Compressive measurement for target tracking in persistent, pervasive surveillance applications

Michael E. Gehm[a,b] and Michael D. Stenner[c]

[a]Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ;
[b]College of Optical Sciences, University of Arizona, Tucson, AZ;
[c]Emerging Technologies Office, MITRE Corp., Boston, MA

ABSTRACT

Motion tracking in persistent surveillance applications enters an interesting regime when the movers are of a size on the order of the image resolution elements or smaller. In this case, for reasonable scenes, information about the movers is a natively sparse signal—in an observation of a scene at two closely separated time-steps, only a small number of locations (those associated with the movers) will have changed dramatically. Thus, this particular application is well-suited for compressive sensing techniques that attempt to efficiently measure sparse signals. Recently, we have been investigating two different approaches to compressive measurement for this application. The first, differential Combinatorial Group Testing (dCGT), is a natural extension of group testing ideas to situations where signal differences are sparse. The second methodology is an ℓ-1-minimization based recovery approach centered on recent work in random (and designed) multiplex sensing. In this manuscript we will discuss these methods as they apply to the motion tracking problem, discuss various performance limits, present early simulation results, and discuss notional optical architectures for implementing a compressive measurement scheme.

Keywords: motion tracking, group testing, compressive sensing

1. INTRODUCTION

In recent years, *persistent, pervasive surveillance* (PPS)—that is, the ability to constantly observe a large geographical region—has emerged as an important tool in a variety of security applications. The challenge in PPS is that applying conventional imaging approaches to these applications can result in overwhelming data bandwidths. Typically, systems will address this by acquiring high-resolution video-rate (or near-video-rate) datastreams and then compressing them using various algorithms to reduce the overall bandwidth to a more manageable level. While this approach does reduce the communications bandwidth of the system, the optics and photodetector hardware must still operate at the native bandwidth, limiting performance and increasing overall system cost. Thus we search for alternative sensing methods that can reduce the bandwidth burden on the entire sensor platform.

An important foothold is provided by the fact that for many applications, the ultimate item of interest is not the actual image of the field-of-view (FOV) but rather information about the motion of objects within the FOV. When the system geometry is such that the moving objects of interest subtend a solid angle that is on the order of the sampling resolution of the imaging system (i.e. in a traditional image, the movers are roughly pixel-scale), an interesting opportunity arises. If we compare two consecutive image frames from a system in this regime, for reasonable scenes we will notice that only a small number of the pixels will exhibit a significant change. Thus, for reasonable scenes, a "difference image" can be considered as *sparse* (no detector noise) or *compressible* (noise small compared to the differences associated with a mover). Further, this difference image retains all of the important information about the movers that was present in the original scenes.

The efficient measurement of sparse/compressive signals has been a particularly active area of signal processing in recent years and rests upon several decades of efforts in this area. The realization that difference images are sparse and retain the relevant information needed for motion tracking applications allows us to bring to bear a

Corresponding address: gehm@ece.arizona.edu

Adaptive Coded Aperture Imaging, Non-Imaging, and Unconventional Imaging Sensor Systems,
edited by Stanley Rogers, David P. Casasent, Jean J. Dolne, Thomas J. Karr, Victor L. Gamiz, Proc. of SPIE
Vol. 7468, 746808 · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.827048

variety of mathematical techniques designed for this sparse signal measurement/reconstruction. Our goal, then, is to use these techniques to design a sensor system that allows us to measure and reconstruct these difference images using fewer system resources (detectors, bandwidth, etc.) than the traditional approach. We note that a number of other groups have explored alternative sparse-signal approaches to motion tracking.[1,2]

In the following sections we first formalize the signal and measurement model that we are assuming for the problem, and then introduce several different mathematical approaches to sparse signal recovery and discuss their relative merits. We conclude with an outline for a notional system to implement one of these approaches and discuss our future plans.

