Static architecture for compressive motion detection in persistent, pervasive surveillance applications

M.D. Stenner[1], D.J. Townsend[1], and M.E. Gehm[2,3]

[1]*MITRE Corp., Bedford, MA*
[2]*Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ*
[3]*College of Optical Sciences, University of Arizona, Tucson, AZ*
gehm@ece.arizona.edu

Abstract: High-resolution, wide-field-of-view airborne imaging produces large optical systems and data streams. Significant simplification is possible if motion detection, rather than full imaging, is the goal. We consider a static, compressive-sensing architecture for this problem.

OCIS codes: (110.1758) Computational imaging; (110.4153) Motion estimation and optical flow

1. Introduction

Recently, *persistent, pervasive surveillance* (PPS)—the ability to continually monitor an extended area—has become central to a number of ongoing security efforts. The primary difficulty in PPS applications is that the use of traditional imaging architectures necessarily produces a tremendous amount of data. Direct transmission of this data would require unrealistic communication bandwidths. As a result, the data is digitally compressed on-platform before transmission or simply stored on the platform. While this reduces (or eliminates) the communication challenge, the optics, photodetectors, and associated electronics must still operate at the underlying bandwidth. When combined with size, weight, and power limitations of available platforms, the result is frequently limited performance and high system cost.

However, many PPS applications do not require an actual image of the field-of-view (FOV), but rather just information about how object move within that FOV. A particular opportunity arises when the moving objects are the same scale as the physical sampling (i. e. in a traditional image, they are pixel-sized or smaller). In this case, the *difference* between two consecutive image frames shows only a small number of significant values (corresponding to the initial and final positions of the movers). This "difference-image" has two useful properties: a) it retains all of the important information regarding the movers, and b) it is *sparse*.

The signal processing subfield of *compressive sensing* is concerned precisely with the efficient measurement of sparse signals [1, 2, 3, 4, 5]. Thus, rather than acquire traditional images and *algorithmically* form the difference images to extract the movers, we consider instead optical implementations where we use the concepts of compressive sensing to *directly infer the difference image* from measurements on the optical field. The expectation is that such an approach will allow determination of the difference image with greatly reduced system resources (detectors, bandwidth, optics, etc.) when compared with the traditional approach. The particular sampling strategy must, of course, be *physically implementable* (i. e. consistent with the physical laws governing electromagnetic fields). This is a central constraint in computational sensing and determining good implementations remains more art then science at the current time. Given this flexibility, it is unsurprising that other groups have also explored alternative sparse-signal approaches to the motion-detection problem [6, 7].

Below, we introduce the basics of compressive sensing and its application to the motion-detection problem. We then discuss the practical appeal of a *static* implementation, before proposing a general static architecture and discussing our current status and future plans.

2. Compressive sensing

A simple mathematical model of sensor operation is:

$$\mathbf{m} = \mathbf{T}\mathbf{s} + \mathbf{n}. \tag{1}$$

Here $\mathbf{s}$ is a vector representation of the scene, $\mathbf{T}$ is a matrix that describes the measurement structure of the sensor, and $\mathbf{n}$ represents additive noise in the system. The vector $\mathbf{m}$ therefore represents the full set of (noise-corrupted) measurements acquired from a given scene. In traditional imaging, each element in the source is measured independently, and therefore $\mathbf{T}$ has the form of the identity matrix. Here, we are concerned with compressive, multiplex sensors where $\mathbf{T}$ is rectangular in form ($u \times v$ with $u < v$) with significant off-diagonal elements.

In general, the scene $\mathbf{s}$ is not sparse, and thus cannot be reconstructed from $\mathbf{m}$ using compressive sensing techniques. However, if we define the difference between two consecutive measurement vectors as $\Delta\mathbf{m}$, we find

$$\Delta\mathbf{m} = \mathbf{m}_2 - \mathbf{m}_1 = \mathbf{T}(\mathbf{s}_2 - \mathbf{s}_1) + (\mathbf{n}_2 - \mathbf{n}_1) = \mathbf{T}\mathbf{d} + \mathbf{n}_\Delta. \tag{2}$$

This equation has the same general form as Eq. 1, but the source vector $\mathbf{d} = \mathbf{s}_2 - \mathbf{s}_1$ *is* sparse given the limitations we discussed above. As such, compressive sensing techniques *can* be used to infer the structure of $\mathbf{d}$ from the vector $\Delta\mathbf{m}$.

