Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 657032, 11 pages
doi:10.1155/2008/657032

Research Article

Joint Video Summarization and Transmission Adaptation for Energy-Efficient Wireless Video Streaming

Zhu Li,[1] Fan Zhai,[2] and Aggelos K. Katsaggelos[3]

[1] *Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong*
[2] *DSP Systems, ASP, Texas Instruments Inc., Dallas, TX 75243, USA*
[3] *Department of Electrical Engineering & Computer Science (EECS), Northwestern University, Evanston, IL 60208, USA*

Correspondence should be addressed to Zhu Li, zhu.li@ieee.org

Received 13 October 2007; Accepted 25 February 2008

Recommended by Jianfei Cai

The deployment of the higher data rate wireless infrastructure systems and the emerging convergence of voice, video, and data services have been driving various modern multimedia applications, such as video streaming and mobile TV. However, the greatest challenge for video transmission over an uplink multiaccess wireless channel is the limited channel bandwidth and battery energy of a mobile device. In this paper, we pursue an energy-efficient video communication solution through joint video summarization and transmission adaptation over a slow fading wireless channel. Video summarization, coding and modulation schemes, and packet transmission are optimally adapted to the unique packet arrival and delay characteristics of the video summaries. In addition to the optimal solution, we also propose a heuristic solution that has close-to-optimal performance. Operational energy efficiency versus video distortion performance is characterized under a summarization setting. Simulation results demonstrate the advantage of the proposed scheme in energy efficiency and video transmission quality.

1. INTRODUCTION

The rapid increase in channel bandwidth brought about by new technologies such as the present third-generation (3G), the emerging fourth-generation (4G) wireless systems, and the IEEE 802.11 WLAN standards is enabling video streaming in personal communications and driving a wide range of modern multimedia applications such as video telephony and mobile TV. However, transmitting video over wireless channels from mobile devices still faces some unique challenges. Due to the shadowing and multipath effect, the channel gain varies over time, which makes reliable signaling difficult. On the other hand, a major limitation in any wireless system is the fact that mobile devices typically depend on a battery with a limited energy supply. Such a limitation is especially of concern because of the high energy consumption rate for encoding and transmitting video bit streams. Therefore, how to achieve reliable video communications over a fading channel with energy efficiency is crucial for the wide deployment of wireless video-based applications.

Energy-efficient wireless communications is a widely studied topic. For example, a simple scheme is to put the device into sleep mode when not in use, as in [1, 2]. Although the energy consumption on circuits is being driven down, as the VLSI design and integrated circuit (IC) manufacturing technologies advance, the communication energy cost is lower bounded by information theory results. In [3], the fundamental tradeoff between average power and delay constraint in communication over fading channels is explored and characterized. In [4], optimal power control schemes for communication over fading channels are developed. In [5, 6], optimal offline and near optimal online packet scheduling algorithms are developed to directly minimize energy usage in transmitting a given amount of information over fading channels with certain delay constraints.

Video streaming applications typically have different quality of service (QoS) requirements with respect to packet loss probability and delay constraints, which differentiate them from traditional data transmission applications. Approaches of cross-layer optimization of video source coding/adaptation and communication decisions have been

widely adopted. Taking advantage of the specific characteristics of video source and jointly adapting video source coding decisions with transmission power, modulation and coding schemes can achieve substantial energy efficiency compared with nonadaptive transmission schemes. Examples of this type of work are reported in [7–11]. In those studies, source-coding controls are mostly based on frame and/or macroblock (MB) level coding mode and parameter decisions.

When both bandwidth and energy are severely limited for video streaming, sending a video sequence over with severe distortion is not desirable. Instead, we consider joint video summarization and transmission approaches to achieve the required energy efficiency. Video summarization is a video adaptation technique that selects a subset of video frames from the original video sequence based on some criterion, e.g., some newly defined frame loss distortion metric [12], specified by the user. It generates a shorter yet visually more pleasing sequence than traditional technologies that usually focus on the optimization of quantization parameters (QP) [12], which can have serious artifacts at reconstruction at very low bit rates.

Video summarization may be required when a system is operating under limited bandwidth conditions, or under tight constraints in viewing time or storage capacity. For example, for a remote surveillance application in which video must be recorded over long lengths of time, a shorter version of the original video sequence may be desirable when the viewing time is a constraint. Video summarization is also needed when important video segments must be transmitted to a base station in real time in order to be viewed by a human operator. Examples of the video summarization and related shot segmentation work can be found in [13–18], where a video sequence is segmented into video shots, and then one or multiple key frames per shot are selected based on certain criterion for the summary.

In this work, we consider the application of video summarization over wireless channels. In particular, we consider using the scheme of video summarization together with other adaptations including transmission power and modulations to deal with problems in uplink wireless video transmission arising from the severe limitation in both bandwidth and transmission energy. Since the summarization process inevitably introduces distortion, and the summarization "rate" is related to the conciseness of the summary, we formulated the summarization problem as a rate-distortion optimization problem in [12], and developed an optimal solution based on dynamic programming. We extended the formulation to deal with the situation where bit rate is used as summarization rate in [19]. In [20, 21], we formulated the energy-efficient video summarization and transmission problem as an energy-summarization distortion optimization problem; the solution of which is found through jointly optimizing the summarization and transmission parameters/decisions to achieve the operational optimality in energy efficiency. In this paper, we further extend the work in [20, 21] to consider the maximum frame drop distortion case for energy-efficient streaming. We also propose a heuristic solution, which is a greedy method that approximates well the performance of the optimal solutions.

The rest of the paper is organized as follows. In Section 2, we describe the assumptions on the communication over fading wireless channels and formulate the problem as an energy-summarization distortion optimization problem. In Section 3, we develop an optimal solution based on Lagrangian relaxation and dynamic programming, as well as a heuristic solution. In Section 4, we present simulation results. Finally, in Section 5 we draw conclusions and discuss the future work in this area.

2. ASSUMPTIONS AND PROBLEM FORMULATION

In this section, we describe the channel model used in this work, carry out delay analysis for video summary packets, and provide the problem formulations.

2.1. Wireless channel models and assumptions

In this work, we assume that the wireless channel can be modeled as a band-limited, additive white Gaussian noise (AWGN) channel with discrete time, and slow block fading. The output y_k is a function of the input x_k as

$$y_k = \sqrt{h_k}x_k + n_k, \tag{1}$$

where h_k is the channel gain for time slot k and n_k is the additive Gaussian noise with power spectrum density N. We assume that the channel gain stays constant for time T_c, the channel coherent time, and that the symbol duration T_s satisfies $T_s \ll T_c$, thus the channel is slow fading and there are many channel uses during each time slot. The variation of the channel state is modeled as a finite state Markov channel (FSMC) [22], which has a finite set of possible states, $H = \{h_1, h_2, \ldots, h_m\}$, and transitions every T_c second with probability given by the transition probability matrix $A = |a_{ij}|$, where $a_{ij} = \text{Prob}\{\text{transition from } h_i \text{ to } h_j\}$.

