Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
20 DEC 2009		**01-09-2008 to 31-08-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Compressive Sensing for Background Subtraction	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rice University,ECE,Houston,TX,77005**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**19**	

Note: In the following paper, there is a typo in the Acknowledgements that was not caught until after publication. After consultation with the grant reporting office (Karen McCauley via aror.reports@us.army.mil on 12/10/09) and with the technical monitor (Dr. John Lavery on 12/16/09), it was agreed to note the correction with the statement:

The acknowledgement correctly refers to the ARO MURI support, but the award number is incorrect in the first digit

"W311NF-07-1-0185"

and should be replaced by

"**W911NF-07-1-0185**".

Robert C. Sharpley {co-PI ARO MURI}
"Model Classes, Approximation, and Metrics for Dynamic Processing of Urban Terrain Data"
December 20, 2009

Compressive Sensing for Background Subtraction

Volkan Cevher[1], Aswin Sankaranarayanan[2], Marco F. Duarte[1], Dikpal Reddy[2], Richard G. Baraniuk[1], and Rama Chellappa[2]

[1] Rice University, ECE, Houston TX 77005
[2] University of Maryland, UMIACS, College Park, MD 20947

Abstract. Compressive sensing (CS) is an emerging field that provides a framework for image recovery using sub-Nyquist sampling rates. The CS theory shows that a signal can be reconstructed from a small set of random projections, provided that the signal is sparse in some basis, e.g., wavelets. In this paper, we describe a method to directly recover background subtracted images using CS and discuss its applications in some communication constrained multi-camera computer vision problems. We show how to apply the CS theory to recover object silhouettes (binary background subtracted images) when the objects of interest occupy a small portion of the camera view, i.e., when they are sparse in the spatial domain. We cast the background subtraction as a sparse approximation problem and provide different solutions based on convex optimization and total variation. In our method, as opposed to learning the background, we learn and adapt a low dimensional compressed representation of it, which is sufficient to determine spatial innovations; object silhouettes are then estimated directly using the compressive samples without any auxiliary image reconstruction. We also discuss simultaneous appearance recovery of the objects using compressive measurements. In this case, we show that it may be necessary to reconstruct one auxiliary image. To demonstrate the performance of the proposed algorithm, we provide results on data captured using a compressive single-pixel camera. We also illustrate that our approach is suitable for image coding in communication constrained problems by using data captured by multiple conventional cameras to provide 2D tracking and 3D shape reconstruction results with compressive measurements.

1 Introduction

Background subtraction is fundamental in automatically detecting and tracking moving objects with applications in surveillance, teleconferencing [1, 2] and even 3D modeling [3]. Usually, the foreground or *the innovation* of interest occupies a sparse spatial support, as compared to the background and may be caused by the motion and the appearance change of objects within the scene. By obtaining the object silhouettes on a single image plane or multiple image planes, a background subtraction algorithm can be performed.

In all applications that require background subtraction, the background and the test images are typically fully sampled using a conventional camera. After the foreground estimation, the remaining background images are either discarded or embedded back into the background model as part of a learning scheme [2]. This sampling process is

inexpensive for imaging at the visible wavelengths as the conventional devices are built from silicon, which is sensitive to these wavelengths; however, if sampling at other optical wavelengths is desired, it becomes quite expensive to obtain estimates at the same pixel resolution as new imaging materials are needed. For example, a camera with an array of infrared sensors can provide night vision capability but can also cost significantly more than the same resolution CCD or CMOS cameras.

Recently, a prototype single pixel camera (SPC) was proposed based on the new mathematical theory of *compressive sensing* (CS) [4]. The CS theory states that a signal can be perfectly reconstructed, or can be robustly approximated in the presence of noise, with sub-Nyquist data sampling rates, provided that it is *sparse* in some linear transform domain [5, 6]. That is, it has K nonzero transform coefficients with $K \ll N$, where N is the dimension of the transform space. For computer vision applications, it is known that natural images can be sparsely represented in the wavelet domain [7]. Then, according to the CS theory, by taking random projections of a scene onto a set of test functions that are incoherent with the wavelet basis vectors, it is possible to recover the scene by solving a convex optimization problem. Moreover, the resulting *compressive measurements* are robust against packet drops over communication channels with graceful degradation in reconstruction accuracy, as the image information is fully distributed.

Compared to conventional camera architectures, the SPC hardware is specifically designed to exploit the CS framework for imaging. An SPC fundamentally differs from a conventional camera by (i) reconstructing an image using only a single optical photodiode (infrared, hyperspectral, etc.) along with a digital micromirror device (DMD), and (ii) combining the sampling and compression into a single nonadaptive linear measurement process. An SPC can directly scale from the visual spectra to hyperspectral imaging with only a change of the single optical sensor. Moreover, enabled by the CS theory, an SPC can robustly reconstruct the scene from much fewer measurements than the number of reconstructed pixels which define the resolution, given that the image of the scene is compressible by an algorithm such as the wavelet-based JPEG 2000.

Conventional cameras can also benefit by processing in the compressive sensing domain if their data is being sent to a central processing location. The naïve approach is to transmit the raw images to the central location. This exacerbates the communication bandwidth requirements. In more sophisticated approaches, the cameras transmit the information within the background subtracted image, which requires an even smaller communication bandwidth than the compressive samples. However, the embedded systems needed to perform reliable background subtraction are power hungry and expensive. In contrast, the compressive measurement process only requires cheaper embedded hardware to calculate inner products with a previously determined set of test functions. In this way, the compressive measurements require comparable bandwidth to transform coding of the raw data. They trade off expensive embedded intelligence for more computational power at the central location, which reconstructs the images and is assumed to have unlimited resources.

