EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

A geometric approach to multi-view compressive imaging

Jae Young Park[1*] and Michael B Wakin[2]

Abstract

In this paper, we consider multi-view imaging problems in which an ensemble of cameras collect images describing a common scene. To simplify the acquisition and encoding of these images, we study the effectiveness of non-collaborative compressive sensing encoding schemes wherein each sensor directly and independently compresses its image using randomized measurements. After these measurements and also perhaps the camera positions are transmitted to a central node, the key to an accurate reconstruction is to fully exploit the joint correlation among the signal ensemble. To capture such correlations, we propose a geometric modeling framework in which the image ensemble is treated as a sampling of points from a low-dimensional manifold in the ambient signal space. Building on results that guarantee stable embeddings of manifolds under random measurements, we propose a "manifold lifting" algorithm for recovering the ensemble that can operate even without knowledge of the camera positions. We divide our discussion into two scenarios, the near-field and far-field cases, and describe how the manifold lifting algorithm could be applied to these scenarios. At the end of this paper, we present an in-depth case study of a far-field imaging scenario, where the aim is to reconstruct an ensemble of satellite images taken from different positions with limited but overlapping fields of view. In this case study, we demonstrate the impressive power of random measurements to capture single- and multi-image structure without explicitly searching for it, as the randomized measurement encoding in conjunction with the proposed manifold lifting algorithm can even outperform image-by-image transform coding.

1. Introduction

Armed with potentially limited communication and computational resources, designers of distributed imaging systems face increasing challenges in the quest to acquire, compress, and communicate ever richer and higher-resolution image ensembles. In this paper, we consider multi-view imaging problems in which an ensemble of cameras collect images describing a common scene. To simplify the acquisition and encoding of these images, we study the effectiveness of non-collaborative Compressive Sensing (CS) [1,2] encoding schemes wherein each sensor directly and independently compresses its image using a small number of randomized measurements (see Figure 1). CS is commonly intended for the encoding of a single signal, and a rich theory has been developed for signal recovery from incomplete measurements by exploiting the assumption that the signal obeys a sparse model. In this paper, we address the problem of how to recover an ensemble of images from a collection of image-by-image random measurements. To do this, we advocate the use of implicitly geometric models to capture the joint structure among the images.

CS is particularly useful in two scenarios. The first is when a high-resolution signal is difficult to measure directly. For example, conventional infrared cameras require expensive sensors, and with increasing resolution such cameras can become extremely costly. A compressive imaging camera has been proposed [3] that can acquire a digital image using far fewer (random) measurements than the number of pixels in the image. Such a camera is simple and inexpensive and can be used not only for imaging at visible wavelengths, but also for imaging at non-visible wavelengths.

A second scenario where CS is useful is when one or more high-resolution signals are difficult or expensive to encode. Such scenarios arise, for example, in sensor networks and multi-view imaging, where it may be feasible to measure the raw data at each sensor, but joint, collaborative

* Correspondence: jaeypark@umich.edu
[1]Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI, USA
Full list of author information is available at the end of the article

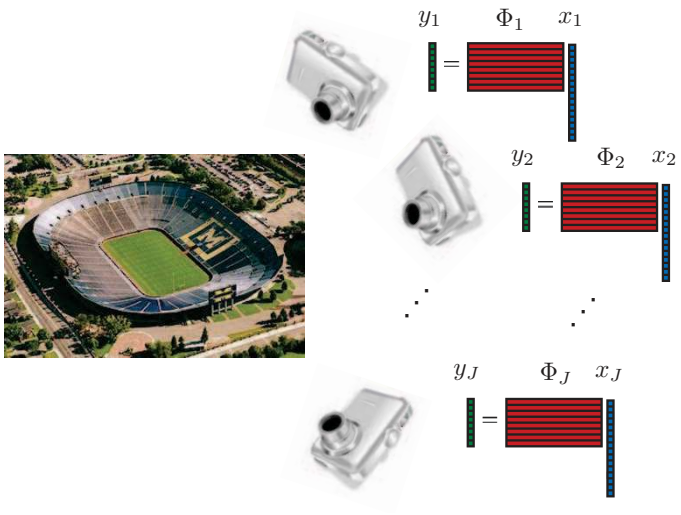


Figure 1 Multi-view compressive imaging setup. A common scene is observed by J cameras from different positions. Each camera j encodes a small number of random measurements y_j of its observed image x_j, and a single decoder jointly reconstructs all images $\{x_j\}$ from the ensemble of compressive measurements $\{y_j\}$.

compression of that data among the sensors would require costly communication. As an alternative to conventional Distributed Source Coding (DSC) methods [4], an extension of single-signal CS known as Distributed CS (DCS) [5] has been proposed, where each sensor encodes only a random set of linear projections of its own observed signal. These projections could be obtained either by using CS hardware as described above, or by using a random, compressive encoding of the data collected from a conventional sensor.

While DCS encoding is non-collaborative, an effective DCS decoder should reconstruct all signals *jointly* to exploit their common structure. As we later discuss, most existing DCS algorithms for distributed imaging reconstruction rely fundamentally on sparse models to capture intra- and inter-signal correlations [5-8]. What is missing from each of these algorithms, however, is an assurance that the reconstructed images have a global consistency, i.e. that they all describe a common underlying scene. This may not only lead to possible confusion in interpreting the images, but more critically may also suggest that the reconstruction algorithm is failing to completely exploit the joint structure of the ensemble.

To better extend DCS techniques specifically to problems involving multi-view imaging, we propose in this paper a general geometric framework in which many such reconstruction problems may be cast. We specifically focus on scenarios where a representation of the underlying scene is linearly related to the observations. This is mainly for simplicity, and there is plenty of room for the development of joint reconstruction algorithms given nonlinear mappings; however, we present a number of scenarios where a linear mapping can be found. For these problems, we explain how viewing the unknown images as living along a low-dimensional manifold within the high-dimensional signal space can inform the design of effective joint reconstruction algorithms. Such algorithms can build on existing sparsity-based techniques for CS but ensure a global consistency among the reconstructed images. We refine our discussion by focusing on two settings: far-field and near-field multi-view imaging. Finally, as a proof of concept, we demonstrate a "manifold lifting" algorithm in a specific far-field multi-view scenario where the camera positions are not known a priori and we only observe a small number of random measurements at each sensor. Even in such discouraging circumstances, by effectively exploiting the geometrical information preserved in the manifold model, we are able to accurately reconstruct both the underlying scene and the camera positions.

2. Background on signal models and compressive sensing

A. Concise signal models

Real-world signals typically contain some degree of structure that can be exploited to simplify their processing and recovery. *Sparsity* is one model of conciseness in which the signal of interest can be represented as a linear combination of only a few basis vectors from some dictionary. To provide a more formal statement, let us consider a signal $x \in \mathbb{R}^N$. (If the signal is a 2D image, we reshape it into a length-N vector.) We let $\Psi \in \mathbb{R}^{N \times N}$ denote an orthonormal basis[a] for $\mathbb{R}^N$, with its columns acting as basis vectors, and we write $x = \Psi\alpha$, where $\alpha := \Psi^T x \in \mathbb{R}^N$ denotes the expansion coefficients of x in the

basis Ψ. We say that x is K-sparse in the basis Ψ if α contains only K nonzero entries. Sparse representations with $K \ll N$ provide exact or approximate models for wide varieties of signal classes, as long as the basis Ψ is chosen to match the structure in x. In the case of images, the 2D Discrete Wavelet Transform (DWT) and 2D Discrete Cosine Transform (DCT) are reasonable candidates for Ψ [9].

As an alternative to sparsity, *manifolds* have also been used to capture the concise structure of multi-signal ensembles [10-14]. Simply put, we can view a manifold as a low-dimensional nonlinear surface within $\mathbb{R}^N$. Manifold models arise, for example, in settings where a low-dimensional parameter controls the generation of the signal (see Figure 2). Assume, for instance, that $x = x_\theta \in \mathbb{R}^N$ depends on some parameter θ, which belongs to a p-dimensional parameter space[b] that we call Θ. One might imagine photographing some static scene and letting θ correspond to the position of the camera: for every value of θ, there is some N-pixel image x_θ that the camera will see. Supposing that the mapping $\theta \to x_\theta$ is well-behaved, then if we consider all possible signals that can be generated by all possible values of θ, the resulting set $\mathcal{M} := \{x_\theta : \theta \in \Theta\} \subset \mathbb{R}^N$ will in general correspond to a nonlinear p-dimensional surface within $\mathbb{R}^N$.

When the underlying signal x is an image, the resulting manifold $\mathcal{M}$ is called an *Image Appearance Manifold* (IAM). Recently, several important properties of IAMs have been revealed. For example, if the images x_θ contain sharp edges that move as a function of θ, the IAM is *nowhere differentiable* with respect to θ [12]. This poses difficulties for gradient-based parameter estimation techniques such as Newton's method because the tangent planes on the manifold (onto which one may wish to project) do not exist. However, it has also been shown that IAMs have a multiscale tangent structure [12,13] that is accessible through a sequence of regularizations of the image, as shown in Figure 3. In particular, suppose we define a spatial blurring kernel (such as a lowpass filter) denoted by h_s, where $s > 0$ indicates the *scale* (e.g., the bandwidth or the cutoff frequency) of the filter. Then, although $\mathcal{M} := \{x_\theta : \theta \in \Theta\}$ will not be differentiable, the manifold $\mathcal{M}_s = \{h_s * x_\theta : \theta \in \Theta\}$ of regularized images will be differentiable, where * denotes 2D convolution. Tangent planes do exist on these regularized manifolds $\mathcal{M}_s$, and as $s \to 0$, the orientation of these tangent planes along a given $\mathcal{M}_s$ changes more slowly as a function of θ. In the past, we have used this multiscale tangent structure to implement a coarse-to-fine Newton method for parameter estimation on IAMs [13].

