HYBRIDCAST: A WIRELESS IMAGE/VIDEO SOFTCAST SCHEME USING LAYERED REPRESENTATION AND HYBRID DIGITAL-ANALOG MODULATION

Zhihai Song, *Ruiqin Xiong*, *Siwei Ma*, *Wen Gao*

Institute of Digital Media, School of EE&CS, Peking University, Beijing 100871, China
Email: {zhsong, rqxiong, swma, wgao}@pku.edu.cn

ABSTRACT

The recently proposed SoftCast scheme employs analog-like transmission for wireless visual communication, providing graceful reconstruction quality degradation for drastically changing channel conditions. However, the transmission in SoftCast is not always efficient in terms of power usage. In this paper, we propose a wireless image/video SoftCast scheme which employs layered representation with hybrid digital-analog modulation. In this scheme, a coarse approximation of the image is coded in a base layer in digital ways, while the residual image details are delivered in an enhancement layer in analog-like way. The outputs from the two layers are superimposed for transmission, using a hybrid digital-analog modulation scheme. Since a major part of the signal is handled by the base layer, the power efficiency of the SoftCast layer is significantly improved. Experimental results show that the proposed scheme outperforms the original SoftCast remarkably, while still preserving the smooth quality degradation characteristic of the SoftCast scheme.

1. INTRODUCTION

In conventional visual communication system, the coding bit rate is usually chosen before the actual transmission process occurs, based on an estimation of channel capacity. If the actual channel capacity drops below a certain threshold, the decoding process tends to break down completely. In wireless scenarios, channel conditions may vary drastically and unpredictably. This brings big challenges to the design of visual communication systems. To make the things worse, in wireless broadcast scenarios, different clients may have different channel conditions. It is difficult to fully utilize the channel resources of each client to provide the best possible video qualities.

Various scalable coding schemes [1][2] have been proposed to handle the above problems. In scalable coding, a visual signal is typically coded into an embedded stream, which can be safely truncated at the time of transmission, in order to make the bit rate match the channel capacity. Hierarchical modulation [6] was also proposed to superimpose the bits from multiple layers into one wireless symbol and allow the clients to decode different numbers of layers according to their channel conditions. However, these schemes usually provide very limited choices of layers and qualities [7].

The recently emerged SoftCast scheme [3]-[5] proposes an analog-like transmission framework, in which the encoding process is simply a series of linear transform, leaving out the conventional quantization and entropy coding. Instead of producing a stream of binary symbols, SoftCast generates a stream of real number coefficients from which exact reconstruction is possible. The coefficients are directly modulated to a dense constellation (e.g. 64k-QAM) for OFDM transmission. The transmission is lossy in nature and the noise level in the received numbers is commensurate with the channel signal-to-noise ratio (CSNR). The most prominent advantage of SoftCast is that it provides graceful quality transition in very wide CSNR range and can serve various clients of different channel conditions simultaneously, using the same transmitted signal in the air. For this reason, SoftCast has attracted much research attention in recent years [8]-[18].

However, the analog-like transmission in SoftCast is not always efficient in terms of power usage. The power distortion functions of digital and analog transmissions indicate that both methods have their own advantages. This motivates us to consider combining the two approaches and exploit both of their advantages.

This paper proposes a wireless image/video SoftCast scheme, using layered representation and hybrid digital-analog modulation for transmission. In this scheme (we call HybridCast), a coarse approximation of the image is coded in a base layer in digital ways, while the residual image details are delivered in an enhancement layer in analog-like way. Since a major part of the signal is handled by the base layer, the power efficiency of the enhancement layer is significantly improved. The outputs from the two layers are superimposed for transmission, using a hybrid digital-analog modulation scheme. To achieve optimal overall transmission performance and guarantee the decoding of hybrid modulated signals, a key point is to allocate transmission power between the digital base layer and the analog-like enhancement layer. To facilitate the power allocation and transmission, a special whitening scheme is also proposed.

The paper is organized as follows. Section 2 briefly reviews the SoftCast scheme and explains the motivation of

This work was supported in part by the National Natural Science Foundation of China (61370114, 61121002, 61073083), Beijing Natural Science Foundation (4132039, 4112026) and Research Fund for the Doctoral Program of Higher Education (20120001110090, 20100001120027).

the work. Section 3 describes the proposed HybridCast scheme. Section 4 discusses the power allocation problem. Section 5 presents experimental results and Section 6 concludes the paper.

