JOINT OBJECT-BASED VIDEO ENCODING AND POWER MANAGEMENT FOR ENERGY EFFICIENT WIRELESS VIDEO COMMUNICATIONS

Haohong Wang, Yiftach Eisenberg, Fan Zhai, and Aggelos. K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
Email: {haohong, yeisenbe, fzhai, aggk}@ece.northwestern.edu

ABSTRACT

In this paper, we consider dynamic resource allocation for object-based wireless video communications. In object-based video coding, a video frame is comprised of objects that are described by their shape as well as their texture. By jointly considering source coding, error concealment, and transmission power management at the physical layer, the proposed framework minimize the expect distortion at the receiver for given energy and delay constraints. In order to provide unequal error protection for the shape and texture information, a new video packetization scheme is proposed. Experimental results indicate that the proposed unequal error protection schemes significantly outperform equal error protection methods.

1. INTRODUCTION

A major problem in video communications is how to efficiently allocate communication resources in order to achieve the best video delivery quality. In wireless video communications, mobile devices normally have a limited battery supply. This limited energy is consumed in the processing, transmission, and displaying of the video sequence. In this paper, we consider how the average transmission power used by a modulation scheme directly affects the probability of packet loss, and therefore affects the received video quality.

Since video packets may have different resource allocation requirements and may have different contribution to video quality, it is natural to use unequal error protection (UEP) when transmitting video data. The idea is to allocate more resources to the parts of the video sequence that have a greater impact on video quality, while spending less resources on parts that are less significant. In [1], a priority encoding transmission scheme is proposed to allow a user to set different priorities of error protection for different segments of the video stream. In [2], an unequal error protection scheme was proposed for layered video coding which provides an optimal bit allocation between source coding and channel coding. In [3], the trade-off between transmission energy consumption and video quality for wireless video communications is studied, where the goal is to minimize the energy needed to transmit a video sequence with an acceptable level of video quality and tolerable delay. By assuming that the transmitter knows the relationship between the transmission power and the probability of the packet loss, the transmission power can be dynamically adjusted to control the level of protection provided for each packet.

Due to the ever-increasing demand for interactive multimedia applications, object-based video has become an important research topic in the field of visual communications. Object-based video is based on the concept of encoding arbitrarily shaped video objects, which are described by their shape and texture. A rate-distortion optimal video encoding scheme was proposed in [4] for object-based video, which enables the optimal bit allocation among shape, texture and motion. In [5], the source coding, packet loss during transmission and error concealment at the decoder are jointly considered, and a robust network-adaptive encoding scheme for object-based video is proposed. In this paper, an optimal unequal error protection scheme is proposed for object-based video communications over wireless channels.

2. PROBLEM FORMULATION

The problem at hand is to choose the source coding and transmission parameters, so as to minimize the total expected distortion, given transmission energy and delay constraint. This objective can also be represented by:

$$\text{Minimize } E[D_{tot}], \text{ Subject to}$$
$$E_{tot} \le E_{\max}, \text{ and } T_{tot} \le T_{\max}, \tag{1}$$

where $E[D_{tot}]$ is the expected total distortion for the frame, E_{tot} is the total transmission energy, T_{tot} is the total transmission delay, E_{max} is the maximum allowable transmission energy, and T_{max} is the maximum amount of time that can be used to transmit the entire frame. We assume that the energy and delay constraints, E_{max} and T_{max}, can vary from frame to frame but are known constants in (1).

A. System model

We consider an MPEG-4 compliant object-based video application, where the video is encoded using different algorithms for shape and texture. As mentioned in [4], compared to texture data, the shape data requires relatively fewer bits to encode but has a very strong impact on the video quality. Therefore, it is natural to

0-7803-8554-3/04/$20.00 ©2004 IEEE.

imagine that the unequal error protection scheme for shape and texture may provide improved performance over an equal error protection scheme. However, implementing an unequal error protection scheme is not straightforward because in the MPEG-4 video packet syntax, the shape and texture data are placed in the same packet (using a combined packetization scheme). If data partitioning is enabled, a motion marker is placed between the shape and texture data for resynchronization. One way to enable unequal error protection is to use a separated packetization scheme, where the shape and texture are placed into separate packets. In a similar way as proposed in [6, 7], we insert an adaptation layer between the MPEG-4 video application and the network, which can reorganize the MPEG-4 compressed bit stream into separate shape packets and texture packets. In addition, the adaptation layer can optimally add some forward error protections to those packets. Figure 1 shows the architecture of the proposed video transmission system. For a wireless network using an H.223 MUX [6], we simply replace the standard adaptation layer in the H.223 multiplexing protocol with our new layer. At the receiver side, the adaptation layer merges the shape and texture packets and makes the output bit stream compatible with the MPEG-4 syntax.

