Visually Weighted Compressive Sensing: Measurement and Reconstruction

Hyungkeuk Lee, Heeseok Oh, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

Abstract—**Compressive sensing (CS) makes it possible to more naturally create compact representations of data with respect to a desired data rate. Through wavelet decomposition, smooth and piecewise smooth signals can be represented as sparse and compressible coefficients. These coefficients can then be effectively compressed via the CS. Since a wavelet transform divides image information into layered blockwise wavelet coefficients over spatial and frequency domains, visual improvement can be attained by an appropriate perceptually weighted CS scheme. We introduce such a method in this paper and compare it with the conventional CS. The resulting visual CS model is shown to deliver improved visual reconstructions.**

Index Terms—**Compressive sensing (CS), modified block compressive sensing, visual compressive sensing, wavelet transform, weighted compressive sampling matching pursuit.**

I. Introduction

DESPITE the rapid development of efficient wireless communication systems, Quality of Service (QoS) management is still a critical issue because of the need to accommodate large increases in the number of mobile users and in the volume of wireless video traffic. One solution may be to compress high rate image and video data more efficiently while ensuring that quality is maintained. Compressive Sensing (CS) is a stable and robust technique that allows for the sub-sampling of data at a given data rate: 'compressive sampling' or 'compressive sensing' at rates smaller than the Nyquist sampling rate [1], [2]. The theory of CS states that if a signal is sparse in a transform domain, then it can be reconstructed exactly from a small set of linear measurements using tractable optimization algorithms. However, since a random measurement may have deficient rank, the matrix must satisfy the so-called Restricted Isometry Property (RIP).

Thus far, most research on CS has focused on increasing the robustness and reducing the computational complexity of recovery algorithms [3]. Yet, while theoretical studies have demonstrated the stability of CS, specific examples of successful and practical applications remain elusive.

Furthermore, it is not certain that applying CS can deliver competitive performance relative to standardized state-of-the-art post-acquisition image and video compression methods, such as the Joint Photographic Experts Group (JPEG), Moving Picture Experts Group (MPEG), and H.264 codecs. The authors of [4] highlighted the potential and limitations of CS as compared to traditional image compression methods, noting that CS is far less efficient than the state-of-the-art JPEG2000 standard. Depending on the sparsity of an image, CS can reduce both the number of measurements needed to reconstruct it and the approximation error. In other words, when the application data are sparse signals that are highly correlated with other data, then superiority of the CS technique over other compression methods can be demonstrated. Suitable likely applications include Distributed Source Coding (DSC) [5], [6] or Multiple Description Coding (MDC) [7], since their target data are generally sparse. If a data set or signal is assumed sparse or compressible, it is a candidate for exploiting the benefits of CS. Wavelet decomposition, which is amenable to analysis over a wide range of applications [8], allows smooth and piecewise smooth signals to be represented by sparse and compressible sets of coefficients and its hierarchical structure provides an excellent framework for capturing global signal features and for generating naturally layered data. Hence the wavelet representation not only provides sparsity, but also admits a natural scalable representation. These features have motivated researchers to apply hierarchical weighting to CS measurement and reconstruction.

The main contribution we make here is as follows: we apply data obtained from a wavelet representation to CS, where the measurement side emphasizes 'visually important' CS data. We leverage ideas from our previous work, where we demonstrated an improvement in compressed image quality by utilizing the concept of 'visual entropy', which is the expected number of bits that is required to map image information onto human visual coordinates [9] expressed in the wavelet domain [10], [11]. Likewise, using the fact that visual entropy aims to maximize visual information delivery within a constrained data rate, we are able to show that applying visual entropy to CS yields visually improved images. Specifically, we use the layered structure of the image wavelet coefficients to automatically assign different compression ratios to different

subbands, thereby enhancing the measurement efficiency and the visual reconstruction performance. We show that the proposed CS approach compares favorably with conventional CS under a visual importance criterion. We also suggest how the proposed technique can be used in practice for 'sensing compressively'-whereby blocks of wavelet coefficients are multiplied by different measurement matrices and a visual weighting matrix.

A number of additional contributions derive from this work: 1) full utilization of sparsity from wavelet transform data, 2) a weight function defined on a visual sensitivity model, 3) image information maximization on CS measurements and reconstruction, and 4) an example of an application of the proposed method. Towards this end, we construct a visual CS framework comprised of a modified block compressive sensing (BCS) model at the CS measurement side and a weighted Compressive Sampling Matching Pursuit (wCoSaMP) model at the CS reconstruction side. Gains acquired from visual CS are analyzed by modeling the wavelet transform data using a Hidden Markov Model (HMM) [12].

II. PROPOSED IDEA IN CS AREA

A. Backgrounds of Compressive Sensing

Given orthonormal basis vectors $\{\psi_i\}_{i=1}^N$, we can represent each image signal $x \in \mathbb{R}^N$ using N coefficients $\{\alpha_i\}_{i=1}^N$ as $x = \sum_{i=1}^N \alpha_i \psi_i$. By stacking the ψ_i as columns in the $N \times N$ matrix $\mathbf{\Psi}$, x can be expressed $x = \mathbf{\Psi}\alpha$ with $\alpha \in \mathbb{R}^N$. Using the transformation $\mathbf{\Psi}$, which could be a wavelet transform, the signal x can be represented as K-sparse data only if $K \ll N$ entries of α are nonzero. If the transform function $\mathbf{\Psi}$ is a discrete wavelet transform, the image signal may be written

$$x(s) = \sum_{i \in \mathbb{Z}^2} \alpha_{j_0,i} \psi_{j_0,i}^{LL}(s) + \sum_{l \in \mathcal{L}} \sum_{j \geq j_0} \sum_{i \in \mathbb{Z}^2} \alpha_{j,i}^l \psi_{j,i}^l(s) \quad (1)$$

where we use the following notations:

s	index over the spatial domain;
j	index of wavelet decomposition level;
i	index of location over the wavelet subbands;
l	index of subbands excluding the LL subband, i.e., $l \in \mathcal{L} = \{LH, HL, HH\}$ where L and H denote the low and high frequencies, respectively;
ψ^{LL}	scaling function;
ψ^l	shifted & dilated functions for the lth subband;
$\alpha_{j_0,i}$	scaling coefficients;
$\alpha_{j,i}^l$	wavelet coefficients.

For brevity, (1) can be expressed in vector form as

$$x = \mathbf{\Psi}\alpha = \alpha_0 \psi_0 + \sum_{\forall j \in \mathcal{J}} \sum_{\forall i \in \mathcal{N}_j} \alpha_{j,i} \psi_{j,i} \quad (2)$$

where $\mathcal{J}$ is the set of all subbands, (i.e., it includes all of the decomposed subbands) and $\mathcal{N}_j$ is the index set of the jth subband. Therefore, the 1-D length-N vector α is constructed by reconciling the wavelet coefficients as

$$\alpha = \begin{bmatrix} \alpha_0 & \alpha_{1,1} & \alpha_{1,2} & \cdots \end{bmatrix}^T = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_N \end{bmatrix}^T \in \mathbb{R}^{N \times 1}. \quad (3)$$
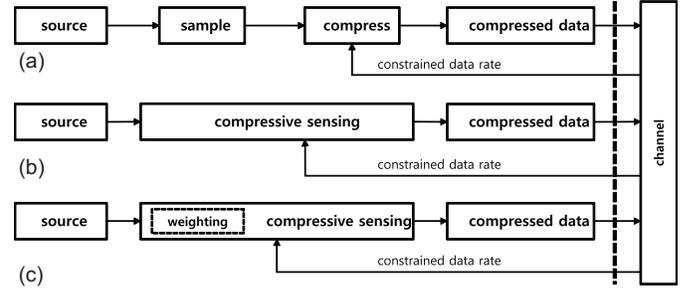

Fig. 1. Conceptual comparison of visually weighted CS with conventional compression and CS.

Assumed that $\hat{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Psi}^{-1}$, x can be measured as

$$u = \hat{\mathbf{\Phi}}x = \mathbf{\Phi}\mathbf{\Psi}^{-1}x = \mathbf{\Phi}\alpha \quad (4)$$

where u is a measurement vector, $u \in \mathbb{R}^M$, and $\mathbf{\Phi}$ is an $M \times N$ random measurement matrix, $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ to satisfy the Restricted Isometry Property (RIP) [13].

Although some natural and man-made signals are not strictly sparse, they can be approximated as sparse signals. Such approximated signals are called "compressible signals". Consider a signal α whose coefficients, when sorted in order of decreasing magnitude, decay according to the power law

$$|\alpha_{\mathcal{I}(i)}| \leq G \cdot i^{-1/r}, \quad i = 1, \ldots N \quad (5)$$

where $\mathcal{I}$ indexes the sorted coefficients, $G = |\alpha_{\mathcal{I}(1)}|$ is the first amplitude of the sorted coefficients and r is a constant that is dependent on the slope of the decay according to the index i. As long as a rapid decay of their coefficients is demonstrated, such signals could be well-approximated by K-sparse signals.

