HERIOT WATT UNIVERSITY

Signal Processing with Side Information

A Geometric Approach via Sparsity

João F. C. Mota

Heriot-Watt University, Edinburgh, UK


Side Information

Signal processing tasks

Denoising

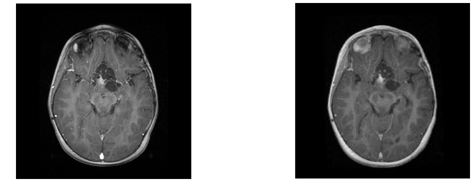
Reconstruction

Demixing (source separation)

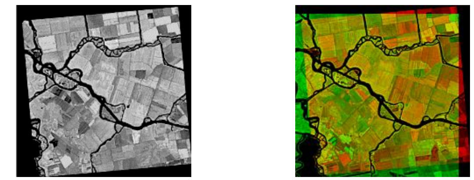
Compression

Inpainting, super-resolution, …

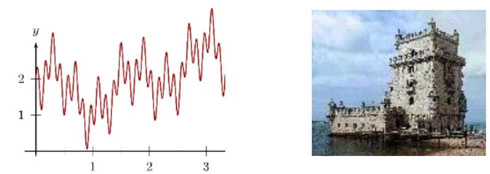
prior information



multi-modal

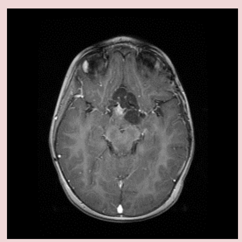


heterogeneous

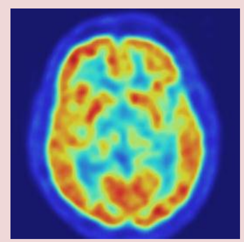


Recommender systems

Medical imaging



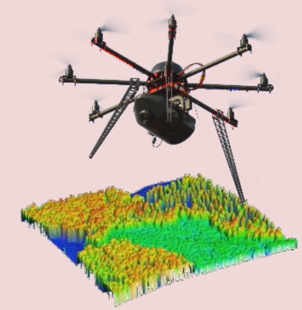
MRI



PET

Consumer electronics

Robotics

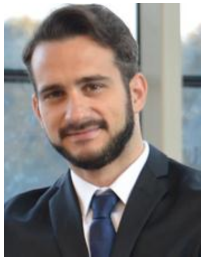


How to represent multi-modal or heterogeneous data ?

How to process it ?

Outline

- Compressed Sensing with Prior Information
- Application: Video Background Subtraction
- X-ray Image Separation
- Conclusions



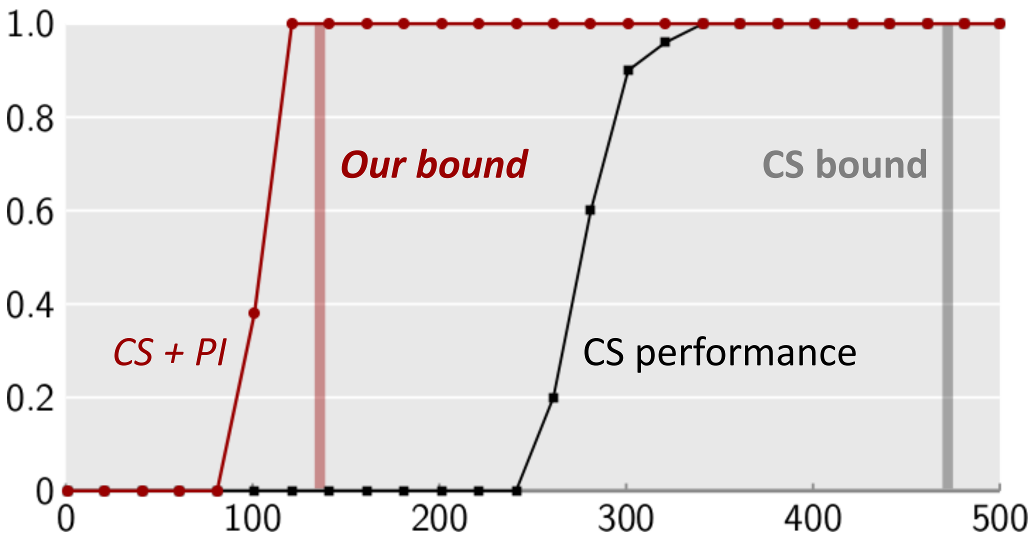
N Deligiannis
VUB-Belgium

M Rodrigues
UCL

Compressed Sensing

$\|\overline{x} - x^\star\|_2 / \|x^\star\|_2 \simeq 0.45$

Sucess rate (50 trials)