To reliably send R information bits over the fading channel in one channel use, the minimum power needed with optimal coding is given as [23]

$$P = N(2^{2R} - 1)/h, \tag{2}$$

where h represents the channel gain. Similarly to the analysis in [5], let $x = 1/R$ be the number of transmissions needed to send one bit over the channel; we can characterize the energy-delay tradeoff as E_b, energy per bit as a function of x as

$$E_b(x, h) = xP = xN(2^{2/x} - 1)/h. \tag{3}$$

Examples of the energy efficiency functions with different fading states are shown in Figure 1. The range of x in Figure 1 corresponds to the received signal-to-noise tatio (SNR) of 2.0 dB to 20 dB, a typical operating range for wireless communication. To send a data packet with B bits and deadline τ, assuming $\tau \gg T_c$, the number of transmissions available is equal to $2W\tau$, where W is the signaling rate. Then

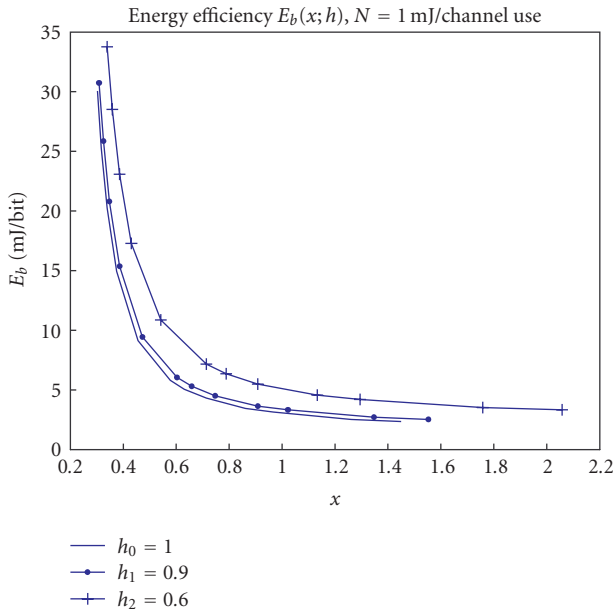


Figure 1: Energy-efficiency over fading channels.

the expected energy cost will be

$$E(B,\tau) = \mathbf{E}_H\{E_b(2W\tau/B,h)B \mid A,H,h_0\}. \quad (4)$$

In (4), the expectation $\mathbf{E}_H$ is with respect to all possible channel states, which are governed by an FSMC specified by the state set H, the transition probability matrix A, and the initial state h_0. The function in (4) can be implemented as a lookup table for a given channel model in simulations. A closed form solution may also be possible, under some optimal coding and packet scheduling assumptions. More details for a 2-state FSMC channel analysis can be found in the appendix.

2.2. Summarization and packet delay constraint analysis

Let a video sequence of n frames be denoted by $V = \{f_0, f_1, \ldots, f_{n-1}\}$ and its video summary of m frames by $S = \{f_{l_0}, f_{l_1}, \ldots, f_{l_{m-1}}\}$. Obviously, the video summarization process has an implicit constraint that $0 \le l_0 < l_1 < \cdots < l_{m-1} \le n-1$. Let the reconstructed sequence $V'_S = \{f'_0, f'_1, \ldots, f'_{n-1}\}$ be obtained by substituting missing frames with the most recent frame that is in the summary S, that is, $f'_k = f_{i=\max(l):\ \text{s.t. } l\in\{l_0,l_1,\ldots,l_{m-1}\},\ i\le k}$. Let the summarization rate be

$$R(S) = \frac{m}{n}, \quad (5)$$

taking values in $\{1/n, 2/n, \ldots, n/n\}$. The summarization distortion can be computed as the average frame distortion between the original sequence and the reconstructed sequence from the summary

$$D(S) = \frac{1}{n}\sum_{k=0}^{n-1} d(f_k, f'_k), \quad (6)$$

where $d(f_k, f'_k)$ is the distortion of the reconstructed frame f'_k and n is the number of frames in the video sequence. Various distortion metrics can be utilized here to capture the impact of frame-loss-induced distortion, $d(f_k, f'_k)$. In this work, we use the Euclidean distance of scaled frames in PCA space, as discussed in [12]. This is an effective metric that matches the perception of frame losses well.

In video summarization studies [24], we also found that in addition to the average frame loss distortion metric, the maximum frame loss distortion-based metric is also very effective in matching the subjective perception, especially the jerkiness in playback. Therefore, the video summarization distortion can also be defined as

$$D(S) = \max_k d(f_k, f'_k). \quad (7)$$

The loss of frames in high activity segments of video sequence will typically result in a large $D(S)$ in this case. The average (l_2) and maximum (l_∞) metrics for video summarization compliment each other in characterizing the distortion.

For the encoding of the video summary frames, we assume a constant Peak SNR (PSNR) or QP coding strategy, with frame bit budget B_{l_j} given by some rate profiler see, for example, [25]. Packets from different summary frames have different delay tolerances. Without loss of generality, we assume that the first frame of the original sequence, f_0, is always selected for the summary and intracoded with some B_0 bits. The delay tolerance τ_0 is determined by how much initial streaming delay is allowed in an application. For packets generated by the summary frame f_{l_j}, with $l_j > 0$, if the previous summary frame $f_{l_{j-1}}$ is decoded at time t_{j-1}, then the packet needs to arrive by the time $t_j = t_{j-1} + (l_j - l_{j-1})/F$, where F is the frame rate of the original video sequence. Therefore, the delay tolerance for frame f_{l_j} is $\tau_{l_j} = (l_j - l_{j-1})/F$. This is a simplified delay model, not accounting for minor variations in frame encoding and other delays. The energy cost to transmit a summary S of m frames is therefore given by

$$E(S) = \sum_{k=0}^{m-1} E(B_{l_k}, \tau_{l_k}) = E(B_0,\tau_0) + \sum_{k=1}^{m-1} E(B_{l_k}, \tau_{l_k}), \quad (8)$$

where B_{l_k} is the number of bits needed to encode summary frame f_{l_k}, and τ_{l_k} is the delay tolerance for frame f_{l_k}.

There are tradeoffs between the summary transmission energy cost, $E(S)$, and the summarization distortion, $D(S)$. The more frames selected into the summary, the smaller the summarization distortion. On the other hand, the more frames in the summary, the more bits needed to be spent in encoding the frames, and the packet arrival pattern gets more dense, which can be translated into higher bit rate and smaller delay tolerance. The transmission of more bits with more stringent deadline can incur higher transmission energy cost.

