this research. To alleviate the effect of transmission errors, we first develop an error propagation model to estimate and track the mean square error (MSE) of each pixel in the decoder. Then, the proposed video decoding algorithm adjusts the reconstruction of each pixel adaptively according to fluctuating channel conditions. More specifically, the decoder reconstructs a pixel in the k th frame F_k by using a weighted sum of two pixels in frames F_{k-1} and F_{k-2} , respectively, where their weights are adaptively selected to minimize the MSE of the reconstructed pixel by using the error propagation model. Extensive simulation results performed on standard H.263 bit streams demonstrate that the MMSE-based concealment algorithm yields a better performance than the conventional method, even if the encoder transmits a single motion vector per block. Moreover, the proposed MMSE decoding algorithm significantly enhances the error resilient capability of the double-vector motion compensation (DMC) algorithm, where two motion vectors are sent per block.

Index Terms—Double-vector motion compensation, error concealment, H.263, minimum mean square error (MMSE) decoding, robust video transmission.

I. INTRODUCTION

THE INCREASING demand of rich media has resulted in a significant amount of research effort in effective transmission and storage of digital video. Compression technologies offer the possibility of transmitting or storing a vast amount of video data in a compact way. They are therefore essential in various applications, such as digital TV/HDTV broadcast, video on demand, and videophone services. A lot of progress has been made in efficient video compression in the last decade, and several international standards have emerged to support a wide range of video applications, including H.261 [2], H.263 [3], and MPEG-4 [4]. These standards achieve a high compression ratio by exploiting spatio-temporal correlations in image sequences through motion compensated prediction, the discrete

cosine transform (DCT), and variable length coding (VLC). However, as a video file gets compressed more, the encoded bit stream becomes more vulnerable to bit errors over wireless channels and packet loss over IP networks. Many techniques have been developed to enhance the error resilience of video bit streams [5]–[7], and a number of error resilient tools have been incorporated in the recent H.263 and MPEG-4 standards [8], [9].

Video signals can be protected against transmission errors at the channel or the source level. At the channel level, forward error correction (FEC) and automatic repeat request (ARQ) are used to detect and correct transmission errors by transmitting redundant data in proactive and reactive manners, respectively. However, FEC faces the challenge of controlling the amount of redundant data adaptively to the channel fluctuation and the source variation. On the other hand, ARQ often fails to satisfy the latency requirement. Thus, source level approaches, where the video encoder sacrifices coding efficiency as a tradeoff for enhanced error resilience, are also popular. They attempt to suppress error propagation in compressed video signals. Transmission errors propagate temporally due to motion compensated prediction, and spatially due to the characteristics of VLC. Intra-coding and periodic insertion of synchronization codewords are effective tools in suppressing temporal and spatial error propagations, respectively. Several algorithms have been proposed to optimize the intra/intermode switching and the placement of synchronization codewords [10]–[14]. Also, the double-vector motion compensation (DMC) method in [1], where each block is predicted from a weighted superposition of two blocks using two motion vectors, can effectively alleviate the effect of temporal error propagation at the cost of additional motion vectors.

At the decoder, the error concealment (EC) technique attempts to hide visible distortion in erroneous blocks by utilizing the information in adjacent blocks. Spatial concealment methods interpolate an erroneous block from spatially adjacent blocks based on the smoothness property of typical video signals [15], [16]. Temporal concealment methods replace an erroneous block temporally after estimating (e.g., using the information in the spatially adjacent blocks) the motion vector [17]–[19]. When all adjacent blocks are also erroneous, spatial and temporal EC methods provide a poor performance. In [20], [21], some feedback mechanism was adopted in association with the EC technique. That is, if errors are detected by the decoder, they are concealed and their locations are transmitted to the encoder via a feedback channel. Then, the encoder tracks the error propagation pattern, and tries to minimize the error impact by intrarefreshing the severely corrupted regions.

Manuscript received May 21, 2002; revised June 5, 2003. This work was supported in part by InterVideo, Inc. This paper was recommended by Associate Editor Q. Tian.

C.-S. Kim is with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: cskim@iee.org).

J. Kim is with the Department of Information and Communications, Kwangju Institute of Science and Technology, Kwangju 500-712, Korea (e-mail: jongwon@kjist.ac.kr).

I. Katsavounidis is with InterVideo Inc., Fremont, CA 94539 USA (e-mail: ioannis@intervideo.com).

C.-C. J. Kuo is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: cckuo@sipi.usc.edu).

Digital Object Identifier 10.1109/TCSVT.2004.839994

However, the feedback mechanism, including ARQ and other means, may introduce extra delay and complicate multicast video transmission.

In this research, we propose a novel video decoding algorithm that can effectively protect the quality of reconstructed video signals against transmission errors based on the minimum mean square error (MMSE) criterion. Whereas conventional EC methods attempt to conceal only erroneous blocks, the proposed decoding algorithm adaptively adjusts the reconstruction methods for error-free blocks and the concealment methods for erroneous blocks according to channel conditions. Here, an error-free block means a block whose compressed data are not lost during transmission, while an erroneous block means a block whose data are lost. It is worthwhile to point out that an error-free block may still be corrupted as a result of error propagation.

With the proposed decoding algorithm, the decoder motion-compensates an error-free block B in the k th frame F_k by using a weighted superposition of two blocks from two previous frames, i.e., B_1 in frame F_{k-1} and B_2 in frame F_{k-2} . Also, the decoder replaces an erroneous block with a weighted superposition of two blocks in a similar way. These weights are adaptively determined to minimize the mean square errors (MSEs) of pixels in B based on the MSEs of pixels in blocks B_1 and B_2 . To this end, we develop an error propagation model to track the MSE of each pixel at the decoder. It is shown that the proposed decoding algorithm provides a better performance than the conventional EC methods, even if the encoder transmits only one motion vector per block as done in the H.263 standard. Furthermore, when combined with the DMC algorithm [1], the proposed MMSE DMC decoding algorithm significantly enhances error resilience.

The rest of the paper is organized as follows. Section II develops the error propagation model. Sections III and IV present the MMSE decoders for the H.263 standard and the DMC algorithm, respectively. Section V describes several issues related to implementational details. Section VI compares the performances of the proposed MMSE decoders with those of conventional decoders. Finally, Section VII concludes this paper.

II. ERROR PROPAGATION MODELING

A. Background and Motivation

In traditional video coders, the k th frame F_k is motion-compensated from the previous frame F_{k-1} in a blockwise fashion, and the corresponding residual signals are encoded by using DCT. Due to the use of motion-compensated prediction, transmission errors in a frame propagate to subsequent frames. It was shown in [13] that transmission errors tend to attenuate as they propagate as a result of the lowpass filtering operation in the prediction loop. The attenuation speed is however relatively slow, and perceived video quality is severely degraded in many cases by error propagation.

To track error propagation accurately, we need the information about error-free reconstructed video, locations of erroneous blocks, concealment methods for those blocks, and the motion vector field [22], [23]. The encoder has the information of error-free reconstruction and the motion vector field. It also knows

concealment methods if the decoder conceals erroneous blocks in a deterministic manner. Error locations are determined by channel conditions and only known to the decoder. Here, we assume that transmission errors are detected at either the transport decoder or the video decoder using the VLC parser and the syntax analyzer [5]. Hence, the exact error tracking mechanism requires an interaction between the encoder and the decoder via a feedback channel [20], [21].

To enhance error resilience of compressed video signals, error propagation has been modeled for the encoder in previous work, e.g., [10]–[12]. Since the encoder does not know error locations, these models consider various combinations of possible error events, and obtain the expected distortion incurred by all the combinations. Then, severely distorted blocks are encoded in the intramode to suppress the expected error propagation. In contrast, we develop an error propagation model for the decoder to alleviate the effect of transmission errors in this work. We define a pixel error as the difference between the error-free reconstruction (or the encoder's reconstruction) and the decoder's reconstruction. Note that the decoder does not know the error-free reconstruction, but it is informed of error locations by the error detector. Therefore, our model treats each pixel error as a random variable, and estimates the variances of the pixel errors that reflect the effect of a specific error event informed by the error detector.

B. Error Variance Propagation and Leaky Factors

Let us first assume that the decoder knows the error variance σ_p^2 for each pixel p in the k th frame F_k . To record this information, the proposed algorithm employs an extra frame buffer, called the *error variance map*, which is of the same size as the video frame. Each value in the error variance map represents the error variance of the corresponding pixel in the video frame. Given the error variance map for frame F_k , the decoder can obtain the error variance map for its subsequent frame F_{k+1} by tracking error propagation.

To be more specific, consider the half-pixel motion compensation scheme adopted by H.263 or MPEG-4. Fig. 1 illustrates the interpolation scheme for the half-pixel motion compensation, where black circles represent ordinary pixels. A pixel is compensated from a virtual pixel depicted by “+,” when either the x or the y component of the motion vector is of half-pixel accuracy. The virtual pixel value is given by the average of the two nearest ordinary pixels. Similarly, when both the x and the y components of the motion vector are of half-pixel accuracy, a pixel is compensated from a virtual pixel “ \times ,” whose value is the average of the four nearest ordinary pixels. Suppose that a pixel q in F_{k+1} is predicted from F_k by using motion vector $\mathbf{v} = (v_x, v_y)$. Then, error propagation can be modeled according to the accuracy of the motion vector as discussed below.

