A Hybrid Source-Channel Coding Scheme for Object-based Wireless Video Communications

Haohong Wang
Qualcomm Inc., San Diego, USA
Email: haohongw@qualcomm.com

Aggelos K. Katsaggelos
Northwestern University, Evanston, USA
Email: aggk@ece.northwestern.edu

Abstract-- **In this paper, we study the joint source-channel coding of object-based video, and propose a data hiding scheme that improves the video error resilience by adaptively embedding the shape and motion information in 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 the knowledge of transmission channel conditions. The problem is solved using Lagarangian relaxation and dynamic programming. Experimental results indicate that the proposed hybrid source-channel coding scheme significantly outperforms methods without data hiding or unequal error protection.**

I. INTRODUCTION

Wireless video communications require advanced video compression techniques due to the limitations of transmission bandwidth and energy. However, these techniques make the coded bitstream very sensitive to channel errors. In this paper, we consider how to conduct forward error correction (FEC) in both the source coding and the channel coding to reach the best quality of error resilience, because the alternative approach, automatic repeat request (ARQ) is not feasible for real-time video communications due to the retransmission delay.

In the source coding, an effective way to stop temporal error propagation is to code the frames or macroblocks in intra-mode periodically. In [1-3], optimal rate-distortion schemes have been proposed to optimally determine the number of intra coded macroblocks and their spatial locations based on the knowledge of the transmission channel conditions. Another way is to provide supportive information to improve the decoder's ability in error concealment. Data hiding (or data embedding) techniques [4] are good examples that have attracted much attention recently, which hide redundancies of some important information, such as edge information and motion vectors, into the coded bitstream for future error concealment purpose. [5] embedded redundant information for protecting motion vectors and coding modes of current frame into the motion vectors of next frame. In [6] the block type and major edge direction of content blocks of an image was embedded into the DCT coefficients of another block. [7] and [8] embedded one or several copies of an approximation version of the original frame inside the frequency coefficients, and [9] embedded the edge information for I-frame, and the motion information, the reference frame, the code modes, and the error concealment strategy for P-frame.

Channel coding involves the insertion of redundant bits into the bitstream to make it possible to detect and correct errors. Recently, joint source-channel coding [10] is widely studied and a practical solution is to keep the source and channel coder separate but optimize their parameters jointly [11]. In [12], a hybrid method for lossy video communications was proposed that combines data hiding and channel coding. This work considers data hiding impact in the optimization, however, neither the source coding parameters is included, nor the data hiding is adaptive.

Object-based video representation comes up recently following the increasing request of content-based interactivity. In this model, the video data is composed of shape, motion and texture information, which have completely different stochastic characteristics and bit rate proportion. A rate-distortion optimal source-coding scheme was proposed in [13] for solving the bit allocation problem, which concludes that for some applications, the shape may have a stronger impact on video quality than texture. This directly motivates the unequal protection of the shape and texture components in video encoding and transmission. However, so far, there is very limited work reported in the joint source-channel coding for object-based video. One important reason is that arbitrarily shaped video objects make the video processing and transmission very complicated. In this paper, we propose a general rate-distortion optimal source-channel scheme for object-based video communications. 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 by using Lagrangian relaxation and dynamic programming. In Section V, some implementation details are demonstrated and experimental results are presented. We draw conclusions in the last section.

II. DATA HIDING FOR VIDEO ENCODING

A. General overview

Data hiding [4] has been widely used in a lot of applications, such as copyright protection, and secure transmission applications. An outstanding feature of data hiding scheme is that it enables the possibility of the retrieval of hidden informtion even without the availability of the original host signal. The general data hiding scheme is shown in Fig. 1, where at the sender side, a binary sequence *m* is embedded in the host signal sequence *x* to form a signal sequence *y*; After transmitted over an noisy channel, the signal *y* could be corrupted and becomes $\hat{y}$ at the receiver. The decoder extracts the estimated embedded signal $\hat{m}$ and reconstructs signal $\hat{x}$ from signal $\hat{y}$.

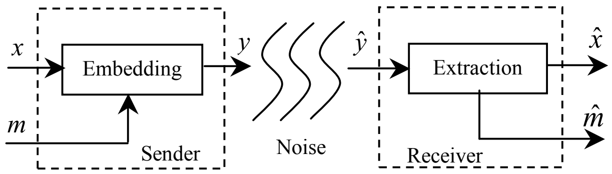


Figure 1. General data hiding scheme

There are two popular data embedding and extracting approaches, spread spectrum method [14] and odd-even method [15]. We describe both methods in the following, and we assume that the host data *x* is the DCT coefficients of the frame, as in the most cases.