Let $\alpha_K \in \Sigma_K$ be the best K-term approximation of α. Such an approximation is obtained by keeping only the first K terms in $\alpha_{\mathcal{I}(i)}$ from (5), i.e.,

$$\alpha_K = \arg \min_{\bar{\alpha} \in \Sigma_K} \|\alpha - \bar{\alpha}\|_p \quad (6)$$

where the ℓ_p norm of the vector α is defined as $\|\alpha\|_p = \left(\sum_{i=1}^N |\alpha_i|^p\right)^{1/p}$ for $1 < p < \infty$. The error of this approximation in the ℓ_p norm is denoted

$$\sigma_K(\alpha)_p \triangleq \|\alpha - \alpha_K\|_p. \quad (7)$$

Then, for $r < p$, the integration of K sorted coefficients will yield

$$\sigma_K(\alpha)_p \leq (rs)^{1/p} GK^{-s}, \quad (8)$$

with $s = \frac{1}{r} - \frac{1}{p}$ [14].

B. Applying Visual Weighting to CS Based on Visual Sensitivity

Figure 1 depicts a conceptual comparison of visually weighted CS with conventional compression and CS. As shown in Fig. 1(a), the encoder side of the general compression model comprises of "sample", which includes possible transformation to a different domain, and "compression", which includes quantization and presumed entropy coding. In the compression step, source information is lost as a function of bandwidth and the channel condition. By contrast with the
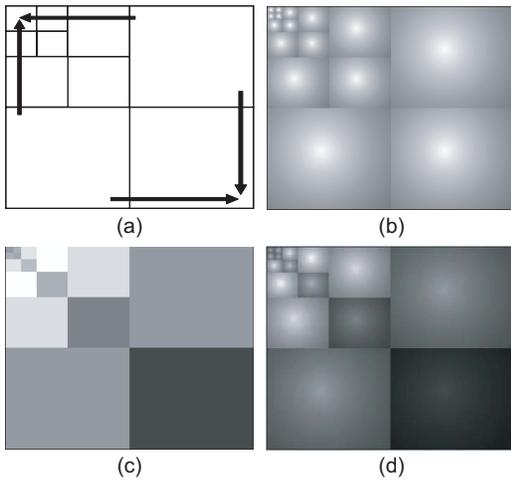
Fig. 2. Visual weights defined over the wavelet domain. (a) Wavelet decomposition. (b) Spatial domain w_i^s weighting. (c) Frequency domain w_i^f weighting. (d) Overall wavelet domain w_i weighting. Brightness: amount of visual information [10].

Fig. 3. Modified BCS method developed in this paper.

general model, the compressive sensing stage in Fig. 1(b) aims to sample the source using a random matrix at a compressed rate, i.e., to sample the source compressively with respect to the number of measurements M. As shown in Fig. 1(c), the proposed visually weighted CS method compressively samples the source as a function of the number of measurements, while discarding visually unimportant information. Visually important information is then packed into measurements of the same length as in conventional (unweighted) CS.

CS can measure wavelet coefficient data uniformly via a random matrix. However, since the wavelet transform decomposes data over space and frequency, each wavelet coefficient may possess a different degree of visual importance from other coefficients regardless of its own amplitude. Fig. 2 depicts a visual sensitivity model in the wavelet domain where brightness represents the visual weight, i.e., brighter regions are deemed to be more perceptually important. The visual weight of the ith wavelet coefficient w_i is characterized by two visual components: a spatial weight w_i^s and a frequency weight w_i^f. The weights well described in [10]. If the visual weights are applied over the wavelet domain in conjunction with the sampling processing of CS, better visual performance can be obtained by the CS measurement and reconstruction sides.

III. VISUAL CS MEASUREMENTS AND RECONSTRUCTION

A. Modified BCS

The authors of [15] proposed a fast block-based sampling algorithm for fast CS of natural images, called Block Compressive Sensing (BCS). In BCS, an original image is divided into small blocks and each block is sampled independently using the same measurement operation with $\tilde{M} = \lfloor \frac{M}{B} \rfloor$, where B is the number of blocks. In BCS, the measurements are easily stored and utilized because of the small block size and it is also unnecessary for the encoder to wait until the entire image is measured.
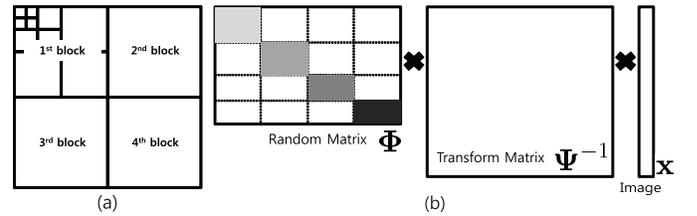
Borrowing from the concept of BCS, whereby the random matrix is divided into blocks, we define a modified BCS method expressed in the wavelet domain that evolves with the characteristics of each subband. Under this different assumption, we show how to obtain the theoretical bound when the statistics of the wavelet coefficients are utilized. Moreover, when visual weighting is included, the attained performance bound yields deeper insights the performance of CS with respect to perceptual quality. Modified BCS proceeds as follows: first, wavelet data from an image is partitioned into four blocks of size $N/4$ as shown in Fig. 3 (a), and CS measurements are obtained for each block using four random sensing matrices. Using $\mathbf{u} = \hat{\mathbf{\Phi}}\mathbf{x} = \mathbf{\Phi}\mathbf{\Psi}^{-1}\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}$, measurements from the modified BCS are obtained by

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \mathbf{u}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & & & \\ & \mathbf{\Phi}_2 & & \\ & & \mathbf{\Phi}_3 & \\ & & & \mathbf{\Phi}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \boldsymbol{\alpha}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1\boldsymbol{\alpha}_1 \\ \mathbf{\Phi}_2\boldsymbol{\alpha}_2 \\ \mathbf{\Phi}_3\boldsymbol{\alpha}_3 \\ \mathbf{\Phi}_4\boldsymbol{\alpha}_4 \end{bmatrix}. \quad (9)$$

Although this approach appears to use the two sequential steps-wavelet transform and measurement, the two steps can be simultaneously expressed as the matrix $\hat{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Psi}^{-1}$. From the perspective of conventional CS, the only difference is that the measurement consists of the random and transform matrices. However, the product of the two matrices can be treated as a random matrix, so this approach treated as a subset of general CS and following all the theory inherited from it. This melds seamlessly with the CS philosophy even while using wavelet coefficients having different statistical characteristics. To exploit this modified BCS method in practice, the random matrix should be altered as shown in Fig. 3 (b), where the blank portions of the random matrix are padded with "zero" values.

Let b denote an index of the divided blocks, N_b denote the number of coefficients and M_b denote the number of measurements in each block respectively. For block b, the measurement vector $\mathbf{u}_b$ becomes $\mathbf{u}_b = \mathbf{\Phi}_b\boldsymbol{\alpha}_b$ where $\boldsymbol{\alpha}_b = [\boldsymbol{\alpha}_{b,1}, \ldots, \alpha_{b,n}, \ldots, \alpha_{b,N/4}]^T$ and $\mathbf{\Phi}_b$ is an $M_b \times (N/4)$ random matrix with $M_b = c \cdot K_b \log(N_b/K_b)$. Depending on the number of non-zero coefficients K_b, the dimension of the measurement matrix is determined to satisfy the K_b-RIP. A constant c is applied to all blocks such that $M_b = c \cdot K_b \log(N_b/K_b)$. Since the modified BCS utilizes four quad-random matrices $\mathbf{\Phi}_b \in \mathbb{R}^{M_b \times (N/4)}$, rather than an $M \times N$ matrix, memory efficiency can be improved. This modified BCS strategy provides a simple structure at the measurement side, while also reducing the number of measurements made. Assume that the number of blocks is 2, the sparsities of
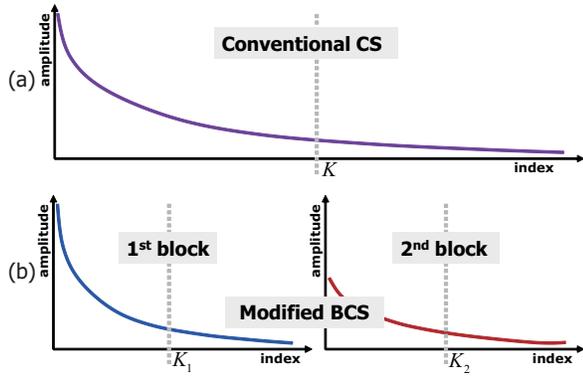
Fig. 4. Sorting of wavelet coefficients with respect to their amplitudes. (a) Conventional CS. (b) Modified BCS with two blocks.

Fig. 5. Modified BCS with visual weights called as visually weighted BCS.

the blocks are K_1 and K_2 and that the decision variable is $\kappa = M / \sum_{\forall b} M_b$ (if $\kappa \geq 1$, then $\sum_{\forall b} M_b \leq M$). To express the gain from using modified BCS in a closed form, we assume that c is constant over all blocks. With $M = c \cdot K \log(N/K)$ and $M_b = c \cdot K_b \log(N_b/K_b)$, κ becomes

$$\kappa = \frac{c \cdot K \log(N/K)}{c\left(K_1 \log(N_1/K_1) + K_2 \log(N_2/K_2)\right)}$$
$$= \frac{(N/K)^K}{(N/2K_1)^{K_1} (N/2K_2)^{K_2}}. \quad (10)$$

Since $K = K_1 + K_2$, (10) becomes

$$= \frac{(2K_1)^{K_1} (2K_2)^{K_2}}{K^K} = \left(\frac{2K_1}{K}\right)^{K_1} \left(\frac{2K_2}{K}\right)^{K_2}.$$

The case $K_1 \geq K_2$ induces $K_1/K \geq 1/2$ and $K_2/K \leq 1/2$. Then (10) becomes

$$= \left(\frac{2K_1}{K}\right)^{K_1} \left(\frac{2K_2}{K}\right)^{K_2} \geq 1. \quad (11)$$

Otherwise, if $K_1 \leq K_2$, (11) still holds. This approximation can be recursively expanded to handle aa larger number of blocks with $K_1 = K_{1,1} + K_{1,2}$ and $K_2 = K_{2,1} + K_{2,2}$. Therefore, the number of measurements required by modified BCS does not exceed the number of measurements required by conventional CS:

$$\sum_{\forall b} M_b \leq M. \quad (12)$$

Even if this assumption is not applicable in general, it can be experimentally demonstrated that Eq. (12) is usually satisfied.

