Compressive Gradient Based Scalable Image SoftCast

Hangfan Liu*, Ruiqin Xiong*, Xiaopeng Fan†, Chong Luo‡, Wen Gao*

*Institute of Digital Media, Peking University, Beijing 100871, China

†Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China

‡Microsoft Research Asia, Beijing 100080, China

Email: {liuhf,rqxiong,wgao}@pku.edu.cn, fxp@hit.edu.cn, cluo@microsoft.com

***Abstract*—In wireless visual communication systems, it is crucial to effectively utilize channel power and bandwidth in the pursue of optimal performance, and it is worthwhile to adapt the transmission scheme to human vision system (HVS) so as to achieve perceptually appealing results. Inspired by observations that visual quality of an image is closely related to the gradient data, this paper proposes to convey visual information by random projection measurements of image gradients in an analog framework. Since HVS is more sensitive to luminance variations of image contents, which are contained in the gradient data, the proposed scheme achieves better perceptual quality than conventional analog uncoded schemes like SoftCast. Besides, the gradient transform removes the low and medium frequency components of the image hence substantially reduces the power of the signal transmitted in the analog channel, thus evidently improves the power-distortion performance of the system. Furthermore, by applying random projection to the gradients, the number of transmitted data can be adjusted according to bandwidth conditions. Another contribution of this paper is to develop an effective optimization scheme for the compressive gradient based reconstruction problem. Experimental results validate the effectiveness of the the proposed transmission and reconstruction scheme under different channel signal-to-noise ratio and bandwidth conditions.**

***Index Terms*—Wireless visual communication, SoftCast, bandwidth, power, channel signal-to-noise ratio, gradient.**

I. Introduction

In the past few years, there has been a rapid growth in the demand for mobile communication, which calls for more efficient wireless visual communication techniques. In wireless scenarios, however, visual communication is fairly challenging due to fluctuant channel condition and relatively high error rate. Under this circumstance, uncoded transmission schemes like SoftCast [1] are invented, and have attracted much research interest in recent years [2]–[6]. SoftCast is endowed with the virtue of graceful quality transition within a broad channel signal-to-noise ratio (CSNR) range, and the ability to treat multiple receivers in different channel conditions simultaneously via a single signal transmission. It applies a linear transform to decorrelate the image signal, and generates a stream of coefficient numbers. In order to transmit these numbers through the wireless channel with minimum distortion, SoftCast scales them individually to balance the energy consumption of coefficients with different significance, and whiten the whole number stream by Walsh-Hardmard Transform, then modulate the resulted numbers directly to a dense constellation for transmission. The scaling factors are determined by a power-distortion optimization procedure, and are sent to the receiver by a limited number of metadata.

The two key steps, decorrelation transform and power allocation, enables SoftCast to attain higher transmission efficiency. One aspect SoftCast has in common with conventional schemes is that, it also aims to minimize mean square error (MSE) of pixels so as to maximize the fidelity of reconstructed images. However, such fidelity metric has long been criticized for not reflecting visual quality perceived by human eyes [7], and the reconstructions of SoftCast exhibit annoying artifacts when CSNR is relatively low, despite of their high peak signal-to-noise (PSNR) scores.

This paper aims to design a wireless visual communication scheme that can achieve appealing perceptual performance, even when transmission power and channel bandwidth are both limited. We propose to transmit random projection measurements of image gradients instead of DCT coefficients, so that the new scheme bears the merits of pleasant visual quality, efficient power usage as well as adaption to limited channel bandwidth conditions. At the receiver, the measurements are reproduced from the wireless channel and used to reconstruct the image. By exploiting both local smoothness and nonlocal similarity of natural images, we integrate the gradient and random projection reconstruction into a single optimization problem, leading to a more efficient and resilient system producing reconstructions with better visual quality.

II. Compressive Gradient Based SoftCast

A. Gradient Based SoftCast: Advantages and Disadvantages

Recent researches on image quality assessment indicate that gradient data are closely related to perceptual similarity of natural images, and most of the visual information is contained in the image gradients [8], [9]. Based on such conclusions, some works consider that gradients can convey visual information more effectively in visual communication systems [2]. Besides, in analog uncoded transmission framework for wireless scenarios, using gradient data is very effective in power usage. In [10], the concept of "activity" of a random source $\mathbf{x}$ is introduced to measure the cost of transmitting $\mathbf{x}$ through a noisy wireless channel, which is defined as $H(\mathbf{x}) \triangleq \sum_i \sqrt{E\left[x_i^2\right]}$, where x_i is the i-th signal element

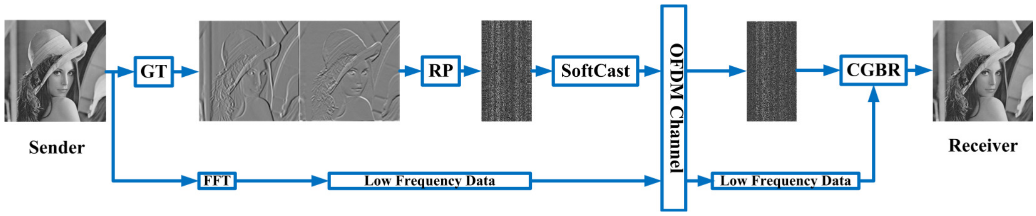

Fig. 1. Illustration of CG-Cast transmission scheme

in $\mathbf{x}$, $E(\cdot)$ stands for the calculation of expectation. Given fixed power (i.e. for a fixed CSNR), higher $H(\mathbf{x})$ leads to lower reconstruction quality. In this regard, the process to generate gradient data sets low frequency data aside and most of the gradients are small numbers close to zero, hence the corresponding $H(\mathbf{x})$ substantially lowered.