2. SIGNAL AND MEASUREMENT MODEL

We mathematically represent the sensor operation as

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

Here $\mathbf{s}$ is a vector representation of the scene, $\mathbf{T}$ is a matrix that describes the measurement structure of the sensor. For a traditional imager, $\mathbf{T}$ would be the identity matrix. We will be concerned with compressive, multiplex sensors where $\mathbf{T}$ is rectangular ($u \times v$ with $u < v$) and has off-diagonal elements. The vector $\mathbf{n}$ represents additive noise in the system, and thus $\mathbf{m}$ is a vector of the measurements that result from a single exposure of the scene in $\mathbf{s}$.

Note the scene vector $\mathbf{s}$ is not sparse/compressible in general. However, defining the difference of two consecutive measurement vectors as $\Delta\mathbf{m}$, we can write

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

We end with a form similar to Eq. 1, but in terms of the difference vector $\mathbf{d} = \mathbf{s}_2 - \mathbf{s}_1$ between two consecutive scenes and with a noise vector drawn from a potentially different distribution (*e.g.* if $\mathbf{n}$ is AWGN, then $\mathbf{n}_\Delta$ is as well, but with zero mean and twice the variance). As we have noted previously, for reasonable conditions, $\mathbf{d}$ *is* sparse/compressible, and is thus amenable to our proposed approach.

3. GROUP TESTING APPROACH

Combinatorial group testing (CGT) was first conceived during World War II as a means of providing a more efficient method of testing servicemembers for syphilis (although the plan was never implemented).[3] Since the incidence of the disease was expected to be relatively rare, performing a blood test on every servicemember would be highly inefficient, as the majority of the tests would be negative. The concept,[3] was to *combine* blood samples into *groups*, which were then tested. If the test on that group was negative, then a single test had replaced the much larger number of individual tests. On the other hand, if the single test was positive, then the individual servicemembers in that group could be tested individually to determine which members were infected. Of course, these groups can be combined in higher-order groups to create a multi-tier approach with extraordinary efficiency gains.

The current formulation of CGT refers to the detection of a small number of *positive* members of a much larger population (the remainder of which are termed *negative*). A proposed set of tests is known as a *design*. The group assignments in a design can be either *deterministic* or *random*. The canonical problem of identifying exactly p positives in a population of n members is known as the (p, n)-problem. A related problem, that of identifying *up to* p positives in a population of n members is known as the $(\bar{p}, n)$-problem. Another way of considering these problems is to realize that they correspond to finding the *single* system configuration that corresponds to the true configuration out of the N possible system configurations. For the (p, n) problem we have

$$N_p = \binom{n}{p} = \frac{n!}{(n-p)!\,p!}, \tag{3}$$

and for the $(\bar{p}, n)$-problem we have

$$N_{\bar{p}} = \sum_{k=0}^{p} \binom{n}{k} = \sum_{k=0}^{p} \frac{n!}{(n-k)!\,k!}. \tag{4}$$

It is this combinatorial growth in the size of potential system configurations that gives the field its name.

As a simple example, we consider the $(1, 8)$-problem. That is, locating the single positive member out of a population of 8 (for the purposes of this example, we will number them as 0-7). Without using group testing, we could individually test all 8 members of the population. A slightly smarter approach (if the testing need not occur simultaneously) would be to consider a sequential test, where we are free to stop after we discover the positive member. On average, this would require 4.5 tests.

Creating a deterministic, CGT design for this problem is actually quite simple. The positive member can be unambiguously identified after 3 tests. In test one, we test the group {4,5,6,7}. In test two, we test the group {2,3,6,7}. Finally, in test three, we test the group {1,3,5,7}. The validity of this design is easy to understand: Each test takes the population (or sub-population) known to contain the positive member, and divides it in half and determines which half contains the positive member. This approach is known as the *binary-division* algorithm.

