

## Adaptive Compressed Image Sensing Using Dictionaries\*

Amir Averbuch<sup>†</sup>, Shai Dekel<sup>‡</sup>, and Shay Deutsch<sup>§</sup>

**Abstract.** In recent years, the theory of compressed sensing has emerged as an alternative for the Shannon sampling theorem, suggesting that compressible signals can be reconstructed from far fewer samples than required by the Shannon sampling theorem. In fact the theory advocates that nonadaptive, “random” functionals are in some sense optimal for this task. However, in practice, compressed sensing is very difficult to implement for large data sets, particularly because the recovery algorithms require significant computational resources. In this work, we present a new alternative method for simultaneous image acquisition and compression called adaptive compressed sampling. We exploit wavelet tree structures found in natural images to replace the “universal” acquisition of incoherent measurements with a direct and fast method for adaptive wavelet tree acquisition. The main advantages of this direct approach are that no complex recovery algorithm is in fact needed and that it allows more control over the compressed image quality, in particular, the sharpness of edges. Our experimental results show, by way of software simulations, that our adaptive algorithms perform better than existing nonadaptive methods in terms of image quality and speed.

**Key words.** compressed sensing, adaptive approximation, nonlinear approximation, wavelet trees

**AMS subject classifications.** 65T60, 65Y10, 68U10

**DOI.** 10.1137/110820579

**1. Introduction.** The Shannon sampling theory is at the core of nearly all signal acquisition protocols. The classic Shannon theorem states that the sampling rate, i.e., the Nyquist rate, must be at least twice the maximum frequency present in the signal (see, e.g., [52]). However, the Nyquist rate is too high and results in a massive data acquisition that must be compressed in order to be either stored or transmitted. In addition, there are important and emerging applications in which high sampling rates are very expensive or limited by physical or physiological constraints.

In recent years, a novel approach for signal sampling, known as *compressed sensing* (CS) [1], [9], [18], [14], [35], has emerged. CS is an alternative to the Nyquist rate for the acquisition of “sparse” signals. Instead of uniformly sampling the signal at Nyquist rate, in the CS paradigm one is given a budget of  $n$  independent “questions” to ask about the signal  $x \in \mathbb{R}^N$ , with  $n \ll N$ . These questions take the form of  $n$  linear functionals  $\langle \phi_i, \cdot \rangle$ , asking for the values  $y_i = \langle \phi_i, x \rangle$ ,  $1 \leq i \leq n$ . The vectors  $\phi_i \in \mathbb{R}^N$  form the sensing matrix  $\Phi = (\phi_1, \phi_2, \dots, \phi_n)$ , such that  $y = \Phi x$ , where the dimension of  $\Phi$  is  $n \times N$ . In the next stage, a numerical optimization method is used to reconstruct the full-length signal from the small amount of

\*Received by the editors January 10, 2011; accepted for publication (in revised form) October 19, 2011; published electronically January 24, 2012.

<http://www.siam.org/journals/siims/5-1/82057.html>

<sup>†</sup>School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel ([amir@math.tau.ac.il](mailto:amir@math.tau.ac.il)).

<sup>‡</sup>GE Healthcare, Herzeliya 46733, Israel ([shai.dekel@ge.com](mailto:shai.dekel@ge.com)).

<sup>§</sup>Department of Applied Mathematics, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel ([shayseut@post.tau.ac.il](mailto:shayseut@post.tau.ac.il)).

collected data. The CS paradigm relies on two key properties: sparsity and incoherency.

A signal  $x \in \mathbb{R}^N$  is  $k$ -sparse in a transform basis  $\Psi = [\psi_1, \psi_2, \dots, \psi_N]$  if the number of nonzero coefficients  $\{\theta_i = \langle x, \psi_i \rangle\}$  is relatively small  $k \ll N$ . It is compressible if the number of significant (e.g., absolute value above some threshold) coefficients is relatively small. The basis can be, for example, a wavelet basis, a Fourier basis, or a local Fourier basis (see, e.g., [31], [52]), depending on the application. There are also the tight frames such as curvelets [10] or contourlets [62] that incorporate directional information, thereby allowing a sparser representation of curve singularities. Incoherency, which is concerned with the acquisition modality, states that the sensing waveforms should have an extremely dense representation in  $\Psi$ .

Based on these two properties, one can design sampling protocols that capture useful information content embedded in a sparse signal and condense it into small data amount. A related condition to incoherency, which also has proved to be very useful in studying the general robustness of CS, is the so-called *restricted isometry property (RIP)* [14], [17]. Given a measurement matrix  $\Phi$  (and a basis  $\Psi$ ), the RIP can provide a sufficient condition for a stable solution for both  $k$ -sparse and compressible (in the basis  $\Psi$ ) signals. In other words, the matrix  $\Phi$  preserves the information in a sparse or compressible signal if it satisfies the RIP condition. In general, constructing a measurement matrix that satisfies the RIP condition for large  $k$  and  $N$  is a hard combinatorial problem [56]. However, the RIP condition can be achieved for a large class of randomized matrices with a high probability [14], [36].

In scenarios where there is no noise present, the recovery of an unknown signal  $x$  from the measurement vector  $y$  would ideally be achieved by searching for the  $l_0$ -sparsest representation that agrees with the measurements

$$(1.1) \quad \hat{\theta} = \arg \min \|\theta\|_0 \text{ subject to } y = \Phi\Psi^{-1}\theta.$$

It is known [26] that if  $x$  is a  $k$ -sparse vector, then with  $\Psi = I$ , one can design a sampling matrix  $\Phi$  of only  $n = 2k$  measurements and recover  $x$  via  $l_0$  minimization. Unfortunately, the  $l_0$  minimization problem leads to a daunting NP-complete combinatorial optimization problem [56].

Instead of solving the  $l_0$  minimization problem, nonadaptive CS theory seeks to solve the “closest possible” tractable minimization problem, i.e., the  $l_1$  minimization

$$(1.2) \quad \hat{\theta} = \arg \min \|\theta\|_1 \text{ subject to } y = \Phi\Psi^{-1}\theta.$$

This modification, known as Basis Pursuit [22], leads to a much simpler convex problem, which approximates the original problem and can be solved by various classical optimization techniques. But at the same time, minimizing the  $l_1$  norm finds a sparse approximation because it prevents diffusion of the signal’s energy over a large number of coefficients. It typically requires a number of  $n = O(k \log(N/k))$  measurements to robustly recover  $k$ -sparse and compressible signals. One can adjust (1.2) to deal with cases where  $x$  or  $y$  is corrupted with additive noise by replacing the equality  $y = \Phi\Psi^{-1}\theta$  with an approximation constraint [22], [16]. Another such algorithm is the minimization of the total variation with quadratic constraints that solves the following problem:

$$(1.3) \quad \min_I \|I\|_{TV} \text{ subject to } \|\Phi x_I - y\|_2 \leq \varepsilon,$$

where

$$\|I\|_{TV} = \sum_{i,j} \sqrt{(I(i+1,j) - I(i,j))^2 + (I(i,j+1) - I(i,j))^2},$$

$x_I$  is the vector created from a digital image  $I$  by concatenating its rows, and  $\varepsilon$  depends on the noise level.

In general, solving the total variation problem (1.3) provides a better image quality for image reconstruction at the expense of greater time complexity.

Although different convex optimization algorithms such as Basis Pursuit are optimal in the sense that they provide a global minimum, they involve expensive computations when applied to large-scale signals. Therefore a second approach, using iterative greedy methods, has been proposed. Examples include Matching Pursuit [39], Orthogonal Matching Pursuit [72], Tree Matching Pursuit (TMP) [48], StOMP [37], Iterative Hard Thresholding Pursuit [5], Subspace Pursuit [30], and CoSaMP [57].

Let us focus on one of the most promising implications of CS derivatives: the enabling and design of new types of *compressed imaging* (CI). One such device, which implements optical CS, is the single-pixel camera [38] inspiring a wide interest in CI applications. The single-pixel camera relies on the use of a *digital micromirror device* (DMD), which is composed of a grid of mirrors, where each mirror of the array is suspended above an individual static random access memory (SRAM) cell (see section 2.2). We also note that there are experimental CI algorithms, designed for measurements taken in the frequency domain in medical imaging [50].

However, the design of an efficient and optimal measurement basis in such a system is still a challenging problem. Although theoretically very powerful, there is still a huge gap between the CS theory and imaging applications. We note a few of these key difficulties:

1. *Control over image quality of the compressed image.* From the theory of wavelet approximation, it is clear that the error of  $n$ -term wavelet approximation is characterized by the “weak-type smoothness” of the image as a function in a Besov space [34]. In terms of image representation, more wavelet coefficients are needed to achieve a specified level of quality if the image contains higher visual activity such as edges and texture parts. Typically, however, existing CS architectures are not adaptive, and the number of measurements is determined before the acquisition process begins, with no feedback during the acquisition process on the improved quality. Yet, there exists work on a “sequential CS” paradigm [51], where the sampling procedure is composed of iterations of acquisition of random projections, reconstruction, testing for convergence, and then acquisition of more random projections if needed. The main challenge with this scheme is that the recovery algorithm needs to be applied at each iteration; thus the total complexity of the acquisition algorithm cannot be of order  $O(n)$ .
2. *Computationally intensive sampling process.* Constructing an efficient and fast sampling operator for nonadaptive CS remains a challenging problem. On the one hand, the CS theory shows that random measurement (i.e., unstructured) systems are in some sense an optimal strategy for acquiring sparse signals. However, sampling operators of random binary patterns are by nature dense matrices that require huge memory allocations and intensive computations for large values of  $N$  and  $n$ . It is impractical even for an image of size  $256 \times 256$ . Some of these practical challenges have already been somewhat addressed. The Scrambled Fourier Ensemble [12] was proposed for an

MRI application and is often employed in CS research because of its fast implementation. Other researchers [4] developed binary sparse measurement matrices with low complexity and near-optimal performance. However, these operators are incoherent only with identity matrices. The Block Hadamard Ensemble [40] was designed to offer a practical measurement operator for implementations.

3. *Computationally intensive reconstruction algorithm.* Although the algorithms designed to solve the  $l_1$  minimization (1.2) are proved to be theoretically efficient, i.e., with polynomial complexity bounds, whenever the problem has a large scale, it requires heavy computations.

In this work, we introduce an architecture that, in accordance with the CS paradigm, aims at providing an acquisition process that captures  $n$  measurements with  $n \ll N$  and  $n = O(k)$ , where  $N$  is the dimension of the full high-resolution image which is assumed to be  $k$ -sparse. However, in practice, it improves upon the existing CS approach by additionally achieving the following:

1. *Adaptivity to the image content.* The acquisition algorithm is still of complexity  $O(n)$ , but it adaptively takes more measurements if needed to achieve some compressed image target quality.
2. *A fast and efficient computational framework.* The algorithm does not require the computational intensive CS recovery algorithm. The decoding process has the computational complexity of existing algorithms in use today such as JPEG or JPEG2000 decoding.

The algorithms described in this paper were inspired by the CS idea of a sensing process that acquires a relatively small number of samples from which a sparse signal can be reconstructed. However, they depart fundamentally from the classic (nonadaptive) CS mathematical framework. Instead of acquiring the visual data using functionals that are random in nature and incoherent with the sparse basis, e.g., wavelets, we suggest a simple, direct adaptive, and fast method for the acquisition of transform coefficients. We show that it is possible to use devices such as the DMD array in a single-pixel camera for the computation of arbitrary transform coefficients, where the computation of each coefficient requires a fixed number of measurements (see section 2.2). With this tool in hand, instead of asking  $n$  universal, nonadaptive questions about the signal in a form of linear functionals, our method takes linear measurements, directly in a sparse and structured basis, with each measurement based on the previous measurements. This allows us to take advantage of the geometric structure of the sparse representation explicitly, such as imposing the modeling of wavelet tree structures over image edges to gather its significant information.

The sampling process is initialized by acquiring a very small number of low-resolution measurements. Then, based on this initial data, by utilizing specific properties of the wavelet structure, our algorithm adaptively extracts the significant information by using a near-minimal number of measurements  $n \ll N$ ,  $n = O(k)$ . This significant information, i.e., wavelet coefficients which correspond to image edges, is gathered based only on the subset of wavelet coefficients that were previously sampled and on an established set of rules. Therefore, our method is called *Adaptive Compressed Sampling (ACS)*. Since the ACS algorithm takes only linear sampling directly in the wavelet domain, there is no need for the standard CS recovery process from “pseudorandom” measurements.

