# A MANIFOLD LIFTING ALGORITHM FOR MULTI-VIEW COMPRESSIVE IMAGING

*Michael B. Wakin*

Division of Engineering, Colorado School of Mines, Golden, CO 80401

## ABSTRACT

We consider a multi-view imaging scenario where a number of cameras observe overlapping, translated subimages of a larger scene. To simplify the acquisition and encoding of these images, we propose a non-collaborative compressive sensing protocol at each camera. We discuss a prototype algorithm for joint reconstruction of the images from the ensemble of random measurements, based on the geometric manifold structure that arises from the varying camera positions. Even when the camera positions are unknown, we demonstrate that it is possible to simultaneously resolve the images and register their positions using only the random measurements.

## 1. INTRODUCTION

To help address the growing challenges of acquiring richer, higher-resolution signals and data sets, a data acquisition protocol known as Compressive Sensing (CS) [1, 2] has recently been proposed, in which only a small number of random, linear measurements need be obtained from each signal. Supposing the signal obeys a concise or sparse model, then from these apparently incomplete measurements, an inverse problem can be solved to recover the full-resolution signal.

CS is particularly useful in two scenarios. The first is when a high-resolution signal is difficult to measure directly. For example, a compressive imaging camera [3] has been proposed that can acquire a digital image using far fewer (random) measurements than the number of pixels in the image. Such a camera can be used not only for imaging at visible wavelengths, but also for imaging at nonvisible wavelengths where conventional imaging hardware can be expensive.

A second scenario where CS is useful is when one or more high-resolution signals may be difficult to encode. Such scenarios arise, for example, in sensor networks and multi-view imaging, where joint, collaborative compression among the sensors would require costly communication. As an alternative, a method known as Distributed CS (DCS) [4] has been proposed, where each sensor encodes only a random set of linear projections of its own observed signal. While the DCS encoding is non-collaborative, the DCS decoding reconstructs all signals jointly to exploit their common structure.

Existing DCS reconstruction algorithms rest on a collection of joint sparsity models, where each signal is assumed sparse in some basis and the sparse coefficient patterns are correlated from signal to signal. However, this is not the only type of joint structure that may arise in multi-signal CS. In this paper we consider a multi-view imaging scenario where a number of cameras observe overlapping, translated subimages of a larger scene. To simplify the acquisition and encoding processes as much as possible, we consider a randomized, non-collaborative measurement protocol for each camera. Because the images under consideration are parameterized by the low-dimensional camera position, it follows that the images live on a low-dimensional submanifold of the ambient signal space. This geometric structure requires novel reconstruction algorithms not treated by the DCS theory.

After introducing the basic problem formulation in Sec. 2, we discuss the geometric manifold perspective in Sec. 3. In Sec. 4, we discuss a prototype "manifold lifting" algorithm for joint reconstruction of the images from the ensemble of random measurements. As we demonstrate in Sec. 5, even when the camera positions are unknown, it is possible to simultaneously resolve the images and register their positions with high accuracy. We conclude in Sec. 6.

## 2. PROBLEM SETUP

To best highlight the concepts in limited space, we will frame our discussion in the context of a specific satellite imaging experiment. Many details can be generalized, and we discuss extensions of the algorithm to other applications in Sec. 6.

Let us consider the following scenario. We wish to acquire a satellite image $x$ such as the $192 \times 192$ image shown in Fig. 1(a). We suppose this image corresponds to 1 square unit of land area. This image will be observed by a collection of 200 satellites, with limited but overlapping fields of view (.44 square units of land area each; see Fig. 1(b)), and with limited resolution ($64 \times 64$ pixel images). For each $j \in \{1, 2, \dots, 200\}$, we let $s_j \in \mathbb{R}^{64\times 64}$ denote the image acquired by camera $j$, and let $p_j = [p_j^V, p_j^H]^T \in \mathbb{R}^2$ denote the vertical and horizontal translation[1] of camera $j$ relative to the center of the desired image $x$. Letting $R_{p_j}$ denote the linear restriction operator that restricts $x$ to a limited field of view (at the camera position $p_j$) and reduces the resolution to $64 \times 64$, we have $s_j = R_{p_j} x$ for each $j \in \{1, 2, \dots, 200\}$.

For a central collection point, given the ensemble of low-resolution satellite images $s_1, s_2, \dots, s_{200}$, and supposing the relative positions $p_1, p_2, \dots, p_{200}$ of the 200 satellites were

Email: mwakin@mines.edu. This research was partially supported by DARPA Grant HR0011-08-1-0078, and portions of this work were performed during the author's affiliation with The University of Michigan.

