

# Cross-Layer Optimization for Video Summary Transmission over Wireless Networks

Dalei Wu, *Student Member, IEEE*, Song Ci, *Senior Member, IEEE*, and Haohong Wang, *Member, IEEE*

**Abstract**—Video summarization has gained increased popularity in the emerging multimedia communication applications, however, very limited work has been conducted to address the transmission problem of video summary frames. In this paper, we propose a cross-layer optimization framework for delivering video summaries over wireless networks. Within a rate-distortion theoretical framework, the source coding, allowable retransmission, and adaptive modulation and coding have been jointly optimized, which reflects the joint selection of parameters at physical, data link and application layers. The goal is to achieve the best video quality and content coverage of the received summary frames and to meet the delay constraint. The problem is solved using Lagrangian relaxation and dynamic programming. Experimental results indicate the effectiveness and efficiency of the proposed optimization framework, especially when the delay budget imposed by the upper layer applications is small, where more than 10% distortion gain can be achieved.

**Index Terms**—universal multimedia access, video summarization, adaptive modulation and coding, automatic repeat request, cross-layer optimization, rate distortion, video communication.

## I. INTRODUCTION

IN recent years, universal multimedia access (UMA) [1], [2] is emerging as one of the most important components for the next generation of multimedia applications. The basic idea of UMA is universal or seamless access to the multimedia content by automatic selection or adaptation of content following user's interaction. As mobile phones have grown in popularity and capability, people have become enthusiastic about watching multimedia content using mobile devices and personalizing the content, for example, summarizing the video for real-time retrieval or for easy transmission. In general, the video summarization algorithm will generate a still-image storyboard, which is composed of a collection of salient images extracted from the underlying video sequence (as shown in Fig. 1).

Although plenty of works on video summarization can be found in literature [3]–[6], the transmission issue of video summary has gained little attention. [7] extends the work of [6] into the wireless video streaming domain, however the packet loss factor due to unsatisfactory wireless channel conditions has not been considered in the framework. In [8], packet loss is considered in the video summary transmission, and the key

frames that minimize the expected end-to-end distortion are selected as the summary frames. However, the source coding has not been optimized in the optimization framework, which might directly impact the perceptual quality of the results. In addition, the algorithm does not guarantee a good content coverage aspect of the selected frames because potential packet loss penalty heavily biases the selection process.

In wireless networks, packet loss is mainly due to the fading effect of time-varying wireless channels. Adaptive modulation and coding (AMC) have been studied extensively and advocated at the physical layer, in order to match transmission rates to time-varying channel conditions. For example, to achieve high reliability at the physical layer, one has to reduce the transmission rate using either small size constellations, or powerful but low-rate error-correcting codes [9]–[11]. An alternative way to decrease packet loss rate is to rely on the automatic repeat request (ARQ) protocol at the data link layer, which requests retransmissions for those packets received in error. Obviously, by allowing a very large retransmission number, ARQ can guarantee a very low packet loss rate. However, to minimize delays and buffer sizes in practice, truncated ARQ protocols have been widely adopted to limit the maximum number of transmissions [12].

A variety of techniques have been proposed to address the problem of multimedia delivery over lossy networks. [13] shows that the problem of rate-distortion optimized streaming of an entire presentation can be reduced to the problem of error-cost optimized transmission of an isolated data unit. Based on this observation, a general framework for rate-distortion optimized streaming of packetized media over a lossy packet network is set up for various transmission scenarios. In [13], distortion-rate performance is measured in an average sense, and practical streaming using window and rate control is proposed to overcome a possible large instantaneous rate. [14] and [15] propose a proxy-driven rate-distortion optimized streaming over a lossy packet network, exploiting a proxy located at the edge of the backbone network to coordinate the streaming process. Traffic load brought from packets lost in the last hop is relieved and end-to-end performance is greatly improved.

In this work, within an expected rate-distortion framework where expectation are taken over channel realizations as in [13], [14] and [15], we focus our study on the cross-layer optimization of the video summary transmission over lossy networks. We assume a video summarization algorithm that can select frames based on some optimality criteria is available in the system. Therefore, a cross-layer approach is proposed to jointly optimize the AMC parameters at the physical layer, the

Manuscript received May 25, 2006; revised November 10, 2006. This paper was presented in part at the International Symposium on Multimedia, San Diego, CA, December 2006.

D. Wu and S. Ci are with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Omaha, NE 68182 USA (email: dwu@unlnotes.unl.edu; sci@engr.unl.edu).

H. Wang is with Marvell Semiconductors, Santa Clara, CA, 95054 USA (email: haohong@ieee.org).

Digital Object Identifier 10.1109/JSAC.2007.0705xx.



Fig. 1. An example of a 10-frame video summary of the Stefan sequence.

ARQ parameters at the data link layer, and the source coding parameters at the application layer to achieve the best video quality of reconstructed video clips from received video summaries. Clearly, due to the spectacular characteristics of video summary data, the general cross-layer optimization schemes recently proposed for normal video sequences [16] and [17] do not automatically cover the summary data transmission. As an example, the neighboring summary frames have typically less correlation in order to cover the content variation of the video clip, and thus the normal temporal-based error concealment algorithm considered in the [16] and [17] would not be efficient in the current scenario. In [18], a cross-layer multi-objective optimized scheduler for video streaming over 1xEV-DO system is presented. With the usage of decodability and semantic importance feedback from the application layer to the scheduler, [18] focuses on determining the best allocation of channel resources (time slots) across users, however, the joint optimization of source coding and transmission parameters has not been considered in the framework.

In the proposed framework, we assume the existence of a system controller, whose responsibility is to communicate with each layer and dynamically determine the corresponding parameters that guarantee the best output video quality, and then drive the application system to perform efficiently. In addition, the framework tries to maintain a good content coverage by providing tunable parameters to avoid cases that a number of consecutive summary frames being lost simultaneously. This task is not trivial due to the complexity of the underlying wireless link protocols. *Compared to the existing systems, the novelty of this work is in twofold: first, this is the first cross-layer optimization framework proposed for the coding and transmission of video summary data; second, AMC and ARQ are jointly considered in the cross-layer design, which gives the controller more flexibility in delivering the summary frames.*

The remainder of this paper is organized as follows. Section II provides a brief description of the background techniques adopted in this paper and then highlights the problem formulation. Section III provides the system model of our framework. Section IV describes the link adaptation principles, while Section V details the algorithm of the controller. Section VI provides experimental results and shows the effectiveness of the proposed framework. Finally, concluding remarks are

given in Section VII.

## II. PROBLEM STATEMENT

The section first provides a brief explanation of background techniques used in this paper, including video summarization at the application layer, link adaptation by combining ARQ at the data link layer and AMC at the physical layer. Then a cross-layer optimization problem is formulated for video summary transmission by jointly determining the optimal parameters for each layer.

### A. Background

Clearly in wireless communication applications, video transmission suffers mainly from unreliable channel conditions and excessive delays. In source coding, setting a finer coding parameter will directly improve the coded video quality, however, might increase the transmission time and increase the chance of getting corrupted by the transmission error. Video summary is a special format of the video clip whose correlation between frames are not as high as normal clips, but losing consecutive or continuous summary frames might cause severe damage for understanding the summary content. The summary could be automatic generated or selected with user's interactions. In this work how to generate the video summary is not within the scope of our framework.

