A Data Embedded Video Coding Scheme for Error-Prone Channels

Jie Song and K. J. Ray Liu, *Senior Member, IEEE*

Abstract—We propose a novel data embedding scheme for fractional-pixel based video coding algorithms such as H.263 and MPEG-2. By modifying the motion estimation procedure at fractional-pel precision, two bits data can be embedded in a motion vector (MV) for a inter-mode coded macroblock (MB). For half-pel precision motion estimation, the resulted bitstream is compatible with the current video coding standards. Performance of the proposed data embedding scheme in terms of compression efficiency is also studied. As an application example of the proposed data embedding scheme, an error-resilient video coding scheme is also presented where some redundant information, which is used to protect MVs and coding modes of MBs in one frame, is embedded into the MVs in the next frame. When errors occur in one group of blocks (GOB), the decoder can exactly recover the MVs of the corrupted GOB if the next frame can be received correctly. Simulations show that the proposed scheme has better performance than standard H.263 coding scheme for transmission over error-prone channels.

Index Terms—Data hiding, digital watermarking, error concealment, video coding.

I. INTRODUCTION

VIDEO compression algorithms such as H.263 can compress video sequences efficiently by exploiting both spatial and temporal redundancy. However, the compressed information is very sensitive to transmission errors and losses caused by channel impairment. Furthermore, the error introduced by the channel may propagate to the following video frames because of inter-frame compensation used in the current video coding standards. The retransmission of damaged frames using automatic repeat request (ARQ) is not suitable in real- time video communication. Forward error control coding (FEC) may protect the data up to certain degree but will take up some precious bandwidth resource when the bandwidth is limited.

Numerous methods have been proposed for error concealment in video communication when some data is lost or corrupted by exploiting the redundancy of the video signal [1]. However, even sophisticated concealment cannot totally avoid image degradation, and the accumulation of several small errors can also result in poor image quality. Normally, the image content of a corrupted group of blocks (GOB) is replaced by simply repeating the GOBs in the previous frame. This method works well for sequences with little motion. However, severe distortions are introduced for image regions containing highly active motion. In [2], one pseudoslice is added to one frame which is an erasure code of other slices in the same frame, such a method works well to recover one slice of errors, but will increase the bit rate by more than 10%. An error tracking combined with feedback channel method is proposed in [3] to deal with bursty errors by sending intra-mode coded macroblocks (MBs) to stop error propagation, but the recovery speed depends on round-trip delay of the network, and in a heavy spatial error propagation case, this method cannot send enough intra-mode MBs in one frame to stop error propagation because of the bit-rate limitation. In [1], when errors occur at the decoder, the locations of the corrupted MBs are sent to the encoder through feedback channel, and the accumulated backward motion track of affected pixels in the following frames are stored. During the period between the information error and the arrival of the retransmitted data, any error concealment methods introduced in [1] can be used. After the decoder receives the information of the corrupted MBs retransmitted from the encoder, the decoding process is redone for the frame with error and lossless recovery is achieved using the previous stored motion track. This method does not require the encoder to encode INTRA-mode MB to stop the error propagation, but this method still cannot provide very good performance when the motion is highly active during the time between the errors occur and lossless recovery. Other techniques for error resilient video coding can be found in [4], [5].

The method proposed here aims at obtaining good recovery performance by protecting important information using data embedding [6]. In the proposed method, when errors occur, the video quality is better than any other error concealment methods because of exact motion vector (MV) recovery. In this paper, we use H.263 video coding standard to present our data embedding scheme for fractional-pixel motion estimation based video coding.

In Section II, we present a new data embedding scheme at fractional-pel MV for video coding. The performance of the proposed scheme in terms of compression efficiency is also studied. In Section III, an error resilient video coding employing the proposed data embedding scheme is also proposed to provide an robust video transmission through error-prone channels. In Section IV, simulation results are presented to show the effectiveness of the data embedding scheme for both video coding efficiency and robustness to error propagation. Finally, conclusions are reached.

J. Song is with Agere Systems, Holmdel, NJ 07733 USA (e-mail: jiesong@agere.com).

K. J. R. Liu is with the Department of Electrical and Computer Engineering, Institute for System Research, University of Maryland, College Park, MD 20742 USA (e-mail: kjrliu@isr.umd.edu).
Publisher Item Identifier S 1520-9210(01)10029-5.
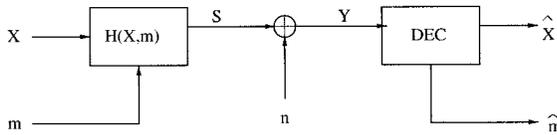
Fig. 1.   General information embedding scheme. An information m is embedded in the host signal $\mathbf{x}$. A noise n corrupts the composite signal $\mathbf{s}$. The decoder extracts the estimate $\hat{m}$ and reconstruct signal $\hat{\mathbf{x}}$ from channel output y.
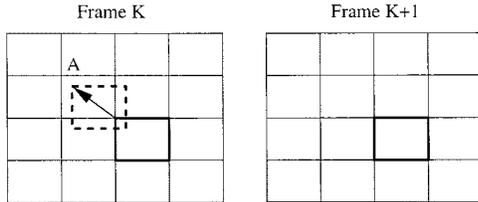
Fig. 2.   Integer-pixel motion prediction in H.263: the block (with dashed line) at integer pixel A in frame K is the motion prediction of the current block (with solid line) in frame $K + 1$.

II. DATA EMBEDDING

Many data embedding methods can be illustrated by Fig. 1 [7]. There is a host signal $\mathbf{x}$ in which we wish to embed some information m. The host signal could be a vector of pixel values or discrete cosine transform (DCT) coefficients from an image, for example. A data embedding function $\mathbf{H}(\mathbf{x}, m)$ maps $\mathbf{x}$ and m to a composite signal $\mathbf{s}$ subject to some distortion constraint. The composite signal is passed through a channel, where $\mathbf{s}$ is subjected to various common signal processing operations such as lossy compression, addition of noise, and attempts to remove the embedded information. The decoder reconstructs $\hat{\mathbf{x}}$ and extracts information $\hat{m}$ after $\mathbf{y}$ is received. We hope that $\hat{m} = m$ under a distortion constraint between $\mathbf{x}$ and $\hat{\mathbf{x}}$, i.e., $D(\mathbf{x}, \hat{\mathbf{x}}) \leq D_{\max}$.

