WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2007; **7**:131–142
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.469

Review of content-aware resource allocation schemes for video streaming over wireless networks

Peshala V. Pahalawatta*,† and Aggelos K. Katsaggelos
Electrical Engineering and Computer Science Department, Northwestern University, Evanston, I.L. 60208, U.S.A.

Summary

As wireless technology evolves towards its fourth generation (4G) of development, the prospect of offering multimedia services such as on-demand video streaming and video conferencing to wireless mobile clients becomes increasingly more viable. The eventual success of such applications depends on the efficient management of the limited system resources while taking into account the time-varying wireless channel conditions as well as the varying multimedia source content. In this paper, we review some of the recent advances in cross-layer design schemes, which aim at providing significant gains in performance for video streaming systems through content-aware resource allocation. Advances in both, real-time video streaming, where the video is encoded and transmitted in real-time, as well as, on-demand video streaming, where the video is pre-encoded in a media server, are considered. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: video streaming; cross-layer optimization; wireless video; multi-user video transmission

1. Introduction

The anticipated popularity of mobile multimedia devices drives wireless technology towards its fourth generation (4G) of development. The 4G networks are expected to provide higher data rates and seamless connectivity, thus enabling users to access, store, and disseminate multimedia content without any restriction on mobility. On-demand video and video conferencing over mobile devices are among the many potential 'killer applications' that stand to gain from the high data rates offered by such emerging wireless networks. The eventual success of such applications depends on the efficient management of system resources such as transmission power, bandwidth, and even time, in the case of scheduling of packets with attached delay constraints. Resource allocation schemes must also take into account the time-varying and error-prone nature of wireless channels in a mobile environment. Great strides have been made in the design of resource allocation schemes that would enable the streaming of high quality video over wireless networks. In this paper, we review some of the recent advances in this field and discuss new opportunities for research that would enhance our current capabilities.

Any wireless video streaming system is composed of three high-level components. They are: (1) the server, which is either a media server that contains a collection of pre-encoded video sequences that can be requested by its clients or a device that acquires video in real-time,

*Correspondence to: Peshala V. Pahalawatta, Electrical Engineering and Computer Science Department, Northwestern University, Evanston, IL 60208, U.S.A.
†E-mail: pesh@eecs.northwestern.edu

compresses it, and then transmits the compressed video to its clients over the network; (2) the scheduler, which allocates network resources that determine the transport channels to be used for data transmission; and (3) the clients that receive the data. It is becoming increasingly apparent that efficient video streaming schemes will require the joint adaptation of source parameters such as coding modes for macroblocks (MBs) at the encoder or server, and network and channel parameters such as packet transmission schedules, and channel bandwidth allocations at the scheduler, while also being aware of the video decoding and error-resilience schemes employed at the receiving client [1–4]. Jointly adapting such parameters controls the allocation of the available resources across the multiple OSI network layers, from the application layer (e.g., through source bit allocation) to the physical layer (e.g., through power allocation). In this paper, we review such 'cross-layer' design schemes that have been proven to provide significant gains over the traditional layered schemes. Readers are also referred to Reference [5], which provides an interesting and important perspective on cross-layer design.

We consider two different scenarios for video streaming where cross-layer design schemes have been shown to be helpful. The first scenario is the point-to-point streaming of video in real-time. Applications that encompass this scenario include video conferencing, as well as applications where mobile devices may be used to record a scene and transmit the video to another device or user in real-time. In such applications, the video must first be transmitted over an uplink channel to a wireless base station and then transmitted downlink to the receiving user. The video is encoded-on-the-fly and the transmitting device has control over the video coding parameters such as quantization step sizes of the encoded video blocks, as well as network parameters such as the allocated transmission power and bandwidth [6,7]. In the case of mobile wireless devices, the available energy in the transmitting device is also a limited resource to be considered in the overall resource allocation scheme [8,9].

The second scenario applies to the downlink video streaming case when the video is pre-encoded and stored at a media server. In this case, the actual video coding parameters are not controllable at the scheduler unless it is capable of decoding, re-encoding, or transcoding the video in real-time. However, in a packet-based system, the scheduler can adapt to channel conditions and control the source rate of the video by opting to delay or eventually throw out video packets instead of transmitting them [10,11]. Another opportunity for adapting the source rate is provided by scalable video coding techniques, which allow the embedding of spatial, temporal, and quality scalability layers within a single bit stream. For example, quality scalability allows for the transmission of a base layer at low quality and bit rate, and then, enhancement layers that progressively improve the received video quality at the cost of additional bits [12,13]. In such schemes, the complete reception of the base layer is vital to avoid error propagation at the decoder. In the case of fine granularity scalability (FGS), in MPEG-4 [14–16] and the emerging scalable extension of H.264 [17], the enhancement layers may be truncated at any point to achieve the corresponding best possible video quality. Such layered encoding schemes also allow for unequal error protection where different levels of protection can be provided to the base layer and enhancement layer video packets [18].

In the following sections, we will elaborate on the above scenarios and point out the recent advances made in each direction. We will begin with a description of a typical wireless video streaming system and the associated control parameters in Section 2. Then, we will describe some important distortion estimation schemes that determine the relative importance of video content, in Section 3. Section 4 will show how the different components of a video streaming system can be utilized together for cross-layer optimization schemes. We will conclude in Section 5 with a discussion of the opportunities for future research.

2. Wireless Video Streaming Systems

Figure 1 provides a high-level overview of a wireless video streaming system. As mentioned before, such a system is composed of a server, which will either encode the video in real-time or contain pre-encoded video, a scheduler, which can be more appropriately termed a network controller, and a client or multiple clients, which will decode the received video streams. The video will be transmitted over time-varying wireless channels with limited data rates.