The rich geometrical information that rests within an IAM makes it an excellent candidate for modeling in multi-view imaging. Letting θ represent camera position, all of the images in a multi-view ensemble will live along a common IAM, and as we will later discuss, image reconstruction in the IAM framework can ensure global consistency of the reconstructed images.

B. Compressive sensing

In conventional signal acquisition devices such as digital cameras and camcorders, we first acquire a full N-dimensional signal x and then apply a compression technique such as JPEG or MPEG [9]. These and other *transform coding* techniques essentially involve computing the expansion coefficients α describing the signal in some basis Ψ, keeping only the K-largest entries of α, and setting the rest to zero. While this can be a very effective way of consolidating the signal information, one could argue that this procedure of "first sample, then compress" is somewhat wasteful because we must measure N pieces of information only to retain $K < N$ coefficients. For certain sensing modalities (such as infrared), it may be difficult or expensive to acquire so many high-resolution samples of the signal.

The recently emerged theory of CS suggests an alternative acquisition scheme. CS utilizes an efficient

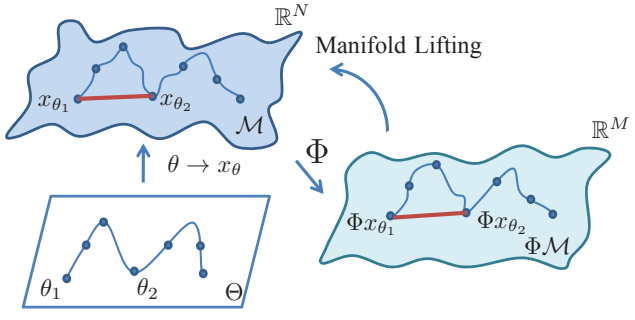


Figure 2 A manifold $\mathcal{M}$ can be viewed as a nonlinear surface in $\mathbb{R}^N$. When the mapping between θ and x_θ is well-behaved, as we trace out a path in the parameter space Θ, we trace out a similar path on $\mathcal{M}$. A random projection Φ from $\mathbb{R}^N$ to a lower dimensional space $\mathbb{R}^M$ can provide a stable embedding of $\mathcal{M}$, preserving all pairwise distances, and therefore preserving the structure within an ensemble of images. The goal of a manifold lifting algorithm is to recover an ensemble of images from their low-dimensional measurements.

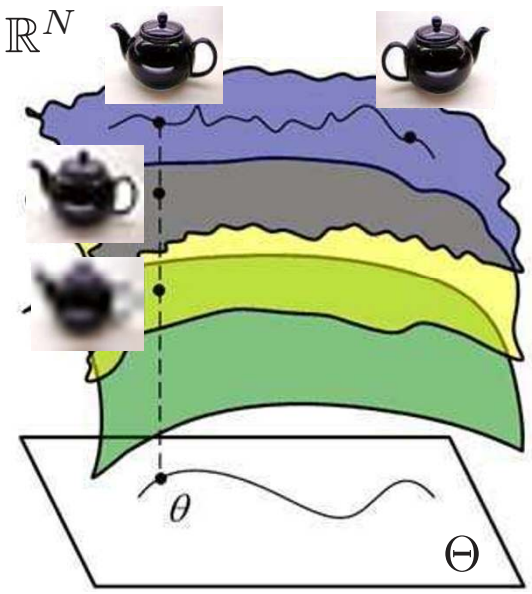


Figure 3 The multiscale structure of manifolds. The top manifold in this figure corresponds to the collection of images of a teapot that could be acquired from different camera positions θ. While manifolds like this containing images with sharp edges are not differentiable, manifolds of images containing smooth images *are* differentiable, and the more one smoothes the images, the smoother the manifold becomes.

encoding framework in which we directly acquire a compressed representation of the underlying signal by computing simple linear inner products with a small set of randomly generated test functions. Let us denote the full-resolution discrete signal as $x \in \mathbb{R}^N$ and suppose that we generate a collection of M random vectors, $\varphi_i \in \mathbb{R}^N$, $i = 1, 2, \ldots, M$. We stack these vectors into an $M \times N$ matrix $\Phi = [\varphi_1\ \varphi_2\ \ldots\ \varphi_M]^T$, which we refer to as a measurement matrix. A CS encoder or sensor produces the measurements $y = \Phi x \in \mathbb{R}^M$, possibly without ever sampling or storing x itself.

At the decoder, given the random measurements y and the measurement matrix Φ, one must attempt to recover x. The canonical approach in CS is to assume that x is sparse in a known basis Ψ and solve an optimization problem of the form [1,2]

$$\min_{\alpha} ||\alpha||_1 \text{s.t.}\, y = \Phi\Psi\alpha, \tag{1}$$

which can be recast as a linear program. When there is bounded noise or uncertainty in the measurements, i.e. $y = \Phi x + n$ with $||n||_2 \le \varepsilon$, it is common to solve a similar problem [15]:

$$\min_{\alpha} ||\alpha||_1 \text{s.t.}\, ||y - \Phi\Psi\alpha||_2 \le \varepsilon, \tag{2}$$

which is again convex and can be solved efficiently.

Depending on the measurement matrix Φ, recovery of sparse signals can be provably accurate, even in noise. One condition on Φ that has been used to establish recovery bounds is known as the Restricted Isometry Property (RIP) [16], which requires that pairwise distances between sparse signals be approximately preserved in the measurement space. In particular, a matrix Φ is said to satisfy the RIP of order $2K$ with respect to Ψ if there exists a constant $0 < \delta_{2K} < 1$ such that for all K-sparse vectors x_1, x_2 in the basis Ψ the following is satisfied,

$$(1 - \delta_{2K})||x_1 - x_2||_2^2 \le ||\Phi x_1 - \Phi x_2||_2^2 \le (1 + \delta_{2K})||x_1 - x_2||_2^2. \tag{3}$$

If Φ satisfies the RIP of order $2K$ with δ_{2K} sufficiently small, it is known that (1) will perfectly recover any K-sparse signal in the basis Ψ and that (2) will incur a recovery error at worst proportional to ε [15]. The performance of both recovery techniques also degrades gracefully if x is not exactly K-sparse but rather is well approximated by a K-sparse signal.

It has been shown that we can obtain an RIP matrix Φ with high probability simply by taking $M = \mathcal{O}(K\log(N/K))$ and populating the matrix with i.i.d. Gaussian, Bernoulli, or more general subgaussian entries [17]. Thus, one of the hallmarks of CS is that this requisite number of measurements M is essentially proportional to the sparsity level K of the signal to be recovered.

In addition to families of K-sparse signals, random matrices can also provide stable embeddings for manifolds (see Figure 2). Letting M denote a smooth[c] p-dimensional manifold, if we take $M = \mathcal{O}(p\log(N))$ and generate Φ randomly from one of the distributions above, we will obtain an embedding $\Phi\mathcal{M} := \{\Phi x : x \in \mathcal{M}\} \in \mathbb{R}^M$ such that all pairwise distances between points on the manifold are approximately preserved [14], i.e. such that (3) holds for all $x_{\theta_1}, x_{\theta_2} \in \mathcal{M}$. Geodesic distances are also approximately preserved. Again, the requisite number of measurements is merely proportional to the information level of the signal, which in this case equals p (the dimension of the manifold), rather than the sparsity level of the signal in any particular dictionary. All of this suggests that manifolds may be viable models to use in CS recovery; see [18] for additional discussion on the topic of using manifold models to recover individual signals.

We see from the above that random measurements have a remarkable "universal" ability to capture the key information in a signal, and this occurs with a number of measurements just proportional to the number of degrees of freedom in the signal. Only the decoder attempts to exploit the signal structure, and it can do so by positing any number of possible signal models.

In summary, in settings where a high-resolution signal x is difficult or expensive to measure directly, CS allows

us to replace the "first sample, then compress" paradigm with a technique for directly acquiring compressive measurements of x. To do this in practice, we might resort to CS hardware that directly acquires the linear measurements y without ever sampling or storing x directly. Several forms of compressive imaging architectures have been proposed, ranging from existing data collection schemes in Magnetic Resonance Imaging (MRI) [19] to more exotic CS-based techniques. One architecture [3], for example, replaces the conventional CCD/CMOS sensor in a digital camera with a digital micromirror device (DMD), which modulates the incoming light and reflects it onto a single photodiode for measurement. Some intriguing uses of this inexpensive "single pixel camera" could include infrared or hyperspectral imaging, where conventional high-resolution sensors can cost hundreds of thousands of dollars.

Before proceeding, however, we note that CS can also be useful in settings where it is possible to *acquire* high-resolution signals, but is difficult or expensive to subsequently *encode* them. For example, x might represent a video signal, for which direct measurement is possible, but for which subsequent compression typically requires exploiting complicated spatio-temporal correlations [7,8]. A more straightforward encoder might simply compute $y = \Phi x$ for some random, compressive Φ. Other scenarios where data are difficult to encode efficiently might be in sensor networks or in multi-view imaging, which is the topic of this paper and is discussed further in the next section.

