Single-Pixel Imaging via Compressive Sampling

Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska
Ting Sun, Kevin F. Kelly, Richard G. Baraniuk
Rice University

Humans are visual animals, and imaging sensors that extend our reach – cameras – have improved dramatically in recent times thanks to the introduction of CCD and CMOS digital technology. Consumer digital cameras in the mega-pixel range are now ubiquitous thanks to the happy coincidence that the semiconductor material of choice for large-scale electronics integration (silicon) also happens to readily convert photons at visual wavelengths into electrons. On the contrary, imaging at wavelengths where silicon is blind is considerably more complicated, bulky, and expensive. Thus, for comparable resolution, a $500 digital camera for the visible becomes a $50,000 camera for the infrared.

In this paper, we present a new approach to building simpler, smaller, and cheaper digital cameras that can operate efficiently across a much broader spectral range than conventional silicon-based cameras. Our approach fuses a new camera architecture based on a digital micromirror device (DMD – see Sidebar: Spatial Light Modulators) with the new mathematical theory and algorithms of *compressive sampling* (CS – see Sidebar: Compressive Sampling in a Nutshell).

CS combines sampling and compression into a single nonadaptive linear measurement process [1–4]. Rather than measuring pixel samples of the scene under view, we measure inner products between the scene and a set of test functions. Interestingly, random test functions play a key role, making each measurement a random sum of pixel values taken across the entire image. When the scene under view is compressible by an algorithm like JPEG or JPEG2000, the CS theory enables us to stably reconstruct an image of the scene from fewer measurements than the number of reconstructed pixels. In this manner we achieve sub-Nyquist image acquisition.

Our "single-pixel" CS camera architecture is basically an optical computer (comprising a DMD, two lenses, a single photon detector, and an analog-to-digital (A/D) converter) that computes random linear measurements of the scene under view. The image is then recovered or

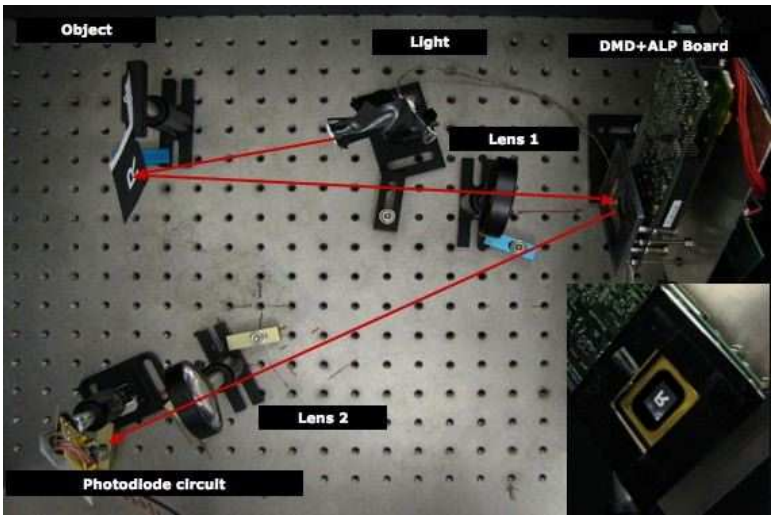


Fig. 1. *Aerial view of the single-pixel compressive sampling (CS) camera in the lab [5].*

processed from the measurements by a digital computer. The camera design reduces the required size, complexity, and cost of the photon detector array down to a single unit, which enables the use of exotic detectors that would be impossible in a conventional digital camera. The random CS measurements also enable a tradeoff between space and time during image acquisition. Finally, since the camera compresses as it images, it has the capability to efficiently and scalably handle high-dimensional data sets from applications like video and hyperspectral imaging.

This article is organized as follows. After describing the hardware, theory, and algorithms of the single-pixel camera in detail, we analyze its theoretical and practical performance and compare it to more conventional cameras based on pixel arrays and raster scanning. We also explain how the camera is information scalable in that its random measurements can be used to directly perform simple image processing tasks, such as target classification, without first reconstructing the underlying imagery. We conclude with a review of related camera architectures and a discussion of ongoing and future work.

I. The Single-Pixel Camera

Architecture

The single-pixel camera is an optical computer that sequentially measures the inner products $y[m] = \langle x, \phi_m \rangle$ between an N-pixel sampled version x of the incident light-field from the scene under view and a set of two-dimensional (2D) test functions $\{\phi_m\}$ [5]. As shown in Fig. 1, the

light-field is focused by biconvex Lens 1 not onto a CCD or CMOS sampling array but rather onto a DMD consisting of an array of N tiny mirrors (see Sidebar: Spatial Light Modulators).

Each mirror corresponds to a particular pixel in x and ϕ_m and can be independently oriented either towards Lens 2 (corresponding to a 1 at that pixel in ϕ_m) or away from Lens 2 (corresponding to a 0 at that pixel in ϕ_m). The reflected light is then collected by biconvex Lens 2 and focused onto a single photon detector (the single pixel) that integrates the product $x[n]\phi_m[n]$ to compute the measurement $y[m] = \langle x, \phi_m \rangle$ as its output voltage. This voltage is then digitized by an A/D converter. Values of ϕ_m between 0 and 1 can be obtained by dithering the mirrors back and forth during the photodiode integration time. To obtain ϕ_m with both positive and negative values (± 1, for example), we estimate and subtract the mean light intensity from each measurement, which is easily measured by setting all mirrors to the full-on 1 position.

