Performance Analysis of Transform in Uncoded Wireless Visual Communication

Ruiqin Xiong*, Feng Wu†, Jizheng Xu†, Wen Gao*
* Institute of Digital Media, School of Electronic Engineering and Computer Science,
Peking University, Beijing 100871, China
† Microsoft Research Asia, Beijing 100080, China

***Abstract*—In wireless scenarios where the channel condition may vary drastically, visual communication systems using source and channel coding generally suffer from threshold effect. An uncoded transmission scheme called SoftCast [1]–[3], however, was recently shown to provide both graceful quality transition and competitive performance. In SoftCast, image signal is directly modulated to a dense constellation using proper power for transmission, solely after employing a transform for energy compaction, leaving out conventional quantization, entropy coding and channel coding. The received signal is lossy in nature, with its noise level commensurate with the channel condition. This paper presents a theoretical analysis for uncoded visual communication, focusing on the role of transform and the quantitative measurement of transform gain in a generalized uncoded transmission framework with optimal power allocation. Our analysis reveal that the energy distribution among signal elements plays an important role in the power-distortion performance. Further analysis show that the energy compaction capability of decorrelation transform can bring significant gain by boosting the energy diversity in signal representation. Numerical analysis results are reported for Markov random signals and natural images, respectively. The performance of typical transforms, e.g. KLT, DCT and DWT, and the effect of different transform sizes or levels are evaluated. These analysis results are verified by simulations.**

I. Introduction

A communication system based on source and channel coding generally requires the channel statistics to be known at the time of coding, in order to choose an appropriate source and channel coding rate. Once the coding process is finished, it works optimally only for a specific channel quality: if the actual channel quality falls below a threshold, the decoding process tends to break down completely; if the channel quality increases beyond that threshold, it cannot provide any improvement in performance. This is known as the "threshold effect". For this reason, accurate channel estimation is desired. However, channel condition may vary drastically and unpredictably, especially in wireless communication scenarios. Therefore, existing communications systems tend to utilize the channel conservatively.

Recently, a scheme called *SoftCast* [1]–[3] was proposed for wireless video. It is essentially a scheme with "*lossless compression and lossy transmission*", as illustrated in Fig. 1. The compression stage is solely a transform to decorrelate the image signal, leaving out the conventional quantization and entropy coding. The transmission stage also abandons the conventional channel coding. Instead, it scales each transform coefficient individually and modulates it directly to a dense constellation for transmission. The scaling operation serves the purposes of both power allocation and unequal signal protection against channel noises. For practical optimization, SoftCast groups the coefficients into a set of chunks and perform scaling at chunk level. At the receiver, the image is reconstructed by demodulating the received signal and inverting the scaling and transform. The scheme was shown to not only provide graceful performance transition in wide channel signal-to-noise range, but also achieve competitive performance compared with the state-of-the-art coding scheme.

Send correspondence to R. Xiong (rqxiong@pku.edu.cn). This work was supported in part by National Natural Science Foundation of China (61073083, 61121002), National Basic Research Program of China (2009CB320904), Beijing Natural Science Foundation (4112026, 4132039) and Research Fund for the Doctoral Program of Higher Education (20100001120027, 20120001110090).

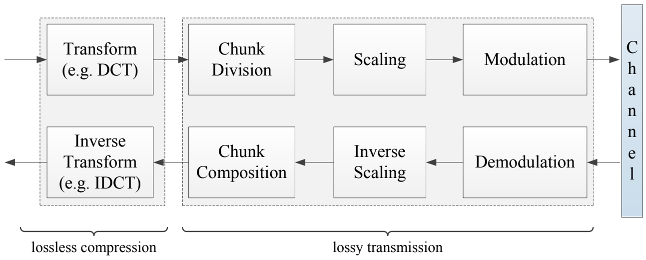

Fig. 1. The flowchart of SoftCast [1]–[3].

The impressive performance of SoftCast motivates us to consider uncoded transmission as a promising direction for robust and efficient wireless visual communication. This paper presents a theoretical analysis for SoftCast, focusing on the role of transform and the quantitative measurement of transform gain in a generalized uncoded transmission framework with optimal power allocation. Our analysis reveal that the energy distribution among signal elements plays an important role in the power-distortion performance. Further analysis show that the energy compaction capability of decorrelation transform can bring significant gain by boosting energy diversity in signal representation. Based on these analysis, the design of an uncoded visual communication system is discussed. In particular, the performance of different transforms (e.g. KLT, DCT or DWT) and different transform sizes or levels are evaluated. The results are verified by transmission simulations.