In spread spectrum method, the data to be embedded is spread over the host data so that the embedded energy for each host frequency bins is negligible, which makes the embedded data imperceptible. Initially, a pseudo random algorithm is used to generate a pseudo-noise signal sequence *w*, which has values in the range of [-1,1] and is of zero mean. Then each embedded signal m_i is spread into *M* different locations inside the host sequence as

$$m_i^j = m_i w_{i \times j}, \ (j=1,2,\ldots, M). \tag{1}$$

The embedding process is $y_{i \times j} = x_{i \times j} + \alpha m_i^j$, (2)

where α is a scaling factor that vary the strength of the embedded information, that is, as α increases, the embedded data becomes more robust, but the difference between *x* and *y* increases. At the receiver side, the seed for generating random sequence is known and thus *w* can be generated. An estimation of the spread version of the embedded signal is performed as follows,

$$\hat{m}_i = \frac{1}{M} \sum_{j=1}^{M} \hat{y}_{i \times j} w_{i \times j} \ . \tag{3}$$

In odd-even embedding method, the data is embedded on the non-zero quantized AC coefficients. If a bit to be embedded is “0”, the AC coefficient is changed to an even number, otherwise, the AC coefficient is changed to an odd number. The scheme can be represented by

$$Embed(p,b) = \begin{cases} p+1 & if\ p > 0\ and\ (p-b) \bmod 2 \neq 0 \\ p-1 & if\ p < 0\ and\ (p-b) \bmod 2 \neq 0 \\ p & otherwise \end{cases}, \tag{4}$$

where *p* is the AC coefficient and *b* is the embedded bit.

The data extraction is quite straightforward, which can be represented by

$$Extract(p,b) = (p-b) \bmod 2 \ . \tag{5}$$

B. Hiding shape/motion in texture for MPEG-4 video

In MPEG-4, data partitioned packetization scheme is applied to increase the error resilience. In this scheme, the shape and texture data are packed in a same packet (see Fig. 2) but separated by a motion marker. This way, when the partition containing texture data is corrupted, the motion vector contained in the other partition (if available) can be used to conceal the corrupted texture. However, since the decoding of the texture partition is relied on the information stored in the shape partition, such as texture motion vector, texture coding mode, and shape BAB type, the whole packet will be discarded when the shape data is corrupted, even if the texture data is correctly received.

Bab_type	MVDs	CR	ST	BAC	COD	MCBPC
MVD	Motion Marker	AC_pred	CBPY	DC	AC	

Figure 2. Data partitioned bitstream syntax for P-VOP

In this work, we consider to embed some information of shape and motion partition 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 also help to paritally recover the lost shape and motion paritition. In Fig. 2, the BAB type, COD, texture coding mode, and CBPC are critical information for the decoding of texture. Besides these, the motion vector and shape data can also be embedded. Normally, the amount of embedded information is restricted by the slots for embedding and also its impact on the texture distortion. Thus, an adaptive embedding scheme is helpful in which the embedded information is divided into several levels, and let the encoder to decide the level of information for embedding. In this work, we use odd-even method, which can guarantee the correct extraction of the embedded information, although the embedding capacity is further reduced to the number of non-zero DCT coefficients.

III. PROBLEM FORMULATION

In this paper, we jointly consider the source coding, channel coding, data embedding and error concealment within a rate-distortion optimization framework. By selecting source coding 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

$$\text{Minimize } E[D_{tot}], \text{ Subject to } R \leq R_{budget}, \qquad (6)$$

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

A. System model

We consider an MPEG-4 [16] compliant object-based video application, where the data hiding mentioned above is applied. In the encoding, the VOP is divided into 16x16 macro blocks, which are numbered in raster scan order and divided into groups (rows) called packets. Each packet is independently decodable; that is, each packet has enough information for decoding and is independent of other packets and their related information. This guarantees that a single bit error only affects the decoding of a single packet. Let I be the number of packets in the given frame and i be packet index. For each macroblock, both shape coding parameters and texture coding parameters are specified. We use μ_{Si} and μ_{Ti} to denote the shape and texture coding parameters for all the macrobocks in the ith packet, and $B_{Si}(\mu_{Si})$ and $B_{Ti}(\mu_{Ti})$ the corresponding total number of bits used to encode these partitions. Let us denote by θ_i the embedding level for the ith packet, and the total number of bits used to encode the texture partition in the ith packet is denoted by $B'_{T_i}(\mu_{T_i}, \theta_i)$.

