

# Joint Source-Channel Coding for Wireless Object-Based Video Communications Utilizing Data Hiding

Haohong Wang, *Member, IEEE*, Sotirios A. Tsaftaris, *Student Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

**Abstract**—In recent years, joint source-channel coding for multimedia communications has gained increased popularity. However, very limited work has been conducted to address the problem of joint source-channel coding for object-based video. In this paper, we propose a data hiding scheme that improves the error resilience of object-based video by adaptively embedding the shape and motion information into the texture data. Within a rate-distortion theoretical framework, the source coding, channel coding, data embedding, and decoder error concealment are jointly optimized based on knowledge of the transmission channel conditions. Our goal is to achieve the best video quality as expressed by the minimum total expected distortion. The optimization problem is solved using Lagrangian relaxation and dynamic programming. The performance of the proposed scheme is tested using simulations of a Rayleigh-fading wireless channel, and the algorithm is implemented based on the MPEG-4 verification model. Experimental results indicate that the proposed hybrid source-channel coding scheme significantly outperforms methods without data hiding or unequal error protection.

**Index Terms**—Data embedding, data hiding, joint source-channel coding, MPEG-4 standard, object-based video, rate-distortion, unequal error protection (UEP), video coding, video communications, wireless channel.

## I. INTRODUCTION

WIRELESS video communication requires advanced video compression techniques due to the limitations of wireless channels. Predictive coding and variable length coding in compression techniques make the coded bitstream very sensitive to channel errors. Automatic repeat request (ARQ) and forward error control coding (FEC) are two major error resilient approaches [1] used by the encoder or sender. However, for real-time video communication, ARQ is not always feasible due to the intolerable delay in retransmission. In this paper, we consider the application of FEC coupled with data hiding to improve the performance of wireless transmissions of object based video.

In recent years, the increasing demand for multimedia applications such as networked video games and interactive digital TV, indicates a growing interest in content-based interactivity.

Manuscript received December 21, 2004; revised August 22, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aria Nosratinia.

H. Wang is with the Marvell Technology Group, Santa Clara, CA 95054 USA (e-mail: haohongw@marvell.com).

S. A. Tsaftaris and A. K. Katsaggelos are with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA (e-mail: stsft@ece.northwestern.edu; aggk@ece.northwestern.edu).

Digital Object Identifier 10.1109/TIP.2006.875194

MPEG-4 [2], [3] has become the first international multimedia communication standard to enable content-based interactivity by supporting object-based video coding. According to the specification the video data are composed of shape, motion, and texture information, which have completely different stochastic characteristics and bit rate allocation. A rate-distortion optimal source-coding scheme was proposed recently for solving the bit allocation problem in object-based video coding [4]. The experimental results indicate that for certain applications, the shape may have a stronger impact on the reconstructed video quality than texture. This result directly motivates the unequal protection of the shape and texture components of the video objects in video encoding and transmission. Unequal protection implies that the important data are 1) given higher priority in allocating bits during source coding, 2) encoded with more redundancy during channel coding, and 3) given higher priority during transmission (for example, it utilizes better quality of service channel in a DiffServ setting, or is allocated higher transmission power [5]). In [6]–[9], unequal error protection (UEP) has been applied to the transmission of MPEG-4 compressed video bit streams over error-prone wireless channels. However, these results are based on pre-encoded video and have not considered video with arbitrarily shaped video objects. So far, joint source-channel coding for object-based video is still an unexplored research area, and very limited work has been reported.

Another class of error resilient source coding methods that enables the decoder to conceal errors upon their detection is data hiding (or data embedding) [10], [11]. Redundant information of important features, such as edge information and motion vectors, is embedded into the coded bitstream for future error concealment purposes [12]–[21]. In [14], redundant information for protecting the motion vectors and coding modes of the current frame is embedded into the motion vectors of the next frame. Although this approach degenerated the coding performance of half-pixel motion compensation back to integer-pixel motion compensation, the overall performance gain is significant because the embedded information can effectively help the decoder recover the motion vectors of the corrupted group of blocks (GOB) and conceal the missing blocks. In [20] and [21], the edge information for an I-frame, the motion vectors, the index of the reference frame, the coding modes, and the error concealment strategy information for P-frames were embedded in the coded bitstream, and the distortion caused by data embedding, quantization, and packet loss during transmission were jointly considered in optimizing the source coding parameters.

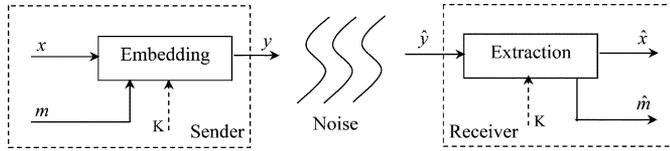


Fig. 1. General data hiding scheme.

Certain efforts on joint channel coding and data hiding have been previously proposed [22]. Such efforts, however, did not consider the optimization over the source coding parameters and the use of adaptive data hiding. On the other hand, joint source-channel coding efforts can be found in literature [23]–[25] that jointly optimize the parameters of separate source and channel coders. However none of them consider data hiding in the framework as an effective error resilient enhancement.

The major contributions of this paper are threefold. First, we propose a novel and general rate-distortion optimal source-channel coding scheme for object-based video communications. The expected distortion seen at the decoder is accurately estimated at the encoder knowing the error concealment strategy employed by the decoder. The parameters in source and channel coding are jointly optimized. Second, a hybrid UEP approach that protects shape and texture differently in both source and channel coding is introduced. Third, data hiding is added in the joint source-channel coding framework. Overall source/channel coding and data hiding are jointly optimized using an efficient and optimal algorithm.

The rest of the paper is organized as follows. Section II provides an overview of data hiding techniques used in video coding. In Section III, the problem of joint source-channel coding for object-based video communications is formulated. Section IV provides the optimal solution to the problem using Lagrangian relaxation and dynamic programming and analytical derivations of expected distortion estimation approaches. Section V presents experimental results and we draw conclusions in the last section.

## II. DATA HIDING FOR VIDEO ENCODING

### A. General Overview

Data hiding [10], [11] has been widely used in a number of applications. In copyright protection or authentication applications [26]–[29], a specific signature information, called the watermark, is (invisibly) hidden, by the owners, inside a host image or video data source before distribution of the media. Later, the signature can be retrieved from a distributed image or video product to prove its ownership or authenticity. In secure transmission applications [30], the secure data (for example, control information) are embedded inside the regular data to be transmitted in a standardized and open form, or over an insecure or readily available medium, such as the Internet, allowing only those authorized at the receiver side to retrieve the additional hidden information. An interesting feature of a certain type of data hiding schemes is the ability of retrieval of the hidden information without the original host.

A general data hiding scheme is shown in Fig. 1. At the sender side, a message  $m$ , optionally encrypted using a key  $K$ , is em-

bedded in the host signal sequence  $x$  to form a signal sequence  $y$  which is transmitted over a noisy channel. The signal  $y$  could be corrupted and become  $\hat{y}$  at the receiver. The decoder extracts the estimated embedded signal  $\hat{m}$ . This process may require the signal  $x$ , and the key  $K$ . In error recovery and similar applications, we should note that a) the key  $K$  may not be necessary, b)  $m$  can be correlated to, or a subset of,  $x$  and c) at the receiver side the signal  $\hat{x}$  is reconstructed from the signals  $\hat{y}$  and  $\hat{m}$ .

There are two popular data embedding and extracting approaches, spread spectrum [27], [31], and quantization index modulation (QIM)[32]. In spread spectrum techniques, the data to be embedded is spread over the spectrum of the host data so that the embedded energy for each host frequency bin is negligible, thus making the embedded data imperceptible. At the receiver side, a correlation-based detection method is usually used.

In our implementation, we use a special case of QIM, the odd-even method [33]. The host data  $x$  are the quantized DCT coefficients of a frame. In odd-even embedding, the data are embedded in the non-zero quantized AC coefficients. If a bit to be embedded is “0,” the quantized AC coefficient is changed to an even number; otherwise, the AC coefficient is changed to an odd number. The data embedding and extraction schemes can be represented by

$$y_i = \text{Embed}(x_i, m_i) = \begin{cases} x_i + 1, & \text{if } x_i > 0 \text{ and } (x_i - m_i) \bmod 2 \neq 0 \\ x_i - 1, & \text{if } x_i < 0 \text{ and } (x_i - m_i) \bmod 2 \neq 0 \\ x_i, & \text{otherwise} \end{cases} \quad (1)$$

$$\text{and } \hat{m}_i = \text{Extract}(\hat{y}_i) = \hat{y}_i \bmod 2 \quad (2)$$

where  $x_i$  and  $\hat{y}_i$  are, respectively, the original and received  $i$ th AC coefficients,  $m_i$  is the embedded bit,  $\hat{m}_i$  is the message bit at the receiver side, and  $i$  is an index of the message bit  $m_i$  in message  $m$ . The output of the mod operations in (1) and (2) is either 0 or 1.