II. Uncoded Transmission with Unequal Protection

A. Distortion-Power Function of Uncoded Transmission

We consider a general uncoded transmission framework. Suppose $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ is a random vector to transmit over a noisy channel. Typically, each element x_i may represent a single pixel or a transform coefficient. To utilize the transmission power efficiently, the encoder scales each signal element x_i by a factor $g_i \in \mathbb{R}^+$ and sends out

$$y_i = g_i \cdot x_i \tag{1}$$

directly using a dense modulation constellation. The signal that arrives at the receiver (after demodulation) is

$$\hat{y}_i = y_i + n, \tag{2}$$

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

$$\hat{x}_i = \hat{y}_i / g_i = x_i + n/g_i. \tag{3}$$

In the above process, the expected distortion in $\hat{x}_i$ is

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

The expected transmission power for sending x_i is

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

Combining (4) and (5), we get the distortion-power relationship:

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

B. Unequal Protection via Power Allocation

To achieve optimal performance, the transmission power is allocated among the elements $\{x_i\}$ by

$$\text{(P1):} \quad \text{minimize} \quad \sum_i D_i \quad \text{s. t.} \quad \sum_i P_i \leqslant P_{\text{total}} \tag{7}$$

This optimization problem can be easily solved by setting the distortion-power slopes of all elements to be equal:

$$\frac{\partial D_i}{\partial P_i} = \frac{-\sigma_n^2 \cdot \mathrm{E}[x_i^2]}{P_i^2} = const \tag{8}$$

This determines the optimal power for sending x_i:

$$P_i = C\sigma_n \sqrt{\mathrm{E}[x_i^2]} \tag{9}$$

Here C is used to control the total transmission power. Substituting (9) into (5), the optimal scaling factors are

$$g_i = \sqrt{C\sigma_n}(E[x_i^2])^{-1/4} \quad \text{or} \quad g_i \propto (E[x_i^2])^{-1/4} \tag{10}$$

C. Overall Performance

Recall the equations (6) and (9), we easily derive

$$D_i = \frac{1}{C}\sigma_n \sqrt{\mathrm{E}[x_i^2]} \tag{11}$$

Here, the normalization factor C is determined by $\sum_i P_i = P_{\text{total}}$ so that $C = P_{\text{total}} / (\sigma_n \sum_i \sqrt{\mathrm{E}[x_i^2]})$. Therefore, the total expected distortion under optimal power allocation is

$$D_{\text{total}} = \sum_i D_i = \frac{\sigma_n^2}{P_{\text{total}}} \left(\sum_i \sqrt{\mathrm{E}[x_i^2]} \right)^2 \tag{12}$$

Consider the definition of channel signal-to-noise ratio and the peak signal-to-noise ratio of reconstructed signal, we have

$$\mathrm{PSNR_{dB}} = c + \mathrm{CSNR_{dB}} - 10\log_{10} \left(\sum_i \sqrt{\mathrm{E}[x_i^2]} \right)^2 \tag{13}$$

with $c = 10\log_{10}(255^2 N)$. We note that the reconstruction PSNR increases linearly with CSNR with a ratio of $1:1$. To reveal the relationship between the signal statistics and the transmission performance, we define

$$H(\mathbf{x}) \triangleq \sum_i \sqrt{\mathrm{E}[x_i^2]} \tag{14}$$

as the "activity" of a random source $\mathbf{x}$. This is analogous to the concept of "entropy", in the sense that it measures the difficulty in transmitting the signal over a noisy channel. For a fixed channel SNR, higher activity $H(\mathbf{x})$ in $\mathbf{x}$ means lower quality in the reconstruction $\hat{\mathbf{x}}$.

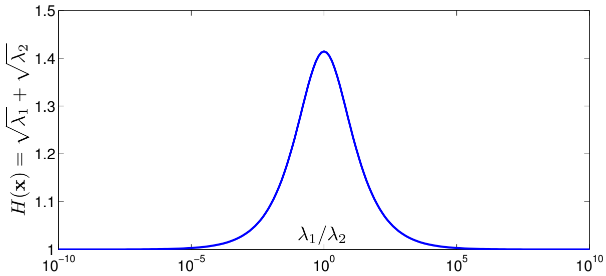

Fig. 2. $H(\mathbf{x})$ of two-variable random source $\mathbf{x}$ with $\mathrm{E}[x_1^2] = \lambda_1$ and $\mathrm{E}[x_2^2] = \lambda_2$, subject to $\lambda_1 + \lambda_2 = \lambda$. $H(\mathbf{x}) = \sqrt{2\lambda}$ is reached when $\lambda_1 = \lambda_2$. In this figure λ=1.

