

# Cross-Layer Optimization for Downlink Wavelet Video Transmission

Hyungkeuk Lee, *Student Member, IEEE*, Sanghoon Lee, *Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

**Abstract**—Cross-layer optimization for efficient multimedia communications is an important emerging issue towards providing better quality-of-service (QoS) over capacity-limited wireless channels. This paper presents a cross-layer optimization approach that operates between the application and physical layers to achieve high fidelity downlink video transmission by optimizing with respect to a quality criterion termed “visual entropy” using *Lagrangian relaxation*. By utilizing the natural layered structure of wavelet coding, an optimal level of power allocation is determined, which permits the throughput of visual entropy to be maximized over a multi-cell environment. A theoretical approach to optimization using the Shannon capacity and the Karush-Kuhn-Tucker (KKT) conditions is explored when coupling the application with the physical layers. Simulations show that the throughput gain for cross-layer optimization by visual entropy is increased by nearly 80% at the cell boundary as compared with peak signal-to-noise ratio (PSNR).

**Index Terms**—Cross-layer optimization, downlink wavelet video, power allocation, visual entropy, wavelet coding.

## I. INTRODUCTION

IT is hoped that the rapid anticipated growth in wireless multimedia demand over the next few years will be matched by corresponding substantial increases in wireless channel capacity via the use of broader bandwidths. In parallel with extensive research on video compression and transmission centered on the standard series of Joint Photographic Experts Group (JPEG), Moving Picture Experts Group (MPEG), and H.264 codecs, considerable effort has been applied to the development of more efficient radio resource utilization over wireless channels. In order to better control multimedia wireless transmission, quality-of-service (QoS) parameters, such as transmission delay, bit error rate, and delay jitter, are widely utilized to characterize multimedia traffic at the physical layer.

Manuscript received May 24, 2009; revised November 28, 2009, June 14, 2010, and November 01, 2010; accepted March 08, 2011. Date of publication April 05, 2011; date of current version July 20, 2011. This work was supported by the Ministry of Knowledge Economy (MKE), Korea, under the national HRD support program for convergence information technology supervised by the National IT Industry Promotion Agency (NIPA) (NIPA-2010-C6150-1001-0013) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0011995). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhihai (Henry) He.

H. Lee and S. Lee are with the Wireless Network Laboratory, Center for Information Technology, Yonsei University, Seoul 120-749, Korea (e-mail punk-tank@yonsei.ac.kr; slee@yonsei.ac.kr).

A. C. Bovik is with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2011.2134840

Recent efforts have highlighted issues of cross-layer optimization for achieving a better QoS over capacity-limited wireless channels. In [1], a modified MPEG-4 coding scheme was employed for progressive data transmission by controlling the number of subcarriers over a multi-carrier system. In [2], a framework was presented for joint source-channel coding and power adaptation. Error resilient source coding, channel coding, and transmission power adaptation were jointly designed to optimize video quality given constraints on the total transmission energy and delay for each video frame. In [3], video quality expressed as minimum total expected mean square error (MSE) was controlled using constraints on transmission cost and delay in a lossy network environment. In [4], the benefits of characterizing video quality by the variance of the end-to-end distortion was explored, when allocating limited source and channel resources. However, the focus of these and other cross-layer approaches has mostly been fixed on the joint source-channel coding schemes, where the source encoding rate is moderated by the rate adaptation. The main contribution in [5] was to propose a distortion aware scheduling scheme for packet-based video transmission over wireless networks. It would be quite useful for resource management if information regarding the incoming video quality and other utility data were made available to the physical layer. Another difficulty is a lack of utility and quality criteria for evaluating performance gains afforded by the use of cross-layer processing. The most widely-used quality criteria is the peak signal-to-noise ratio (PSNR), although it is widely known to correlate poorly with perceptual video quality. This is a common problem for resource management when deciding which bitstream to serve first, when several bitstreams have the same number of bits and arrive simultaneously. Current radio resource control algorithms depend on the PSNR [5], [6] to improve video quality and therefore do not adequately capture perceptual quality, nor account for visual importance.

Fig. 4 illustrates the poor performance of PSNR in this regard. Although the PSNR values shown for Fig. 4(a)–(c) are approximately the same as those shown in Fig. 4(d)–(f), the perceptual quality of these two groups of images are significantly different, since the PSNR criterion does not capture perceptual image errors. In this and other senses, the PSNR as a quality assessment device poorly predicts subjective visual quality [7]. While PSNR remains widely used, a variety of much more powerful image and video quality indices have recently been introduced [8]–[10]. Currently, the Structural SIMilarity (SSIM) index described in [8], and the Visual Information Fidelity (VIF) index described in [9] and [10] are the most competitive, yet none of these have been applied to the cross-layer problem. However, both PSNR and SSIM are full-reference quality metrics, and therefore require reference images or videos for comparison.

What is needed in this context is a quality index that does not require a reference. In this direction, we have previously defined “visual entropy” as the expected number of bits required to represent image information mapped onto foveated human visual coordinates [11]. Using visual entropy, a new quality metric, termed the foveal signal-to-noise ratio (FSNR) was defined and used to optimize a video coding algorithm [12]. Visual entropy was also defined over the wavelet domain in [13] and [14], while [15] was an attempt to apply visual entropy to resource management over wireless networks. Visual entropy has also been deployed in a multiple-input multiple-output (MIMO) system in [16], and in a modified, standard-compliant H.264/advanced video coding (AVC) algorithm in [17] and [18]. In that work, visual entropy was used to estimate the data rate for a bitstream transmitted over a wireless network. One attractive advantage of visual entropy lies in the quantification of visual gain as a concrete quantity such as a bit. Thus visual gain can be displayed as an index to aid resource management in video communications.

In this paper, we explore a theoretical approach to cross-layer optimization between the application and physical layers, using a quality criterion expressed in terms of visual entropy [19]–[21]. The main contributions in this paper are itemized as follows: First, visual entropy is defined as a perceptually relevant utility function or index of visual quality, represented as bits. Second, a novel approach to closely couple the application and physical layers is provided, using a reliable information-passing protocol for quality control in diverse environments, such as orthogonal frequency division multiple access (OFDMA)-based network, MIMO transmission, and multi-hop communication. Third, A new method for reducing inter cell interference (ICI) that may occur in real multi-cell scenario is presented.

## II. OVERVIEW

Here we will use “slice” to represent an independent decodable unit obtained by scanning naturally layered wavelet coefficients based on a wavelet video coding scheme. In general, it can be assumed that the video service accesses each video sequence in the form of multiple slices.

### A. Optimization Based on Channel Feedback

Two critical parameters, i.e., channel quality and visual weight, are utilized in cross-layer optimization. The visual weight of each slice is encapsulated in the header of each packet, while the channel quality is obtained by using periodic channel feedback from all of the mobile stations (MSs) served by the base station (BSs). During the channel update period, the cross-layer optimization is accomplished as follows.

- 1st Step: each MS served reports the channel quality to the BS.
- 2nd Step: the BS determines an optimal power to each packet, using channel quality information from the MSs.
- 3rd Step: the BS transmits video packets using the optimal power allocation mechanism.
- 4th Step: each MS periodically reports its varying channel quality to the BS.
- 5th Step: the procedures of the 3rd and 4th steps are conducted until the channel quality is updated.

We follow the downlink signal flow based on the specifications of OFDMA-based network services [22].

### B. Packetization Including the Visual Weights

Using a progressive wavelet video encoder such as set partitioning in hierarchical trees (SPIHT) [20] or embedded block coding with optimized truncation (EBCOT) [23], each slice can be constructed by scanning the wavelet coefficients. Each wavelet coefficient can be assigned a different visual importance weighting. After summing the visual weights over each slice, the value can be included in the packet header. Since quality of motion is critical to subjective video quality, the visual weight of motion vectors may be significantly larger than certain wavelet coefficients.