Here, we propose a novel data embedding method which is compatible with the half-pixel based video coding standard such as H.263. There are several data hiding methods published for video coding [8]–[10]. But these methods either can only embed several bits in one frame or are not compatible with the video coding standard. The method we propose is suboptimal in terms of H.263 video compression efficiency. However, it is very suitable for video transmission over error-prone channels in terms of better error recovery.

A. Data Embedding at Half-Pixel Motion Estimation

In the H.263 encoder, for every inter-mode coded MB, an integer-pixel MV is found within a search window by motion estimation, as illustrated in Fig. 2, where the block marked with dashed line at integer-pixel A in frame K is the motion prediction of the current MB marked with solid line in frame $K + 1$. Then half-pixel based motion estimation refinement [5] is found by looking for the MV (D_x, D_y) with minimal sum of absolute difference (SAD) among half-pixel locations 1–8 and A, as illustrated in Fig. 3. In our method, we embed two bits data by changing the half-pixel motion estimation as follows.

After Integer-pel motion estimation, MV candidates in H.263 can be classified into four sets S_m, $m \in \{0, 1, 2, 3\}$ which are specified by their horizontal and vertical components. For ex-
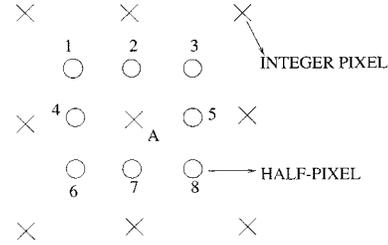
Fig. 3.   Half-pixel motion prediction in H.263: The half-pixel MV is found by looking for the minimum sum of absolute difference (SAD) among the half-pixels 1–8 and integer pixel A.

TABLE I
THE MOTION VECTOR CLASSIFICATION IN FIG. 3 USED FOR DATA EMBEDDING, WHERE I STANDS FOR INTEGER PIXEL LOCATION AND H FOR HALF-PIXEL LOCATION

Set	$b_1 b_0$	(d_x, d_y)	Pixel locations
S_0	00	(I , I)	A
S_1	01	(I , H)	2,7
S_2	10	(H , I)	4,5
S_3	11	(H , H)	1,3,6,8

ample, in Fig. 3, the half-pel locations 1–8 and integer-pel location A can be classified into four sets, as shown in Table I, where I stands for integer pixel and H for half-pixel. Then two bits $b_1 b_0$ can be embedded into an MV by specifying the set S_m that the MV will be in using $m = 2b_1 + b_0$. The final MV (D_x, D_y) of one MB in frame k is then determined by looking for the best MV candidates in S_m, i.e.

$$
\begin{aligned}
(D_x, D_y) & \\
= \arg \min_{(d_x, d_y) \in S_m} & \sum_{i=0}^{15} \sum_{j=0}^{15} \left| f^k(i, j) - f^{k-1}(i + d_x, j + d_y) \right|
\end{aligned}
$$

$$(1)$$

where $f^k(x, y)$ is pixel value at spatial location (x, y) in frame k, and (d_x, d_y) is MV candidates in set S_m.

In other words, we find the MV not from the location with minimum SAD from half-pel locations 1–8 and integer-pel A, but from a MV set specified by the two bits to be embedded. The residual block will be DCT transformed and quantized as normal.

For QCIF video of resolution 176×144, there are 99 MBs of size 16×16 in one frame, which means we can embed at most $99 \times 2 = 198$ bits per frame. In fact, because some MBs in a frame are not coded (use MBs in the previous frame at the same spatial location), or are INTRA-mode coded, the number of inter-mode coded MBs is less than 99. As a result, the maximal number of bits that we can embed in a inter-frame is equal to two times the number of inter-mode coded MBs.

The extraction of the embedded data at decoder is straightforward. For an inter-mode coded MB, its MV is decoded first. Then the two embedded bits $b_1 b_0$ can be extracted from the index of the set that this MV belongs to according to Table II.

The proposed data embedding scheme is very easy to implement and the resulted bitstream is compatible with the half-pixel based video coding standards such as H.263 and MPEG-2.

TABLE II
MOTION VECTOR PROTECTION THROUGH DATA EMBEDDING

	$(x,y)_{1,1}$	$(x,y)_{1,2}$	$\dots$	$(x,y)_{1,11}$
	$(x,y)_{2,1}$	$(x,y)_{2,2}$	$\dots$	$(x,y)_{2,11}$
	$\vdots$	$\vdots$	$\dots$	$\vdots$
$\oplus$	$(x,y)_{9,1}$	$(x,y)_{9,2}$	$\dots$	$(x,y)_{9,11}$
		Parity Code		

B. Performance of Data Embedding at Fractional-Pel Motion Vector

It is desirable to analyze how the performance of video coding will be affected because of data embedding at fractional-pel MV.

A generalized hybrid video coding scheme is shown in Fig. 4. The generalized hybrid coder combines a DPCM algorithm (differential pulse code modulation) along the motion-trajectory of the picture contents with a two-dimensional(2-D) spatial intraframe encoder. The prediction value $\hat{s}$ takes into account a displacement estimate $(\hat{d}_x, \hat{d}_y)$ that is obtained by motion estimation based on the signal s. Since s is not available at the receiver, $(\hat{d}_x, \hat{d}_y)$ has to be transmitted. The prediction error e is encoded by the intra-frame source encoder that eliminates spatial redundancy from signal e. At the receiver, an intraframe source decoder generates the reconstructed prediction error e', which differs from e by some quantization noise. The transmitter contains a replication of the receiver in order to be able to generate the same prediction values s' at the receiver. Regardless of the specific implementation of the intraframe source encoder, the accuracy of the displacement estimate has an important influence on the minimum bit rate that can be achieved by a hybrid coder for a given distortion. In [11], the rate-distortion function for a hybrid coding scheme related to that in Fig. 4 has been presented using an intraframe DCT and Max quantization [11]. Girod derived and evaluated performance bounds for the generalized hybrid coder with motion-compensating prediction (MCP) [12], [13].

