Compressive Coded Apertures for High-Resolution Imaging

Roummel F. Marcia[a], Zachary T. Harmany[b], and Rebecca M. Willett[b]

[a]School of Natural Sciences, University of California, Merced, Merced, CA USA
[b]Department of Electrical and Computer Engineering, Duke University, Durham, NC USA

ABSTRACT

Traditionally, optical sensors have been designed to collect the most directly interpretable and intuitive measurements possible. However, recent advances in the fields of image reconstruction, inverse problems, and compressed sensing indicate that substantial performance gains may be possible in many contexts via computational methods. In particular, by designing optical sensors to deliberately collect "incoherent" measurements of a scene, we can use sophisticated computational methods to infer more information about critical scene structure and content. In this paper, we explore the potential of physically realizable systems for acquiring such measurements. Specifically, we describe how given a fixed size focal plane array, compressive measurements using coded apertures combined with sophisticated optimization algorithms can significantly increase image quality and resolution.

Keywords: Compressive sensing, coded aperture, sparse recovery, model-based sparsity, wavelets

1. INTRODUCTION

Recent advances in the field of compressive sensing (CS)[1–5] indicate that when feasible, judicious selection of the type of image transformation introduced by measurement systems may dramatically improve our ability to extract high-quality images from a limited number of measurements. In optical systems, carefully chosen indirect measurements combined with computational methods allow relatively small focal plane arrays (FPAs) to be used to generate high-resolution imagery. By designing optical sensors to collect measurements of a scene according to CS theory, we can use sophisticated computational methods to infer critical scene structure and content.

While recent progress in the exploitation of CS theory is highly encouraging, there are several key issues in the context of optical systems that must be addressed:

- *Physical constraints.* Photon intensities and CS projections must all be nonnegative in linear optical systems.
- *Practical systems.* Projecting a scene onto a collection of unstructured random vectors, the case most commonly considered in the literature, would typically require very large physical systems or very long total exposure times.
- *Algorithm properties.* Many CS reconstruction algorithms are made fast by exploiting key properties of the sensing matrix. Once the above physical constraints are incorporated, however, algorithms must be adjusted accordingly.
- *Reconstruction time.* Fast image reconstruction algorithms are imperative, particularly in high-throughput or video systems.
- *Noise and quantization.* While CS measurements can lead to impressive results when measurements have arbitrary accuracy, measurement noise and quantization errors play a significant role in determining reconstruction accuracy.

In this paper, we explore practical methods for addressing the above challenges in the context of coded aperture imaging.

Further author information: (Send correspondence to R. F. Marcia)
R. F. Marcia.: E-mail: rmarcia@ucmerced.edu, Telephone: 1-209-228-4874
Z. T. Harmany: E-mail: zth@duke.edu Telephone: 1-919-619-9123
R. M. Willett: E-mail: willett@duke.edu, Telephone: 1-919-660-5544

2. COMPRESSED SENSING

Let $f^\star \in \mathbb{R}_+^n$ be a vector representation of the n-pixel image of interest, and let $\theta^\star \in \mathbb{R}^n$ be the set of basis expansion coefficients for $f^\star$ in some orthonormal basis $W \in \mathbb{R}^{n\times n}$:

$$f^\star = \sum_{i=1}^{n} \theta_i^\star w_i,$$

where w_i is the i^{th} basis vector and $\theta_i^\star$ is the corresponding coefficient. In many settings, the basis $W \triangleq [w_1, \ldots, w_n]$ can be chosen so that only $k \ll n$ coefficients have significant magnitude, i.e., many of the θ_i's are zero or very small for large classes of images; we then say that $\theta^\star$ is *sparse* or *compressible*. The key insight of CS is that, with slightly more than k well-chosen measurements, we can determine which θ_i's are significant and accurately estimate their values. Furthermore, fast algorithms which exploit the *sparsity* of θ make this recovery computationally feasible. Sparsity has long been recognized as a highly useful metric in a variety of inverse problems, but much of the underlying theoretical support was lacking. However, more recent theoretical studies have provided strong justification for the use of sparsity constraints and quantified the accuracy of sparse solutions to these underdetermined systems.[1,6]

The problem of estimating the image $f^\star$ can be formulated mathematically as an inverse problem, where the data collected by an imaging or measurement system are represented as

$$y = Af^\star + \epsilon, \tag{1}$$

where $A \in \mathbb{R}_+^{m\times n}$ linearly projects the scene onto a m-dimensional set of observations, $\epsilon \in \mathbb{R}^m$ is noise or quantization errors associated with the sensor, and $y \in \mathbb{R}_+^m$ is the observed data. Compressive sensing addresses the problem of solving for $f^\star$ when $n \gg m$, i.e., A is severely underdetermined. In general, this is an ill-posed problem as there are an infinite number of candidate solutions for $f^\star$; nevertheless, CS theory provides a set of conditions that, if satisfied, assure an accurate estimation of $f^\star$. We first presuppose that $f^\star$ is sparse or compressible in a basis W. Then given W, we require that A in conjunction with W satisfies a technical condition called the *Restricted Isometry Property* (RIP).[7] More specifically, we say that AW satisfies the RIP of order s if there exists a constant $\delta_s \in (0, 1)$ for which

$$(1 - \delta_s)\|z\|_2^2 \le \|AWz\|_2^2 \le (1 + \delta_s)\|z\|_2^2. \tag{2}$$

holds for all s-sparse $z \in \mathbb{R}^n$. In other words, the energy contained in the projected image, AWz, is close to the energy contained in the original image, z. While the RIP cannot be verified for an arbitrary given observation matrix and basis, it has been shown that observation matrices A drawn independently from many probability distributions satisfy the RIP of order s with high probability for any orthogonal basis W when $m \ge Cs\log(n/s)$ for some constant C.[7] Although generating CS matrices using this procedure is simple in software, building physical systems to measure $AW\theta$ can be notoriously difficult.