The outline of the paper is as follows: in section 2 we recall the basic properties of wavelets and explain how the DMD architecture can be potentially used to acquire any transform coefficients in  $O(1)$ . In section 3 we provide the ACS version of an exact reconstruction result for cartoon images using a wavelet domain sampling scheme. Using basic tools from differential geometry, we prove that our ACS scheme “efficiently” samples all the nonzero wavelet coefficients of cartoon images, which correspond to the edges, thereby providing exact recovery with “optimal” (up to a constant) sampling.

In section 4 we review the Stat-ACS, an adaptive sampling algorithm that applies well-known statistical models of wavelet representations of images. Modeling of wavelet trees over image edges has already been used in the context of CS [2], [44], [48] but in a different and implicit way to speed up a reconstruction algorithm (1.2), as well as to speed up the numeric solution and to reduce the number of measurements. In particular in [2] it was shown that tree-sparse signals can be sampled using traditional CS with  $O(k)$  samples. In contrast, we use adaptively statistical tree models in a direct and explicit way, namely, at the time of the acquisition. The Stat-ACS algorithm is of higher complexity than the algorithm we presented in [33], but is more efficient in predicting significant wavelet coefficients and hence provides a better rate/distortion performance. We also provide comparison results with nonadaptive CS methods. Finally, in section 5, we present a more advanced algorithm, the Tex-ACS, which addresses the model of images as a mix of cartoon and local texture patches. The Tex-ACS algorithm uses a dictionary composed of local Fourier and wavelet basis functions. Experimental results presented for both Stat-ACS and Tex-ACS clearly show that the adaptive form of CS provides state-of-the-art results in terms of higher reconstruction quality for the same number of samples.

## 2. Direct sampling of transform coefficients.

**2.1. The wavelet tree structure.** Recall that the univariate wavelet system (e.g., [31], [52]) is a family of real functions  $\{\psi_{j,k} : (j,k) \in \mathbb{Z}^2\}$  in  $L_2(\mathbb{R})$ , built by dilating and translating a unique mother wavelet function  $\psi(x)$

$$(2.1) \quad \psi_{j,k} = 2^{-j/2} \psi(2^{-j}x - k),$$

where the mother wavelet satisfies

$$(2.2) \quad \int_{\mathbb{R}} \psi(x) dx = 0.$$

The wavelet system is a basis of  $L_2(\mathbb{R})$ , where for each  $f \in L_2(\mathbb{R})$ ,

$$(2.3) \quad f = \sum_{j,k} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k},$$

where  $\tilde{\psi}$  is a dual of  $\psi$ . For special choices of  $\psi$ , the set  $\{\psi_{j,k}\}$  forms an orthonormal basis for  $L_2(\mathbb{R})$ , and then  $\psi = \tilde{\psi}$ . Wavelets are usually constructed from a *multiresolution analysis* (MRA), which is a sequence of closed subspaces of  $L^2(\mathbb{R})$ ,

$$\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \dots, \quad \cap V_j = \{0\}, \quad \cup_j V_j = L_2(\mathbb{R}).$$

Usually, one starts the construction from a scaling function (generator)  $\varphi \in L_2(\mathbb{R})$  that satisfies a two-scale equation

$$\varphi = \sum_k a_k \varphi(2 \cdot -k).$$

In the orthonormal case, one then sets  $V_j = \overline{\text{span}} \{\varphi_{j,k} := 2^{-j/2} \varphi(2^{-j} \cdot -k) : k \in \mathbb{Z}\}$  and constructs  $\psi$  such that  $W_j := \overline{\text{span}} \{\psi_{j,k} : k \in \mathbb{Z}\}$ , with  $V_{j+1} \oplus W_{j+1} = V_j$ . A classic example for an orthonormal system is the Haar scaling function and wavelet:

$$(2.4) \quad \varphi(x) := \begin{cases} 1, & x \in [0, 1], \\ 0 & \text{else,} \end{cases} \quad \psi(x) := \begin{cases} 1, & x \in [0, \frac{1}{2}), \\ -1, & x \in [\frac{1}{2}, 1], \\ 0 & \text{otherwise.} \end{cases}$$

To obtain smooth symmetric wavelets of compact support one needs to construct biorthogonal wavelets, where  $\psi \neq \tilde{\psi}$ . This is achieved by using a dual set of scaling functions and MRAs [31].

The wavelet model can be easily generalized to any dimension [31], [52] via a tensor product of the wavelet and the scaling functions. Assume that the univariate wavelet  $\psi$  and the scaling function  $\varphi$  are given. Then, a bivariate separable basis is constructed using three basic wavelets

$$\psi^1(x_1, x_2) = \varphi(x_1)\psi(x_2), \quad \psi^2(x_1, x_2) = \psi(x_1)\varphi(x_2), \quad \psi^3(x_1, x_2) = \psi(x_1)\psi(x_2).$$

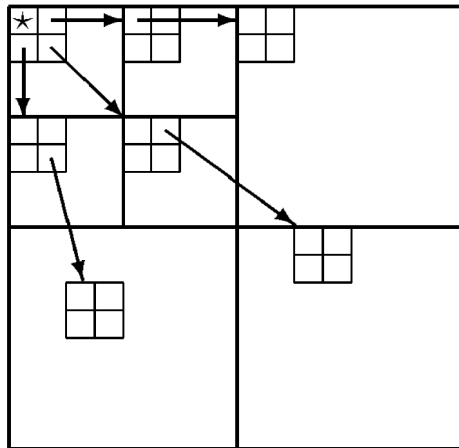
The bivariate discrete transform represents an image  $f \in L_2(\mathbb{R}^2)$ , in terms of the three wavelet functions  $\{\psi^1, \psi^2, \psi^3\}$ :

$$(2.5) \quad f = \sum_{j \in \mathbb{Z}, e=1,2,3} w_{j,k}^e \psi_{j,k}^e,$$

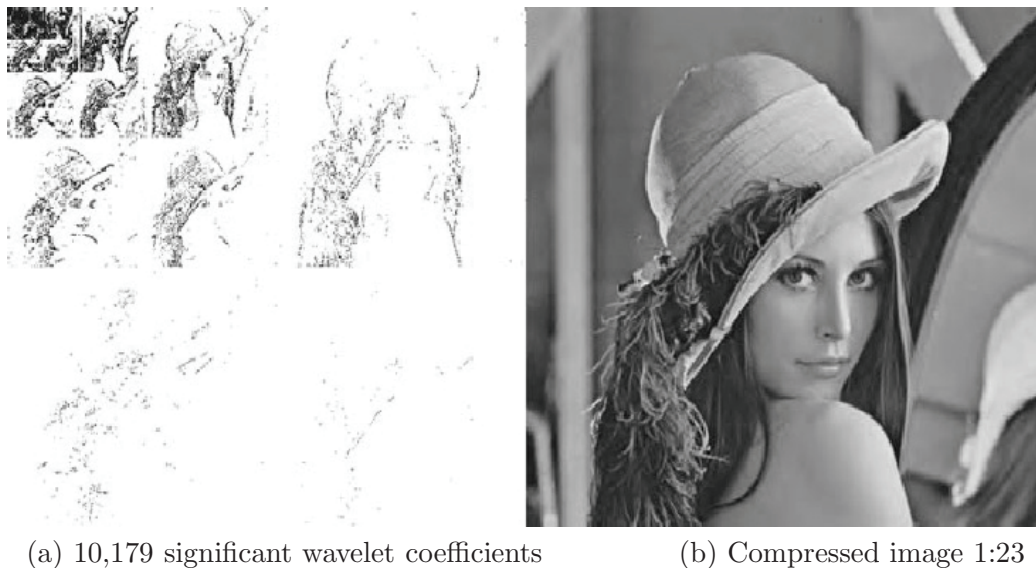
where  $\{w_{j,k}^e\}$  are coefficients. The wavelet decomposition can thus be interpreted as a signal decomposition in a set of three spatially oriented frequency subbands:  $LH$  ( $e = 1$ ) detects horizontal edges,  $HL$  ( $e = 2$ ) detects vertical edges, and  $HH$  ( $e = 3$ ) detects diagonal edges.

A wavelet coefficient  $w_{j,k}^e$  at a scale  $j$  represents the information about the image in the spatial region in the neighborhood of  $2^j k$ ,  $k \in \mathbb{Z}^2$ . At the next finer scale  $j-1$ , the information about this region is represented by four wavelet coefficients, which are described as the children of  $w_{j,k}^e$ . This leads to a natural tree structure organized in a quad tree structure of each of the three subbands as shown in Figure 1. Note that for each finer scale, the coefficients represent a smaller spatial area of the image at higher frequencies. As  $j$  decreases, the child coefficients add finer and finer details into the spatial regions occupied by their ancestors.

Wavelet representations are considered very efficient for image compression: the edges constitute a small portion of a typical image, and a wavelet coefficient is large only if edges are present within the support of the wavelet. Consequently, the image can be approximated well using a few large wavelet coefficients. A significant statistical structure also follows: large/small values of wavelet coefficients tend to propagate through the scales of the quad trees (see section 4.1 for details). As an example, a sparse wavelet representation of the  $512 \times 512$  image Lena and a compressed version of this image are shown in Figure 2, where



**Figure 1.** Wavelet coefficient tree structure across the subbands (MRA decomposition).



**Figure 2.** (a) A sparse wavelet representation of an image. Black—significant coefficient; white—insignificant coefficient. (b) Compressed JPEG2000 image based on the sparse representation of (a).

the compression algorithm was based on the sparse representation. The figure clearly depicts that the significant wavelet coefficients (coefficients with relatively large absolute value) are located on strong edges of the image.

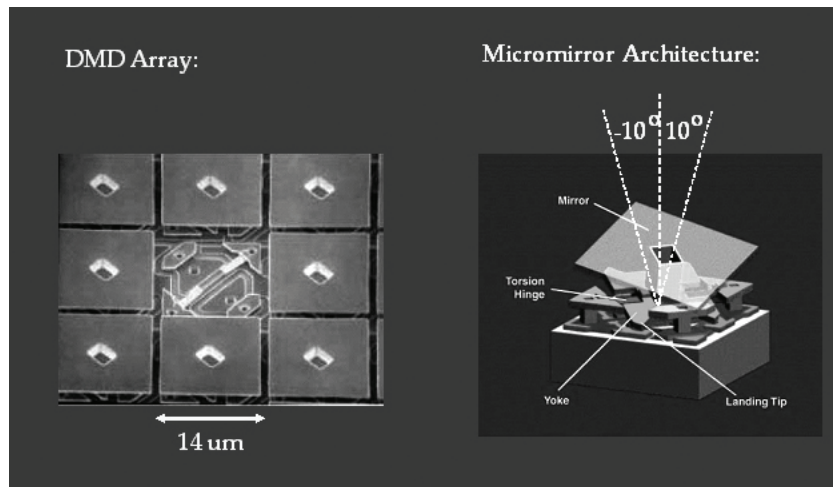
**2.2. Direct sampling of transform coefficients using digital micromirror arrays.** Today's digital camera devices are in the megapixel range thanks to the introduction of *charged coupled devices (CCDs)* and *complementary metal oxide semiconductor (CMOS)* digital technology. Assume that  $I$  is a digital image consisting of  $N$  pixels that we wish to acquire. Some digital cameras perform the acquisition using an array of  $N$  CCDs after exposure to the optical image.

The charges on its pixels are read row by row, are fed to an amplifier, and then undergo an analog-to-digital conversion. CMOS technology allows today's professional cameras to acquire about 12 million pixels ( $3000 \times 4000$ ). Once the digital image  $I$  has been acquired, it is usually compressed using standard compression algorithms such as JPEG and JPEG2000. In most digital cameras, the user has the capability to control the compression/quality tradeoff through the compression algorithm. We now describe the *digital micromirror device (DMD)* hardware architecture, depicted in Figure 3, that was also used for the single-pixel camera of [38].

The DMD consists of an array of electrostatically actuated micromirrors, where each mirror of the array is suspended above an individual SRAM cell. Each mirror rotates about a hinge and can be positioned in one of two states:  $+10$  or  $-10$  degrees from horizontal. Using lenses, all the reflections from the micromirrors at a given state (e.g., the light reflected by mirrors in  $+10$  degree state) are focused and collected by a single photodiode to yield an absolute voltage. The output of the photodiode is amplified through an op-amp circuit and then digitized by an analog-to-digital converter. This value should be interpreted as

$$v = \sum_{i=1}^N x_i 1_{\theta_i=+10} + \text{DC offset},$$

where  $x_I = (x_1, \dots, x_N)$  is the vector formulated by concatenating the rows of the digital image  $I$ , the indicator function  $1_{\theta_i=+10}$  obtains the value 1 if the  $i$ th micromirror is at the state  $+10$  degrees and 0 otherwise, and the DC offset is the value outputted when all the micromirrors are set to  $-10$  degrees.



