Imaging via Three-dimensional Compressive Sampling (3DCS) *

Xianbiao Shu, Narendra Ahuja
University of Illinois at Champaign-Urbana
405 N. Mathews Avenue, Urbana, IL 61801 USA
{xshu2,n-ahuja}@illinis.edu

Abstract

Compressive sampling (CS) aims at acquiring a signal at a sampling rate that is significantly below the Nyquist rate. Its main idea is that a signal can be decoded from incomplete linear measurements by seeking its sparsity in some domain. Despite the remarkable progress in the theory of CS, little headway has been made in the compressive imaging (CI) camera. In this paper, a three-dimensional compressive sampling (3DCS) approach is proposed to reduce the required sampling rate of the CI camera to a practical level. In 3DCS, a generic three-dimensional sparsity measure (3DSM) is presented, which decodes a video from incomplete samples by exploiting its 3D piecewise smoothness and temporal low-rank property. In addition, an efficient decoding algorithm is developed for this 3DSM with guaranteed convergence. The experimental results show that our 3DCS requires a much lower sampling rate than the existing CS methods without compromising recovery accuracy.

1. Introduction

Digital images and videos are being acquired by new imaging sensors with ever increasing fidelity, resolution and frame rate. The theoretical foundation is the Nyquist sampling theorem, which states that the signal information is preserved if the underlying analog signal is uniformly sampled above the Nyquist rate, which is twice its highest analog frequency. Unfortunately, Nyquist sampling has two major shortcomings. First, acquisition of a high resolution image necessitates a large-size sensor. This may be infeasible or extremely expensive in infrared imaging. Second, the raw data acquired by Nyquist sampling is too large to acquire, encode and transmit in short time, especially in the applications of wireless sensor networks, high speed imaging cameras, magnetic resonance imaging (MRI) and etc.

*The support of the Office of Naval Research under grant N00014-09-1-0017 is gratefully acknowledged.

Compressive sensing [8, 4] or compressive sampling (CS), was developed to solve this problem effectively. It is advantageous over Nyquist sampling, because it can (1) relax the computational burden during sensing and encoding, and (2) acquire high resolution data using small sensors. Assume a vectorized image or signal x of size L is sparsely represented as $x = \Psi z$, where z has K non-zero entries (called K-sparse) and Ψ is the wavelet transform. CS acquires a small number of incoherent linear projections $b = \Phi x$ and decodes the sparse solution $z = \Psi^T x$ as follows:

$$\min_z \|z\|_1 \quad \text{s. t.} \quad Az \triangleq \Phi\Psi z = \Phi x = b \qquad (1)$$

where Φ is a random sampling (RS) ensemble or a circulant sampling ensemble. According to [5], CS is capable of recovering K-sparse signal z (with an overwhelming probability) from b of size M, provided that the number of random samples meets $M \geq \alpha K \log(L/K)$. The required sampling rate ($\frac{M}{L}$), to incur lossless recovery, is roughly proportional to $\frac{K}{L}$. A compressive imaging camera prototype using RS is presented in [9]. Recently, circulant sampling (CirS) [25]) was introduced to replace RS with the advantages of easy hardware implementation, memory efficiency and fast decoding. It has been shown that CirS is competitive with RS in terms of recovery accuracy [25].

CS often reduces the required sampling rate by seeking the sparsest representation or by exploring some prior knowledge of the signal. Image CS (called 2DCS) decodes each image independently by minimizing both its sparsity in wavelet domain and total variation (TV2D+2DWT) [15, 16]. However, due to the significant sparsity, the required sampling rate is still quite high. Video CS (called 3DCS) is introduced to further reduce the sampling rate by adding the temporal correlation. Adaptive methods sense a key frame by Nyquist sampling and then sense consecutive frames [23] or frame differences [26] by CS; Sequential methods first decode a key frame and then recover other frames based on motion estimation [13]. Joint methods recover a video by seeking its 3D wavelet sparsity (3DWT) [24], or by minimizing the wavelet sparsity of its first frame and subsequent

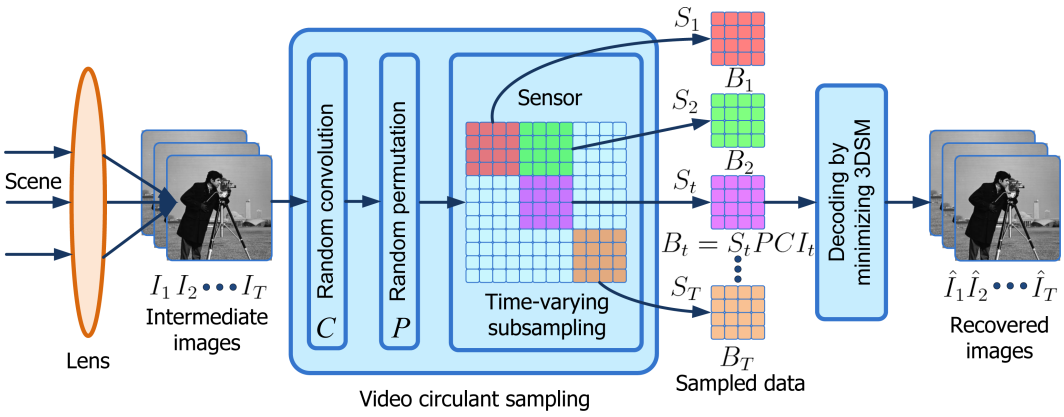


Figure 1. A compressive imaging (CI) camera using the proposed 3DCS. In this CI camera, a photographic lens (for forming image sequences, vectorized as $I_t, 1 \le t \le T$) is followed by video compressive sampling, which consists of optical random convolution (C), random permutation (P) and time-varying subsampling (S_t) on a sensor. From the sampled data sequences $B_t = S_t PCI_t, 1 \le t \le T$, the image sequences $\hat{I}_t, 1 \le t \le T$ is decoded by minimizing the 3DSM.

frame differences (Bk and Ck) [17]. However, none of the existing methods has exploited the major features of videos, i.e. 3D piecewise smoothness and temporal low-rank property.