### C. Visual Weight-Based Scheduling

If the source coding rate is larger than the channel capacity, the optimization scheme reduces the source coding rate to a threshold level so that the channel coding error probability is optimally reduced relative to the reconstructed visual quality. It may be assumed that the bitstreams are constructed in the streaming server before streaming service. When a service is invoked from a client, packets are generated from the bitstream and transmitted to the BS. The BS scheduler then examines the incoming packets, and executes unequal error protection according to the importance of each packet in order to prioritize scheduling for more important packets and to provide the best effort for less important packets.

### D. Error Concealment of Wavelet Coding

In general, transmission errors can occur due to packet loss or bit error over the wireless channel, or to packet dropping from the transmission queue. The quality degradation in the reconstructed video can be mitigated using an error concealment technique by exploited the spatial and temporal correlations in video data. Lost slices can be estimated based on the received slices of the current frame, or on the previous frames. Naturally, the error concealment technique used should rely on the type of picture. For an “I-frame”, if a previous frame is not available, then the frame must be constructed using received data only. The remaining part of an erroneous slice is discarded after the first transmission error. Due to the hierarchical structure of the wavelet decomposition, the distortion is not localized as compared to discrete cosine transform (DCT)-based video coding. If a highly correlated previous frame is available, then error concealment may be applied.

## III. VISUAL ENTROPY FOR WAVELET VIDEO CODING

### A. Definition of Visual Entropy

Visual entropy is defined as the expected number of bits that are required to represent image information that is mapped onto visual coordinates [13]–[15]. The visual weight  $w_m$  is characterized by two visual components: spatial domain weights  $w_m^s$  and frequency domain  $w_m^f$  weights as depicted in Fig. 1. When spatial information of interest, such as an object or objects, are identified, models of the non-uniform foveated sampling process of the human eye can be used to obtain  $w_m^s$  over the spatial domain. Likewise, visual sensitivity

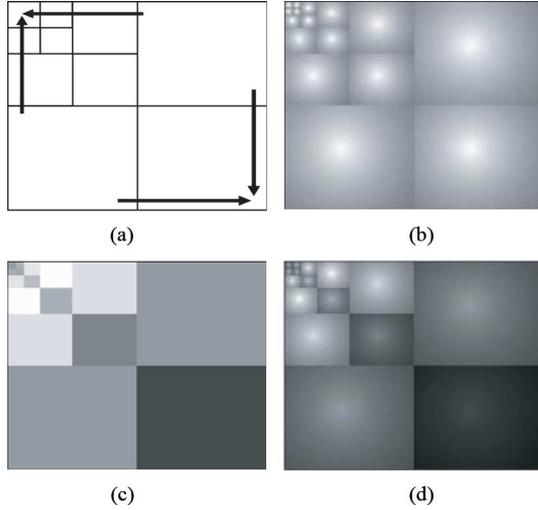


Fig. 1. Depiction of visual entropy defined over the wavelet domain. (a) Wavelet decomposition. (b) Spatial domain  $w_m^s$  weighting. (c) Frequency domain  $w_m^f$  weighting. (d) Total wavelet domain  $w_m$  weighting. The brightness represents the level of visual importance.

can be characterized by  $w_m^f$  over the frequency domain by modeling the contrast sensitivity function (CSF) of the human eye. The total weight over the two domains may be expressed  $w_m = w_m^f \cdot w_m^s$ . In layered video coding based on frequency division without foveation, the weight for each layer becomes  $w_m = w_m^f$ . In region-based, object-based, or foveation-based video coding without a layered structure, the weight simplifies to  $w_m = w_m^s$ . However, in hybrid video coding methods using object-based layered mechanisms, applying weights over both the spatial and frequency domains proves effective. Details regarding the definition  $w_m^f$  and  $w_m^s$  are discussed in [13]–[15]. The distribution of wavelet coefficients from “I- and P-frames” exhibits first-order statistical behavior that is well modeled by a Laplacian distribution [24]. The entropy, which is defined in the wavelet domain, for each frame then becomes  $H(a[m]) = \log_2 \sigma_m + \log_2 \sqrt{2e^2}$ , where  $a[m]$  is the wavelet coefficient,  $m$  is the index ( $1 \leq m \leq M$ ,  $M$  the total number of coefficients),  $\sigma_m$  is the variance, and  $e$  is the base of the natural logarithm.

Let  $\sigma_{I,m}$  and  $\sigma_{P,m}$  be the variances of coefficients for I- and P-frames, respectively. Visual entropy is then obtained by

$$H^w(a[m]) = w_m \left( \log_2 \sigma_m + \log_2 \sqrt{2e^2} \right) \quad (1)$$

where  $\sigma_m = \sigma_{I,m}$  for the I-frame and  $\sigma_m = \sigma_{P,m}$  for the P-frame (which means the residual information’s variance value for processing the P-frame).

### B. Consideration of Temporal Activity

Since the HVS is relatively insensitive to distortions in fast-moving regions, the visual weights for I-frames and P-frames may be modified according to the local temporal activity of the video. This can be computed as the mean value of the motion vectors in the frame. The authors in [25] proposed a quality metric for video quality assessment using the amplitude of motion vectors, and compared it with subjective scoring of the

videos. The temporal activity of the  $i$ th frame  $TA_i$  can be expressed

$$TA_i = \overline{|mv_x^i(x,y)|} + \overline{|mv_y^i(x,y)|} \quad (2)$$

$$1 \leq x \leq W, 1 \leq y \leq H$$

where  $\overline{|mv_x^i(x,y)|}$  and  $\overline{|mv_y^i(x,y)|}$  are the mean values of the horizontal and vertical components of the motion vector at spatial coordinate  $(x,y)$  in the  $i$ th frame, and  $W$  and  $H$  are the width and height of the video sequence, respectively. The visual weights  $w_m^t$  can be redefined to reflect temporal activity as

$$w_m^t = \frac{w_m}{c_1 + \frac{(\max(TA_i, c_2))^2}{c_3}} \quad (3)$$

where  $c_1$ ,  $c_2$ , and  $c_3$  are empirical constants taking values “2.5”, “5”, and “30” in [25].

### C. Localization Region of Interests (ROI)

Simple interactive methods that are effective for determining areas of visual importance include the use of a mouse or a touch screen. For images or video that do not change rapidly, “fixating” using these devices can be easily learned, although they are of limited utility in more generic applications. In controlled applications containing objects of known characteristics, such as faces, the focusing point or region can be automatically chosen and traced by detecting and recognizing the facial shape, color, or motion. Very sophisticated mechanical eye trackers have become commercially available that accurately track the direction of gaze of a human observer by detecting the motions of the eye, either through infrared (IR) reflection or by detecting the pupil. These devices are effective in situations where the user is located in front of a terminal or other display device and within a prescribed or expected physical location. Object recognition and “saliency” detection can also be used. Various approaches for saliency detection have been presented in [26]–[28].

## IV. PROPOSED SYSTEM MODEL

### A. Stepwise Power Allocation

In OFDM-based cellular networks, a frequency reuse factor (FRF) of 1 is desirable for greater channel throughput and ease of deployment. The main advantage of an FRF of 1 is the substantial increase in user capacity per unit bandwidth compared to other FRFs. In spite of this advantage, however, the channel throughput at the cell boundary rapidly decreases due to ICI. In such multi-cell environments, the ICI is a major factor leading to throughput degradation. The signal-to-interference and noise ratio (SINR) and the channel throughput of an MS decreases with distance from the BS [29].

In order to prevent quality degradation as a function of the location of the MSs, it is necessary to develop an interference mitigation technique, e.g., by controlling the power amplitude levels over the broadband. One strategy to achieve a higher FRF is to divide the frequency band into separate subbands as it is done in traditional CDMA systems, which are then assigned to each cell to avoid frequency overlap. By exploiting higher FRFs, it is possible to significantly reduce interference. However, each

cell is then able to use only a fraction of the available bandwidth, and so a decrease in frequency utilization and channel throughput is inevitable. The proposed power allocation scheme uses a stepwise power pattern, where the frequency band is subdivided into subbands and a power level allocated to each subband. However, the band corresponding to the greatest power in one cell is switched by employing an FRF of 7. In such a stepwise power pattern, each band will have reduced interference compared to the ordinary power allocation case, and therefore, ICI of the home cell would also decrease and the throughput increase.

