Focal Plane Array Folding for Efficient Information Extraction and Tracking

Lei Hamilton
Draper Laboratory
Cambridge, MA 02139

Dan Parker*
Northeastern University
Boston, MA 02115

Chris Yu
Draper Laboratory
Cambridge, MA 02139

Piotr Indyk
Massachusetts Institute of Technology
Cambridge, MA 02139

***Abstract*—We develop a novel compressive sensing based approach for detecting point sources in images and tracking of moving point sources across temporal images. One application is the muzzle flash detection and tracking problem. We pursue the concept of lower-dimension signal representation from structured sparse matrices, which is in contrast to the use of random sparse matrices described in common compressive sensing algorithms. The primary motivation is that an approach using structured sparse matrices can lead to efficient hardware implementations and a scheme that we term folding in the focal plane array. This method "bins" pixels modulo a pair of specified numbers across the pixel plane in both the horizontal and vertical directions. Under this paradigm, a significant reduction in the amount of pixel samples is required, which enable high speed target acquisition and tracking while reducing the number of A/D's. Folding is used to acquire a pair of significantly smaller images, in which two different folded images provide the necessary redundancy to uniquely extract location information. We detect the centroid of point sources in each of the two folded images and use the Chinese remainder theorem (CRT) to determine the location of the point sources in the original image. In our work, we successfully demonstrated the correctness of this algorithm through simulation and showed the algorithm is capable of detecting and tracking multiple muzzle flashes in multiple temporal frames. We present both initial results and improvements to the algorithm's robustness, based on robust Chinese remainder theorem (rCRT) in the presence of noise.**

I. Introduction

Draper has pursued the concept of lower-dimension signal representation from structured sparse matrices. This is in contrast to the common starting point for compressive sensing algorithms and the use of random matrices. The primary motivation is that an approach using structured matrices can lead to efficient hardware implementations in a scheme that we term focal plane array folding. Folding in the focal plane array "bins" pixels modulo a specified number across the pixel plane in both the horizontal and vertical directions.

By folding the image we drastically reduce the number of pixels from the original. A key in our approach is that we fold "twice"; that is we fold modulo a second number to give us two different folded images. This provides the necessary redundancy to uniquely extract information. We demonstrate this method on detecting and determining the location of point sources in images, one example is the muzzle flash detection.

*Dan Parker is currently a Draper Laboratory Fellow.

This research was supported in part by the DARPA SCENICC program under contract W911NF-11-C-0210.

We first detect the centroid of point sources in the two folded images. The Chinese remainder theorem (CRT) is then applied to unfold the detected centroids from which the location of the point sources can be ascertained in the original image.

The paper is organized as follows. We start with a brief background on compressive sensing, the proposed folding algorithm for reduced data acquisition, followed by recovery algorithms. We present both initial results and improvements to the algorithm's performance based on robust Chinese remainder theorem (rCRT) in the presence of noise. Preliminary experiments and results for point source detection are shown followed with discussions. Although we show an application of the folding in the focal plane array method, there are no theoretical guarantees for more complex images. It is plausible that more than $K \log(N/K)$ measurements are needed to capture arbitrary K-sparse images using folded matrices. We postulate that by extending the space of folding matrices to include constructions that have more degrees of freedom such as non-uniform folds, rotation, and skewing can lead to a reduction in the number of required measurements. A goal is to achieve $O(K \log(N/K)$ measurement bound, matching that of random matrices. Conclusions and future work in this direction are further discussed in the last section.

II. Theory

In this section, we describe the theoretical background of point source detection. The theory builds on the ideas established in star tracking [1], [2]. Here we first introduce the basic principle of compressive sensing, the folding algorithm for reduced data acquisition, followed by recovery algorithms based on Chinese remainder theorem and robust Chinese remainder theorem to locate muzzle flashes.

