LiSens --- A Scalable Architecture for Video Compressive Sensing

Jian Wang
Mohit Gupta
Aswin C. Sankaranarayanan

ICCP 2015
Motivation

Best result: $128 \times 128$ at video rate

Single pixel camera (SPC)

Why cannot SPC sense videos at very high spatial and temporal resolutions?

Motivation

(1) Natural scenes are increasingly compressible when resolved at higher resolutions.
   - lower resolution, less compressible

(2) SPC has very poor measurement rate, not conducive to sensing at high spatial-temporal resolutions.

measurement rate: # of measurements per second
Motivation

Replace single pixel by multiple pixels
• Achieve required measurement rates
• Capable of sensing at high spatio-temporal resolutions

$128 \times 128$ at video rate

$1024 \times 768$ at video rate
Sparsity of natural images

90 megapixel images of natural scenes
Sparsity of natural images

- Wavelet transform
- Keep largest $K$ coefficients, set others to zero
- Reconstruct the image

$N$ pixels, $N$ coefficients

$K$ - # of non-zero coefficients
$K/N$ - non-zero ratio
Sparsity of natural images

Reconstruction
SNR [dB]

Non-zero ratio
K/N

N [in pixels]

K - # of non-zero coefficients
N - # of pixels

SNR ≥ 25 dB

K/N = 0.01

K/N = 0.1
Sparsity of natural images

Natural images are increasingly compressible at higher spatial resolutions.

- CS techniques perform better for sensing signals at higher resolutions.
- Challenge towards CS @high-res
  Even with a compressive camera, acquisition at high measurement rates is required.
Single-pixel camera

Measurement rate

\[ R = \min(R_{DMD}, R_{ADC}), \quad R_{DMD} \ll R_{ADC} \]

\[ = R_{DMD} \quad \text{Typically orders of 10 KHz} \]
Replace the single pixel by multiple pixels

\[ R = \min(R_{DMD}, R_{ADC}) = R_{DMD} \]

\[ R = \min(R_{DMD} \times F, R_{ADC}) \]
Pixels are extremely costly at SWIR (900 nm – 2.5 µm) bands!

<table>
<thead>
<tr>
<th>Spatio-temporal resolution SWIR sensor</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 megapixel @ 10 fps</td>
<td>$400k</td>
</tr>
<tr>
<td>0.25 megapixel @ 10 fps</td>
<td>$40k</td>
</tr>
</tbody>
</table>

SWIR sensors

[M. E. Gehm and D. J. Brady, Compressive sensing in the EO/IR, Applied Optics, 2015]
Optimal # of pixels

\[ R = \min(R_{DMD} \times F, R_{ADC}) \]

\[ R_{DMD} \ll R_{ADC} \]

\[ F_{\text{optimal}} = \frac{R_{ADC}}{R_{DMD}} \]

typically only 1000s of pixels

Implications. Measurement rate of a full-frame sensor but with a fraction of the number of pixels! (less than 0.1% pixels)
1. Use a linear array of pixels (a line-sensor)
Line-Sensor-based compressive camera (LiSens)

1. Use a linear array of pixels (a line-sensor)
2. Add a cylindrical lens

- objective lens
- relay lens
- cylindrical lens
- line sensor
- digital micromirror device (DMD)
LiSens

(a) Image on DMD
(b) Code on DMD
(c) Coded image
(d) Image on sensor plane
$X \in \mathbb{R}^{L \times N}$

where $\phi_t \in \mathbb{R}^N$, $t = 1, 2, 3 \ldots$

$y_t = X\phi_t + e_t$

(a) Image on DMD  (b) Code on DMD  (c) Coded image  (d) Image on sensor plane
Sensing images

\[ y_t = X\phi_t + e_t \]

\[ Y = X\Phi + E \]

(a) If \( T \geq \# \text{ of unknowns} \)
\[ \hat{X} = Y\Phi^+ \]

(b) If \( T < \# \text{ of unknowns} \)
\[ \min_{X} \text{TV}(X), \text{s.t. } \|Y - X\Phi\|_F \leq \varepsilon \]
\[ \text{TV}(\cdot), \text{2D total variation} \]
Sensing videos

Group together M readouts to reconstruct one frame
Assume sparse spatio-temporal gradients

Limitation: high-speed movements get blurred
Hardware prototype

- Objective lens
- Relay lens
- SPC
- Line-sensor
- Cylindrical lens
- DMD
Measurement rate of LiSens

1MHz
- sensor, 1000 fps
- 1024 pixels

DMD, DLP 7000
Sensor, Hamatsu S11156-2048-01
Line sensor vs. 2D sensor array

- Alignment is significantly easier

(a) Image on DMD  (b) Code on DMD  (c) Coded image  (d) Image on sensor plane
Line sensor vs. 2D sensor array

- Alignment is significantly easier
- 1D sensor profile provides space on top/bottom
  - 100% fill factor
  - Easier and cheaper to obtain additional tech. including frame transfer and per-pixel ADC
Simultaneous exposure and readout
Without frame transfer

Sequential exposure and readout
With frame transfer

Simultaneous exposure and readout
Without frame transfer

Frame transfer
No light loss during readout
Photos and videos
Comparison against SPC

Capture duration: 880ms 440ms 220ms 110ms

LiSens
0.8M meas.

SPC
16k meas.
Summary

Compressive sensing works much better for high-res images.

Replace single pixel in SPC by multiple pixels

- optimal # of pixels = \( \frac{R_{ADC}}{R_{DMD}} \)
- measurement rate = \( \frac{R_{ADC}}{R_{ADC}} \)
Advantages of using line sensor

- Alignment much simpler
- Much space in sensor
  - Easier and cheaper to obtain additional tech.
    - frame transfer, per-pixel ADC;
    - hyperspectral data
  - Without lose of the 100% fill factor
- Widely used in spectroscopy
  - Precise, low-noise, HDR, broad spectral response