In the next subsection, we will characterize the relationship between the summarization distortion and energy cost, and formulate the energy-efficient video summarization

and transmission problem as an energy-distortion (E-D) optimization problem.

2.3. Energy-efficient summarization formulations

The energy-efficient summarization problem can be formulated as a constrained optimization problem. For a given constraint on the summarization distortion, we need to find the optimal summary that minimizes the transmission energy cost, while satisfying the distortion constraint, $D_{\max}$. That is, the Minimizing Energy Optimal Summarization (MEOS) formulation is given by

$$S^* = \arg\min_S E(S), \quad \text{s.t. } D(S) \le D_{\max}. \tag{9}$$

We can also formulate the energy efficiency problem as a Minimizing Distortion Optimal Summarization (MDOS) problem. That is, for a given energy constraint, $E_{\max}$, we want to find the optimal summary that minimizes the summarization distortion:

$$S^* = \arg\min_S D(S), \quad \text{s.t. } E(S) \le E_{\max}. \tag{10}$$

The optimal solutions to the formulations in (9) and (10) can be achieved through Dynamic Programming (DP) for the maximum frame loss distortion case in (7), by exploiting the structure of the summarization problem. As for the average distortion metric case in (6), a convex hull optimal solution can be found via Lagrangian relaxation and DP, which are discussed in more detail in the next section.

3. SOLUTION ALGORITHMS

Solving the constrained problems in (9) and (10) directly is usually difficult due to the complicated dependencies and large searching space for the operating parameters. For the average distortion case, we introduce the Lagrange multiplier relaxation to convert the original problem into an unconstrained problem. The solution to the original problem can then be found by solving the resulting unconstrained problem with the appropriate Lagrange multiplier that satisfies the constraint. This gradient-based approach has been widely used in solving a number of coding and resource allocation problems in video/image compression [8, 26]. For the maximum distortion case, a direct DP solution can provide us with the optimal solution at polynomial computational complexity. Finally, we introduce a heuristic algorithm that approximates the E-D performance of the optimal solutions at a fraction of the computational cost.

3.1. Average distortion problems

Considering the MEOS formulation with the average distortion metric in (4), by introducing the Lagrange multiplier, the relaxed problem is given by

$$S^*(\lambda) = \arg\min_S \{E(S) + \lambda D(S)\}, \tag{11}$$

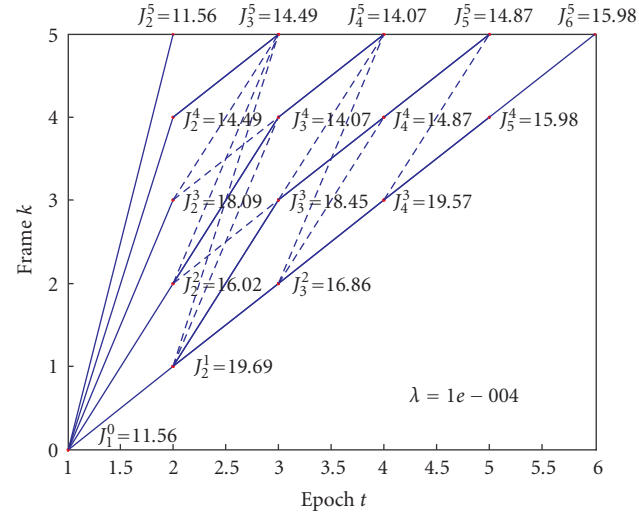


Figure 2: An example of DP trellis for the average distortion minimization problem.

in which the optimal solution S^* becomes a function of λ. From [27], we know that by varying λ from zero to infinity, we sweep the convex hull of the operational E-D function $E(D(S^*(\lambda)))$, which is also monotonic with respect to λ. Therefore, a bisection search algorithm on λ can give us the optimal solution within a convex hull approximation. In real-world applications, the E-D operational point sets are typically convex, and the optimal solution can indeed be found by the algorithm described above.

Solving the relaxed problem in (11) by exhaustive search is not feasible in practice, due to its exponential computational complexity. Instead, we observe that there are built-in recursive structures that can be exploited for an efficient dynamic programming solution of the relaxed problem with polynomial computational complexity.

First, let us introduce a notation on segment distortion introduced by missing frames between summary frame l_t and l_{t+1}, which is given by

$$G_{l_t}^{l_{t+1}} = \sum_{k=l_t}^{l_{t+1}-1} d(f_{l_t}, f_k). \tag{12}$$

Let the *state* of a video summary have t frames, and the last frame f_k be the minimum of the relaxed objective function given by

$$\begin{aligned} J_t^k(\lambda) &= \min_{S:\,\text{s.t. } |S|=t,\, l_{t-1}=k} \{D(S) + \lambda E(S)\} \\ &= \min_{l_1, l_2, \ldots, l_{t-2}} \left\{ G_0^{l_1} + G_{l_1}^{l_2} + \cdots G_{l_{t-2}}^{k} + G_k^n + \lambda \sum_{k=0}^{t-1} E(B_{l_k}, \tau_{l_k}) \right\}, \end{aligned} \tag{13}$$

where $|S|$ denotes the number of frames in S. Note that $l_0 = 0$, as we assume the first frame is always selected. The

minimization process in (11) has the following recursion:

$$
\begin{aligned}
&J_{t+1}^{k}(\lambda)\\
&=\min_{S:\text{ s.t. }|S|=t+1,\, l_t=k}\{D(S)+\lambda E(S)\}\\
&=\min_{l_1,l_2,\ldots,l_{t-1}}\{G_0^{l_1}+G_{l_1}^{l_2}\cdots+G_{l_{t-1}}^{k}+G_k^n\\
&\qquad+\lambda\big[E(B_0,\tau_0)+E(B_{l_1},(l_1-0)/F)\\
&\qquad+\cdots+E(B_{l_{t-1}},(l_{t-1}-l_{t-2})/F)\\
&\qquad+E(B_k,(k-l_{t-1})/F)\big]\}\\
&=\min_{l_1,l_2,\ldots,l_{t-1}}\Bigg\{\underbrace{G_0^{l_1}+G_{l_1}^{l_2}\cdots+G_{l_{t-2}}^{l_{t-1}}+G_{l_{t-1}}^{n}}_{D_t^{l_{t-1}}}-G_{l_{t-1}}^{n}+G_{l_{t-1}}^{k}+G_k^n\\
&\qquad+\lambda\Bigg[\underbrace{E(B_0,\tau_0)+E(B_{l_1},(l_1-0)/F)}_{E_t^{l_{t-1}}}\\
&\qquad\underbrace{+\cdots+E(B_{l_{t-1}},(l_{t-1}-l_{t-2})/F)}_{E_t^{l_{t-1}}}\\
&\qquad+E(B_k,(k-l_{t-1})/F)\Bigg]\Bigg\}\\
&=\min_{l_1,l_2,\ldots,l_{t-1}}\Bigg\{D_t^{l_{t-1}}+\lambda E_t^{l_{t-1}}\\
&\qquad\underbrace{+\lambda E(B_k,(k-l_{t-1})/F)-G_{l_{t-1}}^{n}+G_{l_{t-1}}^{k}+G_k^n}_{e^{l_{t-1},k}}\Bigg\}\\
&=\min_{l_{t-1}}\{J_t^{l_{t-1}}(\lambda)+e^{l_{t-1},k}\}.
\end{aligned}
\tag{14}
$$