B. Channel model

In wireless channels, harsh conditions often present with high bit error rates (BER) (on the order of 10^{-1} to 10^{-3} BERs), thus the channel coding is needed to bring the aggregated BER down to a level so that the error resilient tools at the decode can be effective. In this work, we apply channel coding separately on shape and texture partitions. Since there are a number of channel coding approaches available, without loss of generality, let us denote by r_{S_i} and r_{T_i}, respectively, the channel code rate for shape and texture of the ith packet, thus the total bit rate for the frame can be represented by

$$R = \sum_{i=1}^{I} \left[\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}} + \frac{B_{T_i}'(\mu_{T_i}, \theta_i)}{r_{T_i}} \right]. \qquad (7)$$

We assume the burst errors can be converted into random errors with pre-interleaving [11], thus the bit error is the focus. Let us denote by ρ_{Si} the probability of corruption of the ith shape data, and ρ_{Ti} the probability of corruption of the ith texture partition. Clearly,

$$\rho_{S_i} = 1 - (1 - p_e)^{\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}}}, \text{ and } \rho_{T_i} = 1 - (1 - p_e)^{\frac{B_{T_i}'(\mu_{T_i}, \theta_i)}{r_{T_i}}}, \qquad (8)$$

where p_e is the BER at the decoder, which is usually smaller than the channel bit error rate. It is observable from Eq. (8) that the value of ρ_{Si} highly depends on $B_{Si}(\mu_{Si})$ and r_{S_i}, that is, either increasing source coding bit rate or increasing channel coding bit rate will cause the increase of the probability of partition error.

C. Expected distortion

We assume that the transmitter knows the channel condition. Thus, the distortion at the receiver is a random variable. Let $E[D_i]$ represents the expected distortion at the receiver for the ith packet. So,

$$E[D_i] = (1-\rho_{Si})(1-\rho_{Ti})E[D_{R,i}] + (1-\rho_{Si})\rho_{Ti}E[D_{LT,i}] + \rho_{Si}(1-\rho_{Ti})E[D_{LS,i}] + \rho_{Si}\rho_{Ti}E[D_{L,i}], \qquad (9)$$

where $E[D_{R,i}]$ is the expected distortion for the ith packet if both the shape and texture partitions are received correctly at the decoder, $E[D_{LT,i}]$ 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 the partitions are corrupted. Clearly, $E[D_{R,i}]$ depends only on the source coding parameters for the packet, while $E[D_{LT,i}]$, $E[D_{LS,i}]$ and $E[D_{L,i}]$ may also depend on the embedding level and the decoder concealment strategy.

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. The only assumption we make is that the concealment strategy is also known at the encoder. In our experimental results we use the expected mean squared error (MSE), as is commonly done in the literature [3].

Now, the optimization problem (6) can be rewritten as

$$\underset{\{\mu_{S_i}, \mu_{T_i}, r_{S_i}, r_{T_i}, \theta_i\}}{\text{Minimize}} E[D_{tot}], \qquad (10)$$

$$\text{subject to: } \sum_{i=1}^{I} \left[\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}} + \frac{B_{T_i}'(\mu_{T_i}, \theta_i)}{r_{T_i}} \right] \leq R_{budget}.$$

IV. OPTIMAL SOLUTION

In this section, we present an optimal solution for problem (10). We use the Lagrange multiplier method to relax the bit rate constraint. The relaxed problem can then be solved using a shortest path algorithm.

The Lagrangian relaxation method leads to a convex hull approximation to the constrained problem (10). Let U be the set of all possible decision vectors u_i for the ith packet

(i=1,2, …,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

$$J_\lambda(u)=E[D_{tot}]+\lambda R=\sum_{i=1}^{I}\left\{E[D_i]+\lambda[\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}}+\frac{B_{T_i}'(\mu_{T_i},\theta_i)}{r_{T_i}}]\right\},\quad(11)$$

where λ is the Lagrange multiplier. It can easily be derived from [17] that if there exists a pair λ^* such that $u^*=\arg[\min_u J_{\lambda^*}(u)]$, which leads to $R=R_{budget}$, then u^* is also an optimal solution to (10). Therefore, the task of solving (10) is converted into an easier one, which is to find the optimal solution to the unconstrained problem

$$\text{Min}\sum_{i=1}^{I}\left\{E[D_i]+\lambda[\frac{B_{S_i}(\mu_{S_i})}{r_{S_i}}+\frac{B_{T_i}'(\mu_{T_i},\theta_i)}{r_{T_i}}]\right\}.\quad(12)$$