The band that is allocated the largest amount of power has a relatively high channel gain owing to the reduced ICI, while the band allocated the smallest power is severely interfered with relative other bands. Thus, if a non-uniform channel quality index is efficiently utilized to transmit data having non-uniform visual entropies generated from the layered wavelet coding mechanism, a higher gain in visual throughput should be achievable. For example, the SPIHT and EBCOT codecs use this concept to generate embedded progressive video. After extracting the coefficients from the first sorting and refinement pass, visual entropy of these data is obtained. Using visual entropy, the transmitted power is allocated to the first band while the video data is loaded to it. The data extracted from the next pass is loaded to the second band with power allocation similar to the first step. This procedure is repeated through the last sorting and refinement pass.

To measure the summation of visual entropy after the  $l^{\text{th}}$  sorting and refinement pass, the empirical model for the selected coefficient is defined by

$$H(a[m_l], q_l) = \frac{\alpha_l}{q_l} \left( \log_2 \sigma_{m_l} + \log_2 \sqrt{2e^2} \right) + \beta_l, \quad \forall m_l \in M_l \quad (4)$$

where  $m_l$  is the selected coefficient on the  $l^{\text{th}}$  pass,  $M_l$  is the set of coefficients along the  $l^{\text{th}}$  pass, and  $\alpha_l$ ,  $\beta_l$ , and  $q_l$  are the constants and the quantization step, respectively. The method to obtain the constants  $\alpha_l$  and  $\beta_l$  is described in our recent papers [13]–[18]. Therefore, the quantized bitstream after the  $l^{\text{th}}$  pass becomes

$$\bar{H}(a[m_l], q_l) = \sum_{m \in M_l} H(a[m_l], q_l) \quad (5)$$

and visual entropy for this bitstream also becomes

$$\bar{H}^w(a[m_l], q_l) = \bar{w}_l^t \bar{H}(a[m_l], q_l) \quad (6)$$

where  $\bar{w}_l^t = \sum_{m \in M_l} w_m^t$  represents the visual importance contained in the bitstream, which is obtained by adding the visual weights of the coefficients in [13]–[18].

### B. SINR at the Multi-Carrier-Based Cellular System

Without any loss of generality, hereafter let an MS be located at  $x$  in the  $i^{\text{th}}$  cell (home cell). The pathloss between the MS and the  $i^{\text{th}}$  BS is given by  $G(i, x) = r^{-\alpha}(i, x)10^{\xi(i, x)/10}$ , where  $r(i, x)$  is the distance between  $x$  in the  $i^{\text{th}}$  BS, and  $\xi(i, x)$  is a Gaussian distributed random variable with zero mean and standard deviation representing shadowing over the  $i^{\text{th}}$  BS [30]. Typically, the mean of  $\xi_x$  is zero, and its standard deviation is in

the range of 6–10 dB for signals from adjacent BSs and 2–2.5 dB for signals from the home BS. For frequency-selective fading, the baseband impulse response of a multipath channel can be expressed as

$$h_x(\tau) = \sum_{v=0}^{V-1} h_{x,v} \delta(\tau - \tau_v) \quad (7)$$

where  $h_{x,v}$  is the wide sense stationary channel of the user at the  $x^{\text{th}}$  position and the  $v^{\text{th}}$  multipath,  $V$  is the maximum number of resolvable multipath components, and  $\tau_v$  is the excess delay of the  $v^{\text{th}}$  multipath component. Using (7), the frequency response at the  $c^{\text{th}}$  subcarrier of the user can be expressed

$$H_{x,c} = \frac{1}{\sqrt{C}} \sum_{\tau=0}^{C-1} h_x(\tau) e^{-j2\pi c\tau/C} \quad (8)$$

where  $C$  is the total number of subcarriers. Based on both large-scale and frequency-selective fading the SINR of the  $k^{\text{th}}$  user at the  $c^{\text{th}}$  subcarrier is given by

$$\text{SINR}_{k,x,c} = \frac{G(i, x) \cdot S}{\sigma^2} |H_{x,c}|^2 \quad (9)$$

where  $S$  is the maximum transmit power of each BS and  $\sigma^2$  is the interference and noise power. For simplicity, assume that  $\sum_{v=0}^{V-1} |h_{x,v}|^2 = 1$ ; hence,  $\sum_{c=0}^{C-1} |H_{x,c}|^2 = 1$ . The normalized channel gain  $\gamma_{x,c}$  is then

$$\gamma_x^i = \frac{S \cdot |H_{x,c}|^2}{S \cdot \sum_{c=1}^C |H_{x,c}|^2} = |H_{x,c}|^2. \quad (10)$$

The normalized channel gain  $\gamma_x^i$  can be modeled as an exponential distribution with mean  $\bar{\gamma}$ [31]

$$f_{\Gamma}(\gamma_x^i) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma_x^i}{\bar{\gamma}}\right), \quad 0 \leq \gamma_x^i \leq \infty. \quad (11)$$

The ICI is given by

$$I_{oc} = \sum_{j=1}^{N_{oc}} S \cdot G(j, x) \cdot \gamma_x^j \quad (12)$$

where  $j$  is an index of the adjacent BS,  $N_{oc}$  is the number of neighbor cells ( $N_{oc} = 18$  in 2 tiers. In general, the SINR for the  $k^{\text{th}}$  MS at  $x$  in the  $i^{\text{th}}$  cell without any power allocation is (we omit the index  $c$  based on the assumption that the  $k^{\text{th}}$  user is assigned the  $c^{\text{th}}$  subcarrier)

$$\text{SINR}_{k,x} = \frac{S \cdot G(i, x) \cdot \alpha_{i,k} \cdot \gamma_x^i}{I_{sc} + I_{oc} + N_0 BW_T} \quad (13)$$

where  $G(i, x)$  is the pathloss between the MS and the home BS,  $I_{sc}$  is the intra cell interference ( $I_{sc} = G(i, x) \cdot S \cdot (1 - \alpha_{i,k}) \cdot \gamma_x^i$ ),  $N_0$  is the power spectral density of additive white Gaussian noise (AWGN),  $\alpha_{i,k}$  is the normalized number of carriers of the  $k^{\text{th}}$  user in the  $i^{\text{th}}$  BS, known as the code portion, and  $BW_T$  is the total bandwidth. The energy per bit to noise power spectral density ratio  $E_b/N_0$  then becomes  $\text{SINR}_{k,x} BW_T / R_b$ , where

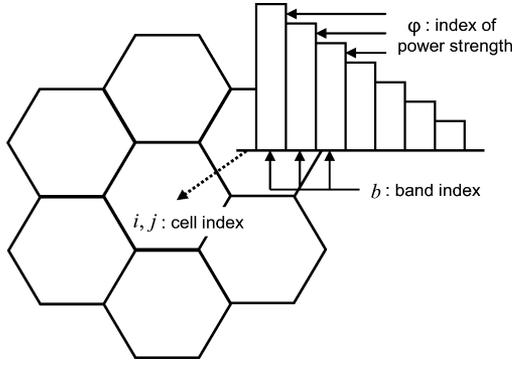


Fig. 2. Depiction of the parameters used.

$R_b$  is the data rate [30], [32]. If each subcarrier is perfectly orthogonal to the others, as in a general OFDM scheme, then  $\alpha_{i,k}$  becomes 1.

### C. Throughput of Visual Entropy With Stepwise Power Allocation

Assume that each cell has a different power pattern, with a periodic power circulation with an FRF of 7, as shown in Fig. 2. Under this system model, the SINR for the  $b$ th band of the  $k$ th MS at position  $x$  is

$$\begin{aligned} SINR_{k,b,x} &= \frac{S_b \cdot \alpha_{i,k} \cdot G(i,x) \cdot \gamma_{x,b}^i}{I_{sc} + I_{oc} + N_0 BW_b} \\ &= \frac{S_{k,b} \cdot \gamma_{x,b}^i}{I_{sc} + I_{oc} + N_0 BW_b} \end{aligned} \quad (14)$$

where  $S_b$  is the maximum transmit power for the  $b$ th band or the  $b$ th subcarrier (i.e.,  $\alpha_{i,k} = 1$ ) of the  $k$ th MS,  $BW_b$  is the bandwidth for the  $b$ th band,  $\gamma_{x,b}^i$  is the normalized channel gain of the  $b$ th band or the  $b$ th subcarrier over the  $i$ th BS,  $I_{sc} = S_b \cdot (1 - \alpha_{i,k}) \cdot G(i,x) \cdot \gamma_{x,b}^i$ , and  $I_{oc} = \sum_{j \in N_{oc}} S_b \cdot G(j,x) \cdot \gamma_{x,b}^j$ .