[1] For this experiment, we ensured that the 200 cameras covered every pixel in $x$, which required that one camera was perfectly situated in each corner of the image.

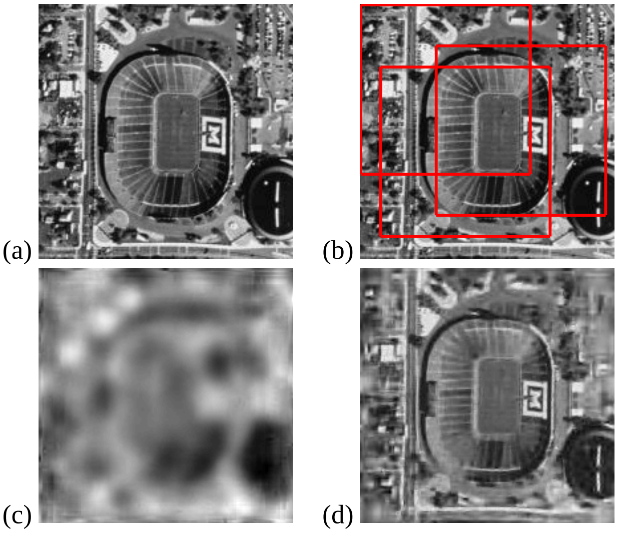

**Fig. 1**. *(a) Original desired image $x$; courtesy USGS. (b) Satellite images with limited field of view. (c) Result of image-by-image reconstruction from random measurements, followed by fusion with exact knowledge of camera positions, PSNR 15.4dB. (d) Result from transform coding of each image, PSNR 22.3dB. Both approaches are inferior to joint recovery using random measurements, see Fig. 4(b).*

known, it would be straightforward to combine this data into a superresolution estimate $\widehat{x}$ of the desired image $x$. In terms of the burden on the satellite sensing and communication systems, this strategy would require each satellite to measure and transmit $64 \cdot 64 = 4096$ numbers describing its image.

However, it may also be the case that the communication bandwidth from each satellite is limited, and so we may wish to minimize the data that it must transmit. (Also, if imaging at a nonvisible wavelength, we may wish to minimize the number of measurements each sensor must take.) As such, we propose that each satellite could measure and report a limited number (say, $96$) of random measurements that summarize its own incident image. In CS notation, we say that each sensor $j$ transmits $y_j = \Phi_j s_j$, where $\Phi_j$ is a measurement matrix of size $96 \times 4096$; for simplicity we take all $\Phi_j = \Phi$ for some fixed $\Phi$. For the random measurements, we use the multiscale noiselet transform [5]; example measurement functions from each of 5 scales are shown in Fig. 2. We take 16 measurements at the coarsest scale and 20 at each finer scale; this enables the coarse-to-fine recovery method described in Sec. 4.

Although the measurement and encoding processes are completely non-collaborative across the multiple satellites, we will see that this protocol makes very efficient use of the available bandwidth. In fact, the ultimate quality of reconstruction we achieve using 96 random measurements per sensor is surprisingly *better* than what would be achieved by image-by-image transform coding (that is, measuring all 4096 pixels of each image, computing the 2D wavelet transform, and transmitting only the 96 largest wavelet coefficients).

Unfortunately, recovering the signals $s_j$ from the random measurements and ultimately recovering an estimate $\widehat{x}$ of the desired image $x$ now becomes more complicated. We have dramatically undersampled each $s_j$; for conventional CS reconstruction on an image-by-image basis, we have far too few measurements to be useful. Figure 1(c) shows an estimate $\widehat{x}$ obtained using standard $\ell_1$-based CS recovery of each image $s_j$, followed by averaging at the correct camera positions.

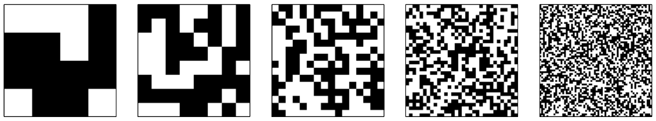

**Fig. 2**. *Size $64 \times 64$ noiselet basis functions used for collecting random measurements at a sequence of 5 scales. Each measurement is the inner product of one such function with the image of interest.*

What is needed is a *joint* recovery algorithm for this multi-signal CS problem, i.e., a method for consolidating all of the measurement vectors $y_j$ and then recovering all of the $s_j$ jointly. To further complicate matters, we will assume that the relative satellite positions $p_j$ are actually *unknown*. As we will see, the correlated information contained in the signals $s_j$ can make joint reconstruction possible with far greater accuracy than image-by-image reconstruction, even when the camera positions $p_j$ are unknown.

