COMPRESSIVE CODED APERTURE SUPERRESOLUTION IMAGE RECONSTRUCTION

Roummel F. Marcia and Rebecca M. Willett

Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708

ABSTRACT

Recent work in the emerging field of compressive sensing indicates that, when feasible, judicious selection of the type of distortion induced by measurement systems may dramatically improve our ability to perform reconstruction. The basic idea of this theory is that when the signal of interest is very sparse (*i.e.,* zero-valued at most locations) or compressible, relatively few incoherent observations are necessary to reconstruct the most significant non-zero signal components. However, applying this theory to practical imaging systems is challenging in the face of several measurement system constraints. This paper describes the design of coded aperture masks for superresolution image reconstruction from a single, low-resolution, noisy observation image. Based upon recent theoretical work on Toeplitz-structured matrices for compressive sensing, the proposed masks are fast and memory-efficient to compute. Simulations demonstrate the effectiveness of these masks in several different settings.

***Index Terms*—** Image reconstruction, Image resolution, Compressive sensing, Coded aperture

1. SUPERRESOLUTION IMAGE RECONSTRUCTION

Superresolution image reconstruction conventionally is the process by which several low resolution, noisy, slightly shifted observations are used to reconstruct an image of the underlying high resolution scene [1]. Mathematically, we can model the $j^{\rm th}$ k-dimensional observation of a high resolution, n^2-dimensional scene f as

$$y_j = R_j f + w_j,$$

where R_j is a $k \times n^2$ matrix representing shifting an image, followed by a blur (such as one induced by imaging optics), and downsampling, and where w_j is zero-mean white Gaussian noise associated with the $j^{\rm th}$ observation. Much of the existing superresolution literature (cf. [1]) assumes we have observed y_j for $j = 1, 2, \ldots J$, for some $J > n/k$, so that the total number of observations is roughly the same as or greater than the number of pixels in f to be reconstructed [1, 2]. The multiple observations can be collected sequentially over time or via a multiplexed imaging system (cf. [2]). In a variety of practical settings, however, collecting several low resolution observations is not feasible because of time limitations, a need to keep the focal plane array small, or data storage restrictions. In light of these constraints, one might ask whether it is possible to collect a single low resolution observation and use it alone to reconstruct f; *i.e.,* we wish to reconstruct f from a single measurement of the form

$$y = Rf + w. \quad (1)$$

Since the dimensionality of y is now significantly less than the dimensionality of f, the resulting inverse problem is highly underdetermined and ill-posed. In fact, even with the benefit of state-of-the-art reconstruction techniques, estimating high resolution features in f such as sharp boundaries and edges is difficult when R is modeled as described above (*i.e.,* blurring and downsampling). These challenges are also present in image upsampling and interpolation settings [3].

This paper addresses the accurate reconstruction of a high resolution image f from a single low resolution observation y via compressive coded aperture imaging. In particular, we describe the design of coded aperture patterns for incoherent imaging systems which significantly improve the accuracy with which we can reconstruct f, and theoretically characterize the performance of the system within the framework of compressive sensing [4]. Section 2 describes coded aperture imaging and the mask design principles which have been used in previous, non-compressive contexts where the dimensionality of the observation (k) is approximately equal to the number pixels to be reconstructed (n^2). In Section 3 we describe recent work in the emerging field of Compressive Sensing (CS), which is based on the idea that we can use a relatively small number of indirect observations of an image and reconstruct it very accurately when that image is sparse in some basis. However, the translation of these ideas to realizable physical systems is nontrivial. Section 4 describes the physical constraints placed upon the CS projection matrix in incoherent optical systems and the subsequent proposed design of coded aperture masks for compressive superresolution image reconstruction. Finally, in Section 5 we present simulation results which demonstrate the efficacy of the proposed mask design compared with conventional imaging systems without coded apertures.

2. CODED APERTURE IMAGING

Coded aperture imaging first arose out of a desire to increase the light hitting a detector in an optical system without sacrificing resolution (by, say, increasing the diameter of a pinhole). The basic idea is to create a mask pattern which introduces a more complicated point spread function than that associated with a pinhole, and exploit this pattern to reconstruct high-quality image estimates. These techniques are particularly popular in astronomical and medical applications because of their efficacy at wavelengths where lenses cannot be used, but recent work has also demonstrated their utility for collecting both high resolution images and object depth information simultaneously [5].