**Figure 3.** The architecture of the DMD.

In [38], the rows of the CS sampling matrix  $\Phi$  are in fact a sequence of  $n$  pseudorandom binary masks, where each mask is actually a scrambled configuration of the DMD array (see also [3]). Thus, the measurement vector  $y$  is composed of inner-products of the digital image  $x$  with pseudorandom masks. In [38] the measurements are collected into a compressed bitstream with possible lossy or lossless compression applied to the elements of  $y$ . At the core of the

decoding process, which takes place at the viewing device, there is the minimization algorithm (1.2). Once a solution is computed, one obtains from it an approximate “reconstructed” image by applying the transform  $\Psi^{-1}$  to the coefficients. In [38], the authors choose the transform  $\Psi$  to be a wavelet transform, but their sampling process has a universal property. This means that any transform  $T$  which is in incoherence with a process of pseudorandom masks  $\Phi$  (almost all reasonable transforms have this property) and in which the image has a sparse representation will lead to a reconstruction with good approximation of the original image.

Recall that our proposed approach, although inspired by the CS concept of acquisition processes of simultaneous sensing and compression of the image, departs fundamentally in the mathematical framework. Instead of acquiring the visual data using a representation that is incoherent with wavelets, we sample directly in the wavelet domain. At a glance, this might seem to be a paradox, since computing the fast wavelet transform of an  $N$ -pixel image requires  $O(N)$  computations, whereas we want to take only  $n$  measurements with  $n \ll N$ . Furthermore, computing even one low-frequency wavelet coefficient requires calculating an integral over a significant portion of the image pixels, which requires  $O(N)$  computations. As we shall see, this paradox is solved by using the DMD in a different way than in [38].

There exist DMD arrays with the functionality that a micromirror can “reflect” a grayscale value (contemporary DMD can produce 1024 grayscale values). In theory, one can use these devices to compute arbitrary real-valued functionals acting on the data, since any functional  $g$  can be represented as a difference of two nonnegative  $g_+, g_-$  functionals, such that  $g = g_+ - g_-$  ( $g_+, g_- \geq 0$ ). This allows us to compute any transform coefficient using only two measurements. Also, we may leverage on the “feedback” architecture of the DMD and make decisions on future measurements based on existing value measurements. As we shall see, this adaptive sampling process relies on a well-known modeling of image edges using a wavelet coefficient tree structure. Then, decisions on which wavelet coefficients should be sampled next are based on the values of the wavelet coefficients obtained so far.

Let us now give the example of the acquisition of an arbitrary Haar wavelet coefficient, computed by using the even simpler binary DMD with only two measurements. The wavelet coefficient associated with a Haar wavelet in subband  $e = 1$  is given by

$$(2.6) \quad \langle f, \psi_{j,k}^1 \rangle = 2^{-j} \left( \int_{2^j k_1}^{2^j(k_1+1)} \int_{2^j k_2}^{2^j(k_2+1/2)} f(x_1, x_2) dx_1 dx_2 - \int_{2^j k_1}^{2^j(k_1+1)} \int_{2^j(k_2+1/2)}^{2^j(k_2+1)} f(x_1, x_2) dx_1 dx_2 \right).$$

This implies that for the purpose of computing a wavelet coefficient of the first type using a binary DMD array, we need to rotate and collect twice, into the photodiode, responses from two subsets of micromirrors, each supported over neighboring rectangular regions corresponding to the scale  $j$  and location  $k$ . From (2.6), the value of the wavelet coefficient we wish to acquire is the difference of these two outputs multiplied by  $2^{-j}$ .

We emphasize that there is ongoing research of compressed image acquisition applications where DMD arrays might have an advantage over conventional CCD architectures. One of these areas is hyperspectral imaging, where the acquisition device samples an image in multiple bands by using configurable optics [47]. In [61], the authors show promising results

in hyperspectral imaging using the DMD array. The adaptive approach presented in this work can potentially improve their nonadaptive scheme.

Finally, as pointed out by the referees, we note that in this work we do not investigate the influence of DMD accuracy on the quality of reconstruction. Typically, when noncompressive sensing is followed by image compression, transform coefficients are first calculated in relatively high precision and then quantized for the purpose of compression. When one acquires transform coefficients directly using the DMD, they are contaminated with noise, due to the relatively low precision of the DMD hardware. We postulate that this has a smaller effect when the goal is low bit-rate compression and transform coefficients are quantized significantly with quantization steps that are larger than the noise level of the hardware system. Obviously, this approach will become even more valid as DMD technology matures and more accurate systems enter the market.

**3. Exact reconstruction of cartoon images using ACS.** The theory of CS contains results on the ability to achieve exact recovery of a sparse signal from a limited number of linear measurements [14], [17], [35]. The prototype “sparse image” used in CS is the piecewise constant “cartoon” Shepp–Logan image (e.g., [14]), since such images are sparse in the wavelet domain and also have “small” total-variation norm. In this section we present a simple theoretical result on exact reconstruction via critical sampling that serves as a motivation for our ACS algorithms. We prove that if a function is a typical cartoon image, i.e., a piecewise constant function over domains with smooth boundaries, then it is possible to reconstruct it exactly by adaptive wavelet coefficient sampling, using an optimal number of measurements (up to a constant). In the following a “sample” is the evaluation of the given function by a linear functional, specifically a wavelet coefficient. In this section, where the given functions are not discrete, we use the convention as in [31] that scales are finer as the index  $j$  is larger (which is the opposite of the usual notation in the discrete setting).

**Theorem 3.1.** *Let  $f(x) = c_1 + c_2 \mathbf{1}_\Omega(x)$ , where the boundary  $\partial\Omega$  is a simple closed  $C^2$  (twice continuously differentiable) curve. Then, there exists a scale  $T \in \mathbb{Z}$ , such that if all the nonzero Haar wavelet (2.4) coefficients  $\langle f, \psi_{j,k}^e \rangle$  are known for  $j \leq T$ , then all nonzero coefficients  $\langle f, \psi_{j,k}^e \rangle$ ,  $j > T$ , can be acquired by an ACS process, where for each resolution*

$$(3.1) \quad \# \{ \text{required samples at resolution } j \} \leq 12 \# \{ \langle f, \psi_{j-1,k}^e \rangle \neq 0 : e = 1, 2, 3, k \in \mathbb{Z}^2 \}.$$

*Remark.* Observe that the above theoretical result is of a different nature than previous “exact reconstruction” results in CS. Most importantly, the setting is not discrete. The simple claim is that, from fine enough resolution, the location of nonzero wavelet coefficients at some resolution is completely predictable from the location of nonzero coefficients at the lower resolution. This motivates the Stat-ACS adaptive sampling algorithm (section 4) that is based on wavelet trees and predicts the location of significant coefficients by using information derived from coefficients already sampled at lower scales. However, since images typically are modeled as the combination of cartoon and texture models, the Tex-ACS algorithm (section 5) is needed as a complement.

We begin with some definitions from differential geometry [43].

**Definition 3.2.** *Given three points  $x = \gamma(s)$ ,  $y = \gamma(\sigma)$ ,  $z = \gamma(\tau)$  on a curve  $\gamma$ , we define the*

local radius of curvature  $\rho(x)$  at each point  $x$  of  $\gamma$  by

$$(3.2) \quad \rho(x) := \lim_{\tau, \sigma \rightarrow x} r(x, y, z),$$

where  $r(x, y, z)$  is the radius of the unique circle which passes through the three points  $x, y, z$ . In the case where a line passes through the points we say that  $r(x, y, z) = \infty$ . Assuming a natural parameterization, the curvature  $k(s)$  at a point  $\gamma(s)$  is

$$(3.3) \quad k(s) := |\gamma''(s)| = \frac{1}{\rho(x)}.$$

The minimal local radius of curvature of the curve  $\gamma$  is

$$(3.4) \quad \rho(\gamma) := \min_{x \in \gamma} \rho(x).$$

**Definition 3.3.** Given a simple closed  $C^2$  curve  $\gamma$ , we define the global radius of curvature  $\rho_G(x)$  at each point  $x$  of  $\gamma$  by

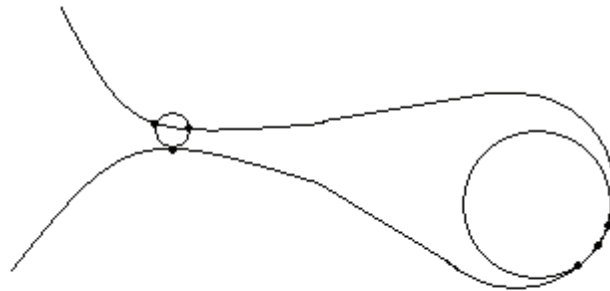
$$(3.5) \quad \rho_G(x) = \min_{y, z \in \gamma} r(x, y, z),$$

$$(3.6) \quad \rho_G(\gamma) := \min_{x \in \gamma} \rho_G(x).$$

It is clear that the global radius of curvature is bounded by the local radius of a curvature, i.e.,

$$(3.7) \quad 0 < \rho_G(x) \leq \rho(x) \quad \forall x \in \gamma.$$

In Figure 4 we see an example that shows the difference between the global radius of a curvature and the local radius of the curvature. In contrast to a local radius of a curvature, the global radius of the curvature contains information about nonlocal parts of the curve.



**Figure 4.** Global radius of curvature (left circle) and local radius of curvature (right circle).

Let  $\Omega$  be a domain with a simple and closed  $C^2$  boundary  $\partial\Omega$ . Assume that the boundary is given in the natural parametric form by

$$(3.8) \quad \gamma(s) = (x(s), y(s)), \quad s \in [0, L],$$

where  $L$  is the length of the curve,

$$(3.9) \quad \|\gamma''(s)\|_\infty = \max_{0 \leq s \leq L} |\gamma''(s)| \leq M,$$

and

$$(3.10) \quad 0 < \rho_G(\gamma) = M_G < \infty.$$

Let  $Q_{j,k}$  denote a dyadic square in  $\mathbb{R}^2$  (recall that in this section, the scales are finer as  $j \rightarrow \infty$ ):

$$(3.11) \quad Q_{j,k} := \left[ \frac{k_1}{2^j}, \frac{(k_1+1)}{2^j} \right] \times \left[ \frac{k_2}{2^j}, \frac{(k_2+1)}{2^j} \right], \quad j \in \mathbb{Z}, \quad k = (k_1, k_2) \in \mathbb{Z}^2.$$

**Lemma 3.4.** *Let  $\gamma$  be a smooth, closed, and simple curve. If for  $j \in \mathbb{Z}$*

$$(3.12) \quad 2^{-j} \leq \rho_G(\gamma),$$

*then the circumscribed circle  $O_{j,k}$  of any cube  $Q_{j,k}$ ,  $k \in \mathbb{Z}^2$ , contains at most one connected segment of the curve  $\gamma$ .*

*Proof.* Since  $\gamma$  is a non-self-intersecting curve, we have  $\rho_G(\gamma) > 0$ . Assume, in contrast, that  $O_{j,k}$  contains two segments of  $\gamma$ ,  $\gamma_1 \equiv \gamma|_{[s_1, s_2]}$  and  $\gamma_2 \equiv \gamma|_{[s_3, s_4]}$ , with  $s_1 < s_2 < s_3 < s_4$ . Thus, the radius  $R$  of the circle passing through  $\gamma(s_1)$ ,  $\gamma(s_2)$ , and  $\gamma(s_3)$  satisfies

$$(3.13) \quad \rho_G(\gamma) \leq R = \frac{\sqrt{2}}{2} 2^{-j} < 2^{-j},$$

which contradicts (3.12). We conclude that  $O_{j,k}$  may contain at most one connected component of  $\gamma$ . ■

Thus, we may assume that if  $j$  satisfies (3.12), then either  $\gamma \cap O_{j,k} = \emptyset$  or  $\gamma \cap O_{j,k}$  is a single connected segment. Let us assume now the second case, where  $\gamma \cap O_{j,k}$  is a single segment, and let the entry and exit points of  $\gamma$  in  $O_{j,k}$  be  $\gamma(s_1)$ , and  $\gamma(s_2)$ , respectively. Define the following line:

$$(3.14) \quad \tilde{\gamma}(s) = \gamma(s_1) + (s - s_1)\gamma'(s_1).$$

**Lemma 3.5.** *If  $j$  satisfies for  $\alpha \geq 1$*

$$(3.15) \quad 2\alpha\pi^2 M \leq 2^j,$$

*with  $M$  given in (3.9), then the distance between  $\gamma$  and the segment  $\tilde{\gamma}$ , with  $\tilde{\gamma}(s) = (\tilde{x}(s), \tilde{y}(s))$ , given in (3.14) obeys*

$$(3.16) \quad \min_{s'} |\gamma(s) - \tilde{\gamma}(s')| \leq \frac{2^{-j}}{\alpha}, \quad s \in [s_1, s_2].$$