Due to the simplicity of embedding and decoding, the odd-even method can be deployed easily. In addition, although there exist spread spectrum methods with high capacity [34], the odd-even method offers high message capacity with less complexity and increased flexibility [35]. Flexibility is important because it enables the encoder to adaptively decide how much information to embed in each block. On the other hand, spread spectrum approaches are more robust to signal alterations (modifications) and various attacks aiming in removing the message or rendering it undetectable. Such robustness although necessary for a copyright protection application and useful in error recovery applications, we consider the traits of odd-even embedding more suited for our application.

### B. Hiding Shape/Motion in Texture for MPEG-4 Video

In MPEG-4, a data partitioned packetization scheme is applied to increase error resilience. In this scheme, the shape/motion and texture data are packed in the same packet (see Fig. 2) but are separated by a motion marker. Given a safe reception of shape and motion data and an error in the texture

Shape and motion data	Motion Marker	Texture data
-----------------------	---------------	--------------

Fig. 2. Simplified MPEG-4 data partitioned syntax for P-VOPs. (a) PSNR = 39.91 dB. (b) PSNR = 37.31 dB, Rate = 5758 bits. (c) PSNR = 23.08 dB.

TABLE I  
INFORMATION TO BE EMBEDDED INTO TEXTURE DATA

		Number of Bits embedded	Necessary for texture decoding
Shape BAB type	Transparent	1	Y
	Opaque	2	
	Boundary	2+4	
COD		1	Y
Texture mode		1	Y
CBPC		2	Y
Texture Motion vector		10	N
Lossy shape		4-16	N

data partition, this partitioning allows for the concealment of the corrupted texture using the available motion vector. However, since the decoding of the texture partition relies on information stored in the shape/motion partition, such as the texture motion vector, the texture coding mode, and the shape binary alpha block (BAB) coding type, the whole packet will be discarded when the shape data is corrupted, even if the texture data is uncorrupted. Readers are referred to [2] and [3] for more detailed information about MPEG-4 video coding.

In this work, we embed shape and motion information into the texture data to make it self-decodable. Thus, the texture data can be used even if the shape partition is corrupted. In addition, the embedded shape and motion data could help to partially recover the lost shape and motion partition. The data proposed to be embedded are shown in Table I, among which are shape BAB type, COD (a 1-bit flag which signals if a macroblock is coded or not), texture coding mode, and coded block pattern for chrominance (CBPC) are critical for the decoding of texture. The integer motion vectors in the range  $[-16, 16]$ , and a lossy version of the shape (using a lower-resolution of  $4 \times 4$  bitmap to represent the original  $16 \times 16$  BAB) can also be potentially embedded.

The exact determination of the amount of embedded information is a critical factor, since over-embedding could bring intolerable texture distortion. Clearly, the embedding capacity is dependent on the number of non-zero quantized DCT coefficients. To efficiently and adaptively allocate the embedded data within limited coefficient locations, we propose five levels of embedding modes, that is: (0) no embedding; (1) embed critical information only; (2) embed critical information and motion vectors; (3) embed critical information and lossy shape; (4) embed critical information, motion vectors, and lossy shape. The embedding mode itself is also embedded. When the number of non-zero quantized DCT coefficients is zero, the decoder automatically determines that there is no embedded information. It is important to emphasize that the selection of embedding modes is optimally determined based on the source coding parameters, the channel coding parameters, as well as the channel conditions and the error concealment strategy.

In the following, we demonstrate the advantage of the scheme of “hiding shape/motion in texture” with a simple example

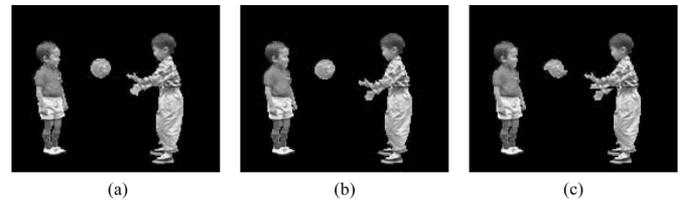


Fig. 3. Reconstructed frames of the “Children” sequence (a) corresponds to the fifteenth frame; (b) and (c) correspond to the sixteenth frame. In (c), the shape of the third packet is corrupted and concealed. (a) PSNR = 37.25 dB, Rate = 5860. (b) PSNR = 23.35 dB, Rate = 5794. (c) PSNR = 37.23 dB, Rate = 5794. (d) PSNR = 23.08 dB. (e) PSNR = 37.19 dB, Rate = 5832. (f) PSNR = 25.54 dB.

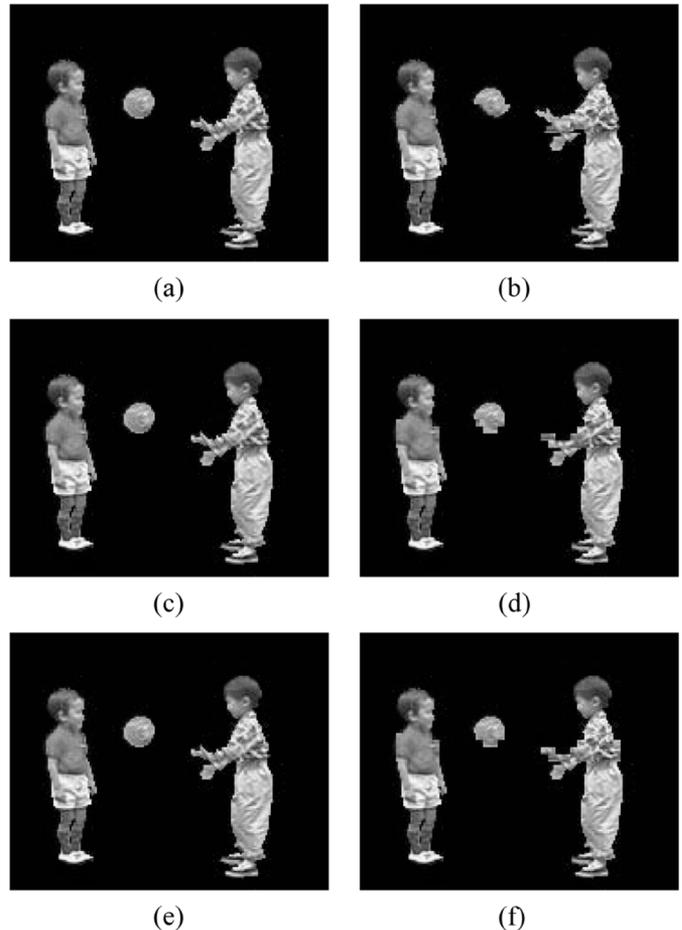


Fig. 4. Example of reconstructing the sixteenth frame of the “Children” sequence using data hiding [(a) and (b), (c) and (d), and (e) and (f) correspond to the case of using data embedding level 2, 3, and 4, respectively]. (a), (c), and (e): No packet corruption has occurred; (b), (d), and (f): the shape partition of the third packet is corrupted and then concealed. (a) R-D curve when  $BER = 10^{-3}$ . (b) R-D curve when  $BER = 10^{-4}$ .

shown in Figs. 3 and 4. For our comparisons, we have used the PSNR metric between the original sequence and the reconstructed sequence at the receiver. We encode the “Children sequence” and consider its transmission over a noisy wireless channel. In Figs. 3(a) and 3(b), frames 15 and 16 are shown, respectively. During transmission, the shape partition of the third packet (third row of macroblocks) of the sixteenth frame is corrupted, and thus the whole packet is discarded. A common concealment strategy is used to recover the lost packet using the motion vectors of the above macroblocks to get predictive data

from the previous frame [shown in Fig. 3(c)]. The proposed data hiding scheme is used with embedding levels 2, 3, and 4, the results of which are shown in Fig. 4(a)–(f), respectively. Clearly, the embedding of shape and motion information only slightly increases the bit rate and causes a very slight reduction in texture quality. However, the embedded information is very helpful in improving the quality of the concealed image. After the embedded lossy shape is extracted and used to recover the lost shape, the PSNR of the reconstructed image is increased by up to 2.5 dB.

### III. PROBLEM FORMULATION

In this paper, we jointly consider source coding, channel coding, data embedding, and error concealment within a rate-distortion optimization framework. By selecting source and channel coding parameters and the level of data embedding, our goal is to minimize the total expected distortion given the frame bit budget, which can also be represented by

$$\text{Minimize } E[D_{\text{tot}}], \text{ Subject to } R \leq R_{\text{budget}} \quad (3)$$

where  $E[D_{\text{tot}}]$  is the expected total distortion for the frame,  $R$  is the actual total bit rate (including source and channel coding rate) for the frame, and  $R_{\text{budget}}$  is the bit budget for the frame.

