Motion Estimation and Classification in Compressive Sensing from Dynamic Measurements

Aug 28, 2014

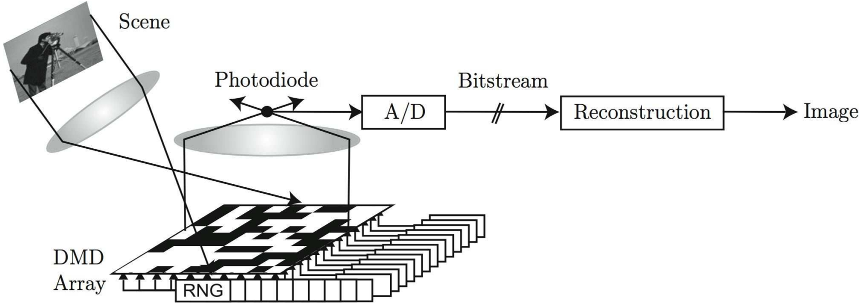
Vijay Rengarajan, *A.N. Rajagopalan*, R. Aravind

Department of Electrical Engineering
Indian Institute of Technology Madras
http://www.ee.iitm.ac.in/ipcvlab

OUTLINE

- Compressed sensing (CS) - an overview
- Motion in CS
- Motion estimation using CS measurements
  - Estimation of a single warp
  - Blockwise motion estimation
- Application: Compressive classification

CS HARDWARE



DMD : Digital micromirror device
RNG : Random number generator

CS MEASUREMENT AND RECONSTRUCTION

Measurement:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{1}$$

Reconstruction:

$$\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\boldsymbol{\alpha}}, \tag{2}$$

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}\mathbf{x} \text{ and } \mathbf{x} = \mathbf{\Psi}\boldsymbol{\alpha}$$

$\mathbf{x}$: signal/image $\in \mathbb{R}^N$

$\mathbf{y}$: measurement vector $\in \mathbb{R}^M, M \ll N$

$\mathbf{\Phi}$: measurement matrix $\in \mathbb{R}^{M \times N}$

$\mathbf{\Psi}$: basis, in which $\mathbf{x}$ is sparse

$\boldsymbol{\alpha}$: basis coefficients $\in \mathbb{R}^N$

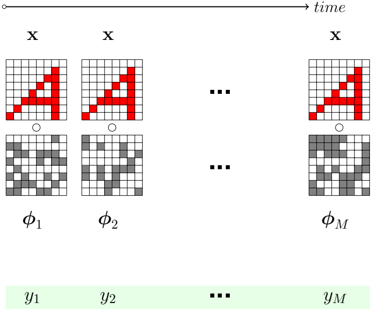
THE PROBLEM

- Sequential acquisition of measurements in a compressed sensing (CS) camera, such as a single pixel camera, lures temporal artifacts
- Manifestation of such artifacts is different from that of a conventional optical camera
- Relative motion between the camera and the scene renders the CS reconstruction process noisy
- We estimate the relative motion using a CS measurement vector against a CS reference vector (which is acquired when there is no motion)
- We assume a planar scene captured with only a global relative motion

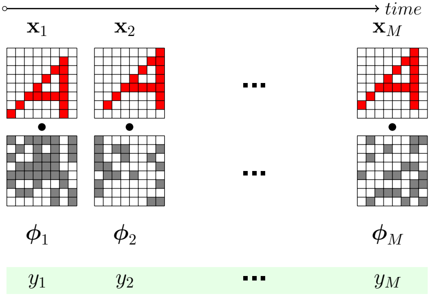
RELATED WORKS

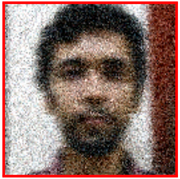
- Davenport et al. (2007) *Electronic Imaging*
  - Compressive classification problem is discussed in which the probe measurement is considered to be affected by only a single warp
  - Gallery contains CS measurements of all possible warps of images
- Duarte et al. (2007) *ICIP*
  - CS measurements are acquired at various image resolutions, and a multi-scale classification is discussed
- Both the methods consider the case where the observed CS vector is affected by a single warp

MOTION IN CS



no motion





CS

conventional

WARP ESTIMATION

- Given an image $\mathbf{x}$ and its warped version $\mathbf{x_p}$, it is possible to estimate the warp $\mathbf{p}$ using an energy minimisation based on Taylor series

$$\widehat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{x_p} - \mathbf{x}(\mathbf{p})\|_2^2 \tag{3}$$

$$\text{with } \mathbf{x}(\mathbf{p}) = \mathbf{x} + \nabla\mathbf{x} \cdot \mathbf{J}(\mathbf{x}) \cdot \mathbf{p}$$

- An iterative algorithm could be developed such that the residual energy $\|\mathbf{x_p} - \mathbf{x}(\mathbf{p})\|_2^2$ reduces in each iteration (Baker and Matthews, 2003)
- **Problem**: Given $\mathbf{y}$, CS measurement with no motion, and $\mathbf{y_p}$, CS measurement with a single warp, how to estimate the parameter vector $\mathbf{p}$?
- Is it possible to develop an algorithm to iteratively estimate $\mathbf{p}$? Will it converge?