To describe the power allocation mechanism, the following parameters are defined:

- $\phi_b^i$  strength of the downlink power allocated to the  $b$ th band in the  $i$ th cell;
- $p_b^k$  transmitted power for the  $b$ th band of the  $k$ th MS;
- $g_b^k$  link gain for the  $b$ th band of the  $k$ th MS, which is a function of the path loss, the mean shadowing, and the ICI.

Under the assumption  $I_{sc} + I_{oc} \gg N_0 BW_b$ , the formula in (14) can be rewritten

$$\begin{aligned} SINR_{k,b,x} &= \frac{\phi_b^i \cdot S_{k,b} \cdot \gamma_{x,b}^i}{I_{sc} + I_{oc}} \\ &= \frac{\phi_b^i \cdot S_b \cdot \alpha_{i,k} \cdot G(i,x) \cdot \gamma_{x,b}^i}{I_{sc} + I_{oc}} \end{aligned} \quad (15)$$

where  $I_{sc} = \phi_b^i \cdot S_b \cdot (1 - \alpha_{i,k}) \cdot G(i,x) \cdot \gamma_{x,b}^i$  and  $I_{oc} = \sum_{j \in N_{oc}} \phi_b^j \cdot S_b \cdot G(j,x) \cdot \gamma_{x,b}^j$ . For brevity, denote

$$p_b^k = \phi_b^i \cdot S_b \cdot \alpha_{i,k}, \quad g_b^k = \frac{G(i,x) \cdot \gamma_{x,b}^i}{I_{sc} + I_{oc}}. \quad (16)$$

The data rate for the  $b$ th band of the  $k$ th user is obtained using the Shannon capacity formula for the Gaussian channel as

$$\bar{R}_b^k = BW_b \log_2(1 + SINR_{k,b,x}). \quad (17)$$

To reduce the gap between the Shannon capacity and the real capacity, a scaling factor  $\zeta_{k,b,x} \in (0, 1]$  is introduced. The scaled data rate for the realistic system then becomes

$$\begin{aligned} \tilde{R}_b^k &= BW_b \log_2(1 + \zeta_{k,b,x} \cdot SINR_{k,b,x}) \\ &= BW_b \log_2(1 + \zeta_{k,b,x} \cdot p_b^k g_b^k) \\ &= BW_b \log_2(1 + p_b^k \tilde{g}_b^k). \end{aligned} \quad (18)$$

In terms of link capacity, the numerical formula in (18) gives a more strict upper bound. In view of implementation, a more achievable rate for the system would be obtained using a coding technique such as turbo code, which nearly achieves Shannon capacity [5]. The scaled data rate in (18) can be interpreted as the average number of bits in a packet to be transmitted for a given bit error probability (BER). When  $M'$ -quadrature amplitude modulation (QAM) is used, the BER of  $SINR_{k,x}$  becomes

$$BER(SINR_{k,x}) = \frac{2}{z} \left(1 - \frac{1}{M'}\right) Q \left( \sqrt{\frac{3z}{M' - 1} \frac{E_b}{N_0}} \right) \quad (19)$$

where  $E_b/N_0 = SINR_{k,x} \cdot BW_b/R_b$ ,  $Q(\cdot)$  is the Q-function, and  $z$  is the number of bits per symbol. The packet error probability (PER) of  $SINR_{k,x}$  then becomes

$$PER(SINR_{k,x}) = 1 - (1 - BER(SINR_{k,x}))^{\bar{H}(a[m_{l,k}], q_{l,k})} \quad (20)$$

where  $a[m_{l,k}]$  is the coefficient from the  $l$ th pass extraction for the  $k$ th MS and  $\bar{H}(a[m_{l,k}], q_{l,k})$  is the number of bits per packet. The equation above is based on the assumption that each transmitted packet contains the bitstream of (5) from the  $l$ th pass extraction and the  $SINR$  during the packet transmission time is a constant. If the packet propagation time is shorter than the transmission interval, the average number of transmitted bits  $R_{l,k}$  in the packet can be obtained by

$$R_{l,k} = (1 - PER(SINR_{k,x})) \bar{H}(a[m_{l,k}], q_{l,k}) \quad (21)$$

and we can consider  $R_{l,k}$  as the number of bits in the  $l$ th transmitted packet for the  $k$ th MS.

The Shannon capacity in (17) is the upper bound of the scaled data rate in (18). In addition, we can make a tight bound by using the scaled data rate, as long as it is always higher than the number of transmitted bits in (21)

$$R_k = \sum_{l=1}^L R_{l,k} \leq \sum_{b=1}^B \tilde{R}_b^k \leq \sum_{b=1}^B \bar{R}_b^k = \bar{R}_k. \quad (22)$$

An appropriate scaling factor  $\zeta_{k,b,x}$  can be obtained when the first equality holds in (22), as follows:

$$\zeta_{k,b,x} \geq \frac{2^{1/BW_b(1 - PER(p_b^k \cdot g_b^k))} \bar{H}(a[m_{l,k}], q_{l,k})}{(p_b^k \cdot g_b^k)}. \quad (23)$$

However,  $\zeta_{k,b,x}$  obtained by (23) varies with position, because it is a function of  $x$ . We need to fix the value of  $\zeta_{k,b,x}$  regardless of position, while keeping the relation in (22). The

smaller the value of  $\zeta_{k,b,x}$  is, the closer the scaled data rate comes to the average number of transmitted bits. To choose an optimal  $\zeta_{k,b,x}$  which leads the scaled data rate to the tight bound of the average number of transmitted bits, the following algorithm is applied:

**Step 1. Obtain  $\zeta_{k,b,x}$  for all  $x$  points using the equality in(23)**

$$\mathbf{X} = \{x_0, x_1, \dots, x_N\}$$

$$\zeta_{k,b} = [\zeta_{k,b,x_0} \zeta_{k,b,x_1} \dots] \text{ for } \forall x \in \mathbf{X}.$$

**Step 2. Screen  $\zeta_{k,b,x}$  according to the condition(22)**

For  $\forall x \in \mathbf{X}$

$$R_{l,k} = (1 - \text{PER}(SINR_{k,b,x})) \bar{H}(a[m_{l,k}], q_{l,k})$$

$$\tilde{R}_b^k = BW_b \log_2(1 + \zeta_{k,b,x} \cdot SINR_{k,b,x})$$

If  $\sum_{l=1}^L R_{l,k} > \sum_{b=1}^B \tilde{R}_b^k$

$$\zeta_{k,b} = \{\zeta_{k,b} \setminus \zeta_{k,b,x}\}.$$

**Step 3. Find the minimal scaling factor  $\zeta_{k,b}^*$**

$$\zeta_{k,b}^* = \arg \min_{\forall \zeta_{k,b,x} \in \zeta_{k,b}} \zeta_{k,b,x}.$$

Once the optimal scaling factor  $\zeta_{k,b}^*$  is obtained, the scaled data rate is exploited as a tight upper bound of  $R_k$  in (22) When each packet which contains the bitstream from the  $l$ th pass is transmitted over the  $b$ th band, we can treat the index  $l$  as the same as the index  $b$ . The sum of transmitted visual entropy, which is the weighted version of the bits transmitted via each packet is then obtained as

$$R_k^w = \sum_{l=1}^L (1 - \text{PER}(SINR_{k,x})) \bar{w}_l^t \bar{H}(a[m_{l,k}], q_{l,k}). \quad (24)$$

## V. OPTIMAL DOWNLINK POWER SET

### A. Main Goal of Optimization

The proposed optimization is intended to transmit video data more reliably. Previous authors have developed methods for maximizing the PSNR, or equivalently to minimizing the MSE [5], [33], [34]. In this paper, visual entropy is considered as a weighted data rate, where the weight is calculated based on a model of human visual system. If the weight is uniform, an equivalently optimal solution is obtained by maximizing the PSNR. Thus, in the final analysis, optimization using visual entropy is a general form of optimization, and maximizing visual entropy is generally equivalent to maximizing a visually-weighted quality metric.