#### A. System Model

We consider an MPEG-4 compliant object-based video application, where the data hiding approach described in Section II-B is applied. During encoding, the video object plane (VOP) is divided into  $16 \times 16$  macro blocks, which are numbered in scan line order and divided into packets. In our implementation each packet consists of a row of data and is independently decodable. Each packet contains the shape and texture partitions uniquely separated. This guarantees that an error might only affect the decoding of a single partition/packet. The amount of errors that can be detected and corrected depends on the deployed FEC. We assume that a partition within a packet will be either received correctly or discarded due to uncorrectable errors. Let  $I$  be the number of packets in the given frame and  $i$  the packet index. For each macroblock, both shape coding parameters and texture coding parameters are specified. We use  $\mu_{S_i}$  and  $\mu_{T_i}$  to denote the shape and texture coding parameters for all the macroblocks in the  $i$ th packet, and  $B_{S_i}(\mu_{S_i})$  and  $B_{T_i}(\mu_{T_i})$  the corresponding total number of bits used to encode these partitions. Let us denote by  $\theta_i$  the embedding level for the  $i$ th packet (shape in here is used to include critical information, motion vectors, and lossy shape as mentioned in the previous section), and  $B'_{T_i}(\mu_{T_i}, \theta_i)$  the total number of bits used to encode the texture partition in the  $i$ th packet. As mentioned in Section II-B,  $\theta_i$  is selected from the set  $\{0, 1, 2, 3, 4\}$ , and if  $\theta_i = 0$ , then  $B'_{T_i}(\mu_{T_i}, \theta_i) \approx B_{T_i}(\mu_{T_i})$ , since only one bit is embedded in the original packet to specify that shape/motion data is not embedded.

#### B. Channel Model

In wireless channels, harsh conditions are often present with very high bit error rates (BER) in the order of  $10^{-1}$  to  $10^{-3}$  [6]. Thus, channel coding is required to bring the aggregate BER

down to a level where error resilient tools at the decoder can be effective. In this work, we apply channel coding separately on the shape and texture partitions. Let us denote by  $r_{S_i}$  and  $r_{T_i}$ , respectively, the channel code rate (the ratio of source bit-rate to total bit-rate) for shape and texture for the  $i$ th packet. The total bit rate then for the frame can be represented by

$$R = \sum_{i=1}^I \left[ \frac{B_{S_i}(\mu_{S_i})}{r_{S_i}} + \frac{B'_{T_i}(\mu_{T_i}, \theta_i)}{r_{T_i}} \right]. \quad (4)$$

We assume that burst errors can be converted into random errors with pre-interleaving [36]. Let us denote by  $\rho_{S_i}$  the probability of corruption of the  $i$ th shape data partition, and by  $\rho_{T_i}$  the probability of corruption of the  $i$ th texture partition. Then

$$\rho_{S_i} = 1 - (1 - p_e)^{\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}}} \quad \text{and} \quad \rho_{T_i} = 1 - (1 - p_e)^{\frac{B'_{T_i}(\mu_{T_i}, \theta_i)}{r_{T_i}}} \quad (5)$$

where  $p_e$  is the BER at the decoder, which is usually smaller than the channel bit error rate. It can be seen from (5) that the value of  $\rho_{S_i}$  depends on  $B_{S_i}(\mu_{S_i})$  and  $r_{S_i}$  in a way that either an increase of the source coding bit rate or an increase of the channel coding bit rate (smaller  $r_{S_i}$ ) will increase the probability of the partition error.

#### C. Expected Distortion

Due to the stochastic channel, the distortion at the receiver is a random variable. Let  $E[D_i]$  represent the expected distortion at the receiver for the  $i$ th packet. We have

$$E[D_i] = (1 - \rho_{S_i})(1 - \rho_{T_i})E[D_{R,i}] + (1 - \rho_{S_i})\rho_{T_i}E[D_{LT,i}] + \rho_{S_i}(1 - \rho_{T_i})E[D_{LS,i}] + \rho_{S_i}\rho_{T_i}E[D_{L,i}] \quad (6)$$

where  $E[D_{R,i}]$  is the expected distortion for the  $i$ th packet if both shape and texture partitions are received correctly,  $E[D_{LT,i}]$  is the expected distortion if the texture partition is corrupted,  $E[D_{LS,i}]$  is the expected distortion if the shape partition is corrupted, and  $E[D_{L,i}]$  is the expected distortion if both partitions are corrupted. Clearly,  $E[D_{R,i}]$  depends on the source coding and data embedding parameters for the packet only, while  $E[D_{LT,i}]$ ,  $E[D_{LS,i}]$ , and  $E[D_{L,i}]$  may also depend on the concealment strategy used at the decoder. It is expected that any data hiding scheme will introduce distortion since signal coefficients are altered. The amount of distortion caused by data hiding is directly dependent on the embedded level determined by the optimization process.

Note that the problem formulation and solution approach presented in this paper are general. Therefore, the techniques developed here are applicable to various concealment strategies used by the decoder. We assume that the channel conditions (as expressed by  $\rho_{S_i}$  and  $\rho_{T_i}$ ) and the concealment strategy are known at the transmitter. In our experimental results, we use the expected mean-squared error (MSE) as distortion metric, as is commonly done in the literature [37].

Based on our previous discussion the overall optimization problem (3) can be rewritten as

$$\begin{aligned} & \text{Minimize} \quad \sum_{i=1}^I E[D_i] \\ & \text{subject to:} \quad \sum_{i=1}^I \left[ \frac{B_{S_i}(\mu_{S_i})}{r_{S_i}} + \frac{B'_{T_i}(\mu_{T_i}, \theta_i)}{r_{T_i}} \right] \leq R_{\text{budget}}. \end{aligned} \quad (7)$$

#### IV. PROPOSED SOLUTION

In this section, we present an optimal solution for the optimization problem expressed in (7). We use the Lagrange multiplier method to relax the bit budget constraint. The relaxed problem can then be solved using the shortest path algorithm. We also study the expected distortion estimation problem with an in-depth analysis based on a popular error concealment strategy at the decoder side.

##### A. Optimal Solution

The Lagrangian relaxation method leads to a convex hull approximation of the constrained problem (7). Let  $U$  be the set of all possible decision vectors  $u_i$  for the  $i$ th packet ( $i = 1, 2, \dots, I$ ), where  $u_i = (\mu_{S_i}, \mu_{T_i}, r_{S_i}, r_{T_i}, \theta_i)$ . We first define a Lagrangian cost function

$$\begin{aligned} J_\lambda(u) &= E[D_{\text{tot}}] + \lambda R \\ &= \sum_{i=1}^I \left\{ E[D_i] + \lambda \left[ \frac{B_{S_i}(\mu_{S_i})}{r_{S_i}} + \frac{B'_{T_i}(\mu_{T_i}, \theta_i)}{r_{T_i}} \right] \right\} \end{aligned} \quad (8)$$

where  $\lambda$  is the Lagrange multiplier. If there exists a  $\lambda^*$  such that  $u^* = \arg[\min_u J_{\lambda^*}(u)]$  leads to  $R(u^*) = R_{\text{budget}}$ , then  $u^*$  is also an optimal solution to (7) [38], [39]. Therefore, the task of solving (7) is equivalent to the easier task of finding the optimal solution to the unconstrained problem that minimizes the cost function  $J_\lambda(u)$  and choosing the appropriate Lagrange multiplier to satisfy the constraint.

Most decoder concealment strategies introduce dependencies between packets. For example, if the concealment algorithm uses the motion vector of the macroblock above to conceal the lost macroblock, then it would cause the calculation of the expected distortion of the current packet to depend on its previous packet. Without loss of generality, we assume that due to the concealment strategy, the current packet will depend on its previous  $a$  packets ( $a \geq 0$ ). To implement the algorithm for solving the optimization problem, we define a cost function  $G_k(u_{k-a}, \dots, u_k)$ , which represents the minimum total bit rate and distortion up to and including the  $k$ th packet, given that  $u_{k-a}, \dots, u_k$  are decision vectors for the  $(k-a)$ th to  $k$ th packets. Therefore,  $G_I(u_{I-a}, \dots, u_I)$  represents the minimum total bitrate and distortion for all the packets of the frame, and thus

$$\min_u J_\lambda(u) = \min_{u_{I-a}, \dots, u_I} G_I(u_{I-a}, \dots, u_I). \quad (9)$$

The key observation for deriving an efficient algorithm is the fact that given  $a+1$  decision vectors  $u_{k-a-1}, \dots, u_{k-1}$  for the  $(k-a-1)$ st to  $(k-1)$ st packets, and the cost function  $G_{k-1}(u_{k-a-1}, \dots, u_{k-1})$ , the selection of the next decision vector  $u_k$  is independent of the selection of the previous decision vectors  $u_1, u_2, \dots, u_{k-a-2}$ . This means that the cost function can be expressed recursively as

$$\begin{aligned} & G_k(u_{k-a}, \dots, u_k) \\ &= \min_{u_{k-a-1}, \dots, u_{k-1}} \left\{ G_{k-1}(u_{k-a-1}, \dots, u_{k-1}) \right. \\ & \quad \left. + E[D_k] + \lambda \left[ \frac{B_{S_k}(\mu_{S_k})}{r_{S_k}} + \frac{B'_{T_k}(\mu_{T_k}, \theta_k)}{r_{T_k}} \right] \right\}. \end{aligned} \quad (10)$$

The recursive representation of the cost function above makes the future step of the optimization process independent from its past step, which is the foundation of dynamic programming.