WARP ESTIMATION

Image domain

$$\mathbf{x_p} = \mathbf{x} + \mathcal{D}(\nabla \mathbf{x})\, \mathbf{p} + \mathbf{e_x} \tag{4}$$

where uth row, $\mathcal{D}(\nabla \mathbf{x})[u,:] = \nabla \mathbf{x}[u,:]\, \mathbf{J}(u),$ for $u = 1, \ldots, N$

CS domain

$$\begin{aligned} \mathbf{y_p} &= \mathbf{\Phi x_p} \\ &= \mathbf{\Phi}(\mathbf{x} + \mathcal{D}(\nabla \mathbf{x})\, \mathbf{p} + \mathbf{e_x}) \\ \mathbf{y_p} &= \mathbf{y} + \mathbf{\Phi}\, \mathcal{D}(\nabla \mathbf{x})\, \mathbf{p} + \mathbf{e_y} \end{aligned} \tag{5}$$

CS problem

$$\widehat{\mathbf{p}} = \arg\min_{\mathbf{p}} ||\mathbf{e_y}||_2 \tag{6}$$

Will this converge?

WARP ESTIMATION

- A set of affine transformed images of the same scene forms a 6D manifold
- Suppose images $\mathbf{x}_1$ and $\mathbf{x}_2$ are points on this manifold $\mathcal{M}$. If $\mathbf{\Phi}$ is an orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$, then the projections of all images in the affine set using $\mathbf{\Phi}$ will form another manifold $\mathbf{\Phi}\mathcal{M}$
- The vectors $\mathbf{y}_1 = \mathbf{\Phi}\mathbf{x}_1$ and $\mathbf{y}_2 = \mathbf{\Phi}\mathbf{x}_2$ are points on this manifold $\mathbf{\Phi}\mathcal{M}$
- For $M = O(d \log(\mu N \epsilon^{-1})/\epsilon^2) < N$, where μ depends on the properties of the manifold such as volume and curvature, d is the dimension of the manifold, and $0 < \epsilon < 1$, the following holds with high probability (Baraniuk and Wakin 2009):

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{||\mathbf{\Phi}\mathbf{x}_1 - \mathbf{\Phi}\mathbf{x}_2||_2}{||\mathbf{x}_1 - \mathbf{x}_2||_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}} \qquad (7)$$

WARP ESTIMATION

- In our case, we have

$$(1-\epsilon)\sqrt{\frac{M}{N}}\|\mathbf{x_p}-\mathbf{x_{\hat{p}}}\|_2^2 \leq \|\mathbf{\Phi x_p}-\mathbf{\Phi x_{\hat{p}}}\|_2^2 \leq (1+\epsilon)\sqrt{\frac{M}{N}}\|\mathbf{x_p}-\mathbf{x_{\hat{p}}}\|_2^2$$

- For $\epsilon \approx 0$,

$$\|\mathbf{\Phi x_p}-\mathbf{\Phi x_{\hat{p}}}\|_2^2 \approx \lambda\|\mathbf{x_p}-\mathbf{x_{\hat{p}}}\|_2^2$$

  for some constant λ. Hence, a monotonic decrease of residual energy is assured with high probability
- Since

$$\begin{aligned} M &= O(d \log(\mu N \epsilon^{-1})/\epsilon^2) \\ &\propto (\log \epsilon^{-1})/\epsilon^2 = O(1/\epsilon^2), \end{aligned}$$

  choose M sufficiently large to ensure that the algorithm converges

WARP ESTIMATION

Algorithm 1: $(\widehat{\mathbf{p}}, e)$ = estimate_motion$(\mathbf{y_p}, \mathbf{y}, \mathbf{\Phi})$

Initialise $\widehat{\mathbf{p}} = [1, 0, 0, 1, 0, 0]^T$. Determine $\mathbf{x}$ from $\mathbf{y}$

repeat

- Warp $\mathbf{x}$ and $\nabla\mathbf{x}$ by $\widehat{\mathbf{p}}$ to get $\mathbf{x}_{\widehat{\mathbf{p}}}$ and $\nabla\mathbf{x}_{\widehat{\mathbf{p}}}$ respectively
- Calculate descent matrix, $\mathbf{S} = \mathbf{\Phi}\ \mathcal{D}(\nabla\mathbf{x}_{\widehat{\mathbf{p}}}) \in \mathbb{R}^{M\times 6}$
- Calculate Hessian, $\mathbf{H} = \mathbf{S}^T\mathbf{S} \in \mathbb{R}^{6\times 6}$
- Calculate $\widehat{\mathbf{y}} = \mathbf{\Phi}\ \mathbf{x}_{\widehat{\mathbf{p}}}$
- Calculate $\Delta\mathbf{p} = \mathbf{H}^{-1}\mathbf{S}^T(\mathbf{y_p} - \widehat{\mathbf{y}})$
- Update $\widehat{\mathbf{p}} = \widehat{\mathbf{p}} + \Delta\mathbf{p}$

until $\widehat{\mathbf{p}}$ converges

return $\widehat{\mathbf{p}}$ and residual energy $e = \|\mathbf{y_p} - \widehat{\mathbf{y}}\|_2^2$