To compute CS randomized measurements $y = \Phi x$ as in (1), we set the mirror orientations ϕ_m randomly using a pseudo-random number generator, measure $y[m]$, and then repeat the process M times to obtain the measurement vector y. Recall from Sidebar: Compressive Sampling in a Nutshell that we can set $M = O(K \log(N/K))$ which is $\ll N$ when the scene being imaged is compressible by a compression algorithm like JPEG or JPEG2000. Since the DMD array is programmable, we can also employ test functions ϕ_m drawn randomly from a fast transform such as a Walsh, Hadamard, or Noiselet transform [6, 7].

The single-pixel design reduces the required size, complexity, and cost of the photon detector array down to a single unit, which enables the use of exotic detectors that would be impossible in a conventional digital camera. Example detectors include a photomultiplier tube or an avalanche photodiode for low-light (photon-limited) imaging (more on this below), a sandwich of several photodiodes sensitive to different light wavelengths for multimodal sensing, a spectrometer for hyperspectral imaging, and so on.

In addition to sensing flexibility, the practical advantages of the single-pixel design include the facts that the quantum efficiency of a photodiode is higher than that of the pixel sensors in a typical CCD or CMOS array and that the fill factor of a DMD can reach 90% whereas that of a CCD/CMOS array is only about 50%. An important advantage to highlight is the fact that each CS measurement receives about $N/2$ times more photons than an average pixel sensor, which significantly reduces image distortion from dark noise and read-out noise. Theoretical advantages

(a)  (b)  (c) 

Fig. 2. *Single-pixel photo album. (a)* 256×256 *conventional image of a black-and-white* R*. (b) Single-pixel camera reconstructed image from* $M = 1300$ *random measurements (*$50\times$ *sub-Nyquist). (c)* 256×256 *pixel color reconstruction of a printout of the Mandrill test image imaged in a low-light setting using a single photomultiplier tube sensor, RGB color filters, and* $M = 6500$ *random measurements.*

that the design inherits from the CS theory include its universality, robustness, and progressivity.

The single-pixel design falls into the class of multiplex cameras [8]. The baseline standard for multiplexing is classical raster scanning, where the test functions $\{\phi_m\}$ are a sequence of delta functions $\delta[n-m]$ that turn on each mirror in turn. As we will see below, there are substantial advantages to operating in a CS rather than raster scan mode, including fewer total measurements (M for CS rather than N for raster scan) and significantly reduced dark noise.

Image acquisition examples

Figure 2 (a) and (b) illustrates a target object (a black-and-white printout of an "R") x and reconstructed image $\widehat{x}$ taken by the single-pixel camera prototype in Fig. 1 using $N = 256 \times 256$ and $M = N/50$ [5]. Fig. 2(c) illustrates an $N = 256 \times 256$ color single-pixel photograph of a printout of the Mandrill test image taken under low-light conditions using RGB color filters and a photomultiplier tube with $M = N/10$. In both cases, the images were reconstructed using Total Variation minimization, which is closely related to wavelet coefficient ℓ_1 minimization [2].

Structured illumination configuration

In a reciprocal configuration to that in Fig. 1, we can illuminate the scene using a projector displaying a sequence of random patterns $\{\phi_m\}$ and collect the reflected light using a single lens and photodetector. Such a "structured illumination" setup has advantages in applications where we can control the light source. In particular, there are intriguing possible combinations of single-pixel imaging with techniques such as 3D imaging and dual photography [9].

Shutterless video imaging

We can also acquire video sequences using the single-pixel camera. Recall that a traditional video camera opens a shutter periodically to capture a sequence of images (called video frames) that are then compressed by an algorithm like MPEG that jointly exploits their spatio-temporal redundancy. In contrast, the single-pixel video camera needs no shutter; we merely continuously sequence through randomized test functions ϕ_m and then reconstruct a video sequence using an optimization that exploits the video's spatio-temporal redundancy [10].

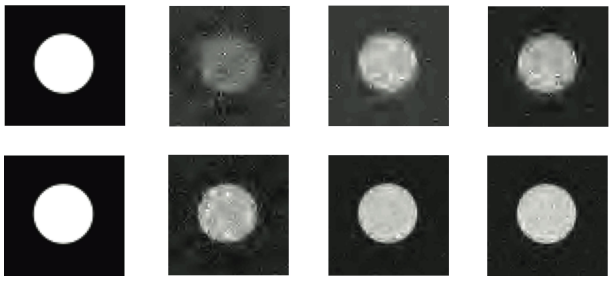
If we view a video sequence as a 3D space/time cube, then the test functions ϕ_m lie concentrated along a periodic sequence of 2D image slices through the cube. A naïve way to reconstruct the video sequence would group the corresponding measurements $y[m]$ into groups where the video is quasi-stationary and then perform a 2D frame-by-frame reconstruction on each group. This exploits the compressibility of the 3D video cube in the space but not time direction.

A more powerful alternative exploits the fact that even though each ϕ_m is testing a different 2D image slice, the image slices are often related temporally through smooth object motions in the video. Exploiting this 3D compressibility in both the space and time directions and inspired by modern 3D video coding techniques [11], we can, for example, attempt to reconstruct the sparsest video space/time cube in the 3D wavelet domain.