A. Fundamentals of CS

Compressive sensing, a novel "linear" approach for acquiring digital images, has recently been discovered [3], [4]. Traditional approaches to image acquisition first capture an entire two dimensional $(2D)$ image and then process it for compression, transmission, or storage. In contrast, CS obtains a compressed representation directly, by acquiring a small number of linear measurements of the signal in hardware. Here we treat an image or higher-dimensional data by vectorizing it into a long one-dimensional vector$x \in \Re^N$. We are solving a linear measurement system

$$y = \Phi x,$$

where the measured signal $y \in \Re^M$ and measurement matrix $\Phi \in \Re^M$. For $M < N$, recovering the signal x is an underdetermined problem. CS states that if the signal x is K-sparse in an orthonormal basis Ψ, then it can be expressed as $s = \Psi^T x$, where s has at most K nonzero elements. The linear measurement system can be rewritten as

$$y = \Phi\Psi s.$$

If the matrix $\Phi\Psi$ satisfies restricted isometry property (RIP), the signal x can still be recovered with high probability. Specifically, $M = O(K \log(N/K))$ measurements suffice to recover the (approximately) top K coefficients in the measured vector x. In the particular case where x is K-sparse, one can recover x from Φx exactly. The advantage of this architecture is that it can use fewer sensors (A/D's), and therefore enable a more efficient solution.

A common choice for Φ is a matrix where each entry is chosen independently and uniformly at random from some distribution (e.g., Gaussian, $0-1$, etc). Unfortunately, matrices with random dense patterns are not easily implementable in the focal plane array, requiring a complex network of wires. More recently. Berinde et al [5] showed that one can instead use sparse random binary matrices (with at most a logarithmic number of ones per column) while preserving the $O(K \log(N/K))$ measurement bound. However, even those matrices can lead to hardware implementation with high connection complexity.

To circumvent this issue, we use the compressive measurement scheme called folding, proposed by Gupta et al. [1], for point source detection. Instead of the most commonly used l_1 minimization for the sparse signal recovery, we use the Chinese remainder theorem and robust Chinese remainder theorem as recovery algorithms, achieving much lower running time. They are described in detail in the following.

B. Measurement using folding algorithm

The compressive measurement, based on folding, is obtained by partitioning the $n_1 \times n_2$ fully sampled muzzle flash image I into block images with dimensions $p_1 \times p_2$. Summation of all non-overlapping block images coordinate-wise yields the measurement vector m consisting of $p_1 \times p_2$ measurements. The folding algorithm can be expressed as

$$FOLD(I, p_1, p_2) = m[x_1, x_2] = \sum_{\substack{y_1 \equiv x_1 (\text{mod } p_1) \\ y_2 \equiv x_2 (\text{mod } p_2)}} I[y_1, y_2]$$

From the compressive sensing perspective, the folding algorithm maps a fully sampled $n_1 \times n_2$ image into a lower-dimensional $p_1 \times p_2$ block image. Figure 1 illustrates this folding process. This process can be viewed as acquiring a reduced amount of measurements. For muzzle flash detection, we acquire two folded images with sizes $p_1 \times p_2$ and $q_1 \times q_2$. The folding algorithm reduces the number of measurements from $n_1 \times n_2$ to $(p_1 \times p_2) + (q_1 \times q_2)$. Here the corresponding mapping matrix is binary and sparse. More importantly, the mapping is highly "local": most of the pairs of adjacent pixels (with the exception of the pairs of pixels adjacent to the fold lines) are mapped to adjacent sensors. This makes it possible to implement this mapping using low-complexity hardware, either optical (e.g., using lenslet arrays [6]) or digital (using CMOS architectures) ones.

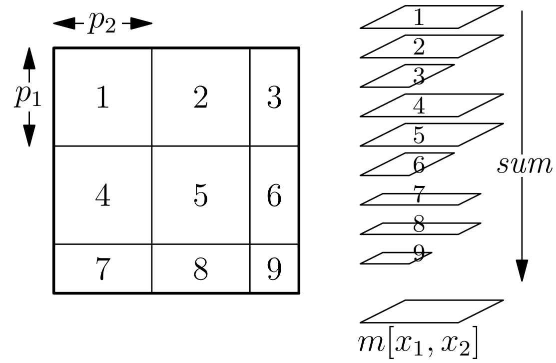

Fig. 1. The folding algorithm.

C. Recovery algorithms

As pointed out in Gupta et al.[1], the local structure of the mapping preserves local structure of the image. This made it possible to design efficient recovery algorithms, with quality of recovery comparable to that of l_1 minimization, but achieving much lower running times. In the recovery stage, we first use a peak detector to locate the centroids of muzzle flashes in two block images with different sizes $\vec{p}$ and $\vec{q}$. Then we apply recovery algorithm, Chinese Remainder Theorem, to detect the location of muzzle flashes in original unfolded image of size $\vec{n}$. Robust CRT (rCRT) was also introduced as a method to mitigate the effects of noise. We introduced CRT and rCRT below and comparisons are discussed as well.