In this paper, a new 3DCS approach is proposed to facilitate a promising CI camera (Figure 1) requiring a very low sampling rate. Without any computational cost, this CI camera acquires the compressed data $B_t, 1 \le t \le T$, from which a video clip is recovered by minimizing a new 3D sparsity measure (3DSM). This 3DSM is motivated by two characteristics of videos. First, videos are often piecewise smooth in 2D image domain and temporal domain. Second, image sequences in a video are highly correlated along the temporal axis, which can be modeled as a temporal low-rank matrix [10] with sparse innovations. Thus, a new 3DSM is constructed by combining 3D total variation (TV3D) and the low-rank measure in the Ψ-transform domain (Psi3D). The contributions of this paper are as follows:

1. A generic 3D sparsity measure (3DSM) is proposed, which exploits the 3D piecewise smoothness and temporal low-rank property in video CS. Extensive experiments demonstrate (1) the superiority of our 3DSM over other measures in terms of much higher recovery accuracy and (2) robustness over small camera motion.

2. An efficient algorithm is developed for the 3DSM with guaranteed convergence, which enables the recovery of large-scale videos.

3. circulant sampling (CirS) is extended from 2DCS to video CirS, by adding time-varying subsampling. The 3DSM and its efficient decoding algorithm, together with the video CirS, constitute the framework of our 3DCS (Figure 1).

This paper is organized as follows. Section 2 presents the proposed 3DSM and video CirS. Section 3 develops the recovery algorithm for the 3DSM. Section 4 describes the experiments in comparison with existing methods. Section 5 gives the concluding remarks.

2. Proposed Method

2.1. Overview of our 3D sparsity measure (3DSM)

In this section, a 3D sparsity measure (3DSM) is given for a fixed CI camera and will be extended to a moving camera later. A video clip is represented as a matrix $\mathbf{I} = [I_1, ..., I_t, ..., I_T]$, where each column I_t denotes one frame. For computational convenience, $\mathbf{I}$ is often vectorized as $I = [I_1; I_2; ...; I_T]$. A 3D sparsity measure (3DSM) is built by combining two complementary measures—3D total variation (TV3D) and 3D Ψ-transform sparsity (Psi3D). TV3D keeps the piecewise smoothness, while Psi3D retains the image sharpness and enforces the temporal low-rank property. As shown in Figure 1, the video I is recovered from the sampled data B by minimizing 3DSM.

$$\min_I \text{TV3D}(I) + \gamma \text{Psi3D}(I) \quad \text{s. t.} \quad \overline{\Phi} I = B \qquad (2)$$

where γ is a tuning parameter, $\overline{\Phi} = \text{diag}(\Phi_1, \Phi_2, ..., \Phi_T)$ is the video circulant sampling matrix, and $B = [B_1; B_2; ...; B_T]$ is the sampled data.

2.2. 3D Total Variation

In this section, TV3D is presented in detail. In 2DCS, total variation (TV) is often used to recover an image from incomplete measurements, by exploiting its piecewise smoothness in the spatial domain. The widely-used form of TV is TVL1L2 [15, 21, 16, 12], denoted as $\text{TV}_{\ell_1\ell_2}(I_t) =$

$\sum_i \sqrt{(D_1 I_t)_i^2 + (D_2 I_t)_i^2}$. In [22], the ℓ_1-norm based TV measure $\text{TV}_{\ell_1}(I_t) = \|D_1 I_t\|_1 + \|D_2 I_t\|_1$ is proven to be better than $\text{TV}_{\ell_1 \ell_2}$ in reducing the sampling rate. By extending TV_{ℓ_1} to the three-dimensional (spatial and temporal) domain, a new measure TV3D is formulated as:

$$\text{TV3D}(I) = \|D_1 I\|_1 + \|D_2 I\|_1 + \rho \|D_3 I\|_1 \quad (3)$$

where (D_1, D_2, D_3) are finite difference operators in 3D domain and ρ is proportional to the temporal correlation.

2.3. 3D Sparsity Measure in Ψ-transform Domain

In a video captured by a fixed camera, most pixels correspond to static scene and almost keep constant value over time. This video is temporally correlated and sparsely innovated (a small number of pixels varies with time), the same as its Ψ-transform coefficients $\mathbf{Z} = [Z_1, ..., Z_T] = [\Psi^T I_1, ..., \Psi^T I_T]$. Motivated by robust principal component analysis (RPCA) [3], $\mathbf{Z}$ is modeled as the sum of a low rank (LR) matrix $\overline{\mathbf{Z}}$ and sparse innovation $\widehat{\mathbf{Z}}$. Thus, Psi3D is formulated as follows:

$$\text{Psi3D}(I) = \min_{\overline{Z}, \widehat{Z}} \mu \text{Rank}(\overline{\mathbf{Z}}) + \eta \|\overline{Z}\|_1 + \|\widehat{Z}\|_1$$
$$\text{s.t.} \quad \overline{\Psi}^T I = \overline{Z} + \widehat{Z} \quad (4)$$

where $\overline{\Psi} = \text{diag}(\Psi, ..., \Psi)$, $\overline{Z} = [\overline{Z}_1; ...; \overline{Z}_T]$ and $\widehat{Z} = [\widehat{Z}_1; ...; \widehat{Z}_T]$ are vectorized versions of $\overline{\mathbf{Z}}$ and $\widehat{\mathbf{Z}}$. The weight coefficient μ tunes the rank of $\overline{\mathbf{Z}}$ and η must be set as $\eta \leq 1$; otherwise, the optimal $\overline{\mathbf{Z}}$ is prone to vanish. Different from the RPCA which seeks $\overline{Z}$ and $\widehat{Z}$ from complete data I, our 3DCS needs to recover $\overline{Z}$ and $\widehat{Z}$ from incomplete projections $B = \overline{\Phi} I$. Thus, both the sparsity and rank of $\overline{\mathbf{Z}}$ are explored to decode I. The proposed Psi3D attempts to minimize the number of nonzero singular values of $\overline{\mathbf{Z}}$, which is NP-hard and no efficient solution is known [7]. In practice, the widely used alternative [6, 3] is the *nuclear norm* $\|\overline{\mathbf{Z}}\|_* = \sum_{k=1} \sigma_k(\overline{\mathbf{Z}})$, which equals the sum of the singular values. Thus, the Psi3D is approximated as:

$$\text{Psi3D}_1(I) = \min_{\overline{Z}, \widehat{Z}} \mu \|\overline{\mathbf{Z}}\|_* + \eta \|\overline{Z}\|_1 + \|\widehat{Z}\|_1$$
$$\text{s.t.} \quad \overline{\Psi}^T I = \overline{Z} + \widehat{Z} \quad (5)$$