We first present some notation and theoretical results from [12], [13]. The motion-compensated prediction error signal is only weakly correlated spatially [12], [14], [15]. This suggests that the prediction error variance

$$\sigma_e^2 = E\{e^2\} - E^2\{e\} \tag{2}$$

can be used to evaluate the performance of MCP. The prediction signal $\hat{s}(x,y)$ is obtained from the samples of the reconstructed previous frame $r(x,y)$, which is only available at sampling location $(x_s, y_s) \in \Pi$, where Π is the set of sampling positions. We assume that the input video signal $s(x,y)$ has a power spectral density $\Phi_{ss}(\omega_x, \omega_y)$, and that the current frame can be predicted upto some residual noise $n(x,y)$ of power spectral density $\Phi_{nn}(\omega_x, \omega_y)$ by translating the reconstructed previous frame $r(x,y)$ by the true displacement (d_x, d_y). If we assume that the noise $n(x,y)$, the signal $s(x,y)$, and the displacement estimation error

$$\begin{pmatrix} \Delta d_x \\ \Delta d_y \end{pmatrix} = \begin{pmatrix} d_x \\ d_y \end{pmatrix} - \begin{pmatrix} \hat{d}_x \\ \hat{d}_y \end{pmatrix} \tag{3}$$
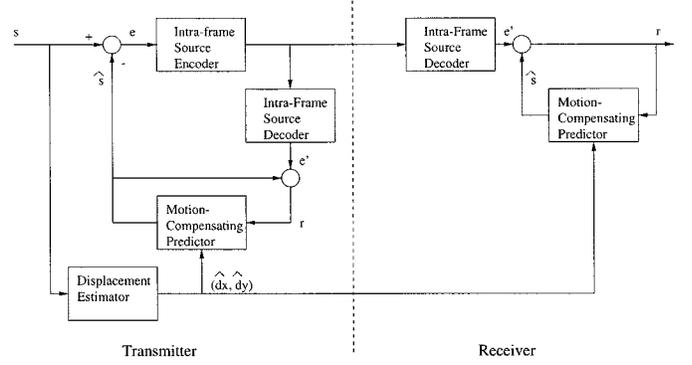
Fig. 4. Block diagram of a generalized hybrid coding scheme with motion-compensating prediction (MCP).

are statistically independent, then the power spectral density of the prediction error e is given by [12]

$$\begin{aligned} \Phi_{ee}(\omega_x, \omega_y) =& \Phi_{ss}(\omega_x, \omega_y) \\ & \cdot \left(1 + |F(\omega_x, \omega_y)|^2 \right. \\ & \left. - 2\mathcal{R}\{F(\omega_x, \omega_y)P(\omega_x, \omega_y)\}\right) \\ & + \Phi_{nn}(\omega_x, \omega_y)|F(\omega_x, \omega_y)|^2 \end{aligned} \tag{4}$$

where $F(w_x, w_y)$ is the Fourier transform of the combined spatial filtering characteristic of the predictor, $\mathcal{R}\{\cdot\}$ denotes the real part of a complex number, and $P(\omega_x, \omega_y)$ is the band-limited 2-D Fourier transform of the continuous probability density function (pdf) $p(\Delta d_x, \Delta d_y)$ of the displacement error $(\Delta d_x, \Delta d_y)$

$$\begin{aligned} P(\omega_x, \omega_y) = \int\int_{x,y} p(x,y)e^{-j\omega_x x/X - j\omega_y y/Y} dx dy \\ \forall |\omega_x| < \pi, |\omega_y| < \pi. \end{aligned} \tag{5}$$

Equation (4) allows us to study the influence of the displacement error pdf on the prediction error variance (2), which can be calculated on the basis of Parseval's relation

$$\sigma_e^2 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Phi_{ee}(\omega_x, \omega_y) d\omega_x d\omega_y. \tag{6}$$

It is pointed in [13] that the precise shape of the displacement error pdf has hardly any influence on the variance of the motion compensated prediction error σ_e^2 as long as the displacement error variance $\sigma_{\Delta d}^2$ does not change.

For a perfect displacement estimator that always estimates the true displacement, the displacement error $(\Delta d_x, \Delta d_y)$ is entirely due to rounding. The displacement error will be uniformly distributed between $\pm(1/2)\beta X$ and $\pm(1/2)\beta Y$, where $\beta = 1$ for integer-pel accuracy, $\beta = 1/2$ for half-pel accuracy, etc. For a sampling grid with balanced horizontal and vertical resolution, $X = Y$, the minimum displacement error variance in moving areas is

$$\begin{aligned} \sigma_{\Delta d}^2 =& E\left[(\Delta d_x)^2 + (\Delta d_y)^2\right] \\ =& \frac{1}{\beta X} \frac{1}{\beta Y} \int_{-\beta X/2}^{\beta X/2} \int_{-\beta Y/2}^{\beta Y/2} (x^2 + y^2) d_x d_y \\ =& \frac{(\beta X)^2}{6}. \end{aligned} \tag{7}$$

Without loss of generality, assume that the half-pel motion estimation based MV after rounding is at integer-pel A as shown in Fig. 3. The displacement error $(\Delta d'_x, \Delta d'_y)$ of the proposed data embedding scheme is due to the rounding to one of those locations specified by the two bits to be embedded with minimal prediction error. Assuming the bits to be embedded are i.i.d. process with $p(b_n = 0) = p(b_n = 1) = 1/2$, $n = 0, 1, \ldots$, we analyze the variance of displacement estimation error with data embedding in four cases under the assumption that the prediction distortion function monotonically increases as we move away from the optimal MV location along any direction in each of the four quadrants [11].

1) If the two bits to be embedded are "00" with probability 1/4, the MV is at A. The corresponding displacement error variance is

$$\sigma_{00}^2 = \frac{(\beta X)^2}{6}. \tag{8}$$

2) If the two bits to be embedded are "01" with probability 1/4, the MV is selected from locations 2 and 7 with smaller prediction error. As mentioned above, we assume that the location with smaller prediction error is always closer to the optimal MV location than other locations. The displacement error is now uniformally distributed as $\Delta d_x \in [-\beta X/2, -\beta X/2]$, $\Delta d_y \in [-\beta Y, -\beta Y/2] \cup [\beta Y/2, \beta Y]$. The corresponding displacement error variance is

$$\sigma_{01}^2 = \frac{1}{\beta X} \frac{1}{\beta Y} 2 \int_{\beta Y/2}^{\beta Y} \int_{-\beta X/2}^{\beta X/2} (x^2 + y^2) dx dy$$
$$= \frac{2(\beta X)^2}{3}. \tag{9}$$

