LAYERED IMAGE/VIDEO SOFTCAST WITH HYBRID DIGITAL-ANALOG TRANSMISSION FOR ROBUST WIRELESS VISUAL COMMUNICATION

Zhihai Song[1], Ruiqin Xiong[1], Siwei Ma[1], Xiaopeng Fan[2], Wen Gao[1]

[1] Institute of Digital Media, Peking University, Beijing 100871, China
[2] Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China
Email: {zhsong, rqxiong, swma, wgao}@pku.edu.cn, fxp@hit.edu.cn

ABSTRACT

Due to the mobility of transceivers and the interference from other signals, the condition of a wireless channel may vary drastically and unpredictably. In such scenario, conventional communication systems usually suffer from cliff effect due to the nature of entropy coding and channel coding. The recently proposed SoftCast scheme, on the contrary, achieves smooth quality degradation by employing analog-like transmission, together with efficient decorrelation and power allocation. However, the analog-like transmission in SoftCast is not always efficient in terms of power usage, when compared with digital approaches. In this paper, we propose a layered image/video SoftCast scheme, in which a coarse approximation of the image is coded in a base layer and transmitted in digital way while the remained image details are represented in an enhancement layer and sent out using the SoftCast way. Since a major part of the signal energy is transmitted in the base layer using digital approach, the efficiency of power usage in the enhancement layer is improved remarkably. Experimental results show that the proposed scheme can outperform the original SoftCast scheme remarkably, while still preserving the smooth quality degradation characteristic of SoftCast.

Index Terms— Wireless communication, SoftCast

1. INTRODUCTION

In conventional communication systems, channel estimation is needed for the encoder to choose a proper coding bit rate. If the channel capacity drops below a certain threshold, the decoding process tends to break down completely. If the channel quality increases beyond that threshold, on the other hand, it cannot provide any improvement in performance. This is known as the *cliff effect*. Due to antenna mobility and occlusion, the quality of wireless channel is usually unpredictable. This brings huge challenges to communication systems in providing reliable performance. Besides, in wireless broadcast scenarios, it is difficult to fully utilize the channel resources of different clients and provide the best video quality matching their individual channel conditions.

To overcome the above shortcomings of traditional schemes, scalable coding schemes [1, 2] have been proposed. In these schemes, visual signal is usually coded into one base layer and several enhancement layers. In transmission, the hierarchical modulation [3] superimposes bits of multiple layers into one wireless symbol and allows the clients to decode different numbers of layers according to their own channel conditions. However, these schemes usually provide limited choices of base layer and enhancement layer rates. DVB-T [4] standard specifies 3 base layer rates and 5 enhancement layer rates, and cliff effect problem is still inevitable.

The recently emerged SoftCast scheme [5–7] provides a soft transmission framework. The SoftCast encoding process is simply a series of linear transform on the visual signal, leaving out the quantization and entropy coding steps. Instead of yielding bit streams, SoftCast generates a stream of real number coefficients from which exact reconstruction is possible. SoftCast also abandons the channel coding. The coefficients are directly modulated into a dense constellation (e.g. 64k-QAM) for OFDM transmission. The transmission in SoftCast 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 [8–14].

However, SoftCast is not always efficient in channel resources utilization. Not all components of the visual signal are of equal efficiency to be transmitted via analog way in SoftCast. Maybe some components of the visual signal are more suitable to be transmitted in SoftCast like way, while the other components may not. On the other hand, digital transmission may be of higher transmission efficiency in transmitting some components of the signal, but not so efficient in transmitting other components. This motivates us to consider integrating the high efficiency part of both systems.

In this paper, based upon the above consideration, we propose a layered SoftCast scheme. In this scheme, image is coded into one base layer and one enhancement layer. The base layer is a low quality version of the original image that is encoded and transmitted in digital ways. The residual between the base layer and the original image is then encoded and transmitted in the analog way like SoftCast.

The paper is organized as follows. Section 2 gives brief review of the SoftCast scheme. Section 3 describes the proposed layered SoftCast scheme. Section 4 presents the experimental results. Finally, Section 5 concludes this paper.

2. REVIEW OF SOFTCAST

Fig. 1 shows the digital communication system. Input image goes through the process of transform, quantization and entropy coding, becoming a stream of bits. Every bit is crucial to the final reconstruction and has to be transmitted correctly to the receiver. Usually, channel coding is performed to provide protection to the bit stream.