1) Chinese Remainder Theorem: The Chinese Remainder Theorem enables one to solve simultaneous equations with respect to different moduli in considerable generality. Here is the statement of the problem that the Chinese Remainder Theorem solves.

Theorem 1. (Chinese Remainder Theorem) *Let $m_1, \cdots m_k$ be integers with $\gcd(m_i, m_j) = 1$, where $i \neq j$. Let m be the product $m = m_1 m_2 \cdots m_k$ and $a_1, \cdots, a_k$ be integers. Consider the system of congruences:*

$$x \equiv a_i (\text{mod } m_i)$$

For $1 \leq i \leq k$, there exists a unique solution modulo m for the unknown x.

The Chinese Remainder Theorem (CRT) coupled with the fact that the fold sizes $\vec{p}$ and $\vec{q}$ are coprimes, allows the recovery to take place. Applying CRT to muzzle flash detection is straightforward. The locations of the detected centroids of the folded images represent the remainders. With these remainders

TABLE I
Comparison between CRT and robust CRT (rCRT).

Original CRT	Robust CRT
Moduli are coprime	Moduli have a greatest common divisor equal to D.
Small errors in folded remainders produce large errors in reconstruction	Reconstructed errors are comparable to folded remainder errors

and the modulus operators (i.e. the fold sizes $\vec{p}$, $\vec{q}$), the location of the original centroid can be calculated. This requires two uses of the CRT in 2D images: one for the row location and one for the column location.

CRT has a simple formula and is easily be implemented. However, two major issues with the Chinese Remainder Theorem (CRT) limit its use in practical applications. First, the modulus operators must be co-prime, leading to cumbersome hardware implementations. Second, the method is very sensitive to noise; that is even small errors from remainders may cause large errors in reconstruction. For the original CRT, some additional error checking capabilities may be added to improve its performance. For example, if the detected centroid location was outside of the image boundaries, indicating an incorrect folded centroid, the process can be repeated with neighboring pixels of the incorrect centroids. This method offers a relatively inexpensive way to improve the robustness of the original CRT.

2) Robust Chinese Remainder Theorem (rCRT): A robust Chinese Remainder Theorem, such as the one introduced in [7], resolve the issues in the original CRT. This technique provides reconstruction errors comparable to the errors in the computed remainders, provided the modulus operators satisfy some mild conditions. The difference is that the modulus operators are no longer required to be co-primes, but rather have some greatest common divisor D. If the remainder errors are bounded by the value τ, then as long as $\tau < D/4$, the reconstruction error will also be bounded by τ. Thus, better error correction can be achieved with higher values of D. rCRT requires a longer computation time, but this is only noticeable when more than two modulus operators are used. Table I. lists comparison between CRT and robust CRT.

3) Recovery Steps: We follow the corresponding recovery steps:
1. In each pair of block images, we compute centroids of all objects $x_i,\ i = 1, \cdots, m$ and $y_j,\ j = 1, \cdots, n$.
2. We choose up to mn different combinations of a pair of centroids, each one from a different block image.
3. Apply the CRT/rCRT to each pair centroids from step 2 in each dimension to recover all candidate coordinates of muzzle flashes in the original image.
4. Ruling out false detections, we keep all possible coordinates, which are locations of muzzle flashes.

III. Experiments and Results

A. Experiments

We simulate two dimensional (2D) single and multiple muzzle flash images for testing our proposed algorithms. A muzzle flash image is a spatially sparse image with all pixels close to zero except for the high-valued blotch of the muzzle flashes themselves. Such an image can be created very simplistically by properly scaling a bivariate Gaussian probability density function (pdf) centered on the desired pixel locations of muzzle flashes. A single muzzle flash example is shown in Figure 2a. The folding algorithm described in the previous section is applied with two selected coprimes forming two different block image sizes (shown in Figure 2b). We use square images in our simulation, so the original image size is $n \times n$ and the set of block image size are $p \times p$ and $q \times q$. The total number of measurements, also the total number of pixels in the folded pair images, is $p^2 + q^2$, the ratio of reduced measurements to the fully sampled image can be expressed as

