Analysis of Decorrelation Transform Gain for Uncoded Wireless Image and Video Communication

Ruiqin Xiong, *Member, IEEE*, Feng Wu, *Fellow, IEEE*, Jizheng Xu, *Senior Member, IEEE*, Xiaopeng Fan, *Member, IEEE*, Chong Luo, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— An uncoded transmission scheme called SoftCast has recently shown great potential for wireless video transmission. Unlike conventional approaches, SoftCast processes input images only by a series of transformations and modulates the coefficients directly to a dense constellation for transmission. The transmission is uncoded and lossy in nature, with its noise level commensurate with the channel condition. This paper presents a theoretical analysis for an uncoded visual communication, focusing on developing a quantitative measurements for the efficiency of decorrelation transform in a generalized uncoded transmission framework. Our analysis reveals that the energy distribution among signal elements is critical for the efficiency of uncoded transmission. A decorrelation transform can potentially bring a significant performance gain by boosting the energy diversity in signal representation. Numerical results on Markov random process and real image and video signals are reported to evaluate the performance gain of using different transforms in uncoded transmission. The analysis presented in this paper is verified by simulated SoftCast transmissions. This provide guidelines for designing efficient uncoded video transmission schemes.**

***Index Terms*— Uncoded visual communication, image representation, decorrelation transform, power allocation, energy diversity, transform gain.**

I. Introduction

VISUAL communication systems nowadays are mostly built upon the *Shannon's separation theorem* [1]–[3], which suggests that source coding and channel coding can be designed and optimized independently without sacrificing the optimality in terms of rate-distortion performance. A general framework of this kind is illustrated by Fig. 1. Input image is compressed into a binary stream of much smaller size, using source coding to remove the statistical and perceptual redundancy in signal. The coded bits are protected by channel codes and then modulated and transmitted over wired or wireless channels.

In the past decades, the source coding camp has made huge efforts in developing all kinds of efficient coding tools, including intra- and inter-frame prediction, transform, quantization, and entropy coding, etc [4]–[8]. As a result, coded transmission achieved great success for digital video communications. However, with the prevalence of wireless networks and mobile devices in recent years, the weakness of this framework also becomes evident.

Firstly, the framework is sensitive to bit errors. As a result of extreme source coding, every bit is exploited to the greatest extent to convey information. Nevertheless, it also makes the stream extremely vulnerable to transmission errors – even a single flip of bit may turn the stream of a whole slice into useless bits. Secondly, the framework lacks the flexibility to handle channel variation. Once the compression is done, the system works optimally only for a specific channel capacity presumed by the source coder – if the actual channel quality falls below a threshold, the decoder tends to break down completely; if the channel quality increases beyond that threshold, the system cannot provide better performance.

To deal with transmission errors, channel codes are widely adopted to protect coded streams [9]–[11]. To maintain efficient channel utilization, unequal error protection (UEP) [12]–[21] can be used to handle bit streams of different importance. In scenarios where feedback is available and delay is allowed, hybrid FEC and retransmission [22]–[24] may be employed to trade off between delay and bandwidth efficiency. These techniques usually require instantaneous channel estimation and intelligent transmission scheduling, all of which make the system design complicated. Even so, error-free transmission cannot be guaranteed, especially for dynamic channels with burst errors.

To handle diversity in channel bandwidth, people investigated various scalable coding schemes, e.g. EZW [25], SPIHT [26] and JPEG2000 [27] for images, and MPEG-4 FGS [28], PFGS [29], 3D-SPIHT [30], MCTF [31]–[35] and H.264 SVC [36] for videos. Such schemes encode images into embedded streams that can be truncated on the fly at the time of transmission. However, it has not been widely adopted in practical systems due to non-negligible performance loss compared with its non-scalable counterpart.

Visual communication in wireless scenarios is particularly challenging, because the error rate of wireless link is relatively

Manuscript received January 20, 2015; revised December 25, 2015; accepted February 17, 2016. Date of publication February 26, 2016; date of current version March 9, 2016. This work was supported in part by the National Basic Research Program of China under Grant 2015CB351800, in part by the National Natural Science Foundation of China under Grant 61370114, Grant 61425026, and Grant 61421062, in part by the Beijing Natural Science Foundation under Grant 4132039, and in part by the Cooperative Medianet Innovation Center. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung.

R. Xiong and W. Gao are with the Institute of Digital Media, School of Electronic Engineering and Computer Science, and National Engineering Laboratory for Video Technology (NELVT), Peking University, Beijing 100871, China (e-mail: rqxiong@pku.edu.cn; wgao@pku.edu.cn).

F. Wu was with Microsoft Research Asia, Beijing 100080, China. He is now with the University of Science and Technology of China, Hefei 230026, China (e-mail: fengwu@ustc.edu.cn).

J. Xu and C. Luo are with Microsoft Research Asia, Beijing 100080, China (e-mail: jzxu@microsoft.com; chong.luo@microsoft.com).

X. Fan is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: fxp@hit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2535288

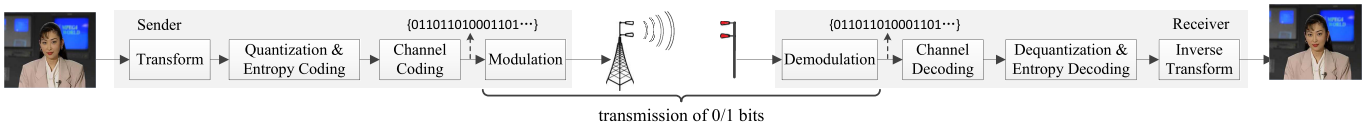


Fig. 1. The coded transmission framework for visual communication.

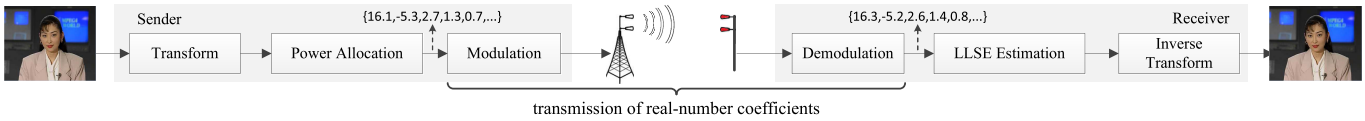


Fig. 2. The uncoded transmission framework for visual communication.

high due to mobility and interference, and the channel condition may fluctuate drastically and unpredictably. To make the things worse, in wireless video broadcast, users at different locations have diversified channel conditions. The conventional framework cannot serve all the client with satisfaction – it has to utilize the channel conservatively by choosing a low video bit rate that can work in the worst cases.

Recently, a novel approach called SoftCast [37]–[39] was proposed for wireless video transmission. It has attracted much attention [40]–[65] since it not only provides an elegant way to deal with transmission errors and channel quality variations, but also achieves competitive performance compared with the state-of-the-art coding scheme at the threshold channel quality, resulting in a superb overall performance for a wide channel SNR range [37], [38], [47], [54], [55].

SoftCast essentially adopts an uncoded transmission framework, as illustrated in Fig. 2. Compared with conventional approaches, the framework has several distinctive features. Firstly, instead of coding the images into a stream of bits, SoftCast applies solely a linear transform to decorrelate the signals, producing a stream of coefficient numbers, and leaving out the conventional quantization and entropy coding. Secondly, instead of struggling for "intact" delivery of discrete information bits, the transmission module in SoftCast only strives to make the numbers get through channel with minimum distortion. SoftCast no longer uses channel coding. Instead, it scales each coefficient individually, applies a Walsh-Hadamard Transform (WHT) to whiten the whole stream, and modulates the resulted numbers directly to a dense constellation for transmission. As illustrated by Fig. 3, a pair of real numbers from the coefficient stream are mapped to a point in the constellation, using the two numbers as the I- and the Q- components, respectively. The transmission is "analog-like" and lossy in nature, with its noise level commensurate with the channel condition (see [37]–[39] for more details). At the receiver side, SoftCast first recovers the coefficient numbers by some linear estimator and then reconstructs the images via inverse transform.

Although the uncoded transmission framework is simple and surely has much lower complexity, it has demonstrated very promising performance. This motivates us to investigate its mechanism with in-depth analysis. Uncoded transmission is not unprocessed transmission. In fact, signal decorrelation and transmission power optimization are two key modules

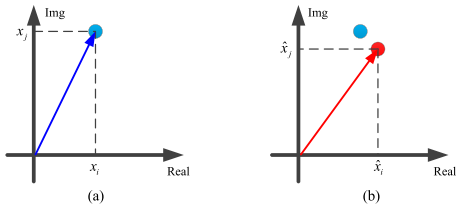


Fig. 3. The modulation for "analog-like" lossy transmission in SoftCast. (a) the sent signal; (b) the received signal.

that determine the efficiency of uncoded transmission. Both coded and uncoded transmission employ transform for signal decorrelation. The most fundamental difference between coded and uncoded transmission is that coded transmission turns the coefficients into equally important bits and then allocates the communication resources (bandwidth and transmission power) to each coded bit, while uncoded transmission allocates the bandwidth and transmission power directly to the coefficients.

This paper presents a theoretical analysis for the transform gain in uncoded transmission. Our analysis reveals that the energy distribution among signal elements is critical for the efficiency of uncoded transmission. A decorrelation transform can potentially bring significant performance gain by boosting the energy diversity in signal representation. That is why SoftCast can achieve very promising performance. More importantly, we demonstrate that the transform gain can be realized only when both the sender and the receiver have good knowledge on the energy diversity in signal elements and fully exploit this diversity for optimal protection. Therefore, the efficiency in describing and sharing the knowledge of energy diversity in signal elements is an important problem for uncoded transmission.

The contributions of the paper are three-fold. Firstly, a formulation for the overall performance of optimized uncoded transmission has been derived, in which the relationship between the reconstruction distortion and the image data complexity is established. Secondly, a quantitative measurement has been developed to measure the efficiency of a decorrelation transform in the context of uncoded transmission. It can be used as a criterion to evaluate different configurations in practical system designs. Thirdly, we revealed the relationship between the energy diversity in signal representation and the transform gain.

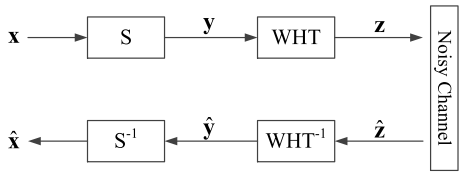


Fig. 4. The general uncoded transmission framework.