2.1. Video Encoding

The raw video is first compressed using a video coding scheme. The scope of this paper will encompass the widely accepted compression standards such as H.263, MPEG-4, and H.264 which use a block-based motion compensation technique for video coding. In

DOI: 10.1002/wcm

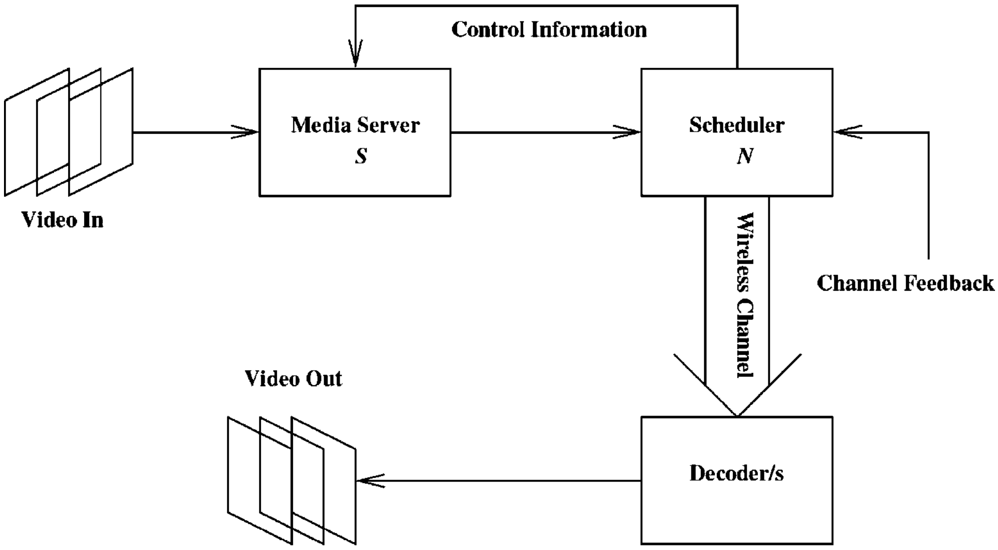

Fig. 1. Wireless video streaming system.

these approaches, the video frame is first divided into MBs, blocks of pixels that are either independently coded (Intra) or coded using temporal prediction from previous frames (Inter). Inter coded MBs use motion vectors (MVs) that specify the location of reference blocks in a previously transmitted frame. The Intra blocks as well as the residual blocks after motion prediction from Inter blocks are then transform coded, followed by quantization of the transform coefficients and entropy coding to complete the encoding process.

To improve coding efficiency, MBs within a frame may also be coded in a spatially predictive manner as, for example, in the predictive intra coding technique used in H.264. Error robustness can be introduced by coding a group of MBs together in a slice, which is independently decodable from all other macroblocks/slices in the same video frame. Therefore, a video packet could consist of a slice of a frame, which would be independently decodable at the receiver and whose loss would not affect the decoding of any other slice in the frame. A recent review of error-resilient video coding techniques is provided in Reference [19]. In addition to the resynchronization ability provided by slices, another important consideration in wireless environments is the packet loss probability. Longer length packets are more susceptible to bit errors and can cause larger localized distortions when lost. Dividing a video frame into multiple shorter length slices, however, can lead to significant overhead due to the included slice headers as well as additional IP/UDP/RTP header data which must be added for data transport. A discussion on the importance of considering packet length in a wireless environment can be found in Reference [20].

The source coding parameter set, S, in Figure 1 could denote the MB modes (such as Intra, Inter, or Skip) or the quantization step sizes for the MBs. In the case of pre-encoded video, multiple video streams of different source bit rates corresponding to different compressed quality levels may be available at the server. Then, the source parameter set could determine the quality level of the downlink bit stream to use. H.264 allows for even greater flexibility by providing the ability to switch between different pre-encoded source bit streams at pre-specified points (*SP* and *SI* frames) in each stream. In scalable coded bit streams, the source parameter can determine the spatial, temporal, and quality scalability levels of the transmitted video. The selection of S will determine the source rate of the video and will also affect the video quality after decoding at the client.

2.2. The Scheduler/Transmitter

The encoded video packets are then sent to the scheduler where the network resources, N, will be allocated for each video packet. The simplest form of scheduling would be a round-robin scheme, such as the time division multiplexing (TDM) scheme used in 2G GSM networks where transmission time is shared equally among all users. Recent research, however, has shown that channel-aware scheduling schemes, which allocate greater resources to users with better channel conditions, can perform significantly better than round-robin schemes by achieving a *multi-user diversity* gain [21]. To avoid favoring users that are closer to the base station, proportionally fair scheduling schemes have been developed that take into account the

DOI: 10.1002/wcm

average throughput of each user as well as the channel conditions [22]. Other schemes such as the *Exponential Rule* [23] attempt to stabilize the queue lengths for each user, which is beneficial for delay sensitive applications such as streaming video transmission. In Reference [24], a gradient-based scheduling scheme is used, which allocates resources such that the user rates weighted by the gradients of a given system utility function will be maximized. This generalized scheme allows for the use of a broad range of utility functions, both queue length based and throughput based.

Improved understanding of the benefits of channel-aware scheduling schemes has led to the development of 3G architectures such as high speed downlink packet access (*HSDPA*) [25], and proposed, possibly 4G wireless architectures such as IEEE 802.16 (commonly known as *WiMAX*) [26] that can support high data rates to multiple mobile users. In these schemes, the scheduler makes use of frequently updated *channel state information* (CSI) to determine the resource allocations. For example, in a CDMA/HDR system [27,22], TDM is used to transmit to a single user at each time-slot. In a wireless network using the HSDPA approach, the scheduler uses a combination of TDM and CDMA, adaptive modulation and coding techniques, and hybrid ARQ to control the network resources provided to each user. Approaches such as HSDPA, CDMA/HDR as well as the 1× EVDV architecture for CDMA 2000 use short feedback intervals and transmission time slots (2 ms for HSDPA, 1.25 ms for 1× EVDV, 1.67 ms for CDMA/HDR) to enable fast adaptation to varying channel conditions.

Most packet scheduling schemes currently used in wireless networks do not achieve the best possible quality for video transmission, since they do not take the video content into consideration when making scheduling decisions. Although they may be more complex, content-aware schemes could perform significantly better in terms of the end-to-end distortion of the video received by the client and more efficiently utilize the available resources at the transmitter. Any such scheduling scheme would need to consider the video encoding method, the channel conditions as well as the decoding and error-resilience methods employed at the decoder to make the best possible scheduling decisions.