3. Problem setup and related work

A. Multi-view imaging using image-by-image random measurements

Let us now turn to the problem of distributed image compression for multi-view imaging. We imagine an ensemble of J distinct cameras that collect images $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$ describing a common scene, with each image x_j taken from some camera position $\theta_j \in \Theta$. We would like to efficiently compress this ensemble of images, but as in any sensor network, we may be limited in battery power, computational horsepower, and/or communication bandwidth. Thus, although we may be able to posit sparse and manifold-based models for concisely capturing the intra- and inter-signal structures among the images in the ensemble, directly exploiting these models for the purpose of data compression may be prohibitively complex or require expensive collaboration among the sensors. This motivates our desire for an effective distributed encoding strategy.

The encoding of multiple signals in distributed scenarios has long been studied under the auspices of the distributed source coding (DSC) community. The Slepian-Wolf framework [4] for lossless DSC states that two sources X_1 and X_2 are able to compress at their conditional entropy rate without collaboration and can be decoded successfully when the correlation model (i.e., the joint probability distribution $p(x_1, x_2)$) is known at the decoder. This work was extended to lossy coding by Wyner and Ziv when side information is available at the decoder [20], and in subsequent years, practical algorithms for these frameworks have been proposed based on channel coding techniques. However, one faces difficulties in applying these frameworks to multi-view imaging because the inter-image correlations are arguably better described geometrically than statistically. Several algorithms (e.g., [21-23]) have been proposed for combining these geometric and statistical frameworks, but fully integrating these concepts remains a very challenging problem.

As a simple alternative to these type of encoding schemes, we advocate the use of CS for distributed image coding, wherein for each sensor $j \in \{1, 2, \ldots, J\}$, the signal $x_j \in \mathbb{R}^N$ is independently encoded using an $M_j \times N$ measurement matrix Φ_j, yielding the measurement vector $y_j = \Phi_j x_j \in \mathbb{R}^{M_j}$. Such an encoding scheme is known in the CS literature as Distributed CS (DCS) [5]. While the primary motivation for DCS is to simplify the *encoding* of correlated high-resolution signals, one may of course bypass the potentially difficult *acquisition* of the high-resolution signals and directly collect the random measurements using CS hardware.

After the randomized encoding, the measurement vectors $y_1, y_2, \ldots, y_J$ are then transmitted to a central node for decoding. Indeed, DCS differs from single-signal CS only in the decoding process. Rather than recover the signals one-by-one from the measurement vectors, an effective DCS decoder should solve a joint reconstruction problem, exploiting the intra- and inter-signal correlations among the signals $\{x_j\}$, while ensuring consistency with the measurements $\{y_j\}$.

The proper design of a DCS decoder depends very much on the type of data being collected and on the nature of the intra- and inter-signal correlations. Ideally, compared to signal-by-signal recovery, joint recovery should provide better reconstruction quality from a given set of measurement vectors, or equivalently, reduce the measurement burden needed to achieve a given reconstruction quality. For example, if each signal in the ensemble is K-sparse, we may hope to jointly recover the ensemble using fewer than the $\mathcal{O}(K \log(N/K))$ measurements per sensor that are required to reconstruct the signals separately. Like single-signal CS, DCS decoding schemes should be robust to noise and to dropped measurement packets. Joint reconstruction techniques should also be robust to the loss of individual sensors, making DCS well-suited for remote sensing applications.

B. Current approaches to DCS multi-view image reconstruction

For signals in general and images in particular, a variety of DCS decoding algorithms have been proposed to date. Fundamentally, all of these frameworks build upon the concept of sparsity for capturing intra- and inter-signal correlations.

One DCS modeling framework involves a collection of joint sparsity models (JSMs) [5]. In a typical JSM, we represent each signal $x_j \in \mathbb{R}^N$ in terms of a decomposition $x_j = z_C + z_j$, where $z_C \in \mathbb{R}^N$ is a "common component" that is assumed to be present in all $\{x_j\}$, and $z_j \in \mathbb{R}^N$ is an "innovation component" that differs for each signal. Depending on the application, different sparsity assumptions may be imposed on z_C and z_j. In some cases, these assumptions can dramatically restrict the space of possible signals. For example, all signals may be restricted to live within the same K-dimensional subspace. The DCS decoder then searches for a signal ensemble that is consistent with the available measurements and falls within the space of signals permitted by the JSM. For signal ensembles well modeled by a JSM, DCS reconstruction can offer a significant savings in the measurement rates. While each sensor must take enough measurements to account for its innovation component z_j, all sensors can share the burden of measuring the common component z_C.

Unfortunately, the applicability of JSMs to multi-view imaging scenarios can be quite limited. While two cameras in very close proximity may yield images having sparse innovations relative to a common background, any significant difference in the camera positions will dramatically increase the complexity of the innovation components. Because conventional JSMs are not appropriate for capturing any residual correlation that may remain among these innovations, we would expect JSM-based recovery to offer very little improvement over independent CS recovery.

Recently, a significant extension of the JSM framework has been proposed specifically for multi-view compressive imaging [6]. This framework assumes that images of a common scene are related by local or global geometrical transformations and proposes an overcomplete dictionary of basis elements consisting of various geometrical transformations of a generating mother function. It is assumed that each image can be decomposed into its own subset of these atoms plus the geometrically transformed atoms of the neighboring images. The benefit of this approach is that information about one image helps reduce the uncertainty about which atoms should be used to comprise the neighboring images. Unfortunately, there seems to be a limit as to how much efficiency may be gained from such an approach. To reconstruct a given image, the decoder may be tasked with solving for, say, K sparse coefficients. While the correlation model may help reduce the measurement burden at that sensor below $\mathcal{O}(K\log(N/K))$, it is not possible to reduce the number of measurements below K. As we will later argue, however, there is reason to believe that alternative reconstruction techniques based on the underlying scene (rather than the images themselves) can succeed with even fewer than K measurements.

Other approaches for multi-view image reconstruction could draw naturally from recent work in CS video reconstruction by ordering the static images $\{x_j\}$ according to their camera positions and reconstructing the sequence as a sort of "fly-by" video. One approach for video reconstruction exploits the sparsity of inter-frame differences [7]. For multi-view imaging, this would correspond to a difference image $x_i - x_j$ having a sparse representation in some basis Ψ. Again, however, this condition may only be met if cameras i and j have very close proximity. We have also proposed a CS video reconstruction technique based on a motion-compensated temporal wavelet transform [8]. For multi-view imaging, we could modify this algorithm, replacing block-based motion compensation with disparity compensation. The challenge of such an approach, however, would be in finding the disparity information without prior knowledge of the images themselves. For video, we have addressed this challenge using a coarse-to-fine reconstruction algorithm that alternates between estimating the motion vectors and reconstructing successively higher resolution versions of the video using the motion-compensated wavelet transform.

What would still be missing from any of these approaches, however, is an assurance that the reconstructed images have a global consistency, i.e. that they all describe a common underlying scene. In the language of manifolds, this means that the reconstructed images do not necessarily live on a common IAM defined by a hypothetical underlying scene. This may not only lead to possible confusion in interpreting the images, but more critically may also suggest that the reconstruction algorithm is failing to completely exploit the joint structure of the ensemble-the images are in fact constrained to live in a much lower-dimensional set than the algorithm realizes.

4. Manifold lifting techniques for multi-view image reconstruction

In light of the above observations, one could argue that an effective multi-view reconstruction algorithm should exploit the underlying geometry of the scene by using an inter-signal modeling framework that ensures global consistency. To inform the design of such an algorithm, we find it helpful to view the general task of reconstruction as what we term a *manifold lifting* problem: we would like to recover each image $x_j \in \mathbb{R}^N$ from its

measurements $y_j \in \mathbb{R}^{M_j}$ ("lifting" it from the low-dimensional measurement space back to the high-dimensional signal space), while ensuring that all recovered images live along a common IAM.

Although this interpretation does not immediately point us to a general purpose recovery algorithm (and different multi-view scenarios could indeed require markedly different algorithms), it can be informative for a number of reasons. For example, as we have discussed in Section 2-B, manifolds can have stable embeddings under random projections. If we suppose that $\Phi_j = \Phi \in \mathbb{R}^{M\times N}$ for all j, then each measurement vector we obtain will be a point sampled from the embedded manifold $\Phi\mathcal{M} \subset \mathbb{R}^M$. From samples of $\Phi\mathcal{M}$ in $\mathbb{R}^M$, we would like to recover samples of (or perhaps all of) $\mathcal{M}$ in $\mathbb{R}^N$, and this may be facilitated if $\Phi\mathcal{M}$ preserves the original geometric structure of $\mathcal{M}$. In addition, as we have discussed in Section 2-A, many IAMs have a multiscale structure that has proved useful in solving non-compressive parameter estimation problems, and this structure may also be useful in solving multi-view recovery problems.