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).

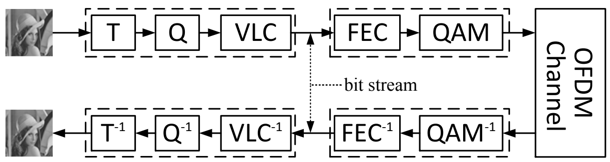

Fig. 1. The framework of conventional communication systems.

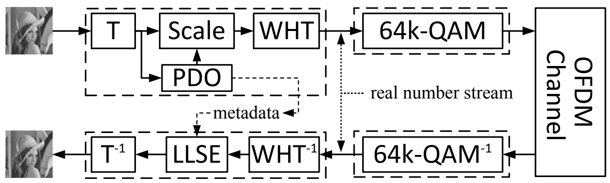

Fig. 2. The framework of SoftCast transmission [11].

Then, the channel coded bits are mapped to a constellation using quadrature amplitude modulation (QAM) (e.g. BPSK, QPSK, 16-QAM, 64-QAM) for OFDM transmission.

In SoftCast, as shown in Fig. 2, the compression stage is solely a transform to decorrelate the image signal, producing a stream of transform coefficient numbers. The transmission stage scales each coefficients individually, applies a Walsh-Hadamard Transform (WHT) to whiten the whole stream. Finally the resulted real numbers are modulated directly to a dense modulation (e.g. 64k-QAM) for OFDM transmission. A pair of real numbers are extracted from the stream each time and mapped to a point in the dense QAM constellation, using the two numbers as the I- and the Q- components, respectively. The scaling factors are determined by a power-distortion optimization (PDO) procedure, and will be communicated to the receiver side via a limited number of metadata.

3. LAYERED SOFTCAST

3.1. Efficiency of Digital/Analog Transmission

Communicating visual signal from the sender to the receiver is a process of diminishing the distortion of the reconstructed signal at the receiver side, or in another word, the uncertainty of the signal. The efficiency of transmission can be measured by how much contribution each unit of transmission power makes to the diminishing of distortion. Therefore, high efficiency transmission means diminishing more distortion with less transmission power. Usually the average power consumption of transmitting each bit is fixed on a transmission device. Thus, high efficiency transmission also means diminishing more distortion with fewer bits.

In digital image transmission, most significant bit plane (MSBP) contains much more visual signal energy than other bit planes, or to say, transmitting the MSBP resolves most of the distortion. The power used in transmitting the MSBP is of higher efficiency than in other bit planes. As the transmission proceeds to the least significant bit plane (LSBP), the power efficiency decreases. According to [15], generally, the power distortion relationship can be described as

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

where σ_I^2 is the variance of the signal. Thus, we know that by sending one more bit plane the distortion is diminished by three quarters.

In analog image transmission, noise is the distortion added by the channel. Consider a random vector to be transmitted over channel, $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^N$. Typically, each x_i may represent a single pixel or coefficient. To fully utilize the transmission energy that the system provides, x_i is scaled 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

$$\hat{y_i} = y_i + n_i = g_i \cdot x_i + n_i \tag{2}$$

where n_i is the channel noise, commonly assumed to be additive-white-Gaussian-noise (AWGN). The receiver inverses the scaling operation and gets an estimation of x_i by:

$$\hat{x_i} = \hat{y_i}/g_i = x_i + n_i/g_i \tag{3}$$

After the above process, the distortion is expected to be

$$D_i = E[(\hat{x_i} - x_i)^2] = \sigma_n^2/g_i^2 \tag{4}$$

where σ_n^2 is the variance of the channel noise. The transmission power expectation for sending x_i is

$$P_i = g_i^2 \cdot E[x_i^2] \tag{5}$$

Combining (4) and (5), we get the power-distortion relation of analog transmission

$$D_i(P_i) = \frac{1}{P_i}\sigma_n^2 \cdot E[x_i^2] \tag{6}$$

From (6) we can conclude that when the transmission power is fixed, signal with smaller variance is less impacted by the channel with distortion.