An important question is: What is the minimum number of tests required for a (p, n)-problem? Based on the binary ideas described above, we are led to what is known as the *information-theoretic bound*. Given that there are N_p (Eq.3) possible system configurations, we require at least as many tests as it takes to allow us to numerically represent any of the possible configurations. Since the outcome of the tests are either *positive* or *negative* the test result can be represented as a binary digit (note that for continuous-valued quantities, we essentially compare the test result to some threshold to determine if the result is positive or negative). Thus we are led to the conclusion that we must have at least as many tests as there are digits in the binary representation of N_p. In other words, the information-theoretic lower bound $B(p, n)$ for the (p, n)-problem is

$$B(p, n) = \lceil \log_2 N_p \rceil, \tag{5}$$

where $\lceil \cdot \rceil$ is the *ceiling* operator. Note that for the $(1, 8)$-problem considered above, we find $B(p, n) = \log_2 8 = 3$, showing that the proposed design is optimal (as any shorter design would violate the information-theoretic bound).

Note that the information-theoretic bound is not an actual prediction of the smallest valid design. Many (if not most) (p, n)-problems do not have solutions of size $B(p, n)$. For this reason, we will define the true minimum number of tests required as $M(p, n)$, with

$$B(p, n) \leq M(p, n). \tag{6}$$

No closed-form expression for $M(p, n)$ is known; however evidence seems to indicate that $M(p, n)$ is often on the close order of $B(p, n)$. Further note that these results only describe the minimal size of the smallest possible design, they do not provide guidance in actually constructing the design of this size. This will be discussed further below. Finally, what of the $(\bar{p}, n)$-problem? In fact, it has been shown[4] that $M(\bar{p}, n)$ is at most one more than $M(p, n)$.

3.1 DIFFERENTIAL COMBINATORIAL GROUP TESTING (dCGT)

Unfortunately, a straightforward application of CGT will not meet our needs. The subtraction in Eq. 2 introduces a fundamental *bipolarity* to the measurement vector $\Delta\mathbf{m}$. We thus implement a bipolar threshold that leaves the results of the individual tests $\in \{-1, 0, 1\}$, corresponding to tests that decreased, remained the same, or increased respectively between the two exposures. The result of any measurement is then a string in *balanced ternary* representation. We can map this string to traditional ternary by adding a '1' to each digit of the string. We call this approach (CGT with bipolar thresholding after subtraction) *differential combinatorial group testing* (dCGT).

In the same way that the binary representation of measurement vectors in CGT drives the logarithm base-2 formulation of the information-theoretic lower bound, the ternary representation in dCGT results in the lower bound

$$B(p, n) = \lceil \log_3 N \rceil, \tag{7}$$

with N again the total number of possible configurations that the sensor must be able to distinguish. In dCGT, the calculation of N is more involved—the states we are attempting to enumerate are the states of a *change map*, that is, the scene locations (FOV elements) that have changed between the two exposures.

Consider a scene with n FOV elements and up to p movers. There are several complications that result from motion: First, movers may enter/leave the FOV. This situation is acceptable as long as the total number of movers in the scene does not exceed p. Second, for a given set of exposures, movers may have a velocity sufficiently small that they do not transition between FOV elements. The result of this is that we have *at most* $2p$ FOV elements that change between exposures. The total number of configurations of the change map is then

$$N = \sum_{k=0}^{2p} \binom{n}{k} = \sum_{k=0}^{2p} \frac{n!}{(n-k)!\,k!}.$$

Again, N grows combinatorially, and does so faster than in the CGT case. This is mitigated somewhat by the base-3 logarithm. A simple numerical example can provide some idea of scale: This result claims that a dCGT system for tracking up to 5 movers in a scene with 10^4 FOV elements requires at least 403 measurements. Achieving anything near the information theoretic bound is therefore a significant efficiency improvement of the traditional approach (which would make 10^4 measurements).

3.1.1 DESIGN AND DECODING CHALLENGES

Despite efforts over several decades, CGT approaches have not become widespread and applications involving CGT rarely achieve efficiencies near the information-theoretic limit. This is largely the result of two interrelated challenges that arise in the context of CGT: design and decoding. The *design challenge* arises because although the simple information-theoretic argument allows us to determine the minimum required number of measurements, it provides no information about the construction of an overall design that produces a one-to-one mapping between possible configurations and the resulting measurement strings.