While this manifold-based interpretation may give us a geometric framework for signal modeling, it may not in isolation sufficiently capture all intra- and inter-signal correlations. Indeed, one cannot disregard the role that concise models such as sparsity may still play in an effective manifold lifting algorithm. Given an ensemble of measurements $y_1, y_2, \ldots, y_J$, there may be many candidates IAMs on which the original images $x_1, x_2, \ldots, x_J$ may live. In order to resolve this ambiguity, one could employ either a model for the intra-signal structure (such as sparsity) or a model for the underlying structure of the scene (again, possibly sparsity). To do the latter, one must develop a representation for the underlying scene or phenomenon that is being measured and understand the mapping between that representation and the measurements $y_1, y_2, \ldots, y_J$. To keep the problem simple, this mapping will ideally be linear, and as we discuss in this section, such a representation and linear mapping can be found in a number of scenarios.

To make things more concrete, we demonstrate in this section how the manifold lifting viewpoint can inform the design of reconstruction algorithms in the context of two generic multi-view scenarios: far-field and near-field imaging. We also discuss how to address complications that can arise due to uncertainties in the camera positions. We hope that such discussions will pave the way for the future development of broader classes of manifold lifting algorithms.

A. Far-field multi-view imaging

We begin by considering the case where the cameras are far from the underlying scene, such as might occur in satellite imaging or unmanned aerial vehicle (UAV) remote sensing scenarios. In problems such as these, it may be reasonable to model each image $x_j \in \mathbb{R}^N$ as being a translated, rotated, scaled subimage of a larger fixed image. We represent this larger image as an element x drawn from a vector space such as $\mathbb{R}^Q$ with $Q > N$, and we represent the mapping from x to x_j (which depends on the camera position θ_j) as a linear operator that we denote as $R_{\theta_j} : \mathbb{R}^Q \to \mathbb{R}^N$. This operator R_{θ_j} can be designed to incorporate different combinations of translation, rotation, scaling, etc., followed by a restriction that limits the field of view.

This formulation makes clear the dependence of the IAM $\mathcal{M}$ on the underlying scene x: $\mathcal{M} = \mathcal{M}(x) = \{R_\theta x : \theta \in \Theta\} \subset \mathbb{R}^N$. Supposing we believe x to obey a sparse model and supposing the camera positions are known, this formulation also facilitates a joint recovery program that can ensure global consistency while exploiting the structure of the underlying scene. At camera j, we have the measurements $y_j = \Phi_j x_j = \Phi_j R_{\theta_j} x$. Therefore, by concatenating all of the measurements, we can write the overall system of equations as $y = \Phi_{\text{big}} R x$, where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \; R = \begin{bmatrix} R_{\theta_1} \\ R_{\theta_2} \\ \vdots \\ R_{\theta_J} \end{bmatrix}, \text{ and } \Phi_{\text{big}} = \begin{bmatrix} \Phi_1 & 0 & \cdots & 0 \\ 0 & \Phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_J \end{bmatrix}. \tag{4}$$

Given y and $\Phi_{\text{big}} R$, and assuming x is sparse in some basis Ψ (such as the 2D wavelet domain), we can solve the usual optimization problem as stated in (1) (or (2) if the measurements are noisy). If desired, one can use the recovered image $\hat{x}$ to obtain estimates $\hat{x}_j : R_{\theta_j}\hat{x}$ of the original subimages. These are guaranteed to live along a common IAM, namely $\mathcal{M}(\hat{x})$.

B. Near-field multi-view imaging

Near-field imaging may generally be more challenging than far-field imaging. Defining a useful representation for the underlying scene may be difficult, and due to effects such as parallax and occlusions, it may seem impossible to find a linear mapping from any such representation to the measurements. Fortunately, however, there are encouraging precedents that one could follow.

One representative application of near-field imaging is in Computed Tomography (CT). In CT, we seek to acquire a 3D volumetric signal x, but the signals x_j that we observe correspond to slices of the Fourier transform of x. (We may assume $y_j = x_j$ in such problems, and so the challenge is actually to recover $\mathcal{M}(x)$, or equivalently just x, rather than the individual $\{x_j\}$.) Given a fixed viewing angle θ_j, this relationship between x and x_j is

linear, and so we may set up a joint recovery program akin to that proposed above for far-field imaging. Similar approaches have been used for joint recovery from undersampled frequency measurements in MRI [19].

For near-field imaging using visible light, there is generally no clear linear mapping between a 3D volumetric representation of the scene and the observed images x_j. However, rather than contend with complicated nonlinear mappings, we suggest that a promising alternative may be to use the *plenoptic function* [24] as a centralized representation of the scene. The plenoptic function f is a hypothetical 5D function used to describe the intensity of light that could be observed from any point in space, when viewed in any possible direction. The value $f(p_x, p_y, p_z, p_\theta, p_\varphi)$ specifies the light intensity that would be measured by a sensor located at the position (p_x, p_y, p_z) and pointing in the direction specified by the spherical coordinates p_θ, and p_φ. (Additional parameters such as color channel can be considered.) By considering only a bounded set of viewing positions, the plenoptic function reduces to a 4D function known as the *lumigraph* [24].

Any image $x_j \in \mathbb{R}^N$ of the scene has a clear relationship to the plenoptic function. A given camera j will be positioned at a specific point (p_x, p_y, p_z) in space and record light intensities arriving from a variety of directions. Therefore, x_j simply corresponds to a 2D "slice" of the plenoptic function, and once the camera viewpoint θ_j is fixed, the mapping from f to x_j is a simple linear restriction operator. Consequently, the structure of the IAM $\mathcal{M} = \mathcal{M}(f)$ is completely determined by the plenoptic function.

Plenoptic functions contain a rich geometric structure that we suggest could be exploited to develop sparse models for use in joint recovery algorithms. This geometric structure arises due to the geometry of objects in the scene: when a physical object having distinct edges is photographed from a variety of perspectives, the resulting lumigraph will have perpetuating geometric structures that encode the shape of the object under study. As a simple illustration, a Flatland-like scenario (imaging an object in the plane using 1D cameras) is shown in Figure 4a. The resulting 2D lumigraph is shown in Figure 4b, where each row corresponds to a single "image". We see that geometric structures in the lumigraph arise due to shifts in the object's position as the camera viewpoint changes. For the 4D lumigraph these structures have recently been termed "plenoptic manifolds" [25] due to their own nonlinear, surface-like characteristics. If a sparse representation for plenoptic functions can be developed that exploits these geometric constraints, then it may be possible to recover plentopic functions from incomplete, random measurements using a linear problem formulation and recovery algorithms such as (1) or (2). One possible avenue to developing such a sparse representation could involve parameterizing local patches of the lumigraph using the wedgelet [26] or surflet [27] dictionaries. Wedgelets (see Figure 4c) can be tiled together to form piecewise linear approximations to geometric features; surflets offer piecewise polynomial approximations.

As a proof of concept, we present a simple experiment in support of this approach. For the lumigraph shown in Figure 4b, which has $J = 128$ 1D "images" that each contain $N = 128$ pixels, we collect $M = 5$ random measurements from each image. From these measurements, we attempt to reconstruct the entire lumigraph using wedgelets [27] following a multiscale technique outlined in Chapter 6 of [28]. The reconstructed lumigraph is shown in Figure 4d and is relatively accurate despite the small number of measurements.

Finally, to illustrate the rich interplay between geometry within the lumigraph and the underlying geometry of the scene, we show that it is actually possible to use the reconstructed lumigraph to estimate the underlying scene geometry. While we omit the precise details of our approach, the estimated wedgelets help us to infer three pieces of information: the positions of each local wedgelet patch in the v and t directions indicate a camera position and viewing direction, respectively, while the orientation of the wedgelet indicates a depth at which a point in the scene belongs to the object. Putting these estimates together, we obtain the reconstruction of the scene geometry shown in Figure 4e. This promising proof of concept suggests that wedgelets or surflets could indeed play an important role in the future for developing improved concise models for lumigraph processing.

C. Dealing with uncertainties in camera positions

In all of our discussions above, we have assumed the camera positions θ_j were known. In some situations, however, we may have only noisy estimates $\widehat{\theta}_j = \theta_j + n_j$ of the camera positions. Supposing that we can define linear mappings between the underlying scene and the images x_j, it is straightforward to extend the CS recovery problem to account for this uncertainty. In particular, letting R denote the concatenation of the mappings R_{θ_j} as in (4), and letting $\widehat{R}$ denote the concatenation of the mappings $R_{\widehat{\theta}_j}$ corresponding to the noisy camera positions, it follows that $y = \Phi_{big} R x = \Phi_{big} \widehat{R} x + n$ for some noise vector n, and so (2) can be used to obtain an approximation $\hat{x}$ of the underlying scene. Of course, the accuracy of this approximation will depend on the quality of the camera position estimates.