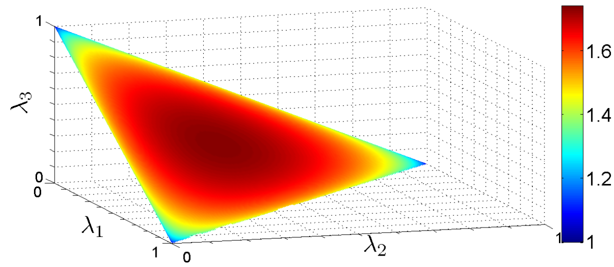

Fig. 3. $H(\mathbf{x})$ of three-variable random source $\mathbf{x}$ with $\mathrm{E}[x_i^2] = \lambda_i, i = 1, 2, 3$, subject to $\lambda_1 + \lambda_2 + \lambda_3 = \lambda$. $H(\mathbf{x}) = \sqrt{3\lambda}$ is reached when $\lambda_1 = \lambda_2 = \lambda_3$. In this figure λ=1.

III. Energy Diversity and Transform Gain

In this section, we reveal the advantages of employing decorrelation transform in SoftCast. We show that a decorrelation transform can bring significant performance gain in uncoded transmission by boosting the energy diversity in signal representation.

A. Energy Diversity

We first study the mathematical property of definition (14). Obviously, $f(x) = \sqrt{x}$ is a strictly upper convex function. Therefore, we have the following remarks: (suppose λ is a constant, $\lambda, \lambda_i \in \mathbb{R}^+$)

Remark 1 If $\lambda_1 <= \lambda_2$, $f(\lambda_1) + f(\lambda_2) > f(\lambda_1 - \epsilon) + f(\lambda_2 + \epsilon)$ holds for $\forall \epsilon > 0$.

Remark 2 Subject to $\lambda_1 + \lambda_2 = \lambda$, $f(\lambda_1) + f(\lambda_2)$ achieves its maximum value only when $\lambda_1 = \lambda_2$.

Fig. 2 illustrates the $H(\mathbf{x})$ value of two-variable random source with different energy distribution. It is clear in Fig. 2 that a higher diversity in the signal's energy distribution corresponds to a lower value in $H(\mathbf{x})$.

The above conclusions can be extended to more general cases of N-variable random source. If $\sum_i \lambda_i = \lambda$ is constant, $\sum_i f(\lambda_i)$ achieves its maximum value only when all $\lambda_i, i = 1, 2, \ldots, N$ are equal. The more diversified these λ_i values are, the smaller $\sum_i f(\lambda_i)$ is. Fig. 3 illustrates how the $H(\mathbf{x})$ value varies with the energy distribution of $\mathbf{x}$, for the case $N = 3$.

B. Effect of Decorrelation Transform

Natural image signals usually exhibit strong correlation among the neighboring pixels. This is commonly exploited in image coding by decomposing the signal using a decorrelation transform. *Karhunen-Loève Transform* (KLT) is the optimal transform to use, when the statistics of signal are known in advance. In practice, DCT is widely used instead as it is a good approximation of KLT [4].

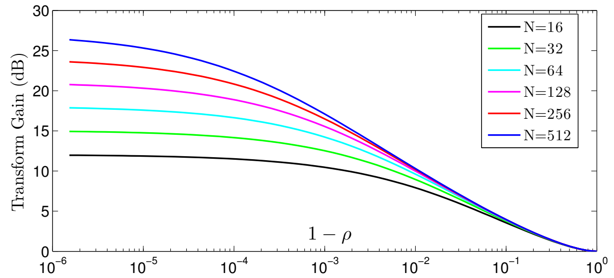

Fig. 4. Transform gain of $1 \times N$ KLT for one-dimensional first-order Markov process with correlation coefficient ρ.