The communication bandwidth and camera hardware limitations make it desirable to directly reconstruct the sparse foreground innovations within a scene without any intermediate image reconstruction. The main idea is that the background subtracted im-

ages can be represented sparsely in the spatial image domain and hence the CS reconstruction theory should be applicable for directly recovering the foreground. For natural images, we use wavelets as the transform domain. Pseudo-random matrices provide an incoherent set of test functions to recover the foreground image. Then, the following questions surface (i) how can we detect targets without reconstructing an image? and (ii) how can we directly reconstruct the foreground without reconstructing auxiliary images?

In this paper, we describe a method based on CS theory to directly recover the sparse innovations (foreground) of a scene. We first show that the object silhouettes (binary background subtracted images) can be recovered as a solution of a convex optimization or an orthogonal matching pursuit problem. In our method, the object silhouettes are learned directly using the compressive samples without any auxiliary image reconstruction. We then discuss simultaneous appearance recovery of objects using the compressive measurements. In this case, we show that it may be necessary to reconstruct one auxiliary image. To demonstrate the performance of the proposed algorithm, we use field data captured by a compressive camera and provide background subtraction results. We also show results on field data captured by conventional CCD cameras to simulate multiple distributed single-pixel cameras and provide 2D tracking and 3D shape reconstruction results.

While the idea of performing background subtraction on compressed images is not novel, there exist no cameras that record MPEG video directly. Both Aggarwal et al. [8] and Lamarre and Clark [9] perform background subtraction on a MPEG-compressed video using the DC-DCT coefficients of I frames, limiting the resolution of the BS images by 64. Our technique is tailored for CS imaging, and not compressed video files. Lamarre et al. [9] and Wang et al. [10] use DCT coefficients from JPEG pictures and MPEG videos, respectively, for representation. Toreyin et al. [11] similarly operate on the wavelet representation. These methods implicitly perform decompression by working on every DCT/wavelet coefficient of every image. We never have to go to the high dimensional images or representations during background subtraction, making our approach particularly attractive for embedded systems and demanding communication bandwidths. Compared to the eigenbackground work of Oliver et al. [12], random projections are universal so there is no need to update bases - the only basis needed is the sparsity basis for difference images, hence no training is required. The very recent work of Uttam, Goodman and Neifeld [13] considers background subtraction from adaptive compressive measurements, with the assumption that the background-subtracted images lie in a low-dimensional subspace. While this assumption is acceptable when image tiling is performed, background-subtracted images are sparse in an appropriate domain, spanning a union of low-dimensional subspaces rather than a single subspace.

Our specific contributions are as follows:

1. We cast the background subtraction problem as a sparse signal recovery problem where convex optimization and greedy methods can be applied. We employ Basis Pursuit Denoising methods [14] as well as total variation minimization [5] as convex objectives to process field data.
2. We show that it is possible to recover the silhouettes of foreground objects by learning a low-dimensional compressed representation of the background image. Hence,

we show that it is not necessary to learn the background itself to sense the innovations or the foreground objects. We also explain how to adapt this representation so that our approach is robust against variations of the background such as illumination changes.

3. We develop an object detector directly on the compressive samples. Hence, no foreground reconstruction is done until a detection is made to save computation.

2 The Compressive Sensing Theory

2.1 Sparse Representations

Suppose that we have an image $\boldsymbol{X}$ of size $N_1 \times N_2$ and we vectorize it into a column vector $\boldsymbol{x}$ of size $N \times 1$ ($N = N_1 N_2$) by concatenating the individual columns of $\boldsymbol{X}$ in order. The nth element of the image vector $\boldsymbol{x}$ is referred to as $x(n)$, where $n = 1, \ldots, N$. Let us assume that the basis $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N]$ provides a K-sparse representation of $\boldsymbol{x}$:

$$\boldsymbol{x} = \sum_{n=1}^{N} \theta(n) \boldsymbol{\psi}_n = \sum_{l=1}^{K} \theta(n_l) \boldsymbol{\psi}_{n_l}, \tag{1}$$

where $\theta(n)$ is the coefficient of the nth basis vector $\boldsymbol{\psi}_n$ ($\boldsymbol{\psi}_n$: $N \times 1$) and the coefficients indexed by n_l are the K-nonzero entries of the basis decomposition. Equation (1) can be more compactly expressed as follows

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\theta}, \tag{2}$$

where $\boldsymbol{\theta}$ is an $N \times 1$ column vector with K-nonzero elements. Using $\|\cdot\|_p$ to denote the ℓ_p norm where the ℓ_0 norm simply counts the nonzero elements of $\boldsymbol{\theta}$, we call an image $\boldsymbol{X}$ as K-sparse if $\|\boldsymbol{\theta}\|_0 = K$.

Many different basis expansions can achieve sparse approximations of natural images, including wavelets, Gabor frames, and curvelets [5, 7]. In other words, a natural image does not result in an exactly K-sparse representation; instead, its transform coefficients decay exponentially to zero. The discussion below also applies to such images, denoted as compressible images, as they can be well-approximated using the K largest terms of $\boldsymbol{\theta}$.

