Video Summarization for Energy Efficient Wireless Streaming

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

[1]Multimedia Research Lab (MRL), Motorola Labs, Schaumburg, IL 60196
[2]Dept. of Electrical & Computer Engineering, Northwestern University, Evanston, IL 60208,
[3]Digital Video Department, HPA, Texas Instruments, Dallas, TX 75243

ABSTRACT

With the proliferation of camera equipped cell phones and the deployment of the higher data rate 2.5G and 3G infra structure systems, providing consumers with video-equipped cellular communication infrastructure is highly desirable, and can drive the development of a large number of valuable applications. However, for an uplink wireless channel, both the bandwidth and battery energy in a mobile phone are limited for video communications. In this paper, we pursue an energy efficient video communication solution through joint video summarization and transmission adaptation over a slow fading wireless channel. Coding and modulation schemes and packet transmission strategy are optimized and adapted to the unique packet arrival and delay characteristics of the video summaries. In additional to the optimal solution, we also propose a heuristic solution that is greedy but has close to optimal performance. Operational energy efficiency – summary distortion performance is characterized under an optimal summarization setting. Simulation results show the advantage of the proposed scheme with respect to energy efficiency and video transmission quality.

Keywords: video summarization, wireless video, energy efficiency.

1. INTRODUCTION

Video summarization is the process of generating a shorter version of an original video sequence. The summarized sequence contains key frames extracted from the original video sequence based on some criterion specified by the user. Video summarization may be required when a system is operating under limited bandwidth conditions, or under tight constraints with regard to 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 view time is a constraint. Video summarization is also needed where important video segments must be transmitted to a based 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 [5, 8, 9, 17, 21, 25], where a video sequence is segmented into video shots, and then one or multiple key frames per shot are selected based on certain heuristic or metric for the summary. In this work, we consider the application of video summarization over wireless channels. Particularly, we consider using the scheme of video summarization to deal with the problem in wireless communications arising from the severe limitation in both bandwidth and energy.

Transmitting video over wireless channels from mobile devices has gained increased popularity in a wide range of applications. For example, dramatic increase in bandwidth has been 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 capability in personal communications. Although wireless video communications is highly desirable in many applications, wireless video communication presents some unique challenges. Due to shadowing and multi-path effect, the channel gain varies over time, which makes the 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 limitation is especially of concern because of the high energy consumption rate in encoding and transmitting video bit streams. Therefore, how to achieve reliable video communication over a fading channel with energy efficiency is crucial for wide deployment of wireless video based applications.

[*] Contact info: 1301 E Algonquin Road, IL02-Room2811, Schaumburg, IL 60196, USA. email: zhu.li@motorola.com; phone: 847 576-6942.

Energy efficient wireless communication is a widely studied topic. For example, a simple scheme is to put the device into sleeping mode when not used, as in [12, 14]. However, as the VLSI design and integrated circuit (IC) manufacturing technologies advance, the energy consumption on circuits side is being driven down, but the communication energy cost is lower bounded by information theory results. In [1], the fundamental trade-off between average power and delay constraint in communication over fading channels is explored and characterized. In [2], optimal power control schemes for communication over fading channels are developed. In [7] and [22], optimal off-line and near optimal on-line 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) requirement with respect to packet loss probability and delay constraint, which differentiate them from the traditional data transmission applications. Taking advantage of the specific source characteristics of video and jointly adapting video source coding decisions with transmission power, modulation and coding schemes can achieve substantial energy efficiency compared with non-adaptive transmission schemes. Examples of this type of work are reported in [3, 6, 11, 13, 18]. In those studies, source-coding controls are mostly based on frame and / or macro block (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 PSNR distortion is not desirable. Instead, we are considering joint video summarization and transmission approaches to achieve the required energy efficiency. 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 [15], 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 [16]. In [26], 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 extend the work in [26] to consider inter-frame coding mechanism. In addition to the energy-distortion optimized solution, we also propose a heuristic solution, which is a greedy method but performs close to the optimal solution.

The rest of the paper is organized as follows. In section 2 we give assumptions on the communication over fading 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 discuss the channel model, delay analysis for video summary packets and the problem formulation.