## 3. GEOMETRIC PERSPECTIVE

As discussed in Sec. 1, prior work in formulating joint recovery algorithms for multi-signal CS [4] utilized joint sparsity models to capture the inter-signal correlations. However, such models are not equipped to capture the joint structure that arises due to the geometry of the present problem.

Instead, we may note that, supposing $x$ is fixed, there are only two degrees of freedom describing any image $s_j$ — these are captured in the camera position vector $p_j$, which describes the 2D offset relative to the center of the image. That is, for any $p \in \mathbb{R}^2$, there corresponds an image $R_p x \in \mathbb{R}^{4096}$, and as $p$ changes the resulting image $R_p x$ changes as a continuous function of $p$. Considering the set of all possible images that can result from all $p \in \mathbb{R}^2$, we define $\mathcal{M} = \{R_p x : p \in \mathbb{R}^2\}$ which corresponds to a nonlinear 2D surface within $\mathbb{R}^{4096}$, also known as a *submanifold* of $\mathbb{R}^{4096}$ [6]. It follows that the $200$ images $s_j$ are simply points drawn from $\mathcal{M}$.

Consequently, the $200$ measurement vectors $y_j$ live along the projection of this manifold within $\mathbb{R}^{96}$, namely $\Phi\mathcal{M} = \{\Phi R_p x : p \in \mathbb{R}^2\} \subset \mathbb{R}^{96}$. (As we have recently shown [7], the geometric structure of signal manifolds can actually be well preserved when projected onto randomized, low-dimensional subspaces.) Thus, the problem of recovering the images $s_j$ from the measurements $y_j$ corresponds to a "manifold lifting" problem: We wish to *lift* each image from its measurement vector in $\mathbb{R}^{96}$ back to its original position in $\mathbb{R}^{4096}$, but meanwhile we wish to preserve the 2D *manifold* structure that relates all of these images together. This is in contrast to many traditional manifold *learning* algorithms, where the objective is to construct a

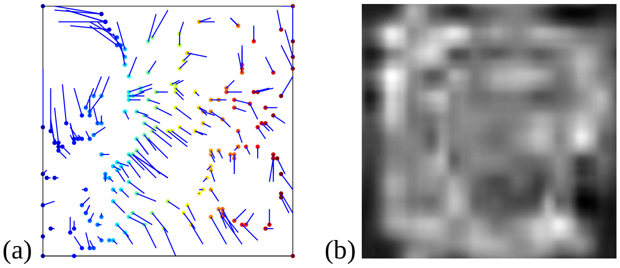

**Fig. 3**. *(a) Initial estimates of camera positions obtained from ISOMAP. (b) Initial estimate $\widehat{x}$ obtained from joint recovery informed by initial estimates of camera positions.*

lower-dimensional embedding of a dataset sampled from a manifold in high-dimensional space; however, as described below, we use one manifold learning known as ISOMAP [8] as an intermediate step in our manifold lifting algorithm.

## 4. MANIFOLD LIFTING ALGORITHM

### 4.1. Initial estimates of satellite positions

We begin with the problem of using the measurement vectors $y_j$ to obtain an initial estimate for the position $p_j$ of each satellite. For this we invoke the manifold perspective: the 200 vectors $y_j$ live along a 2D manifold in $\mathbb{R}^{96}$ as described above, but in addition the camera positions $p_j = [p_j^V, p_j^H]^T$ give a rough coordinate system for the relative positions of the measurement vectors along this 2D manifold.

Algorithms such as ISOMAP [8] are well-equipped for this type of problem. Using ISOMAP, we may pass as input the collection of our measurement vectors $y_1, y_2, \dots, y_{200} \in \mathbb{R}^{96}$, and request as output an embedding $y_1^e, y_2^e, \dots, y_{200}^e \in \mathbb{R}^2$ of the points in 2D Euclidean space that best preserves geodesic distances between those points.

The resulting map of these points $y_j^e$ in 2D is a rough estimate $\widehat{p}_j$ of the relative camera positions. Figure 3(a) shows the result we obtain by passing our data to the ISOMAP algorithm,[2] followed by postprocessing to correctly rotate the embedded coordinates. The colored points represent the estimated camera positions, while the blue vectors represent the error with respect to the true (but unknown) camera position.