Seminal work in coded aperture imaging includes the development of Modified Uniformly Redundant Arrays (MURAs) [6]. These mask patterns, denoted by p, are binary, square patterns with prime integer sidelengths which are designed so that if one observed

$$y = f * p + w,$$

The authors were partially supported by DARPA Contract No. HR0011-04-C-0111, ONR Grant No. N00014-06-1-0610, and DARPA Contract No. HR0011-06-C-0109.

where $*$ denotes convolution, then f could be reconstructed as

$$\widehat{f} = y * \widetilde{p}$$

for some complementary pattern $\widetilde{p}$. In other words, the MURA patterns (and their complements) are specifically designed so that $p * \widetilde{p}$ approximately equals the Kronecker δ function and hence to optimize reconstruction accuracy *subject to the constraint that linear, convolution-based reconstruction methods would be used*.

While MURA coded apertures are successful in the context of linear reconstruction, there exist a wide variety of nonlinear reconstruction methods which can dramatically outperform linear reconstructions when f has a sparse representation in some basis, such as a wavelet basis. However, there currently exist few guiding principles for designing coded aperture masks for nonlinear reconstruction methods.

3. COMPRESSIVE SENSING

Nonlinear image reconstruction based upon sparse representations of images has received widespread attention recently with the advent of "compressive sensing". This emerging theory indicates that, subject to a *Restricted Isometry Property* (RIP) [7] condition on the observation matrix R, very high dimensional vectors (f) can be recovered with astounding accuracy from a much smaller dimensional observation (y) when f has a "sparse" representation in some basis, meaning that only a few non-zero basis coefficients contain the vast majority of the energy in f. Let W denote a basis such that $f = W\theta$ where θ has few non-zero coefficients. Then the $\ell^2 - \ell^1$ minimization

$$\begin{aligned} \hat{\theta} &= \arg\min_{\theta} \frac{1}{2}\|y - RW\theta\|_2^2 + \tau\|\theta\|_1 \qquad (2) \\ \widehat{f} &= W\hat{\theta} \end{aligned}$$

will yield a highly accurate estimate of f with very high probability [8, 9]. The regularization parameter $\tau > 0$ helps to overcome the ill-conditioning of the matrix R for obtaining θ from observations $y = RW\theta + w$. The ℓ^1 penalty term drives small components of θ to zero and helps create sparse solutions.

The observation matrix R is said to satisfy the RIP of order $3m$ if, for $T \subset \{1, 2, \cdots, n\}$ and R_T, a submatrix obtained by retaining the columns of R corresponding to the indices in T, there exists a constant $\delta_{3m} \in (0, 1/3)$ such that for all $z \in \Re^{|T|}$,

$$(1 - \delta_{3m})\|z\|_2^2 \leq \|R_T z\|_2^2 \leq (1 + \delta_{3m})\|z\|_2^2 \qquad (3)$$

holds for all subsets T with $|T| \leq 3m$ [7]. While the RIP cannot be verified for a given observation matrix R, it has been shown that matrices with entries drawn independently from some probability distributions satisfy the condition with high probability when $k \geq Cm\log(n/m)$ for some constant C, where $m \equiv \|\theta\|_{\ell_0}$ is the number of non-zero elements in the vector θ [7].

In a number of practical settings, however, it is not possible to have arbitrary control over the observation matrix R and draw its elements from a given probability distribution. Instead, R is often constrained to have a particular structure associated with the physics of the data collection system. Below, we examine the generation of coded aperture masks which allow the associated observation matrix to satisfy the RIP of order $3m$.

4. MASK DESIGN

4.1. Observation Matrix Structure

In the following, we assume that the observation y is given by $y = D(f*h)+w$, where D is a downsampling operator and h is a point-spread function (PSF). The downsampling operator D corresponds to selecting every other pixel in both the horizontal and vertical directions of an image.

Denote the Fourier transform of h by H. Let F be an $n \times n$ matrix whose entries are $F_{k,l} = \omega_n^{(k-1)(l-1)}$, where ω_n is the n-th root of unity, given by $e^{2\pi i/n}$. Let $\mathcal{F} = F \otimes F$, where $\otimes$ is the matrix Kronecker product. Then $\mathcal{F}h = H$. Let C_H be an $n^2 \times n^2$ diagonal matrix whose diagonal components are the entries in H. Then the observation y is given by $y = Rf$, where R is the linear operator

$$R = D\mathcal{F}^{-1}C_H\mathcal{F}.$$

In a conventional coded aperture imaging setup, we assume the PSF h is given by a mask p, *i.e.,* $h = p$. In a pinhole camera, h corresponds to a δ function and the observation y is merely the blurred and downsampled image. In a Fourier optics setting, the PSF h can be written as

$$h = |\mathcal{F}(p)|^2, \qquad (4)$$

where $\mathcal{F}(p)$ is the Fourier transform of p and $|\cdot|^2$ is understood to be component-wise. If the image is taken with the aperture fully open, *i.e.,* p is a matrix of ones, then h again corresponds to a δ function.