2.3. The Wireless Channel

Schedulers depend on a model of the wireless channel to determine the relative importance of the allocated network resources. Wireless channels typically suffer from multi-path and shadowing effects which lead to high bit error rates and low bandwidth. An important parameter to consider in wireless channels is the signal-to-noise ratio (SNR) at the receiver. A low received SNR leads to more bit errors, and thereby, higher packet loss rates. The SNR is a function of the transmission power and fading in the channel. In CDMA systems, interference from other users will also add to the noise in the channel. In general, the probability of packet loss in a wireless channel can be modeled as

$$\rho_k = f(P_k, R_k, \theta_k) \tag{1}$$

where ρ_k denotes the probability of loss for packet k, P_k the transmission power, R_k the transmission rate, and θ_k the channel fading realization.

In Reference [8], the concept of *outage capacity* discussed in Reference [28] is used to determine $f(\cdot)$ analytically. Assuming a Rayleigh fading channel, outage capacity can be used to determine the packet loss probability as

$$\rho_k = 1 - \exp\left(\frac{1}{P_k \gamma(\theta_k)}\left(2^{R_k/W_k} - 1\right)\right) \tag{2}$$

where $\gamma(\theta_k)$ denotes the normalized expected SNR given θ_k, and W_k denotes the available bandwidth. A Rayleigh fading channel may also be estimated using a Finite State Markov Channel (FSMC) model [29].

In Reference [30], $f(\cdot)$ is derived for a model based on a binary phase-shift keying modulation scheme in a Rayleigh fading channel. The transmission rate is adapted by changing the amount of forward error correction (FEC) applied to each packet using a rate-compatible convolution code. When fast feedback is available as in 3G and 4G wireless networks, the transmission rate can also be indirectly adapted to the channel conditions through ARQ [31].

Another analytical approach would be to use Shannon's capacity theorem directly to determine the maximum rate that can be reliably achieved given the channel conditions and transmission power. This approach is used in Reference [9] for transmission power control and in Reference [24] in a downlink packet-scheduling scheme.

2.4. Video Decoding/Error Concealment

The final stage in Figure 1 is at the client where the video will be decoded. The decoder must take into account the packet errors in the channel, delayed

DOI: 10.1002/wcm

packets as well as packets that may have been dropped due to buffer overflow, and perform some form of error concealment prior to playing back the video. In a video streaming system, any packet that arrives after a given playback deadline will be considered lost by the decoder.

In general, the error-concealment strategies employed at the decoder attempt to exploit the redundancies that are common to most video sequences. Errors in inter coded frames are commonly concealed using temporal concealment techniques that exploit the temporal correlations in the scene [32]. Assuming a MB is lost, the simplest temporal error-concealment method would be to copy the pixels from the MB in the same location in the previous frame. More complicated methods use the MVs from neighboring received MBs in the frame to determine the position of a motion compensated MB in the reference frame. Since more than one candidate MV may be available, a boundary matching technique can be used to determine the best possible candidate [33] or to re-estimate the MVs [32]. Errors in intra frames can be concealed using spatial concealment techniques [34]. Simple but effective spatial concealment methods rely on weighted pixel-averaging schemes [35], where the weight depends on the distance from the concealed pixels. Hybrid spatio-temporal error-concealment schemes also exist [36]. A broad review of error-concealment schemes and their relative benefits can be found in Reference [37].

Figure 2 shows the difference in decoded video quality using the simple MB copy scheme described above and the complex concealment strategy, which uses the boundary matching technique, for a given packet loss realization. Figure 3 shows an example frame, where the packet, which contained the information for the 5th row of MBs, has been lost. Clearly, the complex concealment scheme provides a more visually pleasing image by taking into account the motion within the lost packet. Figures 2 and 3 show that the error-concealment technique used at the decoder has a significant effect on the received video quality. Therefore, knowledge of the actual error-concealment scheme that will be used by the decoder can be very important for content-aware scheduling schemes as it helps guide them towards assigning priority to video packets that, if lost or dropped, would be more difficult to conceal.

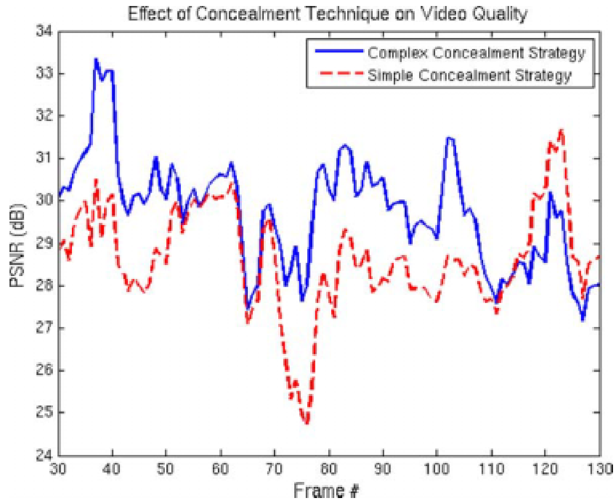

Fig. 2. Variations in video quality for different error concealment strategies.

3. Distortion Estimation

In content-aware schemes, the prioritization of packets must be based on the distortion of the decoded video stream relative to the original. Objective assessment of video distortion, or inversely, video quality is an ongoing area of research in the video signal processing community. Excellent discussions on video quality metrics can be found in References [38–40]. The scope of this review will be restricted to the widely used distortion metrics; those of mean squared error (*MSE*)

(a) (b) (c)

Fig. 3. (a) Decoded frame with no packet losses (b) decoded frame after simple concealment with one lost packet (c) decoded frame after complex error concealment with one lost packet (lost packet composed of row of MBs indicated by the markers).

DOI: 10.1002/wcm

and peak signal-to-noise ratio (*PSNR*), although the algorithms discussed will generally work with any mathematically tractable distortion metric.