When faced with significant uncertainty about the camera positions, the multiscale properties of IAMs help us to conceive of a possible coarse-to-fine reconstruction approach. As in Section 2-A, let h_s denote a

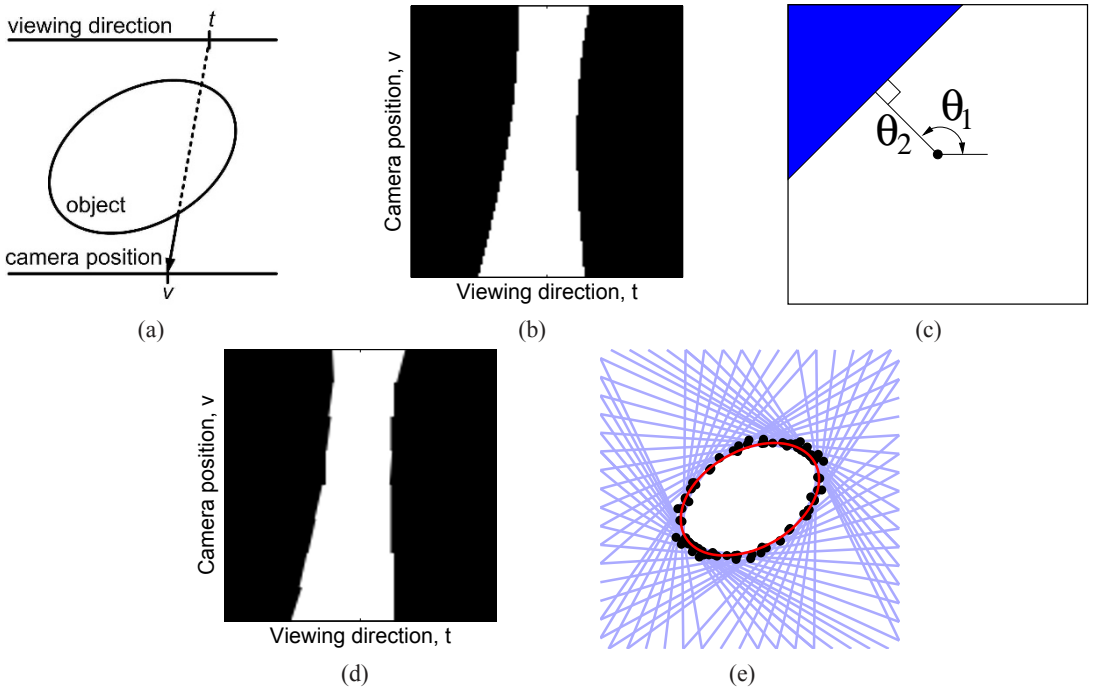


Figure 4 Lumigraph geometry in compressive multi-view imaging. **a** Flatland-like illustration for collecting 1D images of an object in the 2D plane. At each camera position, all viewing directions may be considered. **b** Resulting 128 × 128 lumigraph for the ellipse-shaped object. Each row corresponds to a single "image". (In the real world, each image is 2D and the full lumigraph is 4D.) The lumigraph can be repeated for viewing from all four sides of the object. **c** Wedgelets provide a simple parametric model for local patches of a 2D lumigraph; only two parameters are needed to describe the orientation and offset of the linear discontinuity. **d** Wedgelet-based lumigraph reconstruction from $M = 5$ compressive samples of each image (row of lumigraph). **e** Scene geometry estimated using local edge positions/orientations in the reconstructed lumigraph. Each blue line connects an estimated point on the object to a camera from which that point is visible. The true ellipse is shown in red.

blurring kernel at scale s and suppose for simplicity that $\Theta = \mathbb{R}$. Based on the arguments presented in [13], it follows that for most reasonable mappings $\theta \to x_\theta$, we will have $||\frac{\partial(h_s * x_\theta)}{\partial_\theta}||_2 \to 0$ as $s \to 0$. What this implies is that, on manifolds of regularized images $\mathcal{M}_s = \{h_s * x_\theta : \theta \in \Theta\}$, the images will change slowly as a function of camera position, and so we can ensure that $h_s * (R_{\hat{\theta}_j} x)$ is arbitrarily close to $h_s * (R_{\theta_j} x)$ by choosing s sufficiently small (a sufficiently "coarse" scale). Now, suppose that some elements of each y_j are devoted to measuring $h_s * x_j = h_s * (R_{\theta_j} x)$. We denote these measurements by $y_{j,\,s} = \Phi_{j,\,s}(h_s * x_j)$. In practice, we may replace the convolution operator with a matrix H_s and collect $y_{j,s} = \Phi_{j,s} H_s x_j = \Phi_{j,s} H_s R_{\theta_j} x$ instead. Concatenating all of the $\{y_{j,s}\}_{j=1}^{J}$, we may then use the noisy position estimates to define operators $\{R_{\hat{\theta}_j}\}$ and solve (2) as above to obtain an estimate $\hat{x}$ of the scene. This estimate will typically correspond to a lowpass filtered version of x, since for many reasonable imaging models, we will have $h_s * (R_{\theta_j} x) \approx R_{\theta_j}(h'_s * x)$ for some lowpass filter h'_s, and this implies that $y_{j,s} \approx \Phi_{j,s} R_{\theta_j}(h'_s * x)$ contains only low frequency information about x.

Given this estimate, we may then re-estimate the camera positions by projecting the measurement vectors $y_{j,\,s}$ onto the manifold $\mathcal{M}(\hat{x})$. (This may be accomplished, for example, using the parameter estimation techniques described in [13].) Then, having improved the camera position estimates, we may reconstruct a finer scale (larger s) approximation to the true images $\{x_j\}$, and so on, alternating between the steps of estimating camera positions and reconstructing successively finer scale approximations to the true images. This multiscale, iterative algorithm requires the sort of multiscale randomized measurements we describe above, namely $y_{j,\,s} = \Phi_{j,\,s}(h_s * x_j)$ for a sequence of scales s. In practice, the noiselet transform [29] offers one fast technique for implementing these measurement operators $\Phi_{j,\,s} H_s$ at a sequence of scales. Noiselet scales are also nested, so measurements at a scale s_1 can be re-used as measurements at any scale $s_2 > s_1$.

The manifold viewpoint can also be quite useful in situations where the camera positions are completely

unknown, as they might be in applications such as cryo-electron microscopy (Cryo-EM) [30]. Because we anticipate that an IAM $\mathcal{M}$ will have a stable embedding $\Phi\mathcal{M}$ in the measurement space, it follows that the relative arrangement of the points $\{x_j\}$ on $\mathcal{M}$ will be preserved in $\Phi\mathcal{M}$. Since this relative arrangement will typically reflect the relative arrangement of the values $\{\theta_j\}$ in Θ, we may apply to the compressive measurements[d] any number of "manifold learning" techniques (such as ISOMAP [11]) that are designed to discover such parameterizations from unlabeled data. An algorithm such as ISOMAP will provide an embedding of J points in $\mathbb{R}^p$ whose relative positions can be used to infer the relative camera positions; a similar approach has been developed specifically for the Cryo-EM problem [30]. (Some side information may be helpful at this point to convert these relative position estimates into absolute position estimates.) Once we have these estimates, we may resort to the iterative refinement scheme described above, alternating between the steps of estimating camera positions and reconstructing successively finer scale approximations to the true images.

5. Manifold lifting case study

A. Problem setup

As a proof of concept, we now present a comprehensive multi-view reconstruction algorithm inspired by the manifold lifting viewpoint. We do this in the context of a far-field imaging simulation in which we wish to reconstruct a Q-pixel high-resolution image x of a large scene. Information about this scene will be acquired using an ensemble of J satellites, which will collect N-pixel photographs x_j of the scene from different positions and with limited but overlapping fields of view, as illustrated with red boxes in Figure 5a.

We denote the vertical and horizontal position of satellite j by $\theta_j = (\theta_j^V, \theta_j^H) \in \mathbb{R}^2$. The satellite positions take real values and are chosen randomly except for the caveats that the fields of view all must fall within the square support of x and that each of the four corners of x must be seen by at least one camera. (These assumptions are for convenience but can be relaxed without major modifications to the recovery algorithm.) We let R_{θ_j} denote the $N \times Q$ linear operator that maps x to the image x_j. This operator involves a resampling of x to account for the real-valued position vector θ_j, a restriction of the field of view, and a spatial lowpass filtering and decimation, as we assume that x_j has lower resolution (larger pixel size) than x.

In order to reduce data transmission burdens, we suppose that each satellite encodes a random set of measurements $y_j = \Phi_j x_j \in \mathbb{R}^{M_j}$ of its incident image x_j. Following the discussion in 4-C, these random measurements are collected at a sequence of coarse-to-fine scales $s_1, s_2, \ldots, s_T$ using noiselets. (The noiselet measurements can actually be collected using CS imaging hardware [3], bypassing the need for a conventional N-pixel sensor.) We concatenate all of the measurement vectors $\{y_{j,s_i}\}_{i=1}^{T}$ into the length-M_j measurement vector $y_j = \Phi_j x_j$. Finally, we assume that all satellites use the same set of measurement functions, and so we define $M := M_1 = M_2 = M_J$ and $\Phi := \Phi_1 = \Phi_2 = \ldots = \Phi_J$.