The problem can be converted into a graph theory problem of finding the shortest path in a directed acyclic graph (DAG) [39]. The computational complexity of the algorithm is  $O(I \times |U|^{a+1})$  (where  $|U|$  is the cardinality of  $U$ ), which depends directly on the value of  $a$ . For most cases,  $a$  is a small number, so the algorithm is much more efficient than an exhaustive search algorithm with exponential computational complexity.

At the transmitter, the additional computational complexity due to data hiding can be attributed to the increased number of possible decision vectors  $u_i$  (or the size of  $|U|$ ). This is expected since the adaptive data hiding technique provides  $u_i$  more choices for its component  $\theta_i$ . However, the additional cost is not major for most cases where  $a$  is a small number. At the decoder side the computational overhead attributed to data hiding is minimal since only modulo arithmetic is needed to extract the hidden information.

##### B. Expected Distortion Estimation

As mentioned in the previous section, the techniques developed here can be applied to various error concealment strategies used by the decoder. In order to analyze the calculation of expected distortion in depth and to derive a practical formulation for the distortion estimation, as an example, we take a commonly used error concealment strategy, which provides sufficient information for distortion estimation based on another specific error concealment strategy.

The example error concealment strategy is described as follows.

- 1) If both shape and texture partitions are corrupted, then the decoder uses the motion vector of the neighboring macroblock above the current one as the concealment motion vector. If the concealment motion vector is not available, e.g., if the above macroblock is also lost, then the decoder uses a zero motion vector for concealment.
- 2) If only the shape partition is corrupted, then the decoder checks if enough shape and motion information is embedded inside the texture. If so, then both shape and texture can be recovered; otherwise, if not enough shape information is available, the shape has to be concealed using the motion vector of its neighbor macroblock above, as in (1);

if the embedded motion information is not available, then the texture has to be concealed as in (1).

- 3) If only the texture partition is corrupted, then the texture is recovered using the motion vector embedded in the shape partition.

In object-based video communications, video objects are compressed and transmitted separately. At the receiver, the decoder has the flexibility to decide how to combine the video objects in order to compose the VOP. To evaluate the distortion introduced by a video object, we assume that the transmitter has knowledge of the background VOP on which the transmitted video object will be composed at the receiver. Otherwise, a default background VOP will be used. Therefore, the distortion is calculated as the total intensity error of the composed frame.

We first define some necessary symbols and then present the equations required to compute the expected distortion. Let us denote by  $E[d_j]$  the expected distortion at the receiver at the  $j$ th pixel in the VOP, and  $N$  the total number of pixels in the packet. Let us denote by  $f_n^j$  the original value of pixel  $j$  in VOP  $n$ ,  $\hat{f}_n^j$  its reconstructed value at the decoder,  $s_n^j$  ( $s_n^j = 0$  for transparent or 1 for opaque block—in this work we considered only binary shape) and  $t_n^j$  the corresponding shape and texture components of  $f_n^j$ ,  $\tilde{s}_n^j$  and  $\tilde{t}_n^j$  the corresponding shape and texture components of  $\hat{f}_n^j$ , and  $g_n^j$  the background pixel value at the same position. Clearly

$$f_n^j = s_n^j t_n^j + (1 - s_n^j) g_n^j$$

and

$$\hat{f}_n^j = \tilde{s}_n^j \tilde{t}_n^j + (1 - \tilde{s}_n^j) g_n^j. \quad (11)$$

The expected distortion for the  $i$ th packet, if the expected MSE is used as the objective distortion metric, can be calculated by summing up the expected distortion of all the pixels in the packet

$$E[D_i] = \sum_{j=iN}^{iN+N-1} E[d_j] \quad (12)$$

where

$$\begin{aligned} E[d_j] &= E \left[ \left( f_n^j - \hat{f}_n^j \right)^2 \right] \\ &= (f_n^j)^2 - 2f_n^j E[\tilde{f}_n^j] + E \left[ \left( \tilde{f}_n^j \right)^2 \right] \\ &= (f_n^j)^2 - 2f_n^j E[\tilde{s}_n^j \tilde{t}_n^j] - 2f_n^j g_n^j \\ &\quad + 2f_n^j g_n^j E[\tilde{s}_n^j] + E \left[ \left( \tilde{s}_n^j \tilde{t}_n^j \right)^2 \right] \\ &\quad + 2g_n^j E[\tilde{s}_n^j \tilde{t}_n^j] - 2g_n^j E \left[ \left( \tilde{s}_n^j \right)^2 \tilde{t}_n^j \right] \\ &\quad + (g_n^j)^2 - 2g_n^j E[\tilde{s}_n^j] + (g_n^j)^2 E \left[ \left( \tilde{s}_n^j \right)^2 \right]. \end{aligned} \quad (13)$$

To calculate  $E[d_j]$  in (13), the first and second moments of the reconstructed shape and texture intensity value for the  $j$ th pixel are needed, which can be derived in a similar fashion to the one

presented in [40], [41]. In the following paragraph, we demonstrate how the first moment can be recursively calculated in time. We consider four cases

$$\begin{aligned} E \left[ \hat{s}_n^{j_s} \hat{t}_n^{j_t} \right] (I, I) \\ &= (1 - \rho_{S_i}) (1 - \rho_{T_i}) \hat{s}_n^{j_s} \hat{t}_n^{j_t} \\ &\quad + (1 - \rho_{S_i}) \rho_{T_i} \hat{s}_n^{j_s} E \left[ \hat{t}_{n-1}^{m_t} \right] + \rho_{S_i} (1 - \rho_{T_i}) \xi_n \\ &\quad + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E \left[ \hat{s}_{n-1}^{m_s} \hat{t}_{n-1}^{m_t} \right] \\ &\quad + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E \left[ \hat{s}_{n-1}^{j_s} \hat{t}_{n-1}^{j_t} \right] \end{aligned} \quad (14)$$

$$\begin{aligned} E \left[ \hat{s}_n^{j_s} \hat{t}_n^{j_t} \right] (I, P) \\ &= (1 - \rho_{S_i}) (1 - \rho_{T_i}) \hat{s}_n^{j_s} (\hat{e}_n^{j_t} + E \left[ \hat{t}_{n-1}^{m_t} \right]) \\ &\quad + (1 - \rho_{S_i}) \rho_{T_i} \hat{s}_n^{j_s} E \left[ \hat{t}_{n-1}^{m_t} \right] + \rho_{S_i} (1 - \rho_{T_i}) \xi_n \\ &\quad + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E \left[ \hat{s}_{n-1}^{k_s} \hat{t}_{n-1}^{k_t} \right] \\ &\quad + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E \left[ \hat{s}_{n-1}^{j_s} \hat{t}_{n-1}^{j_t} \right] \end{aligned} \quad (15)$$

$$\begin{aligned} E \left[ \hat{s}_n^{j_s} \hat{t}_n^{j_t} \right] (P, I) \\ &= (1 - \rho_{S_i}) (1 - \rho_{T_i}) E \left[ \hat{s}_{n-1}^{m_s} \right] \hat{t}_n^{j_t} \\ &\quad + (1 - \rho_{S_i}) \rho_{T_i} E \left[ \hat{s}_{n-1}^{m_s} \hat{t}_{n-1}^{m_t} \right] + \rho_{S_i} (1 - \rho_{T_i}) \xi_n \\ &\quad + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E \left[ \hat{s}_{n-1}^{k_s} \hat{t}_{n-1}^{k_t} \right] \\ &\quad + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E \left[ \hat{s}_{n-1}^{j_s} \hat{t}_{n-1}^{j_t} \right] \end{aligned} \quad (16)$$

$$\begin{aligned} E \left[ \hat{s}_n^{j_s} \hat{t}_n^{j_t} \right] (P, P) \\ &= (1 - \rho_{S_i}) (1 - \rho_{T_i}) (\hat{e}_n^{j_t} E \left[ \hat{s}_{n-1}^{m_s} \right] + E \left[ \hat{s}_{n-1}^{m_s} \hat{t}_{n-1}^{m_t} \right]) \\ &\quad + (1 - \rho_{S_i}) \rho_{T_i} E \left[ \hat{s}_{n-1}^{m_s} \hat{t}_{n-1}^{m_t} \right] + \rho_{S_i} (1 - \rho_{T_i}) \xi_n \\ &\quad + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E \left[ \hat{s}_{n-1}^{k_s} \hat{t}_{n-1}^{k_t} \right] \\ &\quad + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E \left[ \hat{s}_{n-1}^{j_s} \hat{t}_{n-1}^{j_t} \right] \end{aligned} \quad (17)$$