In addition to needing a working measurement design, we also need an effective method for mapping the measurement string back to a particular configuration state. This is the *decoding challenge*. The combinatorial growth of potential configuration states renders a lookup table approach completely impractical, and therefore we require some algorithmic method for extracting the meaning from the measurement string. For this reason, the most effective CGT designs to date have been based on specific families of mathematical objects that both solve the design challenge (by provably providing unique mappings) and allow simple decoding of the resulting measurement strings. Unfortunately, the designs generated by these methods fall far short of the information-theoretic bound, and therefore do not result in the full efficiency gains promised by the CGT framework. In formulating dCGT, it is our hope that the underlying ternary structure may prove more amenable to designs operating near the performance limit. Our efforts to date have focused on the design of constrained *random* dCGT designs (as opposed to the deterministic concepts discussed above) as well as methods for efficiently decoding these measurements. These efforts are beyond the scope of this manuscript.

4. ℓ-1-MINIMIZATION APPROACH

An alternative approach to efficient measurement of sparse difference images is provided by the ℓ-1 minimization methods that have attracted so much attention in recent years.[5–9] To understand the utility of the ℓ-1 approach, we first recognize that for systems where $\mathbf{T}$ has fewer rows than columns (as we desire), that Eq. 2 is *underdetermined*. Thus, there are numerous possible solutions that are consistent with the measurements. However, we have additional knowledge that can be used to constrain the solution: The difference image $\mathbf{d}$ is sparse/compressible. Traditional methods such as least-squares solution (minimization of the ℓ-2 norm) are extremely computationally efficient, but do not preferentially seek sparse solutions. Minimization of the ℓ-0 norm would, by definition, find a sparse solution (since the ℓ-0 norm of a vector is the number of non-zero elements), but ℓ-0 minimization of large problems is known to be computationally intractable. Additionally, the applicability of the ℓ-0 norm to compressible (as opposed to sparse) signals is questionable. The ℓ-1 norm (sum of vector elements) occupies an interesting position between the ℓ-2 and ℓ-0 cases. ℓ-1 minimization can be accomplished

via linear programming techniques, and is therefore computationally tractable. Additionally, ℓ-1 minimization preferentially finds sparse/compressible solutions.

The design and decoding challenges that arose in the context of dCGT are not unique to that approach, but are instead more general to the problem of efficient (underdetermined) measurement of sparse signals. The ℓ-1 approach is a potential solution to the decoding challenge, providing a tractable method of converting a set of measurements into an estimate of the underlying sparse state. The great realization of recent years has been the applicability of *random* sampling to the design challenge. In particular, the realization that the reconstruction techniques based on ℓ-1 minimization are effective provided that the sensing matrix $\mathbf{T}$ has the *restricted isometry property* (RIP)

$$(1-\delta_s)||\mathbf{s}||_2^2 \leq ||\mathbf{Ts}||_2^2 \leq (1+\delta_s)||\mathbf{s}||_2^2, \tag{8}$$

for all sufficiently sparse vectors $\mathbf{s}$. Here $||\cdot||_2^2$ represents the square of the ℓ-2 norm. When the RIP holds, the sensing matrix $\mathbf{T}$ approximately maintains the lengths of sufficiently sparse vectors (*i.e.* it is an approximate isometry for those vectors). As a result of the length-preservation, interior angles and inner products between the vectors are also approximately preserved. Further, since any linear functional can be represented as an inner product with a fixed vector, the action of linear operators in the dimensionally-reduced space is similarly preserved. It is this preservation of sparse vector geometry under the action of $\mathbf{T}$ that provides the utility of the ℓ-1 approach.