### 4.2. Initial estimates of satellite images

Once we have obtained a rough estimate $\widehat{p}_j$ of each camera position, we may proceed to find an initial estimate $\widehat{s}_j$ for each satellite image in $\mathbb{R}^{4096}$. While ensuring consistency with the observed measurements, i.e., that $y_j \approx \Phi \widehat{s}_j$, we wish to ensure that these estimates live along a 2D manifold describing a common high-resolution image $\widehat{x} \in \mathbb{R}^{192 \times 192}$, and that the relative positions along this manifold should be roughly consistent with the estimated camera positions $\widehat{p}_j$. In other words, for some $\widehat{x}$, we wish to ensure that $\widehat{s}_j \approx R_{\widehat{p}_j} \widehat{x}$.

Our proposed approach is to directly estimate $\widehat{x}$. We concatenate all of the measurement vectors and operators, letting

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{200} \end{bmatrix}, \widehat{R} = \begin{bmatrix} R_{\widehat{p}_1} \\ R_{\widehat{p}_2} \\ \vdots \\ R_{\widehat{p}_{200}} \end{bmatrix}, \Phi_{\text{big}} = \begin{bmatrix} \Phi & 0 & \cdots & 0 \\ 0 & \Phi & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \Phi \end{bmatrix}.$$

At this point, we may approach the reconstruction as a single-signal CS problem, for example by searching for the image $\widehat{x}$ that is the most sparse in the wavelet domain while maintaining consistency with the observed measurements $y$, i.e., $y \approx \Phi_{\text{big}} \widehat{R} \widehat{x}$. We formulate the reconstruction problem as

$$\widehat{\alpha} = \text{argmin}_{\alpha} \|\alpha\|_1 \text{ subject to } \|y - \Phi_{\text{big}} \widehat{R} \Psi \alpha\|_2 \le \epsilon, \quad (1)$$

where $\Psi$ is a wavelet basis and $\epsilon$ is chosen to reflect the uncertainty in the camera positions $p_j$.[3] Given $\widehat{\alpha}$, we then let $\widehat{x} = \Psi \widehat{\alpha}$, and then let the estimates $\widehat{s}_j = R_{\widehat{p}_j} \widehat{x}$. Figure 3(b) shows the initial reconstruction $\widehat{x}$.

### 4.3. Iterations

Once the estimate $\widehat{x}$ has been obtained, it is possible to return to the question of the camera positions and improve the estimates $\widehat{p}_j$. The estimate $\widehat{x}$ corresponds to a 2D manifold of possible observation vectors in $\mathbb{R}^{96}$, and by registering the measurement vectors $y_j$ with respect to this manifold, we obtain improved estimates. We let $\widehat{p}_j = \arg\min_p \|y_j - \Phi R_p \widehat{x}\|_2$, where the minimization is performed over a local neighborhood of the previous estimate $\widehat{p}_j$.

With the improved estimates $\widehat{p}_j$, it is then possible to refine the estimates $\widehat{s}_j$ as described above in Sec. 4.2. We may then iterate between the two refinement steps until convergence or until reaching a designated stopping criterion.

### 4.4. Multiscale refinements

There is one detail we have omitted to this point, which is that this iterative refinement procedure is best performed in a multiscale, coarse-to-fine fashion. For example, in the ISOMAP estimates described in Fig. 3(a), we do not pass all 96 measurements for each image but rather only the 36 noiselet measurements from the 2 coarsest scales. The reason is that these manifolds have a multiscale structure [6], and at coarse scales they are most flat and most amenable to manifold learning algorithms. Additionally, we use only the 36 noiselet measurements from the 2 coarsest scales for our first estimate of $\widehat{x}$ in Sec. 4.2. As our estimates of camera positions become more accurate, however, we may tolerate more twisting in the manifold, and so we can bring in finer scales of measurements,

[2] As described below in Sec. 4.4, we initially do not pass all 96 measurements as input to ISOMAP, but use only the 36 noiselet measurements from the 2 coarsest scales. This is equivalent to measuring low-pass regularized images, for which the articulation manifold will be more smooth.

[3] In our experiments, we chose the parameter $\epsilon$ as somewhat of an oracle, in particular as $1.25\|y - \Phi_{\text{big}} \widehat{R} x\|_2$ to determine the error that would result if we measured the true image $x$ but with the wrong positions as denoted by $\widehat{R}$. This process should be made more robust in future work.