The paper is organized as follows. Section II presents a theoretical analysis for uncoded transmission with unequal protection. Section III investigated the relationship between decorrelation and energy diversity. A measurement has been developed for decorrelation transform gain in uncoded transmission. Section IV discusses a few practical issues in power optimization. Experimental results are reported in Section V and Section VI concludes the paper.

II. Uncoded Transmission With Unequal Protection

In this section, we present a general uncoded transmission framework and develop a theoretical analysis for its performance under power-distortion optimization. Although part of the discussion appears in the SoftCast papers [37]–[39], the analysis presented here is more comprehensive and we aim to inspire deeper insights.

A. Uncoded Transmission

We consider a general uncoded transmission framework as depicted in Fig. 4. Suppose $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ is a *random* signal to transmit over a noisy channel. In visual communication scenarios, each data element x_i here may typically represent a single pixel in the spatial domain, or a single coefficient in the transform domain. The N elements, $\{x_i\}, i = 1, 2, \ldots, N$, may have similar or different statistics. The goal is to process and send out $\mathbf{x}$ in a proper manner so that the signal can be recovered at the receivers with maximum fidelity, subject to the constraints of channel resources and channel conditions.

The channel we consider in this paper is that of AWGN channels, in which the transmitted signal is contaminated by additive white Gaussian noise. It is known that, for a fixed channel noise level, the signal quality perceived by the receiver can be regulated by controlling the transmission power. In order to utilize the total transmission power efficiently, the encoder scales each element x_i individually by a separate scalar $g_i \in \mathbb{R}^+$ (as depicted by the "S" module in Fig. 4), and sends out

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

directly (via raw OFDM, see [37]–[39] for details) without using any quantization, entropy coding or channel coding techniques. Note that the scaling factor g_i can take different values for different elements.

We temporarily bypass the "WHT" module in Fig. 4 so that the signal $\mathbf{y}$ is directly send to the channel. The signal that arrives at the receiver is $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}$, i.e.,

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

where $\mathbf{n}$ is the channel noise. The decoder inverses the scaling operation (1) and gets an estimation of x_i by[1]:

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

B. Power-Distortion Function

In the above transmission process, we have $\hat{x}_i = x_i + n_i/g_i$ so that the expected distortion in $\hat{x}_i$ is

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

Here σ_n^2 is the power of channel noise. The expected transmission power for sending x_i is

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

Combining (4) and (5), we get the power-distortion function of uncoded transmission:

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

Note that (6) holds irrespective of the g_i value we choose.

It is interesting to note that the scaling operation (1) in uncoded transmission is somehow analogous to the channel coding in conventional communication, in the sense that they both protect the signals against channel noises. In both cases, the error resilience capability is obtained by increasing the distance between source signals before transmission. In the above framework, g_i controls both the transmission power P_i and the distortion D_i. The more power we allocate to one element, the less distortion the channel noise will cause in that element.

C. Power-Distortion Optimization

In current communication systems, a piece of bitstream with higher importance will most likely be assigned a stronger channel code so that it can recover from transmission errors with higher probability [12]–[16], [20]–[24]. One may conjecture that unequal protection is also beneficial for the uncoded transmission of signals with different statistics. Intuitively, more power should be allocated to the elements with larger variance (higher uncertainty, more information).

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

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

Here we assumed the distortion contributed by each element to the ultimate reconstruction is additive with equal importance. This is valid when each x_i represents a pixel or a coefficient of an orthonormal transform.

[1]If decoder knows the power of the signal x_i and the channel noise n_i, linear least square estimator (LLSE) may be used [37]. However, this makes remarkable difference only when channel SNR is very low. Furthermore, the channel noise condition is generally unknown by the sender, especially for the wireless broadcasting scenarios. Therefore, we do not consider the effect of LLSE here.

Recall the distortion-power function (6). Since $D_i(\cdot)$ is strictly decreasing and convex w.r.t. P_i, the optimization problem (P1) can be easily solved by setting the distortion-power slopes of all elements to be equal, i.e.,

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

This determines the optimal power for sending x_i:

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

Here C is a normalization factor to meet the constraint on total transmission power. The formulation (9) confirms our conjecture that a signal element with higher variance (or uncertainty) should be transmitted using more power. Substituting (9) into (5), the optimal scaling factor for sending x_i is

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

D. Overall Performance

Now we can formulate the overall performance for optimized uncoded transmission. Recall the equations (6) and (9), we easily get

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

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

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

Based on the following definition of CSNR (channel signal-to-noise ratio) and PSNR (peak signal-to-noise ratio),

$$\text{CSNR}_{\text{dB}} = 10\log_{10}(\bar{P}/\sigma_n^2), \quad \bar{P} = P_{\text{total}}/N, \tag{13}$$

$$\text{PSNR}_{\text{dB}} = 10\log_{10}(255^2/\bar{D}), \quad \bar{D} = D_{\text{total}}/N, \tag{14}$$

we have

$$\text{PSNR}_{\text{dB}} = c + \text{CSNR}_{\text{dB}} - 20\log_{10}\left(\frac{1}{N}\sum_i \sqrt{E[x_i^2]}\right). \tag{15}$$

Here $c = 20\log_{10}(255)$.

We have a few remarks for (15). 1) The reconstruction PSNR increases linearly with the channel SNR, with a 1:1 ratio. 2) Roughly speaking, a signal $\mathbf{x}$ with higher energy (or higher uncertainty) means it's more difficult to reproduce (to a certain fidelity) at receiver, under a certain channel condition. To make our discussion precise, we define

$$H(\mathbf{x}) = \frac{1}{N}\sum_i \sqrt{E[x_i^2]} \tag{16}$$

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

E. Effect of Signal Whitening

In practical implementation of uncoded transmission (see [37]), a *Walsh-Hadamard Transform* (WHT) may be applied on $\mathbf{y}$, as formulated by $\mathbf{z} = W\mathbf{y}$. The purpose of is to whiten the signal $\mathbf{y}$ so that the transformed signal $\mathbf{z}$ has a significantly lower *peak-to-average ratio* (PAR). As a result, the power of transformed signal is much more stable, which is strongly desired for the implementation of PHY layer. Another advantage is that it scrambles the signal so that each element of $\mathbf{z}$ becomes almost equally important. This makes the framework robust to data losses due to packet drops.

Although the WHT module is very helpful in practice, it can be safely ignored in our analysis, because it does not change the transmission power and channel noise characteristics of the uncoded transmission framework.

1) Transmission Power: WHT is an orthonormal transform, so we have $\text{Power}(\mathbf{z}) = \mathbf{z}^{\text{T}}\mathbf{z} = (W\mathbf{y})^{\text{T}}(W\mathbf{y}) = \mathbf{y}^{\text{T}}W^{\text{T}}W\mathbf{y} = \mathbf{y}^{\text{T}}\mathbf{y} = \text{Power}(\mathbf{y})$. Therefore, WHT does not change the total transmission power of the signal.

2) Channel Noise: If the transmission of $\mathbf{z}$ is formulated by $\hat{\mathbf{z}} = \mathbf{z} + \mathbf{n}$, the effect of noise can be equivalently formulated by $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}'$, where $\mathbf{n}' = W^{-1}\mathbf{n}$. When $\mathbf{n}$ is white Gaussian noise, $\mathbf{n}'$ is also white Gaussian noise with the same statistics. To see this, just check $\text{Cov}(W^{-1}\mathbf{n}) = \sigma_{\text{n}}^2 I$ when $\text{Cov}(\mathbf{n}) = \sigma_{\text{n}}^2 I$. Therefore, we can regard the signal path from $\mathbf{y}$ to $\hat{\mathbf{y}}$ in Fig. 4 as a virtual AWGN channel, which is statistically identical with the underlying actual AWGN channel.

III. Decorrelation, Energy Diversity and Transform Gain for Uncoded Transmission

As mentioned earlier, uncoded transmission does not mean that the signal is unprocessed before transmission. In SoftCast [37], the image signal is transformed from spatial domain to frequency domain before being sent out using the framework in Fig. 4. In fact, utilizing a proper transform turns out to be very critical for the efficiency of SoftCast. In this section, we reveal the advantage of employing a decorrelation transform in uncoded visual communication, and introduce a quantitative measurement for it. 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 in Signal Representation

To see how the data activity $H(\mathbf{x})$ of an image can be reduced by a transform, we first study the mathematical property of the definition (16). Obviously, $f(x) = \sqrt{x}$ is a strictly upper convex function. For upper convex functions in general, we have the following remarks:

Remark 1: For any $\lambda_1 \leqslant \lambda_2$, $f(\lambda_1) + f(\lambda_2) > f(\lambda_1 - \epsilon) + f(\lambda_2 + \epsilon)$ holds for any $\epsilon > 0$.

Remark 2: If $\lambda_1 + \lambda_2$ is constant, $f(\lambda_1) + f(\lambda_2)$ achieves its maximum value only when $\lambda_1 = \lambda_2$.

Remark 3: If $\lambda_1 + \lambda_2$ is constant, $f(\lambda_1) + f(\lambda_2)$ strictly decreases with $\eta(\lambda_1, \lambda_2) = |\log(\frac{\lambda_1}{\lambda_2})|$.

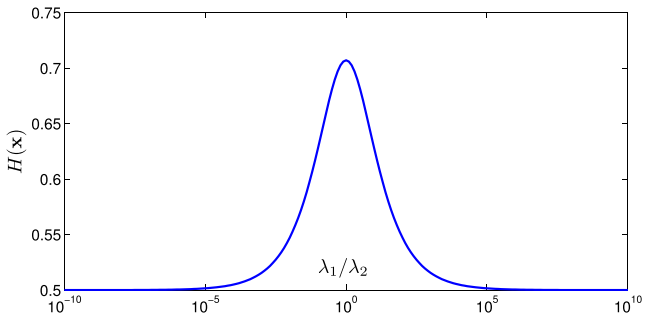


Fig. 5. $H(\mathbf{x})$ of two-variable random source $\mathbf{x} = (x_1, x_2)$ with $E[x_1^2] = \lambda_1$ and $E[x_2^2] = \lambda_2$, subject to $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$ and $\lambda_1 + \lambda_2 = \lambda$. $H(\mathbf{x}) = \sqrt{\lambda/2}$ is reached when $\lambda_1 = \lambda_2$. $H(\mathbf{x}) = \sqrt{\lambda}/2$ is reached when $\lambda_1 = \lambda$ or $\lambda_2 = \lambda$. In this figure $\lambda = 1$.