In real cases, sequential acquisition affects the CS measurement vector continuously

Algorithm 2:

$\left(\{\widehat{\mathbf{p}}^{(j)}\}, \{e^{(j)}\}\right)$ = recursive_estimator($\mathbf{y}_{\mathbf{p}}, \mathbf{y}, \mathbf{\Phi}$)

Let L = length$(\mathbf{y}_{\mathbf{p}})$, $j = 0$

$(\widehat{\mathbf{p}}, e)$ = estimate_motion($\mathbf{y}_{\mathbf{p}}, \mathbf{y}, \mathbf{\Phi}$)

if $e > \tau$ **and** $L \geq 2B_{\text{min}}$ **then**

recursive_estimator($\mathbf{y}_{\mathbf{p}}[1 : \frac{L}{2}], \mathbf{y}, \mathbf{\Phi}[1 : \frac{L}{2}, :]$)

recursive_estimator($\mathbf{y}_{\mathbf{p}}[\frac{L}{2} + 1 : L], \mathbf{y}, \mathbf{\Phi}[\frac{L}{2} + 1 : L, :]$)

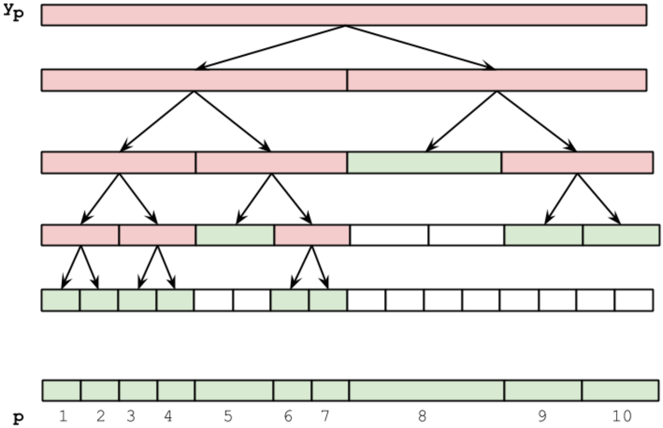
else

$j = j + 1$

return $\widehat{\mathbf{p}}^{(j)} = \widehat{\mathbf{p}}$ and $e^{(j)} = e$

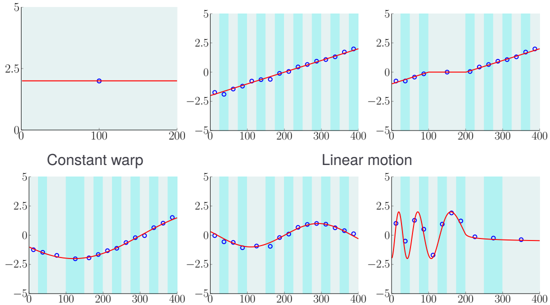
end if

ADAPTIVE BLOCKWISE ESTIMATION



RESULTS

- $\mathbf{x}$: 64 × 64 image from FERET database
  $\mathbf{\Phi}$: scrambled Hadamard matrix
  $\mathbf{y_p}$: horizontal translatory motion

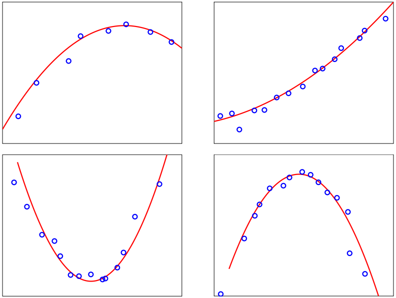


Oscillatory motion

x-axis : measurement number
blue shading : divided blocks
y-axis : camera translation in pixels
blue circle : estimated motion for a block
red line : actual camera motion

RESULTS

- $\mathbf{x}$: 64 × 64 image from FERET database
  $\mathbf{\Phi}$: scrambled Hadamard matrix
  $\mathbf{y_p}$: 2D translation camera motions



red line : actual camera motion
blue circle : estimated motion

COMPRESSIVE CLASSIFICATION

- 200 persons in FERET database
- One CS gallery vector per person
- To assign the incoming probe vector to a gallery person

$$c^* = \arg \min_{1 \leq c \leq C} \{e_c\}, \text{ where } e_c = \frac{1}{M} \sum_{j=1}^{Q} e_c^{(j)}$$

and $j = 1, \ldots, Q$ represents the block number in Algorithm 2.

Recognition results (in %) on FERET database for continuous camera motion during acquisition

	$M = 200$	
Type of motion	Adaptive	No motion estimation
(r_z)	93.5	55.0
(t_x, t_y)	95.5	52.5
(t_x, t_y, r_z)	95.0	53.0
Affine	88.0	48.5

CONCLUSION

- Proposed an algorithm to estimate relative motion between camera and scene during acquisition of CS measurements
- Discussed how a descent algorithm can be formulated to estimate the motion parameter vector from these measurements
- Demonstrated the utility of our motion estimation framework in the CS domain for the face recognition problem