CROSS-LAYER OPTIMIZATION FOR SCALABLE VIDEO CODING AND TRANSMISSION OVER BROADBAND WIRELESS NETWORKS

Jincheol Park, Hyungkeuk Lee and Sanghoon Lee

Wireless Network Lab., Center for IT of Yonsei University, Seoul, Korea, 120-749.
(e-mail: dewofdawn@yonsei.ac.kr, punktank@yonsei.ac.kr and slee@yonsei.ac.kr)

ABSTRACT

For seamless multimedia streaming services, it is very important to overcome severe ICI (InterCell Interference) over wireless networks, particularly in the cell border region. SVC (Scalable Video Coding) has been actively studied due to its advantage of channel adaptation. In this paper, we study an optimal solution for maximizing the expected visual entropy over an OFDM (Orthogonal Frequency Division Multiplexing)-based broadband network from the aspect of cross layer optimization. Based on this approach, an optimal set of source and channel parameters is obtained according to the user location over a multi-cell environment. From the numerical results, we quantify the optimal visual gain attained from the single layer video coding or the SVC according to channel capacity v.s. source data rate.

Index Terms— Scalable Video, Inter-Cell Interference, Visual Entropy, OFDM-based System, Cross-Layer Optimization

1. INTRODUCTION

Migrating toward 4^{th} generation (4G) systems, multimedia services are expected to be provided seamlessly by using a wider band with greatly improved spectral efficiency. The WiMax (World Interoperability for Microwave Access) or 3GPP LTE (Long Term Evolution) systems employ a frequency reuse factor of 1 for the sake of improving the overall link capacity. However, since the influence of ICI (InterCell Interference) is increased toward the cell border, the system performance is severely degraded at the cell boundary [6]. Thus, the seamlessness can not be guaranteed due to severe ICI near the cell border. For realtime video streaming service, several cross layer design approaches, such as [1][2][3], are analyzed by utilizing UEP (Unequal Error Protection) on the SVC. However, there are other issues not taken into account. In the side of channel, there is a sudden outage caused by ICI near the cell border. In the view of source, there is a lack of substantial quality criterion for characterizing the HVS (Human Visual System). Finally, the performances are only evaluated by demonstrating that the each proposed schemes produce a better visual quality based on the simulation result. In this paper, we present a cross-layer optimization approach for maximizing the visual entropy by allocating an optimal number of sub-carriers to each scalable bit-stream generated from the scalable video encoder. In the source side, the optimal set of video coding parameters are taken into account such as a cut-off frequency, QP (Quantization Parameter), and a data rate of layered video. In the channel side, an optimal number of the sub-carriers is obtained for each layered bit-stream.

This work was supported by the Korea Science and Engineering Foundation grant funded by the Korea government (MOST) (No.R01-2007-000-11708-0), and by Seoul Research & Business Development Program (11136M0212351).

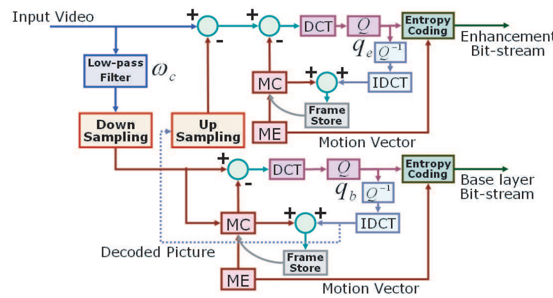

Fig. 1. The scalable video encoder block diagram

Figure 1 depicts the SVC block diagram. From the coder, a video sequence is coded into a base layer and an enhancement layer, as in the former SVC techniques, but a method of dividing the sequence is different. The frequency components of each layer are divided by using the alterable cutoff frequency, ω_c. Meanwhile, the visual weight of each layer can be obtained by the CSF (Contrast Sensitivity Function) defined in [5], according to the frequency components that each layer has. Applying the visual weight to the generated traffic of each layer, it is possible to quantify the visual importance in terms of visual entropy for the base and enhancement layers. Based on this criterion, an UEP scheme can be developed by determining an optimal number of sub-carriers for each layer. In the down link of the cellular networks, ICI relies on a relative position and shadowing factors from BSs (Base Stations). So an E_b/N_o (Bit Energy-to-Noise Density Ratio) formula is derived according to the normalized distance, in terms of the number of sub-carriers and the user bit rate. Using the system parameters (cutoff frequency, QP, number of sub-carriers, user bit rate), an objective function of the cross layer optimization problem can be formulated by the expected visual entropy that is delivered through the OFDM-based system.