3) Similarly, if the two bits to be embedded are "10" with probability 1/4, the MV is selected from location 4 and 5 with smaller prediction error. The displacement error is now uniformally distributed as $\Delta d_x \in [-\beta X, -\beta X/2] \cup [\beta X/2, \beta X]$, $\Delta d_y \in [-\beta Y/2, \beta Y/2]$. The corresponding displacement error variance is

$$\sigma_{10}^2 = \frac{1}{\beta X} \frac{1}{\beta Y} 2 \int_{\beta X/2}^{\beta X} \int_{-\beta Y/2}^{\beta Y/2} (x^2 + y^2) dx dy$$
$$= \frac{2(\beta X)^2}{3}. \tag{10}$$

4) Finally, if the two bits to be embedded are "11" with probability 1/4, the MV is selected from location 1, 3, 6, and 8 with minimal prediction error. The displacement error is now uniformally distributed as $\Delta d_x \in [-\beta X, -\beta X/2] \cup [\beta X/2, \beta X]$, $\Delta d_y \in [-\beta Y, -\beta Y/2] \cup [\beta Y/2, \beta Y]$. The corresponding displacement error variance is

$$\sigma_{11}^2 = \frac{1}{\beta X} \frac{1}{\beta Y} 2 \int_{\beta X/2}^{\beta X} 2 \int_{\beta Y/2}^{\beta Y} (x^2 + y^2) dx dy$$
$$= \frac{7(\beta X)^2}{6}. \tag{11}$$

And the displacement error variance in moving area is

$$\sigma_{\Delta d, \text{EB}}^2 = \frac{1}{4} \sigma_{00}^2 + \frac{1}{4} \sigma_{01}^2 + \frac{1}{4} \sigma_{10}^2 + \frac{1}{4} \sigma_{11}^2$$
$$= \frac{4(\beta X)^2}{6}$$
$$= 4 \sigma_{\Delta d}^2 \tag{12}$$

which shows that data embedding at fractional-pel MV increases the variance of motion estimation error by four times, i.e., the variance of motion compensated prediction error using data embedded motion estimation at βX accuracy is equivalent to the variance of motion compensated prediction error without data embedding at $2\beta X$ motion estimation accuracy.

III. A DATA EMBEDDED VIDEO CODING SCHEME

In this section, we present an error resilient video coding scheme using the proposed data embedding scheme. Now that we can embed some number of bits in a frame, the question now is what kind of information to be embedded. How to efficiently use the information bits for embedding depends on the application environments. In this paper, we protect the important information in video coding such as MV and coding modes {*Inter, Skip, Intra*} for each MB. In the H.263 encoder, each GOB has a starting code for synchronization when errors occur during decoding.

A. Motion Vector and Coding Modes Protection

For QCIF video of size 176×144, there are nine rows of GOB and each GOB has 11 MBs of size 16×16. In inter-frame k, the MV $(D_x, D_y)_{m,l}$ for each MB at location (m, l), $m \in \{1, 2, \ldots, 9\}$, $l \in \{1, 2 \ldots, 11\}$ are differentially Huffman coded as in the H.263 standard. The binary Huffman code for $(D_x, D_y)_{m,l}$ is denoted as $(x, y)_{m,l}$; we call it differentially Huffman coded motion vector (DHC_MV) to distinguish it from the true MV $(D_x, D_y)_{m,l}$. We add one bit prefix for each DHC_MV $(x, y)_{m,l}$ to indicate whether or not this MB has MV. If an MB does not have MV, one more bit is used to specify the coding mode of this MB which can be coded in either *Skip*-mode or *Intra*-mode. The DHC_MVs $(x, y)_{m,l}$, $l = 1, \ldots, 11$ in GOB m are concatenated to a bit-string, and the nine bit-strings are arranged row by row, as shown in Table II. Because the lengths of bit-strings are different, the shorter rows are padded with bit zeros. The rows of bit-strings are then modulo-2 addition coded across rows. The generated parity bit-string will be embedded in the MVs of inter-mode coded MBs in the following frame $k + 1$. The MVs in frame $k + 1$ will undergo the same data embedding process as frame k does, and so on. It should be noted that the number of bits which can be embedded in a frame is dynamic and is determined by the number of inter-mode coded MBs in each frame. In case that the number of bits to be embedded is more than the number of bits one frame can embed, we just drop the additional parity bits because of consistency requirement for data embedding in consecutive frames, i.e., parity bits for frame k can only be embedded in frame $k + 1$ because parity

bits for frame $k + 1$ need to be embedded in frame $k + 2$ and so on. In most cases, the parity code can be embedded in one frame completely which will be shown in later simulations.

B. Error Detection

Before any error-concealment technique can be applied at the decoder, it is necessary to find out whether and where a transmission error has occurred. The techniques developed for this purpose can be divided into two categories: (a) those performed at transport coder/decoder and (b) those performed at video decoder [1].

One way to perform error detection at the transport coder is by adding header information. For example, in packet-based video transmission, the output of the video encoder is packetized into packets, each of which contains a header and payload field. The header contains a sequence number field that is consecutive for sequentially transmitted packets. At the transport decoder, the sequence number can be used for packet-loss detection. Another method for error detection at the transport layer is to use FEC. In this method, error-correction encoding is applied to segments of the output bitstream of the encoder. At the decoder, error-correction decoding is employed to detect and possibly correct some errors.

To accomplish error detection at the video decoder, characteristics of natural video signals have also been exploited. For H.263 coding standard, the set of error detection conditions includes the following.

1) An invalid codeword for MB address increment, MB type, coded block pattern, motion code, or DCT coefficients is found.
2) The total number of decoded DCT coefficients within a block is larger than 64.
3) The total number of decoded MBs within a GOB/slice is not equal to the expected number of MBs within a slice.

Details about error detection techniques can be found in [1] and [16].

C. Error Concealment Using Embedded Information

If there is no error, the decoder just works as a standard H.263 decoder since the bitstream is compatible with H.263 standard. The decoding process at decoder in error free case can be described as

$$f(k) = \mathrm{CLP}[\mathrm{MC}_k(f(k-1)) + r(k)] \qquad (13)$$

where

$f(k)$ reconstructed frame k at the decoder;

$\mathrm{CLP}(\cdot)$ clipping function which operates after summation of prediction and reconstructed prediction errors on resulting pixel values to keep them between 0 and 255;

$\mathrm{MC}_k(\cdot)$ motion compensation operation based on frame $k - 1$ using MVs of frame k;

$r(k)$ residual data of frame k.