While the system of equations in (1) can be grossly underdetermined, CS theory suggests that selecting the *sparsest* solution to this system of equations will yield a highly accurate solution, subject to $\theta^\star = W^T f^\star$ being sufficiently sparse and A satisfying the RIP of order $2k$, where k is the sparsity of $f^\star$.[7] In particular, the ℓ_2-ℓ_1 minimization

$$\widehat{f} = \arg\min_f \ \frac{1}{2}\|y - Af\|_2^2 + \tau\|W^T f\|_1$$

will yield a highly accurate estimate of $f^\star$ with very high probability.[6,8] Alternatively, we can compute the basis expansion coefficients directly by solving

$$\begin{aligned} \widehat{\theta} &= \arg\min_\theta \ \frac{1}{2}\|y - AW\theta\|_2^2 + \tau\|\theta\|_1 \\ \widehat{f} &= W\widehat{\theta} \end{aligned} \tag{3}$$

The regularization parameter $\tau > 0$ helps to overcome the ill-posedness of the problem, and the ℓ_1-penalty term drives small components of θ to zero and helps promote sparse solutions.

3. CONVENTIONAL CODED APERTURES

The basic idea of coded aperture imaging is to use a mask in the aperture of an imaging system composed of a pattern of openings to induce a more complicated point spread function than that associated with a simple pinhole camera. This approach allows a larger fraction of the available photons to hit the detector array; the mask pattern can then be exploited during the image reconstruction process.[9–11] These techniques are particularly popular in astronomical[12,13] and medical[14–16] applications because of their efficacy at wavelengths where lenses cannot be used. Recent work has also demonstrated their utility for collecting both high resolution images and object depth information simultaneously.[17]

Seminal work in coded aperture imaging includes the development of Modified Uniformly Redundant Arrays (MURAs).[18] These mask patterns (which we denote by h^{MURA}) are binary, square patterns, whose *grid size matches the spatial resolution of the photo-detector* and whose sidelength is a prime integer number of grid cells. Each mask pattern is specifically designed to have a complementary pattern h^{recon} such that $h^{\mathrm{MURA}} * h^{\mathrm{recon}}$ is a Kronecker δ function. Here, the operator $*$ denotes 2D convolution, and in a slight abuse of notation, can be applied to vectorized representations of images, such as $f^\star$.

MURA observations can be modeled as

$$y = (\mathcal{D}(f^\star)) * h^{\mathrm{MURA}} + \epsilon, \tag{4}$$

where ϵ corresponds to noise associated with the physics of the sensor, and $\mathcal{D}(f^\star)$ is the downsampling of the scene to the resolution of the detector array. By construction of h^{MURA} and h^{recon}, $\mathcal{D}(f^\star)$ can be estimated using

$$\widehat{f} = y * h^{\mathrm{recon}} = (\mathcal{D}(f^\star)) * h^{\mathrm{MURA}} * h^{\mathrm{recon}} + \epsilon * h^{\mathrm{recon}} = \mathcal{D}(f^\star) + \epsilon * h^{\mathrm{recon}}.$$

However, the resulting resolution is often lower than what is necessary to capture some of the desired details in the image. Clearly, the estimates from MURA reconstruction are limited by the spatial resolution of the photo-detector.

4. COMPRESSIVE CODED APERTURES

Recent studies by the authors[19,20] and others[21] address the accurate reconstruction of a high resolution static image which has a sparse representation in some basis from a single low resolution observation using *compressive* coded aperture (CCA) imaging. In our study, we designed a coded aperture imaging mask such that the corresponding observation matrix A^{CCA} satisfies the RIP of order $2k$ as described in (2) with high probability when $m \geq C_1 k^3 \log(n/k)$ for some constant $C_1 > 0$.[19,22] Stronger results are shown recently,[23] where if $m \geq C_2(k \log(n) + \log^3(k))$ for some constant $C_2 > 0$, then $f^\star$ can be accurately recovered with high probability.

The measurement matrix A^{CCA} associated with compressive coded apertures can be modeled as

$$A^{\mathrm{CCA}} f^\star = \mathcal{D}(f^\star * h^{\mathrm{CCA}}), \tag{5}$$

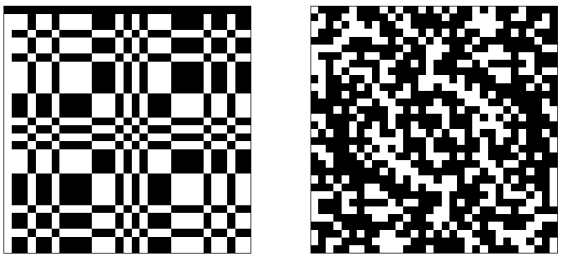
where h^{CCA} is the coding mask and is the size and resolution at which $f^\star$ will be reconstructed, rather than the size and resolution of the FPA as we had with h^{MURA}. The convolution of h^{CCA} with an image $f^\star$ as in (5) can be represented as the application of the Fourier transform to $f^\star$ and h^{CCA}, followed by element-wise matrix multiplication and application of the inverse Fourier transform. In matrix notation, this series of linear operations can be expressed as

$$h^{\mathrm{CCA}} * f^\star = \mathcal{F}^{-1} C_H \mathcal{F} f^\star,$$

where $\mathcal{F}$ is the 2D Fourier transform matrix and C_H is the diagonal matrix whose diagonal elements correspond to the transfer function H^{CCA}, which is the Fourier transform of the h^{CCA}, i.e., $H^{\mathrm{CCA}} \triangleq \mathcal{F}(h^{\mathrm{CCA}})$. The matrix product $\mathcal{F}^{-1} C_H \mathcal{F}$ is block-circulant and each block is in turn circulant. Block-circulant matrices with circulant blocks, whose entries are drawn from an appropriate probability distribution, are known to be CS matrices.[19,22]