The recursion has the initial condition given by

$$J_1^0(\lambda)=G_0^n+\lambda E(B_0,\tau_0). \tag{15}$$

The cost of transition is given by the edge cost $e^{l_{t-1},k}$ in (14), which is a function of λ, l_{t-1} and k as

$$
e^{l_{t-1},k}=\begin{cases}\lambda E(r_k,(k-l_{t-1})/F)-G_{l_{t-1}}^{n}+G_{l_{t-1}}^{k}+G_k^n, & \text{intracoding},\\ \lambda E(r_{k,l_{t-1}},(k-l_{t-1})/F)-G_{l_{t-1}}^{n}+G_{l_{t-1}}^{k}+G_k^n & \text{intercoding},\end{cases}
\tag{16}
$$

where r_k and $r_{k,l_{t-1}}$ are the estimated bit rates obtained from a rate profiler (e.g., [25]) to intracode the frame f_k, and intercode frame f_k with backward prediction from frame $f_{l_{t-1}}$, respectively. The DP solution starts with the initial node J_1^0, and propagates through a trellis with arcs representing possible transitions. At each node, we compute and store the optimal incoming arc and the minimum cost. Once all nodes with the final virtual frame f_n, $\{J_t^n(\lambda)\mid t=1,2,\ldots,n\}$, are computed, the optimal solution to the relaxed problem in (11) is found by selecting the minimum cost

$$S^*(\lambda)=\arg\min_t\{J_t^n(\lambda)\}, \tag{17}$$

and backtracking from the resulting final virtual frame nodes for the optimal solution. This is similar to the Viterbi algorithm [28]. An example of a trellis for $n=5$ and $\lambda=1.0e{-}4$ is shown in Figure 2, where all possible state transitions are plotted. For each state node, the minimum incoming cost is plotted as solid line, while other incoming arcs are plotted as dotted lines. For example, the node J_3^4 is computed as $J_3^4=\min_{j\in\{1,2,3\}}\{J_2^j+e^{j,4}\}$, and its incoming arc with the minimum cost is from node J_2^2. The virtual final frame nodes are all at the top of the trellis.

The Lagrange multiplier controls the tradeoff between summarization distortion and the energy cost in transmitting the summarized video frames. By varying the value of λ and solving the relaxed problem in the inner loop, we can obtain the optimal solution that minimizes the transmission energy cost while meeting certain distortion constraints. Since the operational energy-distortion function $E(D(S^*(\lambda)))$ is monotonic with respect to λ, a fast bisection search algorithm can be applied to find the optimal λ^*, which results in the tightest bound on the distortion constraint $D_{\max}$, that is, $D(S^*(\lambda^*))$ is the closest to $D_{\max}$. The algorithm can perform even faster by reusing the distortion and energy cost results that only need to be computed once in the iteration. The solution to the MEOS formulation can also be solved in the same fashion.

The complexity of the optimal inner loop solution is polynomial in frame number n, and the outer loop bisection search complexity depends on the choice of initial search window size and location. But overall, for small $n<60$, the complexity can be well handled by mobile devices with more powerful modern processors.

3.2. Maximum distortion problems

When the maximum distortion metric in (6) is used, the problem has a simpler structure due to less complex dependencies. Let us consider the MEOS problem first. The objective here is to minimize the energy cost of transmitting a segment of the video summary, with the given constraint on the maximum frame distortion allowed. Unlike the complicated structures in the average distortion case, this given distortion constraint can be used to prune the infeasible edges in the summary state trellis similarly to the previous case, and then a search and back tracking algorithm can be derived.

Let us define the summarization distortion for the video segment between video summary frames l_t and l_{t+1} as

$$D_{l_t}^{l_{t+1}}=\max_{j\in[l_t,\,l_{t+1}-1]}d(f_{l_t},f_j). \tag{18}$$

This is the maximum frame distortion between the previous summary frame l_t, and the subsequent missing frames before

the next summary frame l_{t+1}. It is clear that the placement of summary frames will have a major impact on the resulting video summary distortion. Generally, the larger the distance between the two summary frames l_t and l_{t+1}, the larger the resulting distortion. Where the summary frames are placed is also important. For example, if the summary frames l_t and l_{t+1} astride two different video shots, there will be a spike in the distortion $D_{l_t}^{l_{t+1}}$.

A frame loss distortion larger than $D_{\max}$ is not allowed in this case; we can reflect this constraint by defining the energy cost for the segment as

$$E_{l_t}^{l_{t+1}} = \begin{cases} E(B_{l_{t+1}}, (l_{t+1} - l_t)/F), & \text{if } D_{l_t}^{l_{t+1}} \le D_{\max}, \\ \infty, & \text{otherwise.} \end{cases} \tag{19}$$

With this, any summary frame selections with resulting segment distortion greater than $D_{\max}$ are excluded from the MEOS solution.

For the maximum energy minimization problem, let us also explore the structure of the energy cost of the optimal video summary solution ending with frame l_t:

$$E_{l_t} = \min_{l_1, l_2, \ldots, l_{t-1}} \left\{ E_0^{l_1} + E_{l_1}^{l_2} + \cdots + E_{l_{t-1}}^{l_t} \right\}. \tag{20}$$

This includes any combination of choices of summary frames between f_0 and f_{l_t}. Similarly to the relaxed cost case in average distortion minimization, it also has a recursive structure as

$$\begin{aligned} E_{l_{t+1}} &= \min_{l_1, l_2, \ldots, l_t} \{ E_0^{l_1} + E_{l_1}^{l_2} + \cdots + E_{l_{t-1}}^{l_t} + E_{l_t}^{l_{t+1}} \} \\ &= \min_{l_t} \{ E_{l_t} + E_{l_t}^{l_{t+1}} \} \\ &= \begin{cases} \min\limits_{l_t} \left\{ E_{l_t} + \underbrace{E(r_{l_{t+1}}, (l_{t+1} - l_t)/F)}_{\text{edge cost}} \right\}, & \text{if intracoding,} \\ \min\limits_{l_t} \left\{ E_{l_t} + \underbrace{E(r_{l_{t+1}, l_t}, (l_{t+1} - l_t)/F)}_{\text{edge cost}} \right\}, & \text{if intercoding.} \end{cases} \end{aligned} \tag{21}$$

This recursive relationship is illustrated by an example in Figure 3. A small scale problem with $n = 6$ frames from the "foreman" sequence is considered. The $D_{\max}$ is 15 in this case, which prunes out $[l_t, l_{t+1}]$ summary segments that have resulting distortion $D_{l_t}^{l_{t+1}} > D_{\max}$. The optimal solution is therefore found by searching through all feasible transitions in energy cost trellis, recording the minimum energy cost arcs as we compute the next stage in trellis expansion, and then backtracking for the optimal solution in a Viterbi algorithmic fashion [28]. The optimal summary for the problem in Figure 3 consists of frames f_0 and f_4.