Based on the above analysis, we know that not all components of visual signals are of equal efficiency when transmitted in digital way, nor in analog way. Both digital and analog transmissions have their own relatively high efficiency parts. This makes us to consider integrating the high efficiency parts of both systems, in order to achieve a higher efficiency transmission system.

3.2. Hybrid Design with Layered Coding

The above considerations motivate us to consider extracting and transmitting part of the image signal in digital way, in purpose of resolving most of the distortion with less transmission power and achieving the similar high efficiency of transmitting MSBP in digital transmission. The rest of the signal is of relatively smaller variance that is less impacted by the channel with distortion. Usually, the MSBP contains the low frequency information. Therefore, in practice we extract the low frequency component as the base layer to transmit in digital way, and the enhancement layer is the residual between the original image and the base layer reconstruction image, which is transmitted in analog way.

Fig. 3 shows the diagram of the proposed scheme. The input image is first down sampled to $1/2^{2s}, (s = 0, 1, 2, ..., \log l^{\frac{1}{2}})$ of the original dimension l, then encoded by H.264 as the base layer. In order to adapt to a large CSNR range, base layer part is transmitted by the most reliable channel rate (i.e. BPSK and 1/2 FEC code). The residual image is transformed into DCT coefficients $\mathbb{X}$ and scaled to provide unequal noise protection like SoftCast. At the receiver side the process is inversed. The noisy residual is added on the reconstructed base layer image, forming the final reconstruction image.

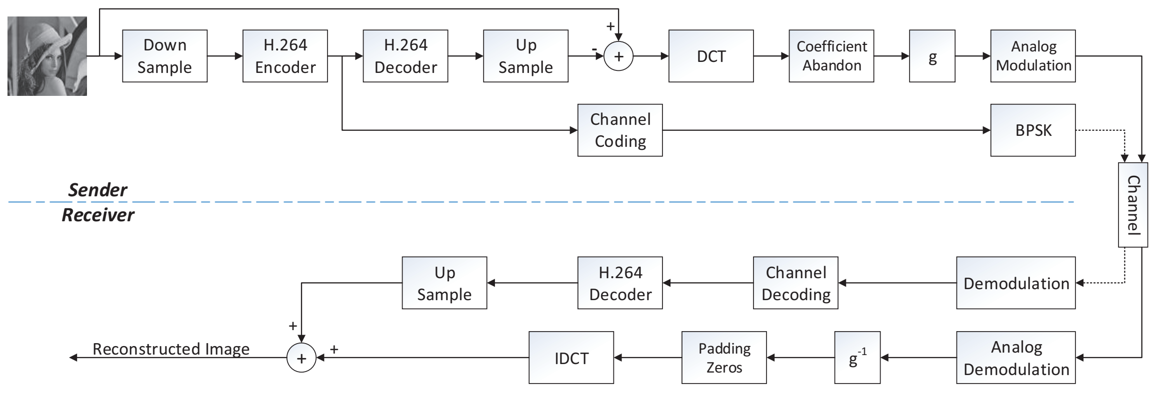

Fig. 3. The proposed layered SoftCast scheme with hybrid digital-analog transmission.

3.3. Overall Performance

Usually, bandwidth is limited in practical communication systems. Therefore, enhancement layer has to discard a certain portion of the coefficients to reserve some bandwidth for the base layer transmission. And the base layer bit rate determines how many coefficients have to be discarded. Higher base layer rate means discarding more residual coefficients, but it also means more image energy is resolved in digital transmission, and the residual is of less energy. So, the base layer rate choice is a tradeoff problem of the transmission resource allocation between the base layer and the enhancement layer. In fact, the overall distortion can be known at the sender side, which can guide the base layer rate choice. The following of this subsection will discuss the overall distortion.

Equation (1) describes the relation of base layer bit rate R and its distortion, which actually equals to the signal energy $\mathbb{E}$ left in the residual

$$\mathbb{E}(R) = f(\sigma_I^2) \cdot 2^{-2R} = \sum_i \mathbb{X}_R[i]^2 \tag{7}$$