At the data link layer, ARQ is widely used to mitigate channel fading and decrease packet error rate (PER). Once some information packets are lost in transmission, retransmission requests are activated and those packets are sent out again. Obviously, by allowing a very large retransmission number, ARQ can guarantee a very low PER. However, large retransmission number means large delay, especially when the round trip time (RTT) of wireless channels is large. Considering the requirement on the allowable maximum delay for transmitting one video summary frame, we need to set a prescribed maximum transmission times for one summary frame.

At the physical layer, AMC has been advocated to enhance the throughput of future wireless data communication systems. In AMC, different size constellations and different rate error-control codes are chosen based on different time-varying channel conditions. For example, in good channel conditions,

AMC schemes with large size constellations and high rate error-control codes can increase the system transmission rate while guaranteeing a good reliability. This means that AMC can effectively decrease transmission delay while satisfying some PER constraint. In this paper, we combine AMC with ARQ to yield a link adaptation to achieve a desirable delay-PER tradeoff for video summary transmissions. With ARQ correcting occasional packet errors at the data link layer, the stringent error control requirement is alleviated for the AMC at the physical layer. In turn, with less stringent error control requirement, AMC schemes with large size constellations and high rate error-control codes are more likely to be chosen, leading to higher transmission rate and smaller transmission delay.

Clearly, the tradeoff among the selected parameters in these layers are mixed, for example, to maintain a reasonable delay, the source coding might choose a coarser parameter, or AMC chooses a larger size constellation or a higher rate FEC channel code, which increases the vulnerability of coded frames and will cause unacceptable video quality. However, if source and channel coding use more bits, with the increase of packet length, the probability of packet loss increases, and ARQ might have to increase the number of retransmission trials to reduce the quality problem due to packet loss, and then will result in excessive delay. Therefore, cross-layer optimization approach is a nature solution to improve the overall system performance.

### B. Problem Formulations

In this work, we propose a cross-layer framework that optimizes the parameter selection in AMC at the physical layer, ARQ at the data link layer and source coding at the application layer to achieve the best video quality of reconstructed video clip from the received video summary.

The following notation will be used. Let us denote by  $n$  the number of frames of a video clip  $\{f_0, f_1, \dots, f_{n-1}\}$ , and  $m$  the number of frames of its video summary  $\{g_0, g_1, \dots, g_{m-1}\}$ . Let  $S_i$  and  $B_i$  be the coding parameters and the resultant consumed bits of the  $i$ th ( $i = 0, 1, \dots, m-1$ ) video summary frame in lossy source coding. The summarization with different coding parameters will produce summary frames with different frame lengths. Large size frames will be fragmented into multiple packets for transmission at lower layers. Let  $Q_i$  denote the number of fragmented packets of the  $i$ th summary frame. Let  $N_{i,q}$  and  $F_{i,q}$  be the number of transmissions and the packet size for the  $q$ th packet of the  $i$ th summary frame, respectively. To improve channel utilization, AMC is designed to update the transmission mode for every transmission and retransmission of each packet. Let  $R_{i,q,n}(A_{i,q,n}, C_{i,q,n})$  be the rate (bits/symbol) of AMC mode used at the  $n$ th transmission attempt when transmitting the  $q$ th packet of the  $i$ th summary frame, where  $A_{i,q,n}$  and  $C_{i,q,n}$  are the corresponding modulation order and coding rate. We assume that the transmission rate of the physical layer channel is fixed, denoted by  $r$  (symbols/second). Clearly, the delay in transmitting the whole summary can be expressed by

$$T = \sum_{i=0}^{m-1} \sum_{q=1}^{Q_i} \sum_{n=1}^{N_{i,q}} \left[ \frac{F_{i,q}(S_i, B_i)}{R_{i,q,n}(A_{i,q,n}, C_{i,q,n}) * r} + T_{\text{RTT}} \right] \quad (1)$$

where  $T_{\text{RTT}}$  is the maximum allowed RTT to get the acknowledgement packet via the feedback channel before a retransmission trial.

Let  $\rho_i(S_i, B_i, N_{\text{max}}, A_{i,q,n}, C_{i,q,n}, \gamma_{i,q,n})$  be the loss probability of the  $i$ th summary frame, where  $N_{\text{max}}$  is the maximum transmission number for one packet and  $\gamma_{i,q,n}$  is the instantaneous channel SNR. Note that any summary frame is possible to get lost during video transmission. However, in order to simplify the problem formulation, we assume the first summary frame would guarantee to be received. Then the expected distortion of the video clip can be calculated by

$$\begin{aligned} E[D] &= \sum_{k=0}^{n-1} E[D(f_k, \tilde{f}_k)] \\ &= \sum_{i=0}^{m-1} \sum_{j=l_i}^{l_{i+1}-1} \sum_{b=0}^i \left\{ (1 - \rho_{i-b}) d[f_j, \tilde{g}_{i-b}(S_{i-b})] \right. \\ &\quad \left. \cdot \prod_{a=0}^{b-1} \rho_{i-a} \right\} \end{aligned} \quad (2)$$

where  $\tilde{f}_k$  is the reconstructed  $k$ th frame from the received summary at the receiver side,  $\tilde{g}_k(S_k)$  is the reconstructed  $k$ th summary frame,  $l_i$  is the index of the summary frame  $g_i$  in the video clip, and function  $d(\cdot)$  is the distortion between two frames. In this work we use the mean squared error (MSE) between the two frames as the metric for calculating the distortion. The same distortion measure for the video summary result has been used in [6]–[8].

The problem at hand can be formulated as

$$\text{Min } E[D], \quad \text{s.t. : } T \leq T_{\text{max}} \quad (3)$$

where  $T_{\text{max}}$  is a given delay budget for delivering the whole video clip.

In this work, we consider the content coverage issue of the received summary. In other words, if a chunk of continuous summary frames are lost due to the channel error, then the coverage of the received summary for the original clip would be degraded significantly. To avoid such a problem but still keep the problem as general as possible, we define  $L$  such that the case of  $L$  or more than  $L$  consecutive summary frames being lost will never happen. For instance, if  $L = 2$  then no neighboring summary frames can be lost together during transmission. So that the distortion can be calculated by

$$\begin{aligned} E[D] &= \sum_{k=0}^{n-1} E[D(f_k, \tilde{f}_k)] \\ &= \sum_{i=0}^{m-1} \sum_{j=l_i}^{l_{i+1}-1} \sum_{b=0}^{\min(i, L-1)} \left\{ (1 - \rho_{i-b}) d[f_j, \tilde{g}_{i-b}(S_{i-b})] \right. \\ &\quad \left. \cdot \prod_{a=0}^{b-1} \rho_{i-a} \right\}. \end{aligned} \quad (4)$$

It is important to realize that the value of  $L$  is a programmable constant by the system, and the introduction of  $L$  does not narrow down the original problem. As you may notice, when we set  $L = m$ , the Eq. (4) is equal to Eq. (2).