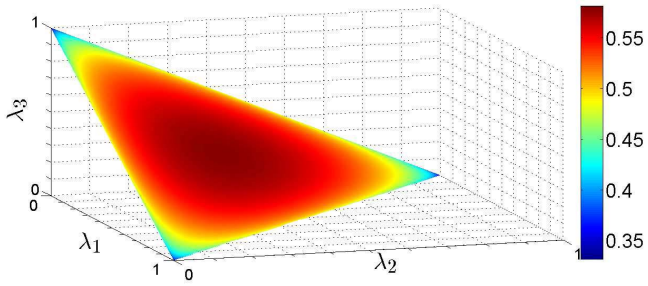


Fig. 6. $H(\mathbf{x})$ of three-variable random source $\mathbf{x} = (x_1, x_2, x_3)$ with $E[x_i^2] = \lambda_i$, subject to $\lambda_i \geqslant 0$ and $\lambda_1 + \lambda_2 + \lambda_3 = \lambda$. $H(\mathbf{x}) = \sqrt{\lambda/3}$ is reached when $\lambda_1 = \lambda_2 = \lambda_3$. $H(\mathbf{x}) = \sqrt{\lambda}/3$ is reached when $\lambda_1 = \lambda$ or $\lambda_2 = \lambda$ or $\lambda_3 = \lambda$. In this figure $\lambda = 1$.

Remark 1 can be easily proved by considering the fact that $f'(x)$ is strictly decreasing. *Remark 2* and *Remark 3* can be derived from *Remark 1*. These remarks mean that, when the sum of λ_1 and λ_2 is constant, the more diversified λ_1 and λ_2 are, the smaller $f(\lambda_1) + f(\lambda_2)$ is. Now we define $\lambda_i = E[x_i^2]$ and study how $H(\mathbf{x})$ varies with different energy distribution specified by $\{\lambda_i\}$. Fig. 5 illustrates the $H(\mathbf{x})$ of two-variable random source $\mathbf{x} = (x_1, x_2)$. It is clear in Fig. 5 that higher diversity in signal energy distribution corresponds to a lower value in $H(\mathbf{x})$.

These conclusions can be extended to the general case of N-variable random source. If $\sum_i^N \lambda_i = \lambda$ is constant, $\sum_i^N f(\lambda_i)$ achieves its maximum value only when all λ_i are equal. The more diversified these λ_i values are, the smaller $\sum_i^N f(\lambda_i)$ is. Fig. 6 illustrates the $H(\mathbf{x})$ value of three-variable random source with different energy distribution.

Furthermore, the achievable range of $H(\mathbf{x})$ can be derived, as in *Remark 4*. This determines the theoretical upper bound of performance improvement one may achieve by transforming the signal from one representation to another, as we shall discuss in the next subsection.

Remark 4: Subject to $\sum_1^N \lambda_i = \lambda$, it holds that $\sqrt{\lambda}/N \leqslant H(\mathbf{x}) \leqslant \sqrt{\lambda/N}$. The lower bound $H(\mathbf{x}) = \sqrt{\lambda}/N$ is reached when all energy falls into a single element x_k, i.e., $\lambda_k = \lambda, \lambda_{i \neq k} = 0$. The upper bound $H(\mathbf{x}) = \sqrt{\lambda/N}$ is reached when the source has uniform energy distribution among elements, i.e., $\lambda_i = \lambda/N, i = 1, 2, \ldots, N$.

B. Transform Gain for Uncoded Transmission

Natural image signals typically exhibit very strong correlation among the neighboring pixels. Such statistical redundancy is commonly exploited in image/video coding by decomposing the signal using decorrelation transform. When the signal is stationary and its covariance matrix is known, *Karhunen-Loève Transform* (KLT) is the optimal transform to use. In practical schemes, DCT is widely used instead as a good approximation of KLT [66].

The typical image decorrelation transforms, including KLT, DCT and DWT, etc., are orthogonal or approximately orthogonal. Therefore, they do not change the total energy of a signal. What they change, however, is the distribution of energy among the signal elements. It is well recognized that, after decorrelation, the signal energy would be compacted to a small part of coefficients – i.e., a small number of coefficients become large while most other coefficients become extremely close to zero.

Based on the discussions in Section III-A, we infer that a decorrelation transform can reduce the data "activity" of a signal, which subsequently leads to higher efficiency in uncoded transmission and better quality in reconstruction.

For an orthonormal transform $\mathcal{T} : X(i) \rightarrow F(u)$, we define the ***transform gain*** (in the context of uncoded transmission) of $\mathcal{T}$ by

$$G(X|\mathcal{T}) = 20\log_{10}\left(\frac{H(X)}{H(F)}\right). \tag{17}$$

To be concrete,

$$G(X|\mathcal{T}) = 20\log_{10}\left(\frac{\sum_i \sqrt{\lambda_X(i)}}{\sum_u \sqrt{\lambda_F(u)}}\right), \tag{18}$$

where $\lambda_X(i) = E[X(i)^2]$ and $\lambda_F(u) = E[F(u)^2]$ represent the energy of signal elements in the spatial domain and the transform domain, respectively.

Based on the *Remark 4* in Section III-A, we have.

Remark 5: For N-element random signals, the upper bound of transform gain is $20\log_{10}(\sqrt{N})$ dB in uncoded transmission. This bound can be reached only if the signal has extremely strong correlation.

C. Some Numerical Results

In this section, we evaluate the transform gain in uncoded transmission and show some numerical results. We first consider first-order Markov (Markov-1) process, a simple but widely used model for natural images. For a stationary Markov-1 process $\{X_i\}$ of length N with correlation coefficient ρ, the covariance matrix is given by

$$C_X = \sigma_X^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{N-1} \\ \rho & 1 & \rho & \ldots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \ldots & 1 \end{bmatrix}. \tag{19}$$

The KLT transform result of X, i.e. $F = \mathrm{KLT}(X)$, has a covariance matrix

$$C_F = \sigma_X^2 \cdot \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}. \tag{20}$$

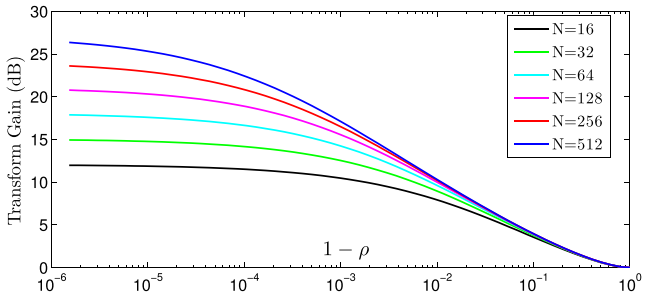


Fig. 7. Transform gain of $1 \times N$ KLT for uncoded transmission of 1D first-order Markov process. Note that the x-axis here shows the value of $1-\rho$.

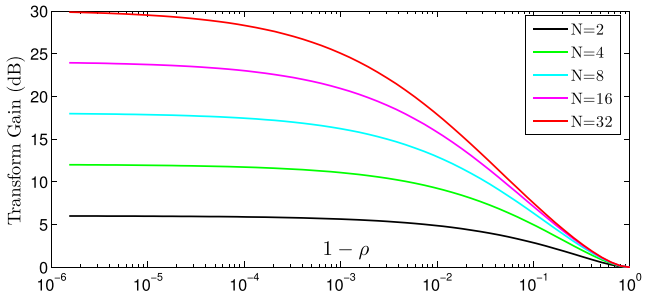


Fig. 8. Transform gain of $N \times N$ KLT for uncoded transmission of 2D Markov process.

The diagonal entries $\{\lambda_i\}$ correspond to the eigenvalues of the Toeplitz matrix in (19). They can be determined by (see [67]):

$$\lambda_i = \frac{1-\rho^2}{1-2\rho\cos\omega_i+\rho^2}, \quad i = 1, 2, \ldots, N, \tag{21}$$

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

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin\omega}{\cos\omega - 2\rho + \rho^2\cos\omega}. \tag{22}$$

In this case, $\lambda_X(i) = \sigma_X^2$, $\lambda_F(i) = \sigma_X^2\lambda_i$ so that the transform gain is

$$G(X|\text{KLT}) = 20\log_{10}\left(\frac{N}{\sum_{i=1}^{N}\sqrt{\lambda_i}}\right), \tag{23}$$

Fig. 7 illustrates the transform gain of KLT on Markov-1 process. We note that the transform gain (or reduction in data activity) depends on the correlation inside the signal. When the signal has strong correlation (i.e. as $\rho \to 1$), the benefit of transform becomes significant. As $\rho \to 0$, the benefit of transform diminish to 0. We also note that the transform gain depends on the transform size N. In general, a KLT transform with larger size provides higher transform gain, when the signal is highly correlated. When $\rho \to 1$, the transform gain increases by 3.01dB ($= 20\log_{10}\sqrt{2}$ dB) when N is doubled. This is because a transform of larger size can exploit the signal correlation at a larger scale. But for $\rho \to 0$, larger transform size N does not bring larger transform gain.

The above observations can be extended to signals of higher dimension. For example, we consider a two-dimensional Markov process, in which the correlation coefficient for any two pixels of distance d is ρ^d. Fig. 8 illustrates the transform gain of $N \times N$ KLT for such 2D Markov process. We can see, when $\rho \to 1$, the transform gain increases by about 6.02dB when N is doubled. When the correlation coefficient ρ is smaller, the transform gain becomes smaller.

We also investigated the transform gain on natural images and real video sequences. In particular, we investigated the influence of different transform sizes. The experimental details and results will be presented in Section V.

IV. Discussions on Related Issues

The analysis of transform gain presented in Section III depends on an assumption that the signal energy distribution is known. Ideally, the sender choose the scaling factor g_i for each element individually according to $E[x_i^2]$ – recall (10). However, the receiver also needs the scaling factors, for the purpose of correct decoding – recall (3). Therefore, the $\{g_i\}$ (or the $\{E[x_i^2]\}$) values must be shared between the sender and the receiver.