*Proof.* Since  $2M \leq 2^j$ , the radius of curvature of  $\gamma$  is bigger than the radius of  $O_{j,k}$ . Therefore, the maximal length of the single connected curve segment contained in  $O_{j,k}$  is bounded by the circumference of  $O_{j,k}$ :

$$(3.17) \quad \text{length}(\gamma \cap O_{j,k}) = s_2 - s_1 \leq \sqrt{2}\pi 2^{-j}.$$

Using the differentiability properties of  $\gamma(s) = (x(s), y(s))$ , we obtain

$$|x(s) - \tilde{x}(s)| \leq \|x''\|_\infty (s - s_1)^2, \quad |y(s) - \tilde{y}(s)| \leq \|y''\|_\infty (s - s_1)^2 \quad \forall s \in [s_1, s_2].$$

Applying (3.15) and (3.17) for the parametric distance  $|\gamma(s) - \tilde{\gamma}(s)|$  yields

$$\begin{aligned} |\gamma(s) - \tilde{\gamma}(s)| &= \sqrt{(x(s) - \tilde{x}(s))^2 + (y(s) - \tilde{y}(s))^2} \\ &\leq M(s_2 - s_1)^2 \leq M(\sqrt{2}\pi 2^{-j})^2 \\ &= 2\pi^2 M 2^{-2j} \leq \frac{2^{-j}}{\alpha}. \end{aligned}$$

Since the geometric distance is always smaller than the parametric distance, we obtain (3.16). ■

Let  $j$  be a sufficiently fine (large) scale which obeys (3.12) and (3.15) for  $\alpha = 16$ . As depicted in Figure 5, let  $I$  be the strip composed from two linear segments  $\tilde{\gamma}_-$  and  $\tilde{\gamma}_+$  that are parallel to  $\tilde{\gamma}$  and obtained by shifts of  $\tilde{\gamma}$  by  $-2^j/16$  and  $+2^j/16$ , respectively, in the direction of the normal of  $\tilde{\gamma}$ . By construction of  $\tilde{\gamma}_-$ ,  $\tilde{\gamma}_+$  and by (3.16), the curve  $\gamma$  is confined to the strip  $I$  in  $Q_{j,k}$ . Define the width of  $I$ ,  $w(I)$ , to be the Hausdorff distance between  $\tilde{\gamma}_-$  and  $\tilde{\gamma}_+$ . Obviously, by construction

$$(3.18) \quad w(I) = 2^{-j}/8.$$

Recall that  $f(x) = c_1 + c_2 \mathbf{1}_\Omega(x)$ . For  $i = 1, \dots, 4$ , let  $S_i = \text{area}(\tilde{Q}_i \cap \Omega)$ , where  $\{\tilde{Q}_i\}$ ,  $i = 1, \dots, 4$ , are the four quadrants of  $Q_{j,k}$ , i.e., the four dyadic cubes at level  $j+1$  that are contained in  $Q_{j,k}$ . Also, let  $S(I) = \text{area}(I \cap Q_{j,k})$ .

In Figure 5, we see an example of a curve component which is contained in a dyadic cube  $Q_{j,k}$  and bounded by a strip  $I$ . Our proof rests on the fact that we can make the width of the strip  $I$  to be relatively sufficiently small, such that the equality  $S_1 = S_2 = S_3 = S_4$  cannot hold.

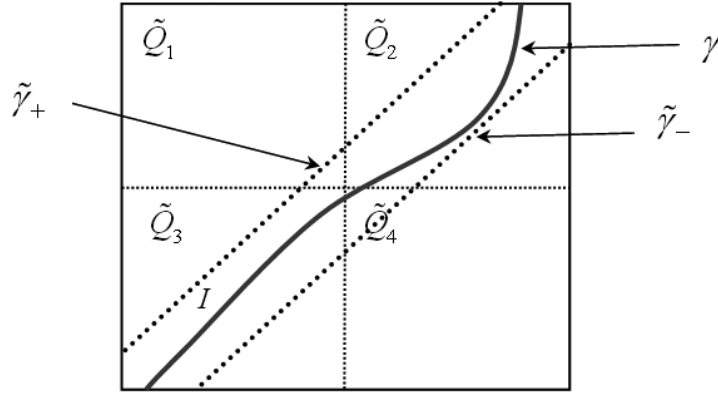
**Lemma 3.6.** *Assume that  $j$  satisfies (3.15) for  $\alpha \geq 16$ . The equality  $S_1 = S_2 = S_3 = S_4$  holds for  $Q_{j,k}$  only if*

$$(3.19) \quad S_i = 0, \quad i \in \{1, 2, 3, 4\},$$

or

$$(3.20) \quad S_i = \frac{2^{-2j}}{4}, \quad i \in \{1, 2, 3, 4\}.$$

*Proof.* Denote the center of  $Q_{j,k}$  by  $o$ . If  $o \in (I)^C$ , then  $\gamma$  does not intersect with at least one of the quadrants  $\tilde{Q}_i$  and therefore  $S_1 = S_2 = S_3 = S_4$  can hold only if  $Q_{j,k} \cap \Omega = \emptyset$ ,



**Figure 5.** A curve component bounded by the strip  $I$  in  $Q_{j,k}$ .

corresponding to (3.19), or  $Q_{j,k} \cap \Omega^c = \emptyset$ , corresponding to (3.20). Otherwise  $o \in I$  (as depicted in Figure 5). By (3.18)

$$S(I) \leq w(I) \cdot \sqrt{2} \cdot \text{side}(Q_{j,k}) = \frac{2^{-j}}{8} \cdot \sqrt{2} \cdot 2^{-j} = \frac{\sqrt{2}}{8} 2^{-2j}.$$

From a simple combinatorial argument, there exists a quadrant  $\tilde{Q}_i$  for which the intersection with  $I$  is equal to or smaller than a quarter of  $S(I)$ :

$$(3.21) \quad S(\tilde{Q}_i \cap I) \leq \frac{\sqrt{2}}{32} 2^{-2j}.$$

Without loss of generality, assume that (3.21) holds for  $i = 1$  and that  $S_1 > 0$ . This implies that

$$(3.22) \quad S_1 \geq \text{area}(\tilde{Q}_1) - S(\tilde{Q}_1 \cap I) \geq \frac{2^{-2j}}{4} - \frac{\sqrt{2} 2^{-2j}}{32} \geq \frac{6 \cdot 2^{-2j}}{32}.$$

On the other hand, for at least one of the other quadrants,  $\tilde{Q}_i$ ,  $i \in \{2, 3, 4\}$ ,

$$(3.23) \quad S_i \leq S(\tilde{Q}_i \cap I) \leq S(I) \leq \frac{\sqrt{2}}{8} \cdot 2^{-2j}.$$

Thus,  $S_i < S_1$  for one of  $i = 2, 3, 4$ , and therefore,  $o \in I$  is not possible. This completes the proof of the lemma. ■

*Proof of Theorem 3.1.* Let  $f(x) = c_1 + c_2 \mathbf{1}_\Omega(x)$ , and let  $\gamma = \partial\Omega$ . Let  $Q_{j,k}$  be a dyadic cube. If  $Q_{j,k} \cap \gamma = \emptyset$ , then  $w_{j,k}^e = \langle f, \psi_{j,k}^e \rangle = 0$  for  $e = 1, 2, 3$ . We claim that if  $Q_{j,k} \cap \gamma \neq \emptyset$ , then for sufficiently large  $j$  it is impossible that  $w_{j,k}^1 = w_{j,k}^2 = w_{j,k}^3 = 0$  simultaneously. Indeed, assume  $Q_{j,k}$ ,  $j \geq T$ , where for  $\alpha \geq 16$

$$(3.24) \quad T := \max(0, \lceil \log_2(2\alpha\pi^2 M) \rceil, \lceil \log_2(\rho_G(\gamma)) \rceil).$$

Assume, by contradiction, that  $Q_{j,k} \cap \gamma \neq \emptyset$  and that  $w_{j,k}^1 = w_{j,k}^2 = w_{j,k}^3 = 0$ . Observe that, using the notation where the scales are finer as  $j \rightarrow \infty$ ,

$$(3.25) \quad w_{j,(k_1,k_2)}^1 = 2^j \left( \int_{2^{-j}k_1}^{2^{-j}(k_1+1)} \int_{2^{-j}k_2}^{2^{-j}(k_2+1/2)} f(x,y) dx dy \right. \\ \left. - \int_{2^{-j}k_1}^{2^{-j}(k_1+1)} \int_{2^{-j}(k_2+1/2)}^{2^{-j}(k_2+1)} f(x,y) dx dy \right) = 0,$$

$$(3.26) \quad w_{j,(k_1,k_2)}^2 = 2^j \left( \int_{2^{-j}k_1}^{2^{-j}(k_1+1/2)} \int_{2^{-j}k_2}^{2^{-j}(k_2+1)} f(x,y) dx dy \right. \\ \left. - \int_{2^{-j}(k_1+1/2)}^{2^{-j}(k_1+1)} \int_{2^{-j}k_2}^{2^{-j}(k_2+1)} f(x,y) dx dy \right) = 0,$$

$$(3.27) \quad w_{j,(k_1,k_2)}^3 = 2^j \left( \int_{2^{-j}k_1}^{2^{-j}(k_1+1/2)} \int_{2^{-j}k_2}^{2^{-j}(k_2+1/2)} f(x,y) dx dy \right. \\ + \int_{2^{-j}(k_1+1/2)}^{2^{-j}(k_1+1)} \int_{2^{-j}(k_2+1/2)}^{2^{-j}(k_2+1)} f(x,y) dx dy \\ - \int_{2^{-j}k_1}^{2^{-j}(k_1+1/2)} \int_{2^{-j}(k_2+1/2)}^{2^{-j}(k_2+1)} f(x,y) dx dy \\ \left. - \int_{2^{-j}(k_1+1/2)}^{2^{-j}(k_1+1)} \int_{2^{-j}k_2}^{2^{-j}(k_2+1/2)} f(x,y) dx dy \right) = 0.$$

With the notation of  $\{S_i\}_{i=1}^4$  given above, the solution of (3.25)–(3.27) is obtained by solving the linear algebraic system

$$(3.28) \quad c_2 \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} = 0.$$

Since we may assume  $c_2 \neq 0$ , we obtain the set of solutions

$$(3.29) \quad S_1 = S_2 = S_3 = S_4.$$

As  $T$  satisfies (3.24) for  $\alpha \geq 16$ , by Lemma 3.6 we obtain that (3.29) is satisfied only if (3.19) or (3.20) holds. But this is a contradiction to our assumption that  $Q_{j,k} \cap \gamma \neq \emptyset$ .

The proof of the theorem can now be completed by a simple induction: at any resolution  $j \geq T$  assume that we have all nonzero wavelet coefficients  $\{w_{j,k}^e\}$ . For any such nonzero coefficient  $w_{j,k}^e$ , we sample the 12 wavelet coefficients

$$\{w_{j+1,l}^e \mid e = 1, 2, 3, l = \{(2k_1, 2k_2), (2k_1 + 1, 2k_2), (2k_1, 2k_2 + 1), (2k_1 + 1, 2k_2 + 1)\}\}.$$

If for the cube  $Q_{j+1,l}$ , we have  $w_{j+1,l}^e = 0$  for  $e = 1, 2, 3$ , then we know  $Q_{j+1,l} \cap \gamma = \emptyset$ , and therefore we do not sample any wavelet coefficients supported in this cube. Otherwise our ACS process continues to sample wavelet coefficients in  $Q_{j+1,l}$ . ■

#### 4. Stat-ACS: An ACS algorithm based on statistical modeling in the wavelet domain.

In [33], we presented a rather simple ACS algorithm, in which wavelet coefficients were sampled based on local Lipschitz smoothness estimates using their ancestors [23], [52]. In this work we elaborate on a more advanced statistical-based ACS algorithm: the *Stat-ACS*. We use information-theoretic analysis tools to estimate statistical dependencies between a wavelet coefficient and related coefficients at adjacent spatial locations and scales. The information obtained from previously sampled coefficients is summarized into a linear predictor which predicts the significance of a coefficient that has not yet been sampled. Since the Stat-ACS algorithm operates on a relatively complex dependence network, we employ hash tables to construct efficient dynamic dictionaries. Roughly speaking, the algorithm, of order  $O(n)$ , captures  $n$  measurements, with  $n \ll N$ , where  $N$  is the number of pixels. The goal of the algorithm is that  $n = O(k)$ , where the image is assumed to be well approximated by  $k$  wavelet terms. While this is motivated by the cartoon model of section 3, it is certainly not guaranteed. The Stat-ACS algorithm is more complex than the adaptive algorithm in [33]; its scheme increases the probability that significant wavelet coefficients will not be “missed” and will be sampled. Thus, the Stat-ACS algorithm suggests a tradeoff between a higher complexity of the algorithm and a higher efficiency in recovering edges and important features in the image.