Assume the total bandwidth is N, the pixel number of the original image. Due to channel rate of base layer transmission (i.e. BPSK modulation and 1/2 channel coding), each bit costs 2 constellation symbols, so $4R$ coefficients of the residual have to be discarded. Usually, to minimize the effect that the discarding operation brings, high frequency coefficients of the DCT plane are discarded first. To be convenient, we denote the index set of the discarded coefficients as $\mathbb{I}_a, |\mathbb{I}_a| = 4R$, the index set of the transmitted coefficient as $\mathbb{I}_t, |\mathbb{I}_t| = N - 4R$. To be concise, we define $\beta = |\mathbb{I}_a|/N$ to identify the discard ratio. Then, the discarded part with the index of $\mathbb{I}_a$ causes the distortion of

$$D_a = \sum_{i \in \mathbb{I}_a} \mathbb{X}_R[i]^2. \tag{8}$$

According to SoftCast, g_i is chosen as

$$g_i = E[\mathbb{X}_R[i]^2]^{-1/4} \cdot \sqrt{\frac{P_0 \cdot |\mathbb{I}_t|}{\sum_{i \in \mathbb{I}_t} \sqrt{E[\mathbb{X}_R[i]^2]}}} \tag{9}$$

where P_0 is the average analog transmission power. The choice of g_i is a power-distortion optimization problem, which is deduced in detail in [5, 12].

Combining (4) and (9), the distortion introduced by channel noise when transmitting coefficients with the index of $\mathbb{I}_t$ is

$$D_t = \frac{\sigma_n^2}{P_0 \cdot |\mathbb{I}_t|} (\sum_{i \in \mathbb{I}_t} \sqrt{E[\mathbb{X}_R[i]^2]})^2 \tag{10}$$

Thus, the total expected distortion is

$$D_{\text{total}} = \sum_{i \in \mathbb{I}_a} \mathbb{X}_R[i]^2 + \frac{\sigma_n^2}{P_0 \cdot |\mathbb{I}_t|} (\sum_{i \in \mathbb{I}_t} \sqrt{E[\mathbb{X}_R[i]^2]})^2 \tag{11}$$

which can be known at the encoder side. According to (11), R can be optimally chosen, at the sender side.

4. EXPERIMENTAL RESULTS

In this section, we conduct some experiments to evaluate the performance of the proposed scheme in comparison with the original SoftCast. In our experiment, AWGN is added onto the analog signal of the coefficients directly to simulate the behavior of channel. For the digital part, since we target our system on the CSNR range of 3dB+, we use BPSK and 1/2 convolutional code for the base layer, expecting a very high probability of correct decoding. So, we assume the base layer bit stream can be transmitted correctly.

The parameters that control R are s and Quantization Parameter (QP) of the H.264 encoder. The system is evaluated by the performance gain in the reconstruction quality. We first evaluate the influence of s.

As Fig. 4 shows, when s is small, R is relatively higher, and the base layer resolves more signal energy. More coefficients of the residual are dropped, which causes saturation at the high end of the curve. The saturation also happens when QP is set to a smaller value and R is high. Fig. 5 shows the effect of QP.

The base layer rate can vary in a large range. When R is near zero, little energy of image signal is resolved in base layer. The residual is near original image, and the system approximates SoftCast. When R is high enough to cover the entire bandwidth, no signal is transmitted in analog way. The system turns into purely digital transmission. Then, the performance curve turns into a horizontal line.

In practice, QP and s can be chosen according to (11), at a certain CSNR. In this paper, we set fixed parameters, $QP = 28$, s=4 and

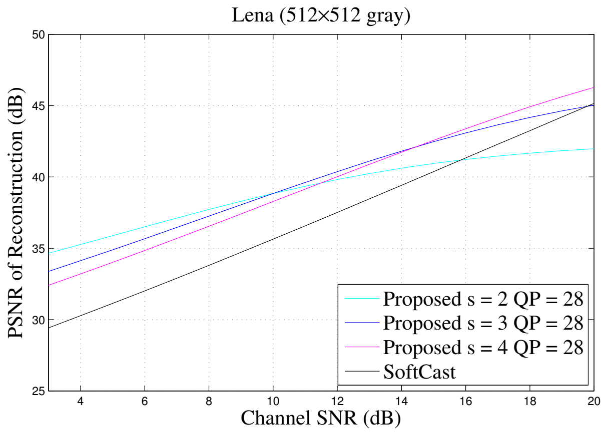

Fig. 4. The performance of different parameter s

QP = 51, s=2, which can handle a large CSNR range. More results are concluded in Fig. 6. Direct transmission of pixel value using analog modulation is added as comparison. As it shows, the proposed system preserves the character of smooth degradation, which make it suitable for unstable wireless channel. Fig. 7 gives subjective comparison.