2.2 Random/Incoherent Projections

In the CS framework, it is assumed that the K-largest $\theta(n)$ are not measured directly. Rather, $M < N$ linear projections of the image vector $\boldsymbol{x}$ onto another set of vectors $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_M']'$ are measured:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta}, \tag{3}$$

where the vector $\boldsymbol{y}$ ($M \times 1$) constitutes the compressive samples and the matrix $\boldsymbol{\Phi}$ ($M \times N$) is called the *measurement matrix*. Since $M < N$, recovery of the image $\boldsymbol{x}$ from the compressive samples $\boldsymbol{y}$ is underdetermined; however, as we discuss below, the additional *sparsity* assumption makes recovery possible.

The CS theory states that when (i) the columns of the sparsity basis $\boldsymbol{\Psi}$ cannot sparsely represent the rows of the measurement matrix $\boldsymbol{\Phi}$ and (ii) the number of measurements M is greater than $\mathcal{O}\left(K \log\left(\frac{N}{K}\right)\right)$, then it is possible to recover the set of nonzero entries of $\boldsymbol{\theta}$ from $\boldsymbol{y}$ [5, 6]. Then, the image $\boldsymbol{x}$ can be obtained by the linear transformation of $\boldsymbol{\theta}$ in (1). The first condition is called the incoherence of the two bases and it holds for many pairs of bases, e.g., delta spikes and the sine waves of the Fourier basis. Surprisingly, incoherence also holds with high probability between an arbitrary basis and a randomly generated one, e.g., i.i.d. Gaussian or Bernoulli/Rademacher ± 1 vectors.

2.3 Signal Recovery via ℓ_1 Optimization

There exists a computationally efficient recovery method based on the following ℓ_1-optimization problem [5, 6]:

$$\widehat{\boldsymbol{\theta}} = \arg\min \|\boldsymbol{\theta}\|_1 \quad \text{s. t. } \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta}. \tag{4}$$

This optimization problem, also known as *Basis Pursuit* [6], can be efficiently solved using polynomial time algorithms.

Other formulations are used for recovery from noisy measurements such as Lasso, Basis Pursuit with quadratic constraint [5]. In this paper, we use Basis Pursuit Denoising (BPDN) for recovery:

$$\widehat{\boldsymbol{\theta}} = \arg\min \|\boldsymbol{\theta}\|_1 + \frac{1}{2}\beta\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2, \tag{5}$$

where $0 < \beta < \infty$ [14]. When the images of interest are smooth, a strategy based on minimizing the total variation of the image works equally well [5].

3 CS for Background Subtraction

With background subtraction, our objective is to recover the location, shape and (sometimes) appearance of the objects given a test image over a known background. Let us denote the background, test, and difference images as $\boldsymbol{x}_b$, $\boldsymbol{x}_t$, and $\boldsymbol{x}_d$, respectively. The difference image is obtained by pixel-wise subtraction of the background image from the test image. Note that the support of $\boldsymbol{x}_d$, denoted as $\mathcal{S}_d = \{n | n = 1, \ldots, N; |\boldsymbol{x}_d(n)| \neq 0\}$, gives us the location and the silhouettes of the objects of interest, but not their appearance (see Fig. 1).

3.1 Sparsity of Background Subtracted Images

Suppose that $\boldsymbol{x}_b$ and $\boldsymbol{x}_t$ are typical real-world images in the sense that when wavelets are used as the sparsity basis for $\boldsymbol{x}_b$, $\boldsymbol{x}_t$, and $\boldsymbol{x}_d$, these images can be well approximated with the largest K coefficients with hard thresholding [15], where K is the corresponding sparsity proportional to the cardinality of the image support. The images $\boldsymbol{x}_b$ and $\boldsymbol{x}_t$ differ only on the support of the foreground, which has a cardinality of $P = |\mathcal{S}_d|$ pixels with $P \ll N$. Moreover, we assume that images have uniform complexity in space. We model the sparsity of the real world images as a function of their

Fig. 1. (*Left*) Example background image. (*center*) Test image. (*Right*) Difference image. Note that the vehicle appearance also shows the curb in the background, which it occludes. The images and are from the PETS 2001 database.

size: $K_{\text{scene}} = K_b = K_t = (\lambda_0 \log N + \lambda_1)N$, where $(\lambda_0, \lambda_1) \in \mathbb{R}^2$. We assume that the difference image is also a real-world image on a restricted support (see Fig. 1(c)), and similarly we approximate its sparsity as $K_d = (\lambda_0 \log P + \lambda_1)P$.

The number of compressive samples M necessary to reconstruct $\boldsymbol{x}_b$, $\boldsymbol{x}_t$, and $\boldsymbol{x}_d$ in N dimensions are then given by $M_{\text{scene}} = M_b = M_t \approx K_{\text{scene}} \log(N/K_{\text{scene}})$ and $M_d \approx K_d \log(N/K_d)$. When $M_d < M_{\text{scene}}$, a smaller number of samples is needed to reconstruct the difference image than the background or foreground images. We empirically show in Section 5 that this condition is almost always satisfied when the sizes of the difference images are smaller than original image sizes for natural images.