For the first case, when both v_x and v_y are of integer pixel accuracy, \mathbf{v} specifies an ordinary pixel p_\bullet in F_k . It is obvious that the error in p_\bullet propagates to q without attenuation. Thus, the error variance σ_{prop}^2 of q as a result of this propagation is given by

$$\sigma_{\text{prop}}^2 = \sigma_{p_\bullet}^2. \quad (1)$$

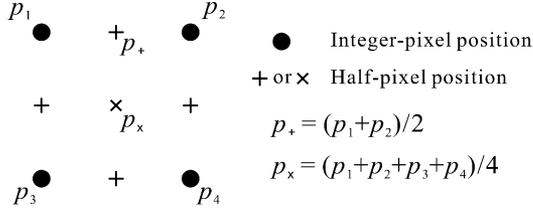


Fig. 1. Interpolation schemes for half-pixel motion compensation.

For the second case, when v_x is of half-pixel accuracy while v_y is of integer accuracy (or vice versa), \mathbf{v} specifies a virtual pixel. Suppose that \mathbf{v} specifies $p_+ = (p_1 + p_2)/2$ as shown in Fig. 1. Let e_1 and e_2 denote the errors in p_1 and p_2 , respectively. Then, q is corrupted by $(e_1 + e_2)/2$. Assuming e_1 and e_2 have zero means, we have

$$\begin{aligned} \sigma_{\text{prop}}^2 &= \sigma_{p_+}^2 = E \left\{ \left(\frac{e_1 + e_2}{2} \right)^2 \right\} \\ &= l_+ \cdot \frac{\sigma_{p_1}^2 + \sigma_{p_2}^2}{2} \end{aligned} \quad (2)$$

where

$$l_+ = \frac{1}{2} \left(1 + 2 \frac{E\{e_1 e_2\}}{E\{e_1^2\} + E\{e_2^2\}} \right)$$

is called the leaky factor. It can be easily shown that the leaky factor l_+ ranges from 0 to 1, and is equal to 1 only if $e_1 = e_2$. When error signals consist of the dc component only, they do not attenuate even if the half-pixel motion compensation is employed. However, error signals contain also ac components in most cases and they become weaker as they propagate.

For the third case, when both v_x or v_y are of half-pixel accuracy, \mathbf{v} specifies a virtual pixel $p_x = (p_1 + p_2 + p_3 + p_4)/4$ as shown in Fig. 1. Then, q is corrupted by $(e_1 + e_2 + e_3 + e_4)/4$, where e_i denotes the error in p_i , $1 \leq i \leq 4$. Similarly, we can derive

$$\begin{aligned} \sigma_{\text{prop}}^2 &= \sigma_{p_x}^2 \\ &= l_x \cdot \frac{\sigma_{p_1}^2 + \sigma_{p_2}^2 + \sigma_{p_3}^2 + \sigma_{p_4}^2}{4} \end{aligned} \quad (3)$$

where

$$l_x = \frac{1}{4} \left(1 + 2 \frac{\sum_{i=1}^3 \sum_{j=i+1}^4 E\{e_i e_j\}}{\sum_{i=1}^4 E\{e_i^2\}} \right).$$

The leaky factor l_x also varies within $[0, 1]$, and is equal to 1 only if $e_1 = e_2 = e_3 = e_4$.

As given by (1)–(3), we see that the values in the error variance map are interpolated in the same way as the corresponding pixels in the video frame, except that they are multiplied by leaky factors l_+ or l_x when half-pixel accuracy motion vectors are employed. The two leaky factors l_+ and l_x can be obtained from training sequences using various error patterns. Typical values of l_+ and l_x are 0.8 and 0.65, respectively.

C. Error Variances Due to Error Concealment and Propagation

By using (1)–(3), the decoder can estimate the error variance of q provided that the data for q in F_{k+1} is not lost during the

transmission. However, when the data for q is lost, its error variance is also affected by the concealment method. For example, suppose that the decoder conceals the loss of q by replacing it with a pixel p in the previous frame. Then, the error in q can be written as

$$e_q = \hat{q} - \tilde{p} = (\hat{q} - \hat{p}) + (\hat{p} - \tilde{p}) \quad (4)$$

where \hat{a} and \tilde{a} denote the error-free and the decoder's reconstructions of pixel a , respectively. Therefore, the first term $(\hat{q} - \hat{p})$ is the concealment error when the referenced pixel p is not corrupted while the second term $(\hat{p} - \tilde{p})$ is the propagation error from the referenced pixel. To simplify the error model, we assume that these two terms are uncorrelated random variables with zero mean. Then, the error variance of q is given by

$$\begin{aligned} \sigma_q^2 &= E\{(\hat{q} - \tilde{p})^2\} \\ &\simeq E\{(\hat{q} - \hat{p})^2\} + E\{(\hat{p} - \tilde{p})^2\} \\ &= \sigma_{\text{conc}}^2 + \sigma_{\text{prop}}^2 \end{aligned} \quad (5)$$

where $\sigma_{\text{conc}}^2 = E\{(\hat{q} - \hat{p})^2\}$ denotes the concealment error variance, and $\sigma_{\text{prop}}^2 = E\{(\hat{p} - \tilde{p})^2\}$ denotes the propagation error variance. Note that the propagation error variance σ_{prop}^2 can be obtained by (1), (2), or (3) according to the accuracy of the motion vector. The concealment error variance σ_{conc}^2 depends on the concealment method, and will be discussed in more detail in later sections.

In this way, the decoder can estimate and track the error variance (or the mean square error) of each pixel. It will be shown in the next two sections that the decoder can effectively alleviate the effect of transmission errors by using the estimated error variance.

III. MMSE DECODING ALGORITHM FOR H.263

In this section, we propose an MMSE decoding algorithm for the H.263 coder. In H.263 [3], each frame is divided into macroblocks (MBs), and each MB is encoded in either the intra or the intermode. For the intramode, an MB is encoded without reference to the previous frame. On the other hand, an inter-MB is predicted from the previous frame with a half-pixel accuracy motion vector, and the residual signals are encoded using DCT.

A. Decoding for I-Frames

In an intraframe (i.e., I-frame), all the MBs are encoded in the intramode. Since the intramode generally requires a larger amount of bits than the intermode, we insert I-frames only after scene cuts. After a scene cut, it is advantageous to conceal erroneous MBs relying only on the spatial correlation. Therefore, in our approach, when an MB in an I-frame is erroneous, it is filled with the pixel values in the lowest row of the upper MB. In other words, the lowest row of the upper MB is vertically expanded to conceal the erroneous MB. If the uppermost MB is erroneous, it is simply filled with a gray pixel value 128. After the concealment, the corresponding pixel values in the error variance map are set to σ_I^2 , which is the expected distortion (or error variance) obtained from training sequences. The error variances for error-free intra-MBs are set to 0, since intrablocks are not affected by the propagation error.

B. Weighted Concealment for Erroneous Inter-coded MB

An erroneous MB in an interframe (i.e., P-frame) is concealed by exploiting the temporal correlation, regardless whether it is encoded in the intra or intermode (the decoder cannot know the encoding mode in many cases). The following two modes are used for the temporal concealment.

- *Mode 1:* The motion vector of the erroneous MB is set to that of the upper MB. If the uppermost MB is erroneous, the motion vector is set to zero vector. Then, the erroneous MB is copied from the previous frame using the estimated motion vector $\tilde{\mathbf{v}}$.
- *Mode 2:* The erroneous MB is copied from the frame before the previous one by using $2\tilde{\mathbf{v}}$, where $\tilde{\mathbf{v}}$ is the estimated vector in mode 1.

Thus, an erroneous pixel q_k in the k th frame is concealed by a pixel c_{k-1} in the $(k-1)$ th frame in mode 1, and by c_{k-2} in the $(k-2)$ th frame in mode 2. It seems that mode 2 is unnecessary, since mode 1 yields a smaller concealment error than mode 2 in general. However, the error variance of a pixel depends on the propagation error as well as the concealment error. Therefore, if c_{k-1} is more severely corrupted than c_{k-2} , mode 2 can be advantageous.

Let $\sigma_{q_k,i}^2$ denote the error variance of q_k , when mode i is used for the concealment ($i = 1, 2$). As given in (5), $\sigma_{q_k,i}^2$ can be modeled as the sum of the concealment error variance and the propagation error variance

$$\begin{aligned}\sigma_{q_k,i}^2 &= E\{(\hat{q}_k - \tilde{c}_{k-i})^2\} \\ &\simeq E\{(\hat{q}_k - \hat{c}_{k-i})^2\} + E\{(\hat{c}_{k-i} - \tilde{c}_{k-i})^2\}, \\ &= \sigma_{\Delta_i}^2 + \sigma_{c_{k-i}}^2\end{aligned}\quad (6)$$

where $\sigma_{\Delta_i}^2 = E\{(\hat{q}_k - \hat{c}_{k-i})^2\}$ is the error variance due to the concealment mode i , and $\sigma_{c_{k-i}}^2$ is the error variance propagated from \tilde{c}_{k-i} . As a simple decoding rule, we can choose mode 1 if $\sigma_{q_k,1}^2 \leq \sigma_{q_k,2}^2$, and mode 2 otherwise.

