AN ARCHITECTURE FOR COMPRESSIVE IMAGING

Michael B. Wakin, Jason N. Laska, Marco F. Duarte, Dror Baron
Shriram Sarvotham, Dharmpal Takhar, Kevin F. Kelly, Richard G. Baraniuk

Department of Electrical and Computer Engineering, Rice University

ABSTRACT

Compressive Sensing is an emerging field based on the revelation that a small group of non-adaptive linear projections of a compressible signal contains enough information for reconstruction and processing. In this paper, we propose algorithms and hardware to support a new theory of *Compressive Imaging*. Our approach is based on a new digital image/video camera that directly acquires random projections of the signal without first collecting the pixels/voxels. Our camera architecture employs a digital micromirror array to perform optical calculations of linear projections of an image onto pseudorandom binary patterns. Its hallmarks include the ability to obtain an image with a single detection element while measuring the image/video fewer times than the number of pixels — this can significantly reduce the computation required for video acquisition/encoding. Because our system relies on a single photon detector, it can also be adapted to image at wavelengths that are currently impossible with conventional CCD and CMOS imagers. We are currently testing a prototype design for the camera and include experimental results.

Index Terms— Data Acquisition, Data Compression, Image Coding, Image Sensors, Video Coding

1. INTRODUCTION

The large amount of raw data acquired in a conventional digital image or video often necessitates immediate compression in order to store or transmit that data. This compression typically exploits a priori knowledge about the data, such as the fact that an N-pixel image can be well approximated as a sparse linear combination of $K \ll N$ wavelets. These appropriate wavelet coefficients can be efficiently computed from the N pixel values and then easily stored or transmitted along with their locations. Similar procedures are applied to videos containing F frames of P pixels each; we let $N = FP$ denote the number of "voxels".

This process has two major potential drawbacks. First, acquiring large amounts of raw image or video data (large N) can be expensive, particularly at wavelengths where CMOS or CCD sensing technology is limited. Second, compressing raw data can be computationally demanding, particularly in the case of video. While there may appear to be no way around this procedure of "sample, process, keep the important information, and throw away the rest," a new theory known as Compressive Sensing (CS) has emerged that offers hope for directly acquiring a compressed digital representation of a signal without first sampling that signal [1–3].

In this paper, we propose algorithms and hardware to support a new theory of Compressive Imaging. Our approach is based on a new digital image/video camera that directly acquires random projections without first collecting the N pixels/voxels [4]. Due to unique measurement approach, it has the ability to obtain an image with a single detection element while measuring the image far fewer times than the number of pixels. Because of this single detector, it can be adapted to image at wavelengths that are currently impossible with conventional CCD and CMOS imagers.

This paper is organized as follows. Section 2 provides an overview of CS, the theoretical foundation for our CI approach. Section 3 overviews our CI framework and hardware testbed and Section 4 presents experimental results.

2. COMPRESSIVE SENSING

CS builds upon a core tenet of signal processing and information theory: that signals, images, and other data often contain some type of *structure* that enables intelligent representation and processing. Current state-of-the-art compression algorithms employ a decorrelating transform to compact a correlated signal's energy into just a few essential coefficients. Such *transform coders* exploit the fact that many signals have a *sparse* representation in terms of some basis Ψ, meaning that a small number K of adaptively chosen transform coefficients can be transmitted or stored rather than $N \gg K$ signal samples. For example, smooth images are sparse in the Fourier basis, and piecewise smooth images are sparse in a wavelet basis; the commercial coding standards JPEG and JPEG2000 directly exploit this sparsity.

The standard procedure for transform coding of sparse signals is to (*i*) acquire the full N-sample signal x; (*ii*) compute the complete set $\{\theta(n)\}$ of transform coefficients $\theta_n = \langle \psi_n, x \rangle$; (*iii*) locate the K largest, significant coefficients and discard the (many) small coefficients; and (*iv*) encode the *values and locations* of the largest coefficients. In cases where N is large and K is small, this procedure can be quite inefficient. Much of the output of the analog-to-digital conversion

Supported by NSF, ONR, AFOSR, DARPA and the Texas Instruments Leadership University Program. Email: {wakin, laska, duarte, drorb, shri, kaka, kkelly, richb}@rice.edu; Web: dsp.rice.edu/cs.

process ends up being discarded (though it is not known a priori which pieces are needed).