Based upon recent theoretical work on Toeplitz- and circulant-structured matrices for CS, compressive mask patterns are fast and memory-efficient to generate[19,22,23] as follows. First, let

$$R \triangleq \mathcal{F}^{-1} C_H \mathcal{F}, \tag{6}$$



(a) (b)

Figure 1. Coded aperture patterns. Here, the white blocks represent the openings in the mask pattern. (a) The 31×31 Modified Uniform Redundant Array (MURA) mask pattern h^{MURA}. The length of the side of the MURA pattern must be a prime number. (b) An example of a 32×32 compressive coded aperture (CCA) mask pattern h^{CCA}. The pattern is not quite random – note the circular symmetry of the CCA pattern about the $(17, 17)$ element.

and let D be the downsampling matrix corresponding to the downsampling operator $\mathcal{D}$. Thus, $A^{\mathrm{CCA}} = DR$. To determine the mask h^{CCA} so that the resulting A^{CCA} satisfies the RIP, we generate the block circulant with circulant blocks matrix R by drawing from an appropriate probability distribution (such as a zero-mean Gaussian or scaled Radamacher distribution). Then, by the way R is defined in (6), we can compute $C_H = \mathcal{F}R\mathcal{F}^{-1}$. The elements of the transfer function H^{CCA} are the elements of the diagonal matrix C_H, from which the mask h^{CCA} can be easily obtained since $h^{\mathrm{CCA}} = \mathcal{F}^{-1}(H^{\mathrm{CCA}})$. The computational bottleneck in this sequence of calculations lies in the formation of C_H from the (large) $n \times n$ matrix R. However, the diagonalizability of block-circulant matrices with circulant blocks by the two-dimensional discrete Fourier transform leads to fast and memory-efficient computations. Specifically, let $\mathscr{F}$ be the one-dimensional Fourier transform matrix so that $\mathcal{F} = \mathscr{F} \otimes \mathscr{F}$, where $\otimes$ is the matrix Kronecker product. If R_j is the j^{th} circulant block of R and if we define the $\sqrt{n} \times \sqrt{n}$ diagonal matrix $M_j \triangleq \mathscr{F}R_j\mathscr{F}^{-1}$ (circulant matrices are diagonalized by the one-dimensional Fourier transform), then the j^{th} $\sqrt{n} \times \sqrt{n}$ diagonal block of C_H is given by

$$(C_H)_j = M_1 + (-1)^{j-1} M_{\sqrt{n}/2+1} + \sum_{t=2}^{\sqrt{n}/2} 2\mathrm{Re}\left(\omega^{-(t-1)(j-1)} M_t\right).$$

(We assume that $\sqrt{n}$ is even.) Since each $(C_H)_j$ and M_t are diagonal, only $\sqrt{n}$ elements need to be stored in each, and so relatively little memory and computational time are needed for this calculation.

The incorporation of the integration downsampling operator $\mathcal{D}$ does not prevent the RIP from being satisfied; a key element of the proof that the RIP is satisfied is a bound on the number of rows of A^{CCA} which are statistically independent. Since the downsampling operator effectively sums rows of a block circulant matrix, downsampling causes the bound on the number of dependent matrix rows to be multiplied by the downsampling factor. Enforcing symmetry on $\mathcal{F}^{-1}C_H\mathcal{F}$ is equivalent to assuring that the transfer function matrix H^{CCA} is symmetric about its center, so that the resulting coding mask pattern $h^{\mathrm{CCA}} \triangleq \mathcal{F}^{-1}(H^{\mathrm{CCA}})$ will be necessarily real.[19] Contrasting mask patterns for MURA coded aperture imaging vs. compressive coded aperture imaging are displayed in Fig. 1.

5. ALGORITHMS FOR NONNEGATIVE CS MINIMIZATION

In optical imaging, we often estimate light intensity, which *a priori* is nonnegative. Thus it is necessary that the reconstruction $\widehat{g} = W\widehat{\theta}$ is nonnegative, which involves adding constraints to the CS optimization problem (3), i.e.,

$$\begin{aligned} \widehat{\theta} \quad = \quad & \arg\min_{\theta} \quad \frac{1}{2}\|y - AW\theta\|_2^2 + \tau\,\mathrm{pen}(\theta), \\ & \text{subject to} \quad W\theta \geq 0, \end{aligned} \tag{7}$$

where $\mathrm{pen}(\theta)$ is a general sparsity-promoting term. The addition of the nonnegativity constraint in (7) makes the problem more challenging than the conventional CS minimization problem, and has been addressed in CS literature recently in the

context of photon-limited compressive sensing.[24,25] Here, we discuss approaches to address the nonnegativity constraints in the CS minimization problem.