B. Visually Weighted Block Compressive Sensing

Since signals represented in the wavelet domain are not strictly sparse, it is necessary to approximate them as compressible signals for analyzing. The slopes of the wavelet coefficients, after being sorted by amplitude, are shown in Fig. 4(a) and (b) for conventional and modified BCS, respectively. The dotted line across the index axis represents the acquired number of non-zero values, i.e., the K largest coefficients selected from the sorted coefficients as shown in Fig. 4(a). To maintain consistency, non-zero values satisfying $K_1 + K_2 = K$
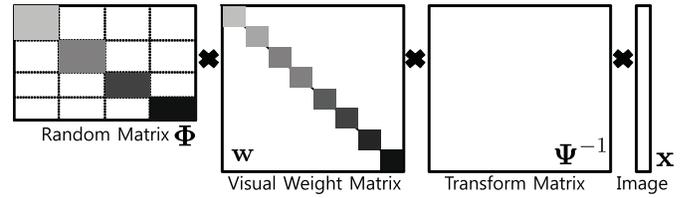
should be chosen. To generate a solution, an optimization problem can be formulated. Non-zero values are selected by minimizing the MSE over each block (this can be interpreted as minimization of the K-term approximation error in (7)). The MSE can be expressed

$$\text{MSE} = \|x - \tilde{x}\|_2^2 = \|\alpha - \tilde{\alpha}\|_2^2 = \sum_{\forall b \in \mathcal{B}} \|\alpha_b - \tilde{\alpha}_b\|_2^2. \quad (13)$$

As shown in Fig. 2, the 1st block is more visually important than the other blocks since the visual function quantifies the extent to which a unit error can affect perception. Thus the weighted MSE is obtained :

$$\text{weighted MSE} = w\|\alpha - \tilde{\alpha}\|_2^2, \quad (14)$$

where w is the diagonal matrix comprised of the visual weight over the wavelet domain, i.e., $w = \text{diag}(w_1, \ldots, w_i, \ldots, w_N)$, and $w_i = w_i^s \cdot w_i^f$.

The measurement matrix for visually weighted BCS is obtained by multiplying the visual weight matrix by the random and transform matrices as shown in Fig. 5. Note that if the visual weight matrix is included in that construction of the measurement matrix, the theoretical analysis derived for conventional CS also seamlessly applied to visually weighted BCS. For a given number of measurements M, the measurements obtained by visually weighted BCS preferentially include visually more important components. Visually weighted BCS also samples the image signal more compressively for a given number of measurements since the measurement matrix is expressed as $u = \Phi_w x = \Phi w \Psi^{-1} x = \Phi w \alpha$.

C. Hidden Markov Wavelet-Tree Model

To deliver a gain in performance by modified BCS, the task of choosing an optimal number of non-zero values is analyzed by modeling the wavelet coefficients as in the following theorem.

Theorem 1: Each wavelet coefficient can be modeled by estimating the power number r in (5):

$$\left|\alpha_{\mathcal{I}(i)}\right| \leq G \cdot i^{-1/r} \leq G \cdot i^{-1/\left(4 - \frac{2N}{G^2}\text{var}(\alpha)\right)}. \quad (15)$$

The detailed proof of Theorem 2 is given in *Appendix*.

Since splitting the wavelet domain into blocks introduces statistical variations between blocks, the modeling of wavelet coefficients for each block should also be modified. HMM have been successfully used to improve the performance of denoising, classification, and segmentation algorithms for wavelet-sparse signals [16], [17].

Statistically, each sparse coefficient can be modeled as one of two types of Gaussian random variables. The first type of variable has a large variance due to the large amplitude of non-zero coefficients. The other type of variable has a small variance due to small or zero coefficients. Let L and S be stages of large and small variance, respectively. These two components are distinguished by associating each wavelet coefficient $\alpha_{j,i}$ with a hidden state $S_{j,i} \in \{S, L\}$. Depending on the state $S_{j,i}$ of $\alpha_{j,i}$, the pdf of $\alpha_{j,i}$ is conditionally different as follows

$$f\left(\alpha_{j,i}|S_{j,i}=S\right) = \mathcal{N}\left(0, \mathrm{var}(\boldsymbol{\alpha}_{S,j,i})\right),$$
$$f\left(\alpha_{j,i}|S_{j,i}=L\right) = \mathcal{N}\left(0, \mathrm{var}(\boldsymbol{\alpha}_{L,j,i})\right) \quad (16)$$

where $\mathrm{var}(\boldsymbol{\alpha}_{S,j,i}) \leq \mathrm{var}(\boldsymbol{\alpha}_{L,j,i})$. Assume that the probability of each state is $\mathrm{Pr}\left(S_{j,i}=S\right) = p_{j,i}^S$ or $\mathrm{Pr}\left(S_{j,i}=L\right) = p_{j,i}^L$. The term $f\left(\alpha_{j,i}\right)$ is then obtained as follows

$$f\left(\alpha_{j,i}\right) = \mathrm{Pr}\left(S_{j,i}=S\right) \cdot \mathcal{N}\left(0, \mathrm{var}(\boldsymbol{\alpha}_{S,j,i})\right)$$
$$+ \mathrm{Pr}\left(S_{j,i}=L\right) \cdot \mathcal{N}\left(0, \mathrm{var}(\boldsymbol{\alpha}_{L,j,i})\right)$$
$$= p_{j,i}^S \cdot g\left(\alpha_{j,i}; 0, \mathrm{var}(\boldsymbol{\alpha}_{S,j,i})\right)$$
$$+ p_{j,i}^L \cdot g\left(\alpha_{j,i}; 0, \mathrm{var}(\boldsymbol{\alpha}_{L,j,i})\right) \quad (17)$$

where $p_{j,i}^S + p_{j,i}^L = 1$ and $g\left(x; \mathrm{mean}(\alpha_i), \mathrm{var}(\alpha_i)\right) = \frac{1}{\sqrt{2\pi \mathrm{var}(\alpha_i)}} \exp\left(-\frac{(x-\mathrm{mean}(\alpha_i))^2}{2\mathrm{var}(\alpha_i)}\right)$.

The Markov model is then completely determined by a set of state transition matrices for coefficients at scales

$$\mathbf{A}_{j,i} = \begin{bmatrix} p_{j,i}^{S\to S} & p_{j,i}^{S\to L} \\ p_{j,i}^{L\to S} & p_{j,i}^{L\to L} \end{bmatrix}. \quad (18)$$

The values of $p_{j,i}^{S\to S}$ and $p_{j,i}^{L\to L}$ are significantly larger than their complements. If the hidden state probabilities are provided for the wavelet coefficient on the coarsest scale p_1^S and p_1^L, then the probability distribution for any hidden state can be obtained recursively:

$$\mathrm{Pr}\left(S_{j,i}=S\right) = p_{\alpha_{P(j,i)}}^S p_{j,i}^{S\to S} + p_{\alpha_{P(j,i)}}^L p_{j,i}^{L\to S}, \quad (19)$$
$$\mathrm{Pr}\left(S_{j,i}=L\right) = p_{\alpha_{P(j,i)}}^S p_{j,i}^{S\to L} + p_{\alpha_{P(j,i)}}^L p_{j,i}^{L\to L}. \quad (20)$$

The HMM parameters include the probabilities for the hidden states $\left\{p_{j,i}^S, p_{j,i}^L\right\}$, the state transition matrices $\mathbf{A}_{j,i}$, and the Gaussian distribution variances $\left\{\mathrm{var}(\boldsymbol{\alpha}_{S,j,i}), \mathrm{var}(\boldsymbol{\alpha}_{L,j,i})\right\}$ for each wavelet coefficient.