This raises a simple question: For a given signal, is it possible to directly estimate the set of large coefficients that will not be discarded by the transform coder? While this seems improbable, the recent theory of *Compressive Sensing* introduced by Candès, Romberg, and Tao [1] and Donoho [2] demonstrates that a signal that is K-sparse in one basis (call it the *sparsity basis*) can be recovered from cK *nonadaptive* linear projections onto a second basis (call it the *measurement basis*) that is incoherent with the first, where where c is a small *overmeasuring* constant. While the measurement process is linear, the reconstruction process is decidedly *nonlinear*.

2.1. Incoherent projections

In CS, we do not measure or encode the K significant $\theta(n)$ directly. Rather, we measure and encode $M < N$ projections $y(m) = \langle x, \phi_m^T \rangle$ of the signal onto a *second set* of basis functions $\{\phi_m\}, m \in \{1, 2, \ldots, M\}$, where ϕ_m^T denotes the transpose of ϕ_m and $\langle \cdot, \cdot \rangle$ denotes the inner product. In matrix notation, we measure

$$y = \Phi x, \tag{1}$$

where y is an $M \times 1$ column vector, and the *measurement basis* matrix Φ is $M \times N$ with each row a basis vector ϕ_m. Since $M < N$, recovery of the signal x from the measurements y is ill-posed in general; however the additional assumption of signal *sparsity* makes recovery possible and practical.

The CS theory tells us that when certain conditions hold, namely that the basis $\{\phi_m\}$ cannot sparsely represent the elements of the sparsity-inducing basis $\{\psi_n\}$ (a condition known as *incoherence* of the two bases [1, 2]) and the number of measurements M is large enough, then it is indeed possible to recover the set of large $\{\theta(n)\}$ (and thus the signal x) from a similarly sized set of measurements $\{y(m)\}$. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of the Fourier basis, or the Fourier basis and wavelets. Significantly, this incoherence also holds with high probability between an arbitrary fixed basis and a randomly generated one (consisting of i.i.d. Gaussian or Bernoulli/Rademacher ± 1 vectors). Signals that are sparsely represented in frames or unions of bases can be recovered from incoherent measurements in the same fashion.

2.2. Signal recovery

The recovery of the sparse set of significant coefficients $\{\theta(n)\}$ can be achieved using *optimization* by searching for the signal with ℓ_0-sparsest[1] coefficients $\{\theta(n)\}$ that agrees with the M observed measurements in y (recall that $M < N$). Unfortunately, solving this ℓ_0 optimization problem is prohibitively complex, and is believed to be NP-hard [5]. The practical revelation that supports the new CS theory is that it is not necessary to solve the ℓ_0-minimization problem to recover the set of significant $\{\theta(n)\}$. In fact, a much easier problem yields an equivalent solution (thanks again to the incoherency of the bases); we need only solve for the ℓ_1-sparsest coefficients θ that agree with the measurements y [1, 2]

$$\widehat{\theta} = \arg\min \|\theta\|_1 \quad \text{s.t. } y = \Phi\Psi\theta. \tag{2}$$

This optimization problem, also known as *Basis Pursuit* [6], is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in N. Although only $K + 1$ measurements are required to recover sparse signals via ℓ_0 optimization [7], with Basis Pursuit one typically requires $M \geq cK$ measurements, where $c > 1$ is an *overmeasuring factor*.

At the expense of slightly more measurements, iterative greedy algorithms have also been developed to recover the signal x from the measurements y. Examples include the iterative Orthogonal Matching Pursuit (OMP) [8], matching pursuit (MP), and tree matching pursuit (TMP) algorithms. The same methods have also been shown to perform well on *compressible signals*, which are not exactly K-sparse but are well approximated by a K-term representation. This is a more realistic model in practice.