The minimization problem (7) can be solved using a sequence of quadratic approximation subproblems that are easier to solve. Specifically, this sequential approach reduces to a series of alternating steps: (a) approximating the objective function with a regularized quadratic objective, and (b) regularized least squares image denoising. (This approach is similar to SpaRSA,[26] which does not address the nonnegativity constraint on the reconstruction.) The resulting minimization subproblem is given by

$$\begin{aligned} \theta^{j+1} &= \underset{\theta}{\arg\min} \quad \phi(\theta^j) + (\theta - \theta^j)^T \nabla\phi(\theta^j) + \frac{\alpha_j}{2}\|\theta - \theta^j\|_2^2 + \tau\,\mathrm{pen}(\theta), \\ &\quad \text{subject to } W\theta \geq 0, \end{aligned} \tag{8}$$

Here, $\phi(\theta) = \frac{1}{2}\|y - AW\theta\|_2^2$ is the quadratic term in (7), and the first three terms in the objective function of (8) correspond to the Taylor expansion of $\phi(\theta)$ at the current iterate θ^j. The second derivative of $\phi(\theta)$ is approximated by a multiple of the identity matrix, namely, $\nabla^2\phi(\theta^j) \approx \alpha_j I$, with $\alpha_j > 0$ computed using the Barzilai-Borwein (spectral) methods.[26–28] This approach is particularly fast and effective for RIP-satisfying CS matrices since the near-isometry condition implies that $A^TA \approx \alpha I$, for some $\alpha > 0$. The constrained subproblem (8) can be written equivalently and more compactly as

$$\begin{aligned} \theta^{j+1} &= \underset{\theta}{\arg\min} \quad \frac{1}{2}\|s^j - \theta\|_2^2 + \frac{\tau}{\alpha_j}\,\mathrm{pen}(\theta) \\ &\quad \text{subject to } W\theta \geq 0, \end{aligned} \tag{9}$$

where

$$s^j \triangleq \theta^j - \frac{1}{\alpha_j}\nabla\phi(\theta^j). \tag{10}$$

The subproblem (9) can be viewed as a nonnegative denoising subproblem applied to s^j (the next gradient descent iterate) to obtain the next iterate θ^{j+1}. This framework allows users to take advantage of the plethora of fast and effective image denoising methods available for various penalty terms.

5.1 ℓ_1 Penalty

When the penalty term $\mathrm{pen}(\theta) = \|\theta\|_1$, (7) is simply the constrained ℓ_2-ℓ_1 CS minimization problem (3), and the corresponding constrained denoising subproblem is given by

$$\begin{aligned} \theta^{j+1} &= \underset{\theta}{\arg\min} \quad \frac{1}{2}\|s^j - \theta\|_2^2 + \frac{\tau}{\alpha_j}\|\theta\|_1 \\ &\quad \text{subject to } W\theta \geq 0, \\ f^{j+1} &= W\theta^{j+1}. \end{aligned} \tag{11}$$

In the canonical basis (i.e., $W = I$), it has an analytic solution that can easily be calculated. In the noncanonical basis, however, (9) does not necessarily have a closed form solution, which, nonetheless, can be estimated quickly using gradient-based methods in the following manner. First, the ℓ_1 norm is non-differentiable, and therefore, a change of variables must be applied to (9) to use derivative information. By letting $\theta = u - v$ where $u, v \geq 0$, we can write (9) as

$$\begin{aligned} (u^{j+1}, v^{j+1}) &= \underset{(u,v)}{\arg\min} \quad g(u,v) \triangleq \frac{1}{2}\|s^j - (u - v)\|_2^2 + \frac{\tau}{\alpha_j}\mathbb{1}^T(u + v) \\ &\quad \text{subject to } u \geq 0,\ v \geq 0,\ W(u - v) \geq 0 \end{aligned} \tag{12}$$

where $\mathbb{1}$ is a vector of ones. The next iterate θ^{j+1} is then defined as $\theta^{j+1} = u^{j+1} - v^{j+1}$. Because the constraints in (12) are nonnegativity bounds not only on the variables u and v but also on $W(u - v)$, solving (12) is not straightforward. However, the *dual* formulation of this minimization problem, i.e., solving for the Lagrange multipliers associated with (12), has simple box constraints and can be easily solved iteratively. Specifically, the Lagrange dual problem is given by

$$\begin{aligned} &\underset{\lambda,\gamma\in\mathbb{R}^m}{\text{minimize}} \ h(\lambda,\gamma) \triangleq \frac{1}{2}\|s^j + \gamma + W^T\lambda\|_2^2 - \frac{1}{2}\|s^j\|_2^2 \\ &\text{subject to } \lambda \geq 0, \quad -\frac{\tau}{\alpha_j}\mathbb{1} \leq \gamma \leq \frac{\tau}{\alpha_j}\mathbb{1} \end{aligned} \tag{13}$$

and at the optimal values $\gamma^\star$ and $\lambda^\star$, the primal iterate $\theta^{j+1} \triangleq u^{j+1} - v^{j+1}$ is given by

$$\theta^{j+1} = s^j + \gamma^\star + W^T\lambda^\star.$$

We note that the minimizers of the primal problem (12) and its dual (13) satisfy $g(u^{j+1}, v^{j+1}) = -h(\gamma^\star, \lambda^\star)$ since the minimization problem (12) satisfies (a weakened) Slater's condition.[29]

The objective function in (13) can be minimized by alternatingly solving for λ and γ, which is accomplished by taking the partial derivatives of $h(\lambda, \gamma)$ and setting them to zero. Each component is then constrained to satisfy the bounds in (13). At the i^{th} iteration, the variables can, thus, be defined as follows:

$$\begin{aligned} \gamma_i &= \text{mid}\left\{-\frac{\tau}{\alpha_j}\mathbb{1}, -s^j - W^T\lambda_{i-1}, \frac{\tau}{\alpha_j}\mathbb{1}\right\} \\ \lambda_i &= \left[-W\left(s^j + \gamma_i\right)\right]_+, \end{aligned}$$

where the operator $\text{mid}(a, b, c)$ chooses the middle value of the three arguments component-wise. Note that at the end of each iteration i, the approximate solution $\theta_i^{j+1} \triangleq s^j + \gamma_i + W^T\lambda_i$ to (11) is feasible with respect to the constraint $W\theta \geq 0$:

$$W\theta_i^{j+1} = Ws^j + W\gamma_i + \lambda_i = W(s^j + \gamma_i) + \left[-W(s^j + \gamma_i)\right]_+ = \left[W(s^j + \gamma_i)\right]_+ \geq 0.$$