3.2 The Background Constraint

Let us assume that we have multiple compressive measurements $\boldsymbol{y}_{bi}$ ($M \times 1$, $i = 1, \ldots, B$) of training background images $\boldsymbol{x}_{bi}$, where $\boldsymbol{x}_b$ is their mean. Each compressive measurement is a random projection of the whole image, whose distribution we approximate as an i.i.d. Gaussian distribution with a constant variance $\boldsymbol{y}_{bi} \sim \mathcal{N}\left(\boldsymbol{y}_b, \sigma^2 \boldsymbol{I}\right)$, where the mean value is $\boldsymbol{y}_b = \boldsymbol{\Phi}\boldsymbol{x}_b$. When the scene changes to include an object which was not part of the background model and we take the compressive measurements, we obtain a test vector $\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{x}_t$, where $\boldsymbol{x}_d = \boldsymbol{x}_t - \boldsymbol{x}_b$ is sparse in the spatial domain.

In general, the sizes of the foreground objects are relatively smaller than the size of the background image; hence, we model the distribution of the *literally* background subtracted vector as $\boldsymbol{y}_d = \boldsymbol{y}_t - \boldsymbol{y}_b \sim \mathcal{N}\left(\boldsymbol{\mu}_d, \sigma^2 \boldsymbol{I}\right)$ ($M \times 1$), where $\boldsymbol{\mu}_d$ is the mean. Note that the appearance of the objects constructed from the samples $\boldsymbol{y}_d$ would correspond to the literal subtraction of the test frame and the background; however, their silhouette is preserved (Fig. 1(c)).

The number of samples M in $\boldsymbol{y}_b$ is greater than M_d as discussed in Sect. 3.1, but is not necessarily greater than or equal to M_b or M_t; hence, it may not be sufficient to reconstruct the background. However, the background image $\boldsymbol{x}_b$ still satisfies the constraint $\boldsymbol{y}_b = \boldsymbol{\Phi}\boldsymbol{x}_b$. To be robust against small variations in the background and noise, we consider the distribution of the ℓ_2 distances of the background frames around their mean $\boldsymbol{y}_b$:

$$\|\boldsymbol{y}_{bi} - \boldsymbol{y}_b\|_2^2 = \sigma^2 \sum_{n=1}^{M} \left(\frac{y_{bi}(n) - y_b(n)}{\sigma} \right)^2 . \tag{6}$$

When M is greater than 30, this sum can be well approximated by a Gaussian distribution due to the central limit theorem. Then, it is straightforward to show that we have $\|\boldsymbol{y}_{bi}-\boldsymbol{y}_b\|_2^2 \sim \mathcal{N}\left(M\sigma^2, 2M\sigma^4\right)$. When we have a test frame with a foreground object, the same distribution becomes $\|\boldsymbol{y}_t-\boldsymbol{y}_b\|_2^2 \sim \mathcal{N}\left(M\sigma^2+\|\boldsymbol{\mu}_d\|_2^2, 2M\sigma^4+4\sigma^2\|\boldsymbol{\mu}_d\|_2^2\right)$.

Since σ^2 scales the whole distribution and $1/M \ll 1$, the logarithm of the ℓ_2 distances in (6) can be approximated quite accurately with a Gaussian distribution. That is, since $u \ll 1$ implies $1+u \approx \mathrm{e}^u$, we have $\mathcal{N}\left(M\sigma^2, 2M\sigma^4\right) = M\sigma^2\mathcal{N}\left(1, \frac{2}{M}\right) = M\sigma^2\left(1+\sqrt{\frac{2}{M}}\mathcal{N}\left(0,1\right)\right) \approx M\sigma^2\exp\left\{\sqrt{\frac{2}{M}}\mathcal{N}\left(0,1\right)\right\}$. This derivation can also motivated by the fact that the square-root of the Chi-squared distribution can be well approximated by a Gaussian [16].

Hence, (6) can be used to approximate

$$\log\|\boldsymbol{y}_{bi}-\boldsymbol{y}_b\|_2^2 \sim \mathcal{N}\left(\mu_{bg}, \sigma_{bg}^2\right), \tag{7}$$

where μ_{bg} is the mean and σ_{bg}^2 is the variance term, which does not depend on the additive noise in pixel measurements. Equation (7) allows some variability around the constraint $\boldsymbol{y}_b = \boldsymbol{\Phi}\boldsymbol{x}_b$ that the background image needs to satisfy in order to cope with the small variations of the background and the measurement noise. However, the samples $\boldsymbol{y}_d = \boldsymbol{y}_t - \boldsymbol{y}_b$ can be used to recover the foreground objects. We learn the log-Normal parameters in (7) from the data using maximum likelihood techniques.

3.3 Object Detector based on CS

Before we attempt any reconstruction, it is a good idea to determine if the test image has any differences from the background. Using the results from Sect. 3.2, the ℓ_2 distance of $\boldsymbol{y}_t$ from $\boldsymbol{y}_b$ can be subsequently approximated by

$$\log\|\boldsymbol{y}_t-\boldsymbol{y}_b\|_2^2 \sim \mathcal{N}\left(\mu_t, \sigma_t^2\right). \tag{8}$$

When the object is small, σ_t^2 should be on the same order size of σ_{bg}^2, while μ_t is different from μ_{bg} in (7). Then, to test the hypothesis of whether there is a new object, the optimal detector would be a simple threshold test since we would be comparing two Gaussian distributions with similar variances. When σ_t^2 is significantly different from σ_{bg}^2, the optimal test can be a two sided threshold test [17]. For our case, we simply use a constant times the standard deviation of the background as a threshold and declare that there is a new object if $\left|\log\|\boldsymbol{y}_t-\boldsymbol{y}_b\|_2^2 - \mu_{bg}\right| \geq c\sigma_{bg}$.