If one GOB_k^m, $m \in \{1, 2, \ldots, 9\}$ in frame k is corrupted, the parity code of DHC_MVs in frame k can be extracted from MVs

TABLE III
THE PROBABILITY OF USING THE PROPOSED ERROR CONCEALMENT SCHEME WITH DATA EMBEDDING WHEN ERRORS OCCUR

p_{GOB}	$D = 1$	$D = 2$	$D = 5$	$D = 10$
0.01	0.8772	0.9501	0.9567	0.9567
0.05	0.5089	0.6608	0.7237	0.7254
0.10	0.2450	0.3399	0.3965	0.4000
0.20	0.0468	0.0610	0.0669	0.0671

in frame $k+1$ after frame $k+1$ is decoded. Then the DHC_MVs of the corrupted GOB m can be recovered by modulo-2 addition using the extracted parity code and the correctly received DHC_MVs in GOB_k^j, $j \neq m$, and $j \in \{1, 2, \ldots, 9\}$ (see Table II), but the residual data r_k^m of the corrupted GOB cannot be recovered. Now the video decoding process is as follows.

1) If the MV of an MB can be recovered, the video decoding process becomes

$$\hat{f}(k) = \mathrm{CLP}[\mathrm{MC}_k(f(k-1)) + \hat{r}(k)] \qquad (14)$$
$$\hat{f}(k+1) = \mathrm{CLP}[\mathrm{MC}_{k+1}(\hat{f}(k)) + r(k+1)] \qquad (15)$$

where $\hat{r}(k) = r_k - r_k^m$ is the residual data of frame k in which the area corresponding to the GOB m is replaced by zero. $\hat{f}(k)$ is the reconstructed frame k using data embedding scheme.

2) If the corrupted MB is coded in *Intra* mode, spatial error concealment techniques as proposed in [17], [18] can be used.

3) If the corrupted MB is coded in *Skip*-mode, the current block is recovered by simply copying the MB at the same spatial location in the previous reconstructed frame.

It can be noted that by embedding the MV and coding mode information, the error concealment at decoder will have more useful information such as MV and coding mode than conventional video coding schemes.

In general, assuming at most one GOB data is corrupted in one frame and the round-trip delay between the encoder and decoder is D frame intervals (we assume no other errors occur during this period for simplicity), the error recovery scheme works as follows.

a) When errors occur at GOB m, $m \in \{1, 2, \ldots, 9\}$ in frame k, the decoder sends a negative acknowledgment (NACK) message to encoder with the location m of the damaged GOB. The Huffman codes of those MVs in GOB j, $j \neq m$, $j \in \{1, 2, \ldots, 9\}$ are arranged row by row, as described in Section III-A.

b) The parity code of the DHC_MVs in frame k can be extracted from the MVs in frame $k + 1$ using the proposed data embedding scheme after frame $k + 1$ is decoded. Then the DHC_MVs of the corrupted GOB m can be recovered using the parity code and other correctly received DHC_MVs by modulo-2 addition. Then the MVs in frame k can be Huffman decoded.

c) The motion compensation proceeds as in (14) and (15) except that the residual data of the corrupted GOBs r_k^m is replaced by zero, the reconstructed frames are $\hat{f}(n)$, $n = k, k+1, \ldots, k + D - 1$.
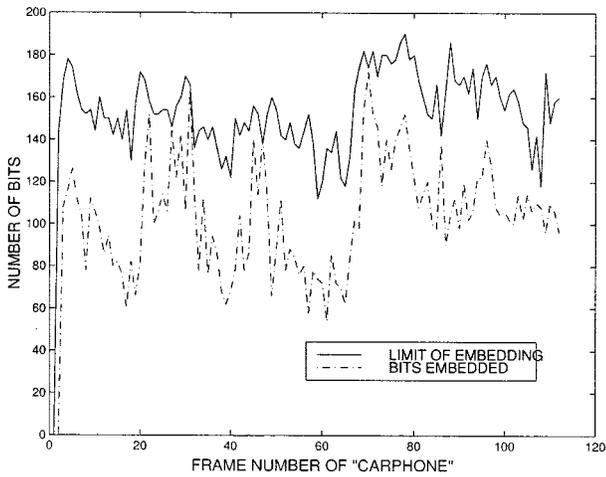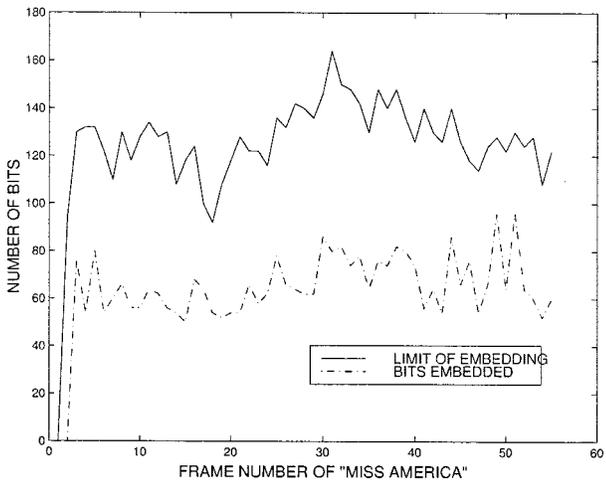
Fig. 5. The number of bits that can be embedded and actual number embedded per frame, (a) is from the sequence *Car phone* which is highly active in motion, and (b) is from the sequence *Miss America* which has relatively smaller motion area.
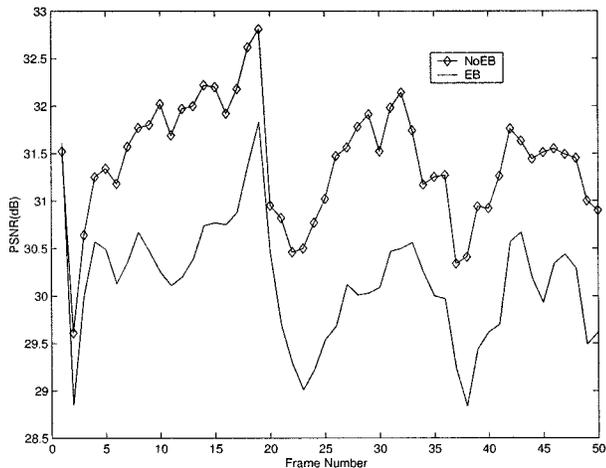
Fig. 6. PSNR comparison of video coding with and without data embedding for *Foreman*.
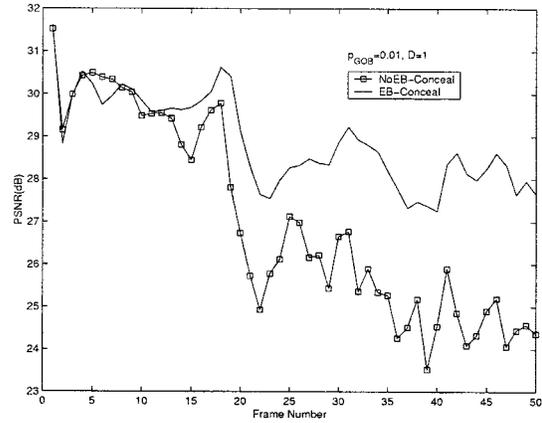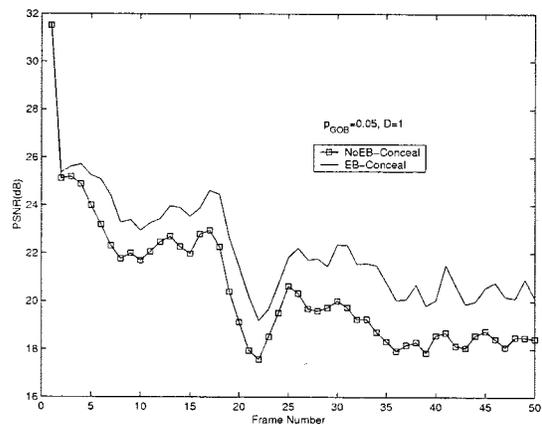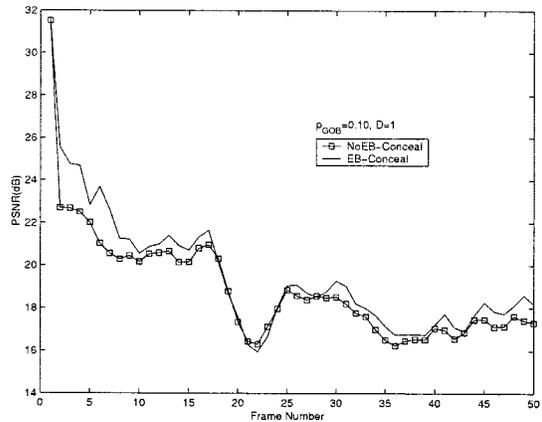
Fig. 7. PSNR comparison of video transmission over error-prone channels for *Foreman*. (a) $p_{\mathrm{GOB}} = 0.01$, (b) $p_{\mathrm{GOB}} = 0.05$, (c) $p_{\mathrm{GOB}} = 0.10$.