2.1. Slow Fading Channel Models and Assumptions

Communication of video over a wireless channel has its unique challenges, namely fading and energy efficiency issues. In this work, we assume that the wireless channel can be modeled as a band-limited, AWGN channel with discrete time, slow block fading,

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

in which h_k is the channel gain for time slot k and it stays constant for time T_c, the channel coherent time. The channel is slow fading because we assume the symbol duration $T_s << T_c$. Thus, there are many channel uses during each time slot. n_k is the additive Gaussian noise with power N. The variation of the channel state is modeled as a finite state Markov channel (FSMC) [24], which has a finite set of possible states, $H=\{h_1, h_2, \dots h_m\}$, and makes transition every T_c second with probability given by the transition probability matrix A, where a_{ij} = Prob{transition from h_j to h_k}.

To reliably send R bits information over the fading channel in one channel use, the minimum power needed with optimal coding is given in information theory [4] as,

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

Similar to [7], let $x=1/R$ be the number of transmission 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,

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

Examples of the energy efficiency functions with different fading states are shown in Fig. 1. The range of x in Fig. 1 corresponds to the received signal-to-noise ratio 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 $\tau >> T_c$, the number of transmission available is $2W\tau$, where W is the signaling rate. Then the expected energy cost will be,

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

In Eq. (4), the expectation 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 look up table for a given channel model in simulation. A closed form solution may also be possible, under some optimal coding and packet scheduling assumption.

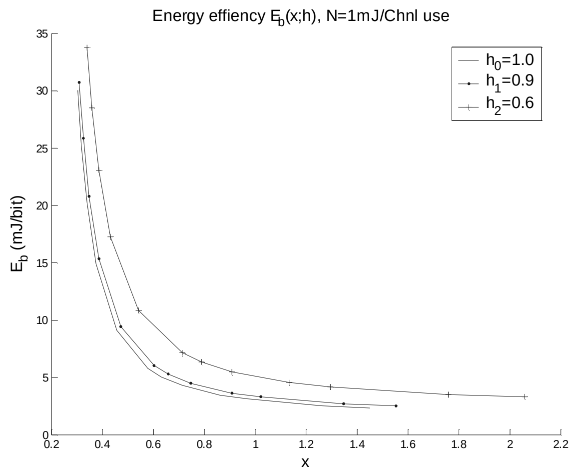


Figure 1. Energy efficiency over fading channels

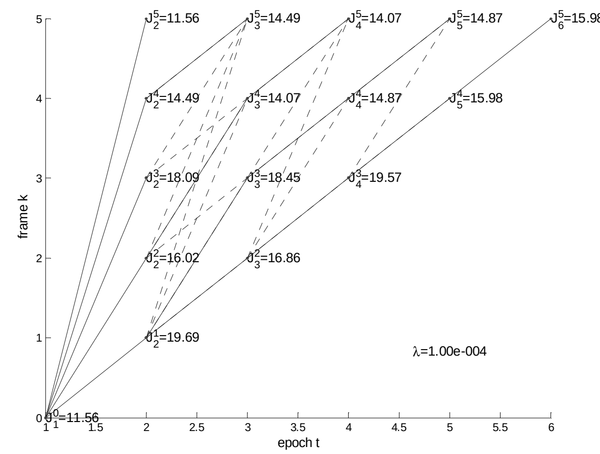


Figure 2. Example of DP trellis

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 be $S=\{f_{l_0}, f_{l_1}, \cdots, f_{l_{m-1}}\}$. Obviously the video summarization process has an implicit constraint that $l_0 < l_1 < \ldots < l_{m-1}$. Let the reconstructed sequence $V_S' = \{f_0', f_1', \cdots f_{n-1}'\}$ be obtained by substituting missing frames with the most recent frame that is in the summary S, that is, $f_k' = f_{i=\max(l):\ s.t.\ l\in\{l_0, l_1, \cdots, l_{m-1}\},\ i\le k}$. Let the summarization rate be,

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

taking values in $\{1/n, 2/n, \ldots, n/n\}$. The summarization rate is a reflection of how many frames out of the original n frames are selected into the summary. The summarization distortion is computed as the average frame distortion between the original sequence and the reconstructed sequence from the summary,

$$D(S) = \frac{1}{n}\sum_{k=0}^{n-1} d(f_k, f_k'). \qquad (5)$$

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. For the encoding of the video summary frames, we assume a constant PSNR coding strategy, with frame bit budget B_{l_j} given by some rate profiler, for example [10]. Packets from different summary frames have different delay tolerances. We assume that the first frame of the original sequence, f_0, is always selected for the summary and intra coded with some B_0 bits. The delay tolerance τ_0 is determined by how much initial streaming delay is allowed.