2. BACKGROUND

2.1. SoftCast Review

Fig. 1 shows the framework of SoftCast. The compression stage is solely a transform to decorrelate the image signal, producing a stream of transform coefficients. The transmission stage scales each coefficient individually and applies a Walsh-Hardmard Transform (WHT) to whiten the whole stream. Finally the resulted real numbers are modulated directly to a dense 64k-QAM modulation for OFDM transmission. The SoftCast decoder inverses the modulation and transforms at encoder, and recovers the original image using a linear least square estimator (LLSE).

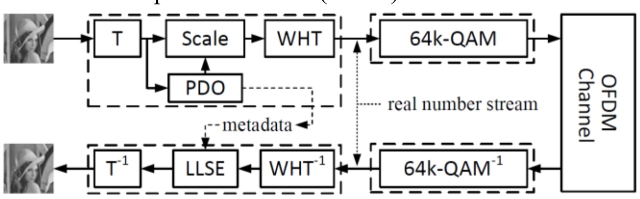

Fig. 1. Diagram of the SoftCast system [3-5].

2.2. Transmission Efficiency Analysis

The purpose of a visual communication system is to reproduce the image signals with maximal quality at the receiver side, subject to the transmission power and channel bandwidth constraints. Given a channel bandwidth condition, the transmission efficiency of such communication system can be measured by its distortion-power characteristics.

Conventional communication systems compress the image signals into a bit stream, and send out the bits using channel coding and modulation techniques. Usually, the power consumption for transmitting each bit can be regarded as constant. Therefore, compression the image signal to a higher bit rate corresponds to a higher power consumption in transmission. Consider a typical rate-distortion model

$$D(R) = f(\sigma_I^2) \cdot 2^{-2R} \tag{1}$$

where σ_I^2 is the variance of the signal, we know that by encoding one more bit, the distortion diminishes very quickly. Therefore the power usage is very efficient at low bit rate.

For the analog-like transmission in SoftCast, however, the efficiency of power usage is different. Suppose $\mathbb{x} = (x_1, x_2, x_3, \ldots, x_N) \in \mathbb{R}^N$ is a random signal to transmit over channel. To utilize the transmission energy efficiently, SoftCast scales each x_i by $g_i \in \mathbb{R}^N$ and sends out $y_i = g_i \cdot x_i$. The signal that arrives at the receiver is $\hat{y}_i = y_i + n_i$, where n_i is the channel noise. The receiver inverses the scaling operation and gets an estimation $\hat{x}_i = \hat{y}_i / g_i = x_i + n_i / g_i$. In this process, the expected distortion is

$$D_i = E[(\hat{x}_i - x_i)^2] = \frac{\sigma_n^2}{g_i^2}, \tag{2}$$

while the transmission power for sending x_i is

$$P_i = g_i^2 \cdot E[{x_i}^2]. \tag{3}$$

To achieve optimal performance, SoftCast determines the scaling factors using power-distortion optimization (PDO). The performance with PDO is (see [13][14] for details)

$$D_{total} = \frac{\sigma_n^2}{P_{total}} \cdot \left(\sum_i \sqrt{E[{x_i}^2]} \right)^2. \tag{4}$$

From (4) we see that the distortion diminishes with transmission power slowly, as $D \propto P^{-1}$.

The above analysis motivates us to consider combining digital transmission and analog-like transmission. We try to encode in a low bit-rate base layer a coarse version of the signal, using an existing coding scheme, and transmit the residue image in the way of SoftCast. Since a major part of the image energy is solved by coded transmission, the energy in the residual signal is greatly reduced, leading to significant improvement in ultimate performance.

3. SYSTEM OVERVIEW OF HYRIBDCAST

Basically, we want to resolve most of the image energy by encoding a low bit rate version of the image, leaving the residual to the analog transmission. In practice, we down sample the image, and encode it using large quantization step into the base layer. Then, the residual between the original image and the base layer reconstruction is processed and transmitted as the enhancement layer.

The most straight forward ways of transmitting the two layers are transmitting them in digital and analog ways separately. However, the bandwidth is usually limited. Introducing a base layer requires extra bandwidth for transmitting the bit stream in it. To avoid dropping enhancement layer coefficients, digital and analog signal have to share the bandwidth. In our scheme, part of the wireless symbols adapts hybrid modulation. As shown in Fig. 2, on a wireless symbol, each component (i.e. I-, Q- component) consists a digital signal s_d, with an analog signal s_a superimposed onto it.

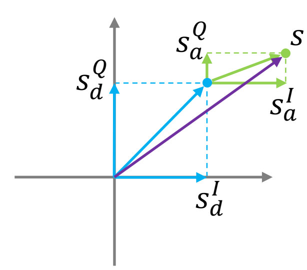

Fig. 2. Hybrid modulation

An important point to note is that, for the decoding of digital signal, the analog signal is considered as noise.