These two approaches are compared in the simulation study illustrated in Fig. 3. We employed simplistic 3D tensor product Daubechies-4 wavelets in all cases. As we see from the figure, 3D reconstruction from 2D random measurements performs almost as well as 3D reconstruction from 3D random measurements, which are not directly implementable with the single-pixel camera.

II. Single-Pixel Camera Tradeoffs

The single-pixel camera is a flexible architecture to implement a range of different multiplexing methodologies, just one of them being CS. In this section, we analyze the performance of CS and two other candidate multiplexing methodologies and compare them to the performance of a brute-force array of N pixel sensors. Integral to our analysis is the consideration of Poisson photon counting noise at the detector, which is image-dependent. We conduct two separate analyses to assess the "bang for the buck" of CS. The first is a theoretical analysis that provides general guidance. The second is an experimental study that indicates how the systems typically perform



(a) frame 32 (b) 2D meas + 2D recon (c) 2D meas + 3D recon (d) 3D meas + 3D recon

Fig. 3. *Frame 32 from a reconstructed video sequence using (top row) $M = 20{,}000$ and (bottom row) $M = 50{,}000$ measurements (simulation from [10]). (a) Original frame of an $N = 64 \times 64 \times 64$ video of a disk simultaneously dilating and translating. (b) Frame-by-frame 2D measurements + frame-by-frame 2D reconstruction; MSE = 3.63 and 0.82, respectively. (c) Frame-by-frame 2D measurements + joint 3D reconstruction; MSE = 0.99 and 0.24, respectively. (d) Joint 3D measurements + joint 3D reconstruction; MSE = 0.76 and 0.18, respectively.*

in practice.

Scanning methodologies

The four imaging methodologies we consider are:

- **Pixel array** (PA): a separate sensor for each of the N pixels receives light throughout the total capture time T. This is actually not a multiplexing system, but we use it as the gold standard for comparison.
- **Raster scan** (RS): a single sensor takes N light measurements sequentially from each of the N pixels over the capture time. This corresponds to test functions $\{\phi_m\}$ that are delta functions and thus $\Phi = I$. The measurements y thus directly provide the acquired image $\widehat{x}$.
- **Basis scan** (BS): a single sensor takes N light measurements sequentially from different combinations of the N pixels as determined by test functions $\{\phi_m\}$ that are not delta functions but from some more general basis [12]. In our analysis, we assume a Walsh basis modified to take the values 0/1 rather than ± 1; thus $\Phi = W$, where W is the 0/1 Walsh matrix. The acquired image is obtained from the measurements y by $\widehat{x} = \Phi^{-1} y = W^{-1} y$.
- **Compressive sampling** (CS): a single sensor takes $M \leq N$ light measurements sequentially from different combinations of the N pixels as determined by random 0/1 test functions $\{\phi_m\}$. Typically, we set $M = O(K \log(N/K))$ which is $\ll N$ when the image is compressible. In our analysis, we assume that the M rows of the matrix Φ consist of randomly

drawn rows from a 0/1 Walsh matrix that are then randomly permuted (we ignore the first row consisting of all 1's). The acquired image is obtained from the measurements y via a sparse reconstruction algorithm such as BPIC (see Sidebar: Compressive Sampling in a Nutshell).

Theoretical analysis

In this section, we conduct a theoretical performance analysis of the above four scanning methodologies in terms of the required dynamic range of the photodetector, the required bit depth of the A/D converter, and the amount of Poisson photon counting noise. Our results are pessimistic in general; we show in the next section that the average performance in practice can be considerably better. Our results are summarized in Table I. An alternative analysis of CS imaging for piecewise smooth images in Gaussian noise has been reported in [13].

Dynamic range: We first consider the photodetector dynamic range required to match the performance of the baseline PA. If each detector in the PA has a linear dynamic range of 0 to D, then it is easy to see that single-pixel RS detector need only have that same dynamic range. In contrast, each Walsh basis test function contains $N/2$ 1's, and so directs $N/2$ times more light to the detector. Thus, BS and CS each require a larger linear dynamic range of 0 to $ND/2$. On the positive side, since BS and CS collect considerably more light per measurement than the PA and RS, they benefit from reduced detector nonidealities like dark currents.

Quantization error: We now consider the number of A/D bits required within the required dynamic range to match the performance of the baseline PA in terms of worst-case quantization distortion. Define the mean-squared error (MSE) between the true image x and its acquired version $\widehat{x}$ as $\mathrm{MSE} = \frac{1}{N}\|x - \widehat{x}\|_2^2$. Assuming that each measurement in the PA and RS is quantized to B bits, the worst-case mean-squared quantization error for the quantized PA and RS images is $\mathrm{MSE} = \sqrt{N}D2^{-B-1}$ [14]. Due to its larger dynamic range, BS requires $B + \log_2 N$ bits per measurement to reach the same MSE distortion level. Since the distortion of CS reconstruction is up to C_N times larger than the distortion in the measurement vector (see Sidebar: Compressive Sensing in a Nutshell), we require up to an additional $\log_2 C_N$ bits per measurement. One empirical study has found roughly that C_N lies between 8 and 10 for a range of different random measurement configurations [15]. Thus, BS and CS require a higher-resolution A/D converter than PA and RS to acquire an image with the same worst-case quantization distortion.