3.4 Foreground Reconstruction

For foreground reconstruction, we use BPDN with a fixed point continuation method [18] and total variation (TV) optimization with an interior point method [5] on the background subtracted compressive measurements. The BPDN solver is the fastest among the proposed algorithms because it solves an unconstrained optimization problem. During the reconstruction, we lose the actual appearance of the objects as the obtained measurements also contain information about the background. Although it is known that the subtracted image is a sum of two components that exclusively appear in $\boldsymbol{x}_b$ and $\boldsymbol{x}_t$, it is difficult, if not impossible, to unmix them without taking enough measurements to

recover $\boldsymbol{x}_b$ or $\boldsymbol{x}_t$. Hence, if the appearances of the objects are needed, a straightforward way to obtain them would be to either reconstruct the test image by taking enough compressive samples and then use the binary foreground image as a mask, or reconstruct and mask the background image and then add the result to the foreground estimate.

3.5 Adaptation of the Background Constraint

We define two types of changes in a background: drifts and shifts. A background drift consists of gradual changes that occur in the background such as illumination changes in the scene and may result in immediate unwanted foreground estimates. A background shift is a major and sudden change in the definition of the background, such as a new vehicle parked within the scene. Adapting to background shifts at the sensing level is quite difficult because high level logical operations are required, such as detecting the new object and deciding that it is uninteresting. However, adapting to background drifts is essential for a robust background subtraction system as it has immediate impacts on the foreground recovery.

The background constraint $\boldsymbol{y}_b$ needs to be updated continuously if the background subtraction system is to be robust against the background drifts. Otherwise, the drifts may accumulate and trigger unwanted detections. In the compressive sensing framework, this can be done as follows. Once we obtain an estimate of the difference image $\widehat{\boldsymbol{x}}_d$ with one of the reconstruction algorithms discussed in the previous section, we determine the compressive samples that should be generated by it: $\widehat{\boldsymbol{y}}_d = \boldsymbol{\Phi}\widehat{\boldsymbol{x}}_d$. Since we already have $\boldsymbol{y}_d = \boldsymbol{y}_t - \boldsymbol{y}_b$, we can substitute the de-noised difference estimate to obtain the background estimate of the current frame: $\widehat{\boldsymbol{y}}_b = \boldsymbol{y}_t - \widehat{\boldsymbol{y}}_d$. Then, a running average can be used to update the background with a learning rate of $\alpha \in (0, 1)$ as follows:

$$\boldsymbol{y}_b^{\{j+1\}} = \alpha \left(\boldsymbol{y}_t^{\{j\}} - \widehat{\boldsymbol{y}}_d^{\{j\}} \right) + (1 - \alpha) \boldsymbol{y}_b^{\{j\}}, \tag{9}$$

where j is the time index.

Unfortunately, this update rule does not suffice for compensating background shifts, such as new stationary targets. Consider a pixel whose intensity value changes because of a background shift. This pixel will then be identified as an outlier in the background model. The corresponding pixel in the background model will not be updated in (9). Hence, for all future frames, the pixel will continue to be classified as part of the foreground. This problem can be handled by allowing for a second moving average of the frames, which updates all pixels within the image as in [19].

Hence, we use the following updates:

$$\begin{aligned} \boldsymbol{y}_{\text{ma}}^{\{j+1\}} &= \gamma \boldsymbol{y}_t^{\{j\}} + (1 - \gamma) \boldsymbol{y}_{\text{ma}}^{\{j\}}, \\ \boldsymbol{y}_b^{\{j+1\}} &= \alpha \left(\boldsymbol{y}_t^{\{j\}} - \widehat{\boldsymbol{y}}_e^{\{j\}} \right) + (1 - \alpha) \boldsymbol{y}_b^{\{j\}}, \end{aligned} \tag{10}$$

where $\boldsymbol{y}_{\text{ma}}$ is the simple moving average, $\gamma \in (0, 1)$ is the moving average learning rate, and $\widehat{\boldsymbol{y}}_e = \boldsymbol{\Phi}\widehat{\boldsymbol{x}}_{\text{ma}}$. Consider a global illumination change. The moving average update integrates the pixel's illumination change over time, whose speed depends on γ. In subsequent frames, the value of the moving average will approach the intensity value

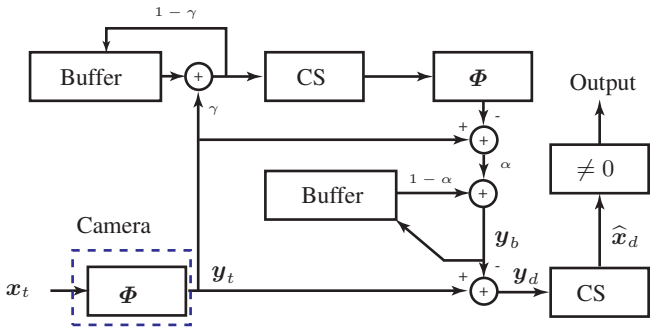

Fig. 2. Block diagram of the proposed method.

observed at the pixel. This implies that when used as a detection image, the moving average will stop detecting the pixel as foreground. Once this happens, the pixel will be updated in the background update, making the background model adaptive to global changes in illumination. A disadvantage of this approach is that if the targets stay stationary for extended periods of time, they become part of the background. However, if they move again, they can be detected. Figure 2 illustrates the outline of the proposed background subtraction method.