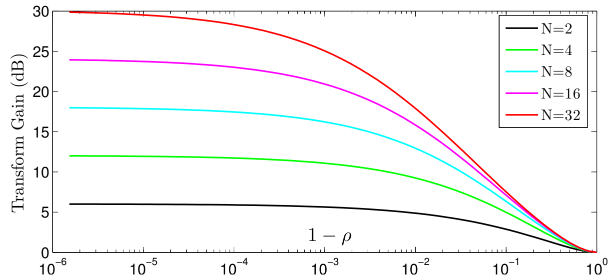

Fig. 5. Transform gain of $N \times N$ KLT for two-dimensional first-order Markov process with correlation coefficient ρ.

Typical decorrelation transforms for images, such as KLT, DCT and DWT, etc., are orthogonal or approximately orthogonal. Therefore, they do not change the total energy of a signal, as long as proper normalization is used. What they change, however, is the distribution of energy among signal elements. After decorrelation, the signal energy is usually compacted to a small part of coefficients: a small number of coefficients become large while most other coefficients become close to zero. Therefore, energy distribution in the signal representation becomes much more diversified after decorrelation.

Based on Section III-A and the above discussions, we infer that a decorrelation transform can reduce the "activity" of a signal, which subsequently leads to higher energy utilization efficiency in uncoded transmission and better quality in reconstruction. For an orthonormal transform $\mathcal{T} : X(i) \to Y(u)$, we define the ***transform gain*** (in dB) of $\mathcal{T}$ in the context of uncoded transmission by

$$G(X|\mathcal{T}) = 20\log_{10}\left(\frac{H(X)}{H(Y)}\right). \tag{15}$$

To be concrete,

$$G(X|\mathcal{T}) = 20\log_{10}\left(\frac{\sum_i \sqrt{\lambda_{X(i)}}}{\sum_u \sqrt{\lambda_{Y(u)}}}\right), \tag{16}$$

where $\lambda_{X(i)} = \mathrm{E}[X(i)^2]$, $\lambda_{Y(i)} = \mathrm{E}[Y(i)^2]$.

IV. Numerical Results of Transform Gain Analysis

In this section, we evaluate the transform gain defined in (16) for some example signals. For this purpose, the knowledge of energy distribution among signal elements is assumed to be perfectly known.

A. Markov Random Signals

First-order Markov random process is a simple but widely used signal model for natural images. For a stationary first-order Markov

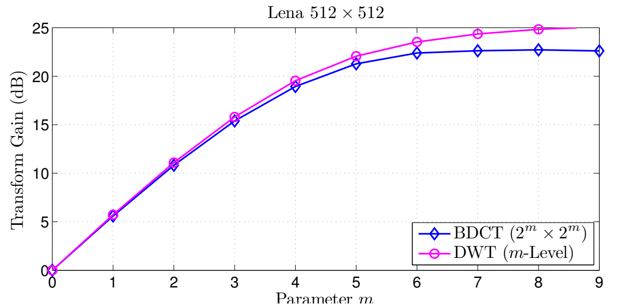

Fig. 6. Transform gain of BDCT and DWT on *Lena* (512×512, gray).

process $\{X_t\}$, $t = 1, 2, \ldots, N$ with correlation coefficient ρ, the covariance matrix is

$$C_X = \sigma_X^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix}. \tag{17}$$

The KLT transform result $Y = \mathrm{KLT}(X)$ has a covariance matrix

$$C_Y = \sigma_X^2 \cdot \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}. \tag{18}$$

The diagonal entries $\{\lambda_i\}$ correspond to the eigen-values of the Toeplitz matrix in (17), which can be determined by (see [5]):

$$\lambda_i = \frac{1-\rho^2}{1-2\rho\cos\omega_i + \rho^2}, \quad i = 1, 2, \ldots, N, \tag{19}$$

where $\{\omega_i\}$ are the positive roots of

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin\omega}{\cos\omega - 2\rho + \rho^2\cos\omega}. \tag{20}$$

Fig. 4 illustrates the KLT transform gain for one-dimensional first-order Markov process and shows how the gain varies with the correlation parameter ρ and the transform size N. For a fixed N, the transform gain increases as ρ goes from 0 to 1 (from low correlation to high correlation). That means applying KLT to a signal with stronger correlation can provide higher transform gain. For a fixed ρ, the transform gain generally increases with N. That corresponds to the fact that a transform of larger size can exploit the signal correlation at a larger scale. When $\rho \to 1$, the transform gain increases by 3.01dB each time N is doubled.