For simplicity, each wavelet coefficient at a scale has the same statistical behavior, i.e., $\mathrm{var}(\alpha_{S,j}) = \mathrm{var}(\boldsymbol{\alpha}_{S,j,i}) = \mathrm{var}(\boldsymbol{\alpha}_{S,j,k})$ for $i \neq k$ and $\mathrm{var}(\boldsymbol{\alpha}_{L,j}) = \mathrm{var}(\boldsymbol{\alpha}_{L,j,i}) = \mathrm{var}(\boldsymbol{\alpha}_{L,j,k})$ for $i \neq k$. The simple model has parameters $\mathbf{A}_j$ and $\left\{\mathrm{var}(\boldsymbol{\alpha}_{S,j}), \mathrm{var}(\boldsymbol{\alpha}_{L,j})\right\}$ for $1 \leq j \leq \mathcal{J}$. In addition, the variances $\mathrm{var}(\boldsymbol{\alpha}_{S,j})$ and $\mathrm{var}(\boldsymbol{\alpha}_{L,j})$ are modeled so that the decay of the coefficient magnitudes across scale may be considered. Thus, the variances decay exponentially as the scale becomes finer

$$\mathrm{var}(\boldsymbol{\alpha}_{S,j}) = C_{\sigma_S} \cdot 2^{-j \cdot a_S}, \quad \mathrm{var}(\boldsymbol{\alpha}_{L,j}) = C_{\sigma_L} \cdot 2^{-j \cdot a_L}$$

where the parameters $C_{\sigma_S}, C_{\sigma_L}, a_S,$ and a_L follow the model in [17]. The term $\mathrm{var}(\boldsymbol{\alpha}_j)$ is obtained as follows

$$\mathrm{var}(\boldsymbol{\alpha}_j) = \left(\mathrm{Pr}\left(S_j=S\right)\right)^2 \mathrm{var}(\boldsymbol{\alpha}_{S,j}) + \left(\mathrm{Pr}\left(S_j=L\right)\right)^2 \mathrm{var}(\boldsymbol{\alpha}_{L,j}). \quad (21)$$

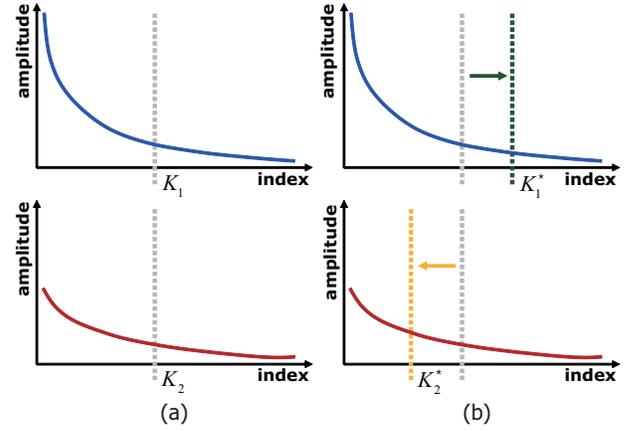Fig. 6. Intuitive explanation of modified BCS. (a) Without visual weights. (b) With visual weights.

For the bth block, the variance of $\boldsymbol{\alpha}_b$ becomes

$$\mathrm{var}\left(\boldsymbol{\alpha}_b\right) = \frac{1}{N_b} \sum_{\forall j \in \mathcal{J}_b} \sum_{\forall i \in \mathcal{N}_j} \left(\alpha_{j,i}\right)^2$$
$$= \frac{1}{N_b} \sum_{\forall j \in \mathcal{J}_b} |\mathcal{N}_j| \cdot \frac{1}{|\mathcal{N}_j|} \sum_{\forall i \in \mathcal{N}_j} \left(\alpha_{j,i}\right)^2$$
$$= \sum_{\forall j \in \mathcal{J}_b} \frac{|\mathcal{N}_j|}{N_b} \mathrm{var}(\boldsymbol{\alpha}_j). \quad (22)$$

Finally, the signal model in each block can be expressed

$$\left|\alpha_{b,\mathcal{I}(i)}\right| \leq G_b \cdot i^{-1/r_b} \leq G_b \cdot i^{-1/\left(4 - \frac{2N_b}{G_b^2}\mathrm{var}(\boldsymbol{\alpha}_b)\right)}, \quad (23)$$

and each block's error in the K_b-term approximation becomes

$$\|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b,K_b}\|_2^2 \leq (r_b s_b)^{1/2} G K_b^{-s_b}. \quad (24)$$

D. Upperbound of Gain from Visually Weighted BCS

Equation (8) can be interpreted as an upper bound on the approximation error due to undersampling. Once the statistics of each subband are obtained, then (8) can be utilized to derive an upper bound on the gain when visually weighted BCS is employed.

The performance difference obtained when sampling non-zero values over each block for the cases of "with visual weight" and "without visual weight" is shown in Fig. 6(a) and (b), respectively. Since the 1st block contains more visually important data than the 2nd block, it is necessary to select more non-zero values from the 1st block, as shown in Fig. 6(b). To find the optimal number of non-zero values, an objective function is used:

$$\min_{K_b} \sum_{b \in \mathcal{B}} \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b,K_b}\|_2^2, \quad \text{s.t.} \sum_{b \in \mathcal{B}} M_b \leq \bar{M} \quad (25)$$

where $\boldsymbol{\alpha}_{b,K_b} = \arg\min_{\bar{\boldsymbol{\alpha}}_b} \|\boldsymbol{\alpha}_b - \bar{\boldsymbol{\alpha}}_b\|_2$ and $\bar{M}$ is the constrained data rate in Fig. 1.

From (24), optimization of (25) can be re-formulated as

$$\min_{K_b} \sum_{b \in \mathcal{B}} (r_b s_b)^{1/2} G \cdot K_b^{-s_b}, \quad \text{s.t.} \sum_{b \in \mathcal{B}} c K_b \log \frac{N}{|\mathcal{B}|K_b} \leq \bar{M}.$$

Then formulate a *Lagrangian relaxation*

$$L\left(K_b, \lambda\right) = \sum_{b \in \mathcal{B}} (r_b s_b)^{1/2} G \cdot K_b^{-s_b}$$
$$+ \lambda \left(\sum_{b \in \mathcal{B}} c K_b \log \frac{N}{|\mathcal{B}| K_b} - \bar{M} \right)$$

where λ is a nonnegative Lagrange multiplier. Taking derivatives with respect to K_b and λ, respectively, yields the Karush-Kuhn-Tucker (KKT) conditions:

$$\frac{\partial L}{\partial K_b} = -s_b \left(r_b s_b\right)^{1/2} G \cdot K_b^{-(s_b+1)}$$
$$+ \lambda \left(c \log \left(\frac{N}{|\mathcal{B}|} \right) - c \log K_b - c \right) \leq 0 \quad (26)$$

$$K_b \cdot \frac{\partial L}{\partial K_b} = 0 \quad (27)$$

$$\lambda \left(\sum_{b \in \mathcal{B}} c K_b \log \frac{N}{|\mathcal{B}| K_b} - \bar{M} \right) = 0 \quad (28)$$

Due to complementary slackness, (26) can be simplified

$$-A \cdot K_b^{-(s_b+1)} + B + C \log K_b = 0, \quad (29)$$

where $A = \sqrt{r_b} s_b^{3/2} G$, $B = \lambda \left(c - c \log \frac{N}{|\mathcal{B}|} \right)$ and $C = c\lambda$. By introducing the Wright Omega function $\omega(\cdot)$, the optimal value of $K_b^\star$ is obtained

$$K_b^\star = \frac{\exp \left(\omega \left(\log \left(\frac{(s_b+1)A}{C} \right) + \frac{(s_b+1)B}{C} \right) / (s_b + 1) \right)}{\exp (B/C)}. \quad (30)$$

Likewise, the weighted MSE-based optimization with (14) in Fig. 6(b) becomes

$$\min_{K_b} \sum_{b \in \mathcal{B}} \boldsymbol{w}_b \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, K_b}\|_2^2, \quad \text{s.t.} \sum_{b \in \mathcal{B}} M_b \leq \bar{M} \quad (31)$$

where $\boldsymbol{w}_b = [w_i, w_{i+1}, \ldots], \forall i \in \mathcal{J}_b$ and $\boldsymbol{\alpha}_{b, K_b} = \arg\min_{\bar{\boldsymbol{\alpha}}_b} \|\boldsymbol{\alpha}_b - \bar{\boldsymbol{\alpha}}_b\|_p$. A re-formulation yields

$$\min_{K_b} \sum_{b \in \mathcal{B}} w_b (r_b s_b)^{1/2} G \cdot K_b^{-s_b}, \quad \text{s.t.} \sum_{b \in \mathcal{B}} c K_b \log \frac{N}{|\mathcal{B}| K_b} \leq \bar{M},$$

where $w_b = \sum_{\forall i \in \mathcal{J}_b} w_i$. The optimal value of $\tilde{K}_b^\star$ is given by

$$\tilde{K}_b^\star = \frac{\exp \left(\omega \left(\log \left(\frac{(s_b+1)A}{C} \right) + \frac{(s_b+1)B}{C} \right) / (s_b + 1) \right)}{\exp (B/C)}, \quad (32)$$

where $A = w_b \sqrt{r_b} s_b^{3/2} G$, $B = \lambda \left(c - c \log \frac{N}{|\mathcal{B}|} \right)$ and $C = c\lambda$.