4 Limitations

In this section, we discuss some of the limitations of the specific compressive sensing approach to the background subtraction presented in this paper. Some of these limitations can be caused by the hardware architecture, whereas others are due to our image models. Note that our formulation is general enough that we do not require an SPC for operation. CS can be used for rateless coding of BS images. If a centralized vision system is used with no background subtraction at the camera, then our methods can be used at conventional cameras for processing in the compressive domain to reduce communication bandwidth and be robust against packet drops.

The SPC architecture uses a DMD to generate a random sampling pattern and sends the resulting inner product of the incident light field from the scene with the random pattern to the optical sensor to create a compressive measurement. By changing the random pattern in time, a set of M consecutive measurements can be made about the scene using the same optical sensor, which form the measurement vector $\boldsymbol{y}$. The current DMD arrays can change their geometric configuration approximately 10 to 40K times per second. For example, with a rate of 30K times per second, we can construct at most a 300×300 resolution background subtracted image with 1% compression ratios at 30fps. Although the resolution may not be sufficient for some applications, it will improve as the capabilities of the DMD arrays increase.

In our background modeling, we assume that the background and foreground images exhibit sparsity. We argued that the background subtracted image has a lower sparsity and hence can be reconstructed with fewer samples that is necessary to reconstruct the background or the foreground images. When the images of interest do not show

sparsity (e.g., they are white noise), our approach can still be applied. That is, the difference image $\boldsymbol{x}_d$ is always sparse regardless of the sparsities of $\boldsymbol{x}_b$ and $\boldsymbol{x}_t$ if its support cardinality P is much smaller than N.

5 Experiments

5.1 Background Subtraction with an SPC

We performed background subtraction experiments with an SPC; in our test, the background $\boldsymbol{x}_b$ consists of the standard test *Mandrill* image, with the foreground $\boldsymbol{x}_t$ consisting of a white rectangular patch as shown in Fig. 3. Both the background and the foreground were acquired using pseudorandom compressive measurements ($\boldsymbol{y}_b$ and $\boldsymbol{y}_t$, respectively) generated by a Mersenne Twister algorithm with a 64×64 pixel resolution [20]. We obtain measurements for the subtraction image as $\boldsymbol{y}_d = \boldsymbol{y}_t - \boldsymbol{y}_b$. We reconstructed both the background, test, and difference images, using TV minimization. The reconstruction is performed using several measurement rates ranging from 0.5% to 50%. In each case, we compare the subtraction image reconstruction with the difference between the reconstructed test and background images. The resulting images are shown in Fig. 3, and show that for low rates the background and test images are not recovered accurately, and therefore the subtraction performs poorly; however, the sparser foreground innovation is still recovered correctly from the difference of the measurements, with rates as low as 1% being able to recover the foreground at this low resolution.

5.2 The Sparsity Assumption

In our formulation, we assumed that the sparsity of natural images has the following form: $K = (\lambda_0 \log N + \lambda_1)N$. To test this assumption, we used the Berkeley Segmentation Data Set (BSDS) as a natural image database [21] and obtained wavelet approximations of various block sizes varying from 2×2 to 256×256 pixels. To approximate the sparsity K of any given tile size, we determined the minimum number of wavelet coefficients that results in a compression with -40dB distortion with respect to the image itself. Figure 4 shows that our sparsity assumption is justified for natural images, and illustrates that the necessary number of compressive samples is monotonic with the tile size. Therefore, if the innovations in the image are smaller than the image, it takes fewer compressive samples to recover them. In fact, the total number of samples necessary to reconstruct is rather close to linear: $M \approx \kappa N^{1-\delta}$ where $\delta \ll 1$. In general, the λ parameters are scene specific (Fig. 4(*Right*)). Hence, the exact number of compressive measurements needed may vary.

5.3 Multi-view Ground Plane Tracking

Background subtraction forms an important pre-processing component for many vision applications. In this regard, it is important to see if the imagery generated using compressive measurements can be used in such applications. In this section, we demonstrate a multi-view tracking application where accurate background subtraction is key in determining overall system performance.

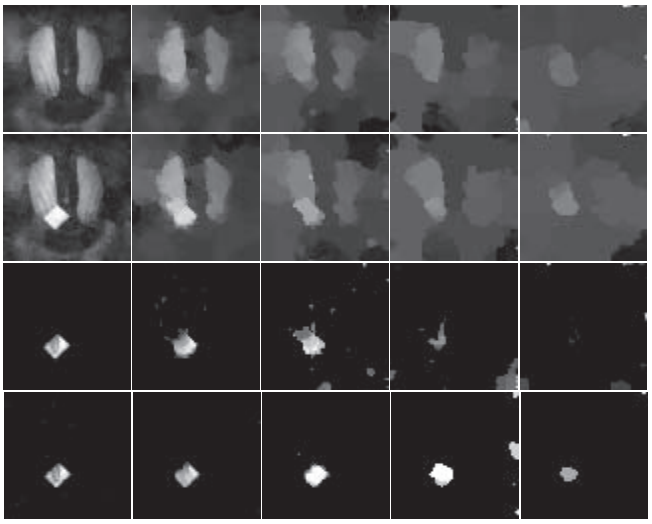

Fig. 3. Background subtraction experimental results using an SPC. Reconstruction of background image (top row) and test image (second row) from compressive measurements. Third row: conventional subtraction using the above images. Fourth row: reconstruction of difference image directly from compressive measurements. The columns correspond to measurement rates M/N of 50%, 5%, 2%, 1% and 0.5%, from left to right. Background subtraction from compressive measurements is feasible at lower measurement rates than standard background subtraction.