Of course, sending one g_i (or $E[x_i^2]$) for each element is impractical since the communication overhead is overwhelming. As the simplest solution, SoftCast groups the coefficients in the same transform band as a data chunk and perform scaling at chunk level. The coefficients in each chunk choose the same g_i value, based on the mean energy of that chunk. This will more or less reduce the transform gain, depending on the size of chunks and the data energy diversity within each chunk. This aspect was firstly discussed in [68].

To fully utilize the benefit of transform but avoid significant overhead at the same time, we need a way to compactly describe the energy distribution in the transform domain. Two energy modeling solutions were proposed in [45] and [50]. In fact, these preliminary works indicate that the actual energy distribution of transform coefficients can be well approximated by certain models using a very small set of meta data. As a result, the theoretical transform gain can be mostly achieved at a low cost of meta data. This aspect has been further studied in our another work (see [69] for more information).

V. Experimental Results

This section reports experimental results. We first show some numerical results based on the theoretical analysis presented in Section III. Then we simulate the uncoded transmission scheme using various transform options, and verify the theoretical results. Both natural images and videos will be considered in our experiments.

A. Results on Natural Images

This section investigates the uncoded transmission of natural images. We first evaluate the transform gain based on formulation (18). In particular, we investigate the influences of using different transform sizes. DCT is used as the decorrelation transform in this paper. The 512×512 images, *Lena*, *Peppers*, *Elaine*, *Fishingboat*, *Barbara*, etc., and the Kodak 768×512 images, *Airplane*, *Cap*, *Parrot*, *Girl*, *Sailboats2*, *Window*, etc., and the 1280×1600 images[2] from the JPEG

[2]Cropped to 1280×1536 in our experiments.

TABLE I

THE TRANSFORM GAIN (dB) OF 2D-DCT IN UNCODED TRANSMISSION OF NATURAL IMAGES (IDEAL CASE)

Test Image		Transform Size of 2D-DCT									Max
Size	Name	1×1	2×2	4×4	8×8	$2^4\times2^4$	$2^5\times2^5$	$2^6\times2^6$	$2^7\times2^7$	$2^8\times2^8$	
512x512	Lena	0.00	5.59	10.82	15.40	18.94	21.28	22.39	22.63	**22.72**	22.72
	Barbara	0.00	5.14	9.94	14.16	17.30	19.21	**19.94**	19.81	19.36	19.94
	Peppers	0.00	5.53	10.67	15.05	18.30	20.22	20.97	21.24	**21.38**	21.38
	Elaine	0.00	5.53	10.74	15.30	18.93	21.47	23.05	23.95	**24.17**	24.17
	Fishingboat	0.00	5.44	10.44	14.65	17.81	19.84	20.89	**21.35**	21.10	21.35
768x512	Airplane	0.00	5.74	11.23	16.22	20.36	23.26	24.86	**25.33**	25.27	25.33
	Cap	0.00	5.64	10.96	15.67	19.42	21.87	23.04	**23.28**	23.22	23.28
	Parrot	0.00	5.69	11.09	15.97	19.98	22.72	24.11	**24.39**	23.15	24.39
	Girl	0.00	5.49	10.56	14.92	18.31	20.50	21.52	**21.68**	21.47	21.68
	Sailboats2	0.00	5.60	10.84	15.44	19.06	21.46	22.70	23.07	**23.09**	23.09
	Window	0.00	5.56	10.69	15.05	18.38	20.41	21.49	**21.88**	21.47	21.88
1280x1536	Bike	0.00	5.50	10.52	14.77	17.92	19.88	20.90	**21.23**	21.08	21.23
	Woman	0.00	5.20	9.83	13.73	16.65	18.47	19.32	**19.38**	18.96	19.38
	Building	0.00	5.53	10.60	14.87	17.91	19.90	21.52	22.75	**23.27**	23.27
	Flower	0.00	5.68	11.05	15.81	19.60	22.12	23.44	23.94	**24.19**	24.19
Average		0.00	**5.52**	**10.66**	**15.14**	**18.59**	**20.84**	**22.01**	**22.39**	**22.26**	**22.39**

TABLE II

THE TRANSFORM GAIN (dB) OF 2D-DCT IN UNCODED TRANSMISSION OF NATURAL IMAGES (PRACTICAL CASE)

Test Image		Transform Size of 2D-DCT									Max
Size	Name	1×1	2×2	4×4	8×8	$2^4\times2^4$	$2^5\times2^5$	$2^6\times2^6$	$2^7\times2^7$	$2^8\times2^8$	
512x512	Lena	0.00	5.34	10.12	14.06	16.99	18.96	20.17	20.94	**21.74**	21.74
	Barbara	0.00	4.69	8.80	12.12	14.44	15.97	16.97	17.64	**18.29**	18.29
	Peppers	0.00	5.18	9.84	13.62	16.35	18.16	19.23	19.99	**20.75**	20.75
	Elaine	0.00	5.37	10.27	14.39	17.56	19.80	21.30	22.35	**23.24**	23.24
	Fishingboat	0.00	5.12	9.61	13.21	15.83	17.57	18.65	19.38	**20.09**	20.09
768x512	Airplane	0.00	5.37	10.30	14.53	17.82	20.11	21.61	22.57	**23.46**	23.46
	Cap	0.00	5.29	10.02	13.93	16.89	18.86	20.07	20.86	**21.60**	21.60
	Parrot	0.00	5.29	10.06	14.10	17.16	19.24	20.50	21.40	**21.94**	21.94
	Girl	0.00	5.13	9.65	13.31	16.02	17.81	18.93	19.67	**20.28**	20.28
	Sailboats2	0.00	5.19	9.75	13.57	16.45	18.36	19.61	20.45	**21.22**	21.22
	Window	0.00	5.12	9.59	13.16	15.81	17.54	18.63	19.36	**19.99**	19.99
1280x1536	Bike	0.00	5.00	9.26	12.60	14.99	16.55	17.50	18.10	**18.53**	18.53
	Woman	0.00	4.80	8.88	12.17	14.51	15.97	16.81	17.29	**17.60**	17.60
	Building	0.00	5.15	9.67	13.36	16.12	17.96	19.15	19.98	**20.61**	20.61
	Flower	0.00	5.50	10.52	14.78	18.04	20.25	21.61	22.40	**22.91**	22.91
Average		0.00	**5.17**	**9.76**	**13.53**	**16.33**	**18.21**	**19.38**	**20.16**	**20.82**	**20.82**

XR evaluation dataset [70], [71], *Bike*, *Woman*, *Building*(p10) and *Flower*(p30), are used as test images.

We first consider the ideal case that the energy statistic of every transform coefficient is perfectly known. The results are presented in Table I. We can see that, for typical natural images, the 2×2, 4×4, 8×8, 16×16 and 32×32 DCT transforms can provide about 5.5dB, 10.7dB, 15.2dB, 18.8dB and 21.1dB gain, respectively. Obviously, the transform gain usually increases with the transform size. In fact, DCT of large size can provide up to $19.9 \sim 25.3$ dB gain. We note that $2^7 \times 2^7$ or $2^8 \times 2^8$ may be the best transform size for most natural images. Frame-size DCT also produces similar results.

We then consider the more practical case that the energy statistics of transform coefficients are only known at band level, as discussed in Section IV. The transform gain results in this setting are presented in Table II. As in the above ideal case, we can get similar conclusion that DCT transform can provide significant gain and the gain typically increases with the transform size. Comparing Table I and Table II, we note that the transform gain in this setting is relatively lower (e.g. by 1.26dB in average for the case of $2^8 \times 2^8$ DCT) than that in the ideal case. This loss is due to the limited accuracy of transform coefficient statistics. But the transform gain in this setting is still up to more than 20dB in average.

Now we simulate the actual transmission procedure and verify the above theoretical results. In the transmission scheme, the sender performs DCT transform, scaling and WHT transform, and the receiver performs inverse WHT transform, LLSE decoding and inverse DCT transform. Several transform size options are considered, including 2×2, 4×4, $8 \times 8, \ldots$, and 128×128. For these cases, the power optimization is performed at band level, i.e., each band chooses the same scaling factor g_i. We also considered a special case (noted by "DirectTx") that no DCT transform is applied. The channel is an AWGN channel and the SNR range considered in this experiment is $0 \sim 20$dB. For each of the scheme and channel configuration, we simulate the transmission process for 20 times and measure the average performance.