By assuming constant background, i.e., $\text{Rank}(\overline{\mathbf{Z}}) = 1$ and $\overline{Z} = [\overline{Z}_c, ..., \overline{Z}_c]$, the Psi3D is simplified by deleting the rank term as follows:

$$\text{Psi3D}_2(I) = \eta T \|\overline{Z}_c\|_1 + \|\widehat{Z}\|_1 \ \text{ s.t. } \Psi^T I_t = \overline{Z}_c + \widehat{Z}_t, \forall t \quad (6)$$

By setting $\eta T = 1$, the simplified Psi3D_2 is similar to the joint sparsity model of multiple signals [1]. In the case of constant background, Psi3D_2 requires less tuning parameters and computational cost than Psi3D_1. However, it is expected that Psi3D_1 achieves higher recovery accuracy than Psi3D_2 in the case of time-varying background ($\text{Rank}(\overline{\mathbf{Z}}) > 1$), e.g., illumination changes.

2.4. Robustness over Camera Motion

In this section, the 3DSM in Eq. (2) will be modified to be robust to camera motion. The 3DSM proposed in Eq. (2) explores the piecewise smoothness and the low-rank property, which is quite effective in improving the recovery accuracy of 3DCS. However, this model assumes a fixed camera or low-texture background. Even small camera motion might cause misalignments among real image sequences $I = [I_1; ...; I_T]$ and increase the temporal rank dramatically. these misalignments of I are modeled as a group of transformations (affine or perspective) $\Omega = \{\Omega_1, ..., \Omega_T\}$ on well-aligned image sequences $\tilde{I} = [\tilde{I}_1; ...; \tilde{I}_T]$ captured by a fixed camera. Specifically, $I_t = \tilde{I}_t \circ \Omega_t, 1 \leq t \leq T$. By introducing a group of transformations Ω, the 3DSM is extended to the moving camera as follows:

$$\min_{I, \Omega} \text{TV3D}(\tilde{I}) + \gamma \text{Psi3D}(\tilde{I}) \quad \text{s. t. } \overline{\Phi} I = B, \ I_t = \tilde{I}_t \circ \Omega_t \quad (7)$$

2.5. Video Circulant Sampling

In this section, a video circulant sampling (CirS) is presented, which, together with a photographic lens, constitutes a compressive imaging camera (Figure 1). It consists of two steps:

1. Random convolution. Video CirS convolves an image I_t by a random kernel H, denoted by CI_t, where C is a circulant matrix with H as its first column. C is diagonalized as $C = \mathcal{F}^{-1} \text{diag}(\widehat{H}) \mathcal{F}$, where $\widehat{H}$ is the Fourier transform of H, denoted by $\widehat{H} = \mathcal{F} H$. It can be easily implemented using Fourier optics [20].

2. Random subsampling, which consists of random permutation (P) and time-varying subsampling (S_t). S_t selects a block of M pixels from all N pixels on PCI_t and obtains the data $B_t = S_t P C I_t \triangleq \Phi_t I_t$. Note that the selected block drifts with time t (Figure 1). To relax the burden of both sensing and encoding, it is desirable to implement a physical mapping from a random subset to a 2D array (sensor). Although it is challenging, a possible solution is to implement random permutation by a bundle of optical fibers, followed by a moving small sensor. The easy way—sensing the whole image CI_t by a big sensor and throwing away the unwanted $N - M$ pixels, does not benefit sensing but yields a method of computation-free encoding.

3. 3DCS Recovery Algorithms

3.1. 3DCS Recovery using Inexact ALM-ADM

In this section, an efficient algorithm is presented to solve the 3DSM for a fixed camera. By introducing weight parameters $\alpha_1 = \alpha_2 = 1, \alpha_3 = \rho$, and auxiliary parameters

$\chi \triangleq (G_i, \overline{Z}, \widehat{Z}, R)$, our 3DCS using Psi3D$_1$ Eq. (2) can be rewritten as:

$$\min_{I,G_i,\overline{Z},\widehat{Z}} \quad \sum_{i=1}^{3} \alpha_i \|G_i\|_1 + \gamma(\mu\|\overline{\mathbf{Z}}\|_* + \eta\|\overline{Z}\|_1 + \|\widehat{Z}\|_1)$$
$$\text{s.t. } G_i = D_i I, \ \overline{Z} + \widehat{Z} = \overline{\Psi}^T I, \ R_t = CI_t, \ S_t P R_t = B_t \quad (8)$$

This linearly constrained problem can be solved by augmented Lagrangian multipliers (ALM) [12]. Given an L1-norm problem $\min \|a\|_1 \quad \text{s.t.} \quad a = b$, its augmented Lagrangian function (ALF) is defined as $\mathcal{L}_\beta(a, b, y) = \|a\|_1 - y\beta(a-b) + \frac{\beta}{2}\|a-b\|_2^2$. Then, the ALF of Eq. (8) is written as

$$\mathcal{L}(I, G_i, \overline{Z}, \widehat{Z}, R, b_i, d, g) = \sum_{i=1}^{3} \alpha_i \mathcal{L}_{\beta_i}(G_i, D_i I, b_i)$$
$$+\gamma(\mu\|\overline{\mathbf{Z}}\|_* + \eta\|\overline{Z}\|_1 + \|\widehat{Z}\|_1 + \frac{\beta_4}{2}\|\overline{Z} + \widehat{Z} - \overline{\Psi}^T I - d\|_2^2)$$
$$+\frac{\beta_5}{2}\|R - \overline{C}I - g\|_2^2 \quad \text{s.t.} \quad S_t P R_t = B_t, 1 \le t \le T \quad (9)$$

where $\beta_i, 1 \le i \le 5$ are over-regularization parameters, $\lambda \triangleq (b_1, b_2, b_3, d, g)$ is Lagrangian multipliers and $\overline{C} = \text{diag}(C, ..., C)$. ALM solves Eq. (8) by iterating between the following two steps:

1. Solve $(I^{k+1}, \chi^{k+1}) \leftarrow \arg\min \mathcal{L}(I, \chi, \lambda^k)$.
2. Update λ^{k+1} with $(\lambda^k, I^{k+1}, \chi^{k+1})$.

Each ALM iteration requires an exact minimization of $\mathcal{L}(I, G_i, \overline{Z}, \widehat{Z}, R)$, which is expensive. Fortunately, at fixed I and λ^k, minimization of $\mathcal{L}(G_i, \overline{Z}, \widehat{Z}, R)$ can be performed independently. In this case, all the variables can be divided into two groups (I and $\chi = \{G_i, \overline{Z}, \widehat{Z}, R\}$), and $\mathcal{L}(I^k, \chi)$ can be minimized by applying the alternating direction method (ADM) [12]. Given λ^{k+1} is updated at a sufficiently slow rate, the exact minimization can be simplified as only one round of alternating minimization (called inexact ADM). As shown in Algorithm 1, the inexact ALM-ADM solves our 3DCS by iterating among three major steps: (1) separate rectification of χ, (2) joint reconstruction of I and (3) update of λ.

Given the exact solution to $\arg\min_\chi \mathcal{L}(I^k, \chi, \lambda^k)$ is reached in the first step, the group of components χ is equivalent to a single variable and the inexact ALM-ADM with respect to two variables $\{\chi, \text{I}\}$ is guaranteed with convergence. Motivated by the convergence analysis in [11], the convergence condition of our recovery algorithm using inexact ALM-ADM is given as follows.

Theorem 1 *If the over-regularization parameters $\beta_i > 0, \forall 1 \le i \le 5$ and the step length $\tau \in (0, (1+\sqrt{5})/2)$, the video sequence I^{k+1} reconstructed by the inexact ALM-ADM (Algorithm 1) will uniquely converge the solution to Eq. (8).*

Algorithm 1 Solving 3DCS using inexact ALM-ADM

Input: $C, \widehat{H}, P, S_t$ and $B_t, 1 \le t \le T$
Output: I^{k+1}

1: $I^0 \leftarrow G_i^0 \leftarrow b_i^0 \leftarrow \text{zeros}(m, n, T), 1 \le i \le 3;$
$\overline{Z}^0 \leftarrow \widehat{Z}^0 \leftarrow d^0 \leftarrow R^0 \leftarrow g^0 \leftarrow \text{zeros}(m, n, T).$
2: **while** I not converged **do**
3: Separate Rectification of: $\chi = \{G_i, \overline{Z}, \widehat{Z}, R\}$
$\chi^{k+1} \leftarrow \arg\min_\chi \mathcal{L}(I^k, \chi, \lambda^k)$
4: Joint Reconstruction:
$(I^{k+1}) \leftarrow \arg\min \mathcal{L}(I, \chi^{k+1}, \lambda^k)$
5: Update λ by "Adding back noise":
$b_i^{k+1} \leftarrow b_i^k - \tau(G_i^{k+1} - D_i I^{k+1})$
$d^{k+1} \leftarrow d^k - \tau(Z^{k+1} - \overline{\Psi}^T I^{k+1})$
$g^{k+1} \leftarrow g^k - \tau(R^{k+1} - \overline{C} I^{k+1})$
6: $k \leftarrow k+1$
7: **end while**

3.1.1 Separate Rectification using Soft Shrinkage

The convergence of the inexact ALM-ADM requires an exact solution to $\arg\min_\chi \mathcal{L}(I^k, \chi, \lambda^k)$ at each iteration, which can be obtained separately with respect to G_i, R and the pair $\{\overline{Z}, \widehat{Z}\}$. Define a soft shrinkage function as $S(X, 1/\beta) = \max\{\text{abs}(X) - 1/\beta, 0\} \cdot \text{sgn}(X)$, where "$\cdot$" denotes elementwise multiplication, then $G_i^{k+1}, 1 \le i \le 3$ are straightforwardly updated by

$$G_i^{k+1} \leftarrow S(D_i I^k + b_i^k, \frac{1}{\beta_i}) \quad (10)$$

As for Psi3D$_1$, $\{\overline{Z}, \widehat{Z}\}$ are rectified from I^k and d^k by singular value shrinkage (SVS)[2]. To meet the convergence condition in Theorem 1, the pair $\{\overline{Z}, \widehat{Z}\}$ need be rectified iteratively until the pair is converged. In practice, the algorithm can be accelerated by just applying one round of rectification.

$$\overline{\mathbf{Z}}^{k+1} \leftarrow US(\Sigma, \mu/\beta_4)V^T \quad (11)$$
$$\overline{Z}^{k+1} \leftarrow S(\overline{Z}^{k+1}, \eta/\beta_4) \quad (12)$$
$$\widehat{Z}^{k+1} \leftarrow S(\overline{\Psi}^T I^k + d^k - \overline{Z}^{k+1}, 1/\beta_4) \quad (13)$$

where $[U, \Sigma, V] = \text{svd}(\overline{\Psi}^T I^k + d^k - \widehat{Z}^k)$. Similarly, as for Psi3D$_2$, $\{\overline{Z}, \widehat{Z}\}$ can be rectified without SVS operation.