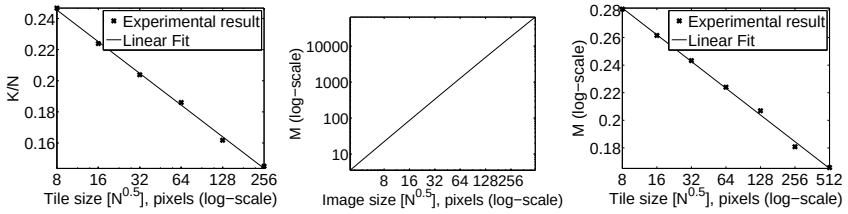

Fig. 4. (*Left*) Average sparsity over N as a function of the tile size for the images in BSDS. (center) Number of compressive measurements needed to reconstruct an image of different sizes from BSDS. (*Right*) Average sparsity over N as a function of the tile size for the images in PETS 2001 data set.

In Figure 5, we show results on a multi-view ground plane tracking algorithm over a sequence of 300 frames with 20% compression ratio. We first obtain the object silhouettes using the compressive samples at each view. We use wavelets as the sparsifying basis $\boldsymbol{\Psi}$. At each time instant, the silhouettes are mapped on to the ground planes and averaged. Objects on the ground plane (e.g., the feet) combine in synergy while those off the plane are in parallax and do not support each other. We then threshold to obtain potential target locations as in [22]. The outputs indicate the background subtracted images are sufficient to generate detections that compare well against the detections generated using the full non-compressed images. Hence, using our method, the com-

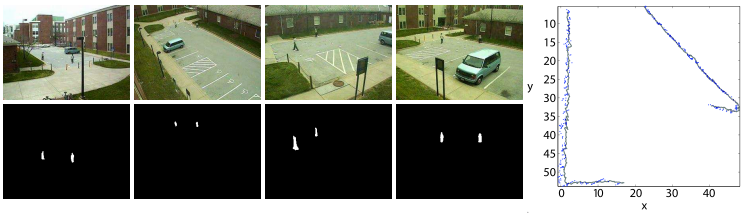

Fig. 5. Tracking results on a video sequence of 300 frames. (*Left*) The first two rows show sample images and background subtraction results using the compressive measurements, respectively. The background subtracted blobs are used to detect target location on the ground plane. The right figure shows the detected points using CS (blue dots) as well as the detected points using full images (black). The distances are in meters.

munication bandwidth of a multi camera localization system can be reduced to one-fifth if the estimation is done at a central location.

5.4 Adaptation to Illumination Changes

To compare the performance of the background constraint adaptations (9) (drift adaptive) and (10) (shift adaptive), we test them on a sequence where there is a global illumination change due to sunlight. To emphasize the differences, we use the delta basis (0/1 in spatial domain) as the sparsifying basis $\boldsymbol{\Psi}$. This basis creates much noisier background subtraction images than wavelets, but it is quite illustrative for the purposes of this comparison.

Figure 6 shows the results of the comparison. The images on top are the original images. The middle row corresponds to the update in (10) whereas the bottom row images correspond to the update in (9). The update in (10) allows the background constraint to keep track of the sudden change in illumination. Hence, the resulting images are cleaner and continue to improve. This results in much lower false alarm rates for the same detection probability (see Fig. 6(*Right*)). For the receiver operating characteristics (ROC) curves, we use the full images, run the background subtraction algorithm proposed in [19], and obtain baseline background subtracted images. We then compare the pixels on the resulting target from different updates to calculate the detection rate. We also compare the spurious detections in the rest of the images to generate the ROC curve.

5.5 Silhouettes vs. Difference Images

We have used a multi camera set up for a 3D voxel reconstruction using the compressive measurements. Figure 7(*Left*) shows the ground truth and the difference image reconstructed using CS, which incorporates elements from the background, such as the camera setup behind the subject, affecting the final reconstruction. Hence, the difference images do not always result in the desired silhouettes. Figure 7(*Right*) shows the voxel reconstruction with four cameras with 40% compression, which is visually satisfactory despite the artifacts in the difference images.

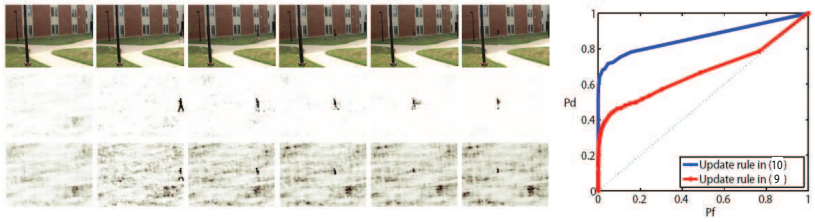

Fig. 6. Background subtraction results on a sequence with changing illumination using (9) and (10) for background constraint updates. Outputs are shown with identical parameters used for both models. Note that for the same detection output, the update rule (10) produces much less false alarm. However, (10) has twice the computational cost as (9).