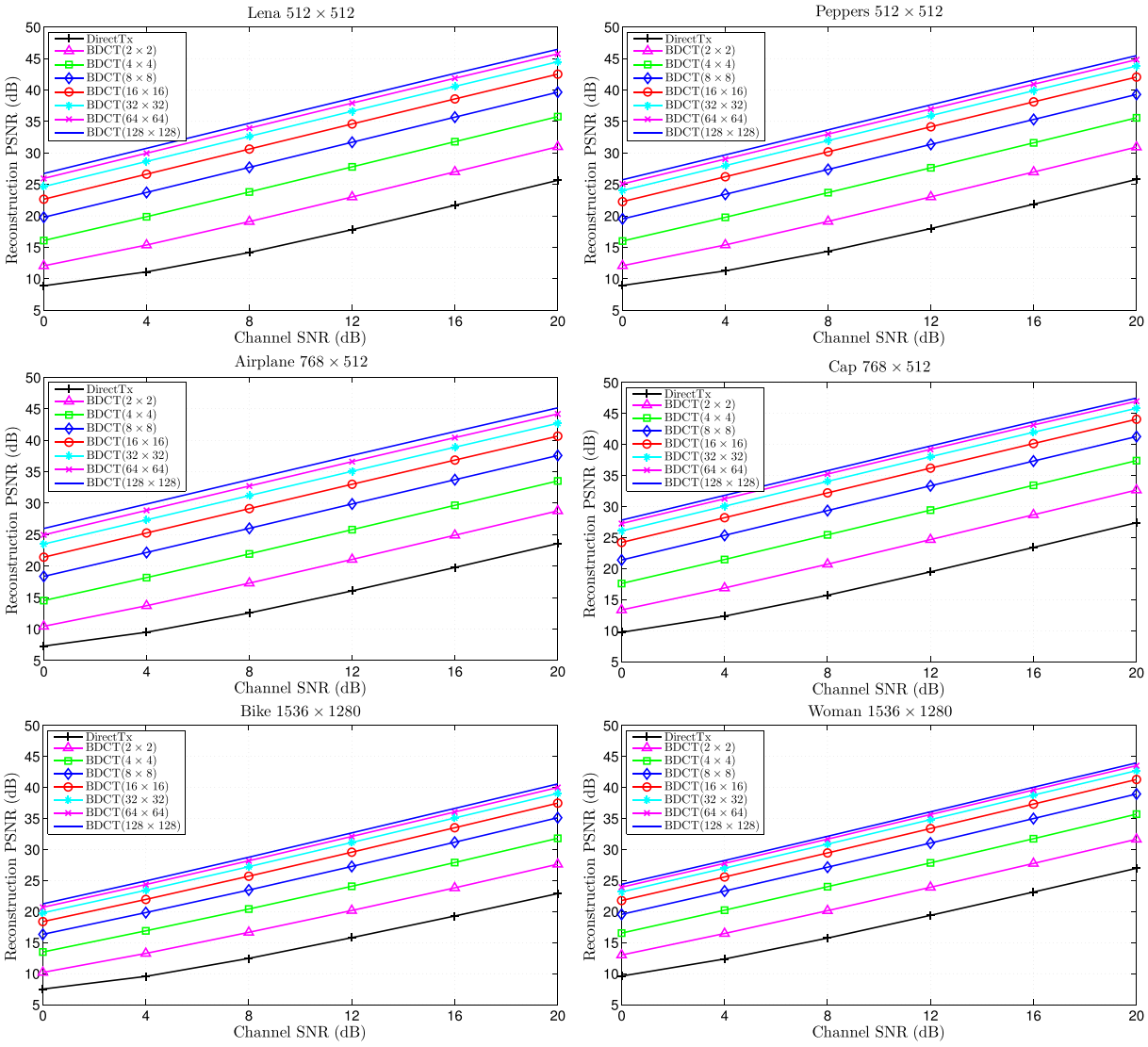


Fig. 9. The PSNR results of simulated uncoded transmission for natural images.

The PSNR results of the simulated transmissions are shown in Fig. 9. These results confirm the linear relationship between PSNR and CSNR, as expressed in Eqn. (15), at least for middle and high CSNR region. The non-linearity at low CSNR region is due to the clipping of decoded pixel values to [0, 255]. The results in Fig. 9 also verify the theoretical analysis results presented in Table II. By comparing the $2^m \times 2^m$ DCT scheme with the DirectTx scheme, we see that the performance gap is highly consistent with the transform gain reported in Table II.

Now we evaluate the subjective performance of the scheme. To clearly show the characteristics of artifacts in reconstructed images, a low channel SNR of 4dB is considered. The reconstruction results for *Lena* are illustrated in Fig. 10. It's clear that the transmission scheme using larger DCT size produces better reconstruction quality, due to the higher transform gain. From the reconstructed images, we notice the different reconstruction error patterns caused by different transform configurations. The DirectTx scheme leaves the white channel noises untouched in spatial domain. The block-DCT based schemes, however, introduce blocking artifacts in the decoded images, since the channel noise is imposed on the transform coefficients instead of on the image pixels. The blocking artifacts are more noticeable when small blocks are used for DCT. When larger block sizes are used, the blocking effect become almost invisible. More results of reconstructed images are shown in Fig. 11 and Fig. 12.

B. Results on Video Sequences

Now we consider uncoded transmission of video sequences. Compared with images, video sequences not only have spatial correlation within each frame, but also have temporal correlation among adjacent frames. For spatial-temporal decorrelation, 3D-DCT is used in SoftCast [37]–[39]. Motion compensated temporal filtering (MCTF) [31]–[35] is later adopted in WaveCast [43] to improve the efficiency of temporal transform based on the motion information. In this part,

Fig. 10. The reconstructed images (of *Lena*) in uncoded image transmission (with CSNR=4dB). From left (1^{th} row) to right (2^{nd} row): Original, DirectTx, Block-DCT with size $2^1 \times 2^1$, $2^2 \times 2^2$, $2^3 \times 2^3$, …, $2^8 \times 2^8$. The PSNR goes from 11.09dB, 15.35dB to 31.45dB. Please enlarge the figure to observe details.

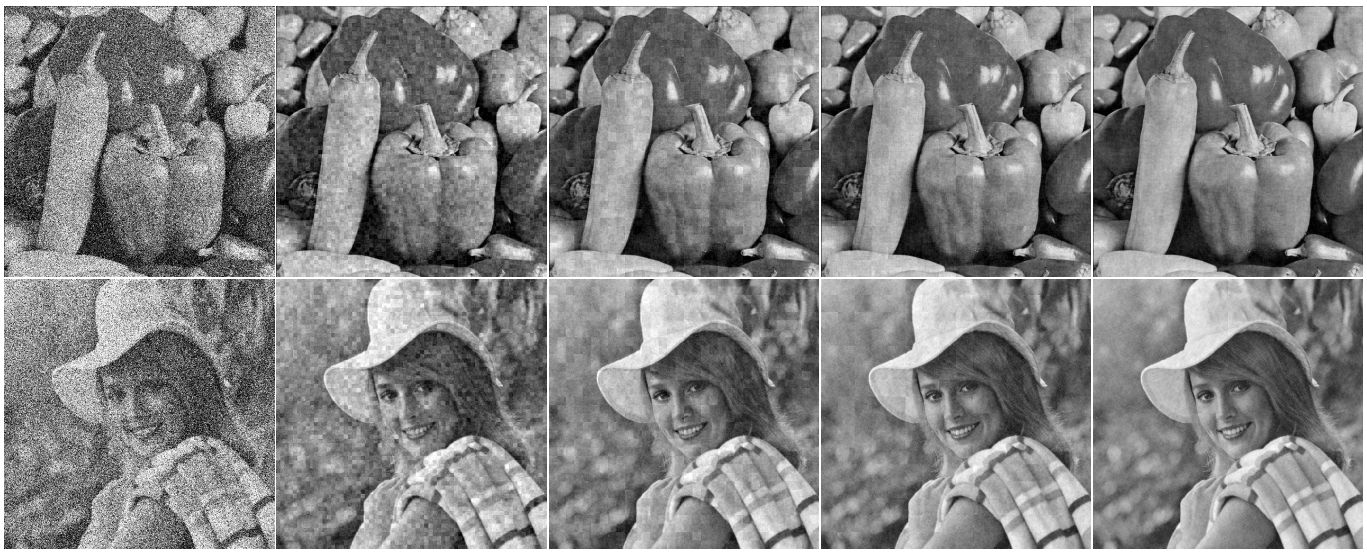


Fig. 11. The reconstructed images (of *Peppers* and *Elaine*) in uncoded image transmission (with CSNR=4dB). From left to right: DirectTx, Block-DCT with size $2^3 \times 2^3$, $2^4 \times 2^4$, $2^5 \times 2^5$ and $2^8 \times 2^8$. Please enlarge the figure to observe details.

we focus on 3D-DCT and evaluate the performance of different spatial- and temporal-transform sizes. But just keep in mind that the same thing can actually also be done for the case of MCTF to analyze its performance. The only difference is that the transform gain of MCTF might be much larger than that of 3D-DCT, since its temporal transform is motion aligned.

In this experiment, we use the CIF videos, *Akiyo*, *Coast-guard*, *Container*, *Foreman*, *Mobile*, *Mother*, *News* and *Silent* as the test sequences. The videos are divided into GOPs (group of pictures) and processed in a GOP-by-GOP manner. The frames in each GOP as a whole are first transformed by DCT in the temporal direction, and then each of the resulting frames is processed by spatial DCT transform. Only the first 128 frames of each video sequence are used in our experiments.

As have done for natural images, we first evaluate the transform gain based on formulation (18), and we assume that the energy statistic of every coefficient is perfectly known. We consider several different spatial transform sizes (2×2, 4×4, 8×8, 16×16, 32×32, and frame-size 352×288, but show only three of them in Table III due to limited space) and six different temporal transform sizes (i.e. GOP size $L = 1, 2, 4, 8, 16, 32$). We see from Table III that, compared with 2D-DCT (i.e. the case that $L = 1$), 3D-DCT provides additional gain by exploiting the inter-frame correlation using DCT in the temporal direction. We note that the transform gain usually increases with the GOP size. We also notice an interesting point – a large spatial transform size is usually desirable when the GOP size is small (e.g. when $L = 1$); however, it is no longer preferred when L is large

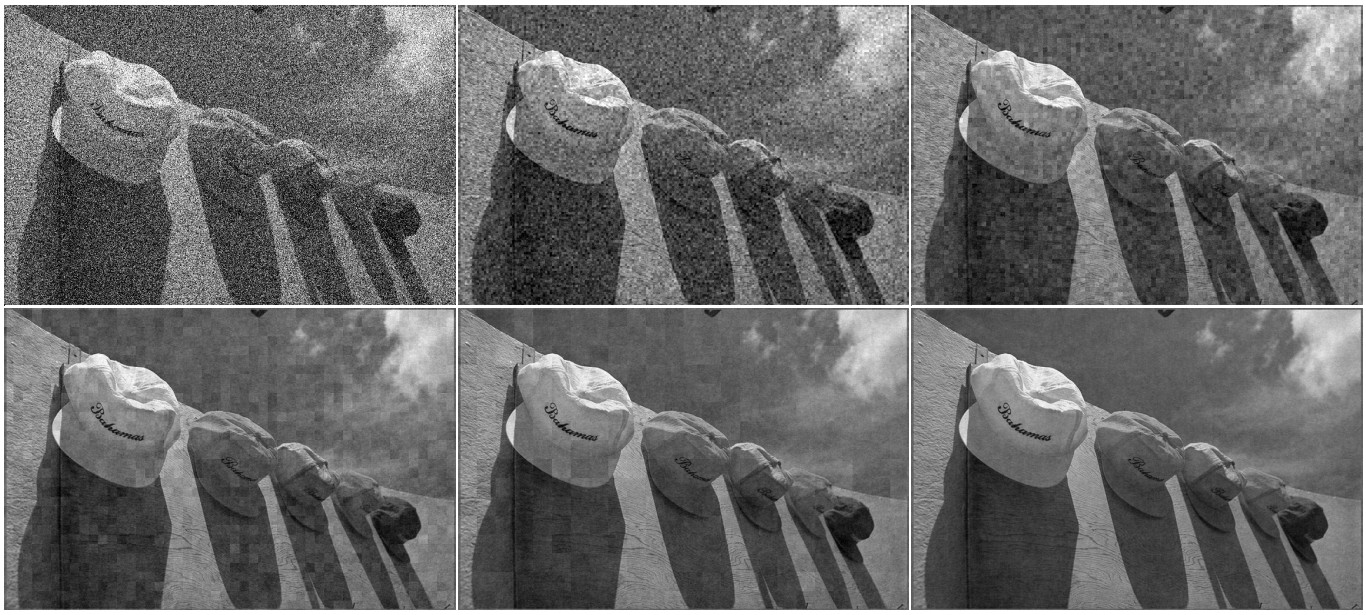


Fig. 12. The reconstructed images (of *Cap*) in uncoded image transmission (with CSNR= 4dB). From left to right: DirectTx, Block-DCT with size $2^2 \times 2^2$, $2^3 \times 2^3$, $2^4 \times 2^4$, $2^5 \times 2^5$ and $2^8 \times 2^8$. Please enlarge the figure to observe details.