In packet-based video streaming where video packets may be lost due to random channel errors, an important measure of video quality is the expected distortion. Given a packet loss probability of ρ_k, the expected distortion for packet k can be expressed as

$$E\left[D_k\right] = (1-\rho_k)\, E\left[D_{\mathrm{R},k}\right] + \rho_k E\left[D_{\mathrm{L},k}\right] \quad (3)$$

where $E[D_{\mathrm{L},k}]$ denotes the expected distortion given the data packet k is lost and $E[D_{\mathrm{R},k}]$ denotes the expected distortion given the packet is received. In an inter coded video frame, distortion can occur even if a data packet is correctly received, since the packet may be dependent through motion compensation on a previously lost data packet. The distortion due to a lost packet will depend on the particular error-concealment algorithm used by the decoder. Based on Equation (3), the end-to-end distortion for a sequence of video packets can be expressed as a function, $g(\cdot)$, of all the individual packet expected distortions as

$$D_{\mathrm{tot}} = g\left(E\left[D_{k=\{1,2,\ldots,K\}}\right]\right) \quad (4)$$

where K is the total number of video packets. Note that, D_{tot} can be a function of both the source coding parameters (S) which affect the source distortion as well as the network parameters (N) which affect the packet loss probability, ρ_k.

Due to their importance in content-aware resource allocation schemes, much effort has been dedicated to the accurate estimation of Equations (3) and (4). In Reference [41], multiple independent packet loss realizations are simulated and the expectation is calculated over the resulting distortions. The accuracy of this method increases with the number of simulations at the cost of increased computation time and storage requirements. A recursive algorithm that calculates the end-to-end distortion in terms of the expected mean absolute difference (*MAD*) of pixels is shown in Reference [7]. In Reference [42], a powerful algorithm called recursive optimal per-pixel estimate (*ROPE*) is introduced that efficiently calculates the distortion in terms of MSE, given a probability of packet loss. The algorithm only relies on a recursive calculation of the first and second moments of the expected distortion and therefore significantly reduces the storage requirements of the system as well as computational complexity. While the initial ROPE algorithm had stringent requirements such as needing full-pixel MVs, which eliminate cross-correlation terms in the calculation of expected distortion, some advancements have since been made in reducing such restrictions by accurately estimating the cross-correlation terms as well [43,44]. Other useful methods for distortion estimation can be found in References [45–47]. Also, a particular distortion estimation technique well suited for streaming pre-encoded video streams is presented in Reference [10] and will be discussed in subsection 4.2.

While the expected distortion is a useful metric in developing a content-aware resource allocation scheme, it is also important to consider the variation of the distortion from its mean. Such variations occurring both spatially within a frame and temporally in a sequence of frames can severely degrade the perceptual quality of the received video. In Reference [48], the variance of the distortion is explicitly considered in the distortion metric as

$$D_{\mathrm{tot}} = \left(\frac{1}{K}\right)\sum_{k=1}^{K}\left\{(1-\alpha)\, E\left[D_k\right] + \alpha \mathrm{Var}\left[D_k\right]\right\} \quad (5)$$

where $\mathrm{Var}[D_k]$ is the variance of the distortion of the kth packet and $\alpha \in [0, 1]$ is a weighting factor that controls the relative importance of the first and second moments.

4. Resource Distortion Optimization

4.1. Resource Distortion Optimization for Real-Time Video Streaming

A general framework for solving the resource distortion optimization problem on systems in which video streams are encoded and transmitted in real-time is provided in Reference [1]. According to this framework, the source and network parameters, S and N, are explicitly controlled in order to achieve the desired objective. For example, minimizing the expected distortion subjected to a constraint on the available resources can be expressed as

$$\begin{aligned} &\min_{\{S,N\}} D_{\mathrm{tot}}(S,N),\\ &\text{s.t.}: C_{\mathrm{tot}}(S,N) \le C_0, \text{ and}\\ &\qquad T_{\mathrm{tot}}(S,N) \le T_0 \end{aligned} \quad (6)$$

where C_0 is the maximum allowable resource consumption and T_0 is the end-to-end delay constraint. A

DOI: 10.1002/wcm

dual formulation to Equation (6), that is

$$\min_{\{S,N\}} C_{\text{tot}}(S, N),$$
$$\text{s.t.,}\ D_{\text{tot}}(S, N) \leq D_0,\ \text{and}\ T_{\text{tot}}(S, N) \leq T_0 \quad (7)$$

is also considered in Reference [1], as a scheme for minimizing the transmission cost subject to video quality and delay constraints. The formulation in Equation (7) is especially important for real-time streaming schemes where the media server may also be a mobile wireless device with limited available energy [8,9].

Solutions to Equations (6) and (7) can be found by converting them to unconstrained minimization problems using Lagrangian relaxation [49,50]. For example, the unconstrained problem for Equation (6) can be expressed as

$$\min_{\{S,N\}} J_{\text{tot}} = D_{\text{tot}}(S, N) + \lambda_1 \cdot C_{\text{tot}}(S, N) + \lambda_2 \cdot T_{\text{tot}}(S, N) \quad (8)$$

where λ_1 and λ_2 are the Lagrange multipliers. Assuming that a video packet only depends on a small number of previously transmitted packets, dynamic programming can be used to determine the optimal values of S and N for a given λ_1 and λ_2. In the case that packets depend only on the previously transmitted packet for error concealment, the complexity of calculating the optimal values of S and N for a given set of λ's becomes, $O(M \times (|N| \times |S|)^2)$, where $|\cdot|$ denotes the cardinality of the set and M denotes the number of packets considered. Various optimization techniques, such as cutting plane and subgradient methods, can be used to efficiently search for the appropriate values of λ_1 and λ_2 [51]. Simulations have shown that the appropriate values of λ can be found within a few iterations.