$$R = \frac{p^2 + q^2}{n^2}$$

We designed several experiments to demonstrate and evaluate our muzzle flash detection algorithms. In the first experiment, original CRT was applied to detect a single muzzle flash location. Two different sizes of muzzle flashes at two different locations were used to test the robustness of the CRT method. In the second experiment, to simulate a more realistic case, random distributed Gaussian noise is added to images. Original CRT and robust CRT results are compared in terms of different noise level. To compare the original and robust CRT, first a large number of candidate co-prime pairs were generated and sorted by increasing the total number of measurements. These co-prime pairs were directly used as the modulus operators for the original CRT, and in the case of rCRT, each pair is multiplied by a specified D. For each moduli pair, 100 simulations were executed to estimate the probability of detection. The performance metric for the original and robust CRT is the probability of accurate detection of the original muzzle flash centroid. Intuitively, the probability of detection should increase as the fold sizes increase, since less “noise” from other pixels is folded onto each other. In each simulation, a single muzzle flash was generated at a different location and a certain amount of white noise was added. The probability of detection, therefore, was calculated as the percentage of simulations where the centroid of the original image was detected, allowing for some error tolerance. In the third experiment, two muzzle flashes were simulated in two continuous temporal frames. Each muzzle flash moves randomly within ± 5 pixels’ radius from successive frames. Random noises with the same variance was added to each frame. For this experiment, rCRT was used for muzzle flash detection.

B. Results

Results from the original CRT are shown in Figure 2 and 3, each has a different centroid and size of muzzle flash with

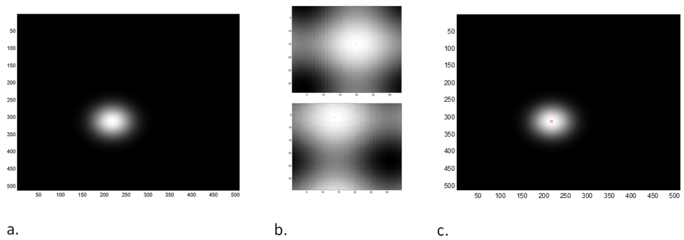

Fig. 2. Original CRT result one. Simulated muzzle flash image is shown in a, located near the middle of the image. Two folded images with size 33×33 and 34×34 are shown in b. Centroids are marked in red in individual folded images. c shows the detected muzzle flash, its centroid location, in red, was put on top of the simulated image, showing a successful detection.

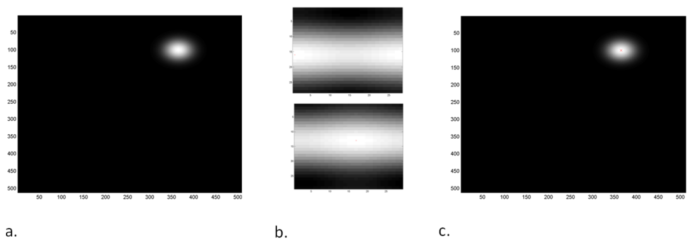

Fig. 3. Original CRT result two. Simulated muzzle flash image (shown in a) locates on the top right of the image. Two folded images with size 28×28 and 29×29 are shown in b. Two Centroids are marked in black in individual folded images. Detected muzzle flash location marked in red was overlapped with simulated image, showing a successful detection.

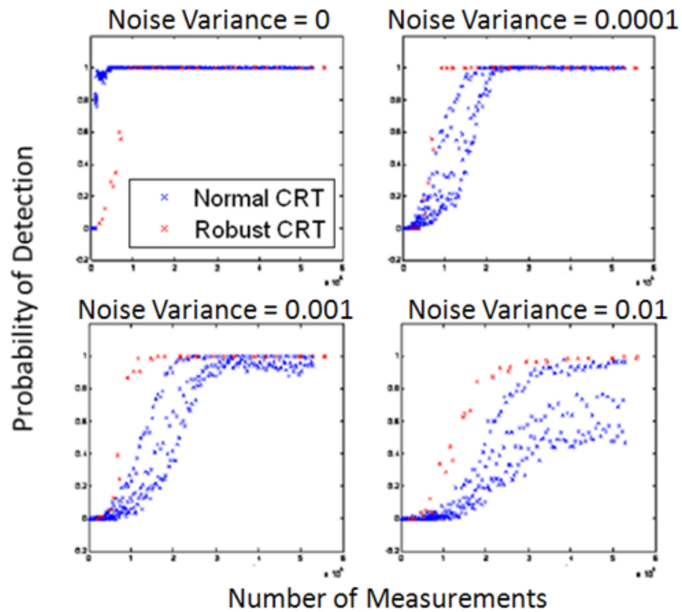

Fig. 4. Probability of detection for both methods, with noise variance increasing and $D = 9$.