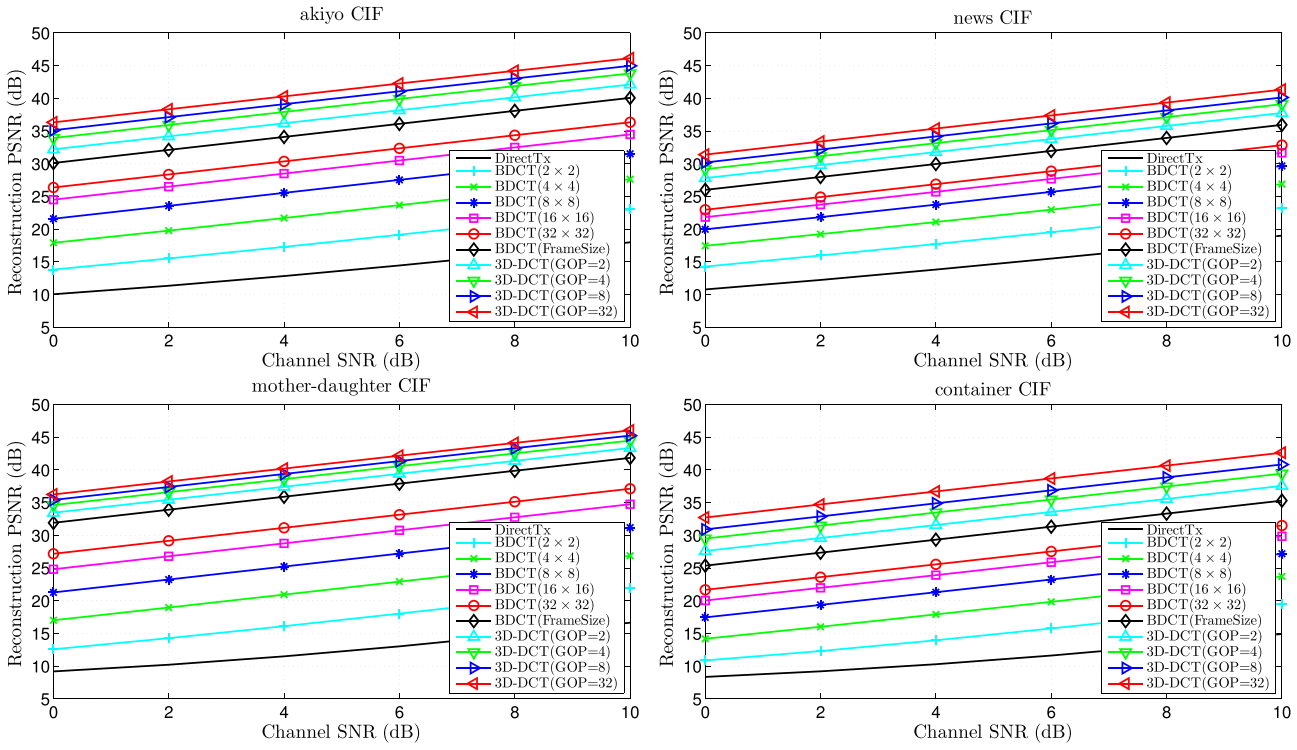


Fig. 13. The PSNR results of simulated uncoded transmission for video sequences.

(e.g. when $L = 32$). To see this, just note that the transform gain of 3D-DCT (32×32, $L = 32$) is 26.49dB but that of 3D-DCT (352×288, $L = 32$) is only 23.10dB, for the *News* sequence.

It is natural to conjecture that the different frames produced by the temporal DCT may require different spatial transform size to obtain optimal performance. To clarify this issue, for each frame after temporal transform, we evaluate the spatial transform gain for every possible spatial transform size. The results are reported in Table IV. It is clear that, a large spatial DCT is preferred for the temporal-DC frame (with frame index 0 in a GOP), but a DCT of size 8×8 or 16×16 is enough for temporal-highpass frames (e.g., with frame index $1 - 31$ in a GOP).

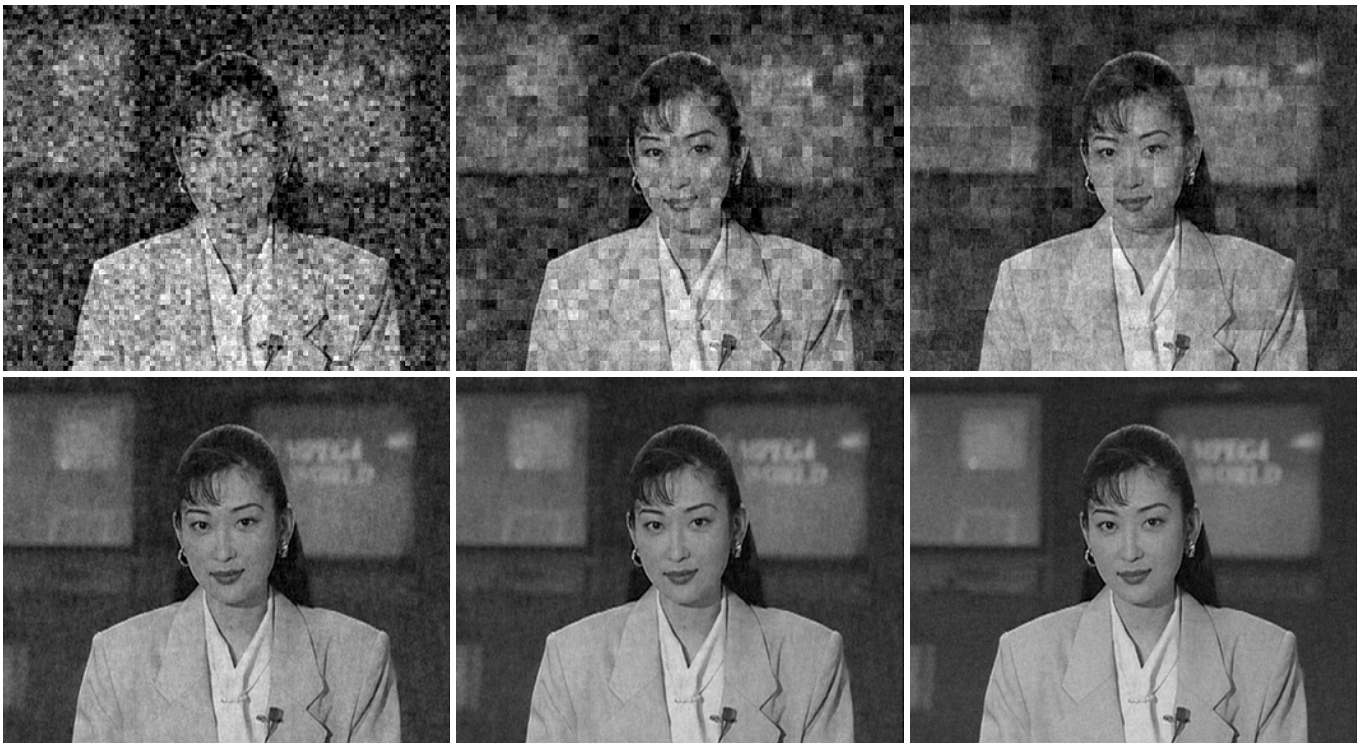


Fig. 14. The reconstructed frames (of *Akiyo*) in uncoded video transmission (with CSNR=0dB). From left to right: GOP=1(4×4), GOP=1(8×8), GOP=1(16×16), GOP=1(Adaptive), GOP=2(Adaptive), GOP=32(Adaptive). Please enlarge the figure to observe details.

TABLE III

THE TRANSFORM GAIN (dB) OF 3D-DCT FOR UNCODED TRANSMISSION OF CIF VIDEO SEQUENCES (FS: FRAME SIZE)

Video Sequence	Spatial DCT Size	Temporal DCT Size (GOP Size L)					
		L=1	L=2	L=4	L=8	L=16	L=32
Akiyo 352×288	8×8	15.73	18.59	21.38	24.02	26.49	28.63
	32×32	21.45	24.11	26.63	28.87	30.80	32.23
	FS	21.73	24.01	26.01	27.53	28.69	29.31
Coastguard 352×288	8×8	13.81	15.85	17.55	18.90	19.81	20.42
	32×32	18.64	20.10	21.23	22.06	22.49	22.70
	FS	19.88	20.87	21.61	22.10	22.35	22.44
Foreman 352×288	8×8	15.64	18.12	20.33	22.21	23.64	24.69
	32×32	21.75	23.69	25.28	26.49	27.26	27.75
	FS	23.70	25.01	25.97	26.64	27.01	27.25
Mobile 352×288	8×8	11.65	13.50	15.11	16.37	17.18	17.51
	32×32	14.29	15.68	16.91	17.95	18.69	19.00
	FS	15.07	16.08	16.86	17.44	17.81	17.98
Mother 352×288	8×8	16.13	18.85	21.45	23.86	25.99	27.84
	32×32	22.89	25.26	27.40	29.24	30.69	31.80
	FS	25.37	27.23	28.76	29.92	30.73	31.30
Silent 352×288	8×8	15.01	17.72	20.31	22.67	24.77	26.47
	32×32	20.48	22.90	25.10	27.00	28.55	29.67
	FS	22.29	24.10	25.61	26.80	27.72	28.36
PeopleOnStreet 2560×1600	8×8	15.27	17.40	19.14	20.38	21.15	21.60
	32×32	20.43	21.93	23.13	23.95	24.38	24.52
	FS	22.08	22.77	23.27	23.64	23.88	24.05
Traffic 2560×1600	8×8	15.25	17.70	19.91	21.82	23.31	24.34
	32×32	20.85	22.81	24.50	25.90	26.92	27.51
	FS	22.25	23.29	24.04	24.61	25.03	25.31
BQTerrace 1920×1080	8×8	15.13	17.49	19.64	21.53	23.07	24.08
	32×32	20.57	22.42	23.99	25.30	26.28	26.71
	FS	21.32	22.74	23.84	24.72	25.35	25.44
Cactus 1920×1080	8×8	15.35	17.65	19.71	21.47	22.91	23.95
	32×32	21.23	22.93	24.40	25.56	26.46	26.99
	FS	22.63	23.33	23.84	24.21	24.51	24.69