[see (18), shown at the bottom of the next page], where shape is intracoded in (14) and (15), and intercoded in (16) and (17); texture is intracoded in (14) and (16), and intercoded in (15) and (17);  $\hat{s}_n^j$  and  $\hat{t}_n^j$  are the encoder reconstructed shape and texture of the  $j$ th pixel, and  $\hat{s}_n^{j'}$  is the encoder reconstructed shape from the embedded information in the texture. Pixel  $j$  in frame  $n$  is predicted by pixel  $m$  in frame  $n-1$  if the motion vector is available, otherwise it is predicted by pixel  $k$  if the concealment motion vector is available. Therefore, in (15) and (17), when the texture is intercoded, the reconstructed residual is equal to  $\hat{e}_n^j = \hat{t}_n^j - \hat{t}_{n-1}^m$ . The subscript  $s$  and  $t$  of  $j_s, j_t, k_s, k_t, m_s,$  and  $m_t$  in (14)–(18) are used to distinguish shape from texture, because the motion vector or concealment motion vector of shape could be different from that of the texture. Clearly, each of the equations (14)–(17) is a summation of the product of possible values of  $\hat{s}_n^{j_s} \hat{t}_n^{j_t}$  and their corresponding probabilities. For example, in (15), the first product term in the expression for the expectation of  $\hat{s}_n^{j_s} \hat{t}_n^{j_t}$  represents the case when both shape and texture data are received correctly at the decoder. Here, the expected value is  $\hat{s}_n^{j_s} (\hat{e}_n^{j_t} + E[\hat{t}_{n-1}^{m_t}])$  because the value of the predicted texture data from the previous frame  $\hat{t}_{n-1}^{m_t}$  is unknown and, therefore, modeled as a random variable by the encoder. Readers interested in the fundamental theory of recursive optimal pixel-based (ROPE) distortion estimation are referred to [37]. The multiplier  $\xi_n$  present in the third product term of equations (14)–(17)

is defined in (18). It represents the case when the shape partition is corrupted but the texture partition is received, therefore the shape/motion data are recovered by the embedded information in the received texture data. In (18), all possible combinations of the embedding levels and texture coding modes are considered. Therefore, for embedding level = 0, no shape/motion information is embedded and thus both shape and texture are concealed using the motion vectors of the above neighboring macroblock. On the other hand, for embedding level = 4 and intercoded texture, the shape data can be reconstructed and the texture data can be concealed by the motion vectors extracted from the embedded data.

Computing  $E[\tilde{s}_n^j \tilde{t}_n^j]$  in (15)–(17) depends on the computation of  $E[\tilde{s}_n^j]$  and  $E[\tilde{t}_n^j]$ , which are calculated recursively as follows:

$$E[\tilde{s}_n^j](I) = (1 - \rho_{S_i}) \hat{s}_n^j + \rho_{S_i} (1 - \rho_{T_i}) \gamma_n \\ + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^k] \\ + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E[\tilde{s}_{n-1}^j] \quad (19)$$

$$E[\tilde{t}_n^j](I) = (1 - \rho_{T_i}) \hat{t}_n^j \\ + \rho_{T_i} (1 - \rho_{S_{i-1}}) E[\tilde{t}_{n-1}^k] \\ + \rho_{T_i} \rho_{S_{i-1}} E[\tilde{t}_{n-1}^j] \quad (20)$$

$$E[\tilde{s}_n^j](P) = (1 - \rho_{S_i}) E[\tilde{s}_{n-1}^m] \\ + \rho_{S_i} (1 - \rho_{T_i}) \gamma_n \\ + \rho_{S_i} \rho_{T_i} (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^k] \\ + \rho_{S_i} \rho_{T_i} \rho_{S_{i-1}} E[\tilde{s}_{n-1}^j] \quad (21)$$

$$E[\tilde{t}_n^j](P) = (1 - \rho_{T_i}) (\hat{e}_n^j + E[\tilde{t}_{n-1}^m]) \\ + \rho_{T_i} (1 - \rho_{S_{i-1}}) E[\tilde{t}_{n-1}^k] \\ + \rho_{T_i} \rho_{S_{i-1}} E[\tilde{t}_{n-1}^j] \quad (22)$$

[see (23), shown at the bottom of the page], where shape and texture are intracoded in (19) and (20), and intercoded in (21) and (22). The second moment can be computed in a similar way.

## V. EXPERIMENTAL RESULTS

In this section, experiments are designed to demonstrate the advantages of using data hiding coupled with the hybrid scheme of joint source coding and channel coding. Our simulations are based on MPEG-4 VM18.0 [2]. The available intramode quantizers are of step size 2, 4, 6, 8, 10, 14, 18, 24, and 30, and the available Intermode quantizers are of step size 1, 3, 7, 11, 15, 19, 25, and 31. The texture component of each macroblock can be coded as intra- or intermode. The shape can be coded as transparent, opaque, or boundary mode. For each boundary BAB, the scan type and resolution (conversion ratio of 1, 1/2, or 1/4) are also selected. The intermode shape coding has not been considered here because it violates the assumption that each packet is independently decodable [3]. We assume that the first frame in the sequence is coded as intramode with a quantization step size of 6 and that enough protection is used so that it arrives correctly at the decoder. This assumption makes the initial conditions identical for all the experiments.

In the first set of experiments, we test the advantages of using adaptive data hiding scheme in source coding. We encode the first 80 frames of the ‘‘Children’’ sequence and transmit them (without channel coding) over the wireless channels with  $BER = 10^{-3}$  and  $BER = 10^{-4}$ . The two approaches, one with adaptive data hiding and one without, are compared in Fig. 5. In both approaches, the source coding is optimized. As expected, the method with adaptive data hiding outperformed the other one. From the figures can be seen that the method using adaptive data hiding starts to gain after the bit rate exceeds a certain value, for example when rate  $> 2400$  in Fig. 5(a). In other words, when the bit rate is very low, the two approaches have

$$\xi_n = \begin{cases} (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^{k_s} \tilde{t}_{n-1}^{k_t}] + \rho_{S_{i-1}} E[\tilde{s}_{n-1}^{j_s} \tilde{t}_{n-1}^{j_t}], & \text{embedding level} = 0 \\ \left\{ (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^{k_s}] + \rho_{S_{i-1}} E[\tilde{s}_{n-1}^{j_s}] \right\} \hat{t}_n^{j_t}, & \text{embedding level} = 1 \text{ or } 2, \text{ intra texture} \\ (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^{k_s} \tilde{t}_{n-1}^{k_t}] + \rho_{S_{i-1}} E[\tilde{s}_{n-1}^{j_s} \tilde{t}_{n-1}^{j_t}], & \text{embedding level} = 1, \text{ inter texture} \\ (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^{k_s} \hat{t}_n^{m_t}] + \rho_{S_{i-1}} E[\tilde{s}_{n-1}^{j_s} \tilde{t}_{n-1}^{m_t}], & \text{embedding level} = 2, \text{ inter texture} \\ \hat{s}_n^{j_s} \hat{t}_n^{j_t}, & \text{embedding level} \geq 3, \text{ intra texture} \\ \left\{ (1 - \rho_{S_{i-1}}) E[\tilde{t}_{n-1}^{k_t}] + \rho_{S_{i-1}} E[\tilde{t}_{n-1}^{j_t}] \right\} \hat{s}_n^{j_s}, & \text{embedding level} = 3, \text{ inter texture} \\ \left\{ (1 - \rho_{S_{i-1}}) E[\tilde{t}_{n-1}^{m_t}] + \rho_{S_{i-1}} E[\tilde{t}_{n-1}^{j_t}] \right\} \hat{s}_n^{j_s}, & \text{embedding level} = 4, \text{ inter texture} \end{cases} \quad (18)$$

$$\gamma_n = \begin{cases} (1 - \rho_{S_{i-1}}) E[\tilde{s}_{n-1}^{k_s}] + \rho_{S_{i-1}} E[\tilde{s}_{n-1}^{j_s}], & \text{shape is not embedded} \\ \hat{s}_n^{j_s}, & \text{shape is embedded} \end{cases} \quad (23)$$

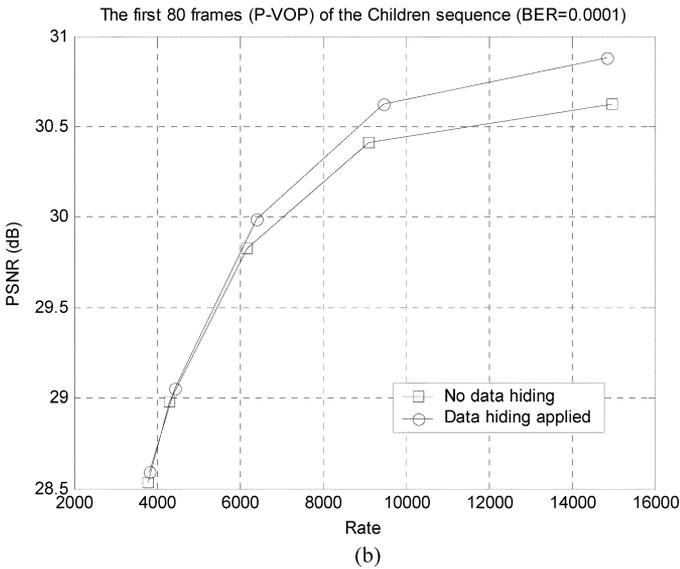
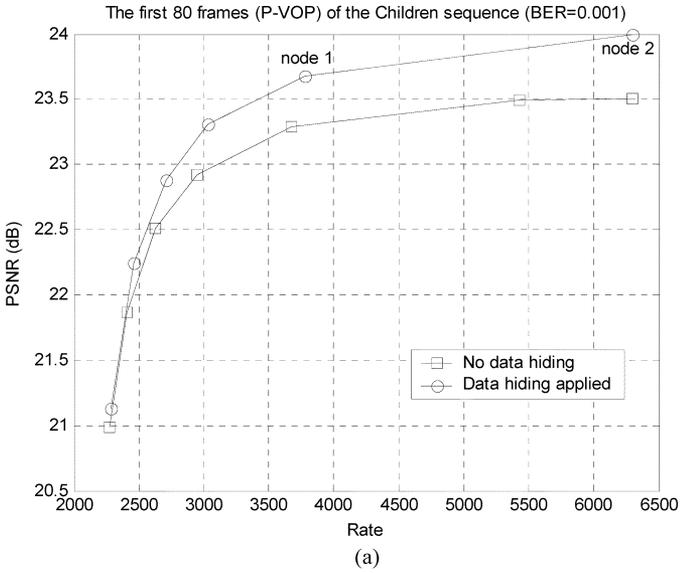