Express the relative gain from the optimal value of $K_b^\star$ when compared to K, the gain $G\left(K, K_b^\star\right)$, in terms of the MSE, becomes

$$G\left(K, K_b^\star\right) = \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_2^2 - \sum_{b \in \mathcal{B}} \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, K_b^\star}\|_2^2. \quad (33)$$

If we want to show the gain between $K_b^\star$ and $\tilde{K}_b^\star$, (33) can be extended as follows:

$$G\left(K_b^\star, \tilde{K}_b^\star\right) = \sum_{b \in \mathcal{B}} \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, K_b^\star}\|_2^2 - \sum_{b \in \mathcal{B}} \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, \tilde{K}_b^\star}\|_2^2. \quad (34)$$

Algorithm 1 Visually Weighted CoSaMP Algorithm

Input: $M \times N$ matrix $\boldsymbol{\Phi}$, sample vector $u = \boldsymbol{\Phi}\boldsymbol{\alpha} + \boldsymbol{n}$ and sparsity of K, visual weight vector w
Output: K-sparse approximation a of α

1:   $a^0 \leftarrow 0$ (Initialization)
2:   $v \leftarrow u$
3:   $k \leftarrow 0$
while halting criterion **do**
4:     $k \leftarrow k + 1$
5:     $y \leftarrow \boldsymbol{\Phi}^* v$ (Signal proxy)
6:     $\Omega \leftarrow \mathbf{supp}\left(\mathbb{W}(y, 2K)\right)$ (Identification)
7:     $\Lambda \leftarrow \Omega \cup \mathbf{supp}\left(a^{k-1}\right)$ (Support merger)

8:     $b|_\Lambda \leftarrow \boldsymbol{\Phi}_\Lambda^\dagger u$ (Estimation)
9:     $b|_{\Lambda^c} \leftarrow 0$

10:    $a^k \leftarrow \mathbb{W}(b, K)$ (Pruning)
11:    $v \leftarrow u - \boldsymbol{\Phi} a^k$ (Sample update)
until   **(while)**

In a similar manner, the visual gain between $K_b^\star$ and $\tilde{K}_b^\star$ is easily obtained

$$G^w\left(K_b^\star, \tilde{K}_b^\star\right) = \sum_{b \in \mathcal{B}} \boldsymbol{w}_b \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, K_b^\star}\|_2^2 - \sum_{b \in \mathcal{B}} w_b \|\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_{b, \tilde{K}_b^\star}\|_2^2. \quad (35)$$

E. Visually Weighted CoSaMP for CS Reconstruction

In the Orthogonal Matching Pursuit (OMP) algorithm, an element that best approximates the residual at each iteration is selected from the dictionary [18]. The CoSaMP algorithm [19], which is an extension of OMP, guarantees the same performance as the best optimization-based CS recovery approaches. Each coefficient in the wavelet domain has a different visual weight. As such, it is advisable to also focus on each coefficient's visual weight and amplitude on the decoder side to achieve an optimized visual CS. Since CoSaMP guarantees that the performance for robust recovery follows the best convex optimization approach, we modified this state-of-the-art CS recovery algorithm using visual CS.

After applying modified BCS encoding to each block separately in the wavelet domain, the visual CS scheme reconstructs each piece of data according to its own visual importance on the decoder side - this process is called visually weighted CoSaMP or vwCoSaMP. The vwCoSaMP algorithm sequentially selects the most important element with respect to the visual weight rather than its own amplitude. The detailed vwCoSaMP procedure is described in Algorithm 1. All of the steps are identical to those of CoSaMP, except for the stages of "identification" and "pruning"; the visual CS decoder applies visual weighting in those stages. We define $\mathbb{W}(\alpha, K)$ as an algorithm that obtains the best K-approximation of α in the subspaces

$$\mathbb{W}(\boldsymbol{\alpha}, K) = \arg\min_{\boldsymbol{\alpha}_K \in \Sigma_K} w \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_2. \quad (36)$$

The performance of vwCoSaMP signal recovery over the wavelet domain is robustly guaranteed by the following theorem.

Theorem 2: Let $\alpha \in \Sigma_K$ and $u = \Phi\alpha + n$ be a set of noisy CS measurements. If Φ has a $4K$-RIP constant $\delta_{4K} \leq 0.1$, then the signal estimate a^k obtained from iteration k of the vwCoSaMP algorithm satisfies

$$\|\alpha - a^k\|_2 \leq \left(\frac{1}{2} + \frac{\|\bar{w}|_{\Omega\setminus\Pi}\|_2}{2\|\bar{w}|_{\Pi\setminus\Omega}\|_2}\right)^k \|\alpha\|_2$$
$$+ \left(5.204\left(1 + \frac{\|\bar{w}|_{\Omega\setminus\Pi}\|_2}{\|\bar{w}|_{\Pi\setminus\Omega}\|_2}\right) + 2.111\right)\|n\|_2.$$

Note that $\bar{w}$ is the diagonal term of W in (36). Thus,

$$w\|\alpha - \bar{\alpha}\|_2 \leq \|\bar{w}\|_2\|(\alpha - \bar{\alpha})\|_2 \qquad (37)$$

where $\bar{w} = [w_1 w_2 \cdots w_N]^T$ and $\Omega\setminus\Pi$ denotes the set difference of Ω and Π. The detailed proof of Theorem 2 is omitted for lack of space, but the procedure to prove Theorem 2 follows the one used to prove the CoSaMP algorithm in Section 4.6 in [3]; a set of six Lemmas in [3] are required. Modifications of the Lemmas in [3] can be applied to the visual CS recovery process.

IV. FEASIBILITY OF THE CS PERFORMANCE

Visually weighted CS could be effectively applied in a specific application where the sender has prior knowledge of the statistical behavior of an image and samples the data with low computational complexity, even if the receiver side is allowed a high computational complexity. For example, in a visual surveillance system, distributed multiple camera nodes can periodically perform a simple measurement and send the information to a master node. To reduce cost, each camera node may have low compute power and conduct only simple processing required to measure each image and send it to the master node. By contrast, the master node may have considerable computing power and storage capacity in order to process and store all the information delivered from the many camera nodes. Moreover, it is not necessary for the master node to reconstruct all of the images transmitted from camera nodes. The reconstruction is performed by the user only when it is necessary to determine what is happening at a certain time and place.

The modified BCS weighting mechanism can be extended to applications of CS where selective nonuniform sampling can produce better performance than uniform sampling, e.g., in medical imaging [20]. For example, CS may enable the reconstruction of MRI image data from undersampled data at relatively low levels of distortion as compared to an original image, while allowing the scan-time to be significantly reduced. Visually weighted CS is an attractive way to improve visual quality via perceptually weighted optimization of reconstruction following random nonuniform sampling on the encoding side. By contrast to conventional MRI reconstruction methods, visually weighted CS has the potential tp deliver improved quality at lower cost.

Here we study the feasibility of CS as applied to image sampling. Specifically, we determine the expected gain attained by visually weighted BCS. If CS is applied to the sampling process, it is necessary to estimate the number of bits required to represent the sampled data. In spite of several theoretical results regarding the rate-distortion performance of CS, it is difficult to find any practical formula. We demonstrate the differences in results achieved when sampling an image using CS and visual CS. We assume that CS senses a signal compressively, quantizes the measurements, then efficiently represents them by entropy coding. In other words, the proposed CS is used to sense signals at a compressed rate at the sampling stage, which generates measurements. Quantifying the CS data rate requires analyzing the statistical properties of the random measurements. The random measurements come from the product of a random matrix with the transformed coefficients. Scalar uniform quantization is employed and the data rate calculated as $\frac{M}{N}h(u)$, where $h(u)$ is the entropy (in bits per pixel) of the quantized measurement data u.

A. Rate-Distortion Model of CS

Among M-length measurements u, the mth measurement u_m is obtained by

$$u_m = \sum_{\forall i \in N} \phi_{i,m}\alpha_i, \qquad (38)$$

where $\phi_{i,m}$ is the mth row element of the ith column Φ_i of Φ. Assume that the entries of Φ are i.i.d. Gaussian random variables with zero mean and unit variance, then the pdf of $\phi_{i,m}$ becomes

$$f(\phi_{i,m}) = g(\phi_{i,m}; 0, 1), \qquad (39)$$

where

$$g(x; \text{mean}(x), \text{var}(x)) = \frac{1}{\sqrt{2\pi \text{var}(x)}}$$
$$\times \exp\left(-\frac{(x - \text{mean}(x))^2}{2\text{var}(x)}\right).$$

In [17], the pdf of the nth coefficient α_n becomes

$$f(\alpha_i) = p_n^S \cdot g(\alpha_i; 0, \text{var}(\alpha_S))$$
$$+ p_i^L \cdot g(\alpha_i; 0, \text{var}(\alpha_L)), \forall i \in N \qquad (40)$$

where $p_i^S + p_i^L = 1$.

Using (39) and (40), the pdf of the mth measurement u_m is obtained

$$f(u_m) = \sum_{\forall i \in N} f(\phi_{i,m}) f(\alpha_i)$$
$$= \sum_{\forall i \in N} g(\phi_{i,m}; 0, 1)\left\{p_i^S \cdot g(\alpha_i; 0, \text{var}(\alpha_S))\right.$$
$$\left. + p_i^L \cdot g(\alpha_i; 0, \text{var}(\alpha_L))\right\}$$
$$= \sum_{\forall i \in N}\left\{p_i^S \cdot g(\phi_{i,m}; 0, 1) g(\alpha_i; 0, \text{var}(\alpha_S))\right.$$
$$\left. + p_i^L \cdot g(\phi_{i,m}; 0, 1) g(\alpha_i; 0, \text{var}(\alpha_L))\right\}.$$

Because the mean and variance values for the product of Gaussian distributions $g(x; \text{mean}(x), \text{var}(x))$ and

$g\left(\mathbf{y}; \operatorname{mean}(\mathbf{y}), \operatorname{var}(\mathbf{y})\right)$ are $\frac{\operatorname{mean}(\mathbf{x})^2 \operatorname{var}(\mathbf{y}) + \operatorname{mean}(\mathbf{y})^2 \operatorname{var}(\mathbf{x})}{\operatorname{var}(\mathbf{x}) + \operatorname{var}(\mathbf{y})}$ and $\frac{\operatorname{var}(\mathbf{x}) \operatorname{var}(\mathbf{y})}{\operatorname{var}(\mathbf{x}) + \operatorname{var}(\mathbf{y})}$ in [21], the pdf becomes

$$
f(u_m) = \sum_{\forall i \in N} \left\{ p_i^S \cdot g\left(\phi_{i,m} \cdot \alpha_i; 0, \frac{\operatorname{var}(\boldsymbol{\alpha}_S)}{1 + \operatorname{var}(\boldsymbol{\alpha}_S)}\right) \right.
$$
$$
\left. + p_i^L \cdot g\left(\phi_{i,m} \cdot \alpha_i; 0, \frac{\operatorname{var}(\boldsymbol{\alpha}_L)}{1 + \operatorname{var}(\boldsymbol{\alpha}_L)}\right) \right\}
$$
$$
= \sum_{\forall i \in N} g\left(\phi_{i,m} \cdot \alpha_i; 0, \left(p_i^S\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_S)}{1 + \operatorname{var}(\boldsymbol{\alpha}_S)} \right.
$$
$$
\left. + \left(p_i^L\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_L)}{1 + \operatorname{var}(\boldsymbol{\alpha}_L)}\right) \tag{41}
$$

