FPA-CS: Focal Plane Array-based Compressive Imaging in Short-wave Infrared

Huaijin Chen,[1] M. Salman Asif,[1] Aswin C. Sankaranarayanan,[2] and Ashok Veeraraghavan[1]
[1]Department of Electrical and Computer Engineering, Rice University.
[2]Department of Electrical and Computer Engineering, Carnegie Mellon University.

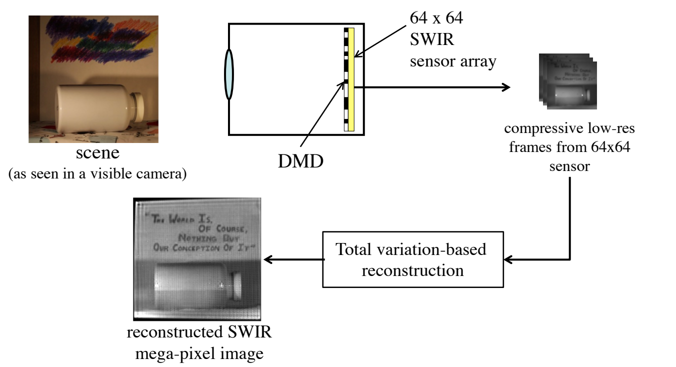

Figure 1: Focal plane array-based compressive sensing (FPA-CS) camera architecture: A 64×64 SWIR sensor array is equivalent to 4096 single pixel cameras (SPCs) operating in parallel. This results in vastly superior spatio-temporal resolutions against what is achievable using the SPC or a traditional camera.

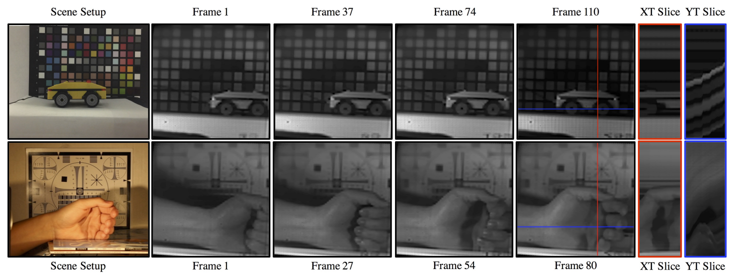

Figure 2: Selected frames from reconstructed SWIR videos. Each frame in the moving car videos is reconstructed using 16 captured images; compression factor $\alpha = 16$, and a consequently 32-fps frame rate. Each frame in the moving hand videos is reconstructed using 22 captured images; compression factor $\alpha = 11.6$, and a consequently 21.8-fps frame rate. Both videos are reconstructed using 3D-TV prior. XT and YT slices for both videos are shown to the right of the images.

Cameras for imaging in short and mid-wave infrared spectra are significantly more expensive than their counterparts for visible imaging. For example, a cellphone camera with a several megapixel sensor costs a few dollars, but a megapixel sensor for short-wave infrared (SWIR) imaging costs tens of thousands dollars. As a result, high-resolution imaging beyond the visible spectrum remains out of reach for many consumers.

Over the last decade, compressive sensing (CS) [1] has emerged as a useful technology for designing high-resolution imaging systems using low-resolution sensors. For instance, a single-pixel camera (SPC) uses a single-pixel detector and a digital micromirror device (DMD) to record coded measurements of a high-resolution image [3]. A computational reconstruction algorithm is then used to recover the high-resolution image from the coded measurements. Unfortunately, the measurement rate of an SPC is insufficient for imaging at high spatial and temporal resolutions [5].

In this paper, we present a focal plane array-based compressive sensing (FPA-CS) architecture that achieves high spatial and temporal resolutions using inexpensive, low-resolution sensors. Our proposed architecture can be viewed as an array of SPCs working in parallel, thereby increasing the measurement rate, and consequently, the achievable spatio-temporal resolution of CS-based cameras. We develop a proof-of-concept prototype SWIR video camera using a low-resolution sensor with 64×64 pixels; the prototype provides a $4096\times$ increase in measurement rate compared to the SPC, and for the first time, achieves megapixel resolution at video rate using CS techniques.

Our prototype FPA-CS camera is constructed using a low-resolution sensor array of 64×64 pixels, each observing a 16×16 patch of micromirrors. The DMD patterns and sensor readout timings are synchronized to record modulated, low-resolution images at a frame rate $F_s = 480$ fps. The sensor image at time t can be described as $y_t = A_t x_t$, where y_t is a vector with 4096 measurements, x_t represents the high-resolution image at the DMD plane, and the matrix A_t encodes modulation of x_t with the DMD pattern and mapping onto the SWIR sensor pixels. To reconstruct video at a desired frame-rate, say F_r fps, we divide low-resolution sensor images into sets of $T = F_s/F_r$ measurements, all of which correspond to the same high-resolution image. Suppose the kth set correspond to $y_t = A_t x_t$ for $t = (k-1)T+1, \ldots, kT$; we assume that $x_t = \mathbf{x}_k$ and stack all the y_t and A_t in the kth set in $\mathbf{y}_k$ and $\mathbf{A}_k$, respectively. Our goal is to reconstruct the $\mathbf{x}_k$ from the noisy and possibly under-determined sets of linear equations $\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k$.

Natural images have been shown to have sparse gradients. We can view a video signal as a 3D object that consists of a sequence of 2D images, and we expect pixels in each image to be similar to their neighbors along horizontal, vertical, and temporal directions. To exploit the spatio-temporal similarity in a video signal, we can use priors for sparse spatio-temporal gradients, and solve an optimization problem of the following form for reconstruction[4]:

$$\text{(TV)} \quad \widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathrm{TV}_{3\mathrm{D}}(\mathbf{x}) \quad \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \varepsilon,$$

where the term $\mathrm{TV}_{3\mathrm{D}}(\mathbf{x})$ refers to the 3D total-variation of $\mathbf{x}$. $\mathrm{TV}_{3\mathrm{D}}$ can be defined as

$$\mathrm{TV}_{3\mathrm{D}}(\mathbf{x}) = \sum_i \sqrt{(D_u \mathbf{x}(i))^2 + (D_v \mathbf{x}(i))^2 + (D_t \mathbf{x}(i))^2},$$

where $D_u\mathbf{x}$ and $D_v\mathbf{x}$ are the spatial gradients along horizontal and vertical dimensions of $\mathbf{x}$, respectively, and $D_t\mathbf{x}$ represents gradient along the temporal dimension of $\mathbf{x}$. We present some of our experimental results in Figure 2, where we used MFISTA [2] for the reconstruction of videos.

FPA-CS provides three advantages over conventional imaging. First, our CS-inspired FPA-CS system provides an inexpensive alternative to achieve SWIR imaging in high spatiotemporal resolution . Second, compared to traditional single-pixel-based compressive cameras, FPA-CS simultaneously records data from 4096 parallel, compressive systems, thereby significantly improves the measurement rate. As a consequence, the achieved spatio-temporal resolution of our device is an order of magnitude better than the SPC.

[1] Richard Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24 (4), 2007.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[3] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, Mar. 2008.

[4] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling and Simulation*, 4(2):460–489, 2005.

[5] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk. CS-MUVI: Video compressive sensing for spatial-multiplexing cameras. In *IEEE International Conference on Computational Photography*, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.