**4.1. Wavelet statistical models.** Our ACS algorithm relies heavily on statistical modeling of images, as is used in tasks such as compression [8], [25], [29], [49], denoising [29], [71], and classification [24]. In most of these works, wavelet coefficients of natural images are modeled by generalized Gaussian marginal statistics. The *probability density function (PDF)* of the generalized Gaussian density is defined as

$$(4.1) \quad p(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta},$$

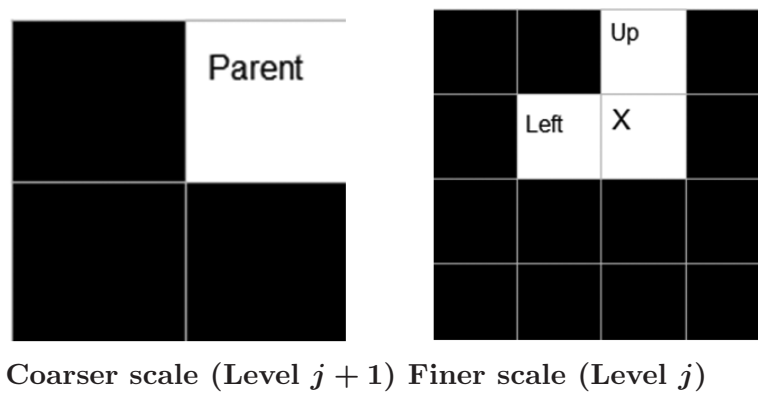
where  $\alpha$  models the width of the PDF peak,  $\beta$  is inversely proportional to the decreasing rate of the peak, and  $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ .  $z > 0$  is the Gamma function.

Also, these works reveal that for images, wavelet coefficients are not statistically independent mainly due to the geometric properties of edge singularities. Thus, at least for natural images, there are significant dependencies between the magnitudes of pairs of wavelet coefficients which correspond to adjacent spatial locations, orientation, and scales.

There are various statistical models which have been developed to capture the dependencies between wavelet coefficients. Image compression schemes that use spatial and scale-to-scale dependencies have proved to be extremely effective where a large magnitude wavelet coefficient may indicate that a neighbor wavelet coefficient at the same subband or at the next finer scale also has a large magnitude. The conditional distribution of a coefficient, given its coarser scale ancestors, was captured by the embedded zero-tree wavelet (EZW) coder [69] to encode entire trees of insignificant coefficients with a single symbol. A predictive method was used in [66] to provide high quality zero-tree coding results. Several studies [60], [59], [63] predict blocks of fine coefficients from blocks of coarse coefficients by using vectorized

look-up tables. Adaptive entropy coding was used in [67] to capture conditional statistics of coefficients that are based on the most significant bits of each of the eight spatial neighbors and the coefficients at the coarser scale. A switch between multiple probability methods that depend on the values of neighboring coefficients was done in [25]. The EZW coder was extended in [74] to use local coefficient contexts. The encoding of classification maps of regions of coefficients based upon the classes of the left and parent regions was described in [46]. The hidden Markov tree (HMT) model [29] also captures the dependencies between a parent and its children. In this work, the statistical correlation of an unsampled wavelet coefficient with neighbors that were already sampled is at the heart of our adaptive sensing algorithms.

We highlight some important wavelet coefficient dependencies as illustrated in Figure 6.  $X$  is a wavelet coefficient,  $P$  is its parent at the coarser scale, and  $NX$  is a predefined set of adjacent coefficients of  $X$  at the same subband.



**Figure 6.** Set of conditioning candidates for a given coefficient  $X$ .

Recall that the mutual information between two random variables  $X$  and  $Y$  with joint density  $p(x, y)$  and marginal densities  $p(x)$  and  $p(y)$  is defined in [68] to be

$$(4.2) \quad I(X; Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy.$$

The mutual information  $I(X; Y)$  can be interpreted as the amount of information  $Y$  conveys about  $X$  and vice versa. A higher value of  $I(X; Y)$  implies a stronger dependency between  $X$  and  $Y$ , and thus, it is easier to estimate  $X$  from  $Y$ . In practice, the random variables are discrete, and in our case the PDFs of  $p(x, y)$ ,  $p(x)$ , and  $p(y)$  are unknown and must be estimated from the empirical data. Studies of this problem resort to nonparametric histogram estimation of the PDFs [8], [58], [70]. The histograms are usually constructed as follows: the ranges of  $X$  and  $Y$  are partitioned into  $N_X$  and  $N_Y$  intervals, respectively. The histogram of  $(X, Y)$  obtained from the empirical data is denoted as  $\{p_{X,Y}(i, j) \mid 1 \leq i \leq N_X, 1 \leq j \leq N_Y\}$  and yields an approximation to the PDF  $p(x, y)$ . Similarly, the marginal distributions  $p(x)$ ,  $p(y)$  are estimated.

Consider a pair of wavelet coefficients  $X$  and  $Y$ . Based on the estimated discrete proba-

bility distribution, we use the following estimator for the mutual information [55]:

$$(4.3) \quad \hat{I}(X; Y) = \sum_i \sum_j p_{X,Y}(i, j) \log \frac{p_{X,Y}(i, j)}{p_X(i)p_Y(j)} = \sum_{i,j} \frac{k_{ij}}{N} \log \frac{k_{ij}N}{k_i k_j},$$

such that  $\frac{k_{ij}}{N}$  replaces  $p_{ij}$ , where  $k_{ij}$  is the number of coefficient pairs in the joint histogram cell  $(i, j)$ , and  $k_i = \sum_j k_{ij}$  and  $k_j = \sum_i k_{ij}$  are the marginal distribution histogram estimates.

Tables 1–3 show averaged mutual information results for a wavelet coefficient (we use the (9,7) wavelet basis), modeled to be dependent on a single other coefficient. The index refers to a relative location of a neighbor with respect to the location of  $X$ . For example,  $NX_{-1,0}$ ,  $NX_{0,-1}$  correspond to the immediate upward and left-hand neighbors of  $X$ . We see that similar results are obtained on a variety of different test images. We measured and compared the mutual information with a single neighbor (scale, location, subband). Roughly speaking, across the three subbands, the parent and the horizontal and vertical neighbors exhibit relatively higher mutual information than the diagonal neighbors.

We conclude this section by noting that recent studies have adopted the mutual information estimation for an efficient characterization of statistical dependencies in an overcomplete

**Table 1**

*Mutual information for scale = 3, horizontal subband.*

$\hat{I}$ (mutual information)	Lena	Peppers	Barbara
$\hat{I}(X; P)$	0.109	0.12	0.08
$\hat{I}(X; NX_{-1,-1})$	0.07	0.05	0.08
$\hat{I}(X; NX_{-1,0})$	0.103	0.08	0.123
$\hat{I}(X; NX_{-1,1})$	0.09	0.05	0.095
$\hat{I}(X; NX_{0,-1})$	0.119	0.115	0.132
$\hat{I}(X; NX_{0,1})$	0.1208	0.114	0.13
$\hat{I}(X; NX_{1,-1})$	0.09	0.04	0.09
$\hat{I}(X; NX_{1,0})$	0.1	0.07	0.11
$\hat{I}(X; NX_{1,1})$	0.07	0.12	0.07

**Table 2**

*Mutual information for scale = 3, vertical subband.*

$\hat{I}$ (mutual information)	Lena	Peppers	Barbara
$\hat{I}(X; P)$	0.127	0.141	0.07
$\hat{I}(X; NX_{-1,-1})$	0.088	0.05	0.1
$\hat{I}(X; NX_{-1,0})$	0.197	0.139	0.22
$\hat{I}(X; NX_{-1,1})$	0.106	0.07	0.11
$\hat{I}(X; NX_{0,-1})$	0.109	0.08	0.12
$\hat{I}(X; NX_{0,1})$	0.107	0.08	0.125
$\hat{I}(X; NX_{1,-1})$	0.109	0.06	0.11
$\hat{I}(X; NX_{1,0})$	0.199	0.14	0.22
$\hat{I}(X; NX_{1,1})$	0.087	0.05	0.1

Table 3

*Mutual information for scale = 3, diagonal subband.*

$\hat{I}$ (mutual information)	Lena	Peppers	Barbara
$\hat{I}(X; P)$	0.107	0.07	0.06
$\hat{I}(X; NX_{-1,-1})$	0.075	0.02	0.12
$\hat{I}(X; NX_{-1,0})$	0.12	0.05	0.18
$\hat{I}(X; NX_{-1,1})$	0.10	0.03	0.124
$\hat{I}(X; NX_{0,-1})$	0.101	0.05	0.147
$\hat{I}(X; NX_{0,1})$	0.101	0.05	0.145
$\hat{I}(X; NX_{1,-1})$	0.108	0.03	0.125
$\hat{I}(X; NX_{1,0})$	0.122	0.04	0.176
$\hat{I}(X; NX_{1,1})$	0.07	0.03	0.124

wavelet representation [45] and multiscale directional transforms such as contourlets [62] and curvelets [7].

#### 4.2. The Stat-ACS Algorithm.

**The linear predictor.** To impose the statistics of wavelet coefficients exhibited by the mutual information into the Stat-ACS algorithm, we utilize a predictor for the absolute value of a coefficient that has not yet been sampled, using the absolute value of coefficients that have already been sampled (or have been determined to be insignificant and not sampled). In our experiments we used the linear predictor

(4.4)

$$\text{Predicted } (|X|) = \begin{cases} 0.5 |P| + 0.28 |NX_{0,-1}| + 0.22 |NX_{-1,0}|, & X \text{ in horizontal or diagonal,} \\ 0.5 |P| + 0.28 |NX_{-1,0}| + 0.22 |NX_{0,-1}|, & X \text{ in vertical,} \end{cases}$$

where  $NX_{0,-1}$ ,  $NX_{-1,0}$ , are the left-hand and upward neighbors and  $P$  is the parent of the unsampled coefficient  $X$ . The fixed set of weights in (4.4) was “learned” from a collection of standard test images via a simple least square error minimization.

We now review in detail the way in which the joint statistical model is used in the framework of our Stat-ACS algorithm. We begin with preliminary notation: given an arbitrary wavelet coefficient corresponding to the index  $(e, j + 1, (k_1, k_2))$  and label “P” (as for parent), its four children at the next finer resolution correspond to the indices  $(e, j, 2k_1, 2k_2)$ ,  $(e, j, 2k_1 + 1, 2k_2)$ ,  $(e, j, 2k_1, 2k_2 + 1)$ ,  $(e, j, 2k_1 + 1, 2k_2 + 1)$ , where we label the indices  $(e, j, 2k_1, 2k_2 + 1)$  and  $(e, j, 2k_1 + 1, 2k_2)$  as “L” (as for left) and “U” (as for up), respectively. In this algorithm, the indices of the sampled coefficients found to be significant are placed in a processing queue, together with their labels parent, left, up, or right. The labels given to each coefficient refer to the relation between the input and output indices of the linear predictor: a coefficient with label “P” indicates that the linear predictor attempts to estimate the significance of its four children at the next higher resolution; a label of “L” indicates that the linear predictor attempts to estimate the right-hand neighbor of that coefficient; and a label of “U” indicates that the linear predictor attempts to estimate the down spatial neighbor of that coefficient. Note that a coefficient can have more than one label. The index and value of the coefficient at the head of the processing queue are passed as input to the linear

predictor given in (4.4), which attempts to assess whether the related unsampled coefficients are potentially significant.

In a manner similar to that of existing wavelet compression and denoising algorithms, the significance of a (predicted) coefficient is determined by comparing its (predicted) absolute value with a predetermined threshold. Thus, the threshold parameter is critical: it controls the rate/distortion tradeoff of the algorithm: a smaller threshold will lead to more samples, revealing more significant coefficients. Once the prediction asserts that an unsampled coefficient is with high probability significant, it is sampled by the sensing apparatus.

The algorithm scans each subband such that the most informative spatial neighbor could be used in the processing queue. Thus, in the horizontal subband, where the left neighbor conveys the highest information, the rows are scanned first (left-to-right order), and in the vertical and diagonal subbands, the columns are scanned first (up-to-right order) since the up coefficient is the most informative.