Figure 4 depicts the results from an example implementation of the problem described in Equation (6), where the Foreman QCIF (176×144) sequence is transmitted at 30 fps, assuming a transmission rate of 360 kbps. The sequence is encoded using an H.263 codec. A method similar to that described in detail in Reference [52] is used to model the channel as an 11-state FSMC with a set of Reed Solomon codes for FEC. An MSE-based calculation as in Equation (3) is used to calculate the expected distortion. System 1 performs joint source channel coding with power allocation where the source parameters, S in Equation (6), are the MB modes such as (inter, intra, or skip) and quantization parameters for each MB, and the network parameters, N, include the coding rates as well as a discrete set of transmission power levels. The available coding rates for the Reed Solomon codes are restricted to the set, (9,9), (11,9), (13,9), (15,9), (17,9), (19,9), and (21,9), where (N, M) denotes that M information symbols are coded to an N symbol codeword. The total energy available for the transmission is assumed to be limited and forms the cost constraint in Equation (6). Systems 2–5 assume a fixed power allocation, which removes the cost constraint from Equation (6) but to achieve a fair comparison, the power allocation is set such that the total consumed energy remains the same as in System 1. System 2 jointly optimizes over the source and channel coding parameters. System 3 does not use any FEC (i.e., only optimizes over the source encoding parameters). System 4 assumes that the optimal channel coding rates are as found in System 2 and then performs source encoding without taking into account the error resilience provided by the channel coding. This is akin to an independent source and channel coding scheme. System 5 keeps the channel coding rate fixed at (13,9) and optimizes over the source coding parameters. It can be seen that the joint source and channel coding scheme with power allocation performs better over the entire range of received SNR and that as the SNR decreases (i.e., probability of packet loss increases), the gain due to jointly considering the source and channel also increases.

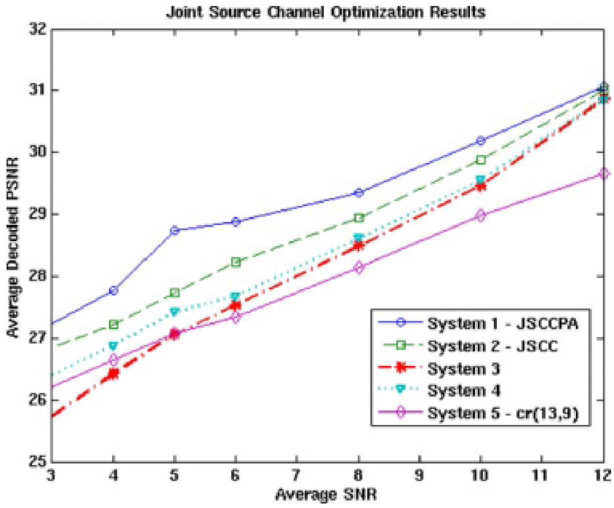

Fig. 4. Experimental results using joint source channel coding versus independent source or channel coding.

4.2. Resource Distortion Optimization for Pre-Encoded Video Streams

An excellent starting point for research on resource distortion optimized streaming of pre-encoded video

DOI: 10.1002/wcm

streams is presented in Reference [10]. An important feature of pre-encoded video streams is that the basic characteristics of the individual video packets are known, or can be calculated in advance for the entire sequence. In Reference [10], the basic characteristics for a data packet k are defined as, the number of bits, B_k, which is known, the *incremental additive distortion* gain due to receiving a packet, Δd_k, which can be calculated as described below, and a timestamp for the maximum delay allowed prior to playback, which depends on the decoder buffer at the receiver.

The additive distortion measure, Δd_k, is a key component of the resource distortion schemes proposed in Reference [10] and subsequent work in the area. In a pre-encoded video sequence, the dependencies between video packets, represented by directed acyclic graphs in Reference [10], are known in advance. For example, in a typical *group of pictures* (GOP) structure the dependencies between I, P, and B frames are known. Therefore, the incremental reduction in distortion due to receiving a particular data packet depends on whether its parent packets have been received, as well as the error-concealment technique used at the decoder. While a simple error-concealment technique is used in Reference [10], approximate solutions that can be used with more complex concealment techniques are proposed in References [53,54]. Another non-additive method for distortion estimation based on dependencies between video packets is discussed in Reference [55].

A rate distortion optimized scheme based on Reference [10] for transmitting a pre-encoded video stream to a single client over a wireless network is presented in Reference [56]. The goal in this scheme is to determine the appropriate transmission policy for the available data packets at a given transmission opportunity. The transmission policy vector, denoted by $\boldsymbol{\pi}$, includes $\{\pi_1, \pi_2, \ldots, \pi_K\}$, which are the individual transmission policies for each data packet in the sequence. Each individual transmission policy represents a decision on whether to transmit/retransmit or not to transmit the corresponding packet. The expected cost

$$C(\boldsymbol{\pi}) = \sum_{k=1}^{K} B_k c\,(\pi_k) \tag{9}$$

can be calculated as a function of the transmission policy vector. In Equation (9), $c(\pi_k)$ denotes the expected cost per source bit for packet k. The incremental additive distortion estimation technique is used to calculate the expected distortion as a function of the policy vector as

$$D\,(\boldsymbol{\pi}) = D_0 - \sum_{k=1}^{K} \Delta d_k \prod_{k' \to k} (1 - \varepsilon\,(\pi_{k'})) \tag{10}$$

where D_0 would be the distortion if no packet is received, $k' \to k$ indicates that k depends on k', and $\varepsilon(\pi_{k'})$ denotes the probability that packet k' is not correctly received within its decoding deadline given the transmission policy.

Based on Equations (9) and (10), the best transmission policy vector can be defined as the one that minimizes $D(\boldsymbol{\pi})$ subject to a constraint on the available resources, $C(\boldsymbol{\pi})$. This can be found using a Lagrangian relaxation technique and then using an iterative descent algorithm to search over the transmission policy vectors [10,56].

The concept of incremental additive distortion is also used in References [57,11] to develop content-aware scheduling schemes for wireless networks with multiple clients. Although data packets of individual users are prioritized in Reference [57], the video content is not considered when scheduling resources over multiple users. In Reference [11], a TDM scheme is used where all the system resources will be given to a single user during a transmission time-slot. The scheme determines the priority across users as a combination of a content-aware importance measure and the delay of the *head-of-line* (HOL) packet. Another heuristic scheme for scheduling between multiple users is presented in Reference [58].