### B. Problem Formulation Using Lagrangian Relaxation

It is assumed that the proposed system employs Automatic Repeat-reQuest (ARQ) and forward error correction (FEC) schemes. It is also assumed that the transmitted rate  $R_b^k$  reflects the header information using the scaling factor in (18). To

maximize the throughput of visual entropy, an optimization problem for transmitting the information of each frame can be formulated:

$$\max_{p_b^k} \sum_{k=1}^K \sum_{l=1}^L R_{l,k}^w = \sum_{k=1}^K \sum_{l=1}^L \bar{w}_l^t R_{l,k}$$

$$\text{subject to } \sum_{k=1}^K \sum_{b=1}^B p_b^k \leq \bar{P} \quad (25)$$

where  $\bar{P}$  is the total transmit power w.r.t. all bands and  $B$  is the number of divided bands.

Since  $\bar{w}_l^t$  is the average visual weight for the packet, the problem of maximally transmitting visual entropy can be solved using this header information. The solution of the optimization problem (25) is an optimal power set,  $\{p_1^k, p_2^k, \dots, p_B^k\} = \{\mathbf{P}^k\}$  for the  $k^{\text{th}}$  MS. Because  $R_{l,k}^w$  is the function of  $(p_b^k$  and  $g_b^k)$ , and  $g_b^k$  is assumed not to change with varying  $p_b^k$ , (25) is a convex problem, which has an optimal solution with respect to  $p_b^k$ .

Formulate a *Lagrangian relaxation*

$$J(p_b^k, \lambda) = \sum_{k=1}^K \sum_{l=1}^L R_{l,k}^w + \lambda \left[ \bar{P} - \sum_{k=1}^K \sum_{b=1}^B p_b^k \right] \quad (26)$$

where  $\lambda$  is a nonnegative *Lagrangian* multiplier. Using (22), the *Lagrangian* function can be written as

$$J(p_b^k, \lambda) = \sum_{k=1}^K \sum_{l=1}^L \bar{w}_l^t R_{l,k} + \lambda \left[ \bar{P} - \sum_{k=1}^K \sum_{b=1}^B p_b^k \right]$$

$$\leq \sum_{k=1}^K \sum_{l=1}^L \bar{w}_l^t \tilde{R}_b^k + \lambda \left[ \bar{P} - \sum_{k=1}^K \sum_{b=1}^B p_b^k \right]$$

$$= \sum_{k=1}^K \sum_{b=1}^B \bar{w}_b^t \cdot BW_b \log_2(1 + p_b^k \zeta_b^k)$$

$$+ \lambda \left[ \bar{P} - \sum_{k=1}^K \sum_{b=1}^B p_b^k \right] = \tilde{J}(p_b^k, \lambda). \quad (27)$$

The new *Lagrangian* function of (27) yields an optimal solution of the power set that is equivalent to the solution of (25), when the equality condition in (22) holds. Therefore, the solution of (27) becomes equivalent to (25) when the total number of transmitted bits is the same as the upper bound of the Shannon capacity. Taking derivatives with respect to  $p_b^k$  and  $\lambda$ , respectively yields the Karush-Kuhn-Tucker (KKT) conditions:

$$\frac{\partial \tilde{J}}{\partial p_b^k} = \bar{w}_b^t \cdot BW_b \frac{\tilde{g}_b^k}{1 + p_b^k \tilde{g}_b^k} - \lambda \leq 0 \quad (28)$$

$$p_b^k \cdot \frac{\partial \tilde{J}}{\partial p_b^k} = 0 \quad (29)$$

$$\lambda \left( \bar{P} - \sum_{k=1}^K \sum_{b=1}^B p_b^k \right) = 0. \quad (30)$$

From (28) and (29), if the power  $p_b^k$  is allocated to band  $b$  (that is,  $p_b^k \geq 0$ ), then the complementary slackness condition is satisfied:

$$\bar{w}_b^t \cdot BW_b \frac{\tilde{g}_b^k}{1 + p_b^k \tilde{g}_b^k} = \lambda. \quad (31)$$

In addition, the optimal values of  $p_b^k$  and its multiplier  $\lambda$  are given by

$$p_b^k = \frac{1}{\tilde{g}_b^k} \left[ -1 + \frac{\bar{w}_b^t \cdot BW_b \cdot \tilde{g}_b^k}{\lambda} \right] \quad (32)$$

$$\lambda = \frac{\sum_{k=1}^K \sum_{b=1}^B \bar{w}_b^t \cdot BW_b \cdot \tilde{g}_b^k}{\bar{P} + \sum_{k=1}^K \sum_{b=1}^B \frac{1}{\tilde{g}_b^k}}. \quad (33)$$

### C. Throughput Gain of the Proposed Scheme

To validate the optimal solution in (32) and to measure the benefit produced by the proposed scheme, the throughput gain is measured in terms of visual entropy. The maximum number of bits transmitted to each MS is expressed using  $C_{pro}$  for the proposed method and  $C_{test}$  for the test method, as follows:

$$C_{pro} = \sum_b^B BW_b \cdot \log_2(1 + \zeta_{k,b,x} SINR_{k,b,x}) \quad (34)$$

$$C_{test} = \sum_b^B \frac{BW_T}{B} \cdot \log_2(1 + \zeta_{k,b,x} SINR_{k,b,x}). \quad (35)$$

The test method measures the volume of transmitted data using the MSE rather than visual entropy [5], [33], [34]. In other words, the PSNR criterion is used for the test method. A modified SPIHT in [21] is employed as the encoding scheme for both methods. The transmitted visual entropy for the proposed and test methods can be expressed

$$C_{pro}^{rw} = \sum_b^B BW_b \cdot \bar{w}_b^t \cdot \log_2(1 + \zeta_{k,b,x} SINR_{k,b,x}) \quad (36)$$

$$C_{test}^{rw} = \sum_b^B \frac{BW_T}{B} \cdot \bar{w} \cdot \log_2(1 + \zeta_{k,b,x} SINR_{k,b,x}) \quad (37)$$

where  $\bar{w}$  can be used for any kind of weights w.r.t the purpose of comparison. For a given capacity  $c$ , the attained gain  $G_{th}$  for the stepwise allocation scheme is obtained using (36) and (37):

$$G_{th} = \frac{(C_{pro}^{rw} - C_{test}^{rw})}{C_{pro}^{rw}}. \quad (38)$$

### D. Extended Environments and Applications

Although we focus on cross-layer optimization over a multi-cell environment (25), the optimization approach can be extended to various realistic situations with some manipulations. In [16], a power allocation method for singular value decomposition (SVD)-based MIMO systems was presented. In the optimization, (25) can be changed into

$$\begin{aligned} \max_{p_l} \quad & \sum_{l=1}^r \bar{w}_l^t \cdot BW_b \log_2 \left( 1 + \frac{p_l}{N_0} \lambda_l \right) \\ \text{subject to} \quad & \sum_{l=1}^r p_l \leq \bar{P} \end{aligned} \quad (39)$$

TABLE I  
SIMULATION PARAMETERS

Parameters	
Number of users	10
Number of cells	19 (considered with two tiers)
Number of divided bands	7
Number of carriers	1024
Total Bandwidth	10 MHz
Modulation order	4 (16-QAM)
Channel frame length	5 msec
Transmission interval	15 msec
Delay constraint	50 msec

where  $l$  is the divided channel after SVD,  $r$  is the total parallel decomposed channels, and  $\lambda_l$  is the eigen value of the  $l$ th channel. Moreover, this method can also be extended to multi-hop transmission where the extension of (25) is obtained as

$$\begin{aligned} \max_{p_n^k} \quad & \sum_{k=1}^K \arg \min_{n \in \mathbb{N}} (\bar{w}_n^t \cdot BW_b \log_2(1 + p_n^k \tilde{g}_n^k)) \\ \text{subject to} \quad & \sum_{k=1}^K \sum_{n \in \mathbb{N}} p_n^k \leq \bar{P} \end{aligned} \quad (40)$$

where  $n$  is the link index composing the multi-hop links and  $\mathbb{N}$  is the entire set of links.