If we set  $L = 2$  in (4), it is very clear that for the  $i$ th summary frame, there are only two possibilities: either it

is received or it is lost but its previous summary frame is received. Let us denote by  $G_i$  the chance of the  $i$ th summary frame being not lost, so  $G_i = 1$  means it is guaranteed to be received, otherwise it is not guaranteed. Based on the constraint, we need  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1$  for all  $i \in [0, m-1]$ .  $G_i$  can be guaranteed and derived by link adaptation, which will be discussed in later section.

For delay issue, we hope that the total delay  $T$  of delivering all summary frames satisfies  $T \leq T_{\max}$ . Therefore, the problem is

$$\begin{aligned} \text{Min}E[D], \quad \text{s.t.: } T \leq T_{\max}, \text{ and} \\ \max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1, \quad i \in [0, m-1]. \end{aligned} \quad (5)$$

Once we work out the problem (5), the optimal parameter combinations, i.e., source coding parameter  $S_i$ , AMC modulation order  $A_{i,q,n}$  and channel coding rate  $C_{i,q,n}$ , and ARQ transmission number  $N_{\max}$  are obtained to transmit the  $i$ th summary frame, which minimizes the whole clip distortion and satisfies certain predefined delay constraint.

### III. SYSTEM MODEL

The system model of the proposed framework is shown in Fig. 2, which consists of a 3-layer structure and a controller. At the application layer, summarization is performed on the target video clip and large size summary frames are fragmented into multiple packets for transmission at lower layers. At the data link layer, ARQ protocol is adopted. If an error is detected in a packet, a retransmission request is generated by the receiver, and is sent to the transmitter via a feedback channel. The transmitter arranges retransmission of the requested packet. If a packet is not received correctly after  $N_{\max}$  transmission attempts, we will declare packet loss, then the summary frame to which the lost packet belongs is also regarded as lost. At the physical layer, we assume that multiple transmission modes are available as shown in Table 1, with each mode consisting of a specific modulation and FEC code pair as in 3GPP, HIPERLAN/2, IEEE 802.11a, and IEEE 802.16 standards [19]–[21]. Based on channel state information (CSI) from the channel estimator, the transmitter updates the AMC mode for the next packet transmission. Coherent demodulation and maximum-likelihood (ML) decoding are used at the receiver. The decoded bit streams are mapped to packets, which are pushed upwards to the data link layer. If all fragmented packets of the summary frame  $g_i$  are correctly delivered, the summary frame  $g_i$  is saved into the buffer, and the video clip frames  $f_{l_i}$  through  $f_{l_{i+1}-1}$  are reconstructed with  $g_i$ . If  $g_i$  does not reach the receiver after some fixed time, the video clip frames  $f_{l_i}$  through  $f_{l_{i+1}-1}$  will be reconstructed with the previously received summary frame. Obviously, the receiver only need a buffer that can contain one summary frame, i.e., the latest received summary frame.

From above description, it is obvious that AMC combined with ARQ performs a link adaptation in a joint approach. For a fixed video summary, say  $\{g_0, g_1, \dots, g_{m-1}\}$ , the link adaptation can guarantee the constraint  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1$  in problem (5), and produce the summary frame error rate (FER)  $\rho_i$  and transmission time  $T_i$  for each summary frame  $g_i$ . The detailed

link adaptation and close-form expressions for  $(\rho_i, T_i)$  will be clarified in the following section.

The controller is the most important part of the system, which is equipped with all possible values of the key parameters of each layer. These parameters include the coding parameter  $S$  at the application layer, the allowed maximum transmission number  $N_{\max}$  at the data link layer, and the available AMC modes with modulation order and FEC code rate pair  $(A, C)$ . Note that here  $S, A$ , and  $C$  are parameter allocation vectors for  $m-1$  summary frames, for example,  $S = \{S_1, S_2, \dots, S_{m-1}\}$ .

The following is a brief list of performing flows of our proposed framework.

- When there is a video clip to transmit, based on the current average SNR  $\bar{\gamma}$  from the channel estimator, from all possible values of parameter set  $\{S, N_{\max}, A, C\}$ , the controller first calculates all possible theoretical values of the pair  $(\rho_i, T_i)$  for all possible summary frames by using the close-form expressions of link adaptation performance with the constraint  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1, i \in [0, m-1]$ .
- With the total delay budget  $T_{\max}$ , the controller use all possible  $(\rho_i, T_i)$ s for the whole summary to solve problem (5). The group values of  $\{S, N_{\max}, A, C\}$  corresponding to the optimal solution of problem (5) are the optimal parameters to transmit the whole video summary.
- The obtained optimal parameters  $\{S, N_{\max}, A, C\}$  are assigned to the corresponding layers, then the whole video summary is sent out frame by frame.
- Corresponding video clip frames are reconstructed with the newly received summary frame.

We next list the operating assumptions adopted in this paper.

- The channel is frequency flat, remains time invariant during a packet, but varies from packet to packet. Thus, AMC is adjusted on a packet-by-packet basis. In other words, AMC scheme is updated for every transmission and retransmission attempt. The channel quality is captured by a single parameter, namely the received SNR  $\gamma$ . we adopt Rayleigh channel model to describe  $\gamma$  statistically. The received SNR  $\gamma$  per packet is thus a random variable with a probability density function (pdf):
- $$p_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right) \quad (6)$$
- where  $\bar{\gamma} := E\{\gamma\}$  is the average received SNR.
- Perfect channel state information (CSI) is available at the receiver. The corresponding mode selection is fed back to the transmitter without error and latency. This assumption could be at least approximately satisfied by using a fast feedback channel with powerful error control information as adopted in IEEE 802.16 [21].
  - Error detection based on CRC is perfect, provided that sufficiently reliable error detection CRC codes are used.

### IV. LINK ADAPTATION

In this section, we explain how link adaptation can guarantee the constraint  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1$ , and derive the close-form expression of  $(\rho_i, T_i)$ .

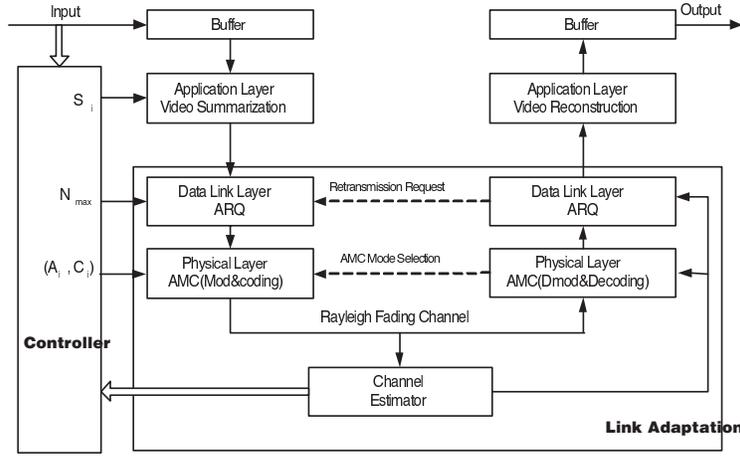


Fig. 2. The system model of cross-layer optimization for video summary transmission

TABLE I  
AMC MODES AT THE PHYSICAL LAYER

	Mode1	Mode2	Mode3	Mode4	Mode5	Mode6
Modulation	BPSK	QPSK	QPSK	16-QAM	16-QAM	64-QAM
Coding Rate $C_m$	1/2	1/2	3/4	9/16	3/4	3/4
$R_m$ (bits/sym.)	0.50	1.00	1.50	2.25	3.00	4.50
$a_m$	1.1369	0.3351	0.2197	0.2081	0.1936	0.1887
$b_m$	7.5556	3.2543	1.5244	0.6250	0.3484	0.0871

Actually, it is impossible to strictly guarantee the constraint  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1$ , due to the fading characteristics of wireless channels. Let  $P_L$  be the probability that  $L$  consecutive summary frames are lost simultaneously. We assume  $P_L$  to be a very small value, say  $10^{-2}$ , to approximate the constraint  $\max(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1$  is satisfied. Then the goal of link adaptation becomes to guarantee  $P_L$  with the least total transmission delay.