number of measurements m

Compressed Sensing (CS)

70-sparse

$$b = A : m \times n \quad x^\star$$

iid Gaussian

1000

Basis pursuit

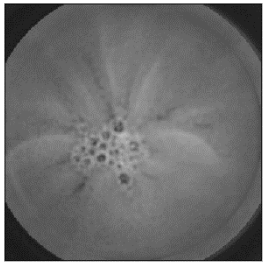
$$\widehat{x} = \underset{x}{\arg\min} \; \|x\|_1 \quad \text{s.t.} \quad b = Ax$$

What if we know $\overline{x} \sim x^\star$? ***prior information***

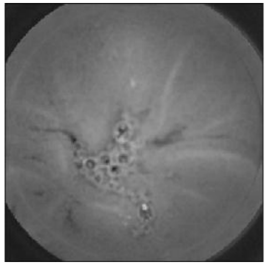
How do we integrate $\overline{x}$ in the problem?

Reconstruction guarantees?

$\overline{z} = \Psi \overline{x}$



$z^\star = \Psi x^\star$



medical images, video, …

Intuition

$Ax^\star = b$ measurements

random orientation

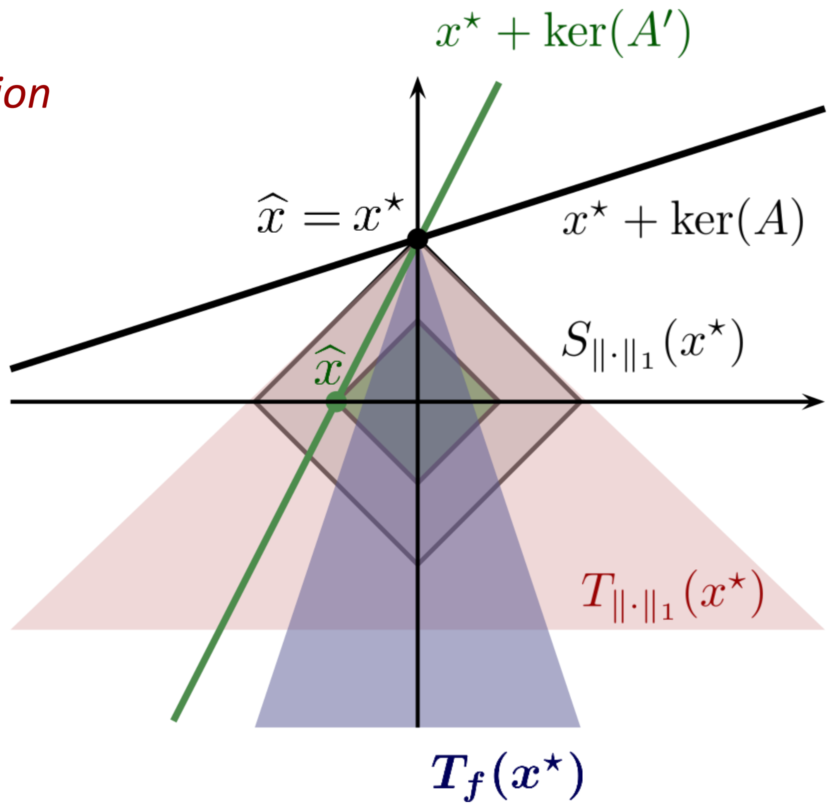
solutions of $Ax = b$: $x^\star + \ker(A)$

$$\widehat{x} = \underset{x}{\operatorname{argmin}} \ \|x\|_1 \quad \text{s.t.} \quad Ax = b$$