2. SYSTEM MODEL

2.1. Visual Entropy Model

In [4], we have defined an empirical rate generated from a macro-block, as below,

$$r(\sigma_x^2) = \alpha_x h(\sigma_x^2) + \beta_x \tag{1}$$

where σ_x^2 is a variance of the macro-block. In addition, the empirical entropy $h(\sigma^2)$ of the quantized parameter q is given by

$$h(\sigma^2) = \begin{cases} \frac{1}{2}\log_2\left(2e^2\frac{\sigma^2}{q^2}\right), & \frac{\sigma^2}{q^2} > \frac{1}{2e} \\ \frac{e}{\ln 2}\frac{\sigma^2}{q^2}, & \frac{\sigma^2}{q^2} \leq \frac{1}{2e}. \end{cases} \tag{2}$$

Suppose that Z_x is a random variable for sample points σ_x^2. The random variable R_x associated with $r(\sigma_x^2)$ can then be expressed by $R_x = \alpha_x Z_x + \beta_x$ where α_x and β_x are constants. The expected value and variance of R are then

$$E[R_x] = \alpha_x E[Z_x] + \beta_x \text{ and } VAR[R_x] = \alpha_x^2 VAR[Z_x]. \tag{3}$$

Therefore, α_x and β_x become

$$\alpha_x = \sqrt{VAR[R_x]/VAR[Z_x]} \text{ and } \beta_x = E[R_x] - \alpha_x E[Z_x]. \tag{4}$$

The variance of each layer can be derived to parameterize by ω_c, where $\omega_c = 2\pi f_c$ and f_c is the cutoff frequency, as in [4]. Thus, the empirical rate can be represented by

$$\begin{aligned} r_e(\omega_c, q_e) &= \alpha_e h(\omega_c, q_e) + \beta_e \\ r_b(\omega_c, q_b) &= \alpha_b h(\omega_c, q_b) + \beta_b \end{aligned} \tag{5}$$

where the subscript, b and e mean "of the base layer" and "of the enhancement layer", respectively.

In [5], the CSF is modeled. However, in the DCT (Discrete Cosine Transform) domain, two dimension frequencies are needed by f_1 and f_2. Thus, the CSF can be written by

$$CS(f_1, f_2) = \frac{1}{CT_0}\exp\left(-\lambda\sqrt{f_1^2 + f_2^2}\right) \tag{6}$$

where λ and CT_0 are constant. For considering the spatial frequency only, the retinal eccentricity is regarded as zero. Using this CSF formula, the visual weight of the base layer, ϕ_b, and the enhancement layer, ϕ_e, can be obtained by

$$\begin{aligned} \phi_b &= \int_0^{f_c}\int_0^{f_c} p(f_1, f_2)\, CS(f_1, f_2)\, df_1 df_2 \\ \phi_e &= \int_{f_c}^{f_{max}}\int_{f_c}^{f_{max}} p(f_1, f_2)\, CS(f_1, f_2)\, df_1 df_2 \end{aligned} \tag{7}$$

where $p(f_1, f_2)$ is the PDF of DCT coefficients. f_c and f_{max} denote the cutoff frequency and the maximum frequency, respectively. By applying the visual weights to the empirical rate, the visual entropy is defined like below,

$$h_b^w(\omega_c, q_b) = \overline{\phi}_b r_b(\omega_b, q_b) \tag{8}$$

$$h_e^w(\omega_c, q_e) = \overline{\phi}_e r_e(\omega_e, q_e) \tag{9}$$

where $\overline{\phi}_b$ and $\overline{\phi}_e$ are the normalized values of the visual weights.

2.2. OFDM-based system

Utilizing the property of OFDM, the bandwidth is given by $BW_l = (N_{sc}^l + 1) \cdot BW_s/2$ where BW_s is the bandwidth of a sub-carrier and N_{sc}^l is the number of sub-carriers. The superscript or subscript, l, means "of the l^{th} layer". Thus, by controlling the number of sub-carriers, the bandwidth is allocated differently.