Since the processing unit of link adaptation is packet, we need to transform  $P_L$  into target packet error rate  $P_{\text{target}}$  of lower layers. Let  $L_s$ ,  $L_f$  and  $L_a$  be summary frame size, fragmentation packet size and the actual packet length of link adaptation, respectively. According to different summary frame sizes, there are two possible cases:

- The summary frame size is smaller than the fragmentation packet size. Since there is no need to do fragmentation, we have  $P_{\text{target}} = P_L^{1/L}$  and  $L_a = L_s$ .
- The summary frame size is larger than the fragmentation packet size, where fragmentation is necessary. A summary frame of length  $L_s$  will be fragmented into  $N_p = \lceil L_s/L_f \rceil$  packets.  $\lceil \cdot \rceil$  is the smallest integer greater than or equal to a given real number. The actual packet size  $L_a$  of the first  $\lceil L_s/L_f \rceil - 1$  packets equals to  $L_f$ , and the actual packet size  $L'_a$  of the final packet is  $L_s - \lfloor L_s/L_f \rfloor \cdot L_f$ . The target PER should be as follows:

$$P_{\text{target}} = 1 - (1 - P_L^{1/L})^{1/N_p} \quad (7)$$

In the above both cases,  $P_{\text{target}}$  can be regarded as the required PER at the data link layer. Next we explain how to guarantee  $P_{\text{target}}$  with transmission packet size  $L_a$  by AMC and ARQ. Let us define a PER upper bound  $P_{\text{AMC}}$  such that the instantaneous PER is guaranteed to be no greater than

$P_{\text{AMC}}$  for each chosen AMC mode at the physical layer. Then the PER at the data link layer after  $N_{\text{max}}$  transmissions is no larger than  $P_{\text{AMC}}^{N_{\text{max}}}$ . To satisfy  $P_{\text{target}}$ , we need to impose

$$P_{\text{AMC}}^{N_{\text{max}}} = P_{\text{target}}, \text{ i.e., } P_{\text{AMC}} = P_{\text{target}}^{1/N_{\text{max}}} \quad (8)$$

We assume each bit inside the packet has the same bit error rate (BER) and bit-errors are uncorrelated, the PER can be related to the BER through

$$\text{PER} = 1 - (1 - \text{BER})^{L_a} \quad (9)$$

for a packet containing  $L_a$  bits. For any AMC mode, to guarantee the upper bound  $P_{\text{AMC}}$ , the required BER to achieve is

$$\text{BER}_{\text{AMC}} = 1 - (1 - P_{\text{AMC}})^{1/L_a} \quad (10)$$

Since exact closed-form BERs for the AMC modes in Table 1 are not available, to simplify the AMC design, we adopt the following approximate BER expression:

$$\text{BER}_m(\gamma) = a_m \exp(-b_m \gamma) \quad (11)$$

where  $m$  is the mode index and  $\gamma$  is the received SNR. Parameters  $a_m$  and  $b_m$  are obtained by fitting (11) to the exact BER. To guarantee  $P_{\text{AMC}}$  with the least delay when transmitting a packet, we set the mode switching threshold  $\gamma_m$  for the AMC mode  $m$  to be the minimum SNR required to achieve  $\text{BER}_{\text{AMC}}$ . By (11)  $\gamma_m$  can be expressed as

$$\gamma_m = \frac{1}{b_m} \ln \left( \frac{a_m}{\text{BER}_{\text{AMC}}} \right), \quad m = 1, 2, \dots, M, \quad (12)$$

$$\gamma_{M+1} = +\infty,$$

where  $M$  is the total number of AMC modes available ( $M = 6$  in this paper).

Since the instantaneous PER is upper-bounded by  $P_{AMC}$  in our AMC design, the average PER at the physical layer will be lower than  $P_{AMC}$ . Taking expectations over channel realizations, the average PER at the physical layer is

$$\begin{aligned}\bar{P} &= \frac{1}{P_T} \sum_{m=1}^M \int_{\gamma_m}^{\gamma_{m+1}} \text{PER}_m(\gamma) p_\gamma(\gamma) d\gamma \\ &= \frac{1}{P_T} \sum_{m=1}^M \int_{\gamma_m}^{\gamma_{m+1}} [1 - (1 - a_m \exp(-b_m \gamma))^{L_a}] \\ &\quad \cdot p_\gamma(\gamma) d\gamma\end{aligned}\quad (13)$$

where  $P_T = \int_{\gamma_1}^{+\infty} p_\gamma(\gamma) d\gamma$  is the probability that channel has no deep fades and at least one AMC mode can be adopted. Similarly, the average delay for one transmission attempt at the physical layer can be expressed as

$$\bar{T} = \frac{1}{P_T} \sum_{m=1}^M \int_{\gamma_m}^{\gamma_{m+1}} \left( \frac{L_a}{R_m \cdot r} + T_{\text{RTT}} \right) p_\gamma(\gamma) d\gamma \quad (14)$$

Then the average number of transmission attempts per packet can be found as [22]

$$\begin{aligned}\bar{N} &= 1 + \bar{P} + \bar{P}^2 + \dots + \bar{P}^{N_{\max}-1} \\ &= \frac{1 - \bar{P}^{N_{\max}}}{1 - \bar{P}}.\end{aligned}\quad (15)$$

Then the actual PER at the data link layer is

$$P_{\text{actual}} = \bar{P}^{N_{\max}}, \quad (16)$$

and the actual transmission delay for each packet at the data link layer is

$$T_{\text{actual}} = \bar{T} \cdot \bar{N}. \quad (17)$$

When to calculate the actual FER  $\rho_i$  and the actual delay  $T_i$  for transmitting the  $i$ th summary frame, two cases should be considered as mentioned before:

- If the summary frame size  $L_s$  is smaller than the fragmentation packet size  $L_f$ , we adopt  $L_a = L_s$  to compute  $\bar{P}(L_a)$  and  $\bar{T}(L_a)$  and we can have  $\rho_i = P_{\text{actual}}(L_a)$  and  $T_i = T_{\text{actual}}(L_a)$  with (13)-(17).
- If the summary frame size  $L_s$  is larger than the fragmentation packet size  $L_f$ , we adopt  $L_a = L_f$  and  $L'_a = L_s - \lfloor L_s/L_f \rfloor \cdot L_f$  to compute  $P_{\text{actual}}(L_a)$ ,  $T_{\text{actual}}(L_a)$ ,  $P'_{\text{actual}}(L'_a)$  and  $T'_{\text{actual}}(L'_a)$ . Then we can have

$$\rho_i = 1 - (1 - P_{\text{actual}})^{N_p-1} \cdot (1 - P'_{\text{actual}}), \quad (18)$$

$$T_i = (N_p - 1) \cdot T_{\text{actual}} + T'_{\text{actual}}. \quad (19)$$

The above closed-form expressions of  $\rho_i$  and  $T_i$  will be used by the controller to calculate all possible  $(\rho_i, T_i)$  to solve problem (5), which we will detail in section V.