In the downlink video sequence transmission, it is necessary to update the power pattern along the temporal axis by relying on the statistical behavior of the visual traffic. To simplify the optimization problem, the power pattern is assumed to be updated according to the frame rate of the video sequence. Thus, an optimal power set is determined having the same frame rate. In addition, it is assumed that the obtained power set for the home cell does not affect the solution of other cells. In other words, the magnitude of ICI is set to an average value over a long duration of temporal power allocation.

## VI. SIMULATION RESULTS

The parameters used for the simulation can be found in Table I, and pathloss, zero-mean shadowing, and frequency flat fading over a subband, to which the power is allocated, are embedded in the channel model as defined in Section IV. Each packet containing encoded data is carried over one frame (frame length = 5 ms), and the round trip time is assumed to 15 ms. The next packet is not transmitted until an acknowledgment (ACK) signal for the previous packet is received by the transmitter. Since the packet propagation time in a typical wireless environment is shorter than the transmission interval, it is assumed that the ACK or negative ACK (NACK) signal arrives in the transmitter before the next slot becomes available. Thus, the transmitter can retransmit the current packet in the next slot whenever channel errors occur in the current packet. The retransmission (a delay constraint) is limited by a maximum queueing delay bound (50 ms). Since the data is transmitted using a packet, the channel throughput can be calculated using the packet error probability which is a function of the SINR, the modulation order, the packet size, and so on.

As described in Section IV-B, (9) is the channel model used in this paper. For ease of analysis, it is assumed that the width

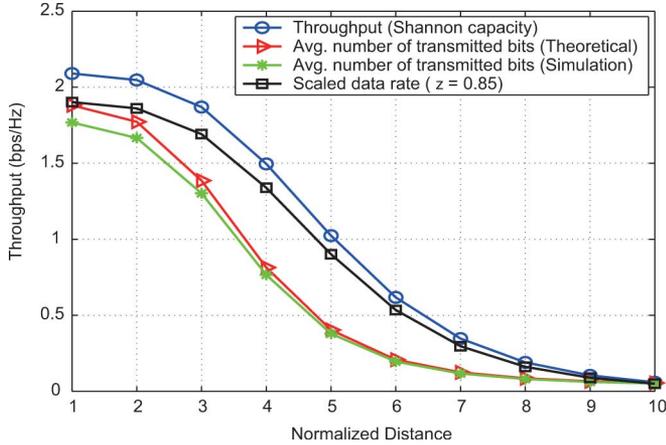


Fig. 3. Comparison between the throughput in (21) and in (34) and the average number of transmitted bits as a function of the normalized distance.

of each divided band for each MS is a constant. Fig. 3 compares the Shannon capacity obtained by (17) and the average number of transmitted bits obtained theoretically by (21) and obtained by simulations with 1 000 000 iteration times. Nevertheless, there exists a gap between the Shannon capacity and the data rate in Fig. 3. To obtain a tighter upper bound, we introduce a scaling factor to reflect the realistic system environment in (18). As shown in Fig. 3, the scaling factor  $\zeta_{k,b,x} = 0.85$  is obtained by the algorithm in Section IV-C and is applied in the following numerical analyses.

Fig. 4 depicts the differences obtained using the PSNR and visual entropy. The figure pairs ((a), (b)), ((c), (d)), and ((e), (f)) compare the differences in visual quality under the same PSNR value. Clearly, in these examples, visual entropy correlates more highly with visual quality than does PSNR. Quality distortions due to transmission errors over the error-prone wireless channel can be expected. Using an appropriate quality metric, we are able to demonstrate the performance of the cross-layer optimization relative to the PSNR.

In the simulation, it is assumed that the image “Lena” is used to measure performance on I-frames. The video sequence “Stefan” is used to measure performance on I-frames and P-frames. The simulation is executed by using the modified SPIHT coder described in [21], which provides pre-processing to obtain weights on the encoder side. We modulate the visual weight using  $w_m$  in (3) as the pre-processing weight. To demonstrate the improved performance of the proposed method, the resource allocation method in [5], [33], and [34] is used as a benchmark. The authors of [5], [33], and [34] defined the utility of each packet when it is transmitted. The utility of the packet is measured as the difference between the values of the PSNR when it is transmitted with errors, against when it is transmitted without errors. In other words, the utility represents the average loss of PSNR caused by the missing packets. The weight of the utility-based method is then obtained by

$$\bar{w}_l^u = \overline{\text{PSNR}}(Y_{a[m]}, Y_{a[\tilde{m}]}) - \text{PSNR}(Y_{a[m]}, Y_{\bar{a}[m_l]}) \quad (41)$$

where  $Y(a[m])$  is the spatial value from the inverse wavelet transform of  $a[m]$ ,  $\overline{\text{PSNR}}(Y_{a[m]}, Y_{a[\tilde{m}]})$  is the maximum value of PSNR, and  $Y_{\bar{a}[m_l]}$  is the spatial value from the inverse wavelet transform without data  $a[m_l]$ , respectively. In the



Fig. 4. Quality assessment using PSNR (in dB) versus visual entropy (in bits).

case of the “Lena” image, the weights in (41) becomes  $\bar{\mathbf{w}}^u = [27.6533 \ 28.1387 \ 20.6066 \ 22.1603 \ 18.8101 \ 3.1478 \ 7.6075]$  and we utilize  $\bar{\mathbf{w}}^u$  as  $\bar{\mathbf{w}}$  in (37) in the following numerical experiments instead of the visual weight. For fair comparison, three different power allocation schemes are conducted to measure performance in view of throughput: stepwise, equal, and utility-based power schemes.

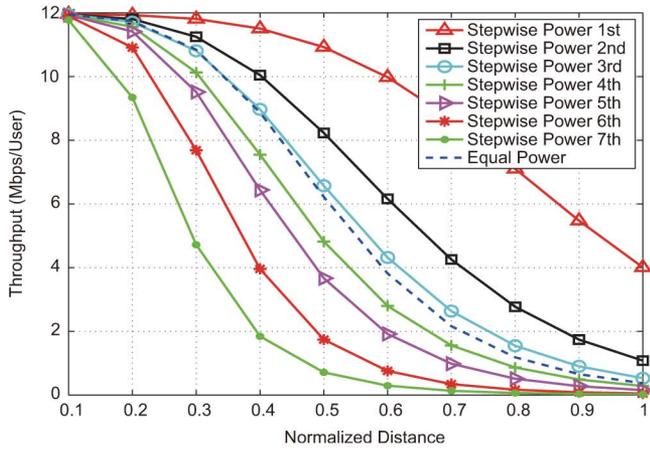
#### A. Performance Measurement for an I-Frame

Table II shows an optimal power set obtained by (32), which is the solution of (27). For the optimal set, it is assumed that one packet is allocated to each band. Although each band delivers the same volume of data, each bitstream has a different visual importance. Since the 1<sup>st</sup> band contains the most significant perceptual information, the highest power is allocated to it. The power patterns for the rest of the bands are determined to minimize ICI.