Since each component $\phi_{i,m}$ is uncorrelated with other components $\phi_{j,m}$, $\forall j \in N, j \neq i$, (41) becomes

$$
f(u_m) \simeq g\left(\phi_{i,m} \cdot \alpha_i; 0, |N|\left(\left(p_i^S\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_S)}{1 + \operatorname{var}(\boldsymbol{\alpha}_S)} \right.\right.
$$
$$
\left.\left. + \left(p_i^L\right)^2 \frac{\operatorname{var}(\alpha_L)}{1 + \operatorname{var}(\alpha_L)}\right)\right).
$$

The differential entropy for u_m then becomes

$$
h(u_m) = \frac{1}{2} \log_2\left(2\pi e |N|\left(\left(p_i^S\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_S)}{1 + \operatorname{var}(\boldsymbol{\alpha}_S)} \right.\right.
$$
$$
\left.\left. + \left(p_i^L\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_L)}{1 + \operatorname{var}(\boldsymbol{\alpha}_L)}\right)\right). \tag{42}
$$

In [22], the classical rate distortion model is modified to yield the data rate and distortion models expressed in terms of the quantization step size Δ. In the codec, the relationship between the quantization step size Δ and the quantization parameter (QP) q is expressed as $\Delta = 2^{q/6}$. Considering the quantized measurements, (42) becomes

$$
h(u_m, \Delta) = \frac{1}{2} \log_2\left(\frac{24\pi e |N|}{\Delta^2}\left(\left(p_i^S\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_S)}{1 + \operatorname{var}(\boldsymbol{\alpha}_S)} \right.\right.
$$
$$
\left.\left. + \left(p_i^L\right)^2 \frac{\operatorname{var}(\boldsymbol{\alpha}_L)}{1 + \operatorname{var}(\boldsymbol{\alpha}_L)}\right)\right).
$$

The empirical data rate is also defined by linearly scaling the entropy in [23] as

$$
r(u_m, \Delta) = c_a \frac{M}{N} h(u_m, \Delta) + c_b \tag{43}
$$

where c_a and c_b are constants. To decide constants c_a and c_b, let Z be a random variable representing the sample points $h(u_m, \Delta)$. The random variable R of $r(u_m, \Delta)$ can then be expressed as $R = c_a(M/N) Z + c_b$. The expected value and variance of R are $\operatorname{mean}(R) = c_a(M/N) \operatorname{mean}(Z) + c_b$ and $\operatorname{var}(R) = c_a^2(M/N)^2 \operatorname{var}(Z)$. Therefore, $c_a = \sqrt{(N \cdot \operatorname{var}(R))/(M \cdot \operatorname{var}(Z))}$ and $c_b = \operatorname{mean}(R) - c_a(M/N) \operatorname{mean}(Z)$.

B. Rate-Distortion Model of JPEG2000

In general, visually weighted CS aims at sampling visually important data from the original image so the basic idea

TABLE I
COMBINATIONS OF ENCODER AND DECODER

Scenario	Measurements	Reconstruction
1	CS	CoSaMP
2	Modified BCS	CoSaMP
3	Modified BCS	vwCoSaMP
4	Visually weighted BCS	vwCoSaMP

is fundamentally different from post-acquisition data compression. Nevertheless, it is necessary to evaluate how much performance gain we are to obtain in practical applications. To make such an evaluation, we benchmark against JPEG2000 [24]. Since JPEG2000 also involves wavelet transformation, quantization and entropy coding, we model rate-distortion for JPEG2000 based on the pdf of coefficients $\boldsymbol{\alpha}$.

The pdf of the coefficient α_n becomes

$$
f(\alpha_i) = p_n^S \cdot g(\alpha_i; 0, \operatorname{var}(\boldsymbol{\alpha}_S)) + p_i^L \cdot g(\alpha_i; 0, \operatorname{var}(\boldsymbol{\alpha}_L))
$$
$$
\simeq g\left(\alpha_i; 0, \left(p_i^S\right)^2 \cdot \operatorname{var}(\boldsymbol{\alpha}_S) + \left(p_i^L\right)^2 \cdot \operatorname{var}(\boldsymbol{\alpha}_L)\right).
$$

The differential entropy and the empirical data rate for α_n are then obtained by

$$
h(\alpha_i, \Delta) = \frac{1}{2} \log_2\left(\frac{24\pi e}{\Delta^2}\left(\left(p_i^S\right)^2 \cdot \operatorname{var}(\boldsymbol{\alpha}_S) \right.\right.
$$
$$
\left.\left. + \left(p_i^L\right)^2 \cdot \operatorname{var}(\boldsymbol{\alpha}_L)\right)\right) \tag{44}
$$
$$
r(\alpha_i, \Delta) = c_c h(\alpha_i, \Delta) + c_d, \tag{45}
$$

where c_c and c_d are constants. To determine the constants c_c and c_d, it is assumed to use the mean and the variance as in the procedure in the previous subsection.

V. EXPERIMENTAL RESULTS

A. Four Scenarios for Measurement and Reconstruction

We employ four experimental scenarios with different measurement and reconstruction combinations, as tabulated in Table I. Measurement data are obtained by producing a random matrix with i.i.d. Gaussian entries.

The model (15) fits actual values as shown in Fig. 7; each solid line represents the actual value of the sorted wavelet coefficients from each block in the 'Lena' image (using the Haar wavelet basis), while each dotted line represents the value estimated from (15). Overall, the statistical model (15) follows the true values quite well and thus we replace the actual wavelet coefficients with the model values in the numerical analysis.

B. Differences in the Number of Measurements

The number of measurements as a function of the inverse sparsity level K/N (the sparsity level is defined here as $1 - K/N$)) for conventional CS encoding and modified BCS is compared in Figs. 8(a) and (b). To obtain values for modified BCS, we use $K_b^\star$ in (30) with respect to $M = \sum_{b \in \mathcal{B}} c K_b \log \frac{N_b}{K_b}$. As shown in Fig. 8, the values of M and $\sum M_b$ in (12) indicate that modified BCS introduces a gain in
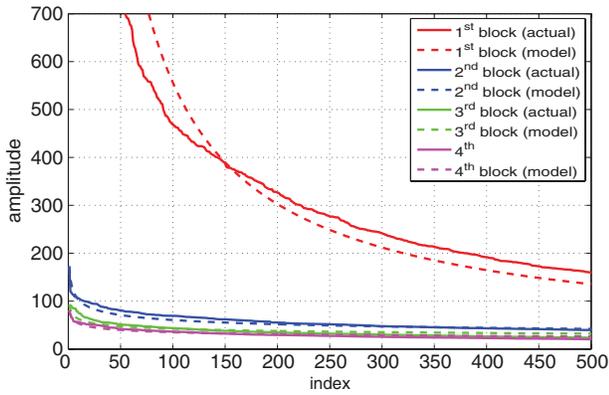
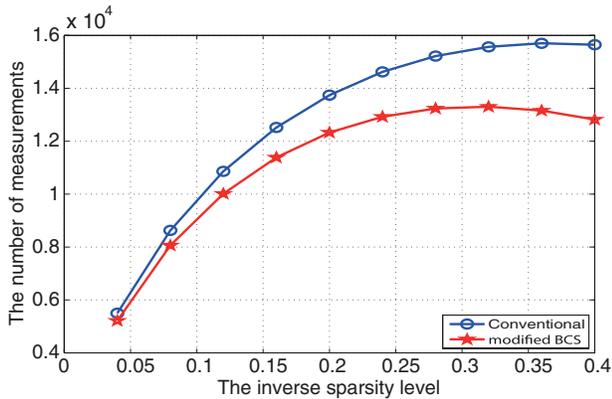Fig. 7. Modeling obtained by using (refeq:modeling) for each block in the wavelet transformed "*Lena*" image.

Fig. 8. Difference of the numbers of measurements between (a) conventional CS and (b) modified BCS.

Fig. 9. Performance comparisons as a function of the number of coefficients for *Scenarios* 1–4. (a) PSNR. (b) Weighted PSNR.

the number of measurements as compared to conventional CS encoding. In other words, while modified BCS uses the same number of non-zero values as conventional CS, it exhibits improved compression efficiency due to a reduction in the number of measurements.