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 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

$$E(S) = \sum_{k=0}^{m-1} E(B_{l_k}, \tau_{l_k}) = E(B_0, \tau_0) + \sum_{k=1}^{m-1} E(B_{l_k}, \tau_{l_k}) \tag{6}$$

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} packets. There are trade offs between the number of frames, m, in the summary versus the summarization distortion. The more frames can be selected into the summary, the smaller the summarization distortion. On the other hand, the more frames in the summary, more bits need to be spent in encoding the frames, and the packet arrival pattern gets more crowed, 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 sub-section, 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 (ED) 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,

$$S^* = \arg\min_S E(S),\ s.t.\ D(S) \le D_{\max}\ . \tag{7}$$

We can also formulate the energy efficiency problem as the 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,

$$S^* = \arg\min_S D(S),\ s.t. E(S) \le E_{\max}\ . \tag{8}$$

The optimal solution to the formulations in Eqs. (7) and (8) is based on Lagrangian relaxation and Dynamic Programming (DP), which is given in the next section.

3. SOLTUION ALGORITHMS

Directly solving the constrained problems in Eqs. (7) and (8) is usually difficult due to the complicated dependencies and large searching space for the operating parameters. Instead, we introduce the Lagrangian multiplier relaxation to convert the original problem in to an unconstrained problem. The solution to the original problem can then be found by solving the resulted unconstrained problem with the appropriate Lagrangian multiplier that satisfies the constraint. This approach has been widely used in solving a number of coding and resource allocation problems in video/image compression [20].

3.1. Dynamic Programming (DP) Solution with Lagrangian Relaxation

Considering the MEOS formulation, by introducing the Lagrangian multiplier, the relaxed problem is therefore given by,

$$S^*(\lambda) = \arg\min_S \{E(S) + \lambda D(S)\}, \tag{9}$$

in which the optimal solution S^* becomes a function of λ. From [19] we know that by varying λ from zero to infinity, we sweep the convex hull of the operational Energy-Distortion (ED) function $E(D(S^*(\lambda)))$, which is also monotonic with respect to λ. Therefore a bi-section search algorithm on λ can give us the optimal solution within a convex hull approximation. In real world applications, the operational ED functions are typically convex, and the optimal solution can indeed be found by the algorithm above.

To solve the relaxed problem in (9) by exhaustive search, which has an exponential computational complexity, is not feasible. Instead, we observe that there are built-in recursive structures that can be exploited for an efficient dynamic programming solution for 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,

$$G_{l_t}^{l_{t+1}} = \sum_{k=l_t}^{l_{t+1}-1} d(f_{l_t}, f_k) \, . \tag{10}$$

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

$$\begin{aligned} J_t^k(\lambda) &= \min_{S:s.t.|S|=t, l_{t-1}=k} \{D(S) + \lambda E(S)\} \\ &= \min_{l_1, l_2, \cdots, l_{t-2}} \{G_0^{l_1} + G_{l_1}^{l_2} + \cdots G_{l_{t-2}}^{k} + G_k^n + \lambda \sum_{k=0}^{t-1} E(B_{l_k}, \tau_{l_k})\} \end{aligned} , \tag{11}$$

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 Eq. (10) has the following recursion,