Our decoder will be presented with the ensemble of the measurement vectors $y_1, y_2, \ldots, y_J$ but *will not* be given any information about the camera positions (save for an awareness of the two caveats mentioned above) and will be tasked with the challenge of recovering the underlying scene x. Although it would be interesting to consider quantization in the measurements, it is beyond

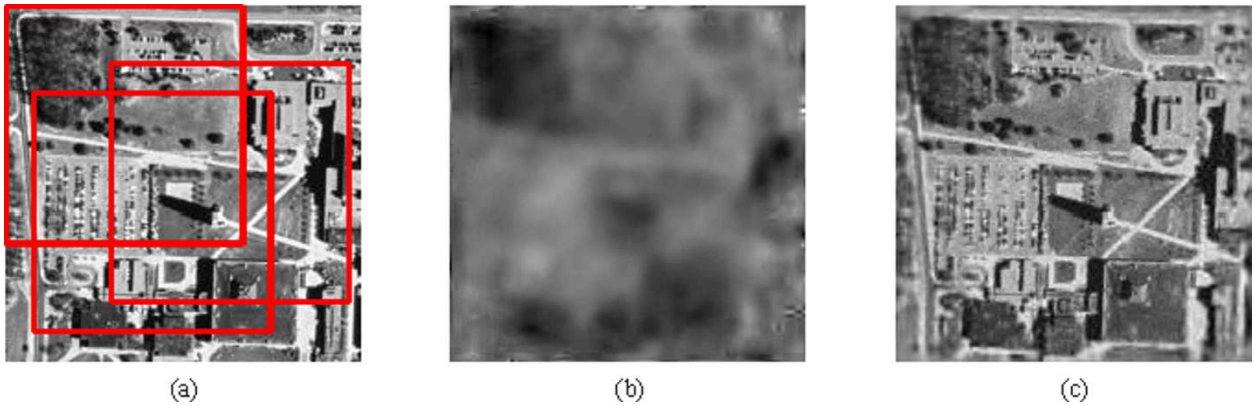


Figure 5 Manifold lifting demonstration. **a** Setup. The original image x (courtesy USGS) has size 192 × 192 and is observed by $J = 200$ satellites. The red boxes illustrate the limited field of view for a few such cameras. **b** Image-by-image reconstruction from random measurements, PSNR 14.4 dB. **c** Joint reconstruction using our manifold lifting algorithm with unknown camera positions, PSNR 23.6 dB.

the scope of this paper and we did not implement any quantization steps in the following simulations.

B. Manifold lifting algorithm

We combine the discussions provided in Sections 4-A and 4-C to design a manifold lifting algorithm that is specifically tailored to this problem.

1) Initial estimates of satellite positions

The algorithm begins by obtaining a preliminary estimate of the camera positions. To do this, we extract from each y_j the measurements corresponding to the two or three coarsest scales (i.e., y_{j,s_1}, y_{j,s_2} and possibly y_{j,s_3}), concatenate these into one vector, and pass the ensemble of such vectors (for all $j \in \{1, 2, \ldots, J\}$) to the ISOMAP algorithm. ISOMAP then delivers an embedding of points $v_1, v_2, \ldots, v_J$ in $\mathbb{R}^2$ that best preserves pairwise geodesic distances compared to the input points; an example ISOMAP embedding is shown in Figure 6a. What can be inferred from this embedding are the relative camera positions; a small amount of side information is required to determine the proper scaling, rotation, and (possible) reflection of these points to correctly align them with an absolute coordinate system. Assuming that we know the correct vertical and horizontal reflections, after reflecting these camera positions correctly, we then rotate and scale them to fill the square support of x.

2) Iterations

Given the initial estimates $\{\widehat{\theta}_j\}$ of our camera positions, we can then define the operators $\{R_{\widehat{\theta}_j}\}$ and consequently $\widehat{R}$. By concatenating the measurement vectors and measurement matrices, initially only those at the coarsest scale (i.e., y_{j,s_1} across all j), we write the overall system of equations as $y = \Phi\widehat{R}x + n$ as in Section 4-A, and solve for

$$\hat{\alpha} = \arg\min_{\alpha} ||\alpha||_1 \text{ subject to } ||y - \Phi_{\text{big}}\widehat{R}x\Psi\alpha||_2 \leq \varepsilon,$$

where Ψ is a wavelet basis and ε is chosen[e] to reflect the uncertainty in the camera positions θ_j. Given $\hat{\alpha}$, we can then compute the corresponding estimate of the underlying scene as $\hat{x} = \Psi\hat{\alpha}$.

After we obtain the estimate $\hat{x}$, we refine the camera positions by registering the measurement vectors y_j with respect to this manifold. In other words, we solve the following optimization problem:

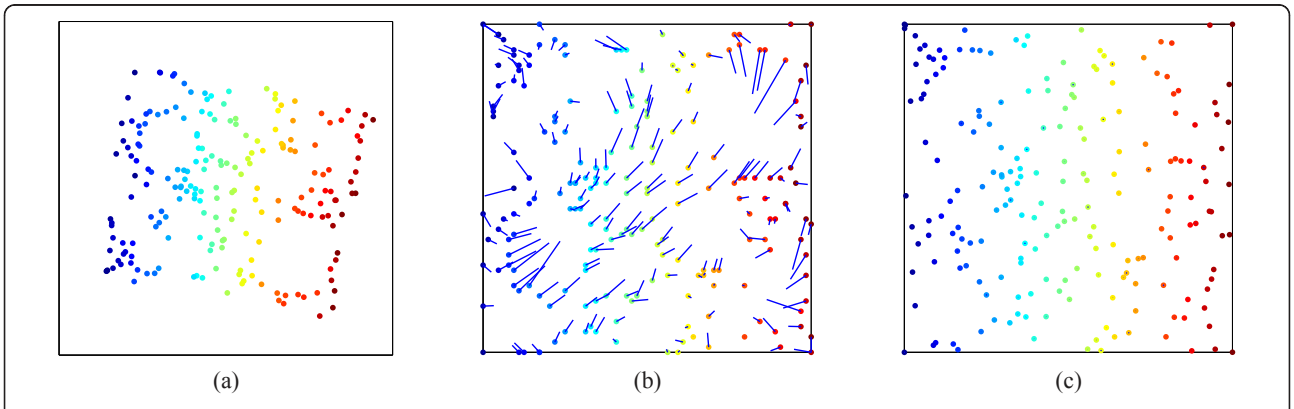
$$\hat{\theta}_j = \arg\min_{\theta} ||y_j - \Phi R_\theta \hat{x}||_2,$$

where again in each y_j we use only the coarse scale measurements. To solve this problem, we use the multiscale Newton algorithm proposed in [13].

With the improved estimates $\hat{\theta}_j$, we may then refine our estimate of $\hat{x}$ but can do so by incorporating finer scale measurements. We alternate between the steps of reconstructing the scene $\hat{x}$ and re-estimating the camera positions $\hat{\theta}_j$, successively bringing in the measurements $y_{j,s_2}, y_{j,s_3}, \ldots, y_{j,s_T}$. (At each scale, it may help to alternate once or twice between the two estimation steps before bringing in the next finer scale of measurements. One can also repeat until convergence or until reaching a designated stopping criterion.) Finally, having brought in all of the measurements, we obtain our final estimate $\hat{x}$ of the underlying scene.

3) Experiments

We run our simulations on an underlying image x of size $Q = 192 \times 192$ that is shown in Figure 5a. We suppose that x corresponds to 1 square unit of land area. We observe this scene using $J = 200$ randomly positioned cameras, each with a limited field of view.



(a) (b) (c)

Figure 6 Camera position estimates in manifold lifting demonstration. a Initial ISOMAP embedding $v_1, v_2, \ldots, v_J$ of the measurement vectors. **b** Initial estimates $\{\widehat{\theta}_j\}$ of camera positions after rotating and scaling the $\{v_j\}$. **c** Final camera position estimates after running the manifold lifting algorithm. In **b** and **c**, the colored points represent the estimated camera positions (color coded by the true θ_j^H value), while the blue vectors represent the error with respect to the true (but unknown) camera position.

Relative to x, each field of view is of size 128 × 128, corresponding to 0.44 square units of land area as indicated by the red boxes in Figure 5a. Within each field of view, we observe an image x_j of size $N = 64 \times 64$ pixels that has half the resolution (twice the pixel size) compared to x. The total number of noiselet scales for an image of this size is 6. For each image, we disregard the coarsest noiselet scale and set $s_1, s_2, \ldots, s_5$ corresponding to the five finest noiselet scales. For each image, we collect 96 random noiselet measurements: 16 at scale s_1, and 20 at each of the scales $s_2, \ldots, s_5$. Across all scales and all cameras, we collect a total of $96 \cdot 200 = 19,200 \approx 0.52Q$ measurements.

Based on the coarse scale measurements, we obtain the ISOMAP embedding $v_1, v_2, \ldots, v_J$ shown in Figure 6a. After rotating and scaling these points, the initial estimates $\{\widehat{\theta_j}\}$ of camera positions are shown in Figure 6b. These initial position estimates have a mean absolute error of 1.8 and 2.0 pixels (relative to the resolution of x) in the vertical and horizontal directions, respectively. Figure 6c shows the final estimated camera positions after all iterations of our manifold lifting algorithm. These estimates have a mean absolute error of 0.0108 and 0.0132 pixels in the vertical and horizontal directions, respectively. The final reconstruction $\hat{x}$ obtained using these estimated camera positions is shown in Figure 5c. We note that the border areas are not as accurately reconstructed as the center region because fewer total measurements are collected near the borders of x. The scale-by-scale progression of the reconstruction of x and the estimated camera positions are shown in Figure 7. Figure 7a shows the reconstructed images of x at each scale $s_1, s_2, \ldots, s_5$, where the left most image is the reconstruction at the coarsest scale s_1 and the right most image is the reconstructed image at the finest scale s_5. Figure 7b shows the corresponding camera position estimates that were used in the reconstruction of the images in Figure 7a. As we have mentioned above, it can help to alternate between reconstruction of the image and estimation of the camera positions at the same scale more than once before moving on to the next finer scale. In this particular simulation, we have alternated between reconstruction and camera position estimation 3 to 4 times at each scale but the finest and 6 times at the finest scale.