This motivates us to employ a kind of adaptive 3D-DCT in practical schemes. In Table V, we evaluate the transform gain of 3D-DCT using adaptive spatial transform size. Comparing

TABLE IV

THE TRANSFORM GAIN (dB) OF SPATIAL DCT ON THE GOP FRAMES TRANSFORMED BY TEMPORAL DCT (GOP SIZE $L = 32$, FS: FRAME SIZE)

Video Sequence	Frame Index	Spatial DCT Transform Size							
		1×1	2×2	4×4	8×8	$2^4 \times 2^4$	$2^5 \times 2^5$	$2^6 \times 2^6$	FS
Akiyo	0	0.00	5.72	11.16	16.04	19.85	22.26	22.09	20.30
	1	0.00	1.92	2.93	3.20	2.47	1.46	-0.27	-3.34
	$2 \sim 3$	0.00	1.84	2.80	3.10	2.38	1.43	-0.16	-2.75
	$4 \sim 7$	0.00	1.27	1.95	2.11	1.42	0.51	-0.96	-3.61
	$8 \sim 15$	0.00	0.83	1.35	1.51	0.97	0.16	-1.11	-3.75
	$16 \sim 31$	0.00	0.45	0.77	0.87	0.51	-0.10	-1.05	-3.41
8 CIF Sequences	0	0.00	5.68	11.07	15.89	19.78	22.38	22.85	21.09
	1	0.00	3.33	5.66	7.00	7.52	7.51	6.85	4.54
	$2 \sim 3$	0.00	2.68	4.51	5.59	6.01	5.99	5.44	3.17
	$4 \sim 7$	0.00	1.92	3.25	4.12	4.46	4.43	3.97	1.67
	$8 \sim 15$	0.00	1.19	2.16	2.85	3.15	3.13	2.83	0.52
	$16 \sim 31$	0.00	0.82	1.49	1.99	2.20	2.12	1.98	-0.40

Table III and Table V, we can see that the adaptive 3D-DCT performs better than the simple 3D-DCT. The transform gain can be up to 32dB for slow motion sequences such as *Akiyo*, and the average gain is 27.8dB for the eight tested sequences. Note that the performance for *Mobile* and *Coastguard* is remarkably lower than that for other sequences since the two videos have relatively higher motion and rich textures.

Now we simulate the actual video transmission procedure and verify the above theoretical results. The scheme for video is similar to that for image, except that 3D-DCT is used instead of 2D-DCT. The PSNR results are shown in Fig. 13. We can clearly see the performance improvement as the spatial transform size increases (for GOP size $L = 1$) and as the GOP size increases (adaptive spatial transform size is used in this case). The subjective performance is illustrated in Fig. 14. The

TABLE V

THE TRANSFORM GAIN (dB) OF 3D-DCT FOR UNCODED TRANSMISSION OF CIF VIDEO SEQUENCES, USING ADAPTIVE SPATIAL TRANSFORM SIZE (64×64 FOR THE TEMPORAL-DCT DC FRAME AND 16×16 FOR ALL TEMPORAL-DCT HIGHPASS FRAMES)

Video Sequence	Spatial DCT Size	Temporal DCT Size (GOP Size L)					
		L=1	L=2	L=4	L=8	L=16	L=32
Akiyo	adaptive	21.22	23.93	26.51	28.85	30.90	**32.50**
Coastguard	adaptive	18.94	20.41	21.51	22.25	22.56	**22.66**
Foreman	adaptive	22.38	24.29	25.84	26.98	27.61	**27.95**
Mobile	adaptive	14.43	15.88	17.18	18.21	18.86	**19.02**
Mother	adaptive	23.04	25.41	27.58	29.46	30.93	**32.07**
Silent	adaptive	21.06	23.48	25.70	27.59	29.10	**30.15**
PeopleOnStreet	adaptive	21.35	22.83	23.86	24.39	**24.50**	24.34
Traffic	adaptive	22.01	23.93	25.52	26.71	27.43	**27.69**
BQTerrace	adaptive	22.09	23.48	24.59	25.38	**25.78**	25.72
Cactus	adaptive	22.67	24.23	25.51	26.47	27.12	**27.47**
Average	adaptive	**20.92**	**22.79**	**24.38**	**25.63**	**26.48**	**26.96**

performance gain in PSNR is confirmed by the improvement in visual quality of the reconstructed video frames.

VI. CONCLUSION AND DISCUSSIONS

Uncoded transmission is a promising way to provide robust and elegant visual communication in wireless and mobile scenarios, especially for wireless video broadcast. It skips entropy coding in source coding and all the subsequent discrete-state operations in channel coding and modulation. However, it can still achieve competitive performance and meanwhile provide graceful quality degradation for a wide channel SNR range.

Uncoded transmission is not unprocessed transmission. In fact, signal decorrelation and transmission power optimization are two key modules that determine the efficiency of uncoded transmission. Both coded and uncoded transmission employ transform for signal decorrelation. The most fundamental difference between coded and uncoded transmission is that coded transmission turns the coefficients into equally important bits and then allocates the communication resources (bandwidth and transmission power) to each coded bit, while uncoded transmission allocates the bandwidth and transmission power directly to the coefficients.

This paper presents a theoretical analysis for the transform gain in uncoded transmission. Our analysis reveals that the energy distribution among signal elements is critical for the efficiency of uncoded transmission. A decorrelation transform can potentially bring significant performance gain by boosting the energy diversity in signal representation. More importantly, we found that the transform gain can be realized only when both the sender and the receiver have good knowledge on the energy diversity in signal elements and fully exploit this diversity for optimal protection [69]. Therefore, the efficiency in describing and sharing the knowledge of energy diversity in signal elements is an important problem for uncoded transmission.

Due to the lossy nature of uncoded transmission, noises always appear in the received signals. Besides using proper transmission power to control the noise level in the received signals, employing appropriate signal estimation (e.g. denoising) approaches [72]–[77] is also an important aspect. This paper only considered a simple linear estimator, since advanced nonlinear estimators may make the power-distortion analysis of reconstruction process too complicated or even infeasible. The optimized reconstruction for uncoded image/video transmission is an interesting topic to study in future works [62].

This paper only considered AWGN channel in order to simplify the analysis. In more realistic settings, in addition to channel noise, channel fading also causes transmission errors. How the fading effect influences the transform gain and changes the optimal power allocation strategy is also an interesting problem that needs further study in future works [49], [54].

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[2] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.

[3] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[4] W. P. Pennebaker and J. L. Mitchell, *JPEG: Still Image Compression Standard*. New York, NY, USA: Van Nostrand, 1993.

[5] D. Taubman and M. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Norwell, MA, USA: Kluwer, 2001.

[6] F. Dufaux, G. J. Sullivan, and T. Ebrahimi, "The JPEG XR image coding standard," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 195–199, Nov. 2009.

[7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[8] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[9] P. G. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Process. Lett.*, vol. 4, no. 7, pp. 189–191, Jul. 1997.

[10] V. Chande and N. Farvardin, "Progressive transmission of images over memoryless noisy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 850–860, Jun. 2000.

[11] H. Yousefi'zadeh, H. Jafarkhani, and F. Etemadi, "Transmission of progressive images over noisy channels: An end-to-end statistical optimization framework," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 220–231, Apr. 2008.

[12] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.

[13] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–386, Mar. 2001.

[14] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 357–372, Mar. 2001.

[15] V. M. Stankovic, R. Hamzaoui, Y. Charfi, and Z. Xiong, "Real-time unequal error protection algorithms for progressive image transmission," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1526–1535, Dec. 2003.

[16] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 121–131, Feb. 2003.

[17] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1890–1901, Nov. 2005.

[18] A. Nosratinia, J. Lu, and B. Aazhang, "Source-channel rate allocation for progressive transmission of images," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 186–196, Feb. 2003.

[19] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Efficient channel code rate selection algorithms for forward error correction of packetized multimedia bitstreams in varying channels," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 240–248, Apr. 2004.

[20] R. Hamzaoui, V. Stankovic, and Z. Xiong, "Optimized error protection of scalable image bit streams [advances in joint source-channel coding for images]," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 91–107, Nov. 2005.

[21] A. H. Banihashemi and A. Hatam, "A distortion optimal rate allocation algorithm for transmission of embedded bitstreams over noisy channels," *IEEE Trans. Commun.*, vol. 56, no. 10, pp. 1581–1584, Oct. 2008.

[22] D. Taubman and J. Thie, "Optimal erasure protection for scalably compressed video streams with limited retransmission," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1006–1019, Aug. 2005.

[23] R. Xiong, D. Taubman, and V. Sivaraman, "Optimal PET protection for streaming scalably compressed video streams with limited retransmission based on incomplete feedback," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2382–2395, Sep. 2010.

[24] R. Xiong, D. S. Taubman, and V. Sivaraman, "PET protection optimization for streaming scalable videos with multiple transmissions," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4364–4379, Nov. 2013.

[25] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[26] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[27] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[28] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[29] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 332–344, Mar. 2001.

[30] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[31] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 572–588, Sep. 1994.

[32] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[33] S.-J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[34] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[35] R. Xiong, J. Xu, F. Wu, and S. Li, "Barbell-lifting based 3-D wavelet coding scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1256–1269, Sep. 2007.