$$\begin{aligned} J_{t+1}^k(\lambda) &= \min_{S:s.t.|S|=t+1, l_t=k} \{D(S) + \lambda E(S)\} \\ &= \min_{l_1, l_2, \cdots, l_{t-1}} \{G_0^{l_1} + G_{l_1}^{l_2} \cdots + G_{l_{t-1}}^{k} + G_k^n + \\ &\qquad \lambda[E(B_0, \tau_0) + E(B_{l_1}, (l_1 - 0)/F) + \cdots + E(B_{l_{t-1}}, (l_{t-1} - l_{t-2})/F) + E(B_k, (k - l_{t-1})/F)]\} \\ &= \min_{l_1, l_2, \cdots, l_{t-1}} \{\underbrace{G_0^{l_1} + G_{l_1}^{l_2} \cdots + G_{l_{t-2}}^{l_{t-1}} + G_{l_{t-1}}^{n}}_{D_t^{l_{t-1}}} - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_k^n + \\ &\qquad \lambda[\underbrace{E(B_0, \tau_0) + E(B_{l_1}, (l_1 - 0)/F) + \cdots + E(B_{l_{t-1}}, (l_{t-1} - l_{t-2})/F)}_{E_t^{l_{t-1}}} + E(B_k, (k - l_{t-1})/F)]\} \\ &= \min_{l_1, l_2, \cdots, l_{t-1}} \{D_t^{l_{t-1}} + \lambda E_t^{l_{t-1}} + \underbrace{\lambda E(B_k, (k - l_{t-1})/F) - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_k^n}_{e^{l_{t-1}, k}}\} \\ &= \min_{l_{t-1}} \{J_t^{l_{t-1}}(\lambda) + e^{l_{t-1}, k}\} \end{aligned} \tag{12}$$

The recursion has the initial condition given by

$$J_1^0(\lambda) = G_0^n + \lambda E(B_0, \tau_0) \, . \tag{13}$$

The cost of transition is given by edge cost $e^{l_{t-1}, k}$ in Eq. (12), which is a function of λ, l_{t-1} and *k*, as,

$$e^{l_{t-1}, k} = \left\{ \begin{array}{ll} \lambda E(r_k, (k - l_{t-1})/F) - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_k^n, & \text{intra - coding} \\ \lambda E(r_{k, l_{t-1}}, (k - l_{t-1})/F) - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_k^n & \text{inter - coding} \end{array} \right\},$$

where r_k and $r_{k, l_{t-1}}$ are the rate profiler (eg, [10]) estimated number of bits to intra-code the summary frame f_k , and inter-code the 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 the nodes with final virtual frame f_n, $\{J_t^n(\lambda) \mid t = 1, 2, \cdots, n\}$, are computed, the optimal solution to the relaxed problem in Eq. (9) is found by selecting the minimum cost

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

and backtracking from the resulted final virtual frame nodes for the optimal solution. This is similar to the Viterbi algorithm [23]. An example with trellis for n=5 and λ=1.0e-4 is shown in Fig. 2, where all possible state transitions are plotted. For each state node, the minimum cost incoming is plotted in solid line, while other incoming arcs are plotted as dotted lines. For example, the note J_3^4 is computed as $J_3^4 = \min_{j\in\{1,2,3\}}\{J_2^j + e^{j,3}\}$, 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 Lagrangian multiplier controls the trade off between the summarization distortion and the energy cost in transmitting the summary. By varying the value of λ and solving the relaxed problem in the inner loop, we can obtain the optimal solution that minimizing the transmission energy cost while meeting certain distortion constraints. Since the operational energy-distortion function $E(D(S^*(\lambda)))$ is monotonic with λ, a fast bi-section search algorithm can be applied to find the optimal λ^*, which results in the tightest bound on the distortion constraint D_{max}, i.e., $D(S^*(\lambda^*))$ is closest to D_{max}. The algorithm can be made even faster by re-using the distortion and energy cost results that only need to be computed once in the iteration. The solution to the MDOS formulation can also be solved in the same fashion.

3.2. Heuristic Greedy Solution

The DP solution has polynomial computational complexity O(n^2), 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.

The heuristic algorithm selects the summary frames such that all summarization distortion segments $G_{l_{t-1}}^{l_t}$, between successive summary frames, satisfying $G_{l_{t-1}}^{l_t} \le \Delta$, for a pre-selected step size Δ. The algorithm is greedy and operates in a one-pass fashion for a given Δ. The pseudo code is shown below,

```
L=0; S={f0}.                    % select the 1st frame
For k=1:n-1
    If  G_L^k > Δ               % check the segment distortion value
        S=S+{fk}
        L=k
    End
End
```

This replaces the DP algorithm in the optimal solution, and a bi-section 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. The simulation results for both the optimal and heuristic algorithms are shown 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. TheThe channel model is shown in Fig. 3, where p and q are the probabilities of channel 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 $\pi_0 = \frac{q}{p+q}$ and $\pi_1 = \frac{q}{p+q}$. Assuming that $\tau >> T_c$, 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, $\frac{p}{p+q}2W\tau$ are in channel state h_1 and $\frac{q}{p+q}2W\tau$ are in channel state h_0.