In order to assess the effectiveness of our algorithm, we compare it to three different reconstruction methods. In all of these methods, we assume that the exact camera positions are known and we keep the total number of measurements fixed to 19,200. First, we compare to image-by-image CS recovery, in which we reconstruct the images x_j independently from their random measurements y_j and then superimpose and average them at the correct positions. As expected, and as shown in Figure 5b, this does not yield a reasonable reconstruction because there is far too little data collected (just 96 measurements) about any individual image to reconstruct it in isolation. Thus, we see the dramatic benefits of joint recovery.

Second, for the sake of completeness, we compare to a non-distributed encoding scheme in which one measures the entire image x using a fully populated $19,200 \times N$

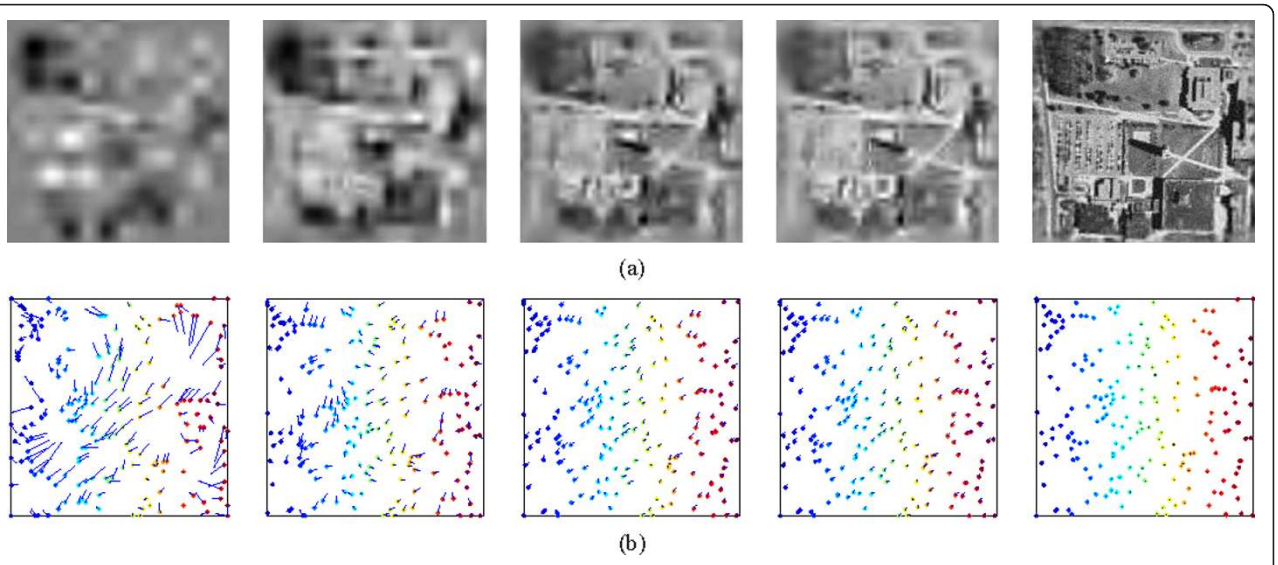


Figure 7 Reconstruction results in manifold lifting demonstration. **a** Scale-by-scale reconstruction of the underlying image proceeding from the coarsest scale s_1 on the left to the finest scale s_5 on the right. **b** The corresponding camera position estimates used in the reconstruction of the images in **a** proceeding from the coarsest scale s_1 on the left to the finest scale s_5 on the right.

Gaussian random matrix. Figure 8a shows the reconstructed image obtained using a single invocation of ℓ_1-minimization. Perhaps surprisingly, the reconstruction quality is actually inferior to that obtained using the manifold lifting algorithm with distributed measurements (shown in Figure 8c). This is somewhat counterintuitive since one would expect that the spatially limited measurement functions would have inferior isometry properties compared to global measurement functions. Although we do not have a concrete theoretical explanation for this phenomenon, we believe that this difference in reconstruction quality is mainly due to the multiscale nature of the measurement functions employed in our manifold lifting example. To support this argument with an experiment, we run the manifold lifting algorithm with spatially limited but non-multiscale measurement functions: for each window, we measure a total of 96 noiselet measurements at the finest scale only, where previously these 96 measurements were spread across several scales. In this case, the reconstructed image has a PSNR of 19.8 dB, which is worse than that obtained using a global Gaussian measurement matrix. This is consistent with our intuition that, when using measurements with limited spatial support, one could pay a penalty in terms of reconstruction quality.

Third, we compare to another alternative encoding scheme, where rather than encode 96 random noiselet measurements of each image, we encode the 96 largest wavelet coefficients of the image in the Haar wavelet basis. (We choose Haar due to its similarity with the noiselet basis, but the performance is similar using other wavelet bases.) This is a rough approximation for how a non-CS transform coder might encode the image, and for the encoding of a single image in isolation, this is typically a more efficient encoding strategy than using random measurements. (Recall that for reconstructing a single signal, one must encode about $K \log(N/K)$ random measurements to obtain an approximation comparable to K-term transform coding.) However, when we concatenate the ensemble of encoded wavelet coefficients and solve (1) to estimate $\hat{x}$, we see from the result in Figure 8b that the reconstructed image has lower quality than that we obtained using a manifold lifting algorithm based on random measurements, even though the camera positions were *unknown* for the manifold lifting experiment. In a sense, by using joint decoding, we have reduced the CS overmeasuring factor from its familiar value of $\log(N/K)$ down to something below 1! We believe this occurs primarily because the images $\{x_j\}$ are highly correlated, and the repeated encoding of large wavelet coefficients (which tend to concentrate at coarse scales) results in repeated encoding of redundant information across the multiple satellites. In other words, it is highly likely that prominent features will be encoded by many satellites over and over again, whereas other features may not be encoded at all. As a result, by examining Figure 8b, we see that strong features such as streets and the edges of buildings (which have large wavelet coefficients) are relatively more accurately reconstructed than, for example, forests or cars in parking lots (which have smaller wavelet coefficients). Random measurements capture more diverse information within and among the images. To more clearly illustrate the specific benefit that random measurements provide over transform coding (for which the camera positions were known), we show in Figure 8c a reconstruction obtained using random measurements with known camera positions.

Finally, we carry out a series of simulations with the same image x using different numbers J of camera positions. We keep the total number of measurements (19,200) and the sizes of the subimages (64 × 64)

Figure 8 Comparative reconstructions for manifold lifting demonstration. **a** Reconstruction using fully dense Gaussian random matrix, PSNR 21.9 dB. **b** Joint reconstruction using transform coding measurements with known camera positions, PSNR 22.8 dB. **c** Joint reconstruction using random measurements with known camera positions, PSNR 24.7 dB.

constant. The results are summarized in Table 1. In all cases, our manifold lifting algorithm without knowledge of the camera positions outperforms transform coding with knowledge of the camera positions. We do note that as J decreases, the performance of transform coding improves. This is likely because each satellite now has more measurements to devote to encoding information about the underlying scene, and there are fewer total cameras to encode redundant information.

6. Discussion and conclusion

In summary, we have discussed in this paper how non-collaborative CS measurement schemes can be used to simplify the acquisition and encoding of multi-image ensembles. We have presented a geometric framework in which many multi-view imaging problems may be cast and explained how this framework can inform the design of effective manifold lifting algorithms for joint reconstruction. We conclude with a few remarks concerning practical and theoretical aspects of the manifold lifting framework.

First, let us briefly discuss the process of learning camera positions when they are initially completely unknown. In our satellite experiments, we have observed that the accuracy of the ISOMAP embedding depends on the relative size of the subimages x_j to the underlying scene x, with larger subimages leading us to higher quality embeddings. As the size of the subimages decreases, we need more and more camera positions to get a reasonable embedding, and we can reach a point where even thousands of camera positions are insufficient. In such cases, and in applications not limited to satellite imaging, it may be possible to get a reliable embedding by grouping local camera positions together. On a different note, once an initial set of camera position estimates has been obtained, it may also be possible to build on an idea suggested in [31] and seek a refinement of these position estimates that minimize the overall ℓ_1 norm of the reconstructed image. A multiscale approach could again help such a technique converge if the initial estimates are far off.

Table 1 Reconstruction results with varying numbers of camera positions J

				ML w/out cam.	
J	Independent CS	TC w/ cam.	ML w/ cam.	PSNR	$\frac{1}{J}\sum_j \lvert\theta_j - \hat{\theta}_j\rvert$
200	14.4	22.8	24.7	23.6	(0.0108, 0.0132)
150	13.7	22.9	24.6	23.7	(0.0110, 0.0148)
100	15.1	23.5	25.1	23.9	(0.0177, 0.0121)
70	15.6	23.7	24.6	23.8	(0.0059, 0.0143)

From left to right, the columns correspond to the PSNR (in dB) of image-by-image CS reconstruction from random measurements, joint reconstruction from transform coding measurements with known camera positions, joint reconstruction from random measurements with known camera positions, and joint reconstruction from random measurements with unknown camera positions. The final subcolumn lists the mean absolute error of the estimated camera positions in the vertical and horizontal directions, respectively

Second, an interesting open question is whether the measurement matrices utilized in DCS multi-view imaging scenarios satisfy the RIP with respect to some reconstruction basis Ψ. Establishing an RIP bound would give a guide for the requisite number of measurements (ideally, at each scale) and also give a guarantee for reconstruction accuracy. Although we do not yet have a definitive answer to this question, we suggest that there may be promising connections between these matrices and other structured matrices that have been studied in the CS literature. For example, the measurement matrix $\Phi_{\text{big}}R$ employed in the satellite experiment is closely related to a partial circulant matrix, where the relative shifts between the rows represent the relative offsets between the camera positions. RIP results have been established for circulant matrices [32] that are generated by a densely populated random row vector. In our case, $\Phi_{\text{big}}R$ has more of a block circulant structure because it is generated by the submatrices Φ_j, and so there may also be connections with the analysis in [33]. However, each row of $\Phi_{\text{big}}R$ will contain a large number of zeros, and it is conceivable that this could degrade the isometric property of $\Phi_{\text{big}}R$. We believe, though, that by collecting multiple measurements from each camera, we are compensating for this degradation. Other possible directions for analysis could be to build on the concentration of measure bounds recently established for block diagonal matrices [34] and Toeplitz matrices [35].