The complete circulent samples $R = [R_1; ...; R_T]$ are rectified by 3D data I^k and partial circulant samples B_t.

$$R_t^{k+1} \leftarrow CI_t^k \quad (14)$$
$$R_t^{k+1}(\text{Picks}_t, :) \leftarrow B_t \quad (15)$$

where Picks_t are the indices of rows selected by S_t.

3.1.2 Efficient Joint Reconstruction using 3D FFT

In this section, an efficient algorithm is presented to recover I^{k+1} jointly from rectified variables $(G_i^{k+1}, \overline{Z}^{k+1}, \widehat{Z}^{k+1}, R^{k+1})$. This algorithm is greatly accelerated by our 3DSM and video circulant sampling. By setting the derivative of $\mathcal{L}$ with respect to I to be zero, the optimal condition of I is induced as follows:

$$\sum_{i=1}^{3} \alpha_i \beta_i D_i^T (G_i - D_i I - b_i) + \gamma \beta_4 \overline{\Psi}(\overline{Z} + \widehat{Z} - \overline{\Psi}^T I - d) + \beta_5 \overline{C}^T (R - \overline{C} I - g) = 0 \quad (16)$$

Eq. (16)is reformulated into the form $\Gamma I = \Theta$, where $\Gamma = \sum_{i=1}^{3} \alpha_i \beta_i D_i^T D_i + \gamma \beta_4 + \beta_5 \overline{C}^T \overline{C}$. Under the periodic boundary condition [18] of finite operators $D_i, 1 \le i \le 3$, both $\overline{C}^T \overline{C}$ and $D_i^T D_i, 1 \le i \le 3$ are block-circulant matrices. The operation Γ on 3D data I is equivalent to the sum of separate convolutions with five point spread functions $\text{PSF}_i, 1 \le i \le 5$, which are given as follows:

$$\text{Horizontal: PSF}_1 = \alpha_1 \beta_1 [1; -2; 1] \quad (17)$$

$$\text{Vertical: PSF}_2 = \alpha_2 \beta_2 [1, -2, 1] \quad (18)$$

$$\text{Temporal: PSF}_3 = \alpha_3 \beta_3 [1 : -2 : 1] \quad (19)$$

$$\text{Dirac delta: PSF}_4 = \alpha_4 \beta_4 \delta(x, y) \quad (20)$$

$$\text{2D: PSF}_5 = \beta_5 \mathcal{F}^{-1}(\hat{H}^* \cdot \hat{H}) \quad (21)$$

where $[A_1 : A_2 : A_3]$ denotes concatenating A_1, A_2 and A_3 along the third (temporal) axis and $\hat{H}^*$ is the complex conjugate of $\hat{H}$. According to the convolution theory, this optimal condition with respect to I can be solved efficiently by applying 3D fast Fourier transform.

$$\sum_{i=1}^{5} \mathcal{F}(\text{PSF}_i) \cdot \widehat{I} = \mathcal{F}(\Theta) \quad (22)$$

where $\widehat{I} = \mathcal{F}(I)$. By defining the optical transfer function OTF $\triangleq \sum_{i=1}^{5} \text{PSF}_i$, given χ^{k+1} and λ^k, Θ^{k+1} can be updated as $\Theta^{k+1} = \sum_{i=1}^{3} \alpha_i \beta_i D_i^T (G_i^{k+1} - b_i^k) + \gamma \beta_4 \overline{\Psi}(\overline{Z}^{k+1} + \widehat{Z}^{k+1} - d^k) + \beta_5 \overline{C}^T (R^{k+1} - g^k)$. Then, I^{k+1} is recovered by applying inverse FFT the element-wise divisor of $\mathcal{F}\Theta^{k+1}$ by OTF. Our decoding algorithm is extremely efficient, for its joint reconstruction only requires two 3D FFT and some simple 1D/2D filtering.

3.2. 3DCS Recovery with Camera Motion

In this section, the inexact ALM-ADM above can be adapted to solve Eq. (7) for the moving camera. The objective function can be relaxed as $\mathbf{F}(I, \Omega)$ as:

$$\mathbf{F}(I, \Omega) = \sum_{i=1}^{3} \alpha_i \|G_i\|_1 + \gamma(\mu \|\overline{\mathbf{Z}}\|_* + \eta \|\overline{Z}\|_1 + \|\widehat{Z}\|_1)$$

$$\text{s.t. } G_i = D_i(I \circ \Omega^{-1}),\ \overline{Z} + \widehat{Z} = \overline{\Psi}^T (I \circ \Omega^{-1}),\ S_t P C I_t = B_t \quad (23)$$

The main difficulty in solving Eq. (23) is the complicated dependence of aligned images $\tilde{I}_t = I_t \Omega_t$ on unknown transformations $\Omega_t, 1 \le t \le T$. It is almost impossible to solve $\tilde{I}_t$ and Ω_t simultaneously. The ADM is applied to solve them by iterating between the following two steps:

1. Given Ω^k, solve $I^{k+1} \leftarrow \min_I \mathbf{F}(I, \Omega^k)$. This can be solved by adding forward and backward transformations in each iteration cycle of inexact ALM-ADM (Algorithm 1), however, it might be time-consuming. To accelerate it, the misalignments Ω_t are modeled as translation $(\Delta x_t, \Delta y_t)$. According to Fourier shift theorem, $\mathcal{F}(I_t)$ is equal to multiplying $\mathcal{F}(\tilde{I}_t)$ by a linear phase $P(\Delta x_t, \Delta y_t)$. it is induced that $CI_t = \mathcal{F}^{-1}(\hat{H} \cdot \mathcal{F}(I_t)) = \mathcal{F}^{-1}(\hat{H} \cdot P(\Delta x_t, \Delta y_t) \cdot \mathcal{F}(\tilde{I}_t)) = \tilde{C}_t \tilde{I}_t$, where $\tilde{C}_t = \mathcal{F}^{-1}(\hat{H} \cdot P(\Delta x_t, \Delta y_t))$. Thus, I is recovered by fast solving $\tilde{I}$, followed by circulant shift $(\Delta x_t, \Delta y_t)$.

$$\min_{\tilde{I}} \sum_{i=1}^{3} \alpha_i \|G_i\|_1 + \gamma(\mu \|\overline{\mathbf{Z}}\|_* + \eta \|\overline{Z}\|_1 + \|\widehat{Z}\|_1)$$
$$\text{s.t. } G_i = D_i \tilde{I},\ \overline{Z} + \widehat{Z} = \overline{\Psi}^T \tilde{I},\ S_t P \tilde{C}_t \tilde{I}_t = B_t \quad (24)$$

2. Given I^{k+1}, solve $\Omega^{k+1} \leftarrow \min_\Omega \mathbf{F}(I^{k+1}, \Omega)$, without considering the constraint $S_t P C I_t = B_t$. It is quite similar to robust image alignment. Readers are referred to [19] for detailed algorithms.