3. COMPRESSIVE IMAGING

In this paper, we develop a new system to support what can be called *Compressive Imaging* (CI). Our system incorporates a microcontrolled mirror array driven by pseudorandom and other measurement bases and a single or multiple photodiode optical sensor. This hardware optically computes incoherent image measurements as dictated by the CS theory; we then apply CS reconstruction algorithms to obtain the acquired images. Our camera can also be used to take streaming measurements of a video signal, which can then be recovered using CS techniques designed for either 2-D frame-by-frame reconstruction or joint 3-D reconstruction.

Other desirable features of our system include the use of a single detector (potentially enabling imaging at new wavelengths too expensive to measure using CCD or CMOS technology), universal measurement bases (incoherent with arbitrary sparse bases), encrypted measurements (tied to a random seed that can be kept secure), and scalable progressive reconstruction (yielding better quality as more measurements are obtained).

3.1. Camera hardware

Our hardware realization of the CI concept is a *single pixel camera*; it combines a microcontrolled mirror array displaying a time sequence of M pseudorandom basis images ϕ_m

[1] The ℓ_0 "norm" $\|\theta\|_0$ merely counts the number of nonzero entries in the vector θ.

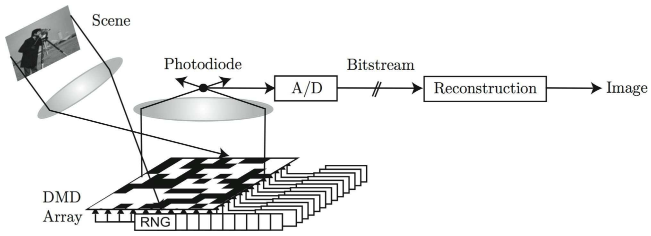


Fig. 1. *Compressive Imaging (CI) camera block diagram. Incident lightfield (corresponding to the desired image x) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern ϕ_m supplied by the random number generators (RNG). Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement $y(m)$.*

with a single optical sensor to compute incoherent image measurements y as in (1) (see Figure 1). By adaptively selecting how many measurements to compute, we trade off the amount of compression versus acquisition time; in contrast, conventional cameras trade off resolution versus the number of pixel sensors.

We employ a Texas Instruments (TI) digital micromirror device (DMD) for generating the random basis patterns. The DMD consists of a 1024×768 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 ($+12$ degrees and -12 degrees from horizontal); thus light falling on the DMD may be reflected in two directions depending on the orientation of the mirrors.

With the help of a biconvex lens, the desired image is formed on the DMD plane; this image acts as an object for the second biconvex lens which focuses the image onto the photodiode. The light is collected from one of the two directions in which it is reflected (e.g., the light reflected by mirrors in the $+12$ degree state). The light from a given configuration of the DMD mirrors is summed at the photodiode to yield an absolute voltage that yields a coefficient $y(m)$ for that configuration. The output of the photodiode is amplified through an op-amp circuit and then digitized by a 12-bit analog-to-digital converter. These photodiode measurements can be interpreted as the inner product of the desired image x with a measurement basis vector ϕ_m. In particular, letting $\rho(m)$ denote the mirror positions of the m-th measurement pattern, the voltage reading from the photodiode v can be written as

$$v(m) \propto \langle x, \phi_m \rangle + \text{DC offset}, \tag{3}$$

where

$$\phi_m = \mathbf{1}_{\{\rho(m)=+12 \text{ degrees}\}} \tag{4}$$

and $\mathbf{1}$ is the indicator function. (The DC offset can be measured by setting all mirrors to -12 degrees; it can then subtracted off.)