**The hash table.** Incorporating a contextual model into a Stat-ACS algorithm, which operates simultaneously on the information of several coefficients (parent, left, up neighbors), requires the use of a “sparse” data structure which supports fast “find” and “add” operations. This is essential in ensuring that our ACS acquisition algorithm is of complexity  $O(n)$ , where  $n$  is the number of measurements, with  $n \ll N$  ( $N$  is the number of pixels in the image). To this end we use a hash table [28] that on average performs find and add operations in  $O(1)$  time. During the adaptive sampling process, we store in the hash table the coefficients that were sampled, and we later retrieve their values for the purpose of predicting the significance of unsampled coefficients. We summarize the algorithm in the pseudocode below.

**Stat-ACS acquisition algorithm.**

1. Acquire the values of all low-resolution coefficients up to a certain low-resolution  $J$ ; each computation is completed using a fixed number of DMD array measurements as in (2.6). The initial resolution  $J$  can be selected as  $\lfloor \log_2 N/2 \rfloor - \text{const}$ . In any case,  $J$  should be bigger if the image is bigger. For images of size  $512 \times 512$ , we used  $J = 4$ . Note that the total number of coefficients at resolutions  $\geq J$  is  $2^{2(1-J)}N$ , which is a small fraction of  $N$ .
2. Initialize a processing queue,  $P(q)$ , containing the indices of each of the wavelet coefficients at resolution  $J$ , with absolute value bigger than a given threshold  $T_{thresh}$ , and label as a parent, “P.”
3. Until the queue  $P(q)$  is exhausted, pop out the coefficient at the head of the queue and process as follows:
  - (a) If labeled “P” with index  $(e, j, k)$  and children coefficients have not been sampled yet:
    - (i) Acquire them now.
    - (ii) If the child with index  $(e, j - 1, (2k_1, 2k_2 + 1))$  is significant, then add its index to the head of  $P(q)$  labeled as “L.”
  - (b) If in the horizontal subband and labeled “L” as left of a coefficient  $X$ , find and read from the hash table the values of “P,” the parent of  $X$ , and of “U,” the upper neighbor of  $X$ . The values are set to zero if coefficients are not in the hash table. Fix the input vector  $\vec{Q} = (\text{value}(\text{“P”}), \text{value}(\text{“L”}), \text{value}(\text{“U”}))$  for

the linear predictor (4.4). If  $|predict(|X|)| \geq T_{thresh}$ , then do the following:

- (i) Sample the group of 4 coefficients to which  $X$  belongs, i.e., having the same parent  $(e, j, k)$ .
- (ii) If a coefficient is significant and at resolution  $j > 1$ , add its index to the end of the queue, labeled as ‘‘P.’’
- (iii) If the coefficient with index  $(e, j - 1, 2k_1, 2k_2 + 1)$  is significant, then add to the head of the queue labeled as ‘‘L.’’

(This step is similar for coefficients from the other 2 subbands.)

- (c) Store the values of any sampled coefficient in the hash table.

**4.3. Stat-ACS: Experimental results.** A good method for evaluating the effectiveness of our approach is to benchmark it using the optimal  $n$ -term wavelet approximation. It is well known [34] that for a given image with  $N$  pixels, the optimal orthonormal wavelet approximation using only  $n$  coefficients is obtained using the  $n$  largest coefficients:

$$\begin{aligned} |\langle f, \psi_{j_1, k_1}^{e_1} \rangle| &\geq |\langle f, \psi_{j_2, k_2}^{e_2} \rangle| \geq |\langle f, \psi_{j_3, k_3}^{e_3} \rangle| \geq \dots, \\ \left\| f - \sum_{i=1}^n \langle f, \psi_{j_i, k_i}^{e_i} \rangle \psi_{j_i, k_i}^{e_i} \right\|_{L_2(\mathbb{R}^2)} &= \min_{\#\Lambda=n} \left\| f - \sum_{(e, j, k) \in \Lambda} \langle f, \psi_{j, k}^e \rangle \psi_{j, k}^e \right\|_{L_2(\mathbb{R}^2)}. \end{aligned}$$

For biorthogonal wavelets this ‘‘greedy’’ approach gives a near-best result, i.e., within a constant factor of the optimal  $n$ -term approximation. Obviously, to find the near-best  $n$ -term approximation one needs to compute all the wavelet coefficients and then select from them the  $n$  coefficients with largest absolute value. Any other threshold method would also require computing each and every coefficient and testing whether its absolute value is above a certain threshold. Thus, any threshold method requires order  $N$  computations, whereas our ACS algorithm is output sensitive and requires only order  $n$  computations.

To simulate our algorithm, we first precompute the entire wavelet transform of a given image. However, we strictly follow the recipe of our ACS algorithm and extract a wavelet coefficient from the precomputed coefficient matrix only if its index was added to the adaptive sampling queue.

In Figure 7(a) we see a ‘‘benchmark’’ near-best  $n = 7000$  term approximation of the image Lena using a biorthogonal (9,7) wavelet, which achieves a peak signal-to-noise ratio (PSNR) of 31.6. In Figure 7(b) we see our Stat-ACS with  $n = 14000$  and PSNR = 30.26. Here,  $J = 4$  and the thresholds  $T_{thresh}$  used were 0.28 for  $j = 1$ , 0.24 for  $j = 2$ , and 0.2 for  $j = 3$ . We see that we have sampled twice as many as in the sparse representation to which we compare.

Typically, in an application where the output is the compressed image, the 14000 sampled wavelet coefficients (Figure 7(b)) should be pruned, with insignificant coefficients removed with a minimal impact on PSNR and the remaining coefficients compressed using a tree-based wavelet compression algorithm [66], [69]. Indeed, in Figures 8(a) and 8(b) we see a 5000-term and an 8000-term approximation, with PSNR = 29.35 and PSNR = 30.1, respectively, extracted from the Stat-ACS approximation obtained in Figure 7(b).



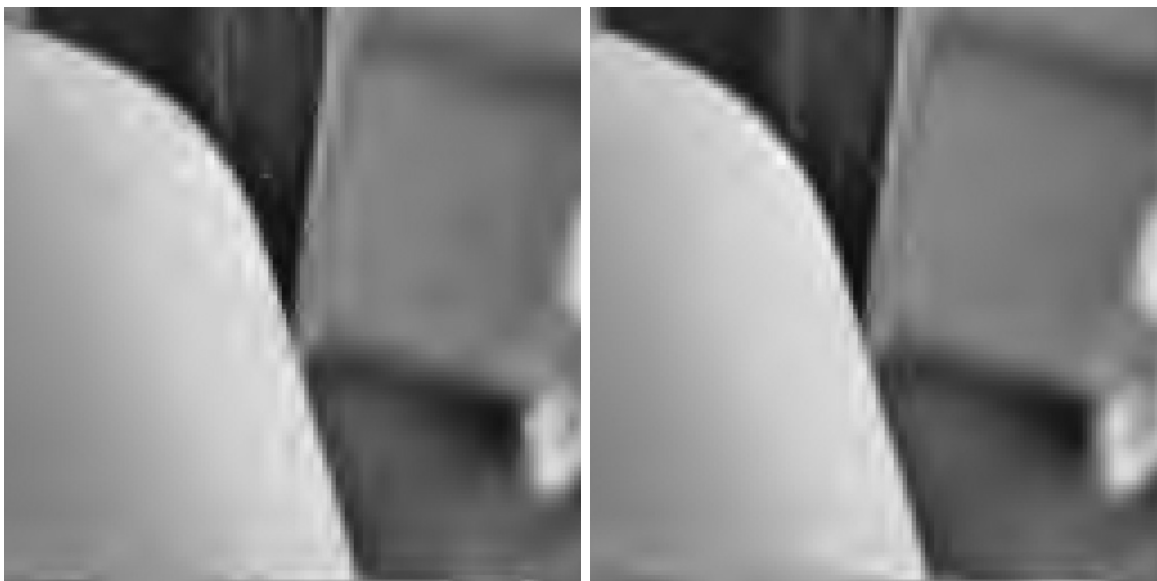
(a) Benchmark,  $n = 262144$ , 7000-term, PSNR=31.6. (b) Stat-ACS,  $n = 14000$ , 14000-term, PSNR=30.26.

**Figure 7.** Comparison of the Stat-ACS algorithm with optimal benchmark on Lena.



(a) Stat-ACS,  $n = 14000$ , 5000-term, PSNR=29.35. (b) Stat-ACS,  $n = 14000$ , 8000-term, PSNR=30.1.

**Figure 8.** Adaptive sampling with Stat-ACS with postprocessing pruning of “insignificant” coefficients.



(a) Benchmark,  $n = 262144$ , 7000-term, PSNR=31.6. (b) Stat-ACS,  $n = 14000$ , 14000-term, PSNR=30.26.

**Figure 9.** Zoom-in from Figure 7 on Lena's shoulder.

Figure 9 shows zooming in on specific edge regions from Figure 7. As Figure 9(b) illustrates, the edges were recovered quite well and are comparable to the near-best 7000-term approximation. In Figure 10(a) we see a benchmark 7000-term biorthogonal (9,7) wavelet approximation of the Peppers image, with PSNR = 30.8. In Figure 10(b) we see our Stat-ACS with  $n = 14000$ , using an 8000-term approximation extracted from the 14000 adaptively sampled wavelet coefficients with PSNR = 29.17.

**4.4. Stat-ACS: Rate/distortion comparison to nonadaptive CS algorithms.** In this section we compare the performance of our Stat-ACS algorithm with that of some of the prominent nonadaptive CS measurement schemes such as the Scrambled Fourier Ensemble (SFE) [11] and the Scrambled Block Hadamard Ensemble (SBHE) [40]. The performances of the SFE and the SBHE have been evaluated using the following reconstruction algorithms: min-TV,  $l_1$  optimization solver, and iterative thresholding. Table 4 tabulates the PSNR values for three  $256 \times 256$  natural images, Lena, Cameraman, and Peppers. Also, results from [13] are included, where random Fourier sampling matrices were applied directly in the wavelet domain. The best result for each image and each number of measurements is given in boldface.

Model-based CS methods [2], [44], [48] use wavelet representation modeling in an implicit way. The CS acquisition process is still pseudorandom, but the recovery algorithms try to integrate the assumptions on wavelet structure, with the goal of faster and more accurate recovery in the presence of noise. Next, we compare the performances of our Stat-ACS method and these methods. In Table 5 we see recovery results for the  $128 \times 128$  test image Peppers (used in [2]) from 5000 samples. We run the Stat-ACS algorithm twice, with the Daubechies ortho-basis D8 that was used in [2] and with our preferred basis the (9,7). The results are reported in root mean square error (RSME) to match the reported results in the references.



(a) Benchmark,  $n = 262144$ , 7000-term, PSNR=30.8. (b) Stat-ACS,  $n = 14000$ , 8000-term, PSNR=29.17.

**Figure 10.** Comparison of the Stat-ACS algorithm with optimal benchmark on Peppers.

**Table 4**

Comparison of Stat-ACS with nonadaptive CS algorithms.

$n$	[13]	$l_1$ optimization		min-TV optimization		Iterative thresholding		Stat-ACS
		SFE	SBHE	SFE	SBHE	SFE	SBHE	
Lena ( $256 \times 256$ )								
10000	26.5	21.5	21.1	27.5	28.0	27.2	27.7	<b>30.83</b>
15000	28.7	23.9	23.7	29.7	30.0	30.4	30.6	<b>33.9</b>
Cameraman ( $256 \times 256$ )								
10000	26.2	20.9	20.7	27.0	27.1	26.5	26.8	<b>28.57</b>
15000	28.7	23.2	23.0	29.3	29.5	29.5	29.7	<b>31.22</b>
Peppers ( $256 \times 256$ )								
10000	26.7	21.6	21.4	28.6	29.0	28.1	28.2	<b>30.75</b>
15000	25.3	22.7	22.6	31.2	31.7	31.2	31.6	<b>34.14</b>

**5. Tex-ACS: An ACS algorithm based on a cartoon/texture model.** In this section we elaborate on an ACS algorithm that relies on a model of images as a mix of cartoon and local texture patches. While the edges are somewhat structured, capturing the textural parts is more difficult and elusive, though easily visually perceived. In general, texture is regarded as a function of the spatial variation in pixel intensities [21].

In the wavelet representation, the significant information of texture typically manifests at the higher frequency scales, i.e., finer resolutions with a smaller index  $j$ . The Stat-ACS algorithm of the previous section relies on a cartoon model of the image and is not equipped to predict significant high-frequency wavelet coefficients which correspond to texture information

Table 5

Comparison of Stat-ACS with nonadaptive model-based methods for  $128 \times 128$  Peppers.