4. Experimental Results

Although the compressive imaging camera has not been built, the 3DCS approach can still be evaluated by feeding the intermediate images $I_t, 1 \le t \le T$ (Figure 1) with three surveillance videos from [14], i.e., an airport video (size: 144×176 pixels), a brighter lobby video (size: 128×160 pixels) and a darker lobby video (size: 128×160 pixels), as well as a video captured by a handheld camera (building video, size: 480×480). Our 3DSM (default: TV3D+Psi3D$_1$) is compared with existing sparsity measures, such as TV2D+2DWT, 3DWT, Bk and Ck. Peak signal-to-noise ratio (PSNR) is used as the measure of recovery accuracy.

4.1. Parameter Selection

Assigning appropriate values to weight parameters $\gamma, \rho, \eta, \mu, \tau$ and $\beta_i, 1 \le i \le 5$ seems quite complicated but it is actually not. Weight parameters γ and ρ are often set to be greater than 1. η is often set to be $\frac{1}{T}$. μ depends on the rank of the background components in the video. Over-regularization parameters $\beta_i, 1 \le i \le 5$ prefer large values. To evaluate our 3DSM and its decoding algorithm, The values of the weight parameters are set empirically through

all experiments, i.e. $\gamma = 4, \rho = 3, \eta = \frac{1}{T}, \tau = 1.6$ and $\beta_i = 100, \forall i$.

4.2. Evaluation of our 3DSM in video CS

4.2.1 Evaluation of 3DSM using Psi3D$_2$

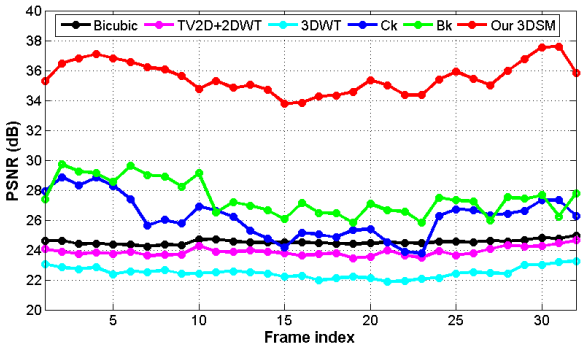


Figure 2. Recovery accuracy of different sparsity measures on a 32-frame airport video at sampling rate $\frac{M}{L} = 25\%$.

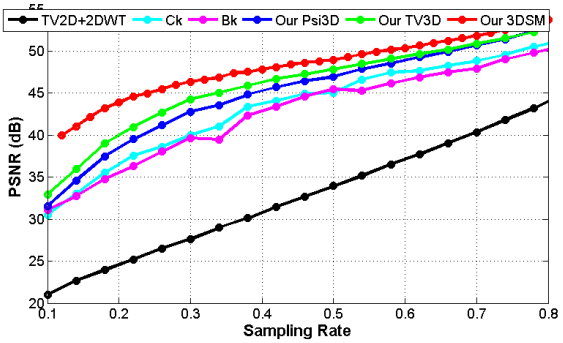


Figure 3. Averaged recovery accuracy of 10 frames at varying sampling rates in the brighter lobby video.

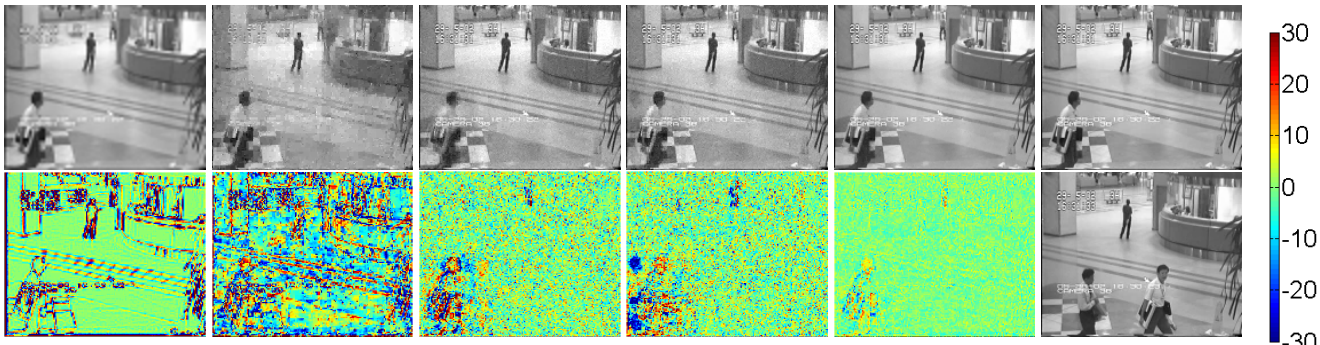
Fixed Sampling Rate. Our 3DSM is evaluated on the airport video (32 frames) at the sampling rate 25%, in comparison with other sparsity measures and a naive method—bicubic interpolation after half downsampling. The computing time is about 123 seconds on a normal computer (Intel E6320 CPU, 3 GB memory). As shown in Figure 2, our 3DSM achieves a recovery accuracy at least 6 dB higher than all the other methods at each frame. 2DCS fails to achieve high recovery accuracy, for the PSNR of its state-of-art (TV2D+2DWT) is lower than bicubic interpolation, which suffers from significant blur. As shown in Figure 4, our 3DSM recovery is much sharper, cleaner and closer to the original 4th frame. The recovery accuracy of our 3DSM decreases as the size of moving foreground grows, e.g., the 4th and 17th frames in Figure 4. Our 3DSM is tested on the brighter and darker lobby videos at the sampling rate of 30%. Figure 5 shows that our 3DSM decodes much better images (PSNR: up to 45 dB) than other measures and the JPEG version of the original image (compression ratio: 30%). Our 3DSM achieves better recovery in the brighter lobby than that in the darker one, due to the significant photon-counting noise under the darker condition.