The magnitude of the *isometry constant* (the smallest value of δ_s for which Eq. 8 remains true) determines the degree to which the ℓ-1 approach produces a result that matches the ℓ-0 solution. In addition to formulating the problem in terms of the RIP, the manuscripts[5–9] showed that instantiations of certain classes of *random matrices* exhibit the RIP with a high degree of probability. This result, therefore, offers a simple solution to the design problem—do no design and use a random sampling matrix instead. This, combined with the computational tractability of the ℓ-1 minimization approach provides a complete solution to the efficient measurement of sparse vectors.

4.1 SIMULATION

To evaluate the utility of the ℓ-1 approach to motion detection in PPS applications, we developed a simple MATLAB simulation. The overall simulation framework consists of three parts: 1) a scene generator, which moves an arbitrary number of single-pixel movers around an image of city streets seen from altitude, 2) a sensor simulator which generates a measurement vector from each time step of the scene by applying a sensing matrix $\mathbf{T}$, and 3) a reconstruction routine that attempts to reconstruct the difference images of the scene (*e.g.* a change map of the scene) from differences of consecutive measurement vectors $\Delta\mathbf{m}$. With this framework we can evaluate the reconstruction accuracy and robustness of various sensing matrices and reconstruction algorithms.

Results from one simple simulation are presented below in Fig. 1. The scene includes 100×100 FOV elements with up to 20 simultaneous movers, observed over 200 timesteps. The reconstruction algorithm is the "L1_LS" package provided by the Boyd group at Stanford.[10] The sensing matrix is $10^3 \times 10^4$ (compression ratio of 10), with the elements randomly drawn from $\{0, 1\}$ and with each row then normalized to maintain the signal sum (equivalent to conservation of energy for the case of an incoherent optical image). The upper left of Fig. 1 shows frame 100 of the scene as shown to the sensor. Note that the movers are in general not easily observed on the background. The upper right shows the actual location of the movers in the scene without the cityscape background. The lower left shows the difference image formed by the difference between frames 100 and 99. Note that in general, as expected, the movers appear as a bipolar signal involving two pixels. The magnitude of the pixels depends on the relative intensity of the mover and the background of the scene. In addition, not all movers have velocities such that they transition from one FOV element to another during a single frame. For these reasons, not all movers appear in the difference image, and of those that do, not all have a clearly defined bipolar signal. The lower right is the reconstructed difference image formed by running the ℓ-1 reconstruction algorithm on the difference of the random multiplex measurements made on frames 100 and 99. The reconstruction is in perfect qualitative agreement with the actual difference image. Quantitatively, the maximum relative error is less than 3% and the RMSE is $\leq 9 \times 10^{-3}$.