Most decoder concealment strategies introduce dependencies among packets. For example, if the concealment algorithm uses the motion vector of the above macroblock 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 the generality, we assume that the concealment strategy will cause the current packet to depend on its previous a packets ($a\geq 0$). To implement the algorithm for solving the optimization problem (12), we define a cost function $G_k(u_{k-a},\dots,u_k)$, which represents the minimum total cost, delay and distortion up to and including the kth packet, given that $u_{k-a},\dots,u_k$ are decision vectors for the ($k-a$)th to kth packets. Therefore, $G_I(u_{I-a},\dots,u_I)$ represents the minimum total cost, delay and distortion for all the packets of the frame, and thus

$$\min_u J_\lambda(u)=\min_{u_{I-a},\dots,u_I} G_I(u_{I-a},\dots,u_I).\quad(13)$$

The key observation for deriving an efficient algorithm is the fact that given $a+1$ decision vectors u_{k-a-1}, …, u_{k-1} for the ($k-a-1$)th to ($k-1$)th 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, …, u_{k-a-2}. This is true since the cost function can be expressed recursively as

$$G_k(u_{k-a},\dots,u_k)=\min_{u_{k-a-1},\dots,u_{k-1}}\{G_{k-1}(u_{k-a-1},\dots,u_{k-1})+E[D_k]+\lambda[\frac{B_{S_k}(\mu_{S_k})}{r_{S_k}}+\frac{B_{T_k}'(\mu_{T_k},\theta_k)}{r_{T_k}}],\quad(14)$$

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) [17]. The computational complexity of the algorithm is $O(I\times|U|^{a+1})$ ($|U|$ 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 which has exponential computational complexity.

V. EXPERIMENTAL RESULTS

In this section, experiments are designed to demonstrate the advantages of using data hiding, and using the hybrid scheme that joins source coding and channel coding. We first present the implementation details such as error concealment and expected distortion calculations, and then report the experimental results.

A. Implementation details

The error concealment strategy in our simulations is described as the follows: If only one partition is corrupted, it can be recovered from the other partition, for example, the texture data can be recovered from the motion vector stored in the shape partition., and the shape data can be recovered from the data embedded in the texture DCT coefficient. However, if the whole packet is corrupted, motion vectors of neighbor macroblocks will be used for data recovery.

The way to calculate the expected end-to-end distortion is almost the same as the method described in [3]. The only difference comes from the inclusion of the data embedding. In the following, we show an example of how to calculate the expected pixel value for the case that shape and texture are both intra-coded, details for other cases are omitted here due to lack of space. Let us denote by $\tilde{s}_n^j$ ($\tilde{s}_n^j$=0 for transparent or 1 for opaque block) and $\tilde{t}_n^j$ the corresponding shape and texture component of reconstructed pixel j in VOP n, then,

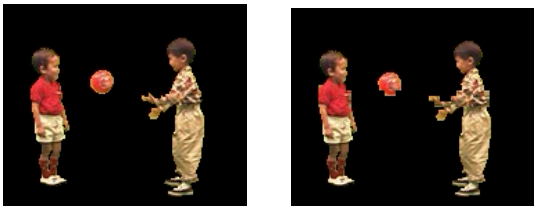
$$\begin{aligned}E[\tilde{s}_n^{j_s}\tilde{t}_n^{j_t}]&=(1-\rho_{S_i})(1-\rho_{T_i})\hat{s}_n^{j_s}\hat{t}_n^{j_t}+(1-\rho_{S_i})\rho_{T_i}\hat{s}_n^{j_s}E[\tilde{t}_{n-1}^{m_t}]+\\&\rho_{S_i}(1-\rho_{T_i})\xi_n+\rho_{S_i}\rho_{T_i}(1-\rho_{S_{i-1}})E[\tilde{s}_{n-1}^{m_s}\tilde{t}_{n-1}^{m_t}]+\\&\rho_{S_i}\rho_{T_i}\rho_{S_{i-1}}E[\tilde{s}_{n-1}^{j_s}\tilde{t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