Let an MS (Mobile Station) be located in an "x" position of the i^{th} BS (Base station, i.e., the home BS). Then, the path loss between the MS of the i^{th} BS and an adjacent BS (here, the j^{th} BS) is given by $L_{(i,x:j)} = r_{(i,x:j)}^{-p} 10^{\xi_{(i,x:j)}/10}$ where p is a path-loss exponent (typically three to four), $r_{(i,x:j)}$ is the distance between x in the i^{th} BS and the j^{th} BS, $\xi_{(i,x:j)}$ is a Gaussian distributed random variable with zero mean and a standard deviation representing shadowing [6]. Therefor, the ICI (I_{oc}) and the intra cell interference (I_{sc}) are given by (here i= 0) $I_{oc} = \sum_{j=1}^{N_{oc}} S \cdot L_{(i,x:j)}$ and $I_{sc} = \theta \cdot S(1 - \varepsilon_{i,k}) \cdot L_{(i,x:i)}$ where S is the total power of each BS, $\varepsilon_{i,k}$ is the normalized power portion of the k^{th} user in the i^{th} BS and θ is orthogonality parameter between users. Thus, $(E_b/N_0)_{i,x,l}$ at the x position in the i^{th} cell over the cellular system is obtained by

$$\left(\frac{E_b}{N_0}\right)_{i,x,l} = \frac{L_{(i,x:i)} \cdot S \cdot \varepsilon_{i,k,l} \cdot \dfrac{BW_l}{R_b^l}}{\sum_{j=1}^{N_{oc}} S \cdot L_{(i,x:j)} + \theta \cdot S(1 - \varepsilon_{i,k}) \cdot L_{(i,x:i)}} \tag{10}$$

where BW_l is the bandwidth and R_b^l is the user bit rate.

Using Eq.(10), the relation between the instantaneous BER (Bit Error Rate) and E_b/N_0 for a Rayleigh fading channel and 4-QAM (Quadrature Amplitude Modulation) modulation is given by

$$P_{i,x,l} = Q\left(\sqrt{2|H|^2\left(\frac{E_b}{N_0}\right)_{i,x,l}}\right) \tag{11}$$

where the channel H is a circularly symmetric complex Gaussian random variable with a mean of 0 and a variance of 1 in [7].

3. OPTIMIZATION OF EXPECTED VISUAL ENTROPY

When the video data is transmitted over a noisy/fading channel, the distortion is not only caused by quantization errors, but also bit errors. The H.264/AVC codec uses entropy coding, so a bit error in coded bit-stream is propagated up to the end of the data, because it can make the decoder lose synchronization [7]. However, an NAL (Network Abstraction Layer) unit consists of a GOB (Group of Block) and it is byte aligned, so the corrupted synchronization by a bit error can be refreshed with the start of new GOB, fortunately.

In the former section, the empirical rate and the visual entropy were derived based on each macro-block. Thus, it can be assumed that, if a bit error occurs in a certain GOB, visual entropy can be preserved up to the previous macro-block that contains the bit error. By accumulating the expected visual entropy of each GOB obtained by the above assumption, the expected visual entropy of a frame becomes

$$\overline{h_l^w} = \sum_{n=1}^{N_l}\sum_{m=2}^{M_l}\sum_{k=1}^{m-1} h_{n,k}^w(\omega_c^l, q_l)(1-P_B^l)^{m-1}P_B^l + \sum_{n=1}^{N_l}\sum_{k=1}^{M_l} h_{n,k}^w(\omega_c^l, q_l)(1-P_B^l)^{M_l}$$

where N_l is the number of GOBs and M_l is the number of macro-blocks in a GOB. P_B^l denotes the probability that a bit error occurs in a certain macro-block, and it can be derived by

$$P_B^l = \sum_{i=1}^{\overline{r}_B^l}(1-P_{i,x,l})^{i-1}\, P_{i,x,l} \tag{12}$$

where $P_{i,x,l}$ is the BER of (11). To make the problem simpler, an average bit length of each macro-block $\overline{r}_B^l$ is used.