In Reference [59], a content-aware scheduling scheme inspired by the gradient-based resource allocation method [24], tackles the downlink multi-user video streaming problem. The key idea in Reference [59] is to maximize the sum of the instantaneous data rates assigned to each user, which will be a function of the network parameters for each user, denoted by N_i, weighted by the gradients of a distortion based utility function. This can be written as

$$\max_{N} \sum_{i} w_i u_i r_i(N_i) \tag{11}$$

where N contains the network parameter set for all users, u_i denotes the gradient of the utility function for user i, w_i denotes an additional weighting parameter to ensure fairness across users, and $r_i(N_i)$ denotes the achievable information rate for user i given the network parameters N_i. In a content-independent scheme, u_i can be a function of the current average

DOI: 10.1002/wcm

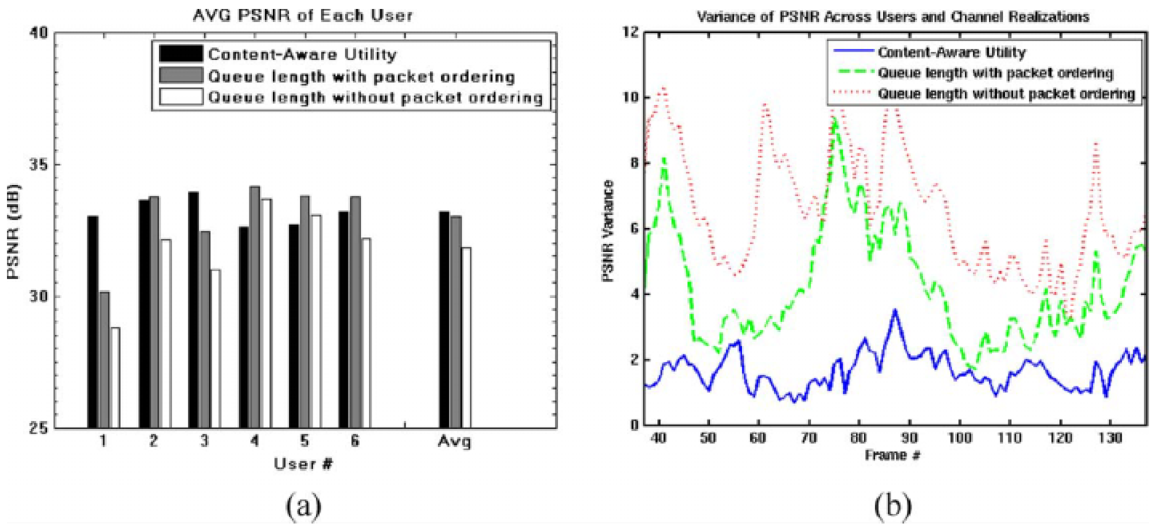

Fig. 5. (a) Average PSNR over five-channel realizations and 100-frame sequence for each user (b) Variance of PSNR across all users.

throughput for user i or the queue-length at user i's transmission buffer but in the content-aware scheme proposed in Reference [59], u_i is defined as the reduction in distortion, after error concealment, in the video sequence of user i due to the transmission of an additional video packet. The video packets for each user are first ordered by decreasing utility gradient to ensure a concave utility function.

Figure 5 shows some results using the gradient-based scheduling scheme described above in a simulated HSDPA network. The user numbers denote a different video sequence (1: Foreman, 2: Mother and Daughter, 3: Carphone, 4: News, 5: Silent, and 6: Hall Monitor). Each video sequence is of QCIF resolution and is encoded using H.264 such that the average decoded PSNR of each compressed sequence is 35 dB. Only the first frame of each sequence is coded as Intra (I frames) and the rest are coded as P frames. Randomly placed intra MBs are added to each frame to reduce error propagation due to dropped packets. Figure 5(a) shows the average decaded PSNR for each user obtained over multiple channel realizations, and averaged over all the frames in the sequence. The three compared schemes correspond to one that uses the gradient of a content-aware utility function in Equation (11), one that orders the video packets by decreasing distortion utility gradient but then uses a content-independent utility function (function of each user's queue length measured in bits) in Equation (11) to determine the network parameter assignments, and finally, one that does not order the video packets and uses a content-independent utility function. Figure 5(a) shows that the two content-aware packet ordering and resource allocation schemes show a significant gain over the purely channel and queue-length dependent scheme. Although from plot (a), packet ordering within each user seems to be sufficient, Figure 5(b), which shows the variance of PSNR across all users, indicates that the content-aware utility gradients tend to provide more consistent quality across users than the queue-length based scheme. The reason for the difference in performance is that sequences such as Foreman and Carphone (users 1 and 3) tend to need more bits in order to achieve the same quality as some lower activity sequences such as Hall Monitor (user 6), an issue which is not addressed by the content-independent schemes.

5. Future Research Directions

Currently very little research has been conducted on adapting channel-dependent multi-user scheduling techniques developed in the wireless communications community to video streaming systems. Such research would involve the development of appropriate buffer management strategies at the transmitter to ensure that all users receive a reasonable *quality of service* (QoS). For example, in the case of pre-encoded video streams, video packets of some users may need to be dropped at the transmitter buffer based on a content-aware packet prioritization scheme. However, the buffer management strategy must also be adaptive to the rapidly changing channel conditions in mobile wireless environments. A significant stumbling block in developing practical content-aware scheduling algorithms is that of reducing the required computational complexity,

DOI: 10.1002/wcm

such that the scheduler can respond dynamically to the rapidly changing channel conditions. Scalable video coding techniques, and especially fine granularity scalability, could be easily adapted to channel-dependent content-aware scheduling schemes, since they provide a natural packet prioritization strategy for the scheduler. Therefore, developing resource distortion optimized scheduling schemes for scalable video coding techniques is an important area of future research.

Another important direction for research is that of obtaining objective measures for video quality assessment that are perceptually relevant. The particular types of errors that can occur due to video packet losses are specific to the block-based motion compensation technique used at the encoder as well as the spatial and temporal error-concealment methods used at the decoder. Video quality measures other than the MSE or PSNR may be better adapted to measuring the perceptual distortions caused by such errors. For example, with scalable video coding techniques, perceptually degrading fluctuations in video quality may be caused due to channel adaptation. Such fluctuations can be avoided by considering other temporal measures of video quality.