These remarks can be extended to signals of higher dimension. Fig. 5 illustrates the transform gain of $N \times N$ KLT for two-dimensional signal which is first-order Markov in each dimension with correlation ρ. When $\rho \to 1$, the transform gain increases by 6.02dB each time N is doubled.

B. Natural Images

For real-world natural images, we consider DCT and DWT (using the 9/7 filter), which are employed in the still image coding standard JPEG and JPEG2000, respectively. Fig. 6 illustrates the transform gain for *Lena* (512×512, gray). Clearly, for both block-DCT and DWT, the transform gain increases with transform block size or transform level. For typical 512×512 natural images, the transform gain can be as high as $20 \sim 25$ dB. In addition, the performance of m-level DWT is slightly better than that of DCT with block size $2^m \times 2^m$.

V. Experimental Results.

We conduct SoftCast simulations using DCT and DWT for *Lena*, *Peppers*, *Elaine*, *Barbara*, etc., to verify the analysis presented in previous sections. To limit the overhead for sending energy distribution information, the coefficients in each band are assumed to have the same statistics, which are delivered to the receiver by band-level meta-data. Results for all tested images exhibit similar trend so that only the results for *Lena* are shown. Fig. 7 and Fig. 8 illustrate the reconstruction PSNR versus channel SNR curves, for SoftCast using block-DCT with various transform sizes and DWT with various transform levels. Obviously, there exists a linear relationship between PSNR and CSNR (except at very low CSNR region where the clip effect become visible). In addition, larger block-DCT transform size or more DWT transform levels lead to better transmission performance. The reconstructed images using these transform configurations are shown in Fig. 9 and Fig. 10.

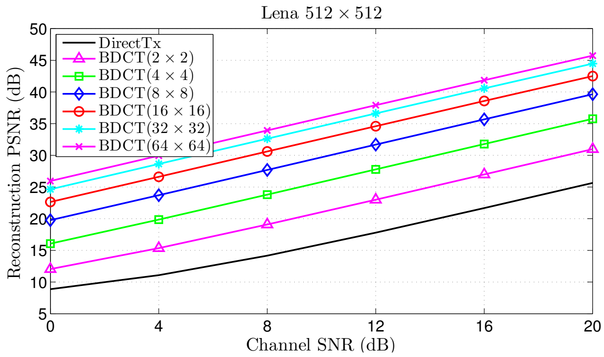

Fig. 7. Simulated performance of block DCT using various transform sizes.

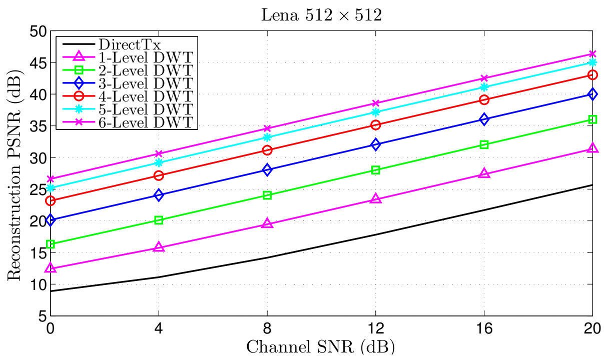

Fig. 8. Simulated performance of DWT using various transform levels.

Fig. 9. Simulation results of SoftCast transmission using $2^m \times 2^m$ DCT (CSNR= 4dB). From top-left to bottom-right: $m = 0, 2, 4, 6$.

Fig. 10. Simulation results of SoftCast transmission using m-level DWT (CSNR= 4dB). From top-left to bottom-right: $m = 0, 2, 4, 6$.

VI. Conclusions and Discussions

This paper presented a theoretical analysis for uncoded visual communication and shown that decorrelation transform can bring significant gain by boosting the energy diversity in the signal representation. Larger transform size in DCT or more transform levels in DWT produces higher energy diversity, but it also requires more meta-data to describe such diversity. More efficient energy diversity describing method will be studied in our future work.

References

[1] 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.

[2] ——, "One-size-fits-all wireless video," in *Proc. Eigth ACM SIGCOMM HotNets Workshop*, New York, NY, USA, 2009.

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

[4] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. San Diego, CA: Academic Press, 1990.

[5] W. Ray and R. Driver, "Further decomposition of the karhunen-loève series representation of a stationary random process," *IEEE Transactions on Information Theory*, vol. 16, no. 6, pp. 663–668, Nov 1970.