If INTRA-mode updating method is used as in [3], then the error propagation will be fully stopped when the frame $k + D$ arrives. If retransmission of residual data method is used, the MVs and the residual data in the affected areas from frame k to frame $k + D - 1$ have to be stored at the decoder. After the residual data r_k^m of the corrupted GOB in frame k arrives with frame $k + D$, the frame k can be losslessly recovered using (13), and the accumulated motion compensations are re-proceeded
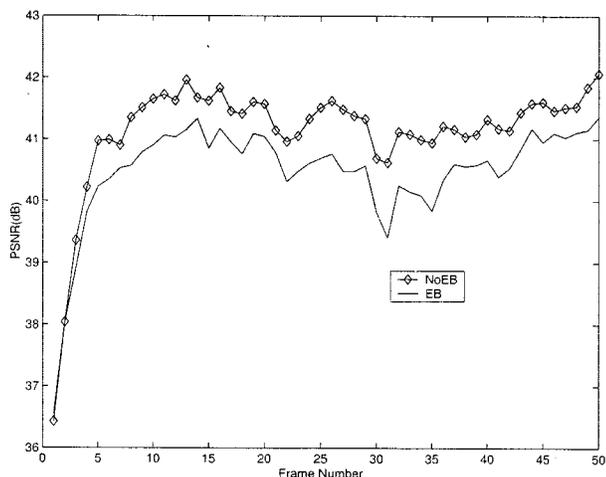
Fig. 8. PSNR comparison of video coding with and without data embedding for *Miss America*.

within the affected area to get the losslessly recovered frame $f(n)$, $n = k, k+1, \ldots, k+D-1$. The losslessly recovered frame $f(k+D-1)$ is used to reconstruct frame $f(k+D)$, and the error propagation is fully stopped.

If more than one GOBs are corrupted in one frame, the proposed MV recovery using embedding cannot be employed because of the limited embedding capacity. It is desirable to investigate that how possible the proposed error concealment scheme can be used in practical situations. Assume that the GOBs go through a memoryless lossy channel with GOB error probability p_{GOB}, and the decoder has a buffer of D frames. We study the possibility $P(\text{EB}|\text{Error})$ of using the proposed error concealment scheme with embedding when errors occur in an inter-frame. The probability that errors occur in any frame k is $1 - (1 - p_{\text{GOB}})^9$ for QCIF. To use the proposed scheme, there has to be one frame l, $l = k+1, \ldots, k+D$ received without error, and each of the frames $k, \ldots, l-1$ has only one GOB error. The probability of no GOB error in one frame is $p_0 = (1 - p_{\text{GOB}})^9$, the probability of one GOB error in one frame is $p_1 = \binom{9}{1} p_{\text{GOB}} (1 - p_{\text{GOB}})^8$. We can verify that

$$P(\text{EB}|\text{Error}) = \frac{\sum_{i=1}^{D-1} p_1^i p_0}{1 - (1 - p_{\text{GOB}})^9}. \tag{16}$$

Table III lists the values of $P(\text{EB}|\text{Error})$ for several combination of p_{GOB} and frame buffer size D. It can be observed that the proposed scheme is more effective when the probability of GOB error is lower than 10% and the frame buffer size larger than two does not help too much.

IV. SIMULATIONS

The simulations have been done using the base-mode H.263 on QCIF sequences *Car phone*, *Foreman*, and *Miss America* that are coded at 10 frames per second with bit-rate 48 kb/s. In the simulation, no ARQ or INTRA-mode updating method is used for clarity.