References

1. Katsaggelos AK, Eisenberg Y, Zhai F, Berry R, Pappas TN. Advances in efficient resource allocation for packet-based streaming video transmission. In *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery* 2005; **93**(1): 135–147.
2. Girod B, Chakareski J, Kalman M, Lang YJ, Setton E, Zhang R. Advances in network- adaptive video streaming, *presented at the Tyrrhenian International Workshop on Digital Communications*, 2002.
3. Wu D, Hou YT, Zhang Y-Q. Transporting real-time video over the Internet: challenges and approaches. In *Proceedings of the IEEE* 2000; **88**(12): 1855–1877.
4. Zheng H. Optimizing wireless multimedia transmission through cross layer design. In *Proceedings of IEEE International Conference on Multimedia and Expo* 2003; **1**: 185–188.
5. Kawadia V, Kumar PR. A cautionary perspective on cross-layer design. *IEEE Wireless Communications* 2005; **12**(1): 3–11.
6. Zhang R, Regunathan SL, Rose K. Video coding with optimal inter/intra-mode switching for packet loss resilience. *IEEE Journal on Selected Areas in Communications* 2000; **18**(6): 966–976.
7. Wu D, Hou YT, Li B, Zhu W, Zhang Y-Q, Chao HJ. An end-to-end approach for optimal mode selection in Internet video communication. *IEEE Journal on Selected Areas in Communications* 2000; **18**(6): 977–995.
8. Eisenberg Y, Luna CE, Pappas TN, Berry R, Katsaggelos AK. Joint source coding and transmission power management for energy efficient wireless video communications. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; **12**(6): 411–424.
9. Luna CE, Eisenberg Y, Pappas TN, Berry R, Katsaggelos AK. Joint source coding and data rate adaptation for energy efficient wireless video streaming. *IEEE Journal on Selected Areas in Communications* 2003; **21**(10): 1710–1720.
10. Chou PA, Miao Z. Rate-distortion optimized streaming of packetized media. *IEEE Transactions on Multimedia* 2006: **8**(2): 390–404.
11. Liebl G, Kalman M, Girod B. Deadline aware scheduling for wireless video streaming. In *Proceedings of IEEE International Conference on Multimedia and Expo* 2005; 470–473.
12. Wu D, Hou YT, Zhang Y-Q. Scalable video coding and transport over broad-band wireless networks. In *Proceedings of the IEEE* 2001; **89**(1): 6–20.
13. Kondi LP, Ishtiaq F, Katsaggelos AK. Joint source-channel coding for motion-compensated DCT-based SNR scalable video. *IEEE Transactions on Image Processing* 2002; **2**(9): 1043–1052.
14. Radha HM, Van der Schaar M, Chen Y. MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP. *IEEE Transactions on Multimedia* 2001; **3**(1): 53–68.
15. Li W. Overview of fine granularity scalability in MPEG-4 video standard. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3): 301–317.
16. *Coding of Audio-Visual Objects, Part-2 Visual, Amendment 4: Streaming Video Profile*, ISO/IEC 14496-2/FPDAM4, July 2000.
17. Reichel J, Schwarz H, Wien M. Joint Scalable Video Model (JSVM)3.0 Reference Encoding Algorithm Description, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, July 2005.
18. Van der Schaar M, Radha H. Unequal packet loss resilience for fine-granular-scalability video. *IEEE Transactions on Multimedia* 2001; **3**(4): 381–394.
19. Wang Y, Wen G, Wenger S, Katsaggelos AK. Review of error resilient techniques for video communications. *IEEE Signal Processing Magazine* 2000; **17**(4): 61–82.
20. Stockhammer T, Hannuksela MM, Wiegand T. H.264/AVC in wireless environments. *IEEE Transactions on Circuits and Systems for Video Technology* 2003; **13**(7): 657–673.
21. Knopp R, Humblet PA. Information capacity and power control in single-cell multiuser communications. In *Proceedings of IEEE International Conference on Communications*, 1995; 331–335.
22. Jalali A, Padovani R, Pankaj R. Data throughput of CDMA-HDR a high efficiency high data rate personal communication wireless system. In *Proceedings of IEEE Vehicular Technology Conference* 2000; 1854–1858.
23. Shakkotai S, Srikant R, Stolyar A. Pathwise optimality and state space collapse for the exponential rule. In *Proceedings of the IEEE International Symposium on Information Theory* 2002; 379.
24. Agrawal R, Subramanian V, Berry R. Joint scheduling and resource allocation in CDMA systems, *2nd Workshop on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2004.
25. *High Speed Downlink Packet Access; Overall description, 3GPP TS 25.308 v6.3.0*, December 2004.
26. *Part 16: Air interface for fixed broadband wireless access systems*, IEEE Std 802.16-2004, October 2004.
27. Bender P, Black P, Grob M, Padovani R, Sindhushayana N, Viterbi A. CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Communications Magazine* 2000; **38**(7): 70–77.
28. Ozarow L, Shamai S, Wyner AD. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology* 1994; **43**(2): 359–378.
29. Arauz J, Krishnamurthy P, Labrador MA. Discrete Rayleigh fading channel modeling. *Wireless Communications and Mobile Computing* 2004; **4**(4): 413–425.
30. Zhai F, Eisenberg Y, Pappas TN, Berry R, Katsaggelos AK. Joint source-channel coding and power allocation for energy efficient wireless video communications. In *Proceedings of the 41st Allerton Conference on Communication, Control and Computing* 2003.