where $\hat{s}_n^j$ and $\hat{t}_n^j$ are the encoder reconstructed shape and texture of the jth pixel, and pixel j in frame n is predicted by pixel m in frame $n-1$ if the motion vector is available, otherwise is predicted by pixel k if the concealment motion vector is available; the value of ξ_n depends on the embedding level, for example, when no information is embedded, $\xi_n=(1-\rho_{S_{i-1}})E[\tilde{s}_{n-1}^{k_s}\tilde{t}_{n-1}^{k_t}]+\rho_{S_{i-1}}E[\tilde{s}_{n-1}^{j_s}\tilde{t}_{n-1}^{j_t}]$, otherwise if the shape data is embedded, $\xi_n=\hat{s}_n'^{j_s}\hat{t}_n^{j_t}$, where $\hat{s}_n'^j$ is the encoder reconstructed shape from embedded information in texture. The subscript s and t of k_s, k_t, m_s, and m_t in above equations are used to distinguish shape from texture, because the motion vector or concealment motion vector of shape could be different from that of texture.

To facilitate adaptive embedding, we propose five embedding levels, 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 vector, and lossy shape. The embedding mode itself is also embedded. Here, the lossy shape means a lower-resolution 4x4 bitmap of the original 16x16 BAB.

Our simulations are based on MPEG-4 VM18.0 [16]. The available Intra mode quantizers are of step size 2, 4, 6, 8, 10, 14, 18, 24 and 30, and the available Inter mode 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 INTER mode. 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 Inter-mode shape coding has not been considered here because it violates the assumption that each packet is independently decodable.

(a) PSNR=*37.31dB*, Rate=5758 (b) PSNR=*23.08dB*



(c) PSNR=*37.19dB*, Rate=*5832* (d) PSNR=*25.54dB*

Figure 3. Reconstructed the 16th "Children" frame

B. Experimental results

We first demonstrate the advantage of the scheme of "hiding shape/motion in texture" with a simple example shown in Fig. 3. We encode the "Children sequence" and transmit it over a noisy wireless channel. During the transmission, the shape partition of the 3rd packet (3rd row of macroblocks) of the 16th frame (as shown in Fig. 3(a)) is corrupted, and thus the whole packet is discarded and a concealment strategy is called to recover the data by using its neighbor's motion vector (as shown in Fig. 3(b)). Figures 3(c) and 3(d) show the cases corresponding to 3(a) and 3(b) when the proposed data hiding scheme is applied. Clearly, the embedding of shape and motion information only increase the bit rate slightly and cause a very slight reduces in texture quality. However, the embedded information is very helpful in improving the concealed image quality. After the embedded lossy shape being extracted and used to recover the lost shape, the PSNR of the reconstructed image is increased by up to *2.5dB*.

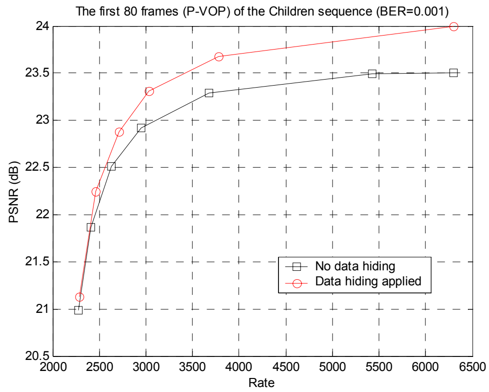


Figure 4. Comparison the method using data hiding and the method without data hiding

In the next experiment, we consider the advantage of using adaptive data hiding scheme in the 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}. Two approaches with optimal source coding are compared in Fig. 4, one uses adaptive data hiding and the other does not. As expected, the method using adaptive data hiding outperformed the other, however, it starts to gain until the bit rate is larger than a certain number (in this case rate>2400). At the low bit rate end, the two approaches have the same performance. This is reasonable because when bit rate is low, the texture data are coded in a coarse visual quality and generates not enough non-zero DCT coefficients for embedding purpose. When the bit rate increases, there is more slots in the DCT coefficients for embedding, and thus show the benefits of using data hiding. In the experiment, we also found that the gain of using adaptive data hiding decreases when a channel with smaller BERs are used, which is easy for understanding.

In the third experiment, we consider the advantage of joint source-channel coding. We compare three approaches, (1) EEP method, where the shape and texture partition within a packet are equally protected (by channel coding) but different packet is allowed to be unequally protected; (2) UEP method, where the shape and texture partitions are unequally protected by channel coding; (3) 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. The data hiding is not used in EEP and UEP methods, but source coding parameters are optimized for all these methods. In the simulation, we use an 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, and the channel parameter is defined as $SNR = \alpha \frac{E_b}{N_0}$. The bit error rates for the same simulation can be found in [18].