However, we can further reduce the error variance of q_k by replacing q_k with the weighted sum of \tilde{c}_{k-1} and \tilde{c}_{k-2} , given by

$$\tilde{q}_k = \alpha \tilde{c}_{k-1} + (1 - \alpha) \tilde{c}_{k-2} \quad (7)$$

where α is a weighting coefficient. The coefficient α is selected to minimize the error variance or the mean square error $E\{(\hat{q}_k - \tilde{q}_k)^2\}$. It can be easily shown that such a value of α is given by

$$\alpha = \frac{\sigma_{q_k,2}^2}{\sigma_{q_k,1}^2 + \sigma_{q_k,2}^2} \quad (8)$$

under the assumption that $(\hat{q}_k - \tilde{c}_{k-1})$ and $(\hat{q}_k - \tilde{c}_{k-2})$ are uncorrelated random variables with zero mean. Also, such a value of α yields the error variance

$$\sigma_{q_k}^2 = E\{(\hat{q}_k - \tilde{q}_k)^2\} = \frac{\sigma_{q_k,1}^2 \cdot \sigma_{q_k,2}^2}{\sigma_{q_k,1}^2 + \sigma_{q_k,2}^2}. \quad (9)$$

Note that $\sigma_{q_k}^2$ is always smaller than $\sigma_{q_k,1}^2$ or $\sigma_{q_k,2}^2$. In summary, with the proposed algorithm, the decoder conceals each pixel by (7), and then updates the corresponding error variance in the error variance map by (9).

C. Weighted Reconstruction of Error-Free Inter-coded MB

An error-free inter-MB is reconstructed by minimizing the effect of error propagation. Suppose that a pixel q_k in the k th frame is predicted from p_{k-1} in the $(k-1)$ th frame with a motion vector \mathbf{v} , and its reconstruction at the encoder is given by

$$\hat{q}_k = \hat{p}_{k-1} + \hat{r}_k$$

where \hat{r}_k is the prediction residual, encoded using DCT. We consider also two modes for the reconstruction of q_k .

- *Mode 1:* We reconstruct q_k via

$$\tilde{q}_k = \tilde{p}_{k-1} + \hat{r}_k$$

as done in the conventional decoder. Furthermore, the error variance of q_k is given by

$$E\{(\hat{p}_{k-1} - \tilde{p}_{k-1})^2\} = \sigma_{p_{k-1}}^2.$$

- *Mode 2:* We employ the pixel \tilde{p}_{k-2} in the $(k-2)$ th frame, specified by $2\mathbf{v}$, as an alternative prediction value. With Mode 2, the error variance of q_k becomes

$$\begin{aligned}E\{(\hat{p}_{k-1} - \tilde{p}_{k-2})^2\} \\ \simeq E\{(\hat{p}_{k-1} - \hat{p}_{k-2})^2\} + E\{(\hat{p}_{k-2} - \tilde{p}_{k-2})^2\} \\ = \sigma_{\Theta}^2 + \sigma_{p_{k-2}}^2\end{aligned}\quad (10)$$

where $\sigma_{\Theta}^2 = E\{(\hat{p}_{k-1} - \hat{p}_{k-2})^2\}$ is the error variance when \hat{p}_{k-1} is replaced by \hat{p}_{k-2} .

As a result, the decoder has two candidates \tilde{p}_{k-1} and \tilde{p}_{k-2} for the prediction value, which render the error variance of q_k equal to $\sigma_{p_{k-1}}^2$ and $\sigma_{\Theta}^2 + \sigma_{p_{k-2}}^2$, respectively. Similar to (7), the proposed algorithm reconstructs q_k via

$$\tilde{q}_k = \alpha \tilde{p}_{k-1} + (1 - \alpha) \tilde{p}_{k-2} + \hat{r}_k. \quad (11)$$

Similar to (8) and (9), the optimal coefficient α that minimizes the error variance $\sigma_{q_k}^2$ is

$$\alpha = \frac{\sigma_{\Theta}^2 + \sigma_{p_{k-2}}^2}{\sigma_{\Theta}^2 + \sigma_{p_{k-1}}^2 + \sigma_{p_{k-2}}^2} \quad (12)$$

and the minimum $\sigma_{q_k}^2$ is given by

$$\sigma_{q_k}^2 = \frac{\sigma_{p_{k-1}}^2 (\sigma_{\Theta}^2 + \sigma_{p_{k-2}}^2)}{\sigma_{\Theta}^2 + \sigma_{p_{k-1}}^2 + \sigma_{p_{k-2}}^2}. \quad (13)$$

D. Parameters for MMSE Decoding Algorithm

Table I summarizes several parameters used for the MMSE decoding. To obtain these parameters, ‘‘Coast guard,’’ ‘‘Container ship,’’ ‘‘Mobile and calendar,’’ ‘‘Mother and daughter,’’ ‘‘Silent voice,’’ and ‘‘Stefan’’ QCIF (176 × 144) sequences are used as the training sequences. In our implementation, each error variance in the error variance map is represented with an 8-bit integer. Therefore, the four error variances, i.e., σ_{Γ}^2 , $\sigma_{\Delta_1}^2$, $\sigma_{\Delta_2}^2$, and σ_{Θ}^2 , are normalized to be within [0, 255]. Note that the error variance due to the spatial concealment ($\sigma_{\Gamma}^2 = 255$) is the largest error variance.

TABLE I
PARAMETERS FOR MMSE DECODING WITH “COAST GUARD,” “CONTAINER SHIP,” “MOBILE AND CALENDAR,” “MOTHER AND DAUGHTER,” “SILENT VOICE,” AND “STEFAN” QCIF SEQUENCES USED AS THE TRAINING SEQUENCES

Modelling	Leaky Factor	Comments
of Error	$l_+ = 0.8$	When either v_x or v_y is of half-pixel accuracy
Propagation	$l_x = 0.65$	When both v_x and v_y are of half-pixel accuracy
Error Variance		
MMSE	$\sigma_{\Gamma}^2 = 255$	Spatial concealment
Decoding	$\sigma_{\Delta_1}^2 = 34$	Temporal concealment of erroneous MB - Mode 1
for H.263	$\sigma_{\Delta_2}^2 = 64$	Temporal concealment of erroneous MB - Mode 2
	$\sigma_{\Theta}^2 = 33$	Reconstruction of error-free inter MB - Mode 2
Error Variance		
MMSE	$\sigma_{\Gamma}^2 = 255$	Spatial concealment
Decoding	$\sigma_{\Delta_1}^2 = 26$	Temporal concealment of erroneous MB - Mode 1
for DMC	$\sigma_{\Delta_2}^2 = 41$	Temporal concealment of erroneous MB - Mode 2
	$\sigma_{\Theta}^2 = 2$	Reconstruction of error-free inter MB

In this work, we adopt the spatial and temporal concealment methods that utilize the information only in the upper macroblock, since they are simple and easy to be implemented in hardware. However, it is worthy to point out that more sophisticated spatial concealment methods [15], [16] and motion vector recovery algorithms [17]–[19] can be incorporated into the MMSE decoding algorithm to enhance the performance. In such a case, it is necessary to retrain the error variance parameters in Table I.

IV. MMSE DECODING ALGORITHM FOR DMC

A. Review of Simple Decoding in DMC

The DMC [1] is an effective algorithm for suppressing error propagation. In DMC, the data for each inter-MB contain two motion vectors \mathbf{v}_1 and \mathbf{v}_2 , which specify the locations of similar regions in the previous frame and the frame before the previous one, respectively. \mathbf{v}_1 is called the first motion vector, and \mathbf{v}_2 the second motion vector. A pixel q_k in F_k is reconstructed at the encoder as

$$\hat{q}_k = \frac{\hat{p}_{k-1} + \hat{p}_{k-2}}{2} + \hat{r}$$

where \hat{p}_{k-1} and \hat{p}_{k-2} are the reconstructed pixels in F_{k-1} and F_{k-2} , specified by \mathbf{v}_1 and \mathbf{v}_2 , respectively, and \hat{r} is the DCT-encoded residual.