Compressive sensing techniques designed to infer the structure of a sparse signal often rely on ℓ-1 minimization methods. To understand the utility of this technique, we first point out that for row-deficient $\mathbf{T}$ (the form we desire) the measurement scheme is essentially an *underdetermined* linear problem. Thus, there are a potentially infinite number of solutions that are consistent with the observations. Compressive sensing techniques constrain the solution through additional knowledge—specifically, the fact that the source is sparse. Traditional least-squares solutions (minimization of ℓ-2) are computationally simple, but do not preferentially find sparse solutions. Thus they do not take advantage of the available additional knowledge. Minimization of the ℓ-0 norm would explicitly find the sparsest solution consistent with the measurements, but ℓ-0 minimization is computationally challenging. However, it was recently realized that, when the source is sufficiently sparse, the minimum ℓ-1 solution is equivalent to the minimum ℓ-0 solution, and that the ℓ-1 minimization can be efficiently found via linear programming methods.

Note that this just addresses the *possibility* of reconstructing the sparse source. Ultimately, whether the technique succeeds also depends on the structure of $\mathbf{T}$. The second great realization of recent years is the applicability of random sampling to the design of $\mathbf{T}$. The success of the ℓ-1 technique relies on $\mathbf{T}$ having the *restricted isometry property* (RIP). It has been shown that instantiations of several different classes of *random matrices* do indeed have the RIP—thereby making them suitable sampling matrices for compressive sensing problems. This, combined with the ℓ-1 technique, then results in a complete solution to the efficient measurement of sparse signals.

3. Static implementation

As mentioned previously, although many forms for $\mathbf{T}$ can be considered, a smaller subset can actually be *implemented*. The most fundamental constraint is that, for incoherent imaging, $\mathbf{T}$ may contain only *non-negative values* in the domain $[0, 1]$. The most-famous compressive sensing implementation, the *single-pixel camera* uses a $\mathbf{T}$ with random 0/1 values [8]. This implementation, like most early compressive sensing implementations, utilizes an active element (e. g. a micro-mirror array, LCD spatial light modulator, etc.) to create the measurements. In this type of approach, the system cycles through the *rows* of $\mathbf{T}$ sequentially, determining the elements of $\mathbf{m}$ one at a time.

This approach is extremely flexible in regard to the possible forms of $\mathbf{T}$. Any arbitrary pattern of values in the domain $[0, 1]$ can be used, and new patterns are a simple matter of reprogramming the control software. However, this flexibility comes with a number of significant costs.

First, as the elements of $\mathbf{m}$ are measured sequentially, there is an inherent assumption that the source remains unchanged as the sequence of measurements is made. As motion-detection is inherently time-sensitive, such a technique is contraindicated for this application. Second, typical architectures of this type generally operate by forming an intermediate image plane which is made coincident with the active element. While the approach reduces the size and complexity of the detector plane, it does so by extending the system track-length and adding additional optical components and their associated mass—a significant drawback on airborne platforms. Finally, the active components are themselves complex and more fragile than monolithic components.

For all these reasons, we were motivated to consider *static* approaches to compressive sensing in the motion-detection problem. Thus, rather than measuring a *sequence* of maps from the source to a *single* pixel, we make a *parallel measurement* that maps the source in different ways onto a *small set* of pixels. This reduces the available coding strategies (i. e. forms of $\mathbf{T}$), and forces us to carefully consider measurement design strategies. We find it useful to address the issue heuristically, extracting certain key features of the random sampling strategies as goals. We believe the following are the central features of a good sampling strategy: 1) Each element (location) in the source contributes to many different measurements (pixels), 2) Each measurement (pixel) combines many different elements (locations), and 3) The coding of a particular location/pixel is relatively uncorrelated with its neighbors.

In our early work [9], we considered an architecture where a detector array was placed slightly behind an image plane containing a diffusing element. In this case, $\mathbf{T}$ took on a block-Toeplitz form with underlying random elements. Work by others had shown that a $\mathbf{T}$ of this form could meet the RIP [10]. However, we found that for the system parameters we considered, there was a residual correlation in the measurements that prevented accurate reconstruction.

We are now working with a system architecture that uses one or more transmissive masks to implement a *visibility function* between the source and the detector array. The resulting $\mathbf{T}$ is truly shift-variant, eliminating the strong correlations. With this change, the approach becomes similar to earlier work on *reference structure tomography*, although that work did not utilize modern compressive sensing approaches or explicitly consider the motion detection problem [11].