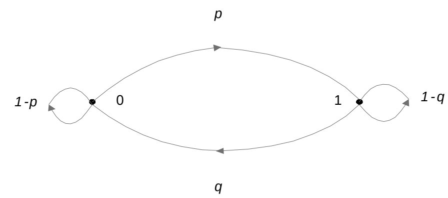


Figure 3. Two state FSMC

Assuming 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,

$$\begin{aligned} E(B,\tau) &= \mathrm{E}_H\{E_b(2W\tau/B,h)B\} \\ &= \min_{0\le x\le 1}\{f(x;B,W,\tau,p,q,h_0,h_1)\} \\ &= \min_{0\le x\le 1}\{xBE_b(\frac{q}{p+q}2W\tau/(xB),h_0)+(1-x)BE_b(\frac{p}{p+q}2W\tau/(B(1-x)),h_1)\} \end{aligned} \quad . \qquad (14)$$

In Eq. (14), we need to find an optimal bits splitting, x, of the total bits B, with xB bits transmitted optimally while channel state is h_0, and (1-x)B bits transmitted optimally while channel state is h_1.

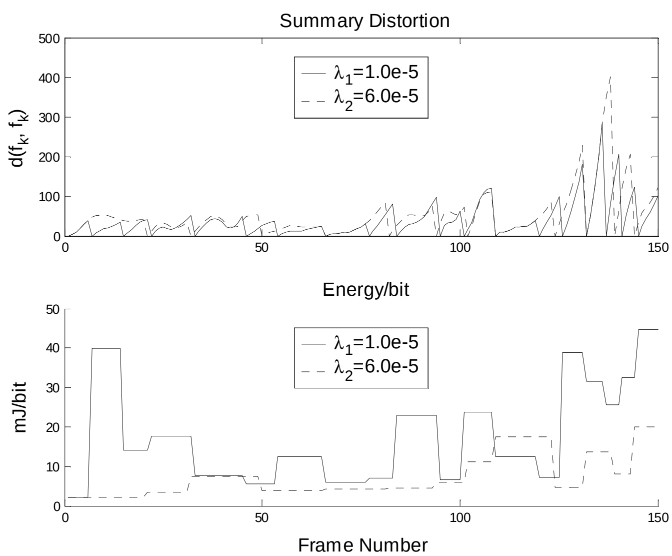


Figure 4. Examples of energy efficient video summarization

Note that Eq. (14) can be implemented as a look up table in a practical system. For simple channel models such as the two-state FSMC, a closed form solution can be derived. Once the 1st and 2nd order conditions (See Appendix for more detail) are checked for the minimization problem in Eq. (14), the optimal splitting of the bits is given by

$$x^* = \frac{w\tau pq}{B(p+q)^2}[\log_2(\frac{h_0}{h_1})+\frac{(p+q)}{w\tau p}B] = \frac{w\tau pq}{B(p+q)^2}\log_2(\frac{h_0}{h_1})+\frac{q}{(p+q)}, \qquad (15)$$

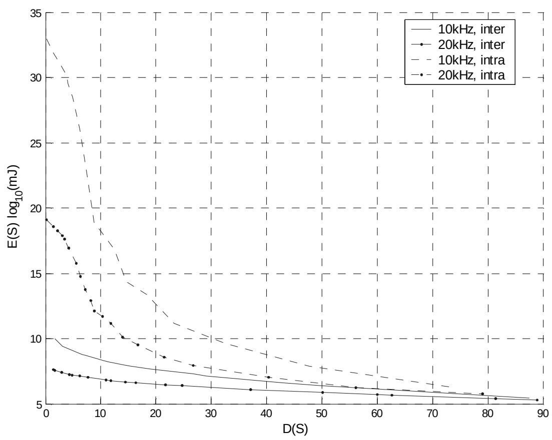
and the minimum energy cost is given by

$$\begin{aligned} E(B,\tau) &= f(x^*;B,W,\tau,p,q,h_0,h_1)\} \\ &= x^*BE_b(\frac{q}{p+q}2W\tau/(x^*B),h_0)+(1-x^*)BE_b(\frac{p}{p+q}2W\tau/(B(1-x^*)),h_1)\} \end{aligned} \qquad (16)$$