TABLE I

Comparison of the four camera scanning methodologies.

	Pixel Array	Raster Scan	Basis Scan	Compressive Sampling
Number of measurements	N	N	N	$M \leq N$
Dynamic range	D	D	$\frac{ND}{2}$	$\frac{ND}{2}$
Quantization (total bits)	NB	NB	$N(B+\log_2 N)$	$M(B+\log_2 N+\log_2 C_N+1)$
Photon counting MSE	$\frac{P}{T}$	$N\frac{P}{T}$	$(3N-2)\frac{P}{T}$	$< 3C_N^2 M\frac{P}{T}$

Photon counting noise: In addition to quantization error from the A/D converter, each camera will also be affected by image-dependent Poisson noise due to photon counting [16]. We compare the MSE due to photon counting for each of the scanning methodologies. The details are worked out in Appendix: Poisson Photon Counting Calculations and are summarized in Table I. We see that the MSE of BS is about three times that of RS. Moreover, when $M < \frac{N}{3C_N^2}$, the MSE of CS is lower than that of RS. We emphasize that in the CS case, we have only a fairly loose upper bound and that there exist alternative CS reconstruction algorithms with better performance guarantees, such as the Dantzig Selector [3].

Summary: From Table I, we see that the advantages of a single-pixel camera over a PA come at the cost of more stringent demands on the sensor dynamic range and A/D quantization and larger MSE due to photon counting effects. Additionally, the linear dependence of the MSE on the number of image pixels N for BS and RS is a potential deal-breaker for high-resolution imaging. The promising aspect of CS is the logarithmic dependence of its MSE on N through the relationship $M = O(K\log(N/K))$.

Experimental results

Since CS acquisition/reconstruction methods often perform much better in practice than the above theoretical bounds suggest, in this section we conduct a simple experiment using real data from the CS imaging testbed depicted in Fig. 1. Thanks to the programmability of the testbed, we acquired RS, BS, and CS measurements from the same hardware. We fixed the number of A/D converter bits across all methodologies. Figure 4 shows the pixel-wise MSE for the capture of a $N = 128 \times 128$ pixel "R" test image as a function of the total capture time T. Here the MSE combines both quantization and photon counting effects. For CS we took $M = N/10$ total measurements per capture and used a Daubechies-4 wavelet basis for the sparse reconstruction.

We make two observations. First, the performance gain of BS over RS contradicts the above

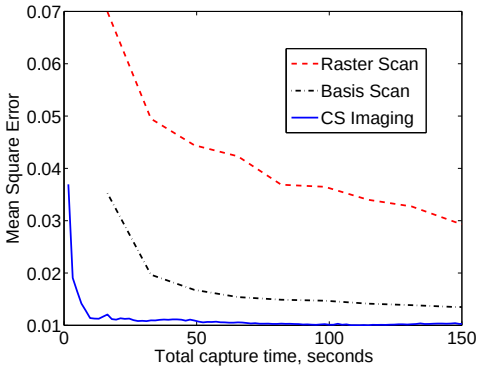


Fig. 4. *Average MSE for Raster Scan (RS), Basis Scan (BS), and Compressive Sampling (CS) single-pixel images as a function of the total image capture time T (real data).*

worst-case theoretical calculations. We speculate that the contribution of the sensor's dark current, which is not accounted for in our analysis, severely degrades RS's performance. Second, the performance gain of CS over both RS and BS is clear: images can either be acquired in much less time for the same MSE or with much lower MSE in the same amount of time.

III. **Information Scalability and the Smashed Filter**

While the CS literature has focused almost exclusively on problems in signal and image reconstruction or approximation, reconstruction is frequently not the ultimate goal. For instance, in many image processing and computer vision applications, data is acquired only for the purpose of making a detection, classification, or recognition decision. Fortunately, the CS framework is *information scalable* to a much wider range of statistical inference tasks [17–19]. Tasks such as detection do not require reconstruction, but only require estimates of the relevant sufficient statistics for the problem at hand. Moreover, in many cases it is possible to directly extract these statistics from a small number of random measurements without ever reconstructing the image.

The *matched filter* is a key tool in detection and classification problems that involve searching for a target template in a scene. A complicating factor is that often the target is transformed in some parametric way — for example the time or Doppler shift of a radar return signal; the translation and rotation of a face in a face recognition task; or the roll, pitch, yaw, and scale of a vehicle viewed from an aircraft. The matched filter detector operates by forming comparisons between the given test data and all possible transformations of the template to find the match that optimizes some performance metric. The matched filter classifier operates in the same way

but chooses the best match from a number of different potential transformed templates.

The naïve approach to matched filtering with CS would first reconstruct the images under consideration and then apply a standard matched filtering technique. In contrast, the *smashed filter* (for dimensionally reduced matched filter) performs all of its operations directly on the random measurements [18].

The two key elements of the smashed filter are the generalized likelihood ratio test and the concept of an image appearance manifold [20]. If the parametric transformation affecting the template image is well-behaved, then the set of transformed templates forms a low-dimensional manifold in the high-dimensional pixel space $\mathbb{R}^N$ with the dimension K equal to the number of independent parameters.[1] Thus, the matched filter classifier can be interpreted as classifying a test image according to the closest template manifold in $\mathbb{R}^N$.