Notice that the summary found is optimal, as compared with the convex-hull approximately optimal in the average distortion case. The resulting distortion $d(f_k, f_k')$ has interesting patterns as shown in Figure 4, for the 120-frame "foreman" sequence segment (frames 120∼249). The

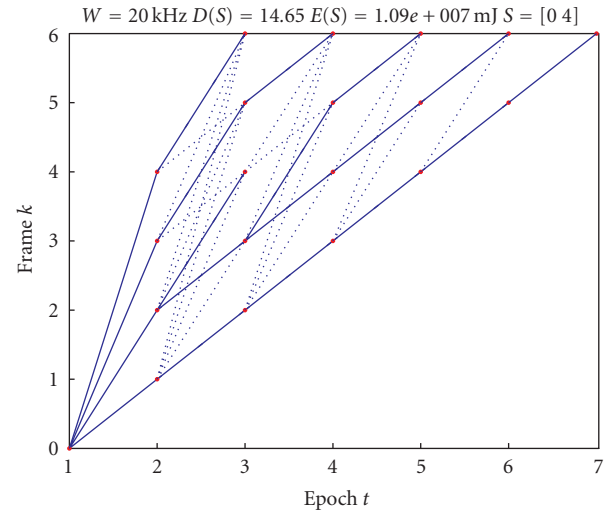


Figure 3: An example of DP trellis for the max distortion minimization problem.

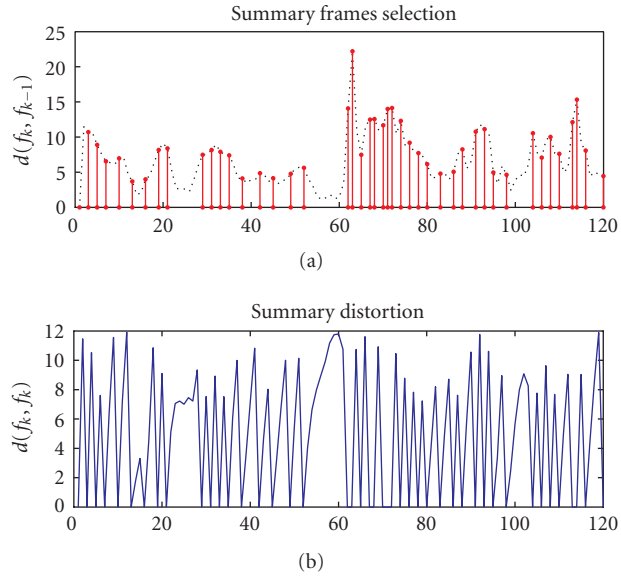


Figure 4: MEOS summary example.

distortion threshold $D_{\max} = 12$, and the resulting summary consists of 45 frames.

Figure 4(a) is the sequence activity level profile as differential frame distance, $d(f_k, f_{k-1})$, and the summary frame selections are plotted in red vertical lines. Figure 4(b) is the summary distortion plot $d(f_k, f_k')$. Notice that the placement of summary frames brings the maximum distortion for each segment below $D_{\max}$ indeed. The density of the summary frames also reflects well the activity level in the sequence, as expected.

To solve the maximum distortion minimization problem, instead of searching on the Lagrange multiplier as in the average distortion case, we develop a bisection search algorithm that searches on the maximum distortion constraint, $D_{\max}$, in

the outer loop, and in the inner loop, and solves the MEOS problem as a function of the threshold $D_{\max}$, that is,

$$S^*(D_{\max}) = \arg\min_S E(S), \quad \text{s.t. } D(S) \le D_{\max}. \tag{22}$$

To find the minimum distortion summary that meets the given energy constraint $E_{\max}$, the bisection search stops when the resulting energy cost $E(S^*(D_{\max}))$ is the closest to the $E_{\max}$. This is similar to the Lagrangian relaxation and DP solution to the average distortion case in structure.

3.3. Heuristic greedy solution

The DP solution has polynomial computational complexity $O(n^2)$, with n the number of frames in the sequence, which may not be practical for mobile devices that usually have limited power and computation capacity. A heuristic solution is thus developed to generate energy-efficient video summaries for both average and maximum distortion cases.

The heuristic algorithm selects the summary frames such that all summarization distortion segments $G_{l_{t-1}}^{l_t}$,

$$G_{l_t}^{l_{t+1}} \begin{cases} \sum_{k=l_t}^{l_{t+1}-1} d(f_{l_t}, f_k), & \text{avg distortion,} \\ \max_{k\in[l_t, l_{t+1}-1]} d(f_{l_t}, f_k), & \text{max distortion,} \end{cases} \tag{23}$$

between successive summary frames satisfy $G_{l_{t-1}}^{l_t} \le \Delta$, for a preselected step size Δ. Notice that this applies to both average and maximum distortions. The algorithm is greedy and operates in an one-pass fashion for a given Δ. The pseudocode of the proposed heuristic algorithm is then shown in Algorithm 1.

This replaces the DP algorithm in the optimal solution, and a bisection search on Δ can find the solution that satisfies the summarization distortion or the energy cost constraints. The computational complexity is $O(n)$ for the greedy algorithm solution. Simulation results with both the optimal and the heuristic algorithms are presented and discussed in Section 4.

4. SIMULATION RESULTS

To simulate a slow fading wireless channel, we model the channel fading as a two-state FSMC with channel states h_0 and h_1. The channel has transition probabilities, p and q, for state transition from h_0 to h_1, and h_1 to h_0, respectively, and the channel state transitional probability is given by $A = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$. The steady-state channel state probability is therefore computed as $\pi_0 = q/(p+q)$ and $\pi_1 = q/(p+q)$. Assuming that the deadline τ is much greater than the channel coherent time, T_c, that is, $\tau \gg T_c$, and the signaling rate is W (W is selected to simulate typical SNR operating range in wireless communications), then out of the total $2W\tau$ channel uses, $(p/(p+q))2W\tau$ are in channel state h_1 and $(q/(p+q))2W\tau$ are in channel state h_0.