Method	CoSaMP [57]	Model-based [2] (D8 wavelet basis)	Stat-ACS, (D8 wavelet basis)	Stat-ACS, ((9,7) wavelet basis)
RMSE for 5000 samples	22.8	11.1	9.4	7.38

from the coefficients at lower resolutions. Thus, our aim is to identify the local texture areas and to sample wavelet coefficients of corresponding frequencies that are supported in these domains.

We address this problem by sampling from a dictionary composed of wavelets and local Fourier-window components over tiles of the image. The role of the local Fourier sampling is to identify the texture information. Once we identify textural areas, we use the correlation between Fourier frequency components and wavelet coefficients with coinciding frequencies to subsequently sample correctly the wavelet coefficients whose support overlaps these areas. In the experimental results, the suggested approach yields significant improvement in extracting the texture parts.

**5.1. Texture and wavelets in the frequency domain.** We very roughly sketch the properties of wavelets from a Fourier side perspective (for more detailed analysis, see, e.g., [31], [52]). In particular we focus on the (9,7) biorthogonal wavelet used in our experiments. Let  $\varphi, \psi$  and  $\tilde{\varphi}, \tilde{\psi}$  be two dual pairs of scaling functions and wavelets that generate biorthogonal wavelet bases of  $L^2(\mathbb{R})$ . Recall that the bivariate wavelet bases we use are created by a tensor product, yielding three types of dual wavelets:

$$(5.1) \quad \tilde{\psi}^1(x) = \tilde{\varphi}(x_1)\tilde{\psi}(x_2), \quad \tilde{\psi}^2(x) = \tilde{\psi}(x_1)\tilde{\varphi}(x_2), \quad \tilde{\psi}^3(x) = \tilde{\psi}(x_1)\tilde{\psi}(x_2).$$

Denote

$$(5.2) \quad \tilde{\psi}_{j,k}^e(x) = 2^{-j}\tilde{\psi}^e(2^{-j}x - k), \quad e = 1, 2, 3, \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z}^2.$$

In the Fourier domain, the separable wavelets from (5.1) become

$$(5.3) \quad \hat{\tilde{\psi}}^1(\omega_1, \omega_2) = \hat{\tilde{\varphi}}(\omega_1)\hat{\tilde{\psi}}(\omega_2), \quad \hat{\tilde{\psi}}^2(\omega_1, \omega_2) = \hat{\tilde{\psi}}(\omega_1)\hat{\tilde{\varphi}}(\omega_2), \quad \hat{\tilde{\psi}}^3(\omega_1, \omega_2) = \hat{\tilde{\psi}}(\omega_1)\hat{\tilde{\psi}}(\omega_2).$$

Roughly speaking,  $|\hat{\tilde{\varphi}}|$  is concentrated around the origin and  $|\hat{\tilde{\psi}}|$  away from the origin. Therefore, by (5.3),  $|\hat{\tilde{\psi}}^1(\omega_1, \omega_2)|$  is large at low horizontal frequencies  $\omega_1$  and high vertical frequencies  $\omega_2$ ,  $|\hat{\tilde{\psi}}^2(\omega_1, \omega_2)|$  is large at high frequencies  $\omega_1$  and low vertical frequencies  $\omega_2$ , and  $|\hat{\tilde{\psi}}^3(\omega_1, \omega_2)|$  is large at high horizontal and vertical frequencies. According to the dominant frequencies computed for the wavelet in scale  $j = 0$ , one can obtain the corresponding dominant frequencies at scale  $j$  by the relation

$$(5.4) \quad \left| \hat{\tilde{\psi}}_{j,k}^e(\omega) \right| = 2^j \left| \hat{\tilde{\psi}}^e(2^j\omega) \right|.$$

Recall the Plancherel formula

$$(5.5) \quad \langle f, \tilde{\psi}_{j,k}^e \rangle = \frac{1}{2\pi} \langle \hat{f}, \hat{\tilde{\psi}}_{j,k}^e \rangle.$$

We wish to examine the amount of oscillatory activity exhibited in the image  $f$  over an arbitrary local area defined by a cube  $Q$ :

$$(5.6) \quad f_Q(x_1, x_2) = \begin{cases} f(x_1, x_2) & \text{if } (x, y) \in Q, \\ 0 & \text{else.} \end{cases}$$

Intensive oscillatory activities may indicate that the support of  $Q$  contains texture and other important features at fine scales. If the support of the wavelet coefficient  $\tilde{\psi}_{j,k}^e$  satisfies  $\text{support}(\tilde{\psi}_{j,k}^e) \subseteq Q$ , then

$$(5.7) \quad \langle f, \tilde{\psi}_{j,k}^e \rangle = \langle f_Q, \tilde{\psi}_{j,k}^e \rangle = \frac{1}{2\pi} \langle \hat{f}_Q, \hat{\psi}_{j,k}^e \rangle.$$

Although the (9,7) wavelet is not compactly supported in the frequency domain, it is “essentially” compactly supported in the frequency domain, and by (5.4) its dilates have “peaks” at the corresponding frequencies. Thus, if, locally,  $f$  contains corresponding frequencies, then we should expect some of the coefficients of wavelets that have essential support around that frequency to have large absolute value.

In practice, we estimate the “energy” of the significant local frequencies by sampling *discrete cosine transform (DCT)* coefficients. The DCT decomposes a local block of an image over  $[x_1, x_1 + \tilde{N}] \times [x_2, x_2 + \tilde{N}]$  in the frequency domain

$$(5.8) \quad \hat{f}(k_1, k_2) = \sum_{n_1=0}^{\tilde{N}-1} \sum_{n_2=0}^{\tilde{N}-1} f(x_1 + n_1, x_2 + n_2) \cos \left[ \frac{\pi}{\tilde{N}} \left( n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[ \frac{\pi}{\tilde{N}} \left( n_2 + \frac{1}{2} \right) k_2 \right].$$

Since the DCT is a linear, real-valued transform, we can compute each DCT coefficient as a combination of two positive functionals which correspond to two DMD measurements, as in (2.6).

**5.2. The Tex-ACS algorithm.** This algorithm is a second step that is applied after the Stat-ACS. For simplicity, we assume that the input image to be processed is of dyadic dimensions,  $N_1 \times N_1 = N$ . We subdivide the image into blocks of  $16 \times 16$  pixels each,  $\{\Omega_{i,l}\}$ ,  $1 \leq i, l \leq N_1/16$ . Over each patch  $\Omega_{i,l}$ , we compute using (5.8), with  $\tilde{N} = 16$ , the three DCT coefficients  $\hat{f}_{\Omega_{i,l}}(1, 8)$ ,  $\hat{f}_{\Omega_{i,l}}(8, 1)$ ,  $\hat{f}_{\Omega_{i,l}}(8, 8)$ , which correspond to the dominant frequencies of the horizontal, vertical, and diagonal wavelet coefficients, respectively, at scale  $j = 2$ . This implies that the second step is initialized by taking about  $0.012 \times N_1^2$  local frequency samples. If any of the three DCT coefficients  $\hat{f}_{\Omega_{i,l}}(k_1, k_2)$  satisfies  $|\hat{f}_{\Omega_{i,l}}(k_1, k_2)| \geq t_j$  for a large threshold  $t_j$ , then we sample the coefficients of wavelets at the corresponding scale and subband whose support significantly overlaps the block  $\Omega_{i,l}$ . Also, if the frequency component  $\hat{f}_{\Omega_{i,l}}(k_1, k_2)$  is significant, it is also likely that significant texture information exists at adjacent scales. Thus in such a case we sample the frequency components  $\hat{f}_{\Omega_{i,l}}(\lceil k_1/2 \rceil, \lceil k_2/2 \rceil)$  and  $\hat{f}_{\Omega_{i,l}}(2k_1, 2k_2)$  to test oscillatory activities in the adjacent frequency domains. If  $|\hat{f}_{\Omega_{i,l}}(\lceil k_1/2 \rceil, \lceil k_2/2 \rceil)| \geq t_j$ , or  $|\hat{f}_{\Omega_{i,l}}(2k_1, 2k_2)| \geq t_j$ , for a scale dependent threshold  $t_j$ , then we subsequently also sample the corresponding wavelet coefficients of scale, compact support, and subband.

The Tex-ACS algorithm is implemented through two separate stages. The first stage, based on the Stat-ACS algorithm, acquires wavelet coefficients which correspond to the cartoon areas. The second stage identifies texture areas using DCT sampling and subsequently acquires fine-scale wavelet coefficients at locations that with high probability contain textures. The Tex-ACS algorithm can be summarized as follows.

**Tex-ACS algorithm. Adaptive sampling for a cartoon-texture image based model.**

**Sampling of the cartoon parts.** Acquire significant wavelet coefficients according to the Stat-ACS algorithm from section 4.2.

**Sampling of the texture parts.** Process each block  $\Omega_{i,l}$ ,  $1 \leq i \leq N_1/16, 1 \leq l \leq N_1/16$ , as follows:

1. Compute the DCT frequency components,  $\hat{f}_{\Omega_{i,l}}(k_1, k_2)$ ,  $(k_1, k_2) = \{(1, 8), (8, 1), (8, 8)\}$ , to detect oscillatory activities at the three subbands of scale  $j = 2$ .
2. If  $|\hat{f}_{\Omega_{i,l}}(k_1, k_2)| \geq t_j$ , then
  - (a) Sample the wavelet coefficients at the corresponding compact support and subband.
  - (b) Compute the frequency components  $\hat{f}_{\Omega_{i,l}}(2k_1, 2k_2)$  and  $\hat{f}_{\Omega_{i,l}}(\lceil k_1/2 \rceil, \lceil k_2/2 \rceil)$  in order to detect oscillatory activities at scales  $j = 1$  and  $j = 3$ , respectively. If  $|\hat{f}_{\Omega_{i,l}}(2k_1, 2k_2)| \geq t_j$  or  $|\hat{f}_{\Omega_{i,l}}(\lceil k_1/2 \rceil, \lceil k_2/2 \rceil)| \geq t_j$ , then sample the wavelet coefficients in the corresponding scale, compact support, and subband.

**5.3. Tex-ACS: Experimental results.** As in the case of the Stat-ACS algorithm, we use the near-best  $n$ -term wavelet approximation as a benchmark. In addition, we compare the efficiency of the Tex-ACS algorithm with our Stat-ACS algorithm from section 4. We test the Tex-ACS algorithm on the Barbara test image, which is an excellent representative of an image that exhibits a mix of cartoon and texture patches. It is evident (illustrated in Figure 12) that the Stat-ACS algorithm based on the cartoon model does not yield good results with regard to the recovery of the texture parts in the image. Moreover, increasing the number of measurements does not yield significant improvement visually or in terms of PSNR. The outcome of the Tex-ACS algorithm (illustrated in Figure 13) shows a significant improvement in terms of PSNR and succeeds in reconstructing texture patches in the image.

In Figure 11(a) we see the Barbara input image. Figure 11(b) depicts a benchmark near-best 12419-term biorthogonal (9,7) wavelet approximation, extracted from the full wavelet representation with PSNR = 27.76. In Figure 12 we show the results of the Stat-ACS algorithm with 12500 terms and PSNR = 24.55, obtained after pruning the insignificant coefficients from 26000 samples. In Figure 13 we see the results using our Tex-ACS algorithm with 12500 terms and PSNR = 25.11, obtained after pruning from 26000 measurements. Note that when we state 26000 samples, this total is composed of both local DCT and wavelet coefficients. In Figures 14 and 15 we zoom in on some textural areas of the result in Figure 13.

We note that the Tex-ACS algorithm works well on Barbara since the texture parts are composed of stripes and checkers that are relatively well modeled by local DCT coefficients. Indeed, it is well known that local Fourier methods usually capture texture in this manner. Possibly, in other cases, we will need to create a different sampling dictionary composed of systems in which we believe the signal will be sparse.



(a) Barbara test image.

(b) Benchmark,  $N = n = 262144$ , 12419-term, PSNR=27.76.

**Figure 11.** Benchmark wavelet approximation of Barbara using 12419 terms.



**Figure 12.**  $n = 26000$ , Stat-ACS (cartoon model based), approximation using 12500 most significant wavelet coefficients, PSNR = 24.55.



**Figure 13.**  $n = 26000$ , *TeX-ACS*, approximation using 12500 most significant wavelet coefficients, PSNR = 25.11.



(a) Zoom in on the chair.

(b) Zoom in on the scarf and legs.

**Figure 14.** Results from Figure 13.



(a) Zoom in on the scarf.

(b) Zoom in on the tablecloth.