The smashed filter exploits a recent result that the structure of a smooth K-dimensional manifold in $\mathbb{R}^N$ is preserved with high probability under a random projection to the lower dimensional space $\mathbb{R}^M$ as long as $M = O(K \log N)$ [21]. This is reminiscent of the number of measurements required for successful CS but with K now the manifold dimension. Thus, to classify an N-pixel test image, we can alternatively compare M random measurements of the test image to the M-dimensional projections (using the same Φ) of the candidate image template manifolds. The upshot is that all necessary computations can be made directly in $\mathbb{R}^M$ rather than in $\mathbb{R}^N$. As in the conventional matched filter, a byproduct of the closest manifold search is an estimate of the template parameters that best match the test image. Previous work in the computer science community (the other "CS") has also employed the Johnson-Lindenstrauss lemma [22] to reduce the data dimensionality prior to computing features for classification; however, they have not considered the intrinsic manifold structure manifested in many image processing and computer vision settings.

Figure 5 demonstrates the effectiveness of smashed filtering for the task of classifying an $N = 128 \times 128$ pixel test image under an unknown translation in the vertical and horizontal directions (hence $K = 2$). The three classes correspond to different translations of a bus, truck, or tank. The test data was generated randomly from one of the three classes. The random measurements were

[1]For the purposes of the discussion here, a K-dimensional manifold can be interpreted as a K-dimensional hypersurface in $\mathbb{R}^N$.

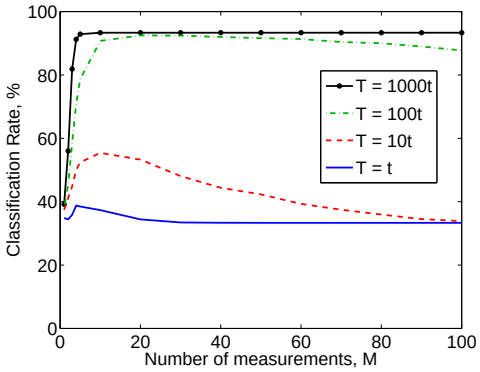


Fig. 5. *Smashed filter image classification performance plotted as a function of the number of random measurements M from a simulated single-pixel CS camera and the total acquisition time T.*

produced using a simulated single-pixel CS camera that takes into account the Poisson photon counting noise associated with a total measurement interval of length T. We average over 10,000 iterations of each experiment. We see that increasing the number of measurements improves performance at first; however, performance then degrades due to the reduced time available to obtain each measurement. Correspondingly, increasing the total capture time improves the performance of the algorithm.

IV. **Other Multiplexing Camera Architectures**

Two notable existing DMD-driven imaging applications involve confocal microscopy (which relates closely to the raster scan strategy studied above) [23, 24] and micro-optoelectromechanical (MOEM) systems [12, 25, 26]. In a MOEM system, a DMD is positioned between the scene and the sensing array to perform column-wise multiplexing by allowing only the light from the desired pixels to pass through. In [12, 25] the authors propose sets of N Hadamard patterns, which enables simple demultiplexing, and randomized Hadamard patterns, which yield a uniform signal-to-noise ratio among the measurements.

Other compressive cameras developed to date include [27, 28], 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 optical projections, group testing, and signal inference. Recent work in [29] has compared several single and multiple pixel imaging strategies of varying complexities; their simulation

results for Poisson counting noise agree closely with those above.

Finally, in [30] the authors use CS principles and a randomizing lens to boost both the resolution and robustness of a conventional digital camera.

V. **Conclusions**

For certain applications, CS promises to substantially increase the performance and capabilities of data acquisition, processing, and fusion systems while lowering the cost and complexity of deployment. A useful practical feature of the CS approach is that it off-loads processing from data acquisition (which can be complicated and expensive) into data reconstruction or processing (which can be performed on a digital computer, perhaps not even co-located with the sensor).

We have presented an overview of the theory and practice of a simple yet flexible single-pixel architecture for CS based on a DMD spatial light modulator. While there are promising potential applications where current digital cameras have difficulty imaging, there are clear tradeoffs and challenges in the single-pixel design. Our current and planned work involves better understanding and addressing these tradeoffs and challenges. Other potential avenues for research include extending the single-pixel concept to wavelengths where the DMD fails as a modulator, such as THz and X-rays.

Sidebar: Compressive Sampling in a Nutshell

Compressive Sampling (CS) is based on the recent understanding that a small collection of nonadaptive linear measurements of a compressible signal or image contain enough information for reconstruction and processing [1–3]. For a tutorial treatment see [4] or the paper by Romberg in this issue.

The traditional approach to digital data acquisition samples an analog signal uniformly at or above the Nyquist rate. In a digital camera, the samples are obtained by a 2D array of N pixel sensors on a CCD or CMOS imaging chip. We represent these samples using the vector x with elements $x[n]$, $n = 1, 2, \ldots, N$. Since N is often very large, e.g. in the millions for today's consumer digital cameras, the raw image data x is often compressed in the following multi-step transform coding process.

The first step in transform coding represents the image in terms of the coefficients $\{\alpha_i\}$ of an orthonormal basis expansion $x = \sum_{i=1}^{N} \alpha_i \psi_i$ where $\{\psi_i\}_{i=1}^{N}$ are the $N \times 1$ basis vectors.