In [1], a binary error map is employed at the decoder. In other words, a pixel is claimed to be corrupted or not, without using the notion of error variance. When the data for an MB are lost during the transmission, its pixels are claimed to be corrupted. On the other hand, when the data are received without error, its pixels are claimed to be not corrupted without considering error propagation. If the data for an inter-MB are received without error, the decoder reconstructs its pixel q_k by

$$\tilde{q}_k = \begin{cases} \tilde{p}_{k-2} + \hat{r}, & \text{if only } \tilde{p}_{k-1} \text{ is corrupted} \\ \tilde{p}_{k-1} + \hat{r}, & \text{if only } \tilde{p}_{k-2} \text{ is corrupted} \\ \frac{\tilde{p}_{k-1} + \tilde{p}_{k-2}}{2} + \hat{r}, & \text{otherwise.} \end{cases}$$

It was shown analytically and experimentally in [1] that DMC provides a better performance than the conventional single-vector motion compensation in an error-prone environment even with this simple decoding rule.

B. MMSE Decoding in DMC

The performance of DMC can be further improved by decoding each pixel based on the MMSE criterion. Specifically, the proposed algorithm reconstructs \tilde{q}_k by

$$\tilde{q}_k = \alpha \tilde{p}_{k-1} + (1 - \alpha) \tilde{p}_{k-2} + \hat{r} \quad (14)$$

where α is an adaptive weighting coefficient. Then, the error variance $\sigma_{q_k}^2$ can be written as

$$\begin{aligned} \sigma_{q_k}^2 &= E\{(\hat{q}_k - \tilde{q}_k)^2\} \\ &= E\left\{\left[\frac{\hat{p}_{k-1} + \hat{p}_{k-2}}{2} - (\alpha \tilde{p}_{k-1} + (1 - \alpha) \tilde{p}_{k-2})\right]^2\right\} \\ &= E\left\{\left[\alpha(\hat{p}_{k-1} - \tilde{p}_{k-1}) + (1 - \alpha)(\hat{p}_{k-2} - \tilde{p}_{k-2})\right.\right. \\ &\quad \left.\left.+ \left(\frac{1}{2} - \alpha\right)(\hat{p}_{k-1} - \hat{p}_{k-2})\right]^2\right\} \\ &\simeq \alpha^2 \sigma_{p_{k-1}}^2 + (1 - \alpha)^2 \sigma_{p_{k-2}}^2 + \left(\frac{1}{2} - \alpha\right)^2 \sigma_{\Theta}^2 \quad (15) \end{aligned}$$

where $\sigma_{\Theta}^2 = E\{(\hat{p}_{k-1} - \hat{p}_{k-2})^2\}$. In (15), it is assumed that

$$\begin{aligned} E\{(\hat{p}_{k-1} - \tilde{p}_{k-1})(\hat{p}_{k-2} - \tilde{p}_{k-2})\} \\ &= E\{(\hat{p}_{k-1} - \tilde{p}_{k-1})(\hat{p}_{k-1} - \hat{p}_{k-2})\} \\ &= E\{(\hat{p}_{k-2} - \tilde{p}_{k-2})(\hat{p}_{k-1} - \hat{p}_{k-2})\} = 0. \end{aligned}$$

Then, the optimal α that minimizes $\sigma_{q_k}^2$ is given by

$$\alpha = \frac{\frac{1}{2} \sigma_{\Theta}^2 + \sigma_{p_{k-2}}^2}{\sigma_{\Theta}^2 + \sigma_{p_{k-1}}^2 + \sigma_{p_{k-2}}^2}. \quad (16)$$

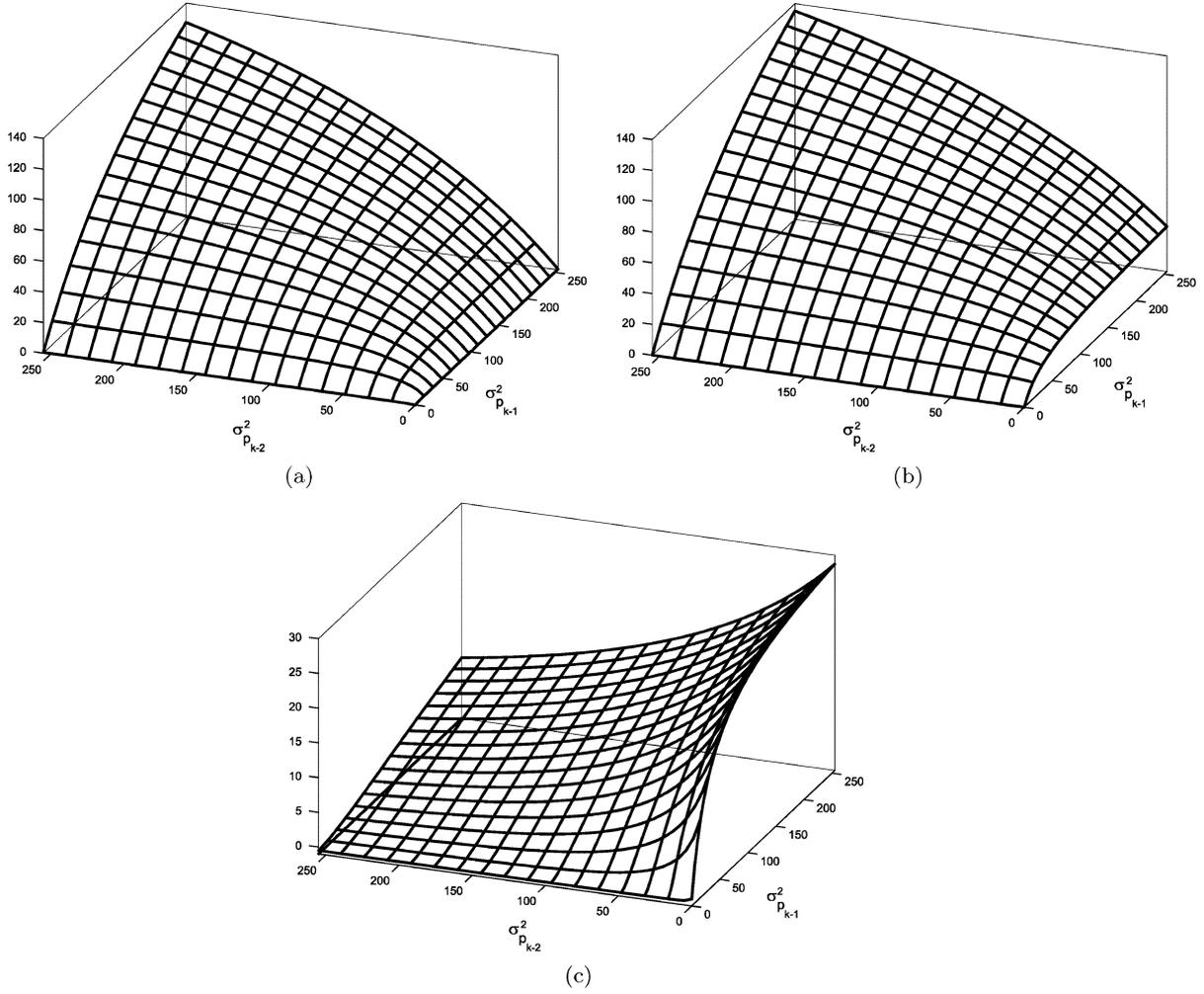


Fig. 2. Perspective plots of $\sigma_{q_k}^2$ with respect to $\sigma_{p_{k-1}}^2$ and $\sigma_{p_{k-2}}^2$. (a) $\sigma_{q_k}^2$ in the DMC decoder. (b) $\sigma_{q_k}^2$ in the H.263 decoder. (c) Difference (b)–(a).

Thus, \tilde{q}_k is reconstructed by inserting (16) into (14), and its error variance is updated by inserting (16) into (15).

As mentioned previously, in DMC, each MB B_k contains the first motion vector \mathbf{v}_1 and the second motion vector \mathbf{v}_2 , which specify the locations of B_{k-1} in the previous frame and B_{k-2} in the frame before the previous one, respectively. The motion vectors are selected such that $\|B_k - B_{k-1}\|$ and $\|B_k - B_{k-2}\|$ are minimized, where $\|\cdot\|$ denotes the sum of absolute differences (SAD). Since

$$\|B_{k-1} - B_{k-2}\| \leq \|B_k - B_{k-1}\| + \|B_k - B_{k-2}\|$$

by the triangle inequality, B_{k-1} and B_{k-2} exhibit high correlation in general. Furthermore, if $\|B_{k-1} - B_{k-2}\|$ is larger than a pre-specified threshold, B_k is encoded in the intramode [1]. Thus, in DMC, \hat{p}_{k-1} and \hat{p}_{k-2} are highly correlated, and the trained value of $\sigma_{\Theta}^2 = E\{(\hat{p}_{k-1} - \hat{p}_{k-2})^2\}$ is a small number as presented in Table I. Therefore, if either \hat{p}_{k-1} or \hat{p}_{k-2} in (14) is not corrupted, \tilde{q}_k can be reconstructed faithfully by setting α close to 1 or 0, respectively.

Fig. 2(a) gives the perspective plot of $\sigma_{q_k}^2$ in (15) with respect to $\sigma_{p_{k-1}}^2$ and $\sigma_{p_{k-2}}^2$. It can be observed that $\sigma_{q_k}^2$ is almost zero when either $\sigma_{p_{k-1}}^2$ or $\sigma_{p_{k-2}}^2$ is equal to zero.