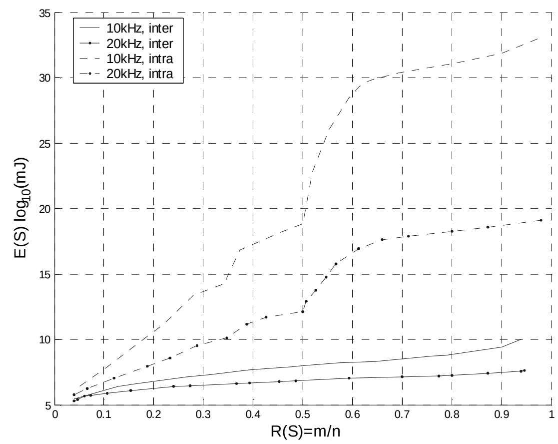
Eq. (16) can be implemented as a look up table for the ED optimization algorithm.

In our simulation, the optimal algorithm is implemented in Matlab and the simulation results using the “foreman” sequence (frames 150~299) are shown in Fig. 4. The channel state is modeled as h_0=0.9, h_1=0.1, p=0.7, q=0.8.

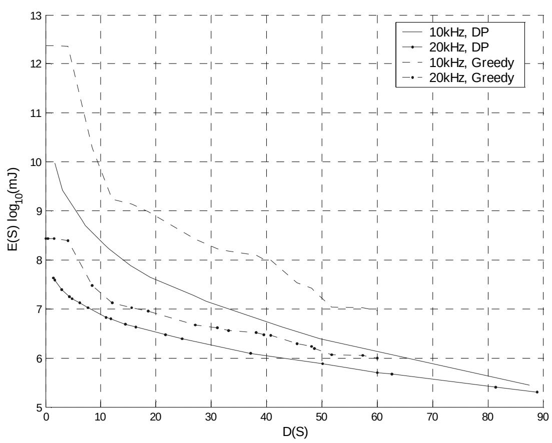
Signaling rate is set as W=20 kHz. The background noise power is assumed to be N=1mJ per channel use. The summary frames are intra-coded with constant PSNR quality using the H.263 codec based on the TMN5 rate control. Summarization distortion and average power during transmission are plotted for different values of Lagrangian multiplier, with λ_1=1.0e-5 and λ_2 =6.0e-5. For larger Lagrangian value, 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 where there is a smaller value λ_1. On the other hand, the summarization distortion is larger for λ_1 than for λ_2, as expected.



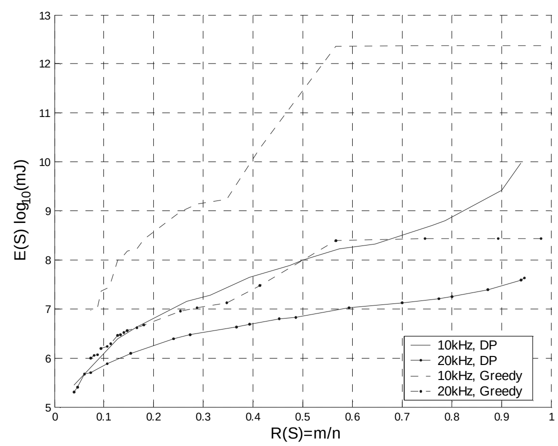
(a) Energy-distortion plots, inter vs intra



(b) Energy-rate plots: inter vs intra



(c) Energy-distortion plots, DP vs greedy



(d) Energy-rate plots: DP vs greedy

Figure 5 . Energy-Distortion performance

The overall performances for the optimal solution are characterized as the Energy-Summarization Distortion (E-D) and Energy-Summarization Rate (E-R) curves in Fig. 5a and Fig. 5b, for both W=10 kHz and 20 kHz, as well as inter and intra coding cases. For the same 150-frame segment of the "foreman" sequence, Fig. 5a 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. Fig. 5b 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 in the order of 2 to 6 magnitudes for the 150 frame segment from "foreman" sequence. This clearly indicates that summarization can be an effective energy conserving scheme for wireless video communications.