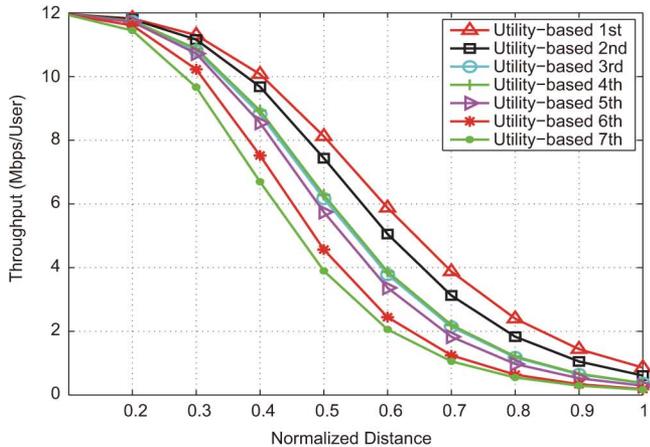
Fig. 5 shows throughput attained by the stepwise power allocation as compared to equal power in (a), utility-based power allocations in (b) over a multi-cell environment. Throughput

TABLE II  
OPTIMAL POWER SET

1	2	3	4	5	6	7
0.451	0.179	0.123	0.0924	0.0721	0.0532	0.0293



(a) Proposed case

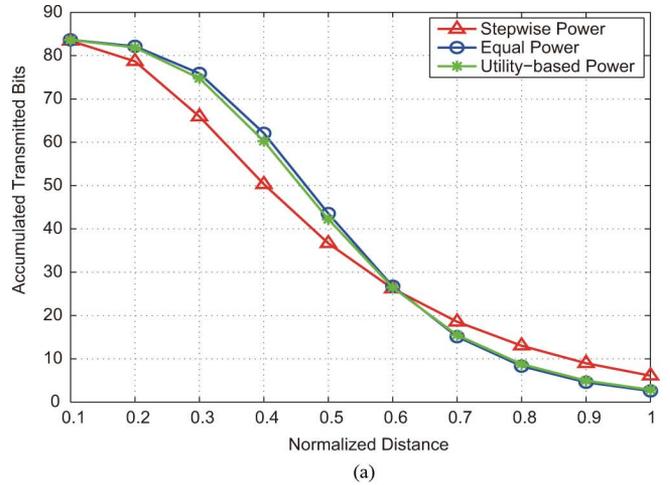


(b) Utility-based case

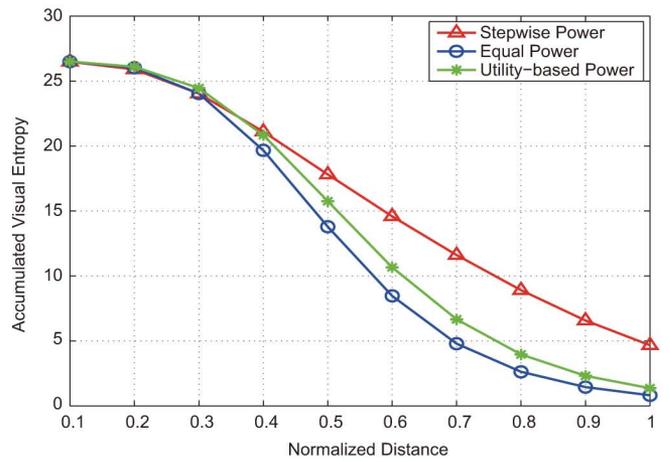
Fig. 5. Throughput for stepwise power allocation, equal power allocation, and utility-based power allocation against a normalized distance. (a) Stepwise power versus equal power. (b) Utility-based power allocations.

as given in (34) is plotted for the stepwise allocation scheme, and throughput as given in (35) is plotted for the equal power and utility-based schemes, as a function of the normalized distance from the BS. The findings show that an increase in the throughput of the 1st band at the edge of the cells results in an improvement in throughput, as shown in Fig. 5. Since the visual weight is largest here among the bands, it is expected that a much higher throughput of visual entropy can be obtained by the stepwise power allocation. By contrast, a decrease in throughput is obtained at the higher frequency bands (the 4th, 5th, 6th, and 7th bands). Even though the result of the utility-based scheme shows similar behavior as the stepwise power scheme, the throughput of the utility-based scheme is not higher than the proposed scheme.

Fig. 6(a) shows the total throughput [using (34) and (35)] and Fig. 6(b) shows the total sum of visual entropy [using (36) and (37)] along the normalized distance. The channel throughput of both equal and utility-based power allocation is higher than that of stepwise allocation in the middle region of the cell, but higher



(a)



(b)

Fig. 6. Transmitted bits versus visual entropy against normalized distance.

visual entropy is achieved using the stepwise allocation, particularly near the cell boundary. As shown in Fig. 6(b), although the proposed method entails a certain loss of transmitted bits in the middle region of the cell, the throughput gains in terms of visual entropy at the cell boundary are increased by as much as 80% and 70% compared to equal and utility-based power allocations.

Fig. 7 shows the relative visual qualities of the reconstructed image using stepwise power allocation of the image “Lena” [Fig. 7(a)], the reconstructed image using equal power allocation [Fig. 7(b)], and the reconstructed image using utility-based power allocation at the cell boundary [Fig. 7(c)]. At the cell boundary, the visual quality using the proposed scheme is apparently much better than obtained using the other quality mechanism. To measure the subjective visual qualities of Fig. 7(a)–(c), too, we put the SSIM indices for each figure.

### B. Performance Measurement for a P-Frame

Similar to the I-frame performance measurement, the solution of the optimization problem for the “Stefan” sequence can be obtained using (32). In the first step, the optimal I-frame solution is obtained on the first frame of “Stefan”. Using the reconstructed frame, the visual quality of the next P-frame can be obtained from the second frame of the video sequence. The optimal power set, [0.2930.1600.1580.1330.1270.1050.024], is obtained using the average visual weight for P-frames.



Fig. 7. (a) Reconstructed image (SSIM index = 0.9863) using the stepwise power allocation. (b) Reconstructed image using the equal power allocation (SSIM index = 0.9241). (c) Reconstructed image (SSIM index = 0.9364) using the utility-based power allocation for an MS located at the cell boundary.

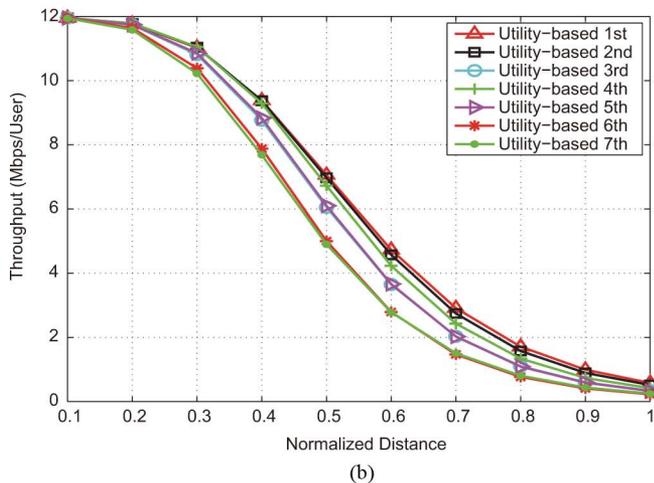
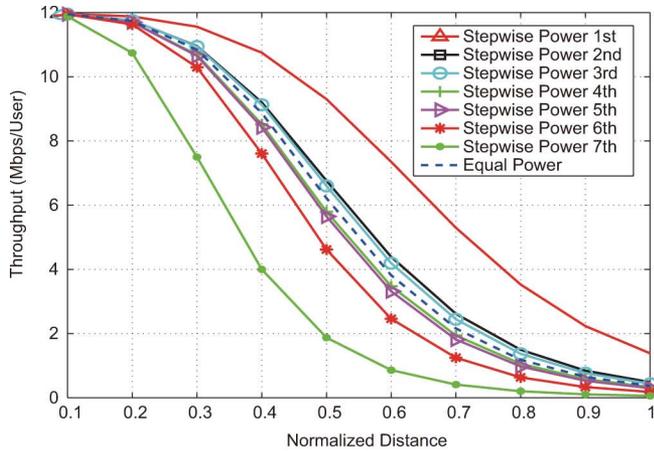


Fig. 8. Throughput for stepwise power allocation, equal power allocation, and utility-based power allocation against a normalized distance. (a) Stepwisepower versus equal power. (b) Utility-based power allocations.

Fig. 8 shows throughput attained by the stepwise power allocation as compared to equal power in (a), utility-based power allocations in (b) over a multi-cell environment. Results generally show the quite similar tendencies with results in Section IV-A. Similar to the I-frame analysis, visual gain is obtained although a loss of transmitted bits occurs over the middle region of the cell as shown in Fig. 9.