Early analysis and simulation validate the idea that the visibility approach yields system responses with the three desired properties, and have led to successful compressive measurement and reconstruction in simulation. Our current experimental efforts are focused on implementing this measurement scheme experimentally, as shown in Fig. 1. Ongoing research is focused on improving measurement diversity, system characterization, and realization of full compressive measurement in the experimental system.

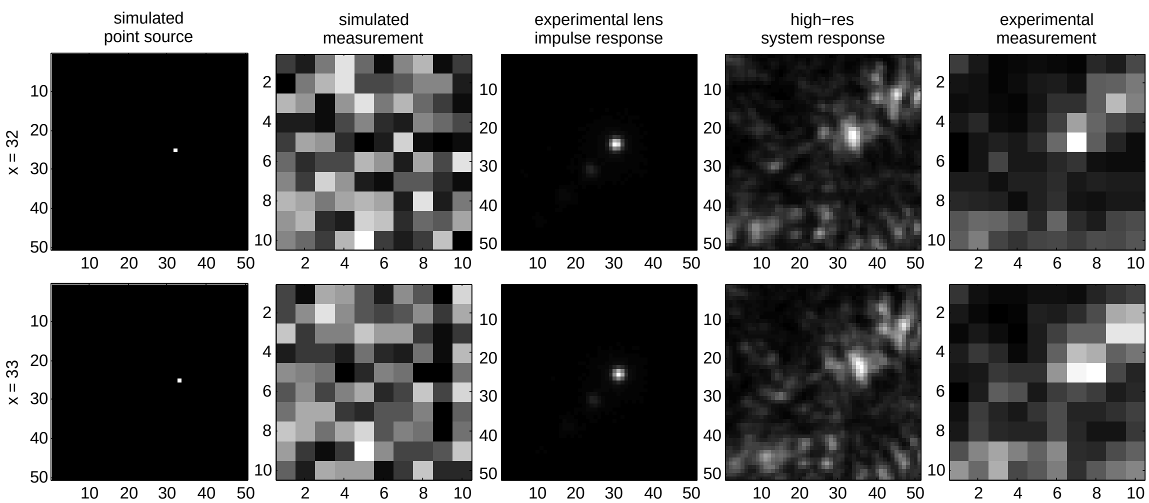

Fig. 1. Simulation and experimental results for measurements of point sources at two neighboring locations. The distance between locations is $1/10$ the distance between pixels in the sensor plane. Shown (from left to right) are the simulated point source, simulated compressive measurement, experimental optical PSF without masks, high-resolution PSF with masks, and compressive measurement. The experiment and simulation use different random mask realizations and are not expected to match quantitatively.

4. Conclusions

Compressive sensing approaches can greatly increase the performance of PPS systems by dramatically reducing the complexity of the system while retaining information related to the motion-detection problem. Despite the success of active approaches in other compressive sensing implementations, they are poorly suited to the particular challenges in motion detection. As such, we are investigating purely static architectures. We are currently optimizing a 2nd-generation system architecture utilizing a visibility-function approach. Experimental proof-of-concept prototypes are under construction. We will report on the final system design and resulting performance.

References

1. D. Donoho, "Compressed Sensing," Information Theory, IEEE Transactions on **52,** 1289–1306 (2006).
2. D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," Information Theory, IEEE Transactions on **52,** 6–18 (2006).
3. E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," Information Theory, IEEE Transactions on **52,** 489–509 (2006).
4. E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math **59,** 1207–1223 (2006).
5. E. Candes and J. Romberg, "Practical signal recovery from random projections," IEEE Trans. Signal Processing (2005).
6. D. Reddy, A. Sankaranarayanan, V. Cevher, and R. Chellappa, *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* (2008), pp. 221–224.
7. S. Uttam, N. Goodman, M. Neifeld, C. Kim, R. John, J. Kim, and D. Brady, "Optically multiplexed imaging with superposition space tracking," Optics Express **17,** 1691–1713 (2009).
8. R. Baraniuk, "Compressive sensing," IEEE Signal Processing Magazine **24,** 118 (2007).
9. M. Gehm and M. Stenner, *Proceedings of SPIE* (2009), Vol. 7468, p. 746808.
10. F. Sebert, Y. Zou, and L. Ying, *Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on* (2008), pp. 47–50.
11. D. Brady, N. Pitsianis, and X. Sun, "Reference structure tomography," Journal of the Optical Society of America A **21,** 1140–1147 (2004).