Tangent cone of f at $x^\star$

$$T_f(x^\star) = \operatorname{cone}\big(S_f(x^\star) - x^\star\big)$$

$\{x : f(x) \le f(x^\star)\}$



Our approach

$\overline{x}$: prior information (PI)

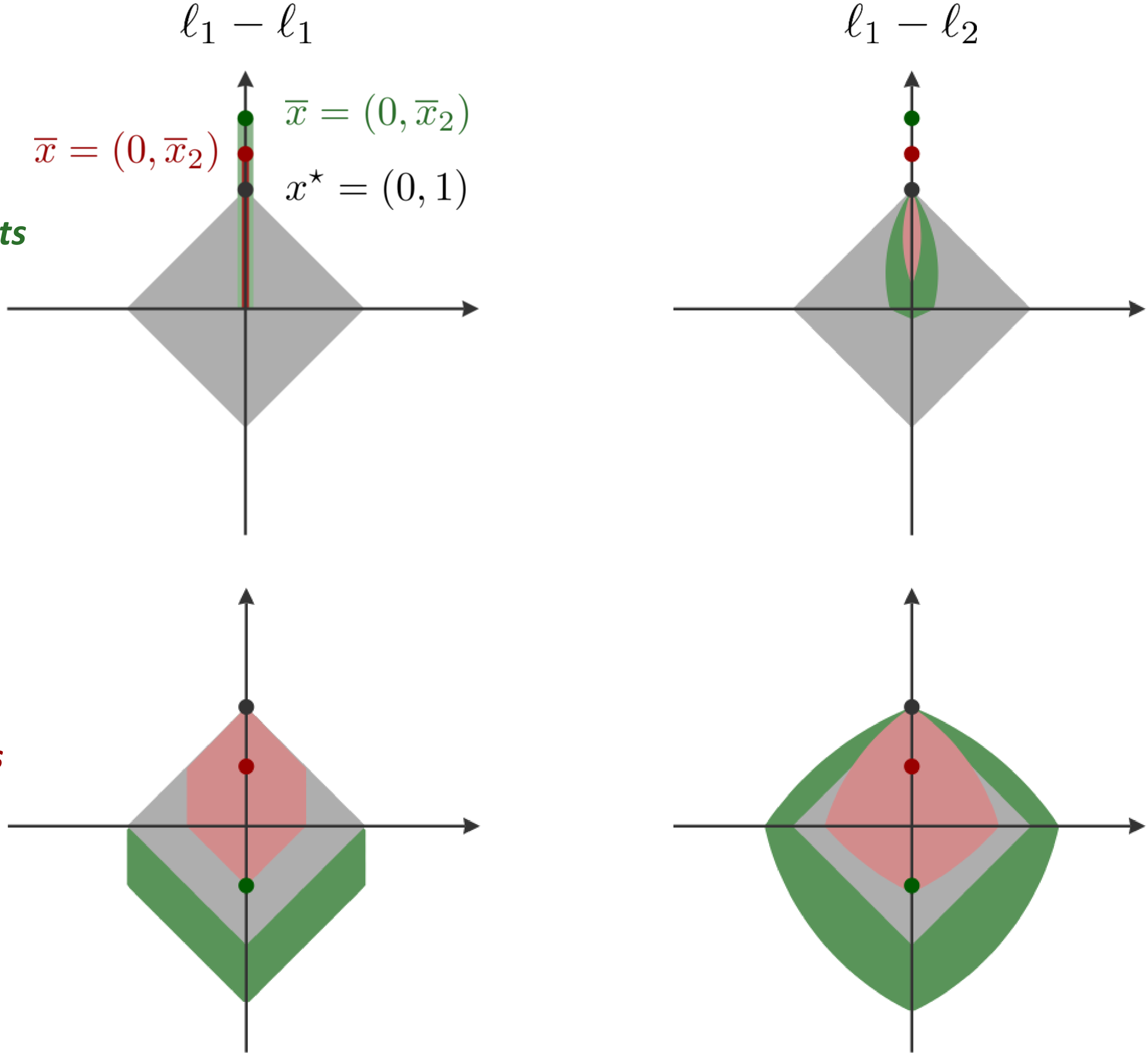
g : model for PI $\quad g(x^\star - \overline{x}) \simeq$ small