The MMSE decoding of I-frames and error-free intra-MBs in P-frames in DMC is the same as done in the MMSE H.263 decoder. Also, each erroneous MB in P-frames is decoded by employing two temporal concealment modes as in the H.263 decoder. Mode 1 is the same as that of the H.263 decoder, i.e., it copies the erroneous MB from the previous frame using the first motion vector $\hat{\mathbf{v}}_1$ of the upper MB. However, mode 2 copies the erroneous MB from the frame before the previous one by using the second motion vector $\hat{\mathbf{v}}_2$ of the upper MB, which was experimentally confirmed to be a more reliable estimate than $2\hat{\mathbf{v}}_1$. Table I presents the error variances $\sigma_{\Delta_1}^2$ and $\sigma_{\Delta_2}^2$ due to these two concealment modes.

C. Comparison of MMSE Decoding for H.263 and DMC

Let us compare the MMSE decoding behavior for H.263 and DMC briefly. Since the H.263 encoder transmits only the first motion vector \mathbf{v} , the H.263 decoder should estimate the second motion vector. In our approach, $2\mathbf{v}$ is employed as the second motion vector. Due to the estimation error in the second motion vector, $\sigma_{\Theta}^2 = E\{(\hat{p}_{k-1} - \hat{p}_{k-2})^2\}$ is much larger in H.263 than in DMC, as presented in Table I.

Fig. 2(b) shows a perspective plot of $\sigma_{q_k}^2$ in (13). Notice that in H.263 $\sigma_{q_k}^2$ has a relatively large value, even if $\sigma_{p_{k-2}}^2$ is zero.

Fig. 2(c) shows the difference graph between Fig. 2(a) and (b). It can be seen that the DMC decoder yields smaller $\sigma_{q_k}^2$ than the H.263 decoder for almost every combination of $\sigma_{p_{k-1}}^2$ and $\sigma_{p_{k-2}}^2$. The exceptional cases are concentrated near the axis $\sigma_{p_{k-1}}^2 = 0$, where the DMC decoder yields at most 0.5 larger error variance than the H.263 decoder, which is however negligible.

Based on the above discussion, we conclude that the DMC algorithm can suppress error propagation more effectively than H.263 by transmitting two motion vectors per MB.

V. IMPLEMENTATION DETAILS

A. Complexity Analysis

Compared to the conventional H.263 decoder, the computational complexities of the MMSE decoders increase modestly. In addition to bit stream parsing, interpolation of image frames, and inverse discrete cosine transform (IDCT), the following additional computations are required in the MMSE H.263 decoder.

- Interpolation of error variance maps. As shown in Fig. 1, there are two p_+ -type pixels and one p_\times -type pixel for each ordinary pixel. The error variance for the p_+ -type pixel can be computed via (2), requiring one addition and one multiplication. Also, the error variance for the p_\times -type pixel can be computed via (3), requiring three additions and one multiplication. Thus, the interpolation of error variance maps requires five additions per pixel (app) and three multiplications per pixel (mpp).
- Concealment of erroneous MBs. To conceal an erroneous pixel, the decoder first computes the weighting coefficient α by (6) and (8), requiring three additions and one multiplication. Then, the reconstruction in (7) needs two additions and two multiplications, and the error variance map updating in (9) needs one multiplication since $\sigma_{q_k}^2 = \alpha \cdot \sigma_{q_{k,1}}^2$. Thus, the concealment requires 5 app and 4 mpp.
- Reconstruction of error-free MBs. Similar to the concealment case, the reconstruction requires 5 app and 4 mpp.

To conclude, the MMSE H.263 decoder requires additional 10 app and 7 mpp in total.

Similarly, for the MMSE DMC decoder, 10 app and 7 mpp are required for the concealment of erroneous MBs, and 15 app and 12 mpp for the reconstruction of error-free MBs. The DMC decoder is computationally more expensive than the H.263 decoder, since the weighting coefficient α in (16) and the error variance map updating rule in (15) have more complex forms. However, it will be shown in simulation results that the DMC decoder provides a much better performance than the H.263 decoder at the cost of a slightly higher computational complexity and a slightly higher bit rate for the additional set of motion vectors.

B. Memory Analysis

The conventional H.263 decoder motion-compensates the k th frame F_k from F_{k-1} , and thus demands two frame buffers for video decoding. The proposed MMSE decoders predict F_k from F_{k-1} and F_{k-2} adaptively by using the error variance maps for F_{k-1} and F_{k-2} . Then, the error variance map for F_k is

updated. Thus, the MMSE decoders need six frame buffers in total; namely, three for image frames and three for error variance maps.

The high memory requirement of MMSE decoders increases the implementational complexity and may not be practical in some applications today, e.g., handheld devices. Reducing the memory requirement is a topic for future research. One idea is to lower the resolution of the error variance map. For example, error variances can be recorded at the block or the MB level instead of at the pixel level. Then, the error propagation model should be modified accordingly. It is expected that there is a tradeoff relation between the resolution of the error variance map and the MMSE decoding performance.

C. Use of Double I-Frames

We use I-frames after scene cuts at the encoder. Suppose that we encode F_k after a scene cut in the intramode, and the next frame F_{k+1} in the intermode. Then, note that the MMSE decoders conceal or reconstruct each pixel p_{k+1} in F_{k+1} by using the weighted sum of two pixels, p_k in F_k and p_{k-1} in F_{k-1} . When the error variance of p_k is higher than that of p_{k-1} , the decoders assign more weight on p_{k-1} that is a pixel before the scene cut. This can lead to a disastrous effect, if the two scenes contain completely different contents.

To overcome this problem, we encode two consecutive frames in the intramode after each scene cut. This approach decreases coding efficiency, especially when an image sequence contains many scenes, but enhances error resilience. In typical image sequences, spatial correlation is lower than temporal correlation so that spatial concealment errors are often larger than temporal concealment errors. Note that the propagation of spatial concealment errors can be alleviated using double I-frames. Specifically, even if one I-frame is corrupted, the MMSE decoders can effectively reconstruct subsequent P-frames with the correctly received information in the other I-frame.

VI. SIMULATION RESULTS

We evaluate the performance of the proposed MMSE decoders with parameters given in Table I. H.263 bit streams are generated by Telenor TMN codec version 2.0. Several coding options [3], such as the advanced prediction mode and the reference picture selection mode, are not employed in our implementation. DMC bit streams are also generated by modifying the TMN codec. Test sequences are ‘‘Carphone,’’ ‘‘Foreman,’’ and ‘‘Glasgow’’ QCIF sequences, consisting of 100, 100, and 750 frames, respectively. Their frame rates are 8.33, 8.33, and, 12.5 f/s. Note that the six training sequences that were utilized to derive the parameters in Table I are not used as test sequences. The peak signal-to-noise ratio (PSNR) is employed as an objective measure of video quality, and is defined as $10 \log(255^2/\varepsilon^2)$ where ε^2 is the mean square error between the decoded and the original luminance frames.

A. Case I: Transient Error Model

First, we compare the error resilience of the MMSE decoders with that of the conventional decoders by simulating a transient



Fig. 3. 10th, 11th, and 35th frames of “Carphone” sequences. (a) Original sequence. Image decoded by using (b) conventional H.263, (c) MMSE H.263, (d) conventional DMC, and (e) MMSE DMC.

and severe transmission error. In this test, the “Carphone” sequence is encoded with the fixed quantizer step sizes of 10 and 12 to generate the H.263 bit stream and the DMC bit stream, respectively. The bit rate for the H.263 bit stream is 43.8 kb/s, and that for DMC is 42.5 kb/s. Fig. 3(a) shows the original 10th, 11th, and 35th frames. Fig. 3(b)–(e) show the reconstructed frames, when all MBs within the white rectangle in Fig. 3(a) are lost during transmission.

Fig. 3(b) is obtained by the conventional H.263 decoder that employs only the temporal concealment mode 1 in Section III.

More specifically, it copies each erroneous MB from the previous frame using the motion vector of the upper MB. It can be seen that the erroneous region contains very severe artifacts, since the estimated motion vectors are not reliable. Moreover, the subsequent 35th frame, which is 3 s after the erroneous 10th frame, is still severely corrupted around the face due to error propagation. Fig. 3(c) is reconstructed by the MMSE H.263 decoder. The erroneous region also contains severe artifacts, since it is too large for both modes 1 and 2 to be effective. However, as compared to Fig. 3(b), the error propagation phenomenon in

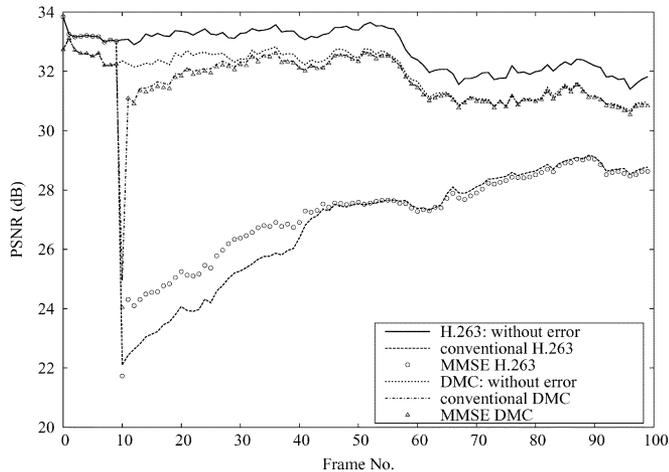


Fig. 4. Error propagation in the “Carphone” sequence when the 2nd–8th GOBs in 10th frame are lost during transmission.

the subsequent 11th frame is alleviated by using the information in the frame before the previous one.