Varying Sampling Rate. By varying the sampling rate, TV3D, Psi3D$_2$ and 3DSM are tested on the brighter lobby video. As shown in Figure 3, either TV3D or our Psi3D is better (3 dB higher PSNR) than other measures at any sampling rate ($\frac{M}{L}$), and their combination (3DSM) is better than each one alone. The superiority of our 3DSM over other measures at $\frac{M}{L} = 25\%$ is up to 7 dB. At $\frac{M}{L} = 10\%$, 3DSM achieves the recovery accuracy (PSNR: 40 dB), which is conventionally considered to be lossless.

Varying Video Size T. As shown in Figure 1, our 3DCS divides a video into short clips of T frames and then decode each short clip by minimizing our 3DSM. To study the influence of T on the decoding accuracy, a hybrid video is built by 16 frames from brighter lobby and 16 frames from from darker lobby. The average accuracy of all T frames using our 3DSM increases quickly with T in the beginning and reaches a stable value when $T \geq 10$. When the lighting changes at the 17th frame, both the accuracy of Psi3D$_1$ and that of Psi3D$_2$ decrease to some extent (Figure 6).

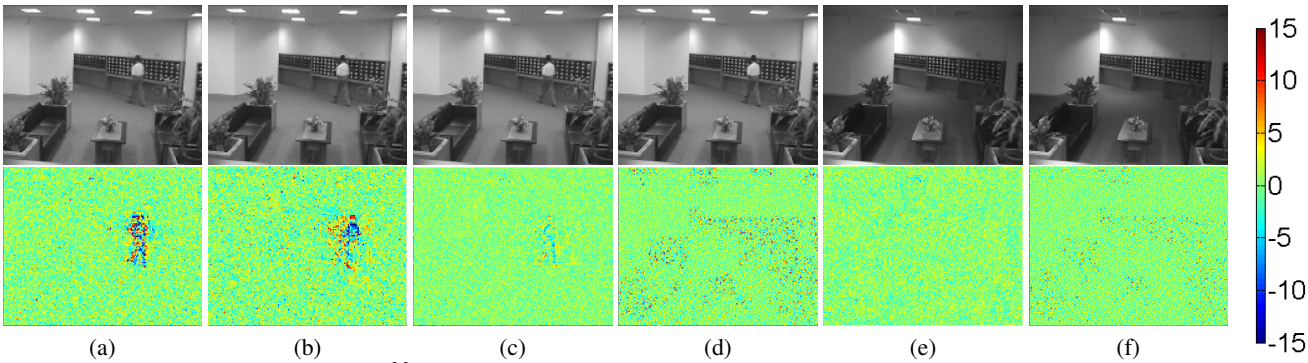
Psi3D$_1$ and Psi3D$_2$. As shown in Figure 6, under constant lighting ($T \leq 16$), our 3DSM using Psi3D$_2$ is better than using Psi3D$_1$ at any tuning parameter μ. However, under varying lighting conditions ($T > 16$), the decoding accuracy of Psi3D$_1$ at $\mu = 100$ remains almost invariant (46 dB $\leq$ PSNR $\leq$ 47 dB) as T increases, and is much higher than that of Psi3D$_2$. This can be explained by the background models of Psi3D$_1$ and Psi3D$_2$. Psi3D$_1$ uses a low-rank background model and can recover rank-2 background images (top row in Figure 7). Thus, the innovation images of Psi3D$_1$ will be sparser than that of Psi3D$_2$, which rigidly assumes rank-1 background. Since the complete data I is unknown, the low-rank images recovered by Psi3D$_1$ are different from the real background (Figure 7).

Robustness over Motion. the robustness of our 3DSM over motion is evaluated by applying it to a video acquired by a moving (up and down) camera. As shown in Figure 8, without image alignment, our 3DSM can still recover the image sequences at $M/L = 30\%$ with acceptable accuracy (top row: PSNR $>$ 31 dB). From this initial recovery, the transformations $\Omega_t, 1 \leq t \leq 12$ are estimated, of which the dominant components are translations $(\Delta x_t, \Delta y_t)$. Given the translation knowledge, our 3DSM improves upon the initial recovery by 2.6 dB in terms of PSNR, less noise (e.g., bricks) and more detailed information (e.g., parking sign), as shown in Figure 8 (mid row). For computational efficiency, the translations are rounded to integer pixels. It is expected that our 3DSM recovery can be further improved by using perspective transformations.



(a) (b) (c) (d) (e) (f)

Figure 4. Reconstructed images (top)of 4th frame and error maps at sampling rate $\frac{M}{L} = 25\%$ using (a) bicubic (PSNR: 24.42 dB), (b) TV2D+2DWT (PSNR: 23.84 dB), (c) Ck (PSNR: 28.84 dB), (d) Bk (PSNR: 29.13 dB), and (e) our 3DSM (PSNR: 37.10 dB). (f) original 4th (top) and 17th (bottom) frames in airport video.