## VII. CONCLUSION

In this paper, a theoretical approach to cross-layer optimization between multimedia and wireless network layers was explored for downlink video transmission. The reference-free

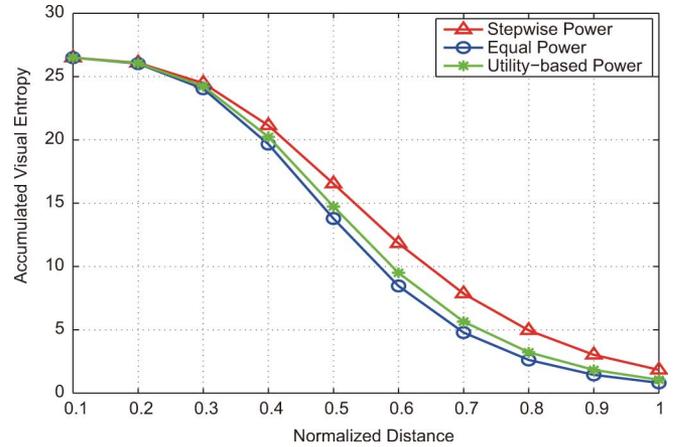


Fig. 9. Visual entropy against normalized distance.

quality criterion visual entropy was utilized for cross-layer optimization to deliver maximal visual information by controlling the power levels in the downlink cellular network. While seeking to maximize visual entropy, the optimal power allocation set was obtained utilizing Lagrangian relaxation. The optimal solution makes it possible to transmit maximum visual information while mitigating ICI.

## REFERENCES

- [1] N. Conci, G. B. Scorza, and C. Sacchi, "A cross-layer approach for efficient mpeg-4 video streaming using multicarrier spread-spectrum transmission and unequal error protection," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Genoa, Italy, Sep. 2005, vol. 1, pp. 11–14.
- [2] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power adaptation for energy efficient wireless video communications," *Signal Process. Image Commun.*, vol. 20, no. 4, Apr. 2005.
- [3] H. Wang, F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Cost-distortion optimized unequal error protection for object-based video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1505–1516, Dec. 2005.
- [4] Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Vapor: Variance-aware per-pixel optimal resource allocation," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 289–299, Feb. 2006.
- [5] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE J. Select. Areas Commun.*, vol. 25, no. 4, pp. 749–759, May 2007.
- [6] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proc. IEEE*, vol. 93, no. 1, pp. 135–147, Jan. 2005.
- [7] Z. Wang and A. C. Bovik, "Why is image quality assessment so difficult?," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, May 2002.
- [8] H. R. Sheikh, Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [9] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [10] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "An evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
- [11] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.
- [12] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.
- [13] H. Lee and S. Lee, "Visual data rate gain for wavelet foveated image coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Genova, Italy, Sep. 2005, vol. 3, pp. 41–44.
- [14] H. Lee and S. Lee, "Visual entropy gain for wavelet image coding," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 553–556, Sep. 2006.

- [15] H. Lee, S. Jeon, and S. Lee, "Cross-layer optimization for downlink wavelet image transmission," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Nov.–Dec. 2006.
- [16] H. Lee, S. Jeon, and S. Lee, "A cross-layer approach for maximizing visual entropy using closed-loop downlink MIMO," *EURASIP J. Adv. Signal Process.*, 2008.
- [17] U. Jang, H. Lee, and S. Lee, "Optimal carrier loading control for the enhancement of visual quality over OFDMA cellular networks," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1181–1196, Oct. 2008.
- [18] J. Park, H. Lee, S. Lee, and A. C. Bovik, "Optimal channel adaptation of scalable video over a multi-carrier based multi-cell environment," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1062–1071, Oct. 2009.
- [19] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- [20] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [21] Z. Wang, L. Lu, and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243–254, Feb. 2003.
- [22] H. Lee and S. Lee, "Definition of a new service class, artPS for video services over WiBro/Mobile WiMAX systems," *Wireless Netw.*, vol. 17, no. 1, pp. 103–117, 2011.
- [23] D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Dordrecht, The Netherlands: Kluwer, 2001.
- [24] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.
- [25] F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective no-reference metric for digital video quality assessment," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 685–688, Oct. 2005.
- [26] U. Rajashekar, I. v. d. Linde, A. C. Bovik, and L. K. Cormack, "Gaffe: A gaze-attentive fixation finding engine," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 564–573, Apr. 2006.
- [27] M. H. Sharif and C. Djeraba, "Exceptional motion frames detection by means of spatiotemporal region of interest features," in *Proc. IEEE Int. Conf. Image Processing*, Cairo, Egypt, Nov. 2009.
- [28] W. Li, C. Pan, and L. Liu, "Saliency-based automatic target detection in forward looking infrared images," in *Proc. IEEE Int. Conf. Image Processing*, Cairo, Egypt, Nov. 2009.
- [29] K. S. Gilhosen, I. M. Jacobs, R. Padovani, A. J. Viterbi, jr, L. A. Weaver, and C. E. Wheatley, III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, May 1991.
- [30] J. Kim, H. Son, and S. Lee, "Frequency reuse power allocation for broadband cellular networks," *IEICE Trans. Commun.*, vol. E89-B, no. 2, pp. 531–538, Feb. 2006.
- [31] M. K. Simon and M. S. Alouini, *Communication over Fading Channels*, 2nd ed. New York: Wiley, 2005.
- [32] H. Son and S. Lee, "Forward-link capacity analysis for MC-CDMA," *IEICE Trans. Commun.*, vol. E88-B, no. 10, pp. 4094–4096, Oct. 2005.
- [33] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, "Joint source adaptation and resource allocation for multi-user wireless video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 582–595, May 2008.
- [34] E. Maani, P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1663–1671, Sep. 2008.



**Hyungkeuk Lee** (S'06) received the B.S. degree and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea in 2005 and 2006, respectively. He is currently pursuing the Ph.D. degree.

In 2008, he was a Visiting Researcher in the Laboratory for Image and Video Engineering (LIVE) of Prof. A. C. Bovik at The University of Texas at Austin. His research interests include wireless resource allocation based on economics, video coding, cross-layer optimization, image/video

quality assessments, and compressive sensing.



**Sanghoon Lee** (M'05) was born in Korea in 1966. He received the B.S. degree in electrical engineering from Yonsei University, Seoul, Korea, in 1989, the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 1991, and the Ph.D. degree in electrical engineering from the University of Texas at Austin in 2000.

From 1991 to 1996, he worked for Korea Telecom. In June–August 1999, he worked for Bell-Lab., Lucent Technologies, on wireless multimedia communications. From February 2000 to December 2002, he worked on developing real-time embedded software and communication protocols for 3G wireless networks, Lucent Technologies. In March 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, where he is an Associate Professor. His research interests include 4G wireless networks, 3G W-CDMA/CDMA networks, multihop sensor networks, wireless multimedia communications, and image/video quality assessments.

Dr. Lee is an Associate Editor of the *Journal of Communications and Networks* (JCN) and the IEEE TRANSACTIONS ON IMAGE PROCESSING.



**Alan Conrad Bovik** (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1980, 1982, and 1984, respectively.

He is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute

for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published nearly 600 technical articles in these areas and holds two U.S. patents. His several books include the most recent companion volumes *The Essential Guides to Image and Video Processing* (New York: Academic Press, 2009).

Dr. Bovik was named the SPIE/IS&T Imaging Scientist of the Year for 2011. He has also received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009); the Education Award (2007); the Technical Achievement Award (2005); and the Meritorious Service Award (1998). He received the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008), the IEEE Third Millennium Medal (2000), and two journal paper awards from the international Pattern Recognition Society (1988 and 1993). He is a Fellow of the Optical Society of America (OSA), a Fellow of the Society of Photo-Optical and Instrumentation Engineers (SPIE), and a Fellow of the American Institute of Medical and Biomedical Engineering (AIMBE). He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996–1998; co-founder and Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; Editorial Board, the PROCEEDINGS OF THE IEEE, 1998–2004; Series Editor for *Image, Video, and Multimedia Processing*, Morgan and Claypool Publishing Company, 2003–present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, TX, in November 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.