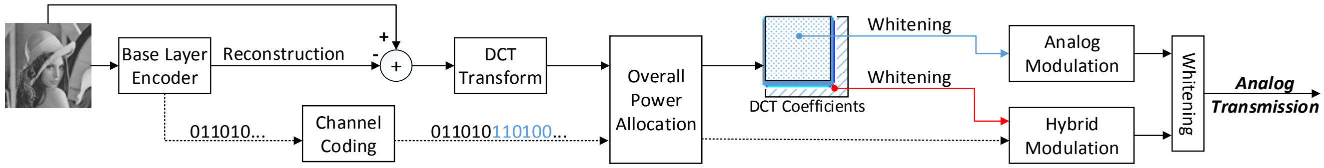

Fig. 3. Framework of the HybridCast scheme.

A very large noise may change the digital signal to a nearby constellation point so that the signal cannot be decoded correctly. Therefore, we only select the high frequency coefficients which are expected to be very small for hybrid modulation. However, impulse values may still reside in high frequency coefficients, ruining the digital signal that it is superimposed on. Therefore, before the hybrid modulation, whitening is performed on the selected coefficients to smooth away the impulsive values.

Fig. 3 depicts the framework of the proposed scheme. Firstly, a base layer bit stream is generated and channel coded into $\mathbb{S}_d$. Then, the analog encoder takes the residual between the base layer reconstruction and the original image as input. The residual is first DCT transformed into $\mathbb{X}$, scaled by overall power allocation factors $\mathbb{G}$ into $\mathbb{S}_a$. Then, we select part of the coefficients in $\mathbb{S}_a$ that correspond to the high frequency sub-bands to do hybrid modulation. To be concise, we denote the hybrid modulated coefficients as $\mathbb{S}_a^h$, the others as $\mathbb{S}_a^r$. Whitening is performed on $\mathbb{S}_a^h$ and $\mathbb{S}_a^r$ separately. In hybrid modulation, $\mathbb{S}_a^h$ and $\mathbb{S}_d$ are modulated together.

After the hybrid modulation, part of the signal is hybrid and the rest is analog. Due to overall power allocation, the amplitude of the two kinds of signals may vary drastically, which is unsuitable to be transmitted in physical devices for their requirements about peak-to-average energy ratio. Therefore, another whitening operation is performed on all the signals, including the hybrid signal and the pure analog signal. Finally, the whitened signals are modulated in analog way and transmitted.

4. RESOURCES ALLOCATION

In this section, we are going to discuss about the overall power allocation, the ratio of the analog power budget and total power budget, $\beta = P_a/P_{total}$, and the chosen of the base layer bit rate R.

The overall power allocation have to meet two requirements. 1) Digital signal should be allocated with enough energy to fight the noise and the superimposed analog signal. 2) The analog power should be optimized according to the power distortion relation described in (4).

To satisfy 2), as discussed in [3][13], when

$$g_i = E[\mathbb{X}_R[i]^2]^{-1/4} \cdot (P_a/\sum\sqrt{E[\mathbb{X}_R[i]^2]})^{-1/2} \qquad (5)$$

the analog part is optimized. And $P_a = \sum E[\mathbb{S}_a(i)^2]$ is the total power used in the analog part, where $\mathbb{S}_a(i) = g_i \cdot \mathbb{X}_R[i]$.

To guarantee the correct decoding of the digital part, the energy of digital signal and analog signal should satisfy

$$A^2 \geq \alpha(\sigma_n^2 + \frac{1}{|\mathbb{I}_h|}\sum_{i\in\mathbb{I}_h} E[\mathbb{S}_a(i)^2]) \qquad (6)$$

where, A is the amplitude of the digital signal, α is the threshold for correct digital decoding, σ_n^2 is the variance of channel noise, $\mathbb{I}_h$ is the index set of hybrid modulated coefficients in $\mathbb{S}_a$. Obviously, we have $P_{total} = |\mathbb{I}_h|A^2 + P_a$. Then, (6) can be written as

$$P_{total}(1-\beta) = \alpha(|\mathbb{I}_h|\sigma_n^2 + \beta\sum_{i\in\mathbb{I}_h} E[g_i'^2\mathbb{X}_R[i]^2]) \qquad (7)$$

where $g_i' = E[\mathbb{X}_R[i]^2]^{-1/4} \cdot (P_{total}/\sum\sqrt{E[\mathbb{X}_R[i]^2]})^{-1/2}$. A^2 is set as minimum in (6) to achieve the largest working CSNR range. Obviously, every base layer bit will be hybrid-modulated with two coefficients, so $|\mathbb{I}_h| = 2NR$, N is the pixel number. Then, we have

$$\beta(R) = \frac{P_a}{P_{total}} = \frac{P_{total} - 2\alpha NR\sigma_n^2}{P_{total} + \alpha\sum_{i\in\mathbb{I}_h} E[g_i'^2\mathbb{X}_R[i]^2]} \qquad (8)$$