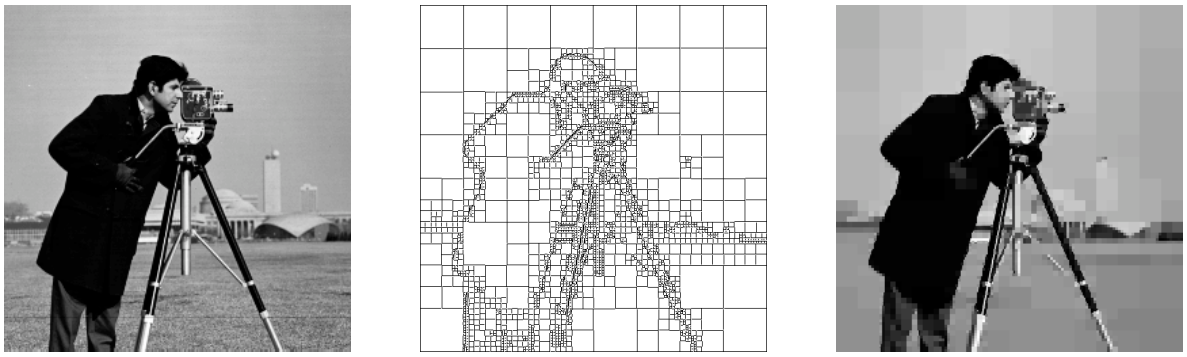
In general, these subproblems are solved approximately (i.e., with a limited number of iterations) as relatively good estimates of each primal iterate θ^j tend to be produced with only few inner iterations.

5.2 Model-Based Sparsity

While the majority of the CS literature has focused on the case where the scene of interest admits a sparse representation in some basis or dictionary, more recent developments have used more sophisticated models of scenes which incorporate key *structure* into the sparsity models. The basic idea has been used previously in the context of image denoising and compression. For example, it is well-known that image denoising can be accomplished via wavelet coefficient-wise thresholding. However, more refined thresholding methods exploit the fact that significant wavelet coefficients tend to cluster near one another within scales and arise at similar locations between scales; this approach can yield significant improvements in accuracy.[30]

CS reconstruction methods have recently been developed based upon similar principles to improve reconstruction results.[4,31,32] For instance, as we noted above, (9) amounts to an image denoising operation conducted during each loop of the reconstruction algorithm. Thus a variety of denoising methods can be used. Unfortunately, some of these models result in nonconvex optimization problems for which finding the globally optimal reconstruction is not computationally tractable – though locally optimal reconstructions are often very good. Nevertheless, recent theoretical work[31] shows that incorporating sparsity models can reduce the number of measurements needed to achieve a desired accuracy level. Another alternative to conventional sparsity assumptions is that scenes of interest lie on a low-dimensional manifold embedded in the n-dimensional pixel space. Similar assumptions have been shown to be useful in analyzing databases of text documents,[33] visualization of high-dimensional data sets,[34] object recognition,[35] and classification and semi-supervised learning.[36,37] In general, however, devising computationally efficient general-purpose methods to perform image reconstruction based on manifold structure is a challenging open problem.[38] In the case where the image of interest is known to be smooth or piecewise smooth in the canonical basis (i.e., it is compressible in a wavelet basis), we can formulate a penalty function which is a useful alternative to the ℓ_1 norm of the wavelet coefficient vector. In particular, we can build on the framework of *recursive dyadic partitions (RDP)*,[39,40] which we summarize here.

It can be shown that partition-based denoising methods such as this are closely related to Haar wavelet denoising with an important hereditary constraint placed on the thresholded coefficients—if a parent coefficient is thresholded, then its children coefficients must also be thresholded.[40] This constraint is akin to wavelet-tree ideas which exploit persistence of significant wavelet coefficients across scales and have recently been shown highly useful in compressive sensing settings.[31] In particular, the partition-based methods calculate image estimates by determining the ideal partition of the domain of observations and by using maximum likelihood estimation to fit a model (e.g., a constant) to each cell in the optimal partition. The space of possible partitions is a nested hierarchy defined through a recursive dyadic partition (RDP) of the image domain, and the optimal partition is selected by pruning a quad-tree representation of the observed data to best fit



(a) (b) (c)

Figure 2. Recursive dyadic partition of the Cameraman image. (a) Original image $f^\star$. (b) RDP P, with larger partition cells corresponding to regions of more homogeneous intensity. (c) Piecewise constant approximation $f(P)$ to original image, with constant pieces corresponding to RDP cells.

the observations with minimal complexity. Each of the terminal squares in the pruned RDP could correspond to a region of homogeneous or smoothly varying intensity. This gives our estimators the capability of spatially varying the resolution to automatically increase the smoothing in very regular regions of the image and to preserve detailed structure in less regular regions.