Forming the coefficient vector α and the $N \times N$ basis matrix $\Psi := [\psi_1|\psi_2|\dots|\psi_N]$ by stacking the vectors $\{\psi_i\}$ as columns, we can concisely write the samples as $x = \Psi\alpha$. The aim is to find a basis where the coefficient vector α is *sparse* (where only $K \ll N$ coefficients are nonzero) or r*-compressible* (where the coefficient magnitudes decay under a power law with scaling exponent $-r$). For example, natural images tend to be compressible in the discrete cosine transform (DCT) and wavelet bases on which the JPEG and JPEG-2000 compression standards are based. The second step in transform coding encodes only the values and locations of the K significant coefficients and discards the rest.

This sample-then-compress framework suffers from three inherent inefficiencies: First, we must start with a potentially large number of samples N even if the ultimate desired K is small. Second, the encoder must compute all of the N transform coefficients $\{\alpha_i\}$, even though it will discard all but K of them. Third, the encoder faces the overhead of encoding the locations of the large coefficients.

As an alternative, CS bypasses the sampling process and directly acquires a condensed representation using $M < N$ linear measurements between x and a collection of test functions $\{\phi_m\}_{m=1}^{M}$ as in $y[m] = \langle x, \phi_m \rangle$. Stacking the measurements $y[m]$ into the $M \times 1$ vector y and the test functions ϕ_m^T as rows into an $M \times N$ matrix Φ we can write

$$y = \Phi x = \Phi\Psi\alpha. \tag{1}$$

The measurement process is non-adaptive in that Φ does not depend in any way on the signal x.

The transformation from x to y is a *dimensionality reduction* and so loses information in general. In particular, since $M < N$, given y there are infinitely many x' such that $\Phi x' = y$. The magic of CS is that Φ can be designed such that sparse/compressible x can be recovered exactly/approximately from the measurements y.

While the design of Φ is beyond the scope of this review, an intriguing choice that works with high probability is a random matrix. For example, we can draw the elements of Φ as independent and identically distributed (i.i.d.) ± 1 random variables from a uniform Bernoulli distribution [22]. Then, the measurements y are merely M different randomly signed linear combinations of the elements of x. Other possible choices include i.i.d., zero-mean, $1/N$-variance Gaussian entries (white noise) [1–3, 22], randomly permuted vectors from standard orthonormal bases, or random

subsets of basis vectors [7], such as Fourier, Walsh-Hadamard, or Noiselet [6] bases. The latter choices enable more efficient reconstruction through fast algorithmic transform implementations. In practice, we employ a pseudo-random Φ driven by a pseudo-random number generator.

To recover the image x from the random measurements y, the traditional favorite method of least squares can be shown to fail with high probability. Instead, it has been shown that using the ℓ_1 optimization [1–3]

$$\widehat{\alpha} = \operatorname{argmin} \|\alpha'\|_1 \quad \text{such that} \quad \Phi\Psi\alpha' = y \tag{2}$$

we can exactly reconstruct K-sparse vectors and closely approximate compressible vectors stably with high probability using just $M \geq O(K \log(N/K))$ random measurements. This is a convex optimization problem that conveniently reduces to a linear program known as *basis pursuit* [1–3]. There are a range of alternative reconstruction techniques based on greedy, stochastic, and variational algorithms [4].

If the measurements y are corrupted by noise, then the solution to the alternative ℓ_1 minimization, which we dub *basis pursuit with inequality constrains* (BPIC) [3]

$$\widehat{\alpha} = \arg\min \|\alpha'\|_1 \quad \text{such that} \quad \|y - \Phi\Psi\alpha'\|_2 < \epsilon, \tag{3}$$

satisfies $\|\widehat{\alpha} - \alpha\|_2 < C_N\epsilon + C_K\sigma_K(x)$ with overwhelming probability. C_N and C_K are the noise and approximation error amplification constants, respectively; ϵ is an upper bound on the noise magnitude, and $\sigma_K(x)$ is the ℓ_2 error incurred by approximating α using its largest K terms. This optimization can be solved using standard convex programming algorithms.

In addition to enabling sub-Nyquist measurement, CS enjoys a number of attractive properties [4]. CS measurements are *universal* in that the same random matrix Φ works simultaneously for exponentially many sparsitfying bases Ψ with high probability; no knowledge is required of the nuances of the data being acquired. Due to the incoherent nature of the measurements, CS is *robust* in that the measurements have equal priority, unlike the Fourier or wavelet coefficients in a transform coder. Thus, one or more measurements can be lost without corrupting the entire reconstruction. This enables a *progressively better reconstruction* of the data as more measurements are obtained. Finally, CS is *asymmetrical* in that it places most of its computational complexity in the recovery system, which often has more substantial computational resources than

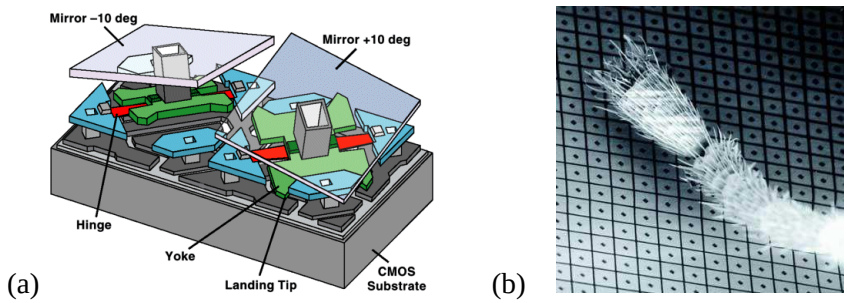


Fig. 6. *(a) Schematic of two mirrors from a Texas Instruments digital micromirror device (DMD). (b) A portion of an actual DMD array with an ant leg for scale. (Image provided by DLP Products, Texas Instruments.)*

the measurement system.