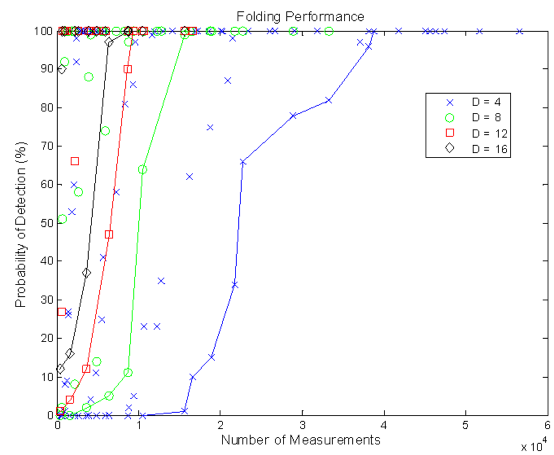

Fig. 5. rCRT performance with different D.

no added noise. Muzzle flash detection results from figure 2.c and 3.c show that the original CRT works well with the ideal noiseless case. The amount of data acquired is only 0.86% and 0.62% of traditional acquisition respectively.

Figure 4 shows the performance of both methods as the variance of added noise increases. The noiseless case shown in the top left reveals the original CRT can guarantee detection with fewer measurements when there are no errors in the computed remainders. However, as soon as noise is added to the original image, noise which compounds itself in the process of folding, the rCRT can reliably guarantee detection with fewer measurements. A greatest common divisor of 9 is used for all these simulations. Note here CRT can choose all different combinations of coprime pairs as sizes of two folded images for simulation, while rCRT can only pick the pairs which have the greatest common divisor 9, therefore rCRT has fewer number of experiments (in red) than CRT (in blue) in the figure 4.

According to [7], $D/4$ corresponds to the maximum allowable error bound for reconstruction. Figure 5 shows rCRT performance for different values of D. Again, 100 simulations were used for each moduli pair. Clearly, as D increases, detection is guaranteed with fewer measurements, and the curve is more stable overall. Note here as D goes up, there are fewer experiments since there are fewer choices for the folded images' sizes.

Figure 6 shows two muzzle flashes with two temporal frame results. The amount of data acquired is 9.67% of the original data. Since muzzle flash can not move too far way between two continuous time frames, this additional constraint may be added to further prevent false detection.

IV. Discussions

A. Folding Algorithm

Experiments shows that the detection algorithm works well in the case of a single muzzle flash. We have also shown that our algorithm can be extended to perform with multiple muzzle flashes with multiple temporal frames simulations. Despite the success of demonstrating the viability of this algorithm, there is still more study needed to improve both the robustness and stability of this algorithm. Also in the case of multiple muzzle flashes, different muzzle flashes can sometimes collide into one object in each block image. This presents a challenge in differentiating multiple muzzle flash tracks. Since

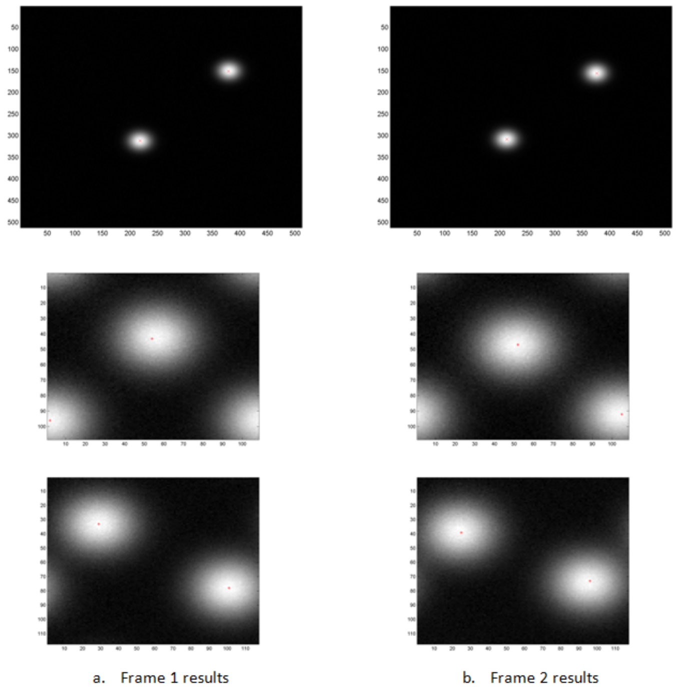

Fig. 6. Two muzzle flashes with two temporal frame results. Figure $3.5a$. shows the first frame result and Figure $3.5b$ shows the second frame result. First row shows detected muzzle flashes (locations marked in red) in original images. Corresponding two block images with centroids marked in red are shown in bottom. Percentage of data used here is 9.67% compared with the traditional methods.