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) = \|x\|_1 + \beta\, g(x - \overline{x}) \\ \text{subject to} & Ax = b \end{array} \Rightarrow T_f(x^\star)$$

$$g_1(x - \overline{x}) = \|x - \overline{x}\|_1$$

$$g_2(x - \overline{x}) = \tfrac{1}{2}\|x - \overline{x}\|_2^2$$

$\beta = 1$



Good components

$\overline{x}_2 > x_2^\star$

Bad components

$\overline{x}_2 < x_2^\star$

L1-L1 minimization

$b = A : m \times n$ *i.i.d.* $\mathcal{N}(0, 1/m)$ $x^\star$ (s-sparse)

Theorem (BP) **[Chandrasekaran, Recht, Parrilo, Willsky, 2012]**

$$m \geq 2s \log\left(\frac{n}{s}\right) + \frac{7}{5}s + 1 \quad \Longrightarrow \quad x^\star = \underset{x}{\operatorname{argmin}} \ \|x\|_1 \ \text{ s.t. } \ b = Ax \qquad \text{w.h.p.}$$

Theorem (L1-L1 minimization) **[M, Deligiannis, Rodrigues, 2017]**

parameter-free

$\beta = 1$

$$\overline{h} := \Big|\{i \,:\, x_i^\star > 0,\, \overline{x}_i < x_i^\star\} \cup \{i \,:\, x_i^\star < 0,\, \overline{x}_i > x_i^\star\}\Big| > 0$$

$$m \geq 2\overline{h} \log\left(\frac{n}{s + \xi/2}\right) + \frac{7}{5}\left(s + \frac{\xi}{2}\right) + 1 \quad \Longrightarrow \quad x^\star = \underset{x}{\operatorname{argmin}} \ \|x\|_1 + \|x - \overline{x}\|_1 \ \text{ s.t. } \ b = Ax \qquad \text{w.h.p.}$$

$$0 \leq \overline{h} \leq s$$

$$\xi = \underbrace{\Big|\{i \,:\, \overline{x}_i \neq x_i^\star = 0\}\Big|}_{\text{support overestimation}} - \Big|\{i \,:\, \overline{x}_i = x_i^\star \neq 0\}\Big|$$