Fig. 3(d) shows the reconstructed frames with the conventional DMC decoder, which copies the erroneous region from the previous frame and does not use the erroneous region in the motion compensation of the subsequent frames [1]. It can suppress error propagation more effectively than both H.263 decoders, since each MB contains the second motion vector specifying a similar region in the frame before the previous one. Note that the subsequent frames are almost free of artifacts. Fig. 3(e) is decoded by the MMSE DMC decoder. It can be seen that the MMSE DMC decoder also suppresses the error propagation effectively.

Fig. 4 shows the PSNR curves. The MMSE H.263 decoder provides a better performance than the conventional H.263 decoder up to 1.9 dB. Also, the DMC decoders effectively localize the effect of transmission errors and outperforms the H.263 decoders by large margins. In this example, the MMSE decoding does not improve the performance of the DMC algorithm, since the transient error can be effectively suppressed even with the simple decoding rule of the conventional DMC decoder. However, as will be shown later, the MMSE decoding also enhances the error resilient capability of the DMC algorithm in more severe error conditions.

B. Case II: GOB Loss Model

Next, we employ a group of block (GOB) loss model, in which the locations of erroneous GOBs are randomly selected according to a given GOB loss rate (GLR).

$$\text{GLR} = \frac{\text{number of erroneous GOBs}}{\text{number of total GOBs}}.$$

For QCIF video, a GOB consists of a single MB row. Furthermore, when the optional slice structured mode [3] is not used, a GOB is a minimum independent decoding unit in H.263. When the slice structured mode is used, an arbitrary number of MBs within a frame can be packetized into one independently decodable packet. If the packet size is fixed, a packet for an I-frame

consists of less MBs than that for a P-frame and the effect of a packet loss is more localized in I-frames. Since I-frames are generally more difficult to conceal than P-frames, error resilience of compressed bit streams often benefits from the slice structured mode and the fixed size packetization. In the packet loss environment, it is expected that the proposed MMSE decoding provides a similar performance improvement as reported in this work.

As mentioned in Section V-C, double I-frames are inserted after each scene cut to enhance the error resilience of the MMSE decoders. However, in the conventional H.263 decoder, the insertion of double I-frames only wastes the bit rate without increasing error resilience, since each frame is reconstructed using only the information in the previous frame. Thus, a single I-frame is inserted after each scene cut for the performance evaluation of the conventional H.263 decoder. The “Carphone” and “Foreman” sequences are composed of one scene, respectively, while the “Glasgow” sequence consists of 16 scenes.

Fig. 5 shows a typical example of the error variance map. The 11th frame of the “Foreman” sequence is reconstructed by the DMC algorithm at a bit rate of 61.1 kb/s. Fig. 5(a) is the error-free reconstruction. Fig. 5(b) is the MMSE decoder reconstruction, when GLR is 0.05. It is corrupted by the propagated error arising from 10 erroneous GOBs in previous frames. Fig. 5(c) depicts squared pixel errors for the case in Fig. 5(b). The decoder does not know the exact pixel errors but estimates the error variance map as shown in Fig. 5(d). From the viewpoint of the decoder, each pixel error is just one sample of a random variable. Therefore, it is not surprising that the error variance in Fig. 5(d) is not very close to the corresponding squared error given in Fig. 5(c). However, we do see that the error variance map captures the overall tendency of the effect of transmission errors. In other words, the error variance map indicates how severely each pixel is corrupted in a probabilistic sense. Using the error variance map, the decoder can reconstruct subsequent frames in a reliable way.

For comparison, Fig. 5(e) shows the reconstruction of the conventional DMC decoder. Comparing Fig. 5(b) with Fig. 5(e), we can see that the MMSE DMC decoder provides a better visual quality (about 2 dB) than the conventional DMC decoder.

Figs. 6–8 show the rate-distortion relations of the H.263 and DMC decoders on the three test sequences. Since the locations of errors affect the quality of the resulting video sequence significantly, errors are randomly generated by 30 different seeds, and the obtained PSNRs are averaged over all frames and all different error patterns. Based on Figs. 6–8, we have the following observations.

- MMSE H.263 versus conventional H.263.

In the error-free condition (GLR = 0), the MMSE H.263 decoder yields a slightly worse performance than the conventional H.263 decoder due to the overhead of double I-frames. However, when GLR is 0.05, the MMSE decoder provides a significantly better performance than the conventional decoder. For example, at 100 kb/s, it provides about 1.1, 0.5, and 0.9 dB gains on the “Carphone,” “Foreman,” and “Glasgow” sequences, respectively. The “Foreman” sequence contains fast and complex motions,

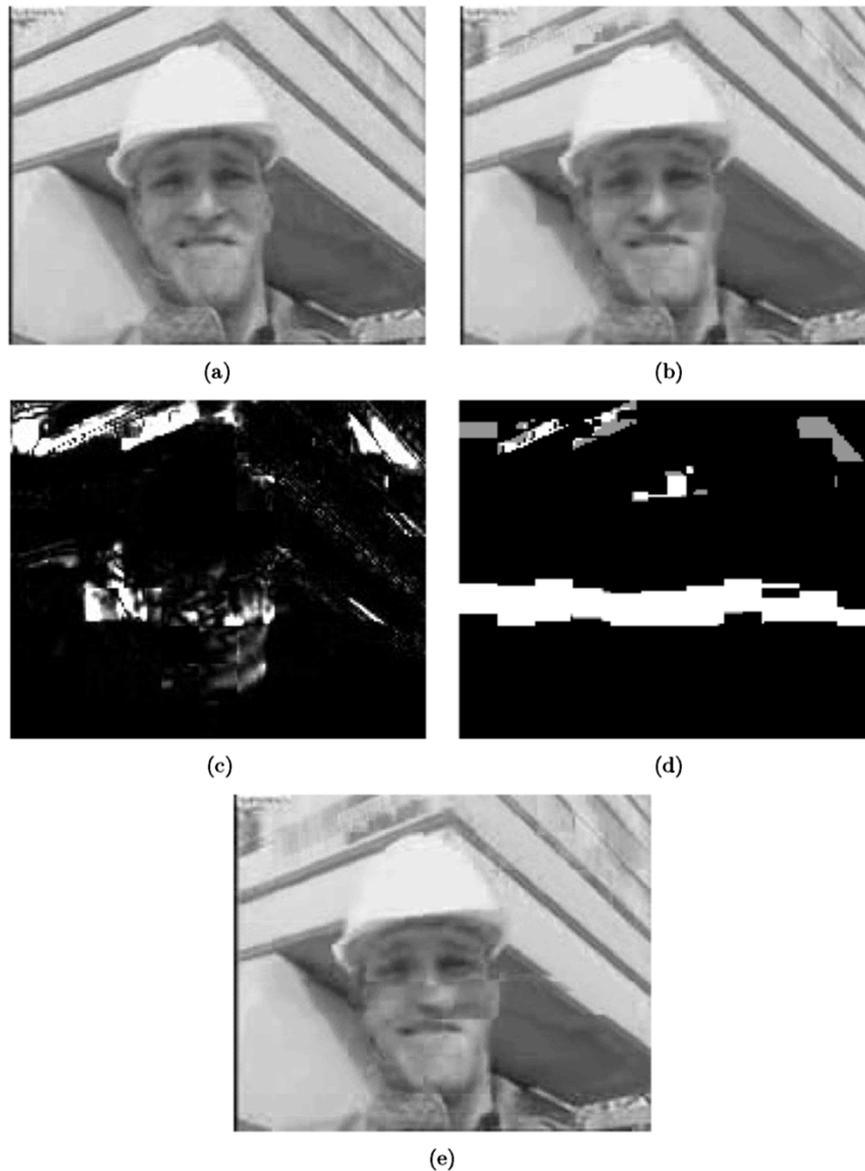


Fig. 5. 11th frame of the “Foreman” sequence. (a) Error-free reconstruction (33.4 dB). (b) MMSE decoder reconstruction at $GLR = 0.05$ (26.9 dB). (c) Squared pixel errors. (d) Error variance map. (e) Conventional decoder reconstruction (24.9 dB).

including shaky camera panning. Therefore, the concealment mode 2 and the reconstruction mode 2 in Section III are less reliable, and the PSNR gain for the “Foreman” sequence is smaller than those for the other two sequences.

- MMSE DMC versus conventional DMC.