Fig. 5. Comparing the performance of the propose methods with and without data hiding.

the same performance. This is expected since the texture data are coded in a very coarse quality, which brings the number of non-zero DCT coefficients below the number needed for the first embedding level. When the bit rate increases, there is more room in the DCT coefficients for embedding, and this is where the benefits of data hiding start to appear. Moreover, the figures indicate that the gain of using adaptive data hiding gets smaller when a better channel (smaller BER), is used. This is reasonable because the gain of using data hiding mainly comes from the ability to recover shape. However, when the BER of the channel decreases, the probability of a corrupted shape partition also decreases, thus the gain decreases. In Fig. 6, the distribution of packets of the frames in each embedding level for two selected points in Fig. 5(a) is shown. We find that the point corresponding to the lower bit rate (point 1) allocates more packets in embedding level 0 (which corresponds to no embedding) than the other point (point 2), while allocates less packets in embedding level 1 and level 2. This proves our major

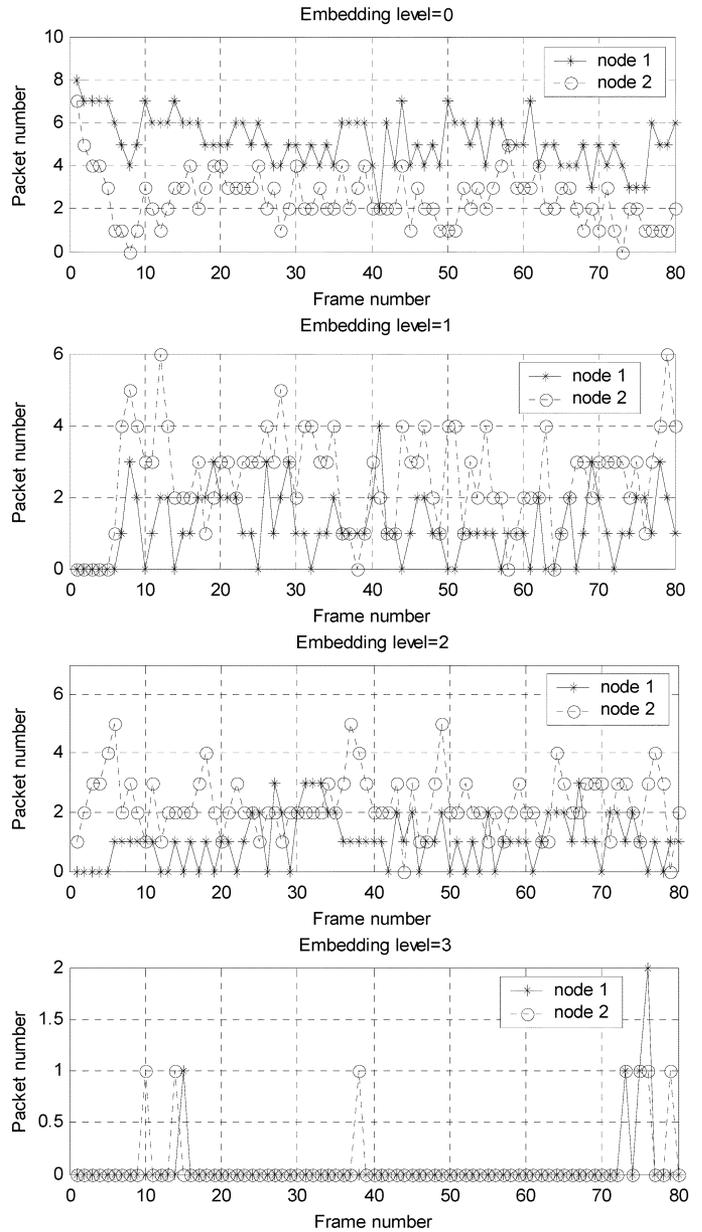


Fig. 6. Distribution of packets in each embedding level for two nodes in Fig. 5(a). (a) SNR = 6 dB. (b) SNR = 10 dB.

argument that bit budget permitting, data embedding is another effective way to help improve the reconstructed video quality in addition to source coding mode selection. However, the above observation should not be interpreted that, the more embedding the better, since more embedding could introduce higher distortion to the texture data. As we pointed out previously, the relationship among source bit rate, distortion, and data embedding is rather complicated. The distortion fundamentally is caused by compression and transmission. When the source bit rate increases, the distortion caused by compression decreases. As the probability of partition error increases, directly causes an increase in transmission errors. In addition, an increase of source bit rate could provide more room for data embedding, which generally decreases the transmission error but increases the compression error. This complicated relationship indirectly

TABLE II  
PERFORMANCE OF RCPC (IN BER) OVER A RAYLEIGH  
FADING CHANNEL WITH INTERLEAVINS

Channel SNR (dB)	6	10	18
Channel rate=4/7	$5.3 \times 10^{-4}$	$4.1 \times 10^{-5}$	$3.8 \times 10^{-6}$
Channel rate=2/3	$7.4 \times 10^{-3}$	$1.7 \times 10^{-4}$	$1.2 \times 10^{-5}$
Channel rate=4/5	$4.0 \times 10^{-2}$	$6.6 \times 10^{-4}$	$3.6 \times 10^{-5}$

proves the importance of our proposed optimal scheme. By dynamically adjusting the parameters in coding and embedding, the optimization procedure finds the best solution to minimize a target function (or distortion).

In the second set of experiments, we consider the advantage of joint source-channel coding. We compare three approaches, namely, 1) an equal error protection (EEP) method, where the shape and texture partition within a packet are equally protected (by channel coding) but different packets are allowed to be unequally protected; 2) a UEP method, where the shape and texture partitions are unequally protected by channel coding; 3) a hybrid method, where the shape and texture partition are unequally protected by channel coding, and the shape and motion data are allowed to be embedded into the texture data with various levels. Data hiding is not used in EEP and UEP methods, but source coding parameters are optimized for all these methods.

In our simulation, we use a rate compatible punctured convolutional (RCPC) channel code with generator polynomials (133,171), mother code rate 1/2, and puncturing rate  $P = 4$ . This mother rate is punctured to achieve the 4/7, 2/3, and 4/5 rate codes. At the receiver, soft Viterbi decoding is used in conjunction with BPSK demodulation. We present experiments on Rayleigh fading channels with channel parameter defined as  $SNR = \alpha(E_b/N_0)$ . The bit error rates for the Rayleigh fading with the assumption of ideal interleaving were obtained experimentally using simulations [42], and are shown in Table II.

We encode the first 80 frames of the ‘‘Children’’ sequence and transmit them over the simulated wireless channels with  $SNR = 6$  dB and  $SNR = 10$  dB, respectively. The experimental results for the three approaches are shown in Fig. 7. For  $SNR = 6$  dB [as shown in Fig. 7(a)] the UEP method outperformed EEP at lower bit rates (when rate is lower than 6000) because it can allocate more channel bits to protect shape data, which has a stronger impact on the decoded video quality. However, when the bit rate goes up, the probability of a corrupted partition increases (recall the probability of a corrupted partition is related to the partition length and the BER of the channel), and in order to control the error, both UEP and EEP methods are forced to choose the channel rate = 4/7 for channel coding, which corresponds to the smallest bit error rate. Again, we observe that data hiding works well when the bit rate is high enough to ensure that there are enough DCT coefficients available to embed the shape and motion information. This way, the hybrid method inherits the virtue of both UEP and data hiding, and thus makes it perform well for all ranges of the bit rate.