[36] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[37] S. Jakubczak, H. Rahul, and D. Katabi, "SoftCast: One video to serve all wireless receivers," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. MIT-CSAIL-TR-2009-005, 2009.

[38] S. Jakubczak, H. Rahul, and D. Katabi, "One-size-fits-all wireless video," in *Proc. ACM HotNets*, New York, NY, USA, 2009.

[39] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2011, pp. 289–300.

[40] X. Fan, F. Wu, and D. Zhao, "D-cast: DSC based soft mobile video broadcast," in *Proc. Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, 2011, pp. 226–235.

[41] X. Fan, F. Wu, D. Zhao, O. C. Au, and W. Gao, "Distributed soft video broadcast (DCAST) with explicit motion," in *Proc. Data Compress. Conf.*, Apr. 2012, pp. 199–208.

[42] X. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "ParCast: Soft video delivery in MIMO-OFDM WLANs," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 233–244.

[43] X. Fan, R. Xiong, F. Wu, and D. Zhao, "WaveCast: Wavelet based wireless video broadcast using lossy transmission," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Nov. 2012, pp. 1–6.

[44] R. Xiong, F. Wu, J. Xu, and W. Gao, "Performance analysis of transform in uncoded wireless visual communication," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 1159–1162.

[45] 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 *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Nov. 2013, pp. 1–6.

[46] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, "Cactus: A hybrid digital-analog wireless video communication system," in *Proc. 16th ACM Int. Conf. Modeling, Anal. Simulation Wireless Mobile Syst. (MSWiM)*, Barcelona, Spain, Nov. 2013, pp. 273–278.

[47] X. Fan, F. Wu, D. Zhao, and O. C. Au, "Distributed wireless visual communication with power distortion optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1040–1053, Jun. 2013.

[48] R. Xiong, H. Liu, S. Ma, X. Fan, and W. Gao, "G-CAST: Gradient based image SoftCast for perception-friendly wireless visual communication," in *Proc. IEEE Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2014, pp. 133–142.

[49] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust uncoded video transmission over wireless fast fading channel," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr./May 2014, pp. 73–81.

[50] Z. Song, R. Xiong, X. Fan, S. Ma, and W. Gao, "Transform domain energy modeling of natural images for wireless SoftCast optimization," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Melbourne, VIC, Australia, Jun. 2014, pp. 1114–1117.

[51] Z. Song, R. Xiong, S. Ma, X. Fan, and W. Gao, "Layered image/video softcast with hybrid digital-analog transmission for robust wireless visual communication," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Chengdu, China, Jul. 2014, pp. 1–6.

[52] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, "ParCast+: Parallel video unicast in MIMO-OFDM WLANs," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2038–2051, Nov. 2014.

[53] F. Wu, X. Peng, and J. Xu, "LineCast: Line-based distributed coding and transmission for broadcasting satellite images," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1015–1027, Mar. 2014.

[54] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust linear video transmission over Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2790–2801, Aug. 2014.

[55] H. Cui, R. Xiong, C. Luo, Z. Song, and F. Wu, "Denoising and resource allocation in uncoded video transmission," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 102–112, Feb. 2015.

[56] S. Yun, D. Kim, X. Lu, and L. Qiu, "Optimized layered integrated video encoding," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2015, pp. 19–27.

[57] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 331–345, Feb. 2014.

[58] L. Yu, H. Li, and W. Li, "Wireless cooperative video coding using a hybrid digital–analog scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 436–450, Mar. 2015.

[59] X. Zhao, H. Lu, C. W. Chen, and J. Wu, "Adaptive hybrid digital-analog video transmission in wireless fading channel," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, p. 1, doi: 10.1109/TCSVT.2015.2444753.

[60] X. Kong and H. Lu, "Joint power allocation for hybrid digital-analog transmission in relay networks," in *Proc. Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[61] A. Wang, B. Zeng, and H. Chen, "Wireless multicasting of video signals based on distributed compressed sensing," *Signal Process., Image Commun.*, vol. 29, no. 5, pp. 599–606, May 2014.

[62] R. Xiong, J. Zhang, F. Wu, and W. Gao, "High quality image reconstruction via non-local collaborative estimation for wireless image/video softcast," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2542–2546.

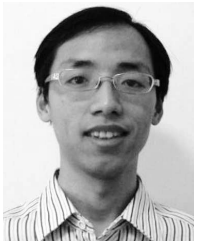
[63] X. Fan, R. Xiong, D. Zhao, and F. Wu, "Layered soft video broadcast for heterogeneous receivers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1801–1814, Nov. 2015.

[64] D. He, C. Luo, C. Lan, F. Wu, and W. Zeng, "Structure-preserving hybrid digital-analog video delivery in wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1658–1670, Sep. 2015.

[65] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Scalable video multicast for MU-MIMO systems with antenna heterogeneity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, p. 1, doi: 10.1109/TCSVT.2015.2430651.

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

[67] W. Ray and R. Driver, "Further decomposition of the Karhunen–Loève series representation of a stationary random process," *IEEE Trans. Inf. Theory*, vol. 16, no. 6, pp. 663–668, Nov. 1970.
[68] R. Xiong, X. Fan, F. Wu, and W. Gao, "Analysis of uncoded image communication over noisy channels with unequal protection," in *Proc. SPIE Electron. Imag. Vis. Inf. Process. Commun. IV*, Feb. 2013, p. 866603.
[69] R. Xiong, J. Zhang, F. Wu, and W. Gao, "Power-distortion optimization for uncoded linear-transformed transmission of images and videos," *IEEE Trans. Image Process.*, to be published.
[70] F. De Simone, L. Goldmann, V. Baroncini, and T. Ebrahimi, "Subjective evaluation of JPEG XR image compression," *Proc. SPIE*, vol. 7443, p. 74430L, Aug. 2009.
[71] *JPEG XR Subjective Assessment: Core Experiments Description 4.1*, document WG1N5001, ISO/IEC JTC1/SC29/WG1 (JPEG), AIC AhG, 2009.
[72] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Image compressive sensing recovery via collaborative sparsity," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 380–391, Sep. 2012.
[73] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, Jun. 2014.
[74] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.
[75] X. Zhang, R. Xiong, W. Lin, S. Ma, J. Liu, and W. Gao, "Video compression artifact reduction via spatio-temporal multi-hypothesis prediction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6048–6061, Dec. 2015.
[76] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, p. 1, doi: 10.1109/TCSVT.2016.2527181.
[77] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "CONCOLOR: Constrained non-convex low-rank model for image deblocking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1246–1259, Mar. 2016.



Ruiqin Xiong (M'08) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 2001, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2007.

He was a Research Intern with Microsoft Research Asia, from 2002 to 2007, and a Senior Research Associate with the University of New South Wales, Australia, from 2007 to 2009. He joined the Peking University in 2010, where he is currently a Professor. He has authored over 80 technical papers in refereed international journals and conference proceedings. His research interests include statistical image modeling, image and video processing, video compression, and multimedia communication.

Feng Wu (M'99–SM'06–F'13) received the B.S. degree in electrical engineering from Xidian University, in 1992, and the M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, in 1996 and 1999, respectively. He was a Principle Researcher and a Research Manager with Microsoft Research Asia. He is currently a Professor with the University of Science and Technology of China.

He has authored or co-authored over 200 high quality papers (including several dozens of IEEE TRANSACTIONS papers) and top conference papers on MOBICOM, SIGIR, CVPR, and ACM MM. He has 77 granted U.S. patents. His research interests include image and video compression, media communication, and media analysis and synthesis. His 15 techniques have been adopted into international video coding standards. As a co-author, he received the best paper award in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY in 2009, PCM 2008, and SPIE VCIP 2007. He serves as an Associate Editor in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and several other international journals. He also received the IEEE Circuits and Systems Society 2012 Best Associate Editor Award. He also serves as the TPC Chair in MMSP 2011, VCIP 2010, and PCM 2009, and Special Sessions Chair in ICME 2010 and ISCAS 2013.

Jizheng Xu (M'07–SM'10) received the B.S. and M.S. degrees in computer science from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China.

He joined Microsoft Research Asia, Beijing, China, in 2003, where he is currently a Lead Researcher. He has authored or co-authored over 100 papers in refereed conferences and journals. He has over 30 U.S. patents granted or pending in image and video coding. His current research interests include image and video representation, media compression, and communication.

Dr. Xu has been an Active Contributor to ISO/MPEG and ITU-T video coding standards. He has over 30 technical proposals adopted by H.264/AVC, H.264/AVC scalable extension, High Efficiency Video Coding (HEVC), HEVC range extension, and HEVC screen content coding standards. He was the Chair and Co-Chair with the ad-hoc group of exploration on wavelet video coding in MPEG, and various technical ad-hoc groups in JCT-VC, e.g., on screen content coding, parsing robustness, lossless coding. He was the Co-Organizer and Co-Chair of special sessions on scalable video coding, directional transform, and high quality video coding at various conferences. He also served as the Special Session Co-Chair of the IEEE International Conference on Multimedia and Expo 2014.

Xiaopeng Fan (M'09) received the B.S. and M.S. degrees from the Harbin Institute of Technology (HIT), in 2001 and 2003, respectively, and the Ph.D. degree from The Hong Kong University of Science and Technology, in 2009.

He was with the Intel China Software Laboratory, Shanghai, China, as a Software Engineer, from 2003 to 2005. He joined the School of Computer Science and Technology, HIT, in 2009, where he is currently a Professor. He has authored or co-authored over 80 technical journal and conference papers. His research interests include image/video processing and wireless communication.

Chong Luo (M'04–SM'14) received the B.S. degree from Fudan University, Shanghai, China, in 2000, the M.S. degree from the National University of Singapore, in 2002, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2012.

She has been with Microsoft Research Asia, Beijing, China, since 2003, where she is currently a Lead Researcher with the Internet Media Group. Her research interests include wireless networking, wireless sensor networks, and multimedia communications.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from The University of Tokyo, Japan, in 1991. He was a Professor of Computer Science with the Harbin Institute of Technology, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, from 1996 to 2006. He is currently a Professor of Computer Science with Peking University, Beijing, China.

He has authored extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

Dr. Gao served or serves on the Editorial Board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON AUTONOMOUSMENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*. He was the Chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ISCAS, ICME, and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.