## V. ALGORITHM OF THE CONTROLLER

### A. Optimal Solution

Since the problem (5) is a constrained minimization problem, it can be solved by Lagrangian relaxation. So the problem can be converted into

$$\begin{aligned}\text{Min}\{E[D] + \lambda T\}, \quad \text{s.t. :} \\ \text{max}(G_i, G_{i-1}, \dots, G_{i+1-L}) = 1, i \in [0, m-1].\end{aligned}\quad (20)$$

The target to be minimized can be derived as the following Lagrangian cost function:

$$\begin{aligned}J_\lambda &= E[D] + \lambda T \\ &= \sum_{i=0}^{m-1} \sum_{j=l_i}^{l_{i+1}-1} \sum_{b=0}^{\min(i, L-1)} \left\{ (1 - \rho_{i-b}) d[f_j, \tilde{g}_{i-b}(S_{i-b})] \right. \\ &\quad \left. \cdot \prod_{a=0}^{b-1} \rho_{i-a} + \lambda T_i \right\}.\end{aligned}\quad (21)$$

Let us define a cost function  $H_i(u_i)$  to represent the sum of distortion and delay for up to  $i$ th summary frame, where  $u_i$  represents the parameter vector  $\{S_i, N_{\max}, A_{i,q,n}, C_{i,q,n}, \gamma_{i,q,n}\}$ . Clearly it can be observed that

$$\begin{aligned}H_i(u_i) &= H_{i-1}(u_{i-1}) + \sum_{j=l_i}^{l_{i+1}-1} \sum_{b=0}^{\min(i, L-1)} \left\{ (1 - \rho_{i-b}) \right. \\ &\quad \left. \cdot d[f_j, \tilde{g}_{i-b}(S_{i-b})] \prod_{a=0}^{b-1} \rho_{i-a} + \lambda T_i \right\},\end{aligned}\quad (22)$$

which means the process of choosing  $u_i$  for the  $i$ th summary frame is independent of  $\{u_0, u_1, \dots, u_{i-2}\}$ , the parameters selected for the first  $i-1$  summary frames. This is the fundamental of dynamic programming (DP). So the optimal solution can be found by a shortest path algorithm.

As a toy example, we assume there are three summary frames  $\{g_0, g_1, g_2\}$  to be sent and assume  $L = 2$ . In addition, we suppose for each summary frame, there are  $k$  different source coding options. Then the path graph will be like Fig. 3. In this figure, each node  $u_i^a$  corresponds to a cost value  $H(u_i^a)$ . The weight  $h(u_i^a u_{i+1}^b)$  on each branch from node  $u_i^a$  to  $u_{i+1}^b$  corresponds the incremental cost value when transmitting the  $(i+1)$ th summary frame with the  $b$ th source coding option.  $h(u_i^a u_{i+1}^b)$  can be computed by the second term of the right hand side of (22).

As discussed before, the solution to problem (5) is to minimize the average distortion  $D$  for a total delay budget  $T_{\max}$  in transmitting a whole video summary. With the path graph like above, the goal of the controller is to find the shortest path in the graph with the forward DP. The obtained shortest path has the minimal distortion  $D$ , and at the same time indicates the optimal choice of parameters  $\{S_i, N_{\max}, A_{i,q,n}, C_{i,q,n}\}$  for source coding and transmitting the  $i$ th summary frame.

For  $J_\lambda$  in (21), it has been shown [23] that if there is a  $\lambda^*$  such that

$$\{S^*, N_{\max}^*, A^*, C^*\} = \arg \min J_{\lambda^*}(S, N_{\max}, A, C) \quad (23)$$

leads to  $T(S, N_{\max}, A, C) = T_{\max}$ , then  $\{S^*, N_{\max}^*, A^*, C^*\}$  is also an optimal solution to (5). It is well known that when  $\lambda$  sweeps from zero to infinity, the solution to problem (23) traces out the convex hull of the distortion delay curve, which is a non-increasing function. Hence  $\lambda^*$  can be obtained via a fast convex recursion in  $\lambda$  using the bisection algorithm.

Next we list the algorithm to find  $\lambda^*$ .

- Step1: We judiciously choose two values of  $\lambda$ ,  $\lambda_l$  and  $\lambda_u$  with  $\lambda_l \leq \lambda_u$  which satisfy the relation:

$$\sum_i T_i^*(\lambda_u) \leq T_{\max} \leq \sum_i T_i^*(\lambda_l) \quad (24)$$

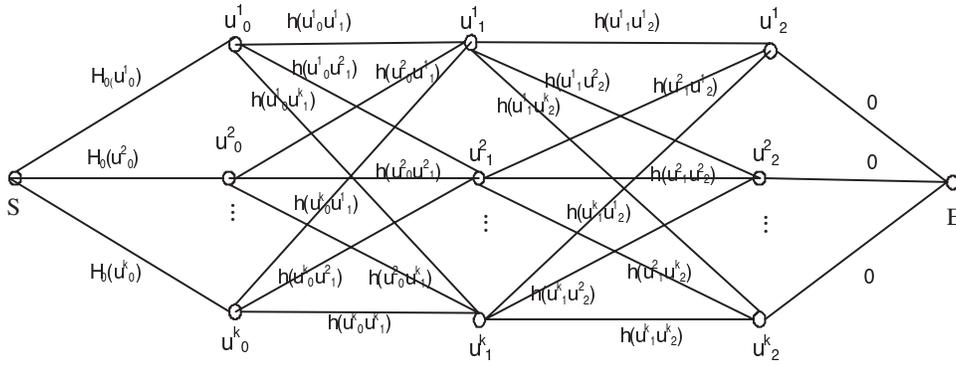


Fig. 3. Path graph of a 3-frame toy video summary transmission

where  $\sum_i T_i^*(\lambda)$  is the total delay corresponding to the shortest path found by forward DP. A conservative choice for a solvable problem would be  $\lambda_l = 0$  and  $\lambda_u = \infty$ ;

- Step 2:  $\lambda_{\text{next}} \leftarrow \frac{\lambda_l + \lambda_u}{2}$ .
- Step 3: Perform forward DP through the path graph for  $\lambda_{\text{next}}$ ;  
 $\implies$  if  $\{\sum_i T_i^*(\lambda_{\text{next}}) = \sum_i T_i^*(\lambda_u)\}$ , then stop,  $\lambda^* = \lambda_u$ ;  
 $\implies$  else if  $(\sum_i T_i^*(\lambda_{\text{next}}) > T_{\text{max}})$ ,  $\lambda_l \leftarrow \lambda_{\text{next}}$ , Go to Step 2,  
 $\implies$  else  $\lambda_u \leftarrow \lambda_{\text{next}}$ , Go to step 2.