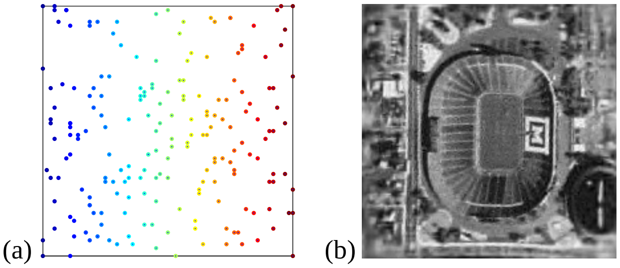

**Fig. 4**. *(a) Final estimated camera positions. (b) Final estimated image $\widehat{x}$ from our manifold lifting algorithm, PSNR 23.8dB.*

which in turn produce better estimates of $x$ and the camera positions, and so on. We incorporate our multiscale measurements into the iterative method described above in Sec. 4.3.

## 5. RESULTS

Figures 4(a),(b) show the estimated camera positions $\widehat{p}_j$ and image $\widehat{x}$ after 10 total iterations where all 5 noiselet scales have gradually been introduced into the reconstruction process. We see that the camera positions have in fact been perfectly estimated, and the reconstructed image quality is far superior to that afforded by image-by-image reconstruction in Fig. 1(c). We attribute this to the strong correlations between the images $s_j$, which our algorithm seeks to exploit.

It is worth emphasizing the apparent benefit of random measurements that we are observing in our experiments. Suppose for each satellite that we were to permit a form of "transform coding", where instead of transmitting 96 random measurements, we request each satellite to transmit its 96 largest wavelet coefficients. On an image-by-image basis, this typically is a *better* way to compress images to minimize PSNR. However, we see in Fig. 1(d) that even if we have perfect knowledge of the camera positions and fuse the available wavelet-based measurements at the collection point using a global $\ell_1$ minimization akin to (1), the ultimate reconstruction is *inferior* to our result in Fig. 4(b). We believe the reason is that the images $s_j$ are highly correlated, and the repeated encoding of large wavelet coefficients (which tend to concentrate at coarse scales) results in the repeated encoding of redundant information across the multiple satellites. For example, transform coding keeps, on average, only 6 coefficients per image $s_j$ at the finest wavelet scale; in contrast, we collect 20 noiselet measurements at each of the finest scales and fewer at the coarsest scale. Thus, random measurements enable more diverse and high frequency information to reach the collection point.

## 6. CONCLUSIONS

We have presented a promising validation of geometric approaches for use in multi-signal CS recovery while highlighting the impressive ability of random measurements to capture single- and multi-signal structure without actively looking for it. However, several important questions remain topics of active research. These include an analysis of the requisite number of measurements (or bits, if quantized) for reconstruction, an understanding of the accuracy required in camera position estimates in order to ensure convergence, and the development of a robust method for choosing $\epsilon$ in (1). With the current simulation parameters, we have found the convergence of the algorithm to be sensitive to the starting camera positions. Also of interest would be to reduce the computational complexity, as solving (1) requires a considerable amount of computation for large $x$.

We note that the algorithm can be simplified in the special case where the camera positions are known in advance. One may, for example, proceed directly to the global recovery program (1). Our initial ISOMAP step is the reason why a large number of camera views are required and also why we choose $\Phi_j = \Phi$ for all $j$. Without the need for this step, one may be able to substantially reduce the number of camera views and also introduce more diversity into the random measurements.

Ultimately, we believe this type of algorithm will be extensible beyond satellite imaging to other multi-view problems such as molecular imaging, light field imaging, etc. The most immediate extensions will be in problems where a linear operator relates the information of interest (in our case, $x$) to the measurements ($y$) as is the case in our global planar translation model where we can write $y_j = \Phi_j R_{p_j} x$; this allows the immediate formulation of an optimization problem such as (1). Algorithms suitable for nonlinear mappings, for example where a multi-camera array captures a 3D scene from a closer distance or from several different angles, remain under development; see [9] for promising preliminary work.

## 7. REFERENCES

[1] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, Feb. 2006.

[3] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[4] D. Daron, M. B. Wakin, M. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Rice University Technical Report TREE-0612, Nov 2006.

[5] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmon. Anal.*, vol. 10, pp. 27–44, 2001.

[6] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, August 2005, SPIE.

[7] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, pp. 51–77, Feb 2009.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, December 2000.

[9] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. Picture Coding Symposium (PCS)*, Chicago, Illinois, May 2009.