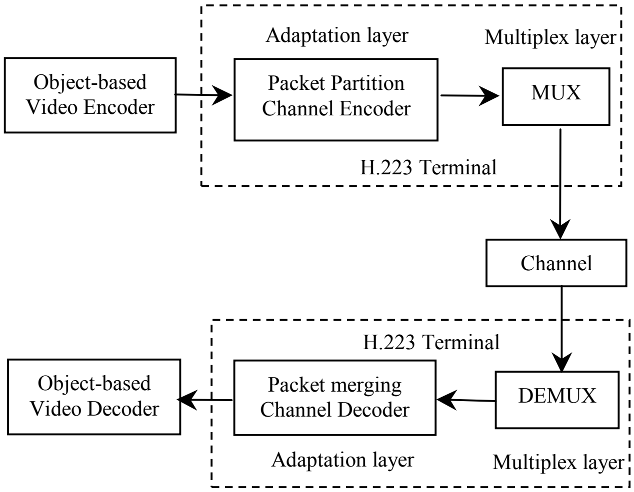

Figure 1. System block diagram

Unless otherwise stated, the separated packetization scheme is used as the default packetization scheme. The coded video frame is divided into 16x16 macro blocks, which are numbered in scan line order and divided into groups called slices. For each slice, there is a corresponding shape packet and a corresponding texture packet. Let I be the number of slices in the given frame and i the slice index. For each macro block, coding parameters are specified for both shape and texture. We use μ_{S_i} and μ_{T_i}, respectively, to denote the coding parameters for all macro blocks in the ith shape and the ith texture packet, and use $B_{S_i}(\mu_{S_i})$ and $B_{T_i}(\mu_{T_i})$, respectively, to denote the corresponding encoding bit rates of these packets. It is important to point out that each packet is independently decodable in our system; that is, each packet has enough information for decoding and is independent of other packets in the same frame. This guarantees that a lost packet will not affect the decoding of other packets in the same frame. Of course, errors may propagate from one frame to the next due to predictive encoding and motion compensation.

B. Channel model

The wireless channel is modeled as a narrow-band block fading channel with additive white Gaussian noise [8]. We assume the channel fading for each packet is independent, and can be modeled by a random variable H. The noise process is modeled as an additive white Gaussian noise process with power spectral density N_0. We assume that H stays fixed during the transmission of a packet, and varies randomly between packets. Each realization h of H is chosen according to the *a priori* distribution $f_H(h|\theta)$, where θ is the channel state information (CSI) parameters known by the transmitter. We assume a packet is dropped if the capacity of the channel realization during that block is less than or equal to the information rate.

C. Transmission energy and transmission delay

Let us denote by R_{S_i} and R_{T_i} the transmission rate for the ith shape and texture packet, and P_{S_i} and P_{T_i} the corresponding transmission power, thus the total energy used to transmit all the packets in a frame is

$$E_{tot} = \sum_{i=1}^{I}[\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}}P_{S_i} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}}P_{T_i}]. \tag{2}$$

Notice in (2) that the energy is a function of the number of bits used to encode each shape and texture packet, the rate at which each packet is transmitted and the power used to transmit each packet. This indicates that by jointly adapting the source coding and communication parameters, we can reach the best video quality with a given energy constraint. The transmission delay can be represented by

$$T_{tot} = \sum_{i=1}^{I}[\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}}]. \tag{3}$$

D. Expected Distortion

We assume that the transmitter only knows the probability with which a packet has arrived at the receiver. Let us denote by ρ_{S_i} and ρ_{T_i} the probability of loss for the ith shape and texture packet. Here we assume $\sqrt{H}$ is Rayleigh distributed, and assume $\theta = E[H]$. Thus, the probability of packet loss can be derived as

$$\rho_{S_i} = 1 - \exp[-\frac{N_0 W (2^{R_{S_i}/W} - 1)}{\theta \cdot P_{S_i}}] \quad , \tag{4}$$

where W is the channel bandwidth. The distortion at the receiver is a random variable. Let $E[D_i]$ represent the expected distortion at the receiver for the ith slice. Thus,

$$\begin{aligned} E[D_i] = (1-\rho_{S_i})(1-\rho_{T_i})E[D_{R,i}] + (1-\rho_{S_i})\rho_{T_i}E[D_{LT,i}] + \\ \rho_{S_i}(1-\rho_{T_i})E[D_{LS,i}] + \rho_{S_i}\rho_{T_i}E[D_{L,i}], \end{aligned} \tag{5}$$