By some manipulations, the expected visual entropy can be represented as an objective function parameterized by ω_c^l, q_l, N_{sc}^l, and R_b^l as

$$\max_{\omega_c^l, q_l, N_{sc}^l, R_b^l} \quad \sum_{l=1}^{L} \overline{h_l^w}(\omega_c^l, q_l, N_{sc}^l, R_b^l) \tag{13}$$

$$\text{subject to} \quad \sum_{l=1}^{L} \frac{(N_{sc}^l+1)BW_s}{2} \le BW_T,\ R_b^l \ge 0$$

$$0 \le \omega_c^l \le \frac{1}{2}\pi,\ 1 \le q_l \le q_{max}$$

where L is the number of layers. The cutoff frequency has to be smaller than $\frac{1}{2}\pi$ because of a decimation by the factor of 2 and the maximum QP is 52 based on H.264/AVC. Since the problem posed above is computationally intractable, a greedy algorithm is proposed to find an approximated solution as follows:

1. Using the total sub-carriers, N_{sc}^T, find a maximum user bit rate, R_b^T, satisfying the target BER.

2. Find an optimal cutoff frequency w_c^{l*} with which the generated traffic is minimized for a given visual entropy, and find q^* by increasing the total generated traffic up to R_b^T. Set $q^* = q_{max}$.

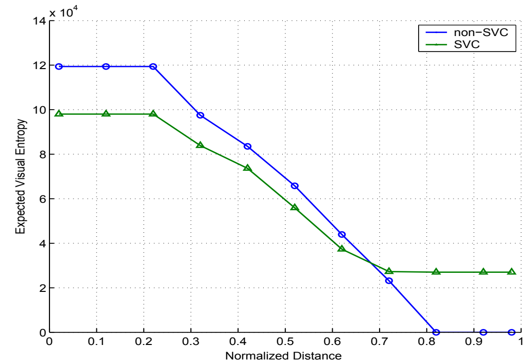

Fig. 2. Expected visual entropy through the normalized distance for a first intra-frame

3. Fixing w_c^{l*} and q^*, allocate R_b^l, and N_{sc}^l, in such a way that the base layer is guaranteed preferentially and the enhancement layers are supported by the best effort.

Figure 2 is the simulation result of this algorithm over the normalized distance. Because of the coding redundancy, the expected visual entropy of the non-scalable video is better than that of the scalable video, when the channel state is reliable. However, if the channel becomes unreliable due to ICI, there is a cross point that the scalable video is better. The reason is that, when the bit errors are occurred, the non-scalable video loses all the visual entropy at the points. However, in the scalable video case, the guaranteed base layer can be preserved at least. Thus, a hybrid scheme at the network can be developed. When the channel state is reliable, the non-scalable video is transmitted. In addition, when the channel state becomes unreliable beyond the cross-point in Fig.2, it is switched to the scalable video.

4. SIMULATION RESULTS

For the simulation, a frequency reuse factor of 1 is used, the entire bandwidth is 10 MHz and the number of sub-carriers is 1024. The hexagonal cellular pattern up to 2^{nd} tier is considered. It is assumed that each BS allocates an even portion of power and sub-carriers to all MSs. The orthogonality parameter is set to 0.05. The target BER is 10^{-6}. The empirical rate model parameters for 'Akiyo', $q = 10$, $\omega_c = 0.3\pi$ can be obtained in an intra frame as follow,

- enhancement layer : $\alpha = 30.7623, \beta = -25.6776$ when $\frac{\sigma^2}{q^2} > \frac{1}{2e}$ and $\alpha = 0.5843, \beta = 62.7377$ when $\frac{\sigma^2}{q^2} \le \frac{1}{2e}$.
- base layer : $\alpha = 25.6007, \beta = -4.5930$ when $\frac{\sigma^2}{q^2} > \frac{1}{2e}$ and $\alpha = 0.3402, \beta = 28.7058$ when $\frac{\sigma^2}{q^2} \le \frac{1}{2e}$.

The normalized visual weights of each layer are $\overline{\phi}_b = 0.98$ and $\overline{\phi}_e = 0.02$, when $\omega_c = 0.3\pi$.

In figure 3, the QP of the non-scalable video can be smaller than that of the scalable video, because of the coding redundancy. Thus, the quality of the non-scalable video is better