An RDP can be obtained by merging neighboring squares of (i.e., pruning) a full quad tree representation of the data to form a data-adaptive RDP, P, and fitting models to the data on the terminal squares of P. Thus the image estimate, $f(P)$, is completely described by P. An image $f^\star$, an RDP P, and an image approximation, $f(P)$, are displayed in Fig. 2. The model coefficients for each partition cell are chosen via nonnegative least squares (addressed below). RDP-based image estimates are computed using a very simple framework for penalized least squares estimation, wherein the penalization is based on the complexity of the underlying partition (i.e., the number of cells in the partition). The goal here is to find the partition which minimizes the penalized squared error function:

$$P^{j+1} = \underset{P\in\mathcal{P}}{\arg\min} \quad \|s^j - f(P)\|_2^2 + \frac{\tau}{\alpha_j}\,\mathrm{pen}(P) \qquad (14)$$

$$\text{subject to } f(P) \geq 0$$

$$f^{j+1} = f(P^{j+1}) \qquad (15)$$

where $\mathcal{P}$ is a collection of all possible RDPs of the image domain and s^j is defined similarly as in (10):

$$s^j = f^j - \frac{1}{\alpha_j}\nabla\phi(f^j) \quad \text{where} \quad \phi(f) = \frac{1}{2}\|y - Af\|_2^2$$

The penalty term $\mathrm{pen}(P)$ in (14) is the penalty associated with the estimate $f(P)$ and is proportional to the number of cells in the partition P.[39–41]

A search over $\mathcal{P}$ can be computed quickly using a dynamic program. When using constant model fits, the nonnegative least-squares fits can be computed non-iteratively in each interval by simply using the maximum of the average of s^j in that interval and zero. Because of this, enforcing nonnegativity constraints is trivial and can be accomplished very quickly.

5.3 Total Variations

Regularization based on a *total variation (TV)* norm has also garnered significant recent attention.[42] In general, this norm measures how much an image varies across pixels, so that a highly textured or noisy image will have a large TV norm, whereas a smooth or piecewise constant image would have a relatively small TV norm. This is often a useful alternative to wavelet-based regularizers, which are also designed to be small for piecewise smooth images but can result in spurious large, isolated wavelet coefficients and related image artifacts.

The problem of estimating the true intensity $f^\star$ from y using a TV-regularized objective function with nonnegativity constraints has recently been addressed by Beck and Teboulle,[43] where an estimate $\widehat{f}$ of $f^\star$ is obtained by solving the following constrained minimization problem:

$$\begin{aligned} \widehat{f} \;=\; & \underset{f}{\arg\min} \quad \frac{1}{2}\|y - Af\|_2^2 + \tau\|f\|_{\mathrm{TV}} \\ & \text{subject to} \quad f \geq 0, \end{aligned} \tag{16}$$

where

$$\|f\|_{\mathrm{TV}} \;\triangleq\; \sum_{k=1}^{\sqrt{n}-1}\sum_{l=1}^{\sqrt{n}} |f_{k,l} - f_{k+1,l}| \;+\; \sum_{k=1}^{\sqrt{n}}\sum_{l=1}^{\sqrt{n}-1} |f_{k,l} - f_{k,l+1}|. \tag{17}$$

In (17), we have used 2D pixel indices instead of vector indices in a small abuse of notation. We have also assumed that $f \in \mathbb{R}^n$ is a vector corresponding to a square $\sqrt{n} \times \sqrt{n}$ image for simplicity of presentation, but this assumption is not necessary for the algorithm. The gradient-based optimization approach proposed by Beck and Teboulle is based on a monotone fast iterative shrinking and thresholding (IST) algorithm and has been shown to outperform other IST methods that do not take into account of the nonnegativity constraints.

5.4 Recentering and Mean Subtraction

Generative models for random projection matrices used in CS involve drawing elements independently from a zero-mean probability distribution,[1,6,7,22,44] and likewise a zero-mean distribution was used to generate the coded aperture masks described in Sec.4. However, a coded aperture mask with a zero mean is not physically realizable in optical systems. We generate our physically realizable mask by taking our randomly generated, zero-mean mask pattern and shifting it so that all mask elements are in the range $[0, 1/m]$, where m is the dimension of the observation.[19] This shifting ensures that the coded aperture corresponds to a valid (i.e., nonnegative and intensity preserving) probability transition matrix which describes the distribution of photon propagation through the optical system.

This shifting, while necessary to accurately model real-world optical systems, negatively impacts the performance of the proposed ℓ_2-ℓ_1 reconstruction algorithm for the following reason. If we generate a non-realizable zero-mean mask (h^0) with elements in the range $[-1/2m, 1/2m]$ and simulate measurements of the form

$$y^0 = \mathcal{D}(f^\star * h^0) \equiv A^0 f^\star, \tag{18}$$

then the corresponding observation matrix A^0 will satisfy the RIP with high probability and $f^\star$ can be accurately estimated from y^0 using ℓ_2-ℓ_1 minimization. In contrast, if we shift h^0 to be in the range $[0, 1/m]$ (by adding $1/2m$ to each element in h^0) and denote this h, then we have a practical and realizable coded aperture mask. However, observations of the form

$$y = \mathcal{D}(f^\star * h) \equiv Af^\star$$

cannot be directly used with ℓ_2-ℓ_1 minimization methods that assume that A is zero mean. For several algorithms,[26,28] including those in this paper, it is crucial that the second derivative of the data-fitting term $\phi(f) = \frac{1}{2}\|y - Af\|_2^2$ is well approximated by a scalar multiple of the identity matrix. In other words, $\nabla^2\phi(f) = A^TA \approx \alpha I$ for some $\alpha > 0$. For the nonzero mean A,

$$h = h^0 + \frac{1}{2m}\mathbb{1}_{\sqrt{n}\times\sqrt{n}},$$

where $\mathbb{1}_{\sqrt{n}\times\sqrt{n}}$ is a $\sqrt{n} \times \sqrt{n}$ matrix of ones, and

$$A^TA \;=\; \left(A^0 + \frac{1}{2m}\mathbb{1}_{m\times n}\right)^T\left(A^0 + \frac{1}{2m}\mathbb{1}_{m\times n}\right) \;\approx\; \alpha I + \left(\frac{C_A}{m} + \frac{1}{4m}\right)\mathbb{1}_{n\times n},$$

where C_A is the sum of each column of A and is known by construction. Note that A^TA is far from being diagonal, so methods which attempt to exploit a near-diagonal structure in A^TA would not result in fast and accurate reconstructions.[26,28]