According to (2)(5)(8), the total distortion is

$$D_{total}(R) = \frac{\sigma_n^2 \cdot \left(\sum\sqrt{E[\mathbb{X}_R[i]^2]}\right)^2}{\beta(R)\cdot P_{total}}. \qquad (9)$$

Thus, in practice the bit rate of the base layer could be optimally chosen by (9). As R increases from zero, system changes from pure analog to hybrid, high efficiency part of digital transmission is taken in, and the performance will increase accordingly. But after a certain point, digital part is switching to its low efficiency stage, and the performance starts to drop.

5. EXPERIMENTAL RESULTS

In this section, we conduct some experiments to evaluate the performance of the proposed scheme in comparison with the original SoftCast. Experiments adopt the test image set of *Lena, Peppers, Elaine, Fishingboat* (512×512, gray).

The base layer bit rate is controlled by the down sample rate r and H.264 Quantization Parameter (QP). In experiment, we set $r = 2$. We first evaluate the influence of the base layer bit rate. Fig. 4 gives the system performance at different base layer rate, when CSNR = 10dB. So, the base layer bit could be optimally chosen at the sender end.

Performance comparisons are concluded in Fig. 5. Direct transmission of pixel value using analog modulation is added as comparison. Fig. 6 gives Objective comparison.

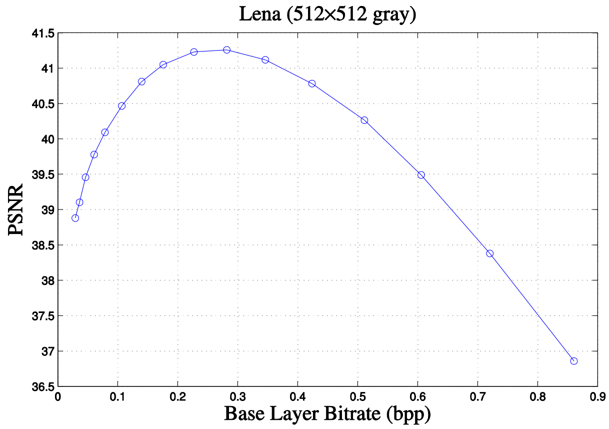

Fig. 4. The influence of choosing different base layer bit rates.

6. CONCLUSIONS AND DISCUSSIONS

Digital and analog transmissions have their own advantages in power utilization efficiency. This paper tries to integrate the high efficiency part of the both systems, and proposes to extract and encode part of visual signal and digitally, while the rest is in analog way. In transmission, hybrid modulation is employed to transmit digital and analog signal at the same time. Thus, the majority of the visual signal distortion is resolved in digital approach, which utilized the advantage of digital transmission. In the meantime, the efficiency of power usage in the enhancement layer is greatly improved. Compared to the original SoftCast, the proposed scheme is more efficiency in power utilization, while still preserves the smooth quality degradation ability. Experimental result shows that the proposed scheme can improve the reconstruction quality remarkably both objectively and subjectively.

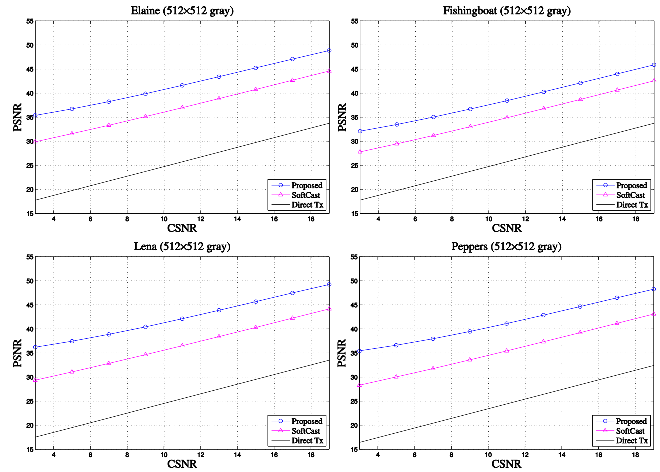

Fig. 5. Performance comparison between the proposed approach and the original SoftCast scheme.