Assuming that the channel state is known to both the transmitter and the receiver, with the optimal coding and packet scheduling, then the expected energy cost of transmitting B bits with delay constraint τ can then be computed as

$$\begin{aligned} E(B,\tau) &= \mathbf{E}_H\{E_b(2W\tau/B, h)B\} \\ &= \min_{0\le z\le 1}\{f(z; B, W, \tau, p, q, h_0, h_1)\} \\ &= \min_{0\le z\le 1}\left\{ zBE_b\left(\frac{q}{p+q}2W\tau/(zB), h_0\right)\right. \\ &\qquad \left. +(1-z)BE_b\left(\frac{p}{p+q}2W\tau/(B(1-z)), h_1\right)\right\}. \end{aligned} \tag{24}$$

In (24), we need to find an optimal bits splitting factor, z in $[0 \quad 1]$, of the total bits B, with zB bits transmitted optimally while the channel state is h_0, and $(1-z)B$ bits transmitted optimally while the channel state is h_1.

Note that (24) can be implemented as a lookup table in a practical system with more complex channel models. For simple channel models such as the two-state FSMC, a closed form solution can be derived. Once the conditions based on the first- and second-order derivatives (see the appendix for more detail) are satisfied for the minimization problem in (24), the optimal splitting of the bits is given by

$$\begin{aligned} z^* &= \frac{w\tau pq}{B(p+q)^2}\left[\log_2\left(\frac{h_0}{h_1}\right) + \frac{(p+q)}{w\tau p}B\right] \\ &= \frac{w\tau pq}{B(p+q)^2}\log_2\left(\frac{h_0}{h_1}\right) + \frac{q}{(p+q)}, \end{aligned} \tag{25}$$

and the minimum energy cost is given by

$$\begin{aligned} E(B,\tau) &= f(z^*; B, W, \tau, p, q, h_0, h_1) \\ &= z^*BE_b\left(\frac{q}{p+q}2W\tau/(z^*B), h_0\right) \\ &\quad + (1-z^*)BE_b\left(\frac{p}{p+q}2W\tau/(B(1-z^*)), h_1\right). \end{aligned} \tag{26}$$

Equation (26) can be implemented as a lookup table for the energy-distortion optimization algorithm.

The performance of the proposed algorithms has been studied in experiments as well. Some representative results are presented next. The implementation of the algorithms was done with a mix of C and Matlab.

In Figure 5, the QCIF-sized "foreman" sequence (frames 150~299) was utilized. The channel state is modeled as $h_0 = 0.9$, $h_1 = 0.1$, $p = 0.7$, $q = 0.8$. Signaling rate is set as $W = 20$ kHz. The background noise power is assumed to be $N = 1$ mJ per channel use. The summary frames are intracoded

```
L = 0; S = {f_0}.                    % select 1st frame
For k = 1: n − 1
        If G_L^k > Δ                 % check the segment distortion value
                S = S + {f_k}
                L = k
        End
End
```

Algorithm 1: Heuristic algorithm pseudo code.

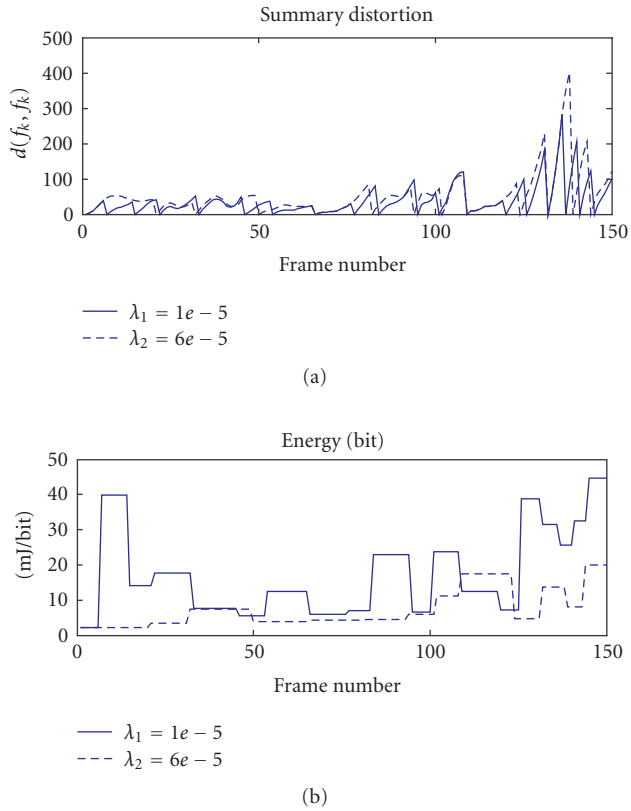


Figure 5: Examples of energy-efficient video summarization for the average distortion case.

with constant PSNR quality using the H.263 codec based on the TMN5 rate control. Summarization distortion and average power during transmissions are plotted for two different values of the Lagrange multiplier, with $\lambda_1 = 1.0e$–5 and $\lambda_2 = 6.0e$–5. For larger Lagrange multiplier, λ_2, more weight is placed on minimizing the energy cost, therefore the associated energy cost (area under the average power plot) is smaller than that of a smaller value λ_1. On the other hand, the summarization distortion is larger for λ_1 than for λ_2, as expected.

In the second set of experiments, the overall performance is characterized as the E-D and Energy-Rate (E-R) curves in Figures 6(a) and 6(b), respectively, for both $W = 10$ kHz and 20 kHz, as well as inter- and intracoding cases. Figure 6(a) characterizes the relationship between the summarization distortion and the total energy cost in $\log_{10}$(mJ) scale. As the summarization distortion goes up linearly, the energy cost drops exponentially. Figure 6(b) characterizes the relationship between the energy cost and the summarization rate. In the typical operating range of the video summarization, for example, $R(S) = [0.1, 0.9]$, the energy cost can change from 2 to 6 orders of magnitude. This clearly indicates that summarization can be an effective energy conserving scheme for wireless video communications.

Table 1: Computational complexity of the DP solution.

$n = 150$	$n = 120$	$n = 90$	$n = 60$	$n = 45$	$n = 30$
$t = 15.47$ s	$t = 9.82$ s	$t = 5.78$ s	$t = 2.78$ s	$t = 1.59$ s	$t = 0.6$ s

Table 2: Energy-summary quality tradeoff subjective evaluation.