Our goal then is to design a mask p in either setting such that the resulting image reconstruction is better than using no mask at all, *i.e.,* the pinhole camera or the fully open aperture in Fourier imaging. This involves defining a p such that the corresponding observation matrix R satisfies an RIP.

4.2. Pseudo-Circulant CS Matrices

Recently, Bajwa et al. [10] showed that random circulant matrices (and Toeplitz matrices, in general) are sufficient to recover f from y exactly with high probability. In particular, they showed that circulant matrices whose elements are drawn independently from an appropriate distribution satisfy an RIP of order $3m$ when $k \geq Cm^3\log(n/m)$ for some constant C. In this paper, we extend these results to pseudo-circulant matrices and use them to motivate our mask design.

Note that $R = D\mathcal{F}^{-1}C_H\mathcal{F}$ is $k \times n^2$, where $k = n^2/d^2$ and d is the downsampling factor in each dimension. Furthermore, R has a specific structure. Let A be such that $R = DA$, and note that A is $n^2 \times n^2$ block-circulant of the form

$$A = \begin{pmatrix} A_n & A_{n-1} & \cdots & A_2 & A_1 \\ A_1 & A_n & \cdots & A_3 & A_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ A_{n-1} & A_{n-2} & \cdots & \cdots & A_n \end{pmatrix}, \qquad (5)$$

where each $A_j \in \Re^{n \times n}$ is circulant; *i.e.,* , A_j is of the form

$$A_j = \begin{pmatrix} a_n & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_n & \cdots & a_3 & a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & \cdots & a_n \end{pmatrix}.$$

The n^2/d^2 rows of R form a subset of the n^2 rows of A corresponding to the downsampling operation. This "pseudo-circulant" structure is a direct result of the fact that F diagonalizes any circulant matrix (such as A_j) and so $\mathcal{F} \equiv F \otimes F$ diagonalizes block-circulant matrices (such as A).

In order to show that R satisfies the RIP, it is necessary that each submatrix R_T of R, where $T \subset \{1, 2, \cdots, n^2\}$ satisfy (3), where R_T is the $k \times |T|$ matrix whose columns are the columns of R with indices in T. In [10], it is shown that a $k \times n^2$ matrix A satisfies the RIP with very high probability if A is circulant and if the elements of the first row of A are drawn from an appropriate probability distribution (such as a zero-mean Gaussian or scaled Radamacher distribution). Central to the proof of this result is the fact that the number of rows of A_T (a $k \times |T|$ submatrix of A) which are statistically dependent upon the i^{th} row of A_T, for any $i \in 1, 2, \ldots, k$, is bounded above by $2|T|(|T| - 1)$. Our observation matrix R is pseudo-circulant rather than circulant (as described above), but it is nevertheless straightforward to demonstrate that the number of rows of R_T which are statistically dependent upon the i^{th} row of R_T is also bounded by $2|T|(|T| - 1)$. Thus the pseudo-circulant R_T has the same degree of dependency structure as a circulant matrix A_T of the same size. It follows directly that because A satisfies the RIP with high probability, R must also satisfy the RIP with high probability.

Finally, since we assume that θ is sparse or compressible in a basis W, it is the product matrix RW that we ultimately require to satisfy the RIP. (Here, W can be, for example, the Haar wavelet transform matrix.) Following the methods outlined in [10], a matrix A_W can be defined such that $A_W W$ is block-circulant. Thus, the product $RW = DA_W W = DA$ will satisfy the RIP with high probability.

4.3. Masks for Pseudo-Circulant CS

The previous subsection showed that we can draw n^2 iid observations from an appropriate distribution (such as a zero-mean Gaussian with variance $1/k$) and use them to form a pseudo-circulant observation matrix R. In this section, we describe how to derive a coded aperture mask from such a matrix in a computationally- and memory-efficient manner.