The "curse" of using gradients is about the channel bandwidth. In the most typical case, gradient data include both vertical and horizontal variations of neighboring pixels, hence the data size of the gradients is twice as large as the image size, which would be a problem when bandwidth is limited. Enlightened by the success of compressive sensing (CS) [11], [12] which shows that a signal can be well recovered at much lower sampling rate than Nyquist rate if the signal is sparse, we consider to compress the gradient data via random projection.

B. The Proposed Transmission Scheme

Based on the discussions above, this paper employs random projection (RP) to compress the image gradients before they are transmitted, so that data size can be reduced to an arbitrary ratio, thus channel bandwidth occupation can be lowered. The random projection matrix Φ is known at both the sender side and the receiver side. Typically, Φ can be a Gaussian random matrix. The sampling rate is easily controlled by adjusting the number of rows in Φ. We name this scheme as compressive gradient based SoftCast (CG-Cast).

The illustration of CG-Cast is shown in Fig. 1. At CG-Cast sender, the gradients are down-sampled by RP after generated by gradient transform (GT). The sample rate is set according to the channel bandwidth condition. Then the produced measurements are sent out for raw OFDM transmission in the similar way as in SoftCast. It is worth noting that, similar with SoftCast, CG-Cast also delivers streams of real numbers rather than streams of bits. In addition, a few low frequency components are also delivered so as to tell the global and regional luminance of the image. The discussions of this paper focus on wireless image SoftCast, which can be directly applied to intra coded frames in videos, and the extension to the whole videos mainly involves exploiting temporal correlation between successive frames.

Let D^{v} and D^{h} be the vertical and horizontal gradient transform operator respectively. The gradients $D^{\mathrm{v}}\mathbf{x}$ and $D^{\mathrm{h}}\mathbf{x}$ are down-sampled by random projections Φ^{v} and Φ^{h}. In wireless transmission process, the transmitted data are prone to be influenced by interference in the air, which is simulated by i.i.d. zero-mean Gaussian noise:

$$\mathbf{m}^{\mathrm{v}} = \Phi^{\mathrm{v}} \cdot D^{\mathrm{v}}\mathbf{x} + \mathbf{n}^{\mathrm{v}}, \; \mathbf{m}^{\mathrm{h}} = \Phi^{\mathrm{h}} \cdot D^{\mathrm{h}}\mathbf{x} + \mathbf{n}^{\mathrm{h}}, \tag{1}$$

where $\mathbf{x}$ is the original image signal, $\mathbf{m}^{\mathrm{v}}$ and $\mathbf{m}^{\mathrm{h}}$ are the noisy measurements, $\mathbf{n}^{\mathrm{v}}$ and $\mathbf{n}^{\mathrm{h}}$ are additive Gaussian white noise. The noise level is commensurate with the channel condition. We write $\Phi = [\Phi^{\mathrm{v}}; \Phi^{\mathrm{h}}]$, $D = [D^{\mathrm{v}}; D^{\mathrm{h}}]$ and $\mathbf{m} = [\mathbf{m}^{\mathrm{v}}; \mathbf{m}^{\mathrm{h}}]$ in this paper for simplicity.

C. Compressive Gradient based Reconstruction

At CG-Cast decoder, the received noisy measurements and relatively clean low frequency data are used for compressive gradient based reconstruction (CGBR) procedure, which requires more sophisticated techniques than that in G-Cast. Similar with CS reconstruction, CGBR entails the image prior that the image signal is sparse in a certain domain. Similar with some recent works on image restoration [4], [5], [13]–[15], this paper utilizes the nonlocal self-similarity of natural images so as to obtain sparser presentation of the image. Specifically speaking, we stack similar patches of the reconstruction image into groups and perform three dimensional (3D) transform, i.e. apply 2D wavelet transform patch-wise and 1D DCT along the third dimension, then the generated coefficients are almost thoroughly decorrelated and highly sparse. Besides, in order to model the local smoothness of images, we use total variation (TV) regularization, which can be seen as sparse presentation in gradient domain. In this way, we actually integrate the gradient and RP reconstruction into a single optimization problem, which leads to an efficient and resilient system.

Employing 3D transform sparsity and TV regularization, the estimate of the original signal $\mathbf{x}$ can be formulated as:

$$\min_{\mathbf{x}} \sum_i \|D_i\mathbf{x}\|_2 + \frac{\eta}{2}\|T^{3D}(\mathbf{x})\|_0 + \frac{\mu}{2}\|\Phi \cdot D\mathbf{x} - \mathbf{m}\|_2^2 \quad \text{s.t. } E \circ \mathcal{F}(\mathbf{x}) = \mathbf{L}, \tag{2}$$

where $D_i\mathbf{x} \in \mathbb{R}^2$ represents the gradients of $\mathbf{x}$ at pixel i and $\|D_i\mathbf{x}\|_2$ is the isotopic TV regularization term, T^{3D} stands for 3D transform applied to patch groups retrieved from the image, $\|\cdot\|_0$ counts the non-zero supports, $\|\Phi \cdot D\mathbf{x} - \mathbf{m}\|_2^2$ is l_2-norm data-fidelity term, μ and η are regularization parameters controlling the trade-off between three competing terms, $\mathbf{L}$ is the low-frequency coefficients of $\mathbf{x}$, $\mathcal{F}$ stands for two-dimensional discrete Fourier transform, E represents the matrix to extract the $M \times M$ block at the top left corner, and $\circ$ denotes component-wise multiplication. Since the techniques