The MMSE DMC decoder provides a better performance than the conventional DMC decoder. For example, it provides up to 1.2 dB gain for the “Foreman” sequence. The gain is the smallest for the “Carphone” sequence, which contains relatively slow motions. In such a case, the error propagation phenomenon can be effectively suppressed even with the simple decoding rule of the conventional DMC decoder.

- MMSE DMC versus conventional H.263.

When GLR is 0.05, the MMSE DMC decoder outperforms the conventional H.263 decoder by a significant margin at the cost of the increased bit rate for motion vectors and double I-frames. For example, at 100 kb/s, it pro-

vides a better performance than the conventional H.263 decoder with a gain of 6.3, 5.2, and 3.8 dB on the “Carphone,” “Foreman,” and “Glasgow” sequences, respectively.

Fig. 9 compares the performance for the “Carphone” sequence at various GLRs, when the total bit rate is fixed to 64 kb/s. As GLR increases, the DMC decoders yield more graceful degradation than the H.263 decoders. Also, the MMSE decoding enhances the robustness of both H.263 and DMC bit streams. Let us suppose that the target image quality is 32 dB. Then, the MMSE DMC decoder can tolerate about 5% loss of GOBs, while the conventional H.263 decoder can tolerate only 1% losses. Fig. 10 compares the performance on the “Foreman” sequence at 64 kb/s, and Fig. 11 on the “Glasgow” sequence at 128 kb/s. They exhibit the similar tendency as observed in Fig. 9.

For comparison, Figs. 9–11 also show the performance of the decoder motion vector estimation (DMVE) algorithm [18], which is one of the most efficient temporal EC methods.

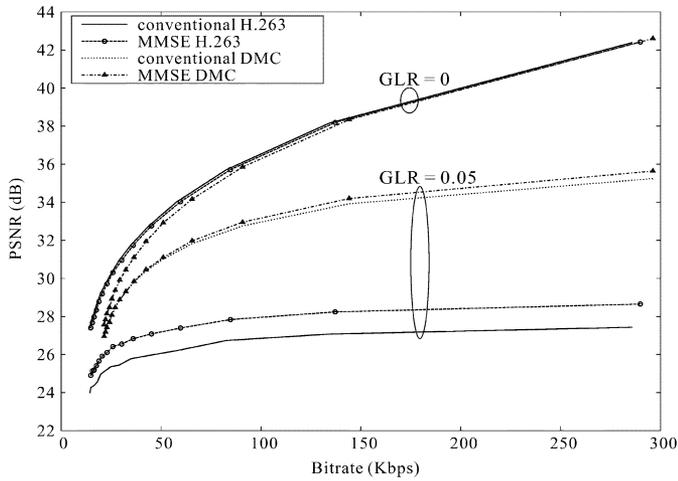


Fig. 6. Performance of the “Carphone” sequence at $GLR = 0$ and 0.05 . At $GLR = 0$, the conventional DMC provides the same performance as the MMSE DMC.

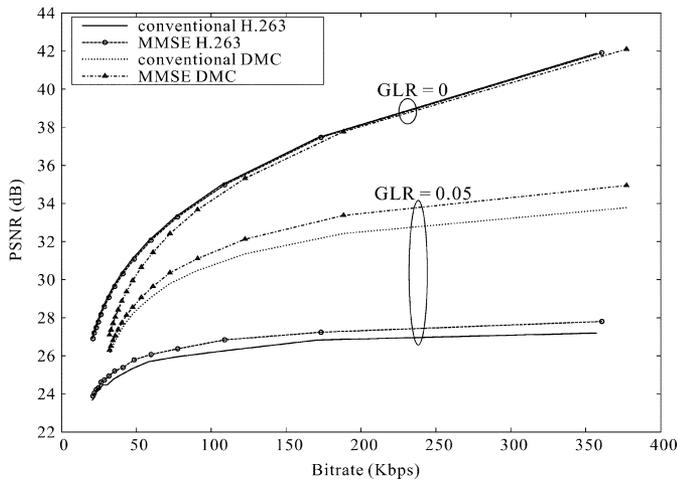


Fig. 7. Performance of the “Foreman” sequence at $GLR = 0$ and 0.05 . At $GLR = 0$, the conventional DMC provides the same performance as the MMSE DMC.

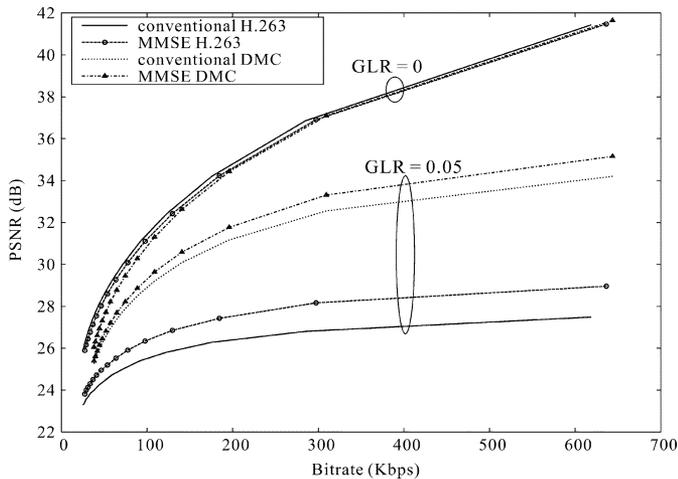


Fig. 8. Performance of the “Glasgow” sequence at $GLR = 0$ and 0.05 . At $GLR = 0$, the conventional DMC provides the same performance as the MMSE DMC.

When an MB is erroneous, DMVE estimates the motion vector for the set of the surrounding pixels, and applies that vector to

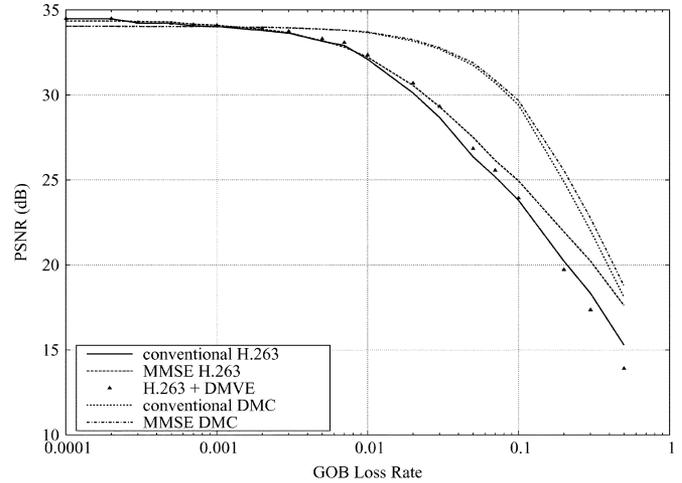


Fig. 9. Performance of the “Carphone” sequence at a bit rate of 64 kb/s.

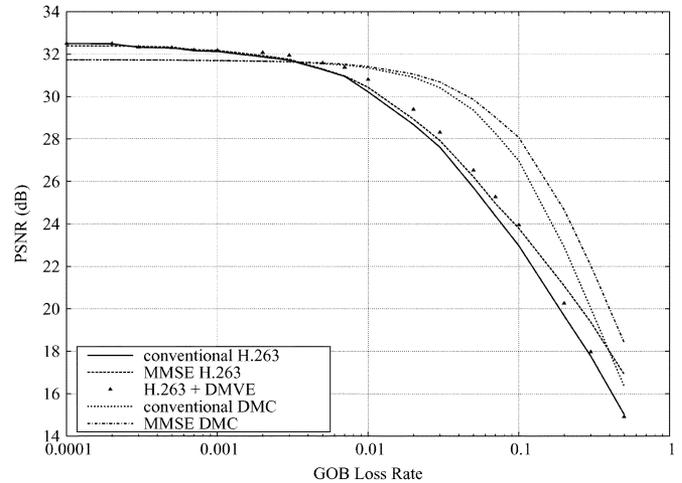


Fig. 10. Performance of the “Foreman” sequence at a bit rate of 64 kb/s.

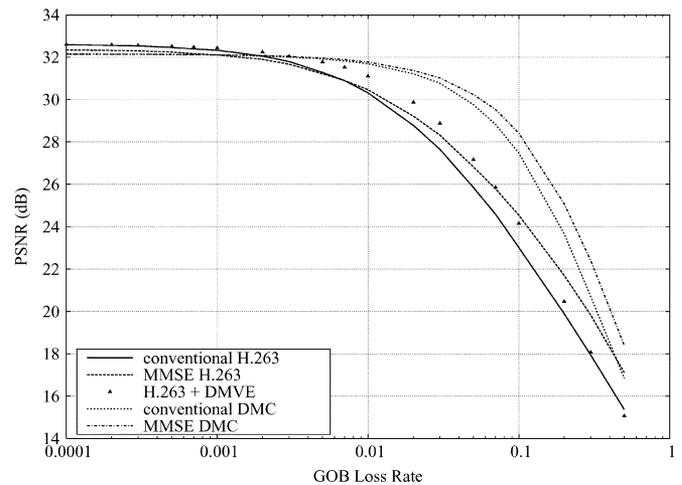


Fig. 11. Performance of the “Glasgow” sequence at a bit rate of 128 kb/s.

the erroneous MB. We can see that DMVE enhances the decoding performance of the H.263 standard in moderate error conditions. However, it provides a worse performance than the MMSE H.263 decoder, when GLR is higher than 0.03, 0.1 and 0.07 for the “Carphone,” “Foreman” and “Glasgow” sequences,

respectively. This is because the surrounding pixels and the estimated motion vector are also unreliable in very severe error conditions. Furthermore, even though DMVE retrieves lost motion vectors at the cost of high computational complexity, its performance is still much worse than those of the DMC decoders.