Finally, another open question in the manifold lifting framework is what could be said about the uniqueness of $\mathcal{M}(x)$ given samples of $\Phi\mathcal{M}(x)$. When all points on the manifold $\mathcal{M}(x)$ are K-sparse, the RIP can be one avenue to proving uniqueness, but since our objective is to sample fewer than $\mathcal{O}(K\log(N/K))$ measurements for each signal, a stronger argument would be preferable. By considering the restricted degrees of freedom that these signal ensembles have, it seems reasonable to believe that we can in fact establish a stronger result. We are currently exploring geometric arguments for proving uniqueness.

Endnotes

[a]It is also possible to consider other more general non-orthonormal dictionaries. [b]Depending on the scenario, the parameter space Θ could be a subset of $\mathbb{R}^p$, or it could be some more general topological manifold such as SO(3), e.g. if θ corresponds to the orientation of some object in 3D space. [c]Although an IAM $\mathcal{M}$ may not itself be smooth, a regularized manifold $\mathcal{M}_s$ will be smooth, and later in this paper we discuss image

reconstruction strategies based on random projections of $\mathcal{M}_s$ at a sequence of scales s. [d]We have found that this process also performs best using measurements of $h_s * x_j$ for s small because of the smoothness of the manifold $\mathcal{M}_s$ at coarse scales. [e]In our experiments, we choose the parameter ε as somewhat of an oracle, in particular as $1.1||y - \Phi_{big}\widehat{R}x||_2$. In other words, this is slightly larger than the error that would result if we measured the true image x but with the wrong positions as used to define $\widehat{R}$. This process should be made more robust in future work.

Acknowledgements
The authors gratefully acknowledge Richard Baraniuk and Hyeokho Choi for many influential conversations concerning the lumigraph and for their help in developing the lumigraph experiments presented here. This research was partially supported by DARPA Grant HR0011-08-1-0078 and AFOSR Grant FA9550-09-1-0465. A preliminary version of some results in this paper originally appeared in [10].

Author details
[1]Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI, USA [2]Department of Electrical Engineering and Computer Science at the Colorado School of Mines, Golden, CO, USA

Competing interests
The authors declare that they have no competing interests.

Received: 12 July 2011 Accepted: 20 February 2012
Published: 20 February 2012

References

1. D Donoho, Compressed sensing. IEEE Trans Inf Theory. **52**(4), 1289–1306 (2006)
2. E Candès, Compressive sampling, in *Proc Int Congress Math*, Madrid, Spain, **3**, 1433–1452 (2006)
3. M Duarte, M Davenport, D Takhar, J Laska, T Sun, K Kelly, R Baraniuk, Single-pixel imaging via compressive sampling. IEEE Signal Process Mag. **25**(2), 83–91 (2008)
4. D Slepian, J Wolf, Noiseless coding of correlated information sources. IEEE Trans Inf Theory. **19**(4), 471–480 (2003)
5. D Baron, MB Wakin, M Duarte, S Sarvotham, RG Baraniuk, Distributed compressed sensing, (Rice University Technical Report TREE-0612, 2006)
6. X Chen, P Frossard, Joint reconstruction of compressed multi-view images, in *Proc IEEE Int Conf Acoustics, Speech, Signal Process (ICASSP)* (2009)
7. R Marcia, R Willett, Compressive coded aperture video reconstruction, in *Proc Eur Signal Process Conf (EUSIPCO)* (2008)
8. JY Park, MB Wakin, A multiscale framework for compressive sensing of video, in *Proc Picture Coding Symp (PCS)* (2009)
9. S Mallat, *A Wavelet Tour of Signal Processing, The Sparse Way, 3rd edn.*, (Academic Press, 2008)
10. MB Wakin, A manifold lifting algorithm for multi-view compressive imaging, in *Proc Picture Coding Symp (PCS)* (2009)
11. JB Tenenbaum, V Silva, JC Langford, A global geometric framework for nonlinear dimensionality reduction. Science. **290**(5500), 2319–2323 (2000). doi:10.1126/science.290.5500.2319
12. DL Donoho, C Grimes, Image manifolds which are isometric to Euclidean space. J Math Imaging Comp Vis. **23**(1), 5–24 (2005). doi:10.1007/s10851-005-4965-4
13. MB Wakin, D Donoho, H Choi, RG Baraniuk, The multiscale structure of non-differentiable image manifolds, in *Proc Wavelets XI at SPIE Optics and Photonics* (August 2005)
14. R Baraniuk, M Wakin, Random projections of smooth manifolds. Found Comput Math. **9**(1), 51–77 (2009). doi:10.1007/s10208-007-9011-z
15. E Candès, The restricted isometry property and its implications for compressed sensing. Comptes rendus de l'Académie des Sciences, Série I. **346**(9–10), 589–592 (2008)
16. EJ Candès, T Tao, Decoding by linear programming. IEEE Trans Inf Theory. **51**(12), 4203–4215 (2005). doi:10.1109/TIT.2005.858979
17. R Baraniuk, M Davenport, R DeVore, M Wakin, A simple proof of the restricted isometry property for random matrices. Constr Approx. **28**(3), 253–263 (2008). doi:10.1007/s00365-007-9003-x
18. M Wakin, Manifold-based signal recovery and parameter estimation from compressive measurements. (2008), http://arxiv.org/abs/1002.1247
19. M Lustig, DL Donoho, JM Santos, JM Pauly, Compressed sensing MRI, in *IEEE Signal Process Mag.* **25**(2), 72–82 (2008)
20. AD Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. IEEE Trans Inf Theory. **22**, 1–10 (1976). doi:10.1109/TIT.1976.1055508
21. I Tošić, P Frossard, Distributed multi-view image coding with learned dictionaries, in *Proc Int ICST Mobile Multimedia Comm Conf* (2009)
22. N Gehrig, PL Dragotti, Geometry-driven distributed compression of the plenoptic function: performance bounds and constructive algorithms. IEEE Trans Image Process. **18**, 457–470 (2009)
23. I Tošić, P Frossard, Geometry-based distributed scene representation with omnidirectional vision sensors. IEEE Trans Image Process. **17**(7), 1033–1046 (2008)
24. SJ Gortler, R Grzeszczuk, R Szeliski, MF Cohen, The lumigraph, in *Proc Ann Conf Computer Graphics Interactive Tech (SIGGRAPH)* (1996)
25. J Berent, PL Dragotti, Plenoptic manifolds. IEEE Signal Process Mag. **24**(6) (2007)
26. DL Donoho, Wedgelets: Nearly-minimax estimation of edges. Ann Statist. **27**, 859–897 (1999). doi:10.1214/aos/1018031261
27. V Chandrasekaran, MB Wakin, D Baron, R Baraniuk, Representation and compression of multi-dimensional piecewise functions using surflets. IEEE Trans Inf Theory. **55**(1), 374–400 (2009)
28. MB Wakin, The geometry of low-dimensional signal models, PhD dissertation, (Department of Electrical and Computer Engineering, Rice University, Houston, TX, 2006)
29. R Coifman, F Geshwind, Y Meyer, Noiselets. Appl Comput Harmon Anal. **10**(1), 27–44 (2001). doi:10.1006/acha.2000.0313
30. A Singer, RR Coifman, FJ Sigworth, DW Chester, Y Shkolnisky, Detecting consistent common lines in cryo-em by voting. J Struct Biol. **169**(3), 312–322 (2010). doi:10.1016/j.jsb.2009.11.003
31. C Huff, R Muise, Wide-area surveillance with multiple cameras using distributed compressive imaging, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. **8055** (April 2011)
32. WU Bajwa, JD Haupt, GM Raz, SJ Wright, RD Nowak, Toeplitz-structured compressed sensing matrices, in *Proc IEEE/SP Workshop Stat Signal Process (SSP)* (2007)
33. RF Marcia, RM Willett, Compressive coded aperture superresolution image reconstruction, in *Proc IEEE Int Conf Acoustics, Speech, Signal Process (ICASSP)* (2008)
34. MB Wakin, JY Park, HL Yap, CJ Rozell, Concentration of measure for block diagonal measurement matrices, in *Proc IEEE Int Conf Acoustics, Speech, Signal Process (ICASSP)* (2010)
35. BM Sanandaji, TL Vincent, MB Wakin, Concentration of measure inequalities for compressive Toeplitz matrices with applications to detection and system identification, in *Proc IEEE Conf Decision and Control (CDC)* (2010)

doi:10.1186/1687-6180-2012-37
Cite this article as: Park and Wakin: **A geometric approach to multi-view compressive imaging.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:37.