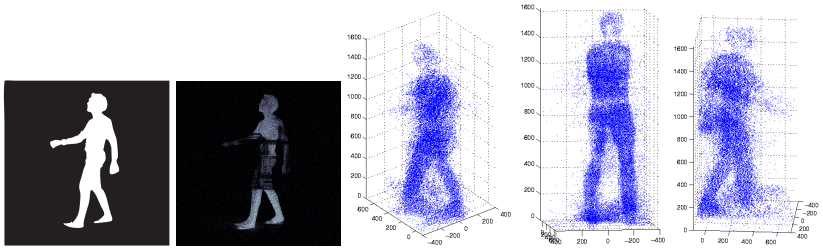

Fig. 7. (*Left*) Ground truth detections marked in white and unthresholded background difference image reconstruction using compressive samples with 40% compression. (*Right*) Reconstructed 3D point clouds of the target.

6 Conclusions

We demonstrated that the CS framework can be used to directly reconstruct sparse innovations on a background scene with a significantly fewer data samples than the conventional methods. As opposed to acquiring the minimum amount of measurements to recover a background and the test image, we can exploit the sparsity of the foreground to perform background subtraction by using even fewer measurements (M_d measurements as opposed to M_b). We illustrated that due to the linear nature of the measurements, it is still possible to adapt to the changes in the background directly in the compressive domain. In addition, it is possible to formulate an object detector. By exploiting sparsity in background subtracted images in multi-view tracking and 3D reconstruction problems, we can reduce sampling costs and alleviate communication and storage burdens while obtaining comparable estimation performance.

Acknowledgements We would like to thank Kevin Kelly and Ting Sun for collecting and providing experimental data, and Nathan Goodman for providing us with a preprint of [13]. VC, MFD and RGB were supported by the grants NSF CCF-0431150, ONR N00014-07-1-0936, AFOSR FA9550-07-1-0301, ARO W911NF-07-1-0502, ARO MURI W311NF-07-1-0185, and the Texas Instruments Leadership University Program. AS, DR and RC were partially supported by Task Order 89, Army Research Laboratory Contract DAAD19-01-C-0065 monitored by Alion Science and Technology.

References

1. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: IEEE FRAME-RATE Workshop, Springer (1999)
2. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics. Volume 4. (2004)
3. Cheung, G.K.M., Kanade, T., Bouguet, J.Y., Holler, M.: Real time system for robust 3 D voxel reconstruction of human motions. In: CVPR. (2000) 714–720
4. Wakin, M.B., Laska, J.N., Duarte, M.F., Baron, D., Sarvotham, S., Takhar, D., Kelly, K.F., Baraniuk, R.G.: An architecture for compressive imaging. In: ICIP, Atlanta, GA (Oct. 2006) 1273–1276
5. Candes, E.: Compressive sampling. In: Proceedings of the International Congress of Mathematicians. (2006)
6. Donoho, D.L.: Compressed Sensing. IEEE Trans. Info. Theory **52**(4) (2006) 1289–1306
7. Mallat, S., Zhang, S.: Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Processing **41**(12) (Dec. 1993) 3397–3415
8. Aggarwal, A., Biswas, S., Singh, S., Sural, S., Majumdar, A.K.: Object Tracking Using Background Subtraction and Motion Estimation in MPEG Videos. In: ACCV, Springer (2006) 121–130
9. Lamarre, M., Clark, J.J.: Background subtraction using competing models in the block-DCT domain. In: ICPR. (2002)
10. Wang, W., Chen, D., Gao, W., Yang, J.: Modeling background from compressed video. In: IEEE Int. Workshop on VSPE of TS. (2005) 161–168
11. Töreyin, B.U., Çetin, A.E., Aksay, A., Akhan, M.B.: Moving object detection in wavelet compressed video. Signal Processing: Image Communication **20**(3) (2005) 255–264
12. Oliver, N., Rosario, B., Pentland, A.: A Bayesian Computer Vision System for Modeling Human Interactions. In: ICVS, Springer (1999)
13. Uttam, S., Goodman, N.A., Neifeld, M.A.: Direct reconstruction of difference images from optimal spatial-domain projections. In: Proc. SPIE. Volume 7096., San Diego, CA (Aug. 2008)
14. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic Decomposition by Basis Pursuit. SIAM Journal on Scientific Computing **20** (1998) 33
15. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press (1999)
16. Cevher, V., Chellappa, R., McClellan, J.H.: Gaussian approximations for energy-based detection and localization in sensor networks. In: IEEE Statistical Signal Processing Workshop, Madison, WI (26–29 August 2007)
17. Van Trees, H.L.: Detection, Estimation, and Modulation Theory, Part I. John Wiley & Sons, Inc. (1968)
18. Hale, E.T., Yin, W., Zhang, Y.: A fixed-point continuation method for ℓ_1-regularized minimization with applications to compressed sensing. Technical Report TR07-07, Rice University Department of Computational and Applied Mathematics, Houston, TX (2007)
19. Joo, S., Zheng, Q.: A Temporal Variance-Based Moving Target Detector. In: Proc. IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS). (2005)
20. Matsumoto, M., Nishimura, T.: Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. ACM Transactions on Modeling and Computer Simulation **8**(1) (1998) 3–30
21. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. Volume 2. (July 2001) 416–423
22. Khan, S.M., Shah, M.: A multi-view approach to tracking people in crowded scenes using a planar homography constraint. In: ECCV. Volume 4. (2006) 133–146