**Figure 15.** Results from Figure 13.

**6. Conclusions.** In this work we proposed a direct and fast method for adaptive transform coefficient acquisition, which relies on statistical modeling of images. While our algorithm uses the same order of measurements as nonadaptive CS methods, the main advantages of our direct approach are that no complex recovery algorithm is needed and that it allows more control over the compressed image quality. Our experimental results show, by way of software simulations, that our adaptive algorithms perform well in terms of image quality and speed. One of the possible hardware architectures on which such a scheme could be implemented is the DMD, although more work is required to fully analyze the influence of the DMD accuracy on the proposed method.

**Acknowledgment.** We sincerely thank all the referees for their patience with us and their numerous comments that have significantly improved the quality of the paper.

## REFERENCES

- [1] R. BARANIUK, *Compressive sensing*, IEEE Signal Processing Mag., 24 (2007), pp. 118–120.
- [2] R. BARANIUK, V. CEVHER, M. DUARTE, AND C. HEGDE, *Model-based compressive sensing*, IEEE Trans. Inform. Theory, 56 (2010), pp. 1982–2001.
- [3] R. BARANIUK, M. DAVENPORT, R. DEVORE, AND M. WAKIN, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx., 28 (2008), pp. 253–263.
- [4] R. BERINDE AND P. INDIK, *Sparse Recovery Using Sparse Random Matrices*, Technical report, Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, 2008.
- [5] T. BLUMENSATH AND M. DAVIES, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 265–274.
- [6] J. BOBIN, J.-L. STARCK, AND R. OTTENSAMER, *Compressed sensing in astronomy*, IEEE J. Sel. Topics Signal Process., 2 (2008), pp. 718–726.

- [7] L. BOUBCHIR AND M. FADILI, *Multivariate statistical modeling of images with the curvelet transform*, in Proceedings of the 8th IEEE International Conference on Signal Processing and Its Applications, 2005, pp. 747–750.
- [8] R. W. BUCCIGROSSI AND E. P. SIMONCELLI, *Image compression via joint statistical characterization in the wavelet domain*, IEEE Trans. Image Process., 8 (1999), pp. 1688–1701.
- [9] E. CANDÈS, *Compressive sampling*, in Proceedings of the International Congress of Mathematics, 2006, pp. 1433–1452.
- [10] E. CANDÈS AND D. DONOHO, *New tight frames of curvelets and optimal representations of objects with piecewise singularities*, Comm. Pure Appl. Math., 57 (2004), pp. 219–266.
- [11] E. CANDÈS AND J. ROMBERG, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–985.
- [12] E. CANDÈS AND J. ROMBERG, *Robust signal recovery from incomplete observations*, in Proceedings of the IEEE International Conference on Image Processing, 2006, pp. 1281–1284.
- [13] E. CANDÈS AND J. ROMBERG, *Practical signal recovery from random projections*, in Wavelet Applications in Signal and Image Processing XI, Proceedings of the SPIE 5914, 2004.
- [14] E. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- [15] E. CANDÈS, J. ROMBERG, AND T. TAO, *Signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2005), pp. 1207–1223.
- [16] E. CANDÈS AND T. TAO, *The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$* , Ann. Statist., 35 (2007), pp. 2313–2351.
- [17] E. CANDÈS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.
- [18] E. CANDÈS AND M. WAKIN, *An introduction to compressive sampling*, IEEE Signal Processing Mag., 25 (2008), pp. 21–30.
- [19] W. CAREY, D. CHUANG, AND S. HEMAMI, *Regularity-preserving interpolation*, IEEE Trans. Image Process., 8 (1999), pp. 1293–1297.
- [20] S. G. CHANG, Z. CVETKOVIĆ, AND M. VETTERLI, *Resolution enhancement of images using wavelet transform extrema extrapolation*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1995, pp. 2379–2382.
- [21] C. CHEN, L. PAU, AND P. WANG, *Handbook of Pattern Recognition and Computer Vision*, World Scientific, Singapore, 1993.
- [22] S. S. CHEN, D. L. DONOHO, AND M. A. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.
- [23] Z. CHEN AND M. A. KARIM, *Forest representation of wavelet transform and feature detection*, Opt. Eng., 39 (2000), pp. 1194–1202.
- [24] H. CHOI AND R. BARANIUK, *Multiscale texture segmentation using wavelet-domain hidden Markov models*, in Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers, Vol. 2, Pacific Grove, CA, 1998, pp. 1692–1697.
- [25] C. CHRYSAFIS AND A. ORTEGA, *Efficient context-based lossy wavelet image coding*, in Proceedings of the Conference on Data Compression, Snowbird, UT, IEEE Computer Society, Washington, DC, 1997.
- [26] A. COHEN, W. DAHMEN, AND R. DEVORE, *Compressed sensing and best  $k$ -term*, J. Amer. Math. Soc., 22 (2009), pp. 211–231.
- [27] R. COIFMAN, F. GESHWIND, AND Y. MEYER, *Noiselets*, Appl. Comput. Harmon. Anal., 10 (2001), pp. 27–44.
- [28] T. CORMEN, C. LEISERSON, R. RIVEST, AND C. STEIN, *Introduction to Algorithms*, 3rd ed., MIT Press, Cambridge, MA, 2009.
- [29] M. CROUSE, R. NOWAK, AND R. BARANIUK, *Wavelet-based statistical signal processing using hidden Markov models*, IEEE Trans. Signal Process., 46 (1998), pp. 886–902.
- [30] W. DAI AND O. MILENKOVIC, *Subspace pursuit for compressive sensing: Closing the gap between performance and complexity*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2230–2249.
- [31] I. DAUBECHIES, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
- [32] S. DEKEL, *Adaptive Compressed Image Sensing Based on Wavelet-Trees*, GE Healthcare technical report, 2008.

- [33] S. DEUTSCH, A. AVERBUCH, AND S. DEKEL, *Adaptive compressed image sensing based on wavelet modeling and direct sampling*, in Proceedings of the 8th International Conference on Sampling Theory and Applications, 2009.
- [34] R. DEVORE, *Nonlinear approximation*, in Acta Numerica, Acta Numer. 7, Cambridge University Press, Cambridge, UK, 1998, pp. 51–150.
- [35] D. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [36] D. DONOHO, *For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math., 59 (2006), pp. 797–829.
- [37] D. DONOHO, I. DRORI, Y. TSAIG, AND J. L. STARCK, *Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit*, preprint, Department of Statistics, Stanford University, Stanford, CA, 2006.
- [38] M. DUARTE, M. DAVENPORT, D. TAKHAR, J. LASKA, T. SUN, K. KELLY, AND R. BARANIUK, *Single-pixel imaging via compressive sampling*, IEEE Signal Processing Mag., 25 (2008), pp. 83–91.
- [39] M. DUARTE, M. WAKIN, AND R. BARANIUK, *Fast reconstruction of piecewise smooth signals from random projections*, in Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations, Rennes, France, 2005.
- [40] L. GAN, T. DO, AND T. TRAN, *Fast compressive imaging using scrambled Hadamard transform ensemble*, in Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008), 2008.
- [41] O. GONZALEZ AND J. MADDOCKS, *Global curvature, thickness, and the ideal shapes of knots*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 4769–4773.
- [42] O. GONZALEZ, J. MADDOCKS, AND J. SMUTNY, *Curves, circles, and spheres*, in Physical Knots: Knotting, Linking, and Folding Geometric Objects in  $\mathbb{R}^3$ , AMS, Providence, RI, 2002, pp. 195–215.
- [43] H. W. GUGGENHEIMER, *Differential Geometry*, Dover, New York, 1977.
- [44] L. HE AND L. CARIN, *Exploiting structure in wavelet-based Bayesian compressive sensing*, IEEE Trans. Signal Process., 57 (2009), pp. 3488–3497.
- [45] K. HUANG AND S. AVIYENTE, *Mutual information based subband selection for wavelet packet based image classification*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, Philadelphia, PA, 2005, pp. 241–244.
- [46] R. L. JOSHI, H. JAFARKHANI, J. H. KASNER, T. R. FISCHER, N. FARVARDIN, M. W. MARCELLIN, AND R. H. BAMBERGER, *Comparison of different methods of classification in subband coding of images*, IEEE Trans. Image Process., 6 (1997), pp. 1473–1486.
- [47] R. KAWAKAMI, J. WRIGHT, Y. TAI, Y. MATSUSHITA, M. BEN-EZRA, AND K. IKEUCHI, *High-resolution hyperspectral imaging via matrix factorization*, in Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 2329–2336.
- [48] C. LA AND M. DO, *Signal reconstruction using sparse tree representation*, in Wavelets XI, Proceedings of SPIE, Vol. 5914, San Diego, CA, 2005.
- [49] J. LIU AND P. MOULIN, *Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients*, IEEE Trans. Image Process., 10 (2001), pp. 1647–1658.
- [50] M. LUSTIG, D. DONOHO, AND J. PAULY, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magn. Reson. Med., 58 (2007), pp. 1182–1195.
- [51] D. MALIOUTOV, S. SANGHAVI, AND A. WILLSKY, *Sequential compressed sensing*, IEEE J. Sel. Topics Signal Process., 4 (2010), pp. 435–444.
- [52] S. MALLAT, *A Wavelet Tour of Signal Processing. The Sparse Way*, 3rd ed., Academic Press, Burlington, MA, 2009.
- [53] S. MALLAT, *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Trans. Pattern Anal. Mach. Intell., 11 (1989), pp. 674–693.
- [54] S. MALLAT AND W. L. HWANG, *Singularity detection and processing with wavelets*, IEEE Trans. Inform. Theory, 38 (1992), pp. 617–642.
- [55] R. MODDEMEIJER, *On estimation of entropy and mutual information of continuous distributions*, Signal Process., 16 (1989), pp. 233–248.
- [56] B. K. NATARAJAN, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
- [57] D. NEEDELL AND J. TROPP, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–321.

- [58] L. PANINSKI, *Estimation of entropy and mutual information*, Neural Comput., 15 (2003), pp. 1191–1253.
- [59] A. PENTLAND AND B. HOROWITZ, *A practical approach to fractal-based image compression*, in Digital Images and Human Vision, A. B. Watson, ed., MIT Press, Cambridge, MA, 1993, pp. 53–59.
- [60] A. PENTLAND, E. SIMONCELLI, AND T. STEPHENSON, *Fractal-Based Image Compression and Interpolation*, U.S. Patent 5, 148, 497, Sept. 15, 1992.
- [61] Y. PFEFFER AND M. ZIBULEVSKY, *A Micro-Mirror Array Based System for Compressive Sensing of Hyperspectral Data*, Technical report CS-2010-01, Department of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel, 2010.
- [62] D. D.-Y. PO AND M. N. DO, *Directional multiscale modeling of images using the contourlet transform*, IEEE Trans. Image Process., 15 (2006), pp. 1610–1620.
- [63] R. RINALDO AND G. CALVAGNO, *Image coding by block prediction of multiresolution subimages*, IEEE Trans. Image Process., 4 (1995), pp. 909–920.
- [64] J. ROMBERG, H. CHOI, AND R. BARANIUK, *Bayesian tree-structured image modeling using wavelet-domain hidden Markov models*, IEEE Trans. Image Process., 10 (2001), pp. 1056–1068.
- [65] F. ROOMS, A. PIZURICA, AND W. PHILIPS, *Estimating image blur in the wavelet domain*, in Proceedings of the IEEE Benelux Signal Processing Symposium (SPS), 2002.
- [66] A. SAID AND W. PEARLMAN, *A new fast and efficient image codec based on set partitioning in hierarchical trees*, IEEE Trans. Circuits Syst. Video Technol., 6 (1996), pp. 243–250.
- [67] E. SCHWARTZ, A. ZANDI, AND M. BOLIEK, *Implementation of compression with reversible embedded wavelets*, in Proceedings of SPIE 2564, 1995.
- [68] C. E. SHANNON, *A mathematical theory of communication*, Bell System Tech. J., 27 (1948), pp. 379–423, 623–656.
- [69] J. SHAPIRO, *Embedded image coding using zerotrees of wavelet coefficients*, IEEE Trans. Signal Process., 41 (1993), pp. 3445–3462.
- [70] E. SIMONCELLI, *Modeling the joint statistics of images in the wavelet domain*, in Proceedings of SPIE 3813, 1999, pp. 188–195.
- [71] K. TIMMERMANN AND R. NOWAK, *Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging*, IEEE Trans. Inform. Theory, 45 (1999), pp. 846–862.
- [72] J. TROPP AND A. GILBERT, *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
- [73] Y. TSAIG AND D. DONOHO, *Extensions of compressed sensing*, Signal Process., 86 (2006), pp. 549–571.
- [74] X. WU AND J. CHEN, *Context modeling and entropy coding of wavelet coefficients for image compression*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 1997, pp. 3097–3100.