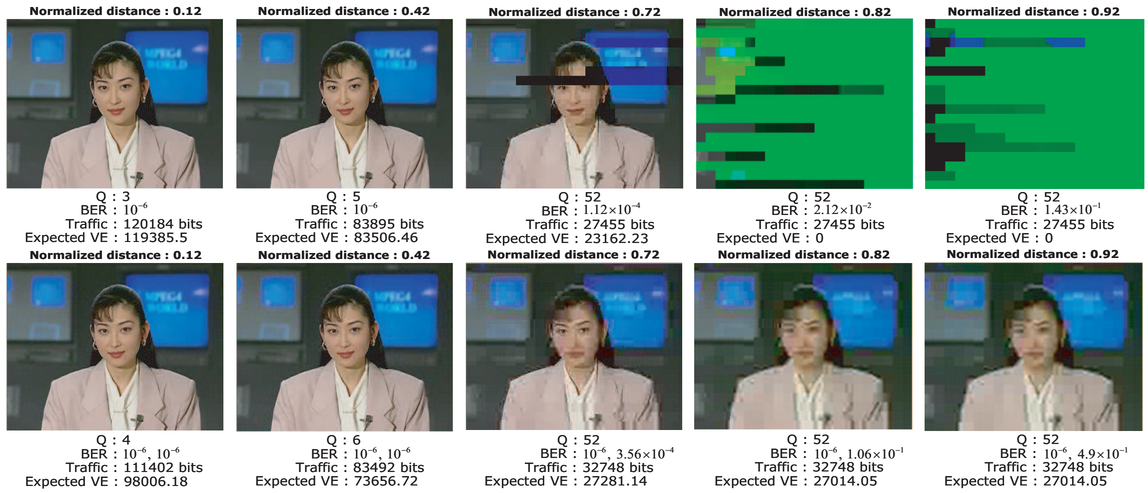

Fig. 3. The reconstructed video quality of non-layered coding (above) and layered coding (below)

when the channel states are reliable. However, when the channel states are unreliable, it can not support the whole traffic generated by using the maximum QP. In such a case, the severe distortion begins to occur. Moreover, the non-scalable video looks like occurring outage even though the generated traffic is less than that of the scalable video when the normalized distances are 0.82 and 0.92. Whereas, the scalable video is just blurred. If the coding redundancy is reduced by the inter layer prediction, the performance can be improved more.

5. CONCLUSION

In this paper, we found a cross-layer optimization approach between the SVC-based source and the OFDM-based channel by allocating sub-carriers to scalable bit-streams to maximize the expected visual entropy. For the source side, we implemented the SVC coder which is adequate to apply visual weights of the HVS, considering formerly proposed SVC techniques. In addition, visual entropy is defined in terms of the cutoff frequency and the QP. For the channel side, the OFDM-based system is modeled according to the normalized distance, in terms of the number of sub-carriers and the user bit rate. Utilizing the system parameters and considering the constraints given by the system properties, an optimization problem can be formulated. Sub-optimal values of the expected visual entropy can be found by using the greedy algorithm. Using this algorithm, it can be observed that the performance of the non-scalable video is better than that of the scalable video in the inner region (i.e., the region with high channel capacity). On the other hand, in the outer region (i.e., the region with low channel capacity), the distortion of non-scalable video is much severe. Whereas, the scalable video is just blurred. Thus, through the cross layer optimization, it was able to obtain a criterion needed to determine which the coding method is more effective for the given channel status.

6. REFERENCES

[1] M. van der Schaar, S. Krishnamachari, S. Choi and X. Xu, Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs, *IEEE J. Select. Areas in Commun.,* vol. 21, no. 10, pp. 1752-1763, Dec. 2003.

[2] Q. Zhang, W. Zhu and Y.-Q. Zhang, Channel-adaptive resource allocation for scalable video transmission over 3G wireless networs, *IEEE Trans. Circuit Syst. Video Technol,* vol. 14, pp. 1049-1063, Aug. 2004.

[3] 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 , *IEEE Int. Conf. Image Processing,* Genova, vol. 1, pp. 11-14, Sept. 2005.

[4] H. Lee and S. Lee, Compression gain measurements by using ROI-based data reduction, *IEICE Trans. FUNDAMENTALS,* vol.E89-A, no.11, pp. 2985-2989, Nov. 2006.

[5] Z. Wang and A.C. Bovik, Embedded foveation iamge coding, *IEEE Trans. Image Processing,* vol. 10, no. 10, pp.1397-1410, Oct. 2001.

[6] H. Son and S. Lee, Forward link capacity analysis for MC-CDMA, *IEICE Trans. COMMUN.,* vol. E88-B, pp. 4094-4096, June 2005.

[7] M. F. Sabir, H. R. Sheikh, R. W. Heath, Jr. and A. C. Bovic, A joint source-channel distortion model for JPEC compressed images, *IEEE Trans. on Iamge Processing,* vol. 15, no. 6, pp. 1349-1364, Jun. 2006.