where $E[D_{R,i}]$ is the expected distortion for the ith slice if both the shape and texture packets are received correctly at the decoder, $E[D_{LT,i}]$ is the expected distortion if the texture packet is lost, $E[D_{LS,i}]$ is the expected distortion if the shape packet is lost, and $E[D_{L,i}]$ is the expected distortion if both the shape and texture packets are lost. Clearly, $E[D_{R,i}]$ depends only on the source coding parameters for the ith packet, while $E[D_{LT,i}]$, $E[D_{LS,i}]$ and $E[D_{L,i}]$ depend on the concealment strategy used at the decoder. See [5] for more details on calculating the expected distortion for object-based video communications.

3. OPTIMAL SOLUTION

In this section, we present an optimal solution for problem (1). We use the Lagrange multiplier method to relax the cost and delay constraints. The Lagrangian relaxation method leads to a convex hull approximation to the constrained problem (1). Let U be the set of all possible decision vectors u_i for the ith slice ($i=1,2, ..., I$), where $u_i = (\mu_{S_i}, \mu_{T_i}, R_{S_i}, R_{T_i}, P_{S_i}, P_{T_i})$. We first define a Lagrangian cost function

$$\begin{aligned} & J_{\lambda_1,\lambda_2}(u) = E[D_{tot}] + \lambda_1 E_{tot} + \lambda_2 T_{tot} \\ & = \sum_{i=1}^{I} \left\{ E[D_i] + \lambda_1 [\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} P_{S_i} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}} P_{T_i})] + \lambda_2 [\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}}] \right\}, \end{aligned} \tag{6}$$

where λ_1 and λ_2 are the Lagrange multipliers. It can easily be derived from [9] that if there exists a pair λ_1^* and λ_2^* such that $u^* = \arg[\min_u J_{\lambda_1^*,\lambda_2^*}(u)]$, which leads to $E_{tot}=E_{max}$ and $T_{tot}=T_{max}$, then u^* is also an optimal solution to (1). Therefore, the task of solving (1) is converted into an easier task, which is to find the optimal solution to the unconstrained problem

$$\begin{aligned} \min_{\mu_{S_i}, \mu_{T_i}, R_{S_i}, R_{T_i}, P_{S_i}, P_{T_i}} \sum_{i=1}^{I} \Bigg\{ E[D_i] + \lambda_1 [\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} P_{S_i} + \\ \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}} P_{T_i})] + \lambda_2 [\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}}] \Bigg\}. \end{aligned} \tag{7}$$

Most decoder concealment strategies introduce dependencies between slices. Without loss of generality, we assume that the concealment strategy will cause the current slice to depend on its previous a slices ($a \geq 0$). To implement the algorithm for solving the optimization problem (7), we define a cost function $G_k(u_{k-a}, ..., u_k)$, which represents the minimum total cost, delay and distortion up to and including the kth slice, given that $u_{k-a}, ..., u_k$ are decision vectors for the (k-a)th to kth slices. Therefore, $G_I(u_{I-a}, ..., u_I)$ represents the minimum total cost, delay and distortion for all the slices of the frame, and thus

$$\min_u J_{\lambda_1,\lambda_2}(u) = \min_{u_{I-a},...,u_I} G_I(u_{I-a},...,u_I). \tag{8}$$

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 $(k\text{-}a\text{-}1)$th to $(k\text{-}1)$th slices, and the cost function $G_{k-1}(u_{k-a-1}, ..., u_{k-1})$, 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

$$\begin{aligned} G_k(u_{k-a},...,u_k) = \min_{u_{k-a-1},..,u_{k-1}} \{ G_{k-1}(u_{k-a-1},...,u_{k-1}) + \\ \lambda_1 \cdot [\frac{B_{S_i}(\mu_{S_i})}{R_{S_i}} P_{S_i} + \frac{B_{T_i}(\mu_{T_i})}{R_{T_i}} P_{T_i}] + \lambda_2 [\frac{B_{S_k}(\mu_{S_k})}{R_{S_k}} + \frac{B_{T_k}(\mu_{T_k})}{R_{T_k}}] + E[D_k] \Bigg\}. \end{aligned} \tag{9}$$

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) [9]. 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.