The performance comparison of the optimal DP solution and the greedy solution is shown in Fig. 5c and Fig. 5d, for the inter coding case with W=10 kHz and 20 kHz. As expected, the E-D performance curves of the greedy solution are lower bounded by the DP solution. In fact the performance of the greedy solution at 20kHz is close to the DP performance at 10 kHz. Compared to the intra coded case, the greedy algorithm with inter coding has a far better performance in terms of energy efficiency.

Table 1 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

To have a better understanding of the trade offs between the energy efficiency and the video summary quality, we have generated multiple video summaries in H.263 bit streams for subjective evaluation. They are available at:

http://ivpl.ece.northwestern.edu/~zli/new_home/demo/vcip05/demo.html

for the same 150-frame "foreman" sequence segment, with W=10 kHz, inter coding case, using the DP algorithm discussed in Section 3.1. The Lagrangian multiplier values, the summary rate, and the energy costs are summarized in Table 1. The results clearly demonstrate the graceful degradation of the video summary visual quality. As the Lagrangian multiplier value increases, more weight is put on the the energy cost during the minimization. In the typical operating range of 0.12 to 0.80 for 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 and proposed an optimal (within a convex hull approximation) algorithm for energy efficient video summarization and transmission. The algorithm is based on the Lagrangian relaxation and dynamic programming. 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 communication energy utilization for video transmission over a slow fading channel.

In the future, we will extend the work using more realistic channel models and simulations. We will also improve the performance of the heuristic algorithm and narrow the gap between the performance of the greedy solution and that of the proposed optimal solution.

APPENDIX

Derivation of the optimal bit split solution is given below:

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,

$$
\begin{aligned}
E(B,\tau) &= \mathrm{E}_H\{E_b(2W\tau/B,h)B\} \\
&= \min_{0\le x\le 1}\{f(x;B,W,\tau,p,q,h_0,h_1)\} \\
&= \min_{0\le x\le 1}\{xBE_b(\frac{q}{p+q}2W\tau/(xB),h_0)+(1-x)BE_b(\frac{p}{p+q}2W\tau/(B(1-x)),h_1)\}
\end{aligned}
$$

So we have,

$$
\begin{aligned}
f(x) &= xBE_b(2W\tau\pi_0/(xB),h_0)+(1-x)BE_b(2W\tau\pi_1/((1-x)B),h_1) \\
&=(2\pi_0W\tau/h_0)(2^{\frac{xB}{\pi_0W\tau}}-1)+(2\pi_1W\tau/h_1)(2^{\frac{(1-x)B}{\pi_1W\tau}}-1)
\end{aligned}
$$

Let,

$$a_0 = 2\pi_0 W\tau / h_0, \qquad a_1 = 2\pi_1 W\tau / h_1,$$
$$b_0 = \frac{B}{\pi_0 W\tau}, \qquad b_1 = \frac{B}{\pi_1 W\tau}$$

We have, $f(x) = a_0(2^{b_0 x} - 1) + a_1(2^{b_1(1-x)} - 1)$. To minimize $f(x)$, let the 1st order condition be zero, this leads to,

$$f'(x) = a_0 b_0 \ln(2) 2^{b_0 x} - a_1 b_1 \ln(2) 2^{b_1(1-x)} = 0, \Rightarrow x^* = \frac{1}{b_0 + b_1}(\log_2(\frac{a_1 b_1}{a_0 b_0}) + b_1),$$

and since the 2nd order condition is always non-negative:

$$f''(x) = a_0 b_0^2 \ln^2(2) 2^{b_0 x} + a_1 b_1^2 \ln^2(2) 2^{b_1(1-x)} \ge 0, \forall 0 \le x \le 1,$$

the optimal bit splitting ratio is,

$$x^* = \pi_0 \pi_1 \log_2(\frac{h_0}{h_1}) \frac{W\tau}{B} + \pi_0,$$

and the optimal energy cost is given by,

$$E(B,\tau) = x^* B E_b(2\pi_0 W\tau/(x^* B), h_0) + (1 - x^*) B E_b(2\pi_1 W\tau/(B(1-x^*)), h_1)$$