C. PSNR, Weighted PSNR, and SSIM Evaluation

The performances of each scenario on the "Phantom" and "Baboon" images (256 × 256) is shown in Table I. For values of M in the range of 4000 to 30000, we calculated the Peak Signal-to-Noise Ratio (PSNR). The PSNR is inversely proportional to the MSE (PSNR = $10 \log_{10}(255^2/\text{MSE})$). Values of the Structural Similarity (SSIM) are benchmarked using the Mean SSIM (MSSIM) index in [25]. A comparison of the PSNR among the scenarios in Table I is shown in Fig. 9(a). It was found that *Scenario 3*, the combination of modified BCS and vwCoSaMP, is superior over the entire measurement range. The PSNR values of *Scenario 4* are not higher than those of *Scenario 3* because of the loss of coefficients due to $\tilde{K}_b^\star$. Another finding is that all of the results have performance bounds that none can overcome. This bound is the unrecoverable energy which comes from the generation of sparse data. In addition, measurements from *Scenario 4* contain more important coefficients in the 1st block. Consequently, *Scenario 4*
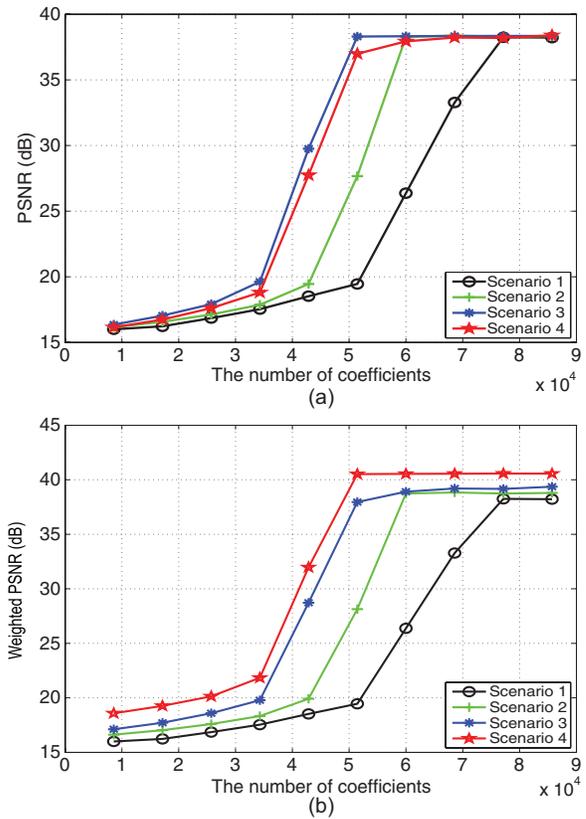
is superior to the other scenarios, with the exception of *Scenario 3*. It is apparent that the performance of vwCoSaMP is better than that of CoSaMP because *Scenario 3* exhibits a better PSNR than *Scenario 2*. In addition, the performance of modified BCS is superior to that of conventional CS since *Scenario 2* exhibits a better PSNR than *Scenario 1*. The values of the weighted PSNR (weighted PSNR = $10 \log_{10}(255^2/\text{weighted MSE})$) are shown in Fig. 9 (b). Since the weighted PSNR reflects visual weights as defined in (14), *Scenario 4* exhibits a definite gain in performance with respect to the other scenarios. In other words, the best visual quality is guaranteed over *Scenario 4* despite the loss in energy. In conventional CS, each coefficient is sampled randomly without applying any weight to the distribution of coefficients. However, each coefficient needs to be adaptively sampled according to its visual importance to achieve a higher visual quality relative to conventional CS. Therefore, our proposed method intuitively shows visually improved performance relative to conventional CS for a given number of measurements.

Figure 10 shows reconstructed images for 'Baboon' when a higher number of measurements ($M = 23000$). Since the encoder chooses more important values from the coefficients over the wavelet domain, *Scenario 4* exhibits visual superiority even with a sufficient number of measurements. In particular, if we focus on an area around the eyes or nose of the 'Baboon' as a Region of Interest (ROI), *Scenario 4* shows improved visual performance. Therefore, for this particular model, i.e., wavelet decomposed data, we can observe that our proposed approach
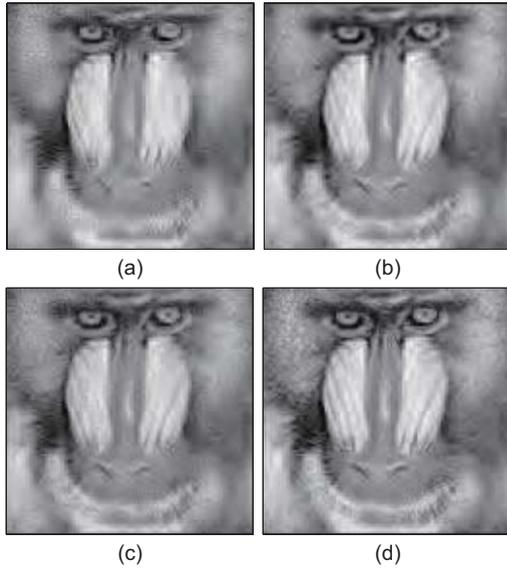
Fig. 10. Reconstruction through by the recovery algorithm (for M=23, 000). (a) *Scenario 1* (PSNR = 28.72 dB and SSIM = 0.89). (b) *Scenario 2* (PSNR = 29.76 dB and SSIM = 0.93). (c) *Scenario 3* (PSNR = 34.79 dB and SSIM = 0.94). (d) *Scenario 4* (PSNR = 34.57 dB and SSIM = 0.97).
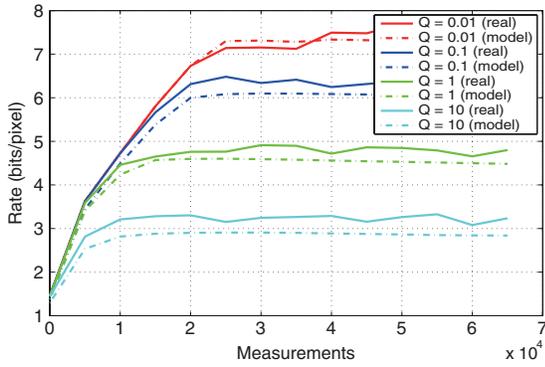

Fig. 11. Data rates plotted against the number of measurements for logarithmic steps in quantization (Q). Solid line values: real rates. Dotted lines: model rates from (43).

Fig. 12. PSNR plotted against data rate for images. (a) "*Lena.*" (b) "*Phantom.*"

Fig. 13. Data rates plotted against the inverse sparsity level (K/N).

generally yields much better quality than other conventional CS methods.

D. Rate Comparison of JPEG2000, CS and Visual CS

Figure 11 plots data rates from CS encoding against the number of measurements for logarithmic stepped quantization step sizes for the 'Baboon' image. It can be seen that the empirical data rates in (43) fit the real ones ($c_a = 37.82$ and $c_b = -15.97$). The modeling is applied to JPEG2000 using (44) and (45).

Fig. 12(a) and (b) plot PSNR against data rate for the 'Lena' and 'Phantom' images when applying JPEG2000, CS and visual CS. As shown in Fig. 12(a), the performance of JPEG2000 generally overwhelms that of CS and visual CS at the same data rate. In addition, it is apparent that the performance of visual CS is higher than CS. Owing to the high complexity of the 'Lena' image, the performance of JPEG2000 is much better than the other two algorithms. Unlike Fig. 12(a),
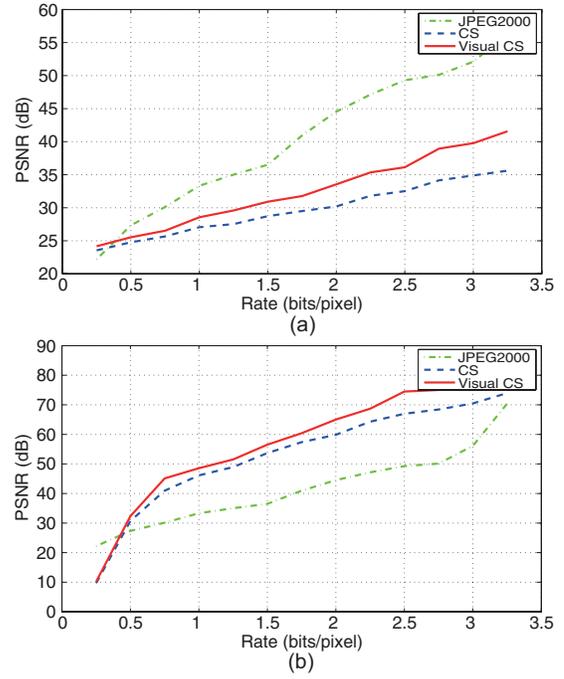
the PSNR of CS is mostly higher than that of JPEG2000 in Fig. 12(b). Thus, it may be observed that CS and visually weighted CS are appropriate for image compression with low sparsity as for the 'Phantom' image. Based on findings from Figs. 12(a) and (b), we measure the outperformance among the three methods using (43) and (45).

Figure 13 shows a comparison of rates w.r.t. the inverse sparsity level ($c_a = 37.82$, $c_b = -15.97$, $c_c = 43.84$ and $c_d = -8.44$ are used). While the rate of JPEG2000 does not vary much as a function of the inverse sparsity level, those of CS and visual CS have a distinct gain at low sparsity, i.e. $K/N < 0.4$. We benchmark JPEG2000 using perceptual distortion control as in [26]. It turns out that the performance of the perceptually optimized JPEG2000 is inferior to CS and visual CS when the original image has low sparsity.

VI. Conclusion

A visually weighted CS framework was proposed. Measurements associated with different levels of visual importance are obtained over the spatial and frequency domains. In the proposed scheme, a modified BCS model on the measurement stage leads not only to the encoding of data with respect to the level of visual importance of information in the wavelet domain, but also reduces the total number of CS measurements. In addition, using vwCoSaMP on the reconstruction side allows the proposed CS to recover sparse data in a manner corresponding to the visual weights. In other words, the visually weighted CS model fully utilizes the sparsity from the wavelet transform data and improves visual quality by combining visual weights with the CS encoder and decoder using a visual weighted BCS and vwCoSaMP. Gains from the visual CS framework were induced by the modeling of wavelet coefficients. The experimental results demonstrated the superiority of the proposed method. In conventional CS, each coefficient is sampled equally without any information about the coefficient; however, each coefficient is sampled according to its visual importance when the volume of visually weighted CS's compressed data is the same as for conventional CS. The visually weighted CS scheme produces the better visual improvements by the introduction of the visual weights.