(a) (b) (c) (d) (e) (f)

Figure 5. Reconstructed images (top) at $\frac{M}{L} = 30\%$ and error maps (bottom) from the bright video clip (10 frames) using (a) Ck (PSNR: 40.62 dB), (b) Bk (PSNR: 39.75 dB) and (c) our 3DSM (PSNR: 46.09 dB), with (d) their original image (top) and error map of its JPEG version (bottom), compression ratio: 30%, PSNR: 40.59 dB). (d) Our 3DSM recovery (top, PSNR: 44.66 dB) and error map (bottom) from 30% circulant samples of the darker clip (10 frames), with (f) its original image (top) and error map of its JPEG version (bottom, compression ratio: 30%, PSNR: 42.53 dB).

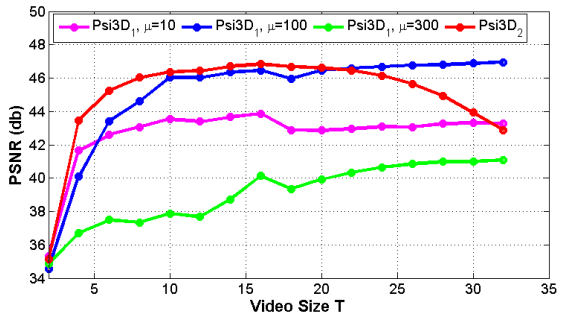


Figure 6. Comparison of Psi3D_1 and Psi3D_2 by varying T. This hybrid lobby video consists of 16 frames from brighter lobby and 16 from the darker one.

(a) (b) (c) (d)

Figure 7. Low rank recovery using our 3DSM (Psi3D_1, $\mu = 100$) from the hybrid lobby video (brighter: 16 frames and darker: 16 frames) at $\frac{M}{L} = 30\%$. Reconstructed 4 frames (bottom) with low rank components (top): (a) (PSNR: 45.98 dB), (b) (PSNR: 46.36 dB), (c) (PSNR: 45.31 dB) and (d) (PSNR: 46.28 dB).

5. Conclusion

In this paper, a 3D compressive sampling (3DCS) approach has been proposed to facilitate a promising compressive imaging camera, which consists of video circulant sampling, 3D sparsity measure (3DSM) and an efficient decoding algorithm with convergence guarantee. By exploiting the 3D piecewise smoothness and temporal low rank property of videos, our 3DSM reduces the required sampling rate to a practical level (e.g. 10% in Figure 3). Extensive experiments have been conducted to show (1) the superiority of our 3DSM over existing sparsity measures in terms of recovery accuracy with respect to the sampling rate, and (2) robustness over small camera motion. Motivated by these exciting results, a real compressive imaging camera will be built, which is very promising for applications such as wireless camera network, infrared imaging, remote sensing and etc.

(a) (b) (c)

Figure 8. Recovery of the 12-frame building video acquired by a handheld camera using our 3DSM at $M/L = 30\%$. Top: initial results without image alignment (a) (PSNR: 31.50 dB), (b) (PSNR: 31.71 dB) and (c) (PSNR: 32.63 dB). From initial results, the estimated translations $(\Delta x, \Delta y)$ are listed as (a)(2.35, 3.11), (b) (0.62, -1.77), (c)(-0.31, 2.36). Middle: final results with estimated $(\Delta x, \Delta y)$ (a) (PSNR: 33.85 dB), (b) (PSNR: 34.10 dB) and (c) (PSNR: 35.18 dB). Bottom: three original frames.

References

[1] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. *Preprint. Available at www.dsp.rice.edu/cs.*, 2005. 3

[2] J. Cai, E. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. *preprint*, 1984. 4

[3] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):article 11, 2011. 3

[4] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1208–1223, 2006. 1

[5] E. Candes and T. Tao. Near-optimal signal recovery from random projections and universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5245, 2006. 1

[6] E. J. Candes and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009. 3

[7] A. L. Chistov and D. Y. Grigoriev. Complexity of quantifier elimination in the theory of algebraically closed fields. *Mathematical Foundations of Computer Science*, 176:17–31, 1984. 3

[8] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289 – 1306, 2006. 1

[9] M. Duarte, M. Davenport, D. Takhar, J. Laska, and T. Sun. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008. 1

[10] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 2

[11] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer-Verlag, 1984. 4

[12] Y. Z. Junfeng Yang and W. Yin. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010. 2, 4

[13] L. Kang and C. Lu. Distributed compressive video sensing. *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, 2009. 1

[14] L. Li, W. Huang, I. Gu, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *SIAM Journal on Scientific Computing*, 13(11), 2004. 5

[15] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed sensing mri. *IEEE Signal Processing Magazine*, 25(2):72–82, 2007. 1, 2

[16] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed mr imaging using total variation and wavelets. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 1, 2

[17] R. Marcia and R. Willett. Compressive coded aperture video reconstruction. *Proc. of European Signal Processing Conference*, 2008. 2

[18] M. K. Ng, R. H. Chan, and W. C. Tang. A fast algorithm for deblurring models with neumann boundary conditions. *SIAM Journal on Scientific Computing*, 21(3):851–866, 2000. 5

[19] Y. Peng, A. Ganesh, J. Wright, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *Proc. of IEEE conference on Computer Vision and Pattern Recognition*, 2010. 5

[20] J. Romberg. Compressive sensing by random convolution. *SIAM Journal on Imaging Science*, 2009. 3

[21] J. K. Romberg. Variational methods for compressive sampling. *SPIE*, 6498:64980J–2–5, 2007. 2

[22] X. Shu and N. Ahuja. Hybrid compressive sampling via a new total variation tvl1. *Proc. of European Conference on Computer Vision*, 2010. 3

[23] V. Stankovic, L. Stankovic, and S. Cheng. Compressive video sampling. *Proc. of European Signal Processing Conference*, 2008. 1

[24] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. Compressive imaging for video representation and coding. *Proc. of Picture Coding Symposium*, 2006. 1

[25] W. Yin, S. P. Morgan, J. Yang, and Y. Zhang. Practical compressive sensing with toeplitz and circulant matrices. *Rice University CAAM Technical Report TR10-01*, 2010. 1

[26] J. Zheng and E. L. Jacobs. Video compressive sensing using spatial domain sparsity. *SPIE Optical Engineering*, 48(8), 2009. 1