Fig. 5 shows the number of bits that can be embedded versus the actual number of parity bits embedded in the inter-frames. It
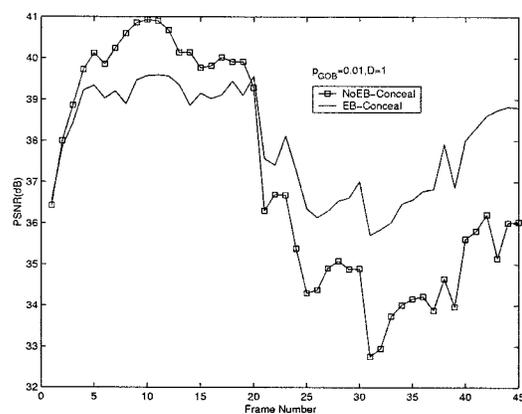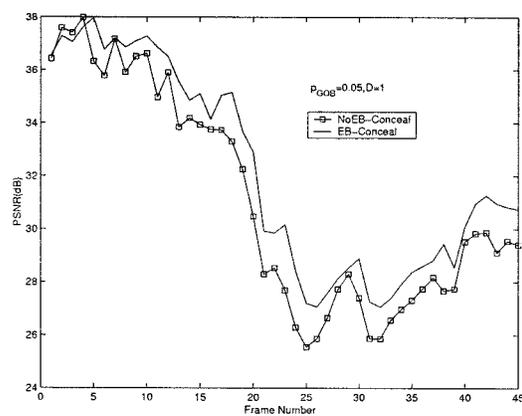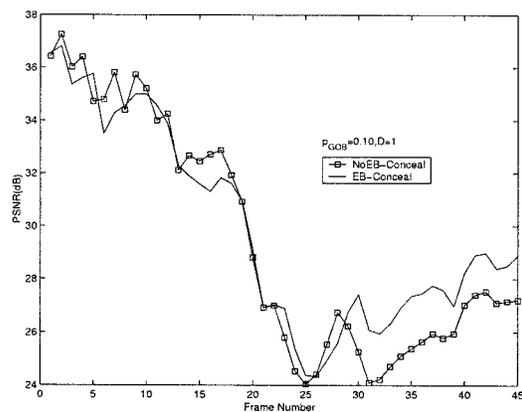
Fig. 9. PSNR comparison of video transmission over error-prone channels for *Miss America*. (a) $p_{\text{GOB}} = 0.01$; (b) $p_{\text{GOB}} = 0.05$; and (c) $p_{\text{GOB}} = 0.10$.

can be seen that in most situations the parity bits can be totally embedded in the next frame.

In the simulations, we use a memoryless lossy channel where each GOB has lossy probability p_{GOB}. The average peak signal-to-noise ratio (PSNR) of 50 runs is used for comparison. For video coding without data embedding, the MVs of the corrupted GOB are simply replaced by the MVs of MBs above
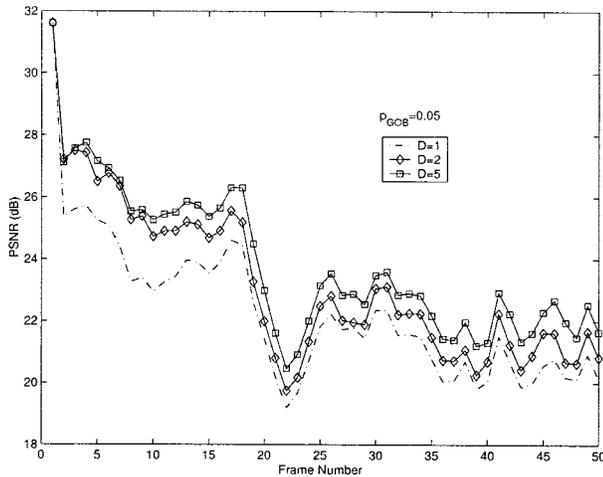
Fig. 10. PSNR comparison of error concealment with data embedding for *Foreman* at different number of frame buffer $D = 1, 2, 5$.

Fig. 11. Comparison between (a) video coding without data embedding and no error; (b) video coding with data embedding and no error; (c) MV recovery using data embedding; and (d) MV estimation from MVs above the damaged GOB. The third GOB is totally corrupted at frame 27.

Fig. 12. Error propagation effects at six frames after errors occur. The comparison between (a) MV recovery using data embedding and (b) MV estimation from MVs above the damaged GOB.

the damaged GOB for error concealment. In Fig. 6, the luminance PSNR of video coding for *Foreman* without embedding (denoted as "NoEB") is about 1 dB better than data embedding scheme (denoted as "EB") for noiseless channel. Fig. 7 shows the PSNR of decoded video after the video bitstreams are transmitted through memoryless lossy channel with $p_{\text{GOB}} = 0.01, 0.05, 0.10$, respectively. In all three cases, the number of frame buffer for error concealment is $D = 1$. It can be noted that the error concealment using proposed data embedding has better performance than conventional concealment schemes without embedding when $p_{\text{GOB}} = 0.01$ and 0.05. However, when the $p_{\text{GOB}} = 0.10$, the proposed scheme has no obvious advantages and the error propagation effects dominate the performance rather than data embedding. The performance is also compared for *Miss America*, as shown in Fig. 8 for noiseless channel and in Fig. 9 for error-prone channels. In Fig. 9(a), the advantage of using data embedding for error concealment is not obvious in the first 20 frames because the motion activity is low. But when the error propagation continues, the simulations show that the loss of residual block data has less significant effect on error propagation than the loss of MV does for both sequences with different motion activity. We also studied the performance of proposed error concealment scheme at different number of frame buffer $D = 1, 2, 5$, as shown in Fig. 10 using *Foreman*. The results verify that it does not help the proposed scheme too much to have more frame buffers at decoder, which follows the numerical analysis in (16) and Table III. In Fig. 11, the third GOB at frame 27 of *Car Phone* is corrupted. The image qualities are compared under four cases:(a) video coding without data embedding and no error, (b) video coding with embedding and no error, (c) MV recovery using data embedding, and (d) MVs of the corrupted GOB are simply replaced by the MVs of MBs above the damaged GOB. The visual quality difference between (a) and (b) is not distinguishable by human visual system. Even though more advanced MV estimation techniques might be used in case (d), the exact MV recovery using data embedding in case (c) will always be better than MV estimation in case (d), especially in highly

active motion area. Fig. 12 shows the error propagation effect at frame 32 which is six frames after the errors occur. When using the data embedding method proposed here, the MVs are exactly recovered and the error propagation because of the loss of the residual data of the corrupted GOB is not serious, which we can see from Fig. 12(a) that the image quality is still acceptable. However, the image quality of Fig. 12(b) which uses MV estimation has obvious spatial error propagation.