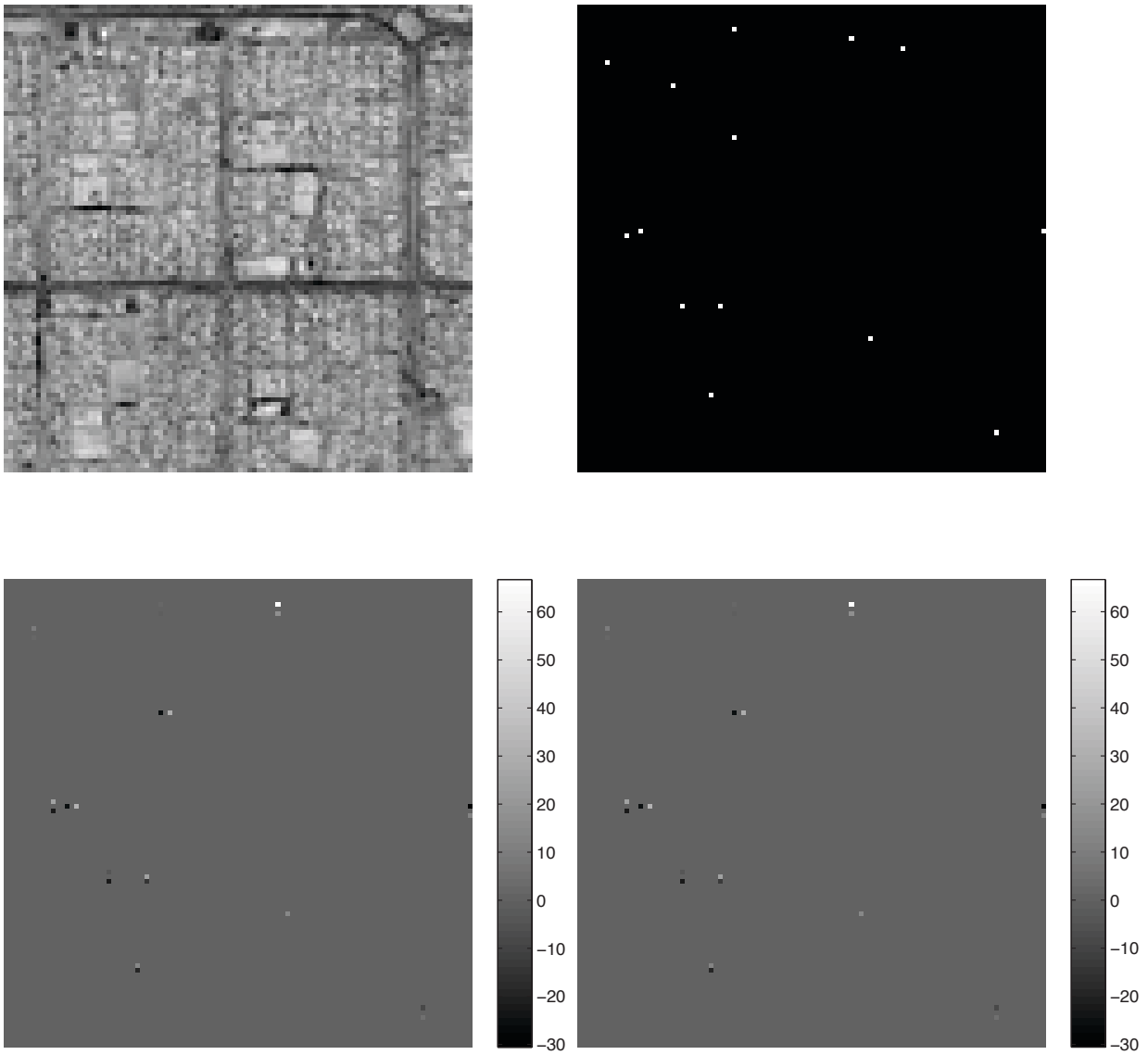

Figure 1. (Upper left) The scene as presented to the sensor in frame 100. The movers are barely visible on the scene background. (Upper right) The actual mover locations in the scene. (Bottom left) The true difference image between frames 99 and 100. (Bottom right) The reconstruction of the difference image from an ℓ-1 minimization algorithm as applied to the difference of random multiplex measurements made on frames 99 and 100. The qualitative agreement is exact, the quantitative agreement is within 3% with an RMSE $\leq 9 \times 10^{-3}$.

4.2 NOTIONAL DESIGN

The simulation results support the idea that ℓ-1 style approaches offer the potential for highly efficient motion tracking in PPS. Of course, at this point, a new challenge arises—to develop an optical architecture capable of implementing the multiplex sensing needed by either the dCGT or the ℓ-1 technique. To date, the majority of experimental systems to implement these approaches have utilized spatial light modulators (SLMs) of one form or another (either an array of fixed masks or a dynamic SLM such as a digital micromirror array (DMM)). These architectures typically involve parallel imaging of the scene (in the case of an array of masks) or temporal sequencing (with the dynamic SLMs). Either approach complicates the optical architecture.