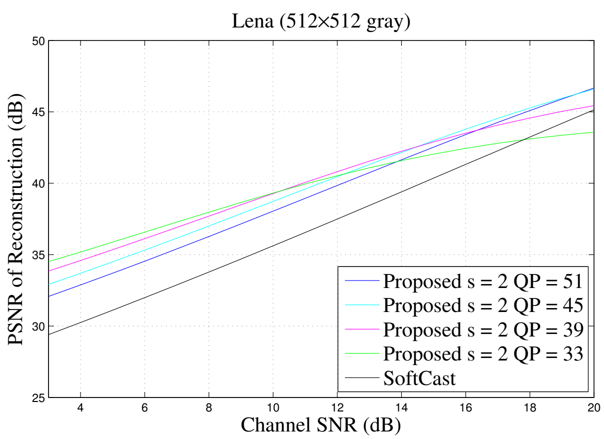

Fig. 5. The performance of different parameter QP

5. CONCLUSIONS AND DISCUSSIONS

Digital and analog transmissions have their own advantages in power utilization efficiency. Different components of the visual signal are more suitable to be transmitted in either digital way or analog way. This paper proposes to extract and transmit part of visual signal digitally, while the rest is transmitted in analog way. Thus, the majority of the visual signal energy 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.

6. 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] K. Ramchandran, A. Ortega, K.M. Uz, and M. Vetterli, "Multiresolution broadcast for digital hdtv using joint source-channel coding," in *IEEE International Conference on Communications (ICC'92)*, Jun 1992, pp. 556–560 vol.1.

[4] "Digital video broadcasting (dvb)," `http://www.etsi.org/deliver/etsien/300700300799/300744/01.06.0160/en300744v010601p.pdf`.

[5] Szymon Jakubczak, Hariharan Rahul, and Dina Katabi, "Softcast: One video to serve all wireless receivers," in *MIT Technical Report, MIT-CSAIL-TR-2009-005*, 2009.

[6] Szymon Jakubczak and Dina Katabi, "Softcast: one-size-fits-all wireless video," in *ACM SIGCOMM 2010 conference*, New York, NY, USA, 2010, SIGCOMM '10, pp. 449–450, ACM.

[7] Szymon Jakubczak and Dina Katabi, "A cross-layer design for scalable mobile video," in *International conference on Mobile computing and networking (MobiCom '11)*, New York, NY, USA, 2011, pp. 289–300.

[8] X. Fan, F. Wu, D. Zhao, O.C. Au, and W. Gao, "Distributed soft video broadcast (dcast) with explicit motion," in *IEEE Data Compression Conference*, Apr. 2012, pp. 199–208.

[9] X. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "Parcast: soft video delivery in mimo-ofdm wlans," in *Internation Conference on Mobile Computing and Networking (Mobicom)*, 2012, pp. 233–244.

[10] X. Fan, R. Xiong, F. Wu, and D. Zhao, "Wavecast: Wavelet based wireless video broadcast using lossy transmission," in *Visual Communications and Image Processing (VCIP), 2012 IEEE*, Nov 2012, pp. 1–6.

[11] R. Xiong, F. Wu, X. Fan, C. Luo, S. Ma, and W. Gao, "Power-distortion optimization for wireless image/video softcast by transform coefficients energy modeling with adaptive chunk division," in *Visual Communications and Image Processing (VCIP), 2013*, Nov 2013, pp. 1–6.

[12] R. Xiong, F. Wu, J. Xu, and W. Gao, "Performance analysis of transform in uncoded wireless visual communication," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, May 2013, pp. 1159–1162.

[13] X. Fan, F. Wu, D. Zhao, and O.C. Au, "Distributed wireless visual communication with power distortion optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 6, pp. 1040–1053, June 2013.

[14] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, "Cactus: A hybrid digital-analog wireless video communication system," in *Proceedings of the 16th ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, 2013, MSWiM '13, pp. 273–278.

[15] K. Takagi, Y. Takishima, and Y. Nakajima, "A study on rate distortion optimization scheme for jvt coder," *SPIE*, vol. 5150, pp. 914–923, 2003.