First, note that R is uniquely defined by its first row entries, e.g., $\{a_j\}_{j=1}^{n^2}$, and that A (where $R = DA$) is similarly uniquely defined by the a_j's. The matrix A can be written as $A = \mathcal{F}^{-1} C_H \mathcal{F}$, and recall that C_H is a diagonal matrix. This means that given A, we can compute $C_H = \mathcal{F} A \mathcal{F}^{-1}$. As discussed in Sec. 4.1, it is straightforward to compute h from C_H. By enforcing symmetry on A (so that the $(i, j)^{\text{th}}$ element of A is the same as the $(j, i)^{\text{th}}$ element of A), we can guarantee that the transfer function H defined by C_H is circularly symmetric and hence h is real-valued. Computing h using this logic is impractical because of limited computational time and system memory. However, it is possible to compute h from the a_j's directly. Thus, even though A is $n^2 \times n^2$, the elements of C can be obtained without having to form A or compute $\mathcal{F} A \mathcal{F}^{-1}$ explicitly.

With some algebraic manipulation, we can show that the elements of C_H can be computed using linear combinations of the diagonal matrices $G_j \equiv F A_j F^{-1}$. Let C_j be the $n \times n$ submatrix of C_H whose first element is the $(n(j-1)+1, n(j-1)+1)$ element of C_H for $j = 1, \ldots, n$. In general, these matrices are real, diagonal and given by

$$C_j = G_1 + (-1)^{j-1} G_{n/2+1} + \sum_{t=2}^{n/2} 2\text{Re}\left(\omega^{-(t-1)(j-1)} G_t\right).$$

Since each G_j and C_j is diagonal, only n elements need to be stored in each, and so relatively little memory and computational time are needed for this calculation.

In a Fourier optics setting, the point-spread function h is given by $h = |\mathcal{F}(p)|^2$, where $|\cdot|^2$ is component-wise. We shift h such that each of its elements is non-negative, and then the mask p can be explicitly calculated. Note that this means that the elements of R are no longer zero mean, but we can easily compensate for this during the reconstruction process defined by (2).

5. SIMULATION RESULTS

5.1. Minimization Algorithm

The compressive imaging problem can be formulated as an optimization problem where the objective function is expressed as a combination ℓ^1 and ℓ^2 minimization program, as described in (2). In our numerical experiments, we use the GPSR (gradient projection for sparse reconstruction) code of Figueiredo et al. [11], which has been shown to outperform several existing optimization algorithms and codes for solving (2), or equivalent formulations of it. The version used in our simulations - GPSR-BB - uses a quasi-Newton search direction, where the Hessian is approximated by a multiple of the identity matrix. Generally, calculating search directions is the computational bottleneck of optimization algorithms. Because the GPSR-BB search direction amounts to the negative gradient, it is generally not expensive to compute. An important feature of GPSR is a post-processing step called *debiasing*. Having obtained a solution (and consequently a sparsity pattern for the solution) to (2), the debiasing phase minimizes the ℓ^2 term while keeping the sparsity pattern of the iterates θ fixed. This step often produces significant improvement over the original sparse solutions. In our numerical experiments, we used the Haar wavelet transform for the basis W.

5.2. Numerical Experiments

To demonstrate the effectiveness of the above compressive coded apertures, we simulate reconstruction of the "Cameraman" image (a magnified section is displayed in Figure 1(a)), from observations collected via several different masks. As a baseline for comparison, we reconstruct the image from observations simulating downsampling without using a coded aperture, *i.e.,* the PSF h is a δ function. This reconstruction in showed in Figure 1(b) and has an MSE of $\|f - \widehat{f}\|_2^2 / \|f\|_2^2 = 0.1011$. Figure 1(c) shows the noisy, coded observation associated with a random mask generated as described above, and Figure 1(d) shows the image reconstruction via GPSR from this observation; its MSE is $\|f - \widehat{f}\|_2^2 / \|f\|_2^2 = 0.0867$.

In some settings, additional simplicity may be desired in the coding mask. For instance, conventional coded aperture masks consist solely of zeros and ones. If we round the mask elements used to generate the observation in Figure 1(c) to satisfy this constraint, generate a new observation, and compute a new reconstruction, we achieve the result shown in Figure 1(e); its MSE is $\|f - \widehat{f}\|_2^2 / \|f\|_2^2 = 0.0897$. Likewise, in Fourier optics one can generate mask elements with values of plus or minus one (for example, by carefully controlling the thickness of an optical element at different locations to induce a phase shift); setting the sign of each element of the mask derived above to achieve this effect (scaling appropriately) and computing the resulting GPSR reconstruction yields the image displayed in Figure 1(f), which has an MSE of $\|f - \widehat{f}\|_2^2/\|f\|_2^2 = 0.0924$.