To address this problem, we note that

$$y = Af^\star = \left(A^0 + \frac{1}{2m}\mathbb{1}_{m\times n}\right) f^\star = y^0 + \frac{\|f^\star\|_1}{2m}\mathbb{1}_{m\times 1},$$

where we exploit the known positivity of $f^\star$. Furthermore, since y is also positive, we can estimate the total signal intensity $\|f^\star\|_1$ using the total intensity of the observations:

$$\mathbb{E}\left[\sum_{i=1}^{m} y_i\right] \;=\; \sum_{i=1}^{m}\sum_{j-1}^{n} A_{i,j} f_j^\star \;=\; \sum_{j=1}^{n}\left(\sum_{i=1}^{m} A_{i,j}\right) f_j^\star \;=\; C_A\|f^\star\|_1,$$

Consequently, $\|f^\star\|_1 \approx \sum_{i=1}^{m} y_i/C_A$. Putting this all together, we can estimate

$$y^0 \approx y - \frac{1}{2mC_A}\left(\sum_{i=1}^{m} y_i\right)\mathbb{1}_{m\times 1}$$

and use it to solve for $f^\star$ in (18). It can readily be seen that solving for $f^\star$ in (18) will produce a solution with zero mean, and so we add $\mu \equiv \sum_{i=1}^{m} y_i/(2mC_A)$ to this result to achieve our final, accurate estimate. Similar shifting operations could be used in cases where the coded aperture mask has a fill factor different from the 50% assumed in the above derivation.

Finally, because we must use the zero-mean A^0 to reconstruct the original signal, the problem formulation (7) must be altered accordingly. Specifically, we must solve

$$\begin{aligned} \widehat{\theta} \;=\; & \arg\min_{\theta} \quad \frac{1}{2}\|y^0 - A^0 W\theta\|_2^2 + \tau\,\mathrm{pen}(\theta) \\ & \text{subject to} \quad W\theta + \mu\mathbb{1} \geq 0 \end{aligned} \tag{19}$$

where $\mathrm{pen}(\theta)$ is one of the various penalty terms described previously. The estimate for $f^\star$ is then given by $\widehat{f} = W\widehat{\theta} + \mu\mathbb{1}$. We note that the algorithms described above can be modified to solve (19) in a straightforward manner.

6. NOISE AND QUANTIZATION ERRORS

While CS is particularly useful when the FPA needs to be kept compact, it should be noted that CS is more sensitive to measurement errors and noise than more direct imaging techniques. The experiments conducted in previous work[20] simulated very high signal-to-noise ratio (SNR) settings and showed that CS methods can help resolve high resolution features in images. However, in low SNR settings CS reconstructions can exhibit significant artifacts that may even cause more distortion than the low-resolution artifacts associated with conventional coded aperture techniques such as MURA. Similar observations are made by Haupt and Nowak.[45] These observations are particularly relevant when considering the bit-depth of focal plane arrays, which corresponds to measurement quantization errors. Related theoretical work[24] shows that in the presence of low SNR photon noise, theoretical error bounds can be large, and thus the expected performance of CS may be limited unless the number of available photons to sense is sufficiently high.

7. CONCLUSION

One of the main tenets of compressed sensing is that relatively few well-chosen observations are needed to form a sparse image using sophisticated image reconstruction algorithms. This suggests that it may be possible to build cameras with much smaller focal plane arrays than are conventionally required for high-resolution imaging. However, the application of these ideas in practical settings poses several challenges. First, directly implementing CS theory by collecting a series of independent pseudo-random projections of a scene requires either (a) a very large physical system or (b) observations collected sequentially over time. In this paper, we describe an alternative snapshot architecture (which capture all observations simultaneously) with a compact form factor, namely coded aperture techniques. These approaches impose structure upon

the pseudo-random projections, most notably by limiting their independence. A second key challenge relates to the nonnegativity of image intensities and measurements which can be collected by linear optical systems. Much of the theoretical literature on CS allows for negative measurements and does not consider nonnegativity during the reconstruction process. In this paper, we show that (a) CS theory can be used in designing coded aperture mask patterns such that high-resolution images can be reconstructed from low-resolution observations, (b) nonnegativity constraints can be incorporated in CS minimization problem reconstruction algorithms without heavy computational costs, and (c) pre-processing observations to account for nonnegative sensing matrices can be applied so that central assumptions underlying some fast CS algorithms are satisfied.

ACKNOWLEDGMENTS

This work was supported by NSF CAREER Award No. CCF-06-43947, DARPA Grant No. HR0011-07-1-003, and NSF Grant DMS-08-11062.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory* **52**(2), pp. 489 – 509, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* **52**(4), pp. 1289–1306, 2006.

[3] R. Willett and M. Raginsky, "Performance bounds on compressed sensing with Poisson noise," in *Proc. IEEE Intl. Symp. Info. Th.*, 2009.