4. EXPERIMENTAL RESULTS

The main objective of the experiments presented here is to compare three error protection schemes: (1) UEP-UST, an unequal error protection scheme using the separated packetization scheme, where the shape and texture data are placed in separate packets and therefore can be transmitted over different service channels; (2) UEP-EST, an unequal error protection scheme using combined packetization where the packets containing both shape and texture data can be transmitted over different service channels; (3) EEP, an equal error protection scheme using combined packetization, where all the packets are transmitted at the same fixed power level.

Our simulations are based on MPEG-4 VM18.0 [10]. In our experiments, the shape is coded in Intra mode because the Inter-mode shape coding in MPEG-4 violates the assumption that each packet is independently decodable. We encode the QCIF "Children" sequence at *10* fps, and set N_0W/θ equals to *6*W, *W=5*MHz, and *R=200*kbits/s for (4). In the experiments, we are looking at power adaptation assuming fixed transmission rate. We use six power levels corresponding to *1, 2, 3, 4, 6*, and *10* Watts. Figure 2 shows the energy-distortion (E-D) curves for the three error protection schemes described above. Note that each point on the E-D curve of the EEP system is obtained by trying each fixed power level and choosing the one that achieves the best quality for each energy constraint. As expected, by jointly adapting the source coding parameters along with the selection of the transmission power, the UEP approaches outperform the EEP approach. In addition, UEP-UST outperformed UEP-EST because the former approach has increased flexibility by providing unequal protection for shape and texture.

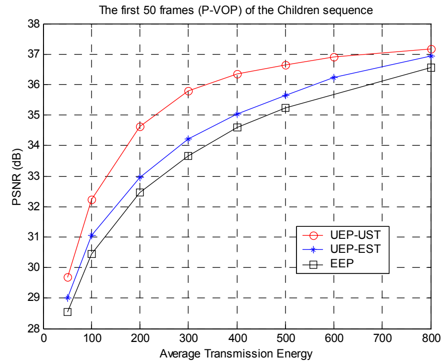

Figure 2. Comparison of Cost-Distortion curves for the UEP-UST, UEP-EST, and EEP schemes

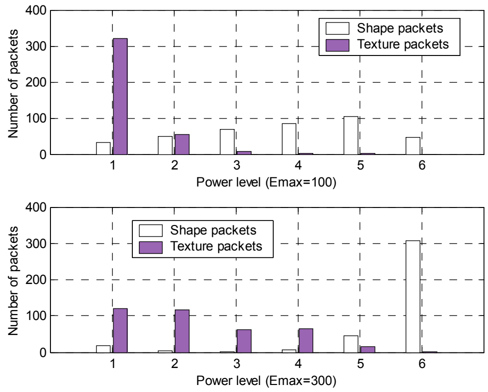

Figure 3. Distribution of shape and texture packets in UEP-UST system

Figures 3 show the distributions of the shape and texture packets among the six transmission channels for the UEP-UST approach when the cost constraint E_{max} equals *100* and *300*. The results indicate that the shape packets are better protected than texture packets, i.e., they are transmitted at higher power levels. During the increasing of E_{max} from *100* to *300*, the shape packets are more frequently selected than the texture packets to be transmitted through the higher-power channels. This is because shape packets have lower bit consumption but strong impact on the video quality. As shown in Fig. 3, when E_{max}*=300,* at least *80%* of the shape packets are transmitted over the highest power channel, while over *60%* of texture packets are transmitted over the two lowest power channels. In other words, the optimization process chooses to allocate more protection to shape, because it greatly impacts the end-to-end distortion.

5. CONCLUSIONS

In this paper, we have proposed an optimal unequal error protection scheme for object-based wireless video communications. The optimization is achieved by jointly considering source coding, transmission power management and error concealment. We applied unequal error protection on shape and texture data. Experimental results indicated that the unequal error protection schemes have significant advantages over equal error protection methods.

REFERENCES

[1] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission", *IEEE. Trans. Information Theory*, Vol. 42, No. 6, pp. 1737-1744, Nov. 1996.

[2] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source/Channel Coding for SNR Scalable Video Processing", *IEEE Transactions on Image Processing*, Vol 11, No. 9, pp. 1043 –1052, Sept 2002.

[3] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications", *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 12, No. 6, pp. 441-424, June 2002.

[4] 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.

[5] 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.

[6] W. R. Heinzelman, M. Budagavi, and R. Talluri, "Unequal error protection of MPEG-4 compressed video", in *Proc Proceedings of the International Conference on Image Processing*, pp. 530-534, Oct. 1999.

[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 WCNC'2000*, pp. 1243-1247, Chicago, Sept. 2000.

[8] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh.* Tech., vol. 43, No. 2, pp. 359-378, May 1994.

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

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