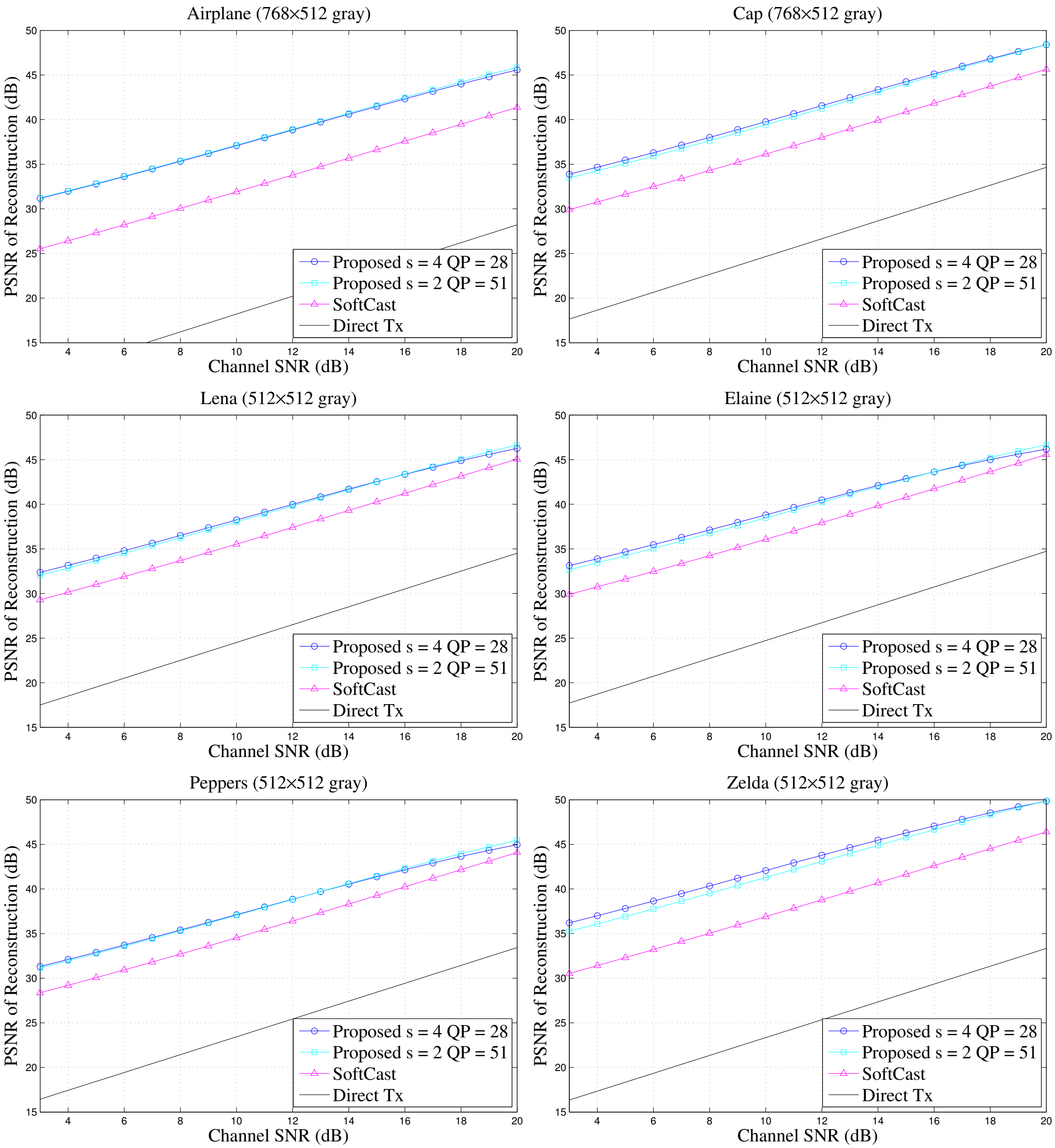

Fig. 6. Performance comparison between the proposed approach, the original SoftCast and direct pixel transmission.

Fig. 7. Reconstructed images by the original SoftCast and the proposed scheme at CSNR = 3dB. In the proposed scheme QP = 51 and s= 2. The columns from left to right: the original SoftCast and the proposed scheme.