[4] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing." Accepted for publication in *IEEE Trans. Sig. Proc.*

[5] S. Jafarpour, R. Willett, M. Raginsky, and R. Calderbank, "Performance bounds for expander-based compressed sensing with Poisson noise," in *Proc. Asilomar*, 2009.

[6] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory* **52**(9), pp. 4036–4048, 2006.

[7] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory* **15**(12), pp. 4203–4215, 2005.

[8] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.* **35**, pp. 2313–2351, 2007.

[9] J. G. Ables, "Fourier transform photography: a new method for X-ray astronomy," *Proc. Astron. Soc. Aust.* **1**, p. 172, 1968.

[10] R. H. Dicke, "Scatter-hole cameras for X-rays and gamma-rays," *Astrophys. J.* **153**, pp. L101–L106, 1968.

[11] L. Mertz and N. O. Young, "Fresnel transformations of images," in *Proc. Conf. Opt. Instrum. Tech.*, (London), 1961.

[12] E. Caroli, J. B. Stephen, G. D. Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in X- and gamma-ray astronomy," *Space Science Reviews* **45**, pp. 349–403, 1987.

[13] G. Skinner, "Imaging with coded-aperture masks," *Nucl. Instrum. Methods Phys. Res.* **221**, pp. 33–40, Mar. 1984.

[14] R. Accorsi, F. Gasparini, and R. C. Lanza, "A coded aperture for high-resolution nuclear medicine planarimaging with a conventional anger camera: experimental results," *IEEE Trans. Nucl. Sci.* **28**, pp. 2411–2417, 2001.

[15] G. R. Gindi, J. Arendt, H. H. Barrett, M. Y. Chiu, A. Ervin, C. L. Giles, M. A. Kujoory, E. L. Miller, and R. G. Simpson, "Imaging with rotating slit apertures and rotating collimators," *Med. Phys.* **9**, pp. 324–339, 1982.

[16] S. R. Meikle, R. R. Fulton, S. Eberl, M. Bahlbom, K. Wong, and M. J. Fulham, "An investigation of coded aperture imaging for small animal SPECT," *IEEE Trans. Nucl. Sci.* **28**, pp. 816–821, 2001.

[17] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. of Intl. Conf. Comp. Graphics. and Inter. Tech.*, 2007.

[18] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. Opt.* **28**, 1989.

[19] R. F. Marcia and R. M. Willett, "Compressive coded aperture superresolution image reconstruction," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Proc.*, 2008.

[20] R. F. Marcia, Z. T. Harmany, and R. M. Willett, "Compressive coded aperture imaging," in *Proc. SPIE Electron. Imag.*, (San Jose, CA), January 2009.

[21] A. Stern and B. Javidi, "Random projections imaging with extended space-bandwidth product," *IEEE/OSA J. Disp. Technol.* **3**(3), pp. 315–320, 2007.

[22] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE Stat. Sig. Proc. Workshop*, 2007.

[23] J. Romberg, "Compressive sampling by random convolution." To appear in *SIAM J. Sci. Comput.*, 2009.

[24] M. Raginsky, Z. Harmany, R. F. Marcia, and R. Willett, "Compressed sensing performance bounds under Poisson noise." submitted, 2009.

[25] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "Sparse Poisson intensity reconstruction algorithms," in *Proc. IEEE Stat. Sig. Proc. Workshop*, 2009.

[26] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.* , 2009 (to appear).

[27] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. of Numer. Anal.* **8**(1), pp. 141–148, 1988.

[28] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Top. Sign. Proces.: Special Issue on Convex Optimization Methods for Signal Processing* **1**(4), pp. 586–597, 2007.

[29] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

[30] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal-processing using hidden markov-models," *IEEE Trans. Sig. Proc.* **46**, pp. 886–902, April 1998.

[31] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Inform. Theory* , 2008. Submitted.

[32] S. Negahban and M. Wainwright, "Phase transitions for high-dimensional joint support recovery," in *NIPS*, 2008.

[33] M. Belkin, P. Niyogi, and V. Singhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research* **7**, pp. 2399 – 2434, 2006.

[34] A. N. Gorban, B. Kégl, D. C. Wunsch, and A. Zinovyev, *Principal Manifolds for Data Visualization and Dimension Reduction*, Springer, 2007.

[35] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. SPIE Elecron. Imag.*, 2007.

[36] A. Goldberg, X. Zhu, A. Singh, Z. Xu, and R. Nowak, "Multi-manifold semi-supervised learning," in *AISTATS*, 2009.

[37] Q. Liu, X. Liao, H. Li, J. R. Stack, and L. Carin, "Semisupervised multitask learning," *IEEE Tran. Patt. Anal. Mach. Int.* **31**(6), pp. 1074–1086, 2009.

[38] M. B. Wakin, "Manifold-based signal recovery and parameter estimation from compressive measurements." submitted, 2009.

[39] R. Nowak, U. Mitra, and R. Willett, "Estimating inhomogeneous fields using wireless sensor networks," *IEEE J. Sel. Areas Commun.* **22**(6), pp. 999–1006, 2004.

[40] R. Willett and R. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Trans. Inform. Theory* **53**(9), pp. 3171–3187, 2007.

[41] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. Med. Imaging* **22**(3), pp. 332–350, 2003.

[42] T. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*, Society for Industrial and Applied Mathematics, 2005.

[43] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.* **18**(11), pp. 2419–34, 2009.

[44] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices." To appear in *Constructive Approximation*, 2007.

[45] J. Haupt and R. Nowak, "Compressive sampling vs. conventional imaging," in *Proc. IEEE Intl. Conf. on Imag. Proc.*, pp. 1269–1272, 2006.