Sidebar: Spatial Light Modulators

A spatial light modulator (SLM) modulates the intensity of a light beam according to a control signal. A simple example of a transmissive SLM that either passes or blocks parts of the beam is an overhead transparency. Another example is a liquid crystal display (LCD) projector.

The Texas Instruments (TI) digital micromirror device (DMD) is a reflective SLM that selectively redirects parts of the light beam [31]. The DMD consists of an array of bacterium-sized, electrostatically actuated micro-mirrors, where each mirror in the array is suspended above an individual static random access memory (SRAM) cell (see Fig. 6). Each mirror rotates about a hinge and can be positioned in one of two states ($+10$ degrees and -10 degrees from horizontal) according to which bit is loaded into the SRAM cell; thus light falling on the DMD can be reflected in two directions depending on the orientation of the mirrors.

The DMD micro-mirrors in our lab's TI DMD 1100 developer's kit[2] and accessory light modulator package (ALP)[3] form a pixel array of size 1024×768. This limits the maximum native resolution of our single-pixel camera. However, mega-pixel DMDs are already available for the display and projector market.

Appendix: Poisson Photon Counting Calculations

This appendix derives the average mean squared error (MSE) of the four camera schemes studied in the Single-Pixel Camera Tradeoffs Section under Poisson counting noise (the last row

[2]Tyrex Services Group Ltd., `http://www.tyrexservices.com`
[3]ViALUX GmbH, `http://www.vialux.de`

of Table I). Let $\widehat{x}$ be the estimated version of the ideal image x. Assuming that the pixel estimates are unbiased and independent with variances $\sigma^2_{\widehat{x}[n]}$, we calculate that

$$E\left[\text{MSE}\right] = E\left[\frac{1}{N}\|x-\widehat{x}\|_2^2\right] = \frac{1}{N}\sum_{n=1}^{N}\sigma^2_{x[n]}. \tag{4}$$

We now briefly review the Poisson model of photon detection. Consider a point in the scene under view that emits photons at a rate of P photons/s; then over τ s the number of photons follows a Poisson distribution with mean and variance $\lambda = P\tau$. To form an unbiased estimate $\widehat{P}$ of the rate, we collect and count photons with a photodetector over τ s and then normalize the count by dividing by τ. The variance of this estimator is then $\sigma^2_{\widehat{P}} = P/\tau$. To simplify the analysis we assume that the photon rate of each image pixel $x[n],\ n = 1, 2, \ldots, N$ is an i.i.d. Gaussian random variable with mean $\mu_{x[n]} = P$. Let the total image capture time be T s.

In the PA, each pixel $x[n]$ has its own dedicated sensor that counts photons for the entire period T. The number of received photons $p[n]$ is Poisson with mean $\lambda[n] = Tx[n]$. The time normalized measurement $\widehat{x}[n] = p[n]/T$ has variance $\sigma^2_{\widehat{x}[n]} = \frac{x[n]}{T}$, and thus as $N \to \infty$ we have

$$E\left[\text{MSE}\right] = \frac{1}{N}\sum_{n=1}^{N}\sigma^2_{\widehat{x}[n]} = \frac{1}{N}\sum_{n=1}^{N}\frac{x[n]}{T} \approx \frac{P}{T}. \tag{5}$$

The RS is similar, except that one sensor is time-shared over each of the N pixels, meaning that the time for each measurement is reduced from T to T/N. Accounting for this, we obtain as $N \to \infty$ that $E\left[\text{MSE}\right] \approx N\frac{P}{T}$.

In BS and CS, the single sensor measures the sum of a group of $N/2$ pixels for T/N s in BS and T/M s in CS. As $N \to \infty$, the photon rate incident at the sensor is approximately $\frac{NE[x[n]]}{2}$, and so each measurement follows a Poisson distribution with mean and variance $\frac{PT}{2}$ for BS and $\frac{PTN}{2M}$ for CS. The time-normalized measurement $y[m]$ thus has variance $\frac{N^2P}{2T}$ for BS and $\frac{NMP}{2T}$ for CS. Continuing for BS, we estimate the image as $\widehat{x} = W^{-1}y$. The variance of the pixel estimates is $\frac{2(N+1)P}{T}$ and thus the average MSE equals $E\left[\text{MSE}\right] = \frac{(3N-2)P}{T}$.

For CS, we estimate the image via a nonlinear optimization that requires that the entries of Φ be $\pm\frac{1}{\sqrt{M}}$ rather than 0/1. Hence, we correct the measurements y by recentering and scaling, which changes the measurement variances to $\frac{3NP}{T}$ and results in the measurement average squared error norm $E[\|y-\widehat{y}\|_2^2] = \sum_{n=1}^{M}\sigma^2_{\widehat{y}[n]} = \frac{3MNP}{T}$. Assuming that we reconstruct the image using

a technique such as BPIC (see Sidebar: Compressive Sampling in a Nutshell), we obtain $E\left[\text{MSE}\right] \leq 3C_N^2 M \frac{P}{T}$.