V. Conclusions

A data embedded video coding scheme is proposed which can embed two bits at each position of half-pixel MV for video coding standards such as H.263 and MPEG-2. We have shown that the data embedded video coding scheme degenerates the coding performance of half-pel motion compensation back to integer-pel motion compensation. Furthermore, based on the proposed data embedding scheme, we also presented an error resilient video coding method to deal with error propagation of video transmission over error-prone channels. Even though the data embedded video coding scheme has lower coding performance compared with video coding without data embedding, simulations show that the proposed method provides better video quality when errors occur by exactly recovering the MVs of damaged GOB in error-prone channels when the GOB error probability is lower than 0.10. The proposed method has advantages in terms of better error recovery performance and compatibility with half-pixel motion estimation based video coding standards.

References

[1] Y. Wang and Q. Zhu, "Error control and concealment for video communication," *Proc. IEEE*, pp. 974–997, May 1998.
[2] M. Khansari and V. Bhaskaran, "A low-complexity error-resilient H.263 coder," in *Proc. IEEE ICASSP*, 1997, pp. 2737–2740.
[3] E. Steinbach, N. Farber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 872–881, Dec. 1997.
[4] N. Farber, B. Girod, and J. Villasenor, "Extensions of ITU-T recommendation H.324 for error-resilient video transmission," *IEEE Commun. Mag.*, pp. 120–128, June 1998.
[5] ITU H.263, Vers. 2 *Video Coding for Low Bitrate Communication*, Jan. 1998.
[6] M. B. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking techniques," *Proc. IEEE*, vol. 86, pp. 1064–1087, June 1998.
[7] B. Chen and C. W. Wornell, "Digital watermarking and information embedding using dither modulation," in *Proc. IEEE 2nd Workshop Multimedia Signal Processing*, 1998, pp. 273–278.
[8] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," in *Proc. IEEE ICASSP*, 1997, pp. 2621–2624.
[9] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Data hiding for video in video," in *Proc. IEEE ICIP*, 1997, pp. 676–679.
[10] J. Lacy, S. Quackenbush, A. R. Reibman, D. Shur, and J. H. Snyder, "On combining watermarking with perceptual coding," in *Proc. IEEE ICASSP*, 1998, pp. 3725–3728.
[11] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. COM-29, pp. 1799–1804, Dec. 1981.
[12] B. Girod, "The efficiency of motion-compensation prediction for hybrid coding of video sequence," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1140–1154, Aug. 1987.
[13] ——, "Motion-compensation prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–611, Apr. 1993.
[14] M. Kaneko, V. Hatori, and A. Kloike, "Improvements of transform coding algorithm for motion-compensated interframe prediction errors-DCT/SQ coding," *IEEE. J. Select. Areas Commun.*, vol. SAC-5, pp. 1068–1078, Aug. 1987.
[15] M. Gilge, "A high quality videophone coder using hierarchical motion estimation and structure coding of the prediction error," in *Proc. SPIE Conf. Visual Communications and Image Processing*, vol. 1001, Nov. 1988, pp. 864–874.
[16] H. Shyu and J. Leou, "Detection and concealment of transmission errors in MPEG-2 images—A genetic algorithm approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 937–948, Sept. 1999.
[17] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex sets," *IEEE Trans. Image Processing*, vol. 4, pp. 470–477, Apr. 1995.
[18] W. Kwok and H. Sun, "Multi-directional interpolation for spatial error concealment," *IEEE Trans. Consumer Electron.*, vol. 39, pp. 455–460, Aug. 1993.

Jie Song received the B.S. degree from Beijing University, Beijing, China, in 1990, the M.S. degree from Beijing University of Posts and Telecommunications in 1993, and the Ph.D. degree from the University of Maryland, College Park, in 2000, all in electrical engineering.

From 1993 to 1996, he was a Lecturer and Researcher in the Information Engineering Department at Beijing University of Posts and Telecommunications. From 1997 to 1999, he was a part-time Consultant of multimedia technologies in Oddyssey Technologies, Inc., Jessup, MD, where he worked on H.323/H.324 videophone, portable multimedia terminal design, and multichannel video capturing systems. Since 2000, he has been working on research, design, and implementation for broad-band communication systems in the Microelectronics Group, Lucent Technologies (now Agere Systems), Holmdel, NJ. His research interests include signal processing for digital communications and multimedia communications.

K. J. Ray Liu (S'86–M'86–SM'93) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., and the Ph.D. degree from the University of California, Los Angeles, both in electrical engineering.

He is currently a Professor in the Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park. He was a Visiting Professor at Stanford Univesity, Stanford, CA; IT University, Copenhagen, Denmark; and Hong Kong Polytechnic University, Hong Kong. His research interests span broad aspects of signal processing algorithms and architectures; image/video compression, coding, and processing; wireless communications; security; and medical and biomedical technology. He has published over 200 papers, more than 60 of which are in refereed journals.

Dr. Liu has received numerous awards including the 1994 National Science Foundation Young Investigator, the IEEE Signal Processing Society's 1993 Senior Award (Best Paper Award), the IEEE Vehicular Technology Conference Best Paper Award, Amsterdam, 1999, and the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education and the Outstanding Systems Engineering Faculty Award in 1996 in recognition of outstanding contributions in interdisciplinary research, both from the University of Maryland. He is Editor-in-Chief of *EURASIP Journal on Applied Signal Processing*, and has been an Associate Editor of IEEE Transactions on Signal Processing, a Guest Editor of Special Issues on Multimedia Signal Processing of Proceedings of the IEEE, a Guest Editor of a Special Issue on Signal Processing for Wireless Communications of IEEE Journal of Selected Areas in Communications, a Guest Editor of a Special Issue on Multimedia Communications over Networks of IEEE Signal Processing Magazine, a Guest Editor of a Special Issue on Multimedia over IP of IEEE Transactions on Multimedia, and an Editor of the *Journal of VLSI Signal Processing Systems*. He currently serves as the Chair of Multimedia Signal Processing Technical Committee of IEEE Signal Processing Society and the Series Editor of the Marcel Dekker series on Signal Processing and Communications.