Thus  $\sum_i T_i^*(\lambda)$  is made successively closer to  $T_{\text{max}}$  and finally we obtain the expected  $\lambda^*$ . With  $\lambda^*$ , we perform DP for the last time and obtain the optimal shortest path. The values of  $\{S_i, N_{\text{max}}, A_{i,q,n}, C_{i,q,n}\}$  corresponding to the shortest path are just the optimal parameter values for source coding and transmitting the  $i$ th summary frame.

### B. Implementation Considerations

From the above analysis, we can say that problem (5) is converted into a graph theoretic problem of finding the shortest path in a directed acyclic graph (DAG) [24]. The computational complexity of the above algorithm is  $O(N \times |U|^L)$ , with  $|U|$  denoting the cardinality of  $U$ , which depends on the number of the optional values of parameters  $\{S, N_{\text{max}}, A, C\}$ , but is still much more efficient than the exponential computational complexity of an exhaustive search algorithm. Clearly for cases with smaller  $L$ , the complexity is quite practical to perform the optimization. On the other hand, for larger  $L$ , the complexity can be limited by reducing the cardinality of  $U$ . The practical solution would be an engineering decision and tradeoff between the computational capability and optimality of the solution. For storage issue, it is important to emphasize that the problem formulation and the proposed solution are quite generic and flexible for devices with various storage and computational capabilities. For the transmitter with a buffer size that only allows to store some portion of the video clip, the clip has to be divided into a number of segments and problem (5) is solved for each segment. In such cases, although the solution is not full optimal for the video clip, the optimization would still bring sufficient gains compared to those without optimization.

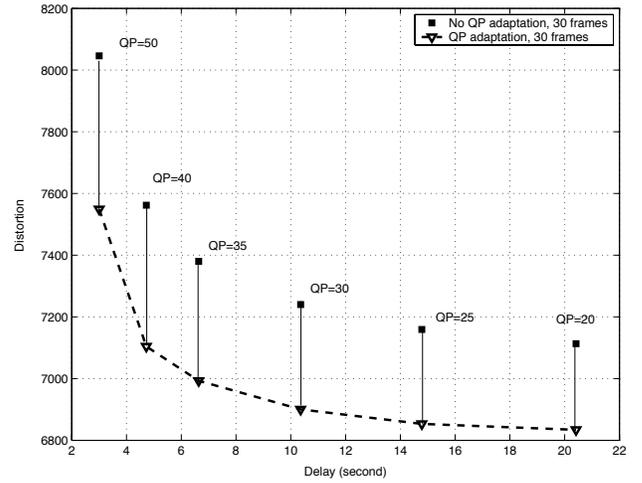


Fig. 4. Distortion vs. delay comparison between QP adaptation and no QP adaptation with 30 summary frames

## VI. EXPERIMENTAL RESULTS

In this section, experiments are designed using H.264/AVC JM 10.2 for the video clip called "Glasgow", which is a typical test clip. For comparison, we summarize the first 300 frames into 30 and 60 summary frames, respectively. The case with more summary frames means higher sampling rate thus less distortion. To simplify the problem, we compress the summary by choosing different QP (quantization step size), and we consider 10 possible QPs (5, 10, 15, 20, 25, 30, 35, 40, 45, 50). According to each QP, the frames have different rates and distortion values. The video summary is coded with intra-coding mode for each summary frame due to the less correlation between neighbor frames. In addition, without loss of generality, we consider the case of  $L = 2$ , in other words, we impose the constraint  $\max(G_i, G_{i-1}) = 1$  ( $i \in [0, m-1]$ ) which needs to be guaranteed by the link adaptation. Besides parameters to be optimized, we assume fixed channel transmission rate  $r = 6 \times 10^6$  symbols/second and fixed round trip time  $T_{\text{RTT}} = 100$  milliseconds in our experiment.

Fig. 4 and Fig. 5 are comparisons between QP adaptation and No QP adaptation. In both figures, the average channel SNR  $\bar{\gamma}$  is 25dB and we fix  $P_L = 10^{-2}$  and  $N_{\text{max}} = 3$ .

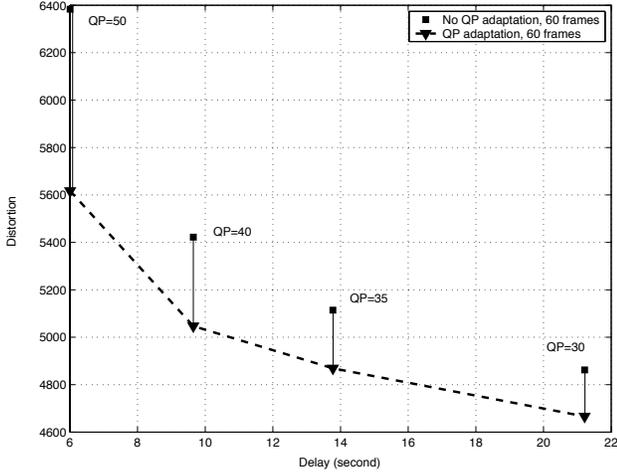


Fig. 5. Distortion vs. delay comparison between QP adaptation and no QP adaptation with 60 summary frames

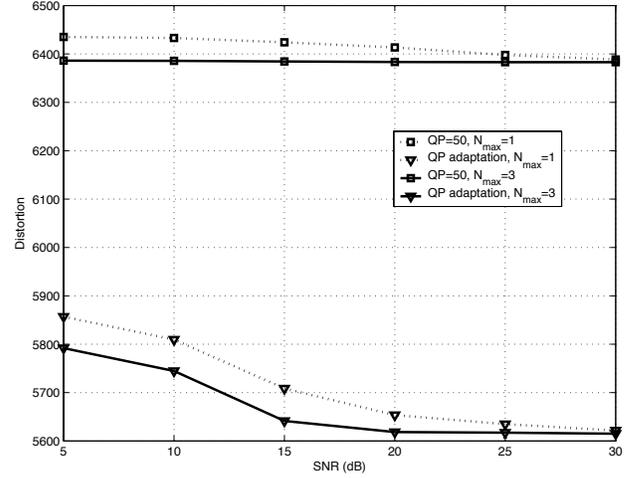


Fig. 7. Distortion vs. average SNR comparison between QP adaptation and QP=50 with different  $N_{\max}$  for a 60-frame video summary

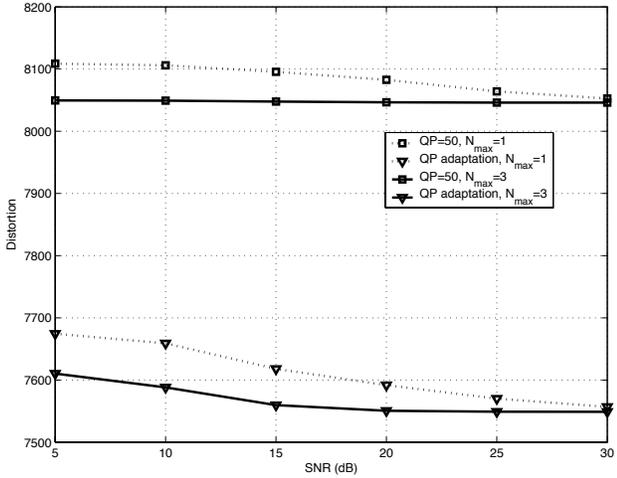


Fig. 6. Distortion vs. average SNR comparison between QP adaptation and QP=50 with different  $N_{\max}$  for a 30-frame video summary