REFERENCE

1. R. Berry and R. Gallager, “Communication over fading channels with delay constraints”, *IEEE Trans. on Information Theory*, May 2002.
2. G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels”, *IEEE Trans. on Information Theory*, July 1999.
3. Y. S. Chan and J. W. Modestino, “Transport of scalable video over CDMA wireless networks: A joint source coding and power control approach”, *Proceedings of Int’l Conference on Image Processing (ICIP),* Thesaloniki, Greece, 2001.
4. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunication, New York, 1991.
5. N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias, “Video Content Representation Using Optimal Extraction of Frames and Scenes”, *Proc. of Int’l Conference on Image Processing (ICIP),* Chicago, Illinois, 1998.
6. Y. Eisenberg, C. E. Luna, T.N. Pappas, R. Berry, and A. K. Katsaggelos, “”, IEEE Trans. on Circuits & System for Video Technology, vol. 12, no. 8, June 2002.
7. A. El. Gamal, C. Nair, B. Prabhakar, E. Uysal-biyikoglu, and S. Zahedi, “Energy efficient scheduling of packet transmission over wireless networks”, *Proceedings of INFOCOM*, 2002.
8. A. Hanjalic and H. Zhang, “An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis”, *IEEE Trans. on Circuits and Systems for Video Technology,* vol.9, December 1999.
9. A. Hanjalic, “Shot-Boundary Detection: Unraveled and Resolved?”, *IEEE Trans. on Circuits and Systems for Video Technology,* vol.12, No. 2, Feb. 2002.
10. Z. He and S. K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Trans. on Circuits and System for Video Technology*, vol. 11, pp. 1221 –1236, December 2001
11. Z. He, J. Cai, and C.-W. Chen, “Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding”, *IEEE Trans. on Circuits and Systems for Video Technology,* vol. 12, no. 6, June 2002.
12. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY), Spec. of IEEE 802.11 Standard, 1998.
13. I.-M. Kim, and H.-M. Kim, “An optimum power management scheme for wireless video service in CDMA systems”, *IEEE Trans. on Wireless Communication*, vol. 2, no. 1, Jan. 2003.
14. R. Kravets and P. Krishnan, “Application-driven power management for mobile communication”, Wireless Networks, vol. 6, no. 4, September, 2000.
15. Z. Li, A. K. Katsaggelos, G. Schuster, and B. Gandhi, “Rate-Distortion Optimal Video Summary Generation”, *IEEE Trans. on Image Processing*, to appear.
16. Z. Li, G. Schuster, A. K. Katsaggelos, and B. Gandhi, “Bit Constrained Optimal Video Summarization”, *Proceedings of International Conference on Image Processing (ICIP),* Singapore, 2004.
17. 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.

18. 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, December 2003.
19. K. Ramchandran and M. Vetterli, “Best wavelet packet bases in a rate-distortion sense”, *IEEE Trans. on Image Processing*, vol. 2, no. 2, April 1993.
20. G. M. Schuster and A. K. Katsaggelos, Rate-Distortion Based Video Compression, Optimal Video Frame Compression and Object Boundary Encoding. Norwell, MA: Kluwer, 1997.
21. H. Sundaram and S-F. Chang, “Constrained Utility Maximization for Generating Visual Skims”, *IEEE Workshop on Content-Based Access of Image & Video Library*, 2001.
22. E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, “Energy Efficient Packet Transmission Over a Wireless Link”, *IEEE/ACM Trans. on Networking*, vol. 10, no. 4, August 2002.
23. A. J. Viterbi, “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm”, *IEEE Trans. on Information Theory*, vol. IT-13, pp. 260-269, April 1967.
24. H. S. Wang, and N. Moayeri, “Finite-state Markov channel- a useful model for radio communication channels”, *IEEE Trans. on Vehicular Tech*, vol. 44, February 1995.
25. Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, “Adaptive Key Frame Extracting Using Unsupervised Clustering”, *Proc. of Int’l Conference on Image Processing (ICIP)*, Chicago, Illinois, 1998
26. Z. Li, F. Zhai, A. K. Katsaggelos, and T. N. Pappas, “Energy Efficient Video Summarization and Transmission Over a Slow Fading Wireless Channel”, *Proc. Of SPIE Image & Video Communication & Processing (IVCP)*, San Jose, CA, 2005.