Experimental Results

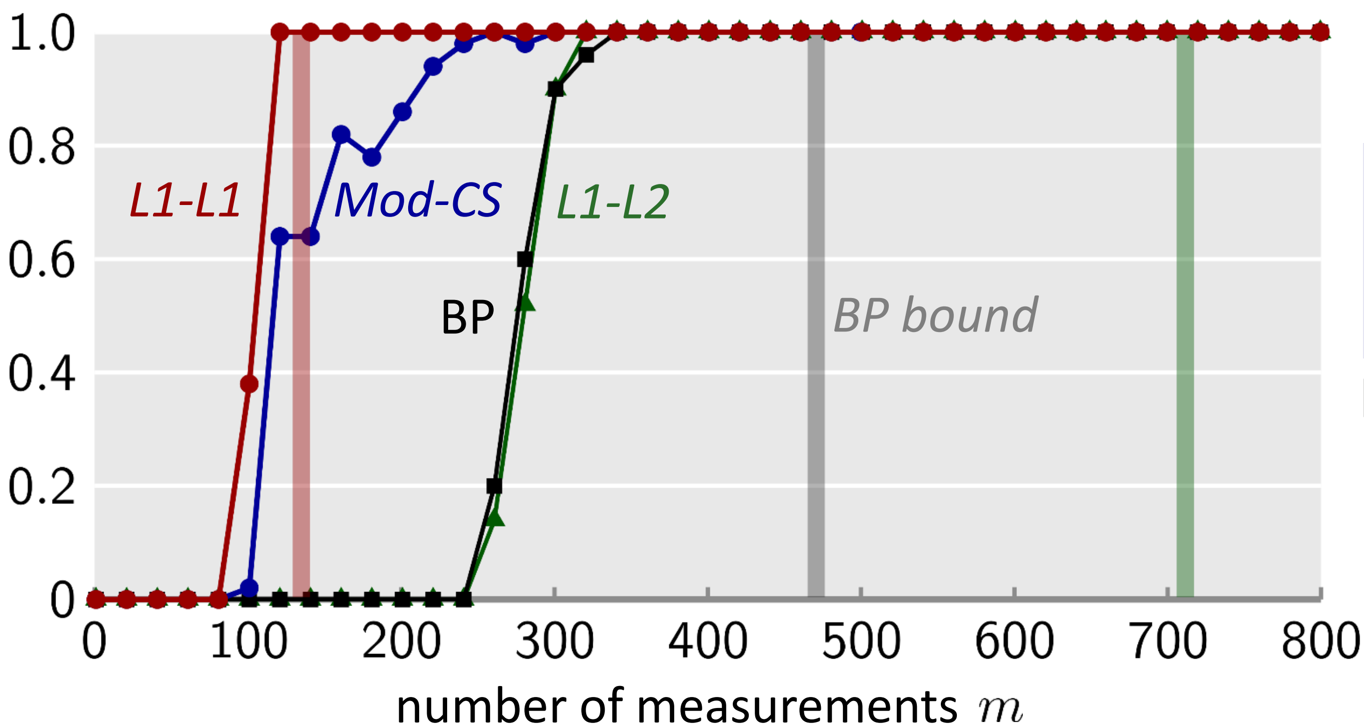
A_{ij} : Gaussian

$x^\star : \; n = 1000 \qquad s = 70 \qquad x_i^\star \sim \mathcal{N}(0,1) \qquad \|\overline{x} - x^\star\|_2 / \|x^\star\|_2 \simeq 0.45$

$\overline{x} : \; \overline{x} = x^\star + z \qquad \mathrm{card}(z) = 28 \qquad z_i \sim \mathcal{N}(0, 0.8) \qquad \overline{h} = 11 \quad \xi = -42$

$\|\hat{x} - x^\star\|_\infty / \|x^\star\|_\infty \leq 10^{-2}$

Sucess rate (50 trials)

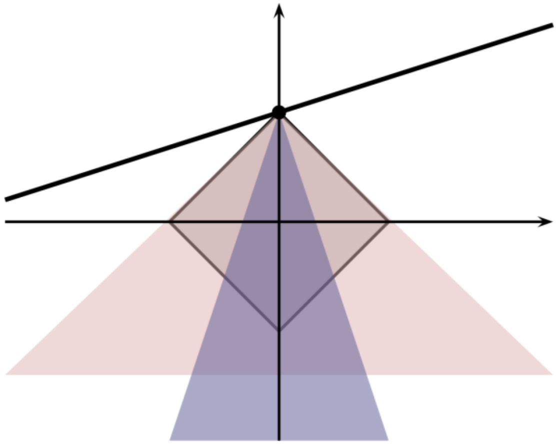


Mod-CS

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x_{T^c}\|_1 \\ \text{subject to} & Ax = b \end{array}$$

[Vaswani and Lu, 2010]

Summarizing

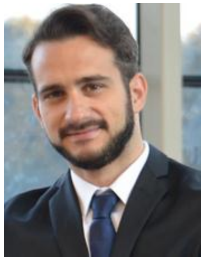


- Prior Information *can help*, but *can also hinder*
- L1-L1 works better than L1-L2 (theory and practice)
- (Computable) bounds are tight for L1-L1, but not for L1-L2
- Theory predicts optimal β ; indicates how to improve $\overline{x}$
- Limitations: Gaussian matrices; bounds depend on unknown parameters

$$m \geq 2\overline{h} \log\left(\frac{n}{s+\xi/2}\right) + \frac{7}{5}\left(s + \frac{\xi}{2}\right) + 1$$