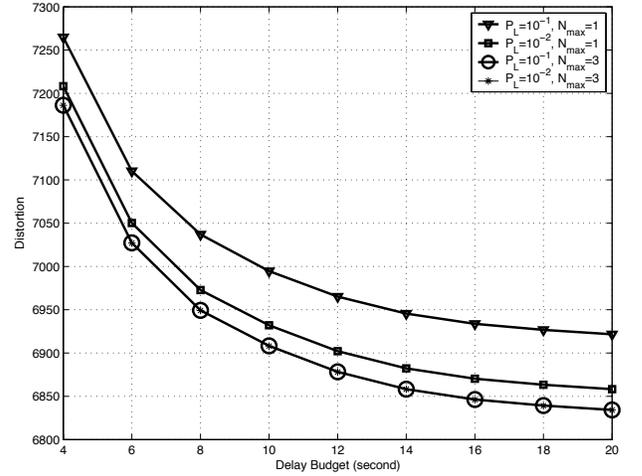


Fig. 8. Comparisons of distortion vs. delay budget with different  $P_L$

$P_L$  is the target probability that  $L = 2$  consecutive frames are being lost simultaneously, which should be small enough to approximately satisfy the constraint  $\max(G_i, G_{i-1}) = 1$ . In Fig. 4 where the total summary frame number is 30, the square nodes show the distortion-delay pairs when the video summary is source coded with the labelled QPs. The 'v' nodes refer to the distortion-delay budget pairs with QP adaptation when delay budget is set equal to the delay time that the corresponding labelled single QP takes to transmit the video summary. We can observe that QP adaptation, i.e., the proposed cross-layer framework, has much distortion gain up to 6.2% over fixed QP video transmission when the delay is small. In the case of summary frame number equal to 60 as in Fig. 5, much more significant distortion gain up to 12% can be obtained in small delay regions.

Fig. 6 shows the distortion vs. SNR comparisons between QP adaptation and QP=50 with different prescribed maximum transmission number for ARQ. Due to link adaptation performed by ARQ and AMC in a cross-layer fashion, both QP adaptation and QP=50 have a stable distortion level along all SNR values. Of course here the delay difference of different

schemes are not considered. We also notice that for either QP adaptation or QP=50,  $N_{\max} = 3$  has better performance than  $N_{\max} = 1$ . This is because the case with  $N_{\max} = 3$  can achieve lower actual PER than with  $N_{\max} = 1$  even though they both aim to guarantee  $P_L = 10^{-2}$ . The same conclusion goes to Fig. 7 where the total summary frame number is 60.

Different distortion vs. delay budget with different  $P_L$  ( $L=2$  in this paper) is shown in Fig. 8. We observe that there is a large distortion-delay difference between  $P_L = 10^{-1}$  and  $P_L = 10^{-2}$ . Once  $P_L$  achieves  $10^{-2}$ , there is no big distortion vs. delay difference even though  $N_{\max}$  is different. However, in the two cases of with different  $P_L$  and same  $N_{\max} = 3$ , the difference in distortion vs. delay is marginal. This is because with larger  $N_{\max}$ , the actual PER is much lower than  $P_L$ . From this figure, we can conclude that the maximum transmission number impacts much the video transmission quality with our cross-layer optimization framework. With same delay budget, a larger allowed maximum transmission number leads to better video transmission quality.

Fig. 9 shows the distortion vs. delay budget with our proposed framework when the summary frame number is 30 and 60 respectively both with  $\bar{\gamma} = 25$ dB. With the same delay

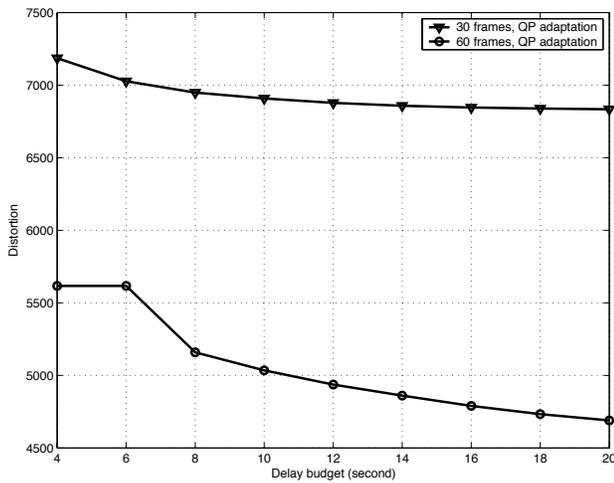


Fig. 9. Distortion vs. delay budget comparison with different summary frame number

budget, the case with 60 frames has better performance than the case with 30 frames. This is because with higher sampling rate (that is, using 60 summary frames instead of 30), the similarity and correlation between neighbor summary frames have increased. Therefore, the distortion caused by losing one frame is reduced in this case because the lost frame would be concealed by its neighbor summary frame with higher similarity.

## VII. CONCLUSION

In this paper, we have proposed an optimization framework for delivering video summary frames over wireless networks in a cross-layer fashion. The proposed framework seamlessly integrates the source coding at the application layer, ARQ at the data link layer and adaptive modulation and coding schemes at the physical layer. Within the delay-distortion theoretical framework, all major parameters at each layers are jointly optimized in a way to achieve the best video quality while satisfy the delay budget imposed by the video summary frames. Both theoretical and experimental results have been presented in this paper. Simulation results show that the proposed optimization framework can achieve more than 10% distortion gain, especially when the delay budget is small.

Although the framework in this paper operates in an optimal sense, some assumptions may not always hold true, such as perfect CSI is available at the receiver and the feedback channel is error and latency free. In addition, the paper only considered the scenario of a single-user link with single-transmit and single-receive antennas. Therefore, there are many potential extensions of this work to carry out in the future. One direction is to generalize this work to situations where multiuser and multiple links exist with multitransmit and multireceive antennas. Another direction that is worth pursuing is to extend this work from summary transmission to normal video transmission with acceptable complexity, coupling more video coding parameters in the cross-layer optimization framework.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and valuable suggestions which improved the quality of this paper.