We have begun considering an alternative architecture, which dispenses with the ability to generate controlled multiplex measurements in exchange for optical simplicity and increased measurement SNR. A schematic of the architecture is shown in Fig. 2. The key idea is that the scene is imaged onto a random diffuser (random phase mask) with features smaller than the desired image resolution. Photons striking the diffuser are then scattered over an angular range before being intercepted by a detector array. The diffuser thus randomly distributes energy from one resolution element into a small neighboorhood of resolution elements that are downsampled by the detector array. In this way, the output of each detector element is proportional to a weighted sum of intensities from a number of resolution elements. Unlike the matrices discussed above, however, the elements present in any particular weighted sum are not drawn globally from all of the available resolution elements, but from a local neighborhood. In this way, the architecture owes much to *extended PSF imagers* discussed previously in the literature.[11] Despite this non-global behavior, early simulations support the idea that ℓ-1 based methods can accurately reconstruct scenes measured in this way.

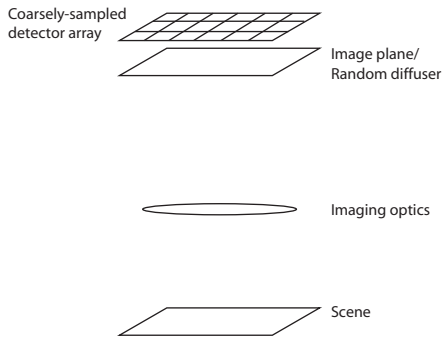

Figure 2. Simple schematic of a notional optical architecture. The scene is imaged onto a random phase mask. The mask scatters the photons in a pseudorandom manner over a small spatial neighborhood which is downsampled by large detector elements. In contrast to SLM based systems, the architecture is particularly simple. Early simulations support the idea that ℓ-1 based methods can provide accurate reconstruction despite the non-global nature of the sensing matrix $\mathbf{T}$.

5. CONCLUSION AND FUTURE WORK

We have begun investigating methods for efficient motion tracking in PPS applications. We have formulated a variation on CGT, which we term dCGT suitable to the motion tracking task and have developed a simple information-theoretic bound on its performance. We have also investigated ℓ-1 based methods for sparse signal recovery in these kinds of systems and have performed basic simulations showing the viability of the general approach. In addition, we have begun formulating a practical system architecture that is potentially more suitable to the application than other experiments to date. We are continuing our research into both the dCGT and ℓ-1 systems, in particular the new system architecture. Future manuscripts will be forthcoming as results warrant.

REFERENCES

[1] Reddy, D., Sankaranarayanan, A., Cevher, V., and Chellappa, R., "Compressed sensing for multi-view tracking and 3-D voxel reconstruction," in [*Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*], 221–224 (2008).

[2] Uttam, S., Goodman, N., Neifeld, M., Kim, C., John, R., Kim, J., and Brady, D., "Optically multiplexed imaging with superposition space tracking," *Optics Express* **17**(3), 1691–1713 (2009).

[3] Dorfman, R., "The detection of defective members of large populations," *Ann. Math. Statist.* **14**, 436–440 (1943).

[4] Hwang, F., Song, T., and Du, D., "Hypergeometric and generalized hypergeometric group testing," *SIAM J. Alg. Disc. Methods* **2**, 426–428 (1981).

[5] Donoho, D., "Compressed Sensing," *Information Theory, IEEE Transactions on* **52**(4), 1289–1306 (2006).

[6] Donoho, D., Elad, M., and Temlyakov, V., "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on* **52**(1), 6–18 (2006).

[7] Candes, E., Romberg, J., and Tao, T., "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on* **52**(2), 489–509 (2006).

[8] Candes, E., Romberg, J., and Tao, T., "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math* **59**(8), 1207–1223 (2006).

[9] Candes, E. and Romberg, J., "Practical signal recovery from random projections," *IEEE Trans. Signal Processing* (2005).

[10] Kim, S., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D., "An interior-point method for large-scale l1-regularized least squares.," *Selected Topics in Signal Processing, IEEE Journal of* **1**(4), 606–617 (2007).

[11] Ashok, A. and Neifeld, M., "Pseudorandom phase masks for superresolution imaging from subpixel shifting," *Appl. Opt.* **46**(12), 2256–2268 (2007).