It is worthy to point out that the MMSE decoding algorithm has complementary relation to many other error resilient schemes. For example, we can select the best two modes among several advanced EC techniques, including DMVE, and then combine them according to the MMSE criterion. Also, the proposed MMSE decoders can be combined with the encoder-based error resilient schemes [7], such as reversible variable length coding (RVLC) and adaptive intrarefreshing [10], to enhance the error-resilience further.

VII. CONCLUSION

We proposed a robust video decoding algorithm based on an MMSE criterion, which can alleviate the effect of transmission errors effectively. The proposed MMSE decoding algorithm uses an error propagation model for the decoder to estimate the MSE of each pixel. Then, it reconstructs both error-free and erroneous blocks in the k th frame F_k by using a weighted superposition of two blocks in F_{k-1} and F_{k-2} . The weights are adaptively determined to minimize the mean square error of each pixel based on the error propagation model.

The MMSE algorithm was implemented in association with the H.263 coding standard and the DMC coder, respectively. Experimental results on H.263 bit streams demonstrated that the MMSE algorithm provides a better performance than the conventional method. Furthermore, when combined with the DMC coder, the proposed MMSE decoding algorithm significantly enhances error resilience. It is our belief that the proposed algorithm is suitable for many video delivery applications, especially for low-latency lossy wireless channels.

REFERENCES

- [1] C.-S. Kim, R.-C. Kim, and S.-U. Lee, "Robust transmission of video sequence using double-vector motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 1011–1021, Sep. 2001.
- [2] *Video Codec for Audiovisual Services at $p \times 64$ kbit/s*, 1993. ITU-T Recommendation H.261.
- [3] *Video Coding for Low Bitrate Communication*, 1998. ITU-T Recommendation H.263.
- [4] *Information Technology—Generic Coding of Audio-Visual Objects, Part 2: Visual*, 1998. ISO/IEC 14496-2.
- [5] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
- [6] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.
- [7] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, pp. 61–82, Jul. 2000.
- [8] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 7, pp. 867–877, Nov. 1998.
- [9] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 standard," *IEEE Commun. Mag.*, vol. 36, pp. 112–119, Jun. 1998.
- [10] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intramode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.
- [11] G. Côté, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 952–965, Jun. 2000.
- [12] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1050–1062, Jun. 2000.
- [13] N. Färber, K. Stuhlmüller, and B. Girod, "Analysis of error propagation in hybrid video coding with application to error resilience," in *Proc. ICIP*, vol. 2, Oct. 1999, pp. 550–554.
- [14] G. de los Reyes, A. R. Reibman, S.-F. Chang, and J. C.-I. Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1063–1074, Jun. 2000.
- [15] Y. Wang, Q.-F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1544–1551, Oct. 1993.
- [16] J. W. Park and S. U. Lee, "Recovery of corrupted image data based on the NURBS interpolation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 7, pp. 1003–1008, Oct. 1999.
- [17] M.-J. Chen, L.-G. Chen, and R.-M. Weng, "Error concealment of lost motion vectors with overlapped motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 560–563, Jun. 1997.
- [18] J. Zhang, J. F. Arnold, and M. R. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 659–665, Jun. 2000.
- [19] S. Tsekeridou and I. Pitas, "MPEG-2 error concealment based on block-matching principles," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 646–658, Jun. 2000.
- [20] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 5, pp. 807–814, Jun. 1989.
- [21] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 6, pp. 872–881, Dec. 1997.
- [22] M. Ghanvari, "Postprocessing of late cells for packet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 6, pp. 669–678, Dec. 1996.
- [23] C.-S. Kim, R.-C. Kim, and S.-U. Lee, "Robust transmission of video sequence over noisy channel using parity-check motion vector," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 7, pp. 1063–1074, Oct. 1999.



Chang-Su Kim (S'95-M'01) was born in Seoul, Korea, in 1971. He received the B.S. and M.S. degrees in control and instrumentation engineering in 1994 and 1996, respectively, and the Ph.D. degree in electrical engineering in 2000, all from Seoul National University (SNU), Seoul, Korea.

From 2000 to 2001, he was a Visiting Scholar with the Signal and Image Processing Institute, University of Southern California, Los Angeles, and a Consultant for InterVideo Inc., Los Angeles. From 2001 to 2003, he was a Postdoctoral Researcher with the School of Electrical Engineering, SNU. In August 2003, he joined the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong, as an Assistant Professor. His research topics include video and 3-D graphics processing and multimedia communications.



JongWon Kim (S'88-M'95-SM'01) was born in Sangju, Korea, on September 4, 1964. He received the B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, in 1987, 1989, and 1994, respectively, all in control and instrumentation engineering.

During 1994–1999, he was with the Department of Electronics Engineering at the Kongju National University, Kongju, Korea, as an Assistant Professor. From 1997 to 2001, he was visiting the Signal and Image Processing Institute (SIPI) of Electrical Engineering–Systems Department at the University of Southern California, Los Angeles, where he has served as a Research Assistant Professor since December 1998. From September 2001, he joined the Department of Information and Communications, Kwang-Ju Institute of Science and Technology (K-JIST), Gwangju, Korea as an Associate Professor. Under the slogan of "reliable and flexible delivery system for integrated multimedia over wired/wireless IP networks," he is focusing on networked media systems and protocols including multimedia signal processing and communications. He has coauthored around 150 technical publications in international/domestic journals and conferences.

Dr. Kim is a member of ACM, SPIE, KICS, and IEEE.



Ioannis Katsavounidis (M'95) was born in Veria, Greece, in 1968. He received the Diploma degree in electrical engineering in 1991 from the Aristotle University of Thessaloniki (A.U.Th.), Thessaloniki, Greece. He received the M.S. degree in 1993 and the E.E.E. and Ph.D. degrees in electrical engineering in 1997 and 1998, respectively, all from the University of Southern California, Los Angeles, working within the Signal and Image Processing Institute (SIPI) and Integrated Media Systems Center (IMSC).

From 1996 to 2000, he worked in Italy as an engineer for the High-Energy Physics Department of the California Institute of Technology (MACRO experiment). In July 2003, he joined InterVideo, Inc., Fremont, CA, where he is currently a Software Manager. His research topics include image and video compression, video processing, and algorithm optimization for multimedia applications.



C.-C. Jay Kuo (S'86–M'87–SM'92–F'99) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is with the Department of Electrical Engineering, the Signal and Image Processing Institute (SIPI), and the Integrated Media Systems Center (IMSC) at the University of Southern California (USC), Los Angeles, as Professor of Electrical

Engineering and Mathematics. His research interests are in the areas of digital media processing, multimedia compression, communication and networking technologies, and embedded multimedia system design. He has guided 55 students to their Ph.D. degrees and supervised 15 postdoctoral research fellows. Currently, his research group at USC consists of around 30 Ph.D. students and five postdoctors (please visit website <http://viola.usc.edu>), which is one of the largest academic research groups in multimedia technologies. He is coauthor of more than 700 technical publications in international conferences and journals as well as the following seven books: *Content-based Audio Classification and Retrieval for Audiovisual Data Parsing* (with Tong Zhang, New York: Kluwer, 2001), *Semantic Video Object Segmentation for Content-based Multimedia Applications* (with Ju Guo, New York: Kluwer, 2001), *Intelligent Systems for Video Analysis and Access over the Internet* (with Wensheng Zhou, Englewood Cliffs, NJ: Prentice-Hall, 2002), *Video Content Analysis Using Multimodal Information* (with Ying Li, New York: Kluwer, 2003) *Quality of Service for Internet Multimedia* (with Jitae Shin and Daniel Lee, Englewood Cliffs, NJ: Prentice-Hall, 2003), *Radio Resource management for Multimedia QoS support in Wireless Cellular Networks* (with Huan Chen, Lei Huang, and Sunil Kumar, New York: Kluwer, 2003) and *High Fidelity Multichannel Audio Coding* (with Dai Tracy Yang and Chris Kyriakakis, Sylvania, OH: Hindawi, 2004).

Dr. Kuo is a Fellow of SPIE. He received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively. He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, Editor for the *Journal of Information Science and Engineering*, and the *EURASIP Journal of Applied Signal Processing*. He is also on the Editorial Board of the *IEEE Signal Processing Magazine*. He served as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING in 1995–1998, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1995–1997 and IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING in 2001–2003.