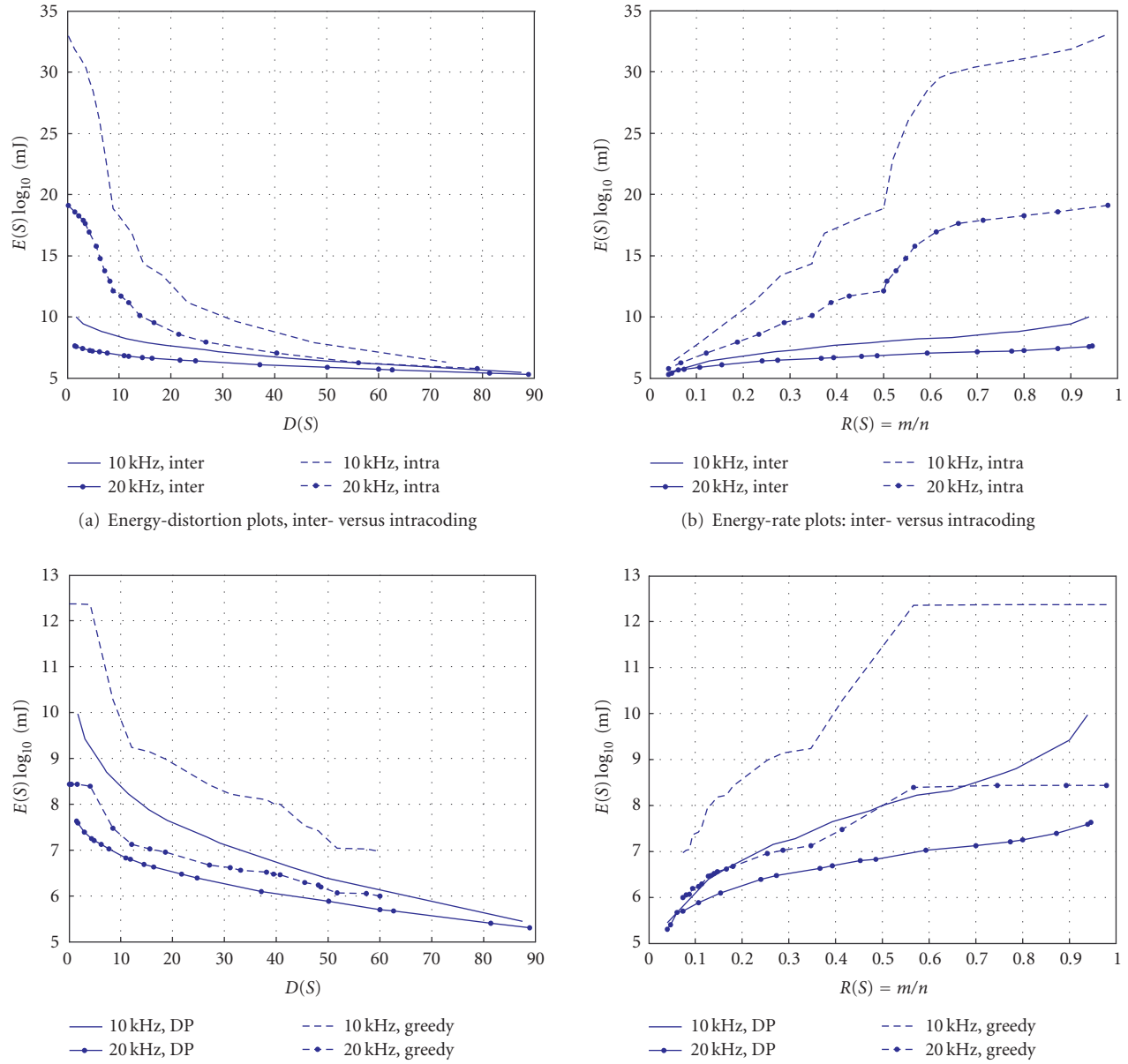
Summary name	λ	$R(S)$	$D(S)$	$E(S)$
"S1.263"	4.8e − 08	0.80	06.32	7.55e + 08
"S2.263"	2.0e − 07	0.68	09.75	2.62e + 08
"S3.263"	6.0e − 07	0.55	13.14	1.18e + 08
"S4.263"	3.0e − 06	0.39	18.91	4.46e + 07
"S5.263"	1.0e − 05	0.26	29.08	1.44e + 07
"S6.263"	1.0e − 04	0.12	49.68	2.53e + 06

The E-D performance for the maximum distortion metric is also summarized in Figure 7 for the optimal DP and greedy algorithms. Notice that the greedy solution performs closer to the optimal solution in this case.

The computational complexity of the DPsolution is indeed significantly larger than that of the greedy solution, especially as the size of the problem becomes larger. The execution times for the DP algorithm for various video segment lengths are summarized in Table 1.

These results are obtained with nonoptimized Matlab code running on a 2.0 GHz Celeron PC. Notice that the average execution time for the greedy algorithm is 0.11 s on the same computer for $n = 150$.

In Table 2 the summary rate, distortion, and energy cost are shown for various values of the Lagrange multiplier, along with the corresponding names of the summary sequences (based on the same 150-frame "foreman" sequence segment, intercoding, with $W = 10$ kHz) generated with the optimal DP algorithm. The sequences are also available for subjective evaluation of the tradeoffs between visual quality and energy cost in transmitting the sequence.



(a) Energy-distortion plots, inter- versus intracoding

(b) Energy-rate plots: inter- versus intracoding

(c) Energy-distortion plots, DP versus greedy, with intercoding

(d) Energy-rate plots: DP versus greedy, with intercoding

Figure 6: Energy-distortion performance for the average distortion minimization case.

Based on the visual evaluation of the results in Table 2, the graceful degradation of the video summary visual quality is clearly demonstrated. As the Lagrange multiplier value increases, more weight is placed on the energy cost during minimization. In the typical operating range of 0.12 to 0.80 for the video summarization rate, the energy cost differs by a factor of around 300 times. This demonstrates that video summarization is indeed an effective energy conservation scheme for wireless video streaming applications.

5. CONCLUSION AND FUTURE WORK

In this work, we formulated the problem of energy-efficient video summarization and transmission and proposed an optimal (within a convex hull approximation) algorithm for solving it. The algorithm is based on Lagrangian relaxation and dynamic programming in the average distortion metric case, and bisection search on distortion threshold and dynamic programming in the maximum distortion metric case. A heuristic algorithm to reduce the computational complexity has also been developed. The simulation results indicate that this is a very efficient and effective method in energy-efficient video transmission over a slow fading wireless channel.

The next step of the work is to have more realistic channel models for commercially deployed wireless systems, for example, WiMAX, and consider a multiuser setup and exploit diversity gains among users.