Equation (3) holds the key for implementing a CI system. For a given image x, we take M measurements $\{y(1), y(2), \ldots, y(M)\}$ corresponding to mirror configurations $\{\rho(1), \rho(2), \ldots, \rho(M)\}$. Since the patterns $\rho(m)$ are programmable, we can select them to be incoherent with the sparsity-inducing basis (e.g., wavelets or curvelets). As mentioned previously, random or pseudorandom measurement patterns enjoy a useful universal incoherence property with any fixed basis, and so we employ pseudorandom ± 12 degree patterns on the mirrors. These correspond to pseudorandom $0/1$ Bernoulli measurement vectors $\phi_m = \mathbf{1}_{\{\rho(m)=+12 \text{ degrees}\}}$. (The measurements may easily be converted to ± 1 Rademacher patterns by setting all mirrors in $\rho(1)$ to $+12$ degrees and then letting $y(m) \leftarrow 2y(m) - y(1)$ for $m > 1$.) Other options for incoherent CI mirror patterns include $-1/0/1$ group-testing patterns [9]. Mirrors can also be duty-cycled to give the elements of ϕ finer precision, for example to approximate Gaussian measurement vectors [2, 3].

This system directly acquires a reduced set of M incoherent projections of an N-pixel image x *without* first acquiring the N pixel values. Since the camera is "progressive," better quality images (larger K) can be obtained by taking more measurements M. Also, since the data measured by the camera is "future-proof," new reconstruction algorithms based on better sparsifying image transforms can be applied at a later date to obtain even better quality images.

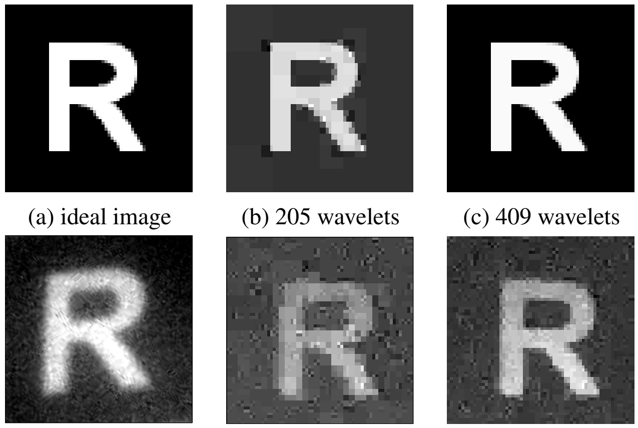
3.2. Related work

Other efforts on CI include [10, 11], which employ optical elements to perform transform coding of multispectral images. These designs obtain sampled outputs that correspond to coded information of interest, such as the wavelength of a given light signal or the transform coefficients in a basis of interest. The elegant hardware designed for these purposes uses concepts that include optical projections, group testing [9], and signal inference. Two notable previous DMD-driven applications involve confocal microscopy [12] and micro-optoelectromechanical (MOEM) systems [13]. For more related work, see the references in [4].

4. EXPERIMENTAL RESULTS

4.1. Imaging results

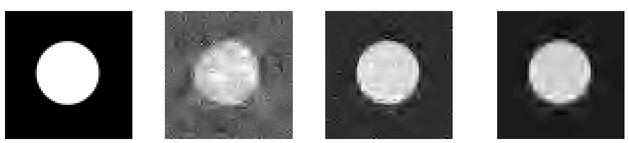
For our imaging experiment, we displayed a printout of the letter "R" in front of the camera; Figure 2(a) shows the printout. Since our test image is piecewise constant (with sharp edges) it can be sparsely represented in the wavelet domain. Figures 2(b) and 2(c) show the best K-term Haar wavelet approximation of the idealized image in Figure 2(a) with $K = 205$ and 409, respectively. Using $M = 819$ and 1600 measurements (roughly $4\times$ the K used in (b) and (c)), we reconstructed the images shown in Figures 2(e) and 2(f) using OMP [8]. This preliminary experiment confirms the feasibility of the CI approach; we are currently working to resolve



(a) ideal image (b) 205 wavelets (c) 409 wavelets

(d) image on DMD (e) 819 meas. (f) 1600 meas.