VI. **Acknowledgements**

Thanks to Dave Brady for suggesting the Poisson noise analysis, to Dennis Healy and Courtney Lane for many enlightening discussions, and to Michael Wakin for his many contributions. This work was supported by the grants DARPA/ONR N66001-06-1-2011 and N00014-06-1-0610, NSF CCF-0431150, ONR N00014-07-1-0936, AFOSR FA9550-07-1-0301, ARO W911NF-07-1-0502, ARO MURI W911NF-07-1-0185, and the Texas Instruments Leadership University Program. Special thanks to TI for providing the DMD developer's kit and accessory light modulator package (ALP).

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Sept. 2006.

[2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory*, vol. 52, pp. 5406–5425, Dec. 2006. Preprint.

[3] E. J. Candès, "Compressive sampling," in *International Congress of Mathematicians*, vol. 3, (Madrid, Spain), pp. 1433–1452, 2006.

[4] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–120, 124, July 2007.

[5] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Computational Imaging IV*, vol. 6065, (San Jose, CA), pp. 43–52, Jan. 2006.

[6] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comp. Harmonic Analysis*, vol. 10, pp. 27–44, 2001.

[7] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, June 2007.

[8] D. Brady, "Multiplex sensors and the constant radiance theorem," *Optics Letters*, vol. 27, no. 1, pp. 16–18, 2002.

[9] P. Sen, B. Chen, G. Garg, S. Marschner, M. Horowitz, M. Levoy, and H. Lensch, "Dual photography," *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 745–755, 2005.

[10] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Picture Coding Symposium*, (Beijing, China), Apr. 2006.

[11] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *IEEE Int. Conf. Image Proc.*, vol. 2, (Thessaloniki, Greece), pp. 1029–1032, Oct. 2001.

[12] 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 *Spectral Imaging: Instrumentation, Applications, and Analysis II*, vol. 4959, pp. 12–22, 2003.

[13] J. Haupt and R. Nowak, "Compressive sampling versus conventional imaging," in *IEEE Int. Conf. Image Proc.*, (Atlanta, GA), pp. 1269–1272, Oct. 2006.

[14] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[15] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[16] R. Constantini and S. Susstrunk, "Virtual sensor design," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V*, vol. 5301, pp. 408–419, Jun. 2004.

[17] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *IEEE Int. Conf. Acoustics, Speech and Signal Proc.*, (Toulouse, France), pp. III–872–875, May 2006.

[18] M. A. Davenport, M. F. Duarte, D. Takhar, J. N. Laska, K. K. Kelly, and R. G. Baraniuk, "The smashed filter: Compressive sensing for signal classification," in *Computational Imaging V*, vol. 6498, (San Jose, CA), Jan. 2007.

[19] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, "Compressive sampling for signal classification," in *Asilomar Conf. Signals, Systems and Computers*, (Pacific Grove, CA), Oct. 2006.

[20] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Wavelets XI*, vol. 5914, (San Diego, CA), August 2005.

[21] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, 2008. To appear.

[22] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, 2008. To appear.

[23] P. M. Lane, R. P. Elliott, and C. E. MacAulay, "Confocal microendoscopy with chromatic sectioning," in *Spectral Imaging: Instrumentation, Applications, and Analysis II*, vol. 4959, pp. 23–26, 2003.

[24] V. Bansal, S. Patel, and P. Saggau, "High-speed confocal laser scanning microscopy using acousto-optic deflectors and a digital micromirror device," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XI*, vol. 5324, pp. 47–54, 2004.

[25] G. L. Davis, M. Maggioni, F. J. Warner, F. B. Geshwind, A. C. Coppi, R. A. DeVerse, and R. R.Coifman, "Hyper-spectral analysis of normal and malignant microarray tissue sections using a novel micro-optoelectrical-mechanical system," *Modern Pathology*, vol. 17 Suppl. 1:358A, 2004.

[26] R. Muise and A. Mahalanobis, "Target detection using integrated hyper spectral sensing and processing." IMA Workshop on Integration of Sensing and Processing, December 2005.

[27] N. P. Pitsianis, D. J. Brady, and X. Sun, "Sensor-layer image compression based on the quantized cosine transform," in *Visual Information Processing XIV*, vol. 5817, (Orlando, FL, USA), p. 250, SPIE, 2005.

[28] D. J. Brady, M. Feldman, N. Pitsianis, J. P. Guo, A. Portnoy, and M. Fiddy, "Compressive optical MONTAGE

photography," in *Photonic Devices and Algorithms for Computing VII*, vol. 5907, (San Diego, CA, USA), pp. 44–50, SPIE, 2005.

[29] J. Ke and M. Neifeld, "Optical architectures for compresssive imaging," *Applied Optics*, vol. 46, pp. 5293–5303, Aug. 2007.

[30] R. Fergus, A. Torralba, and W. T. Freeman, "Random lens imaging," Tech. Rep. MIT-CSAIL-TR-2006-058, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, Sept. 2006.

[31] J. Sampsell, "An overview of the digital micromirror device (DMD) and its application to projection displays," in *SID Int. Symp. Digest of Technical Papers*, vol. 24, p. 1012, 1993.