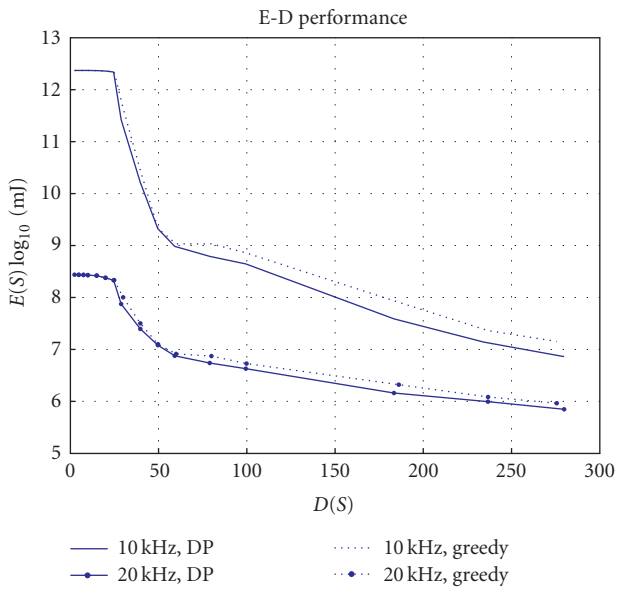


Figure 7: Energy-distortion performance for the maximum distortion case.

APPENDIX

DERIVATION OF THE OPTIMAL SPLIT IN TRANSMISSION

Assuming the channel state is known to both the transmitter and the receiver, the expected energy cost of transmitting B bits with delay τ is computed as

$$\begin{aligned} E(B,\tau) &= \mathbf{E}_H\{E_b(2W\tau/B,h)B\} \\ &= \min_{0\le z\le 1}\{f(z;B,W,\tau,p,q,h_0,h_1)\} \\ &= \min_{0\le z\le 1}\left\{zBE_b\left(\frac{q}{p+q}2W\tau/(zB),h_0\right)\right. \\ &\qquad \left.+(1-z)BE_b\left(\frac{p}{p+q}2W\tau/(B(1-z)),h_1\right)\right\}. \end{aligned} \tag{A.1}$$

Consequently, we have

$$\begin{aligned} f(z) &= zBE_b(2W\tau\pi_0/(zB),h_0) \\ &\quad +(1-z)BE_b(2W\tau\pi_1/((1-z)B),h_1) \\ &= (2\pi_0W\tau/h_0)(2^{zB/\pi_0W\tau}-1) \\ &\quad +(2\pi_1W\tau/h_1)(2^{(1-z)B/\pi_1W\tau}-1). \end{aligned} \tag{A.2}$$

Let

$$\begin{aligned} a_0 &= 2\pi_0W\tau/h_0, \qquad a_1 = 2\pi_1W\tau/h_1, \\ b_0 &= \frac{B}{\pi_0W\tau}, \qquad b_1 = \frac{B}{\pi_1W\tau}. \end{aligned} \tag{A.3}$$

We have $f(z) = a_0(2^{b_0z}-1)+a_1(2^{b_1(1-z)}-1)$. To minimize $f(z)$, let the first-order derivative be zero, which leads to

$$\begin{aligned} f'(z) &= a_0b_0\ln(2)2^{b_0z} - a_1b_1\ln(2)2^{b_1(1-z)} \\ &= 0, \quad \Longrightarrow z^* = \frac{1}{b_0+b_1}\left(\log_2\left(\frac{a_1b_1}{a_0b_0}\right)+b_1\right). \end{aligned} \tag{A.4}$$

Because the second-order derivative is always nonnegative as below

$$\begin{aligned} f''(z) &= a_0b_0^2\ln^2(2)2^{b_0z} \\ &\quad +a_1b_1^2\ln^2(2)2^{b_1(1-z)} \ge 0, \quad \forall 0\le z\le 1, \end{aligned} \tag{A.5}$$

the optimal bit splitting ratio is then

$$z^* = \pi_0\pi_1\log_2\left(\frac{h_0}{h_1}\right)\frac{W\tau}{B}+\pi_0, \tag{A.6}$$

and the optimal energy cost is given by

$$\begin{aligned} E(B,\tau) &= z^*BE_b(2\pi_0W\tau/(z^*B),h_0) \\ &\quad +(1-z^*)BE_b(2\pi_1W\tau/(B(1-z^*)),h_1). \end{aligned} \tag{A.7}$$

ACKNOWLEDGMENT

Part of this work was presented at SPIE VCIP 2005.

REFERENCES

[1] Wireless LAN Medium Access Control (MAC) Physical Layer (PHY), Specification of IEEE 802.11 Standard, 1998.

[2] R. Kravets and P. Krishnan, "Application-driven power management for mobile communication," *Wireless Networks*, vol. 6, no. 4, pp. 263–277, 2000.

[3] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[4] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[5] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '02)*, vol. 3, pp. 1773–1782, New York, NY, USA, June 2002.

[6] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 487–499, 2002.

[7] Y. S. Chan and J. W. Modestino, "Transport of scalable video over CDMA wireless networks: a joint source coding and power control approach," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 973–976, Thesaloniki, Greece, October 2001.

[8] 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 Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 411–424, 2002.

[9] 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 Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 511–523, 2002.

[10] I.-M. Kim and H.-M. Kim, "An optimum power management scheme for wireless video service in CDMA systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 1, pp. 81–91, 2003.

[11] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy-efficient wireless video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1710–1720, 2003.

[12] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Rate-distortion optimal video summary generation," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1550–1560, 2005.

[13] N. D. Doulamis, A. D. Doulamis, Y. S. Avrithis, and S. D. Kollias, "Video content representation using optimal extraction of frames and scenes," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 875–879, Chicago, Ill, USA, October 1998.

[14] A. Hanjalic and H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1280–1289, 1999.

[15] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 90–105, 2002.

[16] R. Lienhart, "Reliable transition detection in videos: a survey and practioner's guide," *International Journal of Image and Graphics*, vol. 1, no. 3, pp. 469–486, 2001.

[17] H. Sundaram and S.-F. Chang, "Constrained utility maximization for generating visual skims," in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '01)*, pp. 124–131, Kauai, Hawaii, USA, December 2001.

[18] Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, "Adaptive key frame extracting using unsupervised clustering," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 866–870, Chicago, III, USA, October 1998.

[19] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Bit constrained optimal video summarization," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '04)*, Singapore, October 2004.

[20] Z. Li, F. Zhai, A. K. Katsaggelos, and T. N. Pappas, "Energy-efficient video summarization and transmission over a slow fading wireless channel," in *Image and Video Communications and Processing*, vol. 5685 of *Proceedings of SPIE*, pp. 940–948, San Jose, Calif, USA, January 2005.

[21] Z. Li, F. Zhai, and A. K. Katsaggelos, "Video summarization for energy-efficient wireless streaming," in *Visual Communications and Image Processing*, vol. 5960 of *Proceedings of SPIE*, pp. 763–774, Beijing, China, July 2005.

[22] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunication, John Wiley & Sons, New York, NY, USA, 1991.

[24] Z. Li, G. M. Schuster, and A. K. Katsaggelos, "MINMAX optimal video summarization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1245–1256, 2005.

[25] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1221–1236, 2001.

[26] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video Frame Compression and Object Boundary Encoding*, Kluwer Academic Publishers, Norwell, Mass, USA, 1997.

[27] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 160–175, 1993.

[28] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.