DOI: 10.1002/wcm

31. Nanda S, Balachandran K, Kumar S. Adaptation techniques in wireless packet data services. *IEEE Communications Magazine* 2000; **38**(1): 54–64.
32. Hong MC, Kondi L, Scwab H, Katsaggelos AK. Video error concealment techniques. *Signal Processing: Image Communications* 1999; **14**(6–8): 437–492.
33. Wang Y-K, Hannuksela MM, Varsa V, Hourunranta A, Gabbouj M. The error concealment feature in the H.26L test model. In *Proceedings of the International Conference on Image Processing* 2002; **2**: 729–732.
34. Sun H, Kwok W. Concealment of damaged block transform coded images using projection onto convex sets. *IEEE Transactions on Image Processing* 1995; **4**(4): 470–477.
35. Salama P, Shroff NB, Delp EJ. Error concealment in encoded video streams. In *Signal Recovery Techniques for Image and Video Compression and Transmission*, Katsaggelos AK, Galatsanos NP (eds). Kluwer: Dordrecht, Netherlands, 1998.
36. Belfiore S, Grangetto M, Magli E, Olmo G. Spatio-temporal video error concealment with perceptually optimized mode selection. In *Proceedings of the IEEE International Conference on Multimedia and Expo* 2003; 169–172.
37. Wang Y, Zhu Q-F. Error control and concealment for video communication: A review, *Proceedings of the IEEE* 1998; **86**(5): 974–997.
38. Wang Z, Sheikh HR, Bovik A. Objective video quality assessment. In *The Handbook of Video Databases: Design and Applications*, Furht B, Marqure O (eds). CRC Press: Boca Raton, Florida, 2003: 1041–1078.
39. Yuen M, Wu HR. A survey of hybrid MC/DPCM/DCT video coding distortions. *Signal Processing* 1998; **70**: 139–147.
40. Winker S. Issues in vision modeling for perceptual video quality assessment. *Signal Processing* 1999; **78**: 231–252.
41. Stockhammer T, Wiegand T, Wegner S. Optimized transmission of H.26L/JVT coded video over packet-switched networks. In *Proceedings of the IEEE International Conference on Image Processing*, 2002: 173–176.
42. Zhang R, Regunathan SL, Rose K. Video coding with optimal inter/intra-mode switching for packet loss resilience. *IEEE Journal on Selected Areas in Communications* 2000; **18**(6): 966–976.
43. Leontaris A, Cosman PC. Video compression with intra/inter mode switching and a dual frame buffer. In *IEEE Data Compression Conference* 2003.
44. Yang H, Rose K. Recursive end-to-end distortion estimation with model-based cross-correlation approximation. In *Proceedings of the IEEE International Conference on Image Processing* 2003; **3**: 469–472.
45. Cote G, Shirani S, Kossentini F. Optimal mode selection and synchronization for robust video communications over error-prone networks. *IEEE Journal on Selected Areas in Communications* 2000; **18**(6): 952–965.
46. Wiegand T, Farber N, Stuhlmuller K, Girod B. Error-resilient video transmission using long-term memory motion-compensated prediction. *IEEE Journal on Selected Areas in Communications* 2000; **18**(6): 1050–1062.
47. He Z, Cai J, Chen CW. 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* 2002; **12**(6): 511–523.
48. Eisenberg Y, Zhai F, Luna CE, Pappas TN, Berry R, Katsaggelos AK. VAPOR: Variance-aware per-pixel optimal resource allocation. *IEEE Transactions on Image Processing* 2006; **15**(2): 289–299.
49. Schuster GM, Katsaggelos AK. *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*. Kluwer: Norwell, MA, 1997.
50. Ortega A, Ramchandran K. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine* 1998; **15**(6): 23–50.
51. Bertsekas DP. *Nonlinear Programming* (2nd edn). Athena Scientific: Belmount MA, 1999.
52. Maani E, Katsaggelos AK. Joint source-channel coding and power allocation for video transmission over wireless fading channels. In *Proceedings of the IEEE Consumer Communications and Networking Conference* 2006: 949–953.
53. Zhang R, Regunathan SL, Rose K. Optimized video streaming over lossy networks with real-time estimation of end-to-end distortion. In *Proceedings of the IEEE International Conference on Multimedia and Expo* 2002: 861–864.
54. Cheung G, Tan W-T. Directed acyclic graph based source modeling for data unit selection of streaming media over QoS networks. In *Proceedings of the International Conference on Multimedia and Expo* 2002: 81–84.
55. Chakareski J, Apostolopoulos J, Tan W-T, Wee S, Girod B. Distortion chains for predicting the video distortion for general packet loss patterns. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing* 2004; **5**: 17–21.
56. Chakareski J, Chou PA. Application layer error-correction coding for rate-distortion optimized streaming to wireless clients. *IEEE Transactions on Communications* 2004; **52**(10): 1675–1687.
57. Liebl G, Stockhammer T, Buchner C, Klein A. Radio link buffer management and scheduling for video streaming over wireless shared channels. In *Proceedings of the Packet Video Workshop* 2004.
58. Tupelly RS, Zhang J, Chong EKP. Opportunistic scheduling for streaming video in wireless network. In *Conference on Information Sciences and Systems* 2003.
59. Pahalawatta PV, Berry R, Pappas TN, Katsaggelos AK. A content-aware scheduling scheme for video streaming to multiple users over wireless networks. *European Signal and Image Processing Conference* 2006.

Authors' Biographies

Aggelos K. Katsaggelos received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees both in electrical engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985 he joined the Department of Electrical Engineering and Computer Science at Northwestern University, where he is currently working as a professor. He is also the Director of the Motorola Center for Seamless Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital. Dr. Katsaggelos is a member of the Publication Board of the *IEEE Proceedings*, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as editor-in-chief of the *IEEE Signal Processing Magazine* (1997–2002), a member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, an Associate editor for the *IEEE Transactions on Signal Processing* (1990–1992), an area editor for the journal *Graphical Models and Image Processing* (1992–1995), a member of the Steering Committees of the *IEEE Transactions on Image Processing*

DOI: 10.1002/wcm

(1992–1997) and the *IEEE Transactions on Medical Imaging* (1990–1999), a member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the editor of *Digital Image Restoration* (Springer-Verlag 1991), co-author of *Rate-Distortion Based Video Compression* (Kluwer, 1997), co-editor of *Recovery Techniques for Image and Video Compression and Transmission*, (Kluwer, 1998), and co-author of Super-Resolution for Images and Video, (Claypool, 2006). He was the holder of the Ameritech Chair of Information Technology (1997–2003), and he is the co-inventor of 12 international patents, a Fellow of the IEEE, and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), and an IEEE ICME Best Paper Award (2006).

Peshala Pahalawatta received the B.S. degree in electrical engineering from Lafayette College, Easton, Pennsylvania, in 2000, and the M.S. degree in electrical engineering from Northwestern University, Evanston, Illinois, in 2002. He is currently pursuing a Ph.D. degree in electrical enginnering at Northwestern University. His primary research interests include image and video compression and transmission, wireless communications, and computer vision.