However, the advantage of using a UEP or a hybrid method decreases when channel conditions improve. For the  $SNR =$

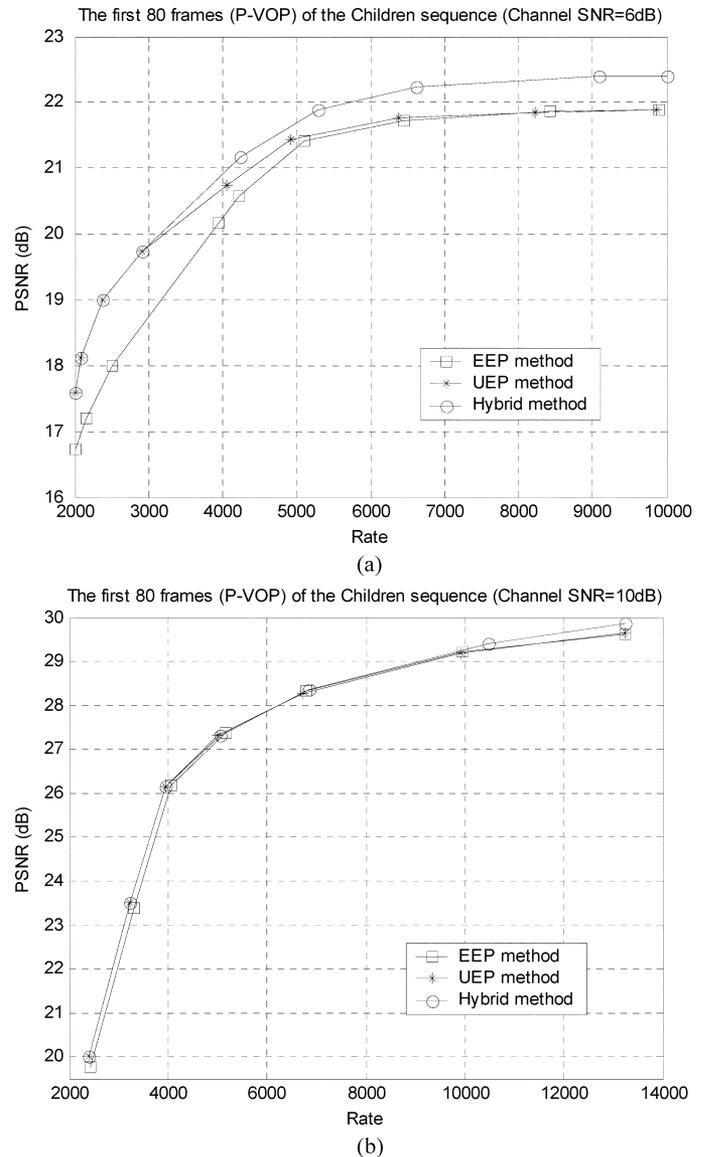


Fig. 7. R-D curves for Children sequence. (a)  $SNR = 6$  dB. (b)  $SNR = 10$  dB. (c)  $SNR = 18$  dB.

10 dB channel shown in Fig. 7(b), the hybrid method has a maximum gain of 0.3 dB over EEP. The result is expected, since the differences between the BERs corresponding to various channel code rates are getting smaller, which directly affect the gain of UEP over EEP. On the other hand, better channel conditions decrease the probability of a corrupted shape partition, which directly affects the gain of using data hiding.

In another experiment, we test the advantages of the proposed approach on sequences with higher motion activity. We encode the Bream sequence (frames 100 to 130), which corresponds to the flipping action of the Bream, and transmit it over wireless channels with  $SNR = 6$  dB,  $SNR = 10$  dB, and  $SNR = 18$  dB. The experimental results are shown in Fig. 8. When a channel  $SNR$  of 6 dB is used [as shown in Fig. 8(a)], the whole advantage is attributed to UEP. It is understandable that data hiding has no contribution in this case since the bit rate is too low to find enough coefficients to embed the lowest level shape/motion information. When the  $SNR = 10$  dB channel is used [as shown

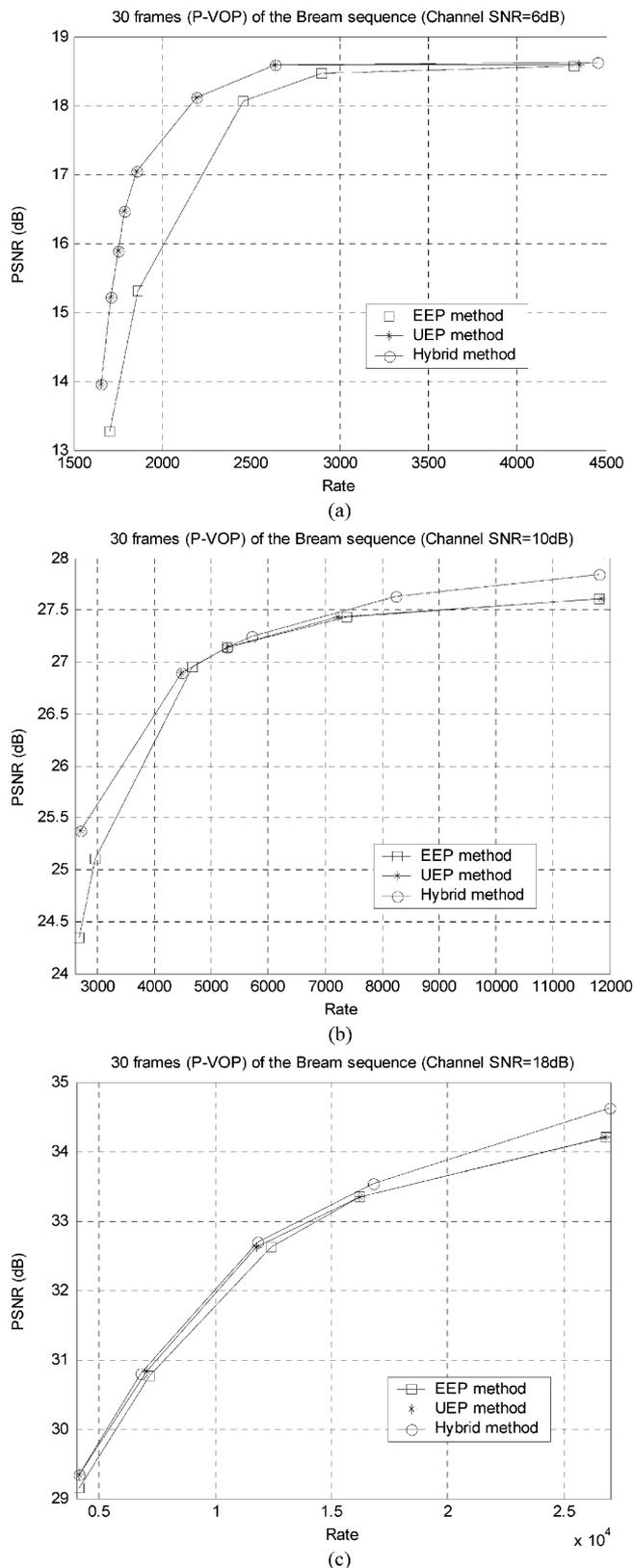


Fig. 8. R-D curves for Bream sequence.

in Fig. 8(b)], the hybrid approach performs better at lower bit rates due to UEP and at higher bit rates due to the data hiding. When SNR = 18 dB [as shown in Fig. 8(c)], the proposed hybrid approach has a gain of 0.2–0.5 dB over EEP. The results

of this experiment indicate that the proposed approach works better for video sequences containing higher motion and morphing activities. This is expected since the efficient encoding of these sequences relies highly on the shape and motion information, which makes the error resilience of such data very important and very beneficial.

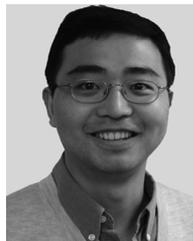
## VI. CONCLUSION

In this paper, a hybrid source-channel coding scheme is proposed for wireless object-based video communications. In the source coding, an adaptive data hiding method is adopted to embed shape and motion information into texture DCT coefficients for decoder concealment purposes. The level of embedding is jointly optimized with the selection of source coding parameters and channel coding rates to achieve the minimum expected distortion. The formalized optimization problem is solved by Lagrangian relaxation and dynamic programming. We tested the proposed scheme by encoding and transmitting object-based video sequences over a simulated Rayleigh fading wireless channel, based on the MPEG-4 verification model. The results indicate that the proposed hybrid scheme has performance advantages at all bit rates. The improved performance at lower bit rates is attributed to UEP while at higher bit rates to data hiding.