Fig. 2. *CI DMD imaging of a* 64×64 *(*$N = 4096$ *pixel) image. Ideal image (a) of full resolution and approximated by its (b) largest 400 wavelet coefficients and (c) largest 675 wavelet coefficients. (d) Conventional* 320×240 *camera image acquired at the DMD plane. CS reconstruction from (e) 1600 random measurements and (f) 2700 random measurements. In all cases, Haar wavelets were used for approximation or reconstruction.*



(a) frame 32 (b) 2D CS (c) 3D CS (d) 3D wavelets

Fig. 3. *(a) Frame 32 of a 64-frame video sequence (64x64 images of a disk moving from top to bottom, corresponding to 262,144 3-D voxels). (b) CS frame-by-frame reconstruction using 20,000 total 2-D random projections (313 independent 2-D projections for each image in the sequence). (c) Full 3-D video reconstruction from 20,000 3-D random projections (using 3-D wavelets as the sparsity-inducing basis). (d) Result of 3-D wavelet thresholding to 2000 total coefficients.*

minor calibration and noise issues to improve the reconstruction quality.

4.2. Video results

In principle, our camera can also be used to take streaming measurements of video sequences. Figure 3 shows a simulation comparing three different schemes for video acquisition/coding: frame-by-frame acquisition that independently acquires each image of the video sequence using 2-D random projections, with frame-by-frame reconstruction at the decoder; joint acquisition that acquires 3-D random projections of the entire video sequence, with 3-D reconstruction at the decoder using 3-D wavelets as a sparsity-inducing basis; and (for comparison) 3-D wavelet encoding that thresholds the 3-D wavelet transform of the entire video sequence. As we see from Figure 3, 3-D reconstruction significantly outperforms 2-D frame-by-frame reconstruction, because a 3-D video wavelet transform is significantly more sparse than a collection of 2-D image wavelet transforms. Current work focuses on extending the camera design to take fully 3-D incoherent projections and exploiting more advanced models for 3-D video structure (beyond 3-D wavelets) for reconstruction.

5. DISCUSSION AND CONCLUSIONS

In this paper, we have presented a prototype imaging system that successfully employs compressive sensing (CS) principles. The camera has many attractive features, including simplicity, universality, robustness, and scalability, that should enable it to impact a variety of different applications. Another interesting and potentially useful practical feature of our system is that it off-loads processing from data collection into data reconstruction. Not only will this lower the complexity and power consumption of the device, but it will enable new adaptive new measurements schemes. The most intriguing feature of the system is that, since it relies on a single photon detector, it can be adapted to image at wavelengths that are currently impossible with conventional CCD and CMOS imagers.[2]

6. REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," 2004, Preprint.

[2] D. Donoho, "Compressed sensing," 2004, Preprint.

[3] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," 2004, Preprint.

[4] D. Takhar, J. N. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. K. Kelly, and R. G. Baraniuk, "A new camera architecture based on optical-domain compression," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, Jan. 2006, vol. 6065.

[5] E. Candès and T. Tao, "Error correction via linear programming," 2005, Preprint.

[6] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[7] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005, Available at http://www.dsp.rice.edu/cs.

[8] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," Apr. 2005, Preprint.

[9] G. Cormode and S. Muthukrishnan, "Towards an algorithmic theory of compressed sensing," *DIMACS Tech. Report 2005-25*, 2005.

[10] N. P. Pitsianis, D. J. Brady, and X. Sun, "Sensor-layer image compression based on the quantized cosine transform," in *SPIE Visual Information Processing XIV*, 2005.

[11] D. J. Brady, M. Feldman, N. Pitsianis, J. P. Guo, A. Portnoy, and M. Fiddy, "Compressive optical MONTAGE photography," in *SPIE Photonic Devices and Algorithms for Computing VII*, 2005.

[12] P. M. Lane, R. P. Elliott, and C. E. MacAulay, "Confocal microendoscopy with chromatic sectioning," in *Proc. SPIE*, 2003, vol. 4959, pp. 23–26.

[13] R. A. DeVerse, R. R. Coifman, A. C. Coppi, W. G. Fateley, F. Geshwind, R. M. Hammaker, S. Valenti, and F. J. Warner, "Application of spatial light modulators for new modalities in spectrometry and imaging," in *Proc. SPIE*, 2003, vol. 4959.

[2]Thanks to Texas Instruments for providing the TI DMD developer's kit and accessory light modulator package (ALP). Thanks also to Dave Brady and Dennis Healy for enlightening discussions.