Fig. 6. Reconstructed images by the original SoftCast and the proposed scheme at CSNR = 3dB. The columns from left to right: the original SoftCast and the proposed scheme.

7. REFERENCES

[1] F. Wu, S. Li, and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 3, pp. 332 –344, Mar 2001.

[2] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the h.264/avc standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103 –1120, sept. 2007.

[3] S. Jakubczak, H. Rahul, and D. Katabi, “Softcast: One video to serve all wireless receivers,” in MIT Technical Report, MIT-CSAIL-TR-2009-005, 2009.

[4] S. Jakubczak, H. Rahul, and D. Katabi,, “One-size-fits-all wireless video,” in Proc. Eigth ACM SIGCOMM HotNets Workshop, New York, NY, USA, 2009.

[5] S. Jakubczak and D. Katabi, “A cross-layer design for scalable mobile video,” in Proc. international conference on Mobile computing and networking (MobiCom ’11) , pp. 289–300, New York, NY, USA, 2011.

[6] K. Ramchandran, A. Ortega, K. Uz, and M. Vetterli, “Multiresolution broadcast for digital hdtv using joint source-channel coding,” in IEEE International Conference on Communications, vol.1, pp. 556 –560, June 1992.

[7] “Digital Video Broadcasting (DVB),” Website, 2009, http://www.etsi.org/deliver/etsien/300700300799/300744/01.06.01 60/en 300744v010601p.pdf.

[8] Lei Yu, Houqiang Li, Weiping Li, “Hybrid Digital-Analog Scheme for Video Transmission over Wireless,” in IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1163–1166, May 2013.

[9] Xiaopeng Fan, Feng Wu, Debin Zhao, C. Oscar Au, and Wen Gao, “Disributed soft video broadcast (dcast) with explicit motion,” in IEEE Data Compression Conference (DCC), pp. 199– 208, 2012.

[10] Xiaolin Liu, Wenjun Hu, Qifan Pu, Feng Wu, and Yongguang Zhang, “Parcast: soft video delivery in mimo-ofdm wlans,” in Internation Conference on Mobile Computing and Networking (Mobicom), 2012, pp. 233–244.

[11] Xiaopeng Fan, Ruiqin Xiong, Feng Wu, and Debin Zhao, “Wavecast: Wavelet based wireless video broad-cast using lossy transmission,” in IEEE Visual Communications and Image Processing (VCIP), 2012, pp. 1–6.

[12] Ruiqin Xiong, Xiaopeng Fan, Feng Wu, and Wen Gao, “Analysis of uncoded image communication over noisy channels with unequal protection,” in SPIE Electronic Imaging 2013, Visual Information Processing and Communication IV, Feb. 2013, p. 866603.

[13] Ruiqin Xiong, Feng Wu, Jizheng Xu, and Wen Gao, “Performance analysis of transform in uncoded wireless visual communication,” in IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1159–1162, May 2013.

[14] Ruiqin Xiong, Feng Wu, Xiaopeng Fan, Chong Luo, Siwei Ma, and Wen Gao, “Power-distortion optimization for wireless image/video softcast by transform coefficients energy modeling with adaptive chunk division,” in IEEE Visual Communications and Image Processing (VCIP), 2013, pp. 1–6.

[15] Ruiqin Xiong, Hangfan Liu, Siwei Ma, Xiaopeng Fan, and Wen Gao, “G-cast: Gradient based image softcast for perception friendly wireless visual communication,” in IEEE Data Compression Conference (DCC), pp. 133–142, Snowbird, Utah, USA, Mar. 2014.

[16] Zhihai Song, Ruiqin Xiong, Xiaopeng Fan, Siwei Ma, and Wen Gao, “Transform domain energy modeling of natural images for wireless softcast optimization,” in IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1114–1117,Melbourne, Australia, June 2014.

[17] Zhihai Song, Ruiqin Xiong, Siwei Ma, Xiaopeng Fan, and Wen Gao, “Layered image/video softcast with hybrid digital analog transmission for robust wireless visual communication,” in IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, Jul. 2014.

[18] Hangfan Liu, Ruiqin Xiong, Siwei Ma, Xiaopeng Fan, and Wen Gao, “Gradient based image transmission and reconstruction using non-local gradient sparsity regularization,” in IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, Jul. 2014.

[19] Koichi Takagi, Yasuhiro Takishima and Yasuyuki Nakajima, “A Study on Rate Distortion Optimization Scheme for JVT Coder”, in Proc. SPIE, vol. 5150, 2003, pp. 914–923.