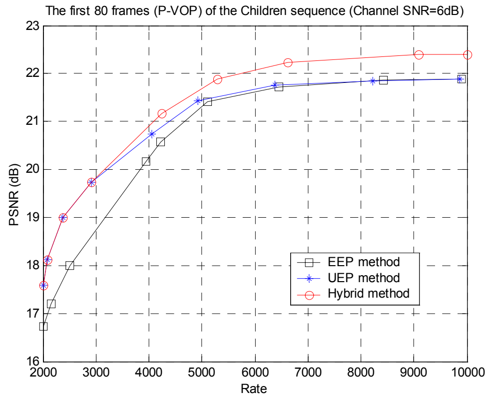


Figure 5. R-D curves for Children sequence

We encode the first *80* frames of the "Children" sequence and transmit them over the simulated wireless channels with SNR=*6*dB. The experimental results for the three approaches are shown in Fig. 5. UEP method outperformed EEP at lower bit rate (when rate is lower than *6000*) because it can use 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 partition corrupted become larger and larger, 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. An interesting observation is that data hiding works well when bit rate is higher enough that there is enough DCT coefficients available for embedding shape and motion information. This way, the hybrid method inherits the virtue from both UEP and data hiding, and thus makes it work well for all ranges of the bit rate. In the experiments, we found that the advantage of proposed method reduces when the channel condition gets better.

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 purpose. The level of embedding is jointly optimized with the selection of source coding parameters and channel coding rates to achieve the minimum expected distortion. Experimental results indicate that the hybrid scheme inherits virtues from both data hiding and joint source-channel coding, which makes it significantly outperform other approaches.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Fan Zhai of Northwestern University for his helpful discussion.

REFERENCES

[1] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error prone networks", *IEEE J. Select. Areas Commun.*, Vol. 18, NO. 6, pp. 952-965, June 2000.

[2] R. Zhang, S.L.Regunathan, K.Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience", *IEEE J. on Selected Areas in Communications*, Vol. 18, NO. 6, pp..966-976, June 2000.

[3] H. Wang, A. K. Katsaggelos, "Robust network-adaptive object-based video encoding", in *Proc IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004.

[4] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding – a survey", *Proc. IEEE*, Vol. 87, No. 7, July 1999. pp. 1062-1078.

[5] J. Song and K. J. R. Liu, "A data embedded video coding scheme for error-prone channels", *IEEE Trans. Multimedia,* Vol. 3, No. 4, Dec. 2001. pp. 415-423.

[6] P. Yin, B. Liu, and H. H. Yu, "Error concealment using data hiding", in *Proc.IEEE Int. Conf. on Acoustics, Speech, and Signal Processing,* 2002. pp. 729-732.

[7] 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.

[8] C. S. Lu, "Wireless multimedia error resilience via a data hiding technique", in *Proc. Int. Workshop Multimedia & Expo*, St. Thomas, USA, Dec. 2002.

[9] L. Kang and J. Leou, "An error resilient coding scheme for H.264 video transmission based on data embedding", In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004.

[10] 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, Oct. 1999. pp. 1734-1750.

[11] B. Girod and N. Farber, "Wireless video", in *Compressed video over networks*, M. –T. Sun and A. R. Reibman, Eds. New York, Marcel Dekker Inc., 2000.

[12] C.B. Adsumilli, M. Carli, M.C.Q. de Farias, S.K. Mitra, "A hybrid constrained unequal error protection and data hiding scheme for packet video transmission", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2003, Hong Kong.

[13] H. Wang, G. M. Schuster, and A. K. Katsaggelos, "Object-based video compression scheme with optimal bit allocation among shape, motion and texture", in *Proc. IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.

[14] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia", *IEEE Trans. Image Processing*, Vol. 6, pp. 1673-1687, Dec. 1997.

[15] M. Wu, H. Yu and A. Gelman, "Multi-level data hiding for digital image and video", *SPIE*, Vol. 3854, 1999.

[16] MPEG-4 video VM 18.0, ISO/IEC JTC1/SC29/WG11 N3908, Pisa, Jan. 2001.

[17] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion based video compression: optimal video frame compression and object boundary encoding*, Kluwer Academic Publishers, 1997.

[18] 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 *Proc. 41st Allerton Conf. Communication, Control, and Computing*, Oct. 2003.