Qualitatively, we see that the codes developed based on the principles of compressive sensing yield higher resolution reconstruction and less spatial aliasing in several key image features, such as the columns in one of the background buildings, the features of the man's face, and the crispness of strong edges such as around the legs of the tripod. Furthermore, we find that modifying the masks slightly to aid in ease of manufacturing does not significantly detract from the quality of the reconstructions.

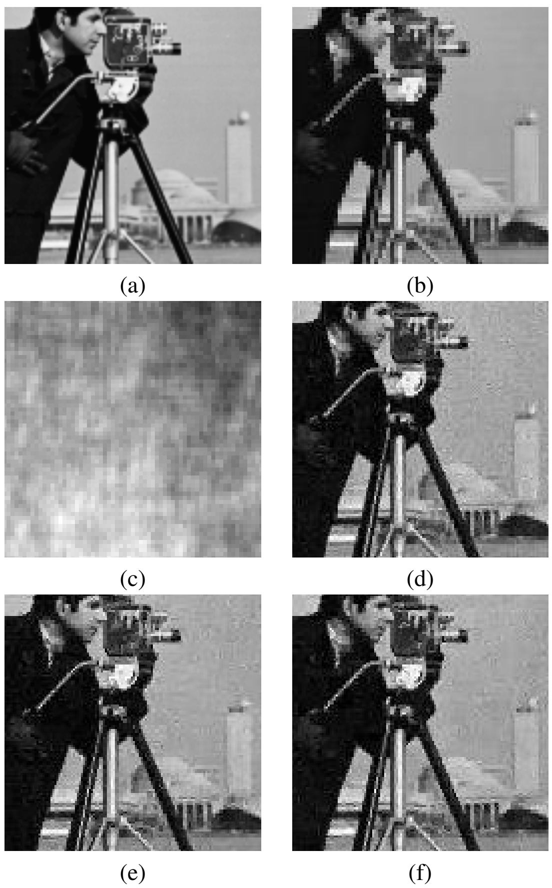

Fig. 1. Compressive coded aperture superresolution results. (a) Original cameraman image (f). (b) Reconstruction without coding (*i.e.,* h is the δ function); MSE = 0.1011. (c) Observed coded aperture image (y), with one quarter as many pixels as f, using a code derived with the proposed method. (d) Reconstruction from observations in (c); MSE = 0.0867. (e) Reconstruction after rounding elements of h to be either 0 or 1 before collecting observations; MSE = 0.0897. (f) Reconstruction in Fourier imaging setting after setting elements of p to be either -1 or 1 before collecting observations; MSE = 0.0924. Note the improved resolution in the background building columns, the man's face, and the edges of the tripod legs.

6. CONCLUSIONS

This paper has demonstrated that coded apertures designed to meet the Restricted Isometry Property [7] can improve our ability to perform superresolution image reconstruction from noisy, low resolution observations. In particular, building from the theory of RIPs for Toeplitz-structured matrices for compressive sensing [10], we establish a method for generating coded aperture masks in both the conventional coded aperture setting and a Fourier imaging setting; these random masks can be shown to result in an observation matrix which, with high probability, satisfies the RIP. Furthermore, simulations demonstrate that these masks combined with $\ell^2 - \ell^1$ minimization reconstruction methods yield superresolution reconstructions with crisper edges and improved feature resolution over reconstructions achieve without the benefit of coded apertures.

7. REFERENCES

[1] R. C. Hardie, K. J. Barnard, and E.E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1633, 1997.

[2] M. Shankar, R. Willett, N. Pitsianis, T. Schulz, R. Gibbons, R. Te Kolste, J. Carriere, C. Chen, D. Prather, and D. Brady, "Thin infrared imaging systems through multi-channel sampling," Submitted to *Applied Optics*, 2007.

[3] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," in *Handbook of medical imaging*, pp. 393–420. Academic Press, Inc., Orlando, FL, USA, 2000.

[4] E. J. Candes, J. Romberg, and T. Tao, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, November 2004, Submitted.

[5] 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.

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

[7] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 15, no. 12, pp. 4203–4215, 2005.

[8] E. J. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, 2005, To appear.

[9] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

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

[11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," To appear in *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing*, 2007.