Appendix

A. Proof of Theorem 2

Recall from (7),

$$\sigma_K\left(\boldsymbol{\alpha}\right)_p \triangleq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_p = \left(\sum_{i \in N} \left(\alpha_i - \alpha_{K,i}\right)^p\right)^{1/p}$$
$$= \left(\sum_{i \in N} \left(\alpha_{\mathcal{I}(i)} - \alpha_{K,\mathcal{I}(i)}\right)^p\right)^{1/p} \quad (46)$$

where $\alpha_{K,i}$ is the ith element for the best K-term approximation of $\boldsymbol{\alpha}$ and $\alpha_{K,\mathcal{I}(i)}$ is the ith element when sorted in order of decreasing magnitude for $\alpha_{K,i}$.

Assume $p = 2$ and $K = 1$. Then, (46) may be interpreted as the result of removing the largest coefficient from the original set

$$\sigma_1\left(\boldsymbol{\alpha}\right)_2 = \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_1\|_2 = \sqrt{\sum_{i \in N}\left(\alpha_i - \alpha_{1,i}\right)^2} \quad (47)$$

where $\alpha_{1,i}$ is an element of α_1. For example, if $\alpha = [1\ 7\ 10\ 5\ 12\ 9]$, then $\alpha_1 = [0\ 0\ 0\ 0\ 12\ 0]$. Also, using (8), (47) becomes

$$\sigma_1\left(\boldsymbol{\alpha}\right)_2 \leq (rs)^{1/2}\, G \cdot 1^{-s} = (rs)^{1/2}\, G, \quad (48)$$

and the square of (48) becomes

$$\left(\sigma_1\left(\alpha\right)_2\right)^2 = \sum_{i=2}^{N} \alpha_{\mathcal{I}(i)}^2 + \left(\alpha_{\mathcal{I}(1)} - \alpha_{1,\mathcal{I}(1)}\right)^2$$
$$= \sum_{i=2}^{N} \alpha_{\mathcal{I}(i)}^2 \leq (rs)\, G^2.$$

By adding the largest value $\left(\alpha_{\mathcal{I}(1)}\right)^2$ to both sides,

$$\left(\sigma_1\left(\boldsymbol{\alpha}\right)_2\right)^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2 = \sum_{i=2}^{N} \alpha_{\mathcal{I}(i)}^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2$$
$$= \sum_{i \in N} \alpha_{\mathcal{I}(i)}^2 \leq (rs)\, G^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2.$$

Since $\left|\alpha_{\mathcal{I}(1)}\right| \leq G^2$ from (5) with $i = 1$, $\left(\sigma_1\left(\boldsymbol{\alpha}\right)_2\right)^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2$ becomes

$$\left(\sigma_1\left(\boldsymbol{\alpha}\right)_2\right)^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2 = \sum_{i \in N} \alpha_{\mathcal{I}(i)}^2 = \sum_{i \in N} \alpha_i^2$$
$$\leq (rs)\, G^2 + \left(\alpha_{\mathcal{I}(1)}\right)^2 \leq (rs)\, G^2 + G^2 = (1 + rs)\, G^2.$$

If $\boldsymbol{\alpha}$ is modeled as obeying a zero-mean Gaussian distribution with variance $\mathrm{var}(\boldsymbol{\alpha}) = \frac{1}{N}\sum_{i \in N} \alpha_i^2$ [12], then

$$\mathrm{var}(\boldsymbol{\alpha}) = \frac{1}{N}\sum_{i \in N} \alpha_i^2 \leq \frac{G^2}{N}\left(1 + rs\right) = \frac{G^2}{N}\left(2 - \frac{r}{2}\right), \quad (49)$$

where r is the power number and $s = 1/r - 1/2$. Thus, the power number r is bounded as

$$r \leq 4 - \frac{2N}{G^2}\mathrm{var}(\alpha). \quad (50)$$

Using (50), (5) can be expressed

$$\left|\alpha_{\mathcal{I}(i)}\right| \leq G \cdot i^{-1/r} \leq G \cdot i^{-1/\left(4 - \frac{2N}{G^2}\mathrm{var}(\boldsymbol{\alpha})\right)}. \quad (51)$$

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[4] A. Schulz, L. Velho, and E. A. B. da Silva, "On the empirical rate-distortion performance of compressive sensing," in *Proc. IEEE Int. Conf. Image Process.*, Cairo, Egypt, Nov. 2009, pp. 3049–3052.

[5] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.

[6] R. Puri, A. Majumdar, and P. I. K. Ramchandran, "Distributed video coding in wireless sensor networks," *IEEE Signal Procees. Mag.*, vol. 23, no. 4, pp. 94–106, Jul. 2006.

[7] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1999.

[9] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[10] H. Lee and S. Lee, "Visual entropy gain for wavelet image coding," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 553–556, Sep. 2006.

[11] H. Lee, S. Lee, and A. C. Bovik, "Cross-layer optimization for downlink wavelet video transmission," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 813–823, Aug. 2011.

[12] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.

[13] E. J. Candès, "Conmpressive sensing," in *Proc. Int. Congr. Math.*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.

[14] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[15] L. Gan, "Block compressed sensing of natural images," in *Proc. Int. Conf. Digit. Signal Process.*, Cardiff, U.K., Jul. 2007, pp. 403–406.

[16] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden Markov tree model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, LV, Mar.–Apr. 2008, pp. 5137–5140.

[17] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[19] D. Needle and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and incaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–302, May 2009.

[20] H. Oh and S. Lee, "Visually weighted reconstruction of compressive sensing MRI," *Magn. Reson. Imag.*, to be published.

[21] P. A. Bromiley, "Products and convolutions of Gaussian distributions," Medical School, Univ. Manchester, Manchester, U.K., Tech. Rep. 2003-003, 2003.

[22] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application. Part I: Fundamental theory," vol. 7, no. 2, pp. 287–298, Apr. 1997.

[23] J. Park, H. Lee, S. Lee, and A. C. Bovik, "Optimal channel adaptation of scalable video over a multi-carrier based multi-cell environment," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1062–1071, Oct. 2009.

[24] D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Dordrecht, The Netherlands: Kluwer, 2001.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[26] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1763–1778, Jul. 2006.

Hyungkeuk Lee received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2005, 2006, and 2011, respectively.

He is currently with Samsung Electronics, Inc., Seoul, Korea. In 2008, he was a Visiting Researcher with the Laboratory for Image and Video Engineering, University of Texas at Austin, Austin, involved in research under guidance of Prof. A. C. Bovik. His current research interests include wireless resource allocation based on economics, video coding, cross-layer optimization, image/video quality assessments, compressive sensing, 3-D image processing, and computer vision.

Heeseok Oh received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree.

His current research interests include 3-D image/video quality and visual fatigue assessment, and compressive sensing.

Sanghoon Lee (SM'12) was born in Korea in 1966. He received the B.S. degree from Yonsei University, Seoul, Korea, the M.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, and the Ph.D. degree from the University of Texas at Austin, Austin, in 1989, 1991, and 2000, respectively, all in electrical engineering.

He was with Korea Telecom, Seoul, from 1991 to 1996. In 1999, he was with the Bell Laboratories, Lucent Technologies, where he was engaged in research on wireless multimedia communications, and from 2000 to 2002, he was engaged on developing real-time embedded software and communication protocols for 3G wireless networks. In 2003, he joined the Department of Electrical and Electronics Engineering, Yonsei University, where he is currently a Full Professor. His current research interests include image and video quality assessments, image and video processing, wireless multimedia communications, and 4G wireless networks.

Dr. Lee was a recipient of the 2012 Special Service Award from the IEEE Broadcast Technology Society. He is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, an Editor of the *Journal of Communications and Networks*, the Chair of the IEEE *Standard Working Group for 3-D quality assessment*, a Guest Editor of the IEEE *Transactions on Image Processing* 2012, and the General Chair of the 2013 IEEE IVMSP Workshop.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) is the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, Austin, where he is the Director of the Laboratory for Image and Video Engineering, and he is a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. He has authored or co-authored more than 650 technical articles, and he holds two U.S. patents. His several books include the recent companion volumes The Essential Guides to Image and Video Processing (Academic Press, 2009). His current research interests include image and video processing, computational vision, and visual perception.

Dr. Bovik was a recipient of the Journal Paper Award from the International Pattern Recognition Society in 1988 and 1993, the IEEE Third Millennium Medal in 2000, the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana in 2008, the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin, a number of major awards from the IEEE Signal Processing Society, including the Meritorious Service Award in 1998, the Technical Achievement Award in 2005, the Education Award in 2007, and the Best Paper Award in 2009, and the SPIE/IS&T Imaging Scientist of the Year Award in 2011. He is a fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He was involved in numerous professional society activities, including: on the Board of Governors of the IEEE Signal Processing Society from 1996 to 1998; the Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002; on the Editorial Board of The Proceedings of the IEEE from 1998 to 2004; and the Founding General Chairman of the First IEEE International Conference on Image Processing held in Austin, Texas, in 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.