## REFERENCES

- [1] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 4, pp. 61–82, Jul. 2000.
- [2] *MPEG-4 video VM 18.0*, ISO/IEC JTC1/SC29/WG11 N3908, Jan. 2001, Pisa, Italy.
- [3] N. Brady, "MPEG-4 standardized methods for the compression of arbitrarily shaped video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1170–1189, Dec. 1999.
- [4] H. Wang, G. M. Schuster, and A. K. Katsaggelos, "Rate-distortion optimal bit allocation scheme for object-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [5] H. Wang, F. Zhai, Y. Eisenburg, and A. K. Katsaggelos, "Optimal object-based video communications over differentiated services networks," presented at the IEEE Int. Conf. Image Processing, Singapore, Sep. 2004.
- [6] W. R. Heinzelman, M. Budagavi, and R. Talluri, "Unequal error protection of MPEG-4 compressed video," in *Proc. Conf. Image Processing*, Oct. 1999, pp. 530–534.
- [7] J. Cai, Q. Zhang, W. Zhu, and C. W. Chen, "An FEC-based error control scheme for wireless MPEG-4 video transmission," in *Proc. IEEE Wireless Communications and Networking Conference*, Chicago, Sep. 2000, pp. 1243–1247.
- [8] M. G. Martini and M. Chiani, "Proportional unequal error protection for MPEG-4 video transmission," in *Proc. IEEE Conf. Communications*, Helsinki, Jun. 2001, pp. 1033–1037.
- [9] S. Worrall, S. Fabri, A. H. Sadka, and A. M. Kondoz, "Prioritization of data partitioned MPEG-4 video over mobile networks," *Eur. Trans. Telecommun.*, vol. 12, no. 3, pp. 169–174, May/June 2001.
- [10] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding—A survey," *Proc. IEEE*, vol. 87, no. 7, pp. 1062–1078, Jul. 1999.
- [11] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. New York: Morgan Kaufmann, 2001.
- [12] D. L. Robie and R. M. Mersereau, "Video error correction using steganography," *EURASIP J. Appl. Signal Process.*, vol. 20002, no. 2, pp. 164–173, Feb. 2002.
- [13] F. Bartolini, A. Manetti, A. Piva, and M. Barni, "A data hiding approach for correcting errors in H.263 video transmitted over a noisy channel," in *Proc. IEEE 4th Workshop on Multimedia Signal Processing*, Cannes, France, Oct. 2001, pp. 65–70.
- [14] J. Song and K. J. R. Liu, "A data embedded video coding scheme for error-prone channels," *IEEE Trans. Multimedia*, vol. 3, no. 4, pp. 415–423, Dec. 2001.

- [15] P. Yin, B. Liu, and H. H. Yu, "Error concealment using data hiding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2002, pp. 729–732.
- [16] M. Carli, D. Bailey, M. Farias, and S. K. Mitra, "Error control and concealment for video transmission using data hiding," in *Proc. Wireless Personal Multimedia Comm.*, Oct. 2002, pp. 812–815.
- [17] C. S. Lu, "Wireless multimedia error resilience via a data hiding technique," presented at the Int. Workshop Multimedia & Expo, Dec. 2002.
- [18] M. Kurosaki, K. Munadi, and H. Kiya, "Error correction using data hiding technique for JPEG 2000 images," in *Proc. Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, vol. III, pp. 473–476.
- [19] A. Giannoula and D. Hatzinakos, "Compressive data hiding for video signals," in *Proc. Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, vol. I, pp. 529–532.
- [20] L. W. Kang and J. Leou, "An error resilient coding scheme for H.264 video transmission based on data embedding," presented at the IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, May 2004.
- [21] L. W. Kang and J. J. Leou, "A new error resilient coding scheme for H.263 video transmission," in *Proc. IEEE Pacific-Rim Conf. Multimedia*, Hsinchu, Taiwan, 2002, pp. 814–822.
- [22] C. B. Adsumilli, M. Carli, M. C. Q. de Farias, and S. K. Mitra, "A hybrid constrained unequal error protection and data hiding scheme for packet video transmission," presented at the Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Hong Kong, Apr. 2003.
- [23] R. E. V. Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proc. IEEE*, vol. 87, no. 10, pp. 1734–1750, Oct. 1999.
- [24] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source/Channel Coding for SNR Scalable Video Processing," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1043–1052, Sep. 2002.
- [25] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
- [26] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," *Syst. Video Commun.*, pp. 205–213, Oct. 1996.
- [27] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
- [28] S. Craver, N. Memon, B. Yeo, and M. Yeoung, "Can invisible watermarks resolve right ownership?," *Proc. SPIE*, vol. 3022, pp. 310–321, 1997.
- [29] D. Simitopoulos, S. A. Tsafaris, N. V. Boulgouris, G. A. Triantafyllidis, and M. G. Strintzis, "Intelligent Integrated Media Communication Techniques," in *Digital Watermarking for the Copyright Protection of Compressed Video*, M. Najim, M. Ansoorge, and J. F. Tasic, Eds. Norwell, MA: Kluwer, Dec. 2003.
- [30] D. Mukherjee, J. J. Chae, S. K. Mitra, and B. S. Manjunath, "A source and channel-coding framework for vector-based data hiding in video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 630–645, Jun. 2000.
- [31] P. Campisi, M. Carli, G. Giunta, and A. Neri, "Tracing watermarking for multimedia communication quality assessment," in *Proc. IEEE Int. Conf. Communications*, New York, Apr. 2002, pp. 1154–1158.
- [32] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [33] M. Wu, H. Yu, and A. Gelman, "Multi-level data hiding for digital image and video," *Proc. SPIE*, vol. 3854, 1999.
- [34] M. L. Miller, G. J. Doerr, and I. J. Cox, "Applying informed coding and embedding to design a robust, high capacity watermark," *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 792–807, Jun. 2004.
- [35] M. Wu and B. Liu, "Modulation and multiplexing techniques for multimedia data hiding," in *Proc. SPIE ITcom Conf.*, Denver, Co, Aug. 2001, vol. 4518.
- [36] B. Girod and N. Farber, *Compressed video over Networks* (in Wireless video), M.-T. Sun and A. R. Reibman, Eds. New York: Marcel Dekker, 2000.
- [37] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.
- [38] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, pp. 399–417, 1963.
- [39] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.
- [40] H. Wang and A. K. Katsaggelos, "Robust network-adaptive object-based video encoding," presented at the IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, May 2004.
- [41] H. Wang, F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Cost-distortion optimal unequal error protection for object-based video communications," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [42] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless video communications," in *41st Allerton Conf. Communication, Control, and Computing*, Oct. 2003.
- [43] H. Wang and A. K. Katsaggelos, "A hybrid source-channel coding scheme for object based wireless video communications," presented at the Proc. 13th Int. Conf. Computer Communications and Networks, Chicago, IL, Oct. 2004.



**Haohong Wang** (S'03–M'04) received the B.S. degree in computer science and the M.Eng. degree in computers and their applications from Nanjing University, China, the M.S. degree in computer science from University of New Mexico, Albuquerque, and the Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, IL.

He is currently with Marvell Technology Group at Santa Clara, CA. Prior to joining Marvell, he held various technical positions at AT&T, Catapult Communications, and Qualcomm. Dr. Wang's

research involves the areas of computer graphics, human computer interaction, image/video analysis and compression, and multimedia signal processing and communications. He has published more than 30 articles in peer-reviewed journals and International conferences. He is the inventor of more than 20 pending U.S. patents.

Dr. Wang is a member of the IEEE Visual Signal Processing and Communications Technical Committee, IEEE Multimedia and Systems Applications Technical Committee, and the IEEE Multimedia Communications Technical Committee. He is currently serving as a Guest Editor for the *IEEE Communications Magazine* Special Issue on Advances in Visual Content Analysis and Adaptation for Multimedia Communications, a Guest Editor for the *Journal of Wireless Communications and Mobile Computing* Special Issue on Video Communications for 4G Wireless Systems, and he is the Co-Editor of *Computer Graphics* (Publishing House of Electronics Industry, 1997). He has served as the Technical Program Co-Chair of the 2005 IEEE WirelessCom Symposium on Multimedia over Wireless (Maui, HI) and the 2006 International Symposium on Multimedia over Wireless (Vancouver, BC, Canada). He will serve as the Technical Program Co-Chair of the 16th International Conference on Computer Communications and Networks (ICCCN07) in 2007.



**Sotirios A. Tsafaris** (S'01) was born in Thessaloniki, Greece, in 1978. He received the ECE diploma from the Aristotle University of Thessaloniki in 2000 and the M.Sc. degree in electrical and computer engineering from Northwestern University, Evanston, IL, in June 2003, where he is currently pursuing the Ph.D. degree.

He is with the Department of Electrical Engineering and Computer Science and the Image and Video Processing Laboratory, Northwestern University, where he works on bio-molecular computing applications and bioinformatics. He held a research position at the Informatics and Telematics Institute, Greece, where he worked on digital rights management (DRM). He has published more than ten articles in the areas of DNA-based digital signal processing and digital watermarking. His research interests include biomolecular computing and bioinformatics, microarray image compression, DRM, and image and video processing.

Mr. Tsafaris is the Technical Chair of NU Nanoalliance. He is a member of the International Society for Nanoscale Science, Computation and Engineering (INSNCE) and the Technical Chamber of Greece. He has held research and teaching assistantship positions and was the recipient of the Murphy Fellowship (2001–2002). His work is also funded by the Alexander S. Onassis post-graduate fellowship (2001–present) from the Alexander S. Onassis Public Benefit Foundation.



**Aggelos K. Katsaggelos** (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently a Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital. He is the editor of *Digital Image Restoration* (New York: Springer-Verlag, 1991), coauthor of *Rate-Distortion Based Video Compression* (Norwell, MA: Kluwer, 1997), and co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Norwell, MA: Kluwer, 1998), and the co-inventor of eight international patents

Dr. Katsaggelos is a member of the Publication Board of the IEEE PROCEEDINGS, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002), member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), Area Editor for the journal *Graphical Models and Image Processing* (1992–1995), member of the Steering Committees of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), and an IEEE Signal Processing Society Best Paper Award (2001).