the analysis in paper [1] for star tracking was experimental, the performance of folding algorithm in a practical system is not fully understood. There are no theoretical guarantees for more complex images. In fact, with increased number of muzzle flashes and temporal frames, it is plausible that more than $K \log(N/K)$ measurements are needed to capture arbitrary K-sparse images using folding matrices. A part of the reason is that folding matrices have few "degrees of freedom", which makes it hard to handle all possible "pathological" image patterns. Moreover, we have also observed that the folding matrix can have "degenerate" structure, that can lead to numerical instability for recovery procedures such as l_1 minimization. Further theoretical investigation and tests will be needed to evaluate this hypothesis.

The possible solution can be extending the space of folding matrices to include constructions that have more degrees of freedom and thus provide more design flexibility. At the same time, the constructions are still local, and therefore amenable to simple hardware implementations. Non-uniform folds (sizes of block images are selected randomly from some distribution), rotation (allow a rotation of an image before folding) or skew (an alternative to a rotation) can be investigated as extensions to the basic folding constructions. Using non-uniform folds can help in the case of images that have periodic structure, as then the folds become "incoherent" with the image. At the same time, non-uniform folds still retain the local structure of the basic construction. Using rotation enables dealing with images that consist of many vertical or horizontal stripes. In such cases folding would make many pixels overlap, which would make the recovery process difficult. In contrast, rotating an image by a "random" amount makes it resilient to the "dominating" image directions. A random skew serves a role similar to that of the random rotation, making the image more "generic".

B. Recovery Algorithm

One of the basic questions is whether using non-uniform folding and rotation can lead to $O(K \log N)$ measurement bound for arbitrary K-sparse images. It appears that a combination of the two provides enough "randomness" to ensure that any pair of pixels is mapped together with only small (about $1/M$) probability. This property is known to be a sufficient condition for a simple thresholding recovery method to work, both in theory and in practice. It would be even more interesting to achieve the $O(K \log(N/K))$ measurement bound, matching that of the random matrices. This would probably require using more advanced recovery methods, such as the l_1 minimization. We believe that the amount of randomness in the matrix should be sufficient to render the numerically unstable configurations unlikely. Moreover, using l_1 minimization would likely lead to a higher quality of the recovered image (although at a price of higher running time), which provides additional motivation for this line of research. Last but not least, it would be very interesting to design matrices and recovery algorithms that work if an image is sparse in non-standard (e.g., wavelet) basis.

V. Conclusions and Future work

In this paper, we have introduced a point source target detection algorithm based on compressive sensing. We successfully demonstrated the correctness of this algorithm through simulation and showed the algorithm is capable of detecting and localizing multiple muzzle flashes in multiple temporal frames with a reduced amount of acquired data. Furthermore, we have shown that using the robust Chinese Remainder Theorem for reconstruction provides resiliency against noise.

As addressed in discussions, we expect that adding more degrees of freedom to folding constructions and design better recovery algorithms will yield improvements both in theory and practice. Additionally, extending compressive sensing algorithm (e.g compressive particle filtering [8]) in the time dimension may be considered in this work.

References

[1] R. Gupta, P. Indyk, E. Price, and Y. Rachlin, "Compressive sensing with local geometric features," in *Symposium on Computational Geometry*, 2011, pp. 87–96.

[2] R. Gupta, "A compressive sensing algorithm for attitude determination," Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2011.

[3] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, aug 2006.

[5] R. Berinde, A. C. Gilbert, P. Indyk, H. J. Karloff, and M. J. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, 2008, p. 798.

[6] S. Uttam, N. A. Goodman, M. A. Neifeld, C. Kim, R. John, J. Kim, and D. Brady, "Optically multiplexed imaging with superposition space tracking," *Optical Express*, vol. 17, no. 3, pp. 1691–1713, 2009.

[7] X. Li, H. Liang, and X. gen Xia, "A robust chinese remainder theorem with its applications in frequency estimation from undersampled waveforms," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4314–4322, Nov 2009.

[8] E. Wang, J. Silva, and L. Carin, "Compressive particle filtering for target tracking," in *IEEE 15th Workshop on Statistical Signal Processing*, 2009, pp. 233–236.