## REFERENCES

- [1] P. V. Beek *et al.*, "Metadata-driven multimedia access," *IEEE Signal Processing Mag.*, pp. 40–52, Mar. 2003.
- [2] B. L. Tseng, C. Lin, and J. R. Smith, "Using MPEG-7 and MPEG-21 for personalizing videoMetadata-driven multimedia access," *IEEE Multimedia*, vol. 11, pp. 40–53, Jan. 2004.
- [3] D. DeMenthon, V. Kobla, and D. Doermann, "Video summarization by curve simplification," in *Proc. ACM Multimedia*, July 1998.
- [4] A. Hanjalic and H. Zhang, "An integrated scheme for automatic video abstraction based on unsupervised cluster-validity analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, Dec. 1999.
- [5] Y. Gong and X. Liu, "Video summarization using singular value decomposition," *Computer Vision and Pattern Recognition*, vol. 2, pp. 13–15, June 2000.
- [6] Z. Li *et al.*, "Rate-distortion optimal video summary generation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, pp. 1550–1560, Oct. 2005.
- [7] Z. Li, F. Zhai, and A. K. Katsaggelosothers, "Video Summarization for Energy Efficient Wireless Streaming," in *Proc. SPIE VCIP'99*, Beijing, China, 2005.
- [8] P. V. Pahalawatta *et al.*, "Rate-Distortion Optimized Video Summary Generation and Transmission over Packet Lossy Networks," in *Proc. SPIE IVCP'05*, San Jose, US, 2005.
- [9] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, pp. 844–855, June 1999.
- [10] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.
- [11] M. S. Alouini and A. J. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels," *Kluwer J. Wireless Communications*, vol. 13, pp. 119–143, May 2000.
- [12] E. Malkamaki and H. Leib, "Performance of truncated type-II hybrid ARQ schemes with noisy feedback over block fading channels," *IEEE Trans. Commun.*, vol. 48, pp. 1477–1487, Sept. 2000.
- [13] P. A. Chou and Z. Miao, "Rate-Distortion Optimized Streaming of Packetized Media," *IEEE Trans. Multimedia*, vol. 8, pp. 390–404, Apr. 2006.
- [14] J. Chakareski, P. A. Chou, and B. Girod, "Information Capacity and Power Control in Single-Cell Multiuser Communications," in *IEEE Workshop on Multimedia Signal Processing*, S. Thomas, Ed., Dec. 2002, pp. 49–52.
- [15] J. Chakareski and P. A. Chou, "RaDiO Edge: Rate-distortion optimized proxy-driven streaming from the network edge," *IEEE/ACM Trans. Networking*, 2006.
- [16] F. Zhai *et al.*, "Rate-distortion optimized hybrid error control for packetized video communications," *IEEE Trans. Image Processing*, vol. 15, pp. 40–53, Jan. 2006.
- [17] H. Wang, F. Zhai, Y. Eisenburg, and A. K. Katsaggelosothers, "Cost-distortion optimal unequal error protection for object-based video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 1505–1516, Dec. 2005.
- [18] T. Ozcelebi, F. D. Vito, M. O. Sunay, M. R. C. A. M. Tekalp, and J. D. Martin, "Cross-Layer Scheduling with Content and Packet Priorities for Optimal Video Streaming over 1xEV-DO," in *Proc. VLBV'05*, Sardinia, Italy, Sept. 2005, pp. 76–83.
- [19] *Physical Layer Aspects of UTRA High Speed Downlink Packet Access (release 4)*, 3GPP TR 25.848 V4.0.0 Std. V4.0.0, 2004.
- [20] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, "A Comparison of the HIPERLAN/2 and IEEE 802.11a Wireless LAN Standards," *IEEE Commun. Mag.*, vol. 40, pp. 172–180, May 2002.
- [21] *Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, 2002.
- [22] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer Combining of Adaptive Modulation and Coding with Truncated ARQ over Wireless Links," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1746–1755, Sept. 2004.
- [23] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

- [24] *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.



**Dalei Wu** (S'05) received his B.S. and M.Eng. degrees in electrical engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively. He is currently working toward the Ph.D. degree in the Department of Computer and Electronics Engineering, the University of Nebraska-Lincoln, Lincoln, USA.

He was a research assistant with the Department of Electronic Engineering, City University of Hong Kong, during the winter of 2004. From 2004 to 2005, he was a member of OFDMA Technical Staff

with ZTE Communications, Inc., Shenzhen, China. His research interests include cross-layer design and optimization over wireless networks, wireless multimedia communications.

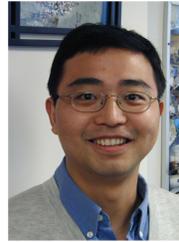


**Song Ci** (S'98-M'02-SM'06) is an assistant professor of computer and electronics engineering at the University of Nebraska-Lincoln. He received his B.S. from Shandong University of Technology (now Shandong University), Jinan, China, in 1992, M.S. from Chinese Academy of Sciences, Beijing, China, in 1998, and Ph.D. from the University of Nebraska-Lincoln in 2002, all in Electrical Engineering.

He also worked with China Telecom(Shandong) as a telecommunication engineer from 1992 to 1995, and with Wireless Connectivity Division of 3COM

Cooperation, Santa Clara, CA, as a R&D co-op in 2001. Prior to joining University of Nebraska-Lincoln, he was an assistant professor of computer science at the University of Massachusetts Boston and the University of Michigan-Flint. He is the founding director of Intelligent Pervasive Computing Laboratory (IPC Lab) at the Peter Kiewit Institute of the University of Nebraska.

His research interests include: cross-layer design for QoS Provisioning in wireless data networks, resource allocation in heterogeneous wireless networks, sensor network management and data mining, and low-power embedded real-time sensor system design. He has published more than fifty research papers in referred journals and international conferences in those areas. He currently serves as an Associate Editor in the Editorial Board of Wiley Wireless Communications and Mobile Computing (WCMC) and a Guest Editor of IEEE Network Magazine Special Issue on Wireless Mesh Networks: Applications, Architectures and Protocols. He also serves as the TPC co-chair of IEEE ICCCN 2007, the TPC co-chair of IEEE WLN 2007, the TPC co-chair of Wireless Applications track at IEEE VTC 2007 Fall, and as reviewers for numerous referred journals and technical committee members for many international conferences. He is a senior member of the IEEE and a member of the ACM.



**Haohong Wang** (S'03-M'04) received the B.S. degree in computer science, and the M.Eng. degree in computer & its application both from Nanjing University, China, and the M.S. degree in computer science from University of New Mexico, Albuquerque, and the Ph.D. degree in Electrical and computer engineering from Northwestern University, Evanston, Illinois.

He is currently a System Architect at Marvell Semiconductors at Santa Clara, California. Prior to joining Marvell, he held various technical positions at AT&T, Catapult Communications, and Qualcomm. Dr. Wang's research areas are multimedia information processing and communications. He has published more than 30 articles in peer-reviewed journals and International conferences. He is the inventor of more than 20 U.S. patents and pending applications.

Dr. Wang is a member of the IEEE Visual Signal Processing and Communications Technical Committee, IEEE Multimedia and Systems Applications Technical Committee, and the IEEE Multimedia Communications Technical Committee. He is the Founding Chair of the Steering Committee of the annual International Symposium on Multimedia over Wireless. He has served as a Guest Editor of the IEEE Communications Magazine Special Issue on Advances in Visual Content Analysis and Adaptation for Multimedia Communications, the journal of Wireless communications and Mobile Computing Special Issue on Video Communications for 4G Wireless Systems, and the journal of Advances in Multimedia Special Issue on Cross-layer Optimized Wireless Multimedia Communications. He is also the co-editor of Computer Graphics (Publishing House of Electronics Industry, 1997). He has served as the Technical Program Co-Chair of the 2005 IEEE WirelessCom Symposium on Multimedia over Wireless (Maui, Hawaii), and the 2006 International Symposium on Multimedia over Wireless (Vancouver, Canada). He is currently serving as the Technical Program Chair of the 16th IEEE International Conference on Computer Communications and Networks (ICCCN 2007) and the General Co-Chair of the First International Workshop on Multimedia Analysis and Processing (IMAP 2007) to be held in Honolulu, Hawaii, August 2007.