Outline

- Compressed Sensing with Prior Information
- *Application: Video Background Subtraction*
- X-ray Image Separation
- Conclusions



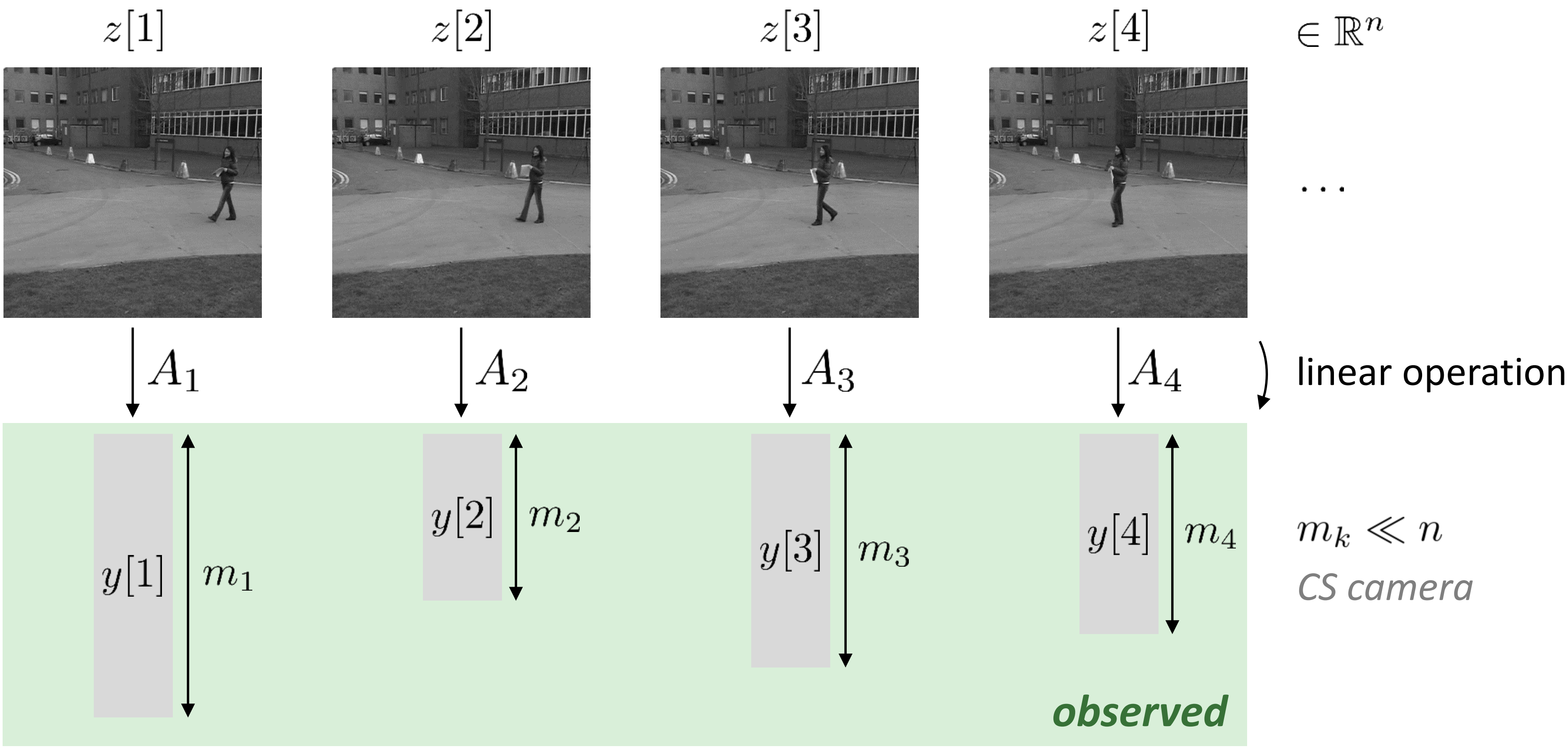
N Deligiannis
VUB-Belgium

M Rodrigues
UCL

A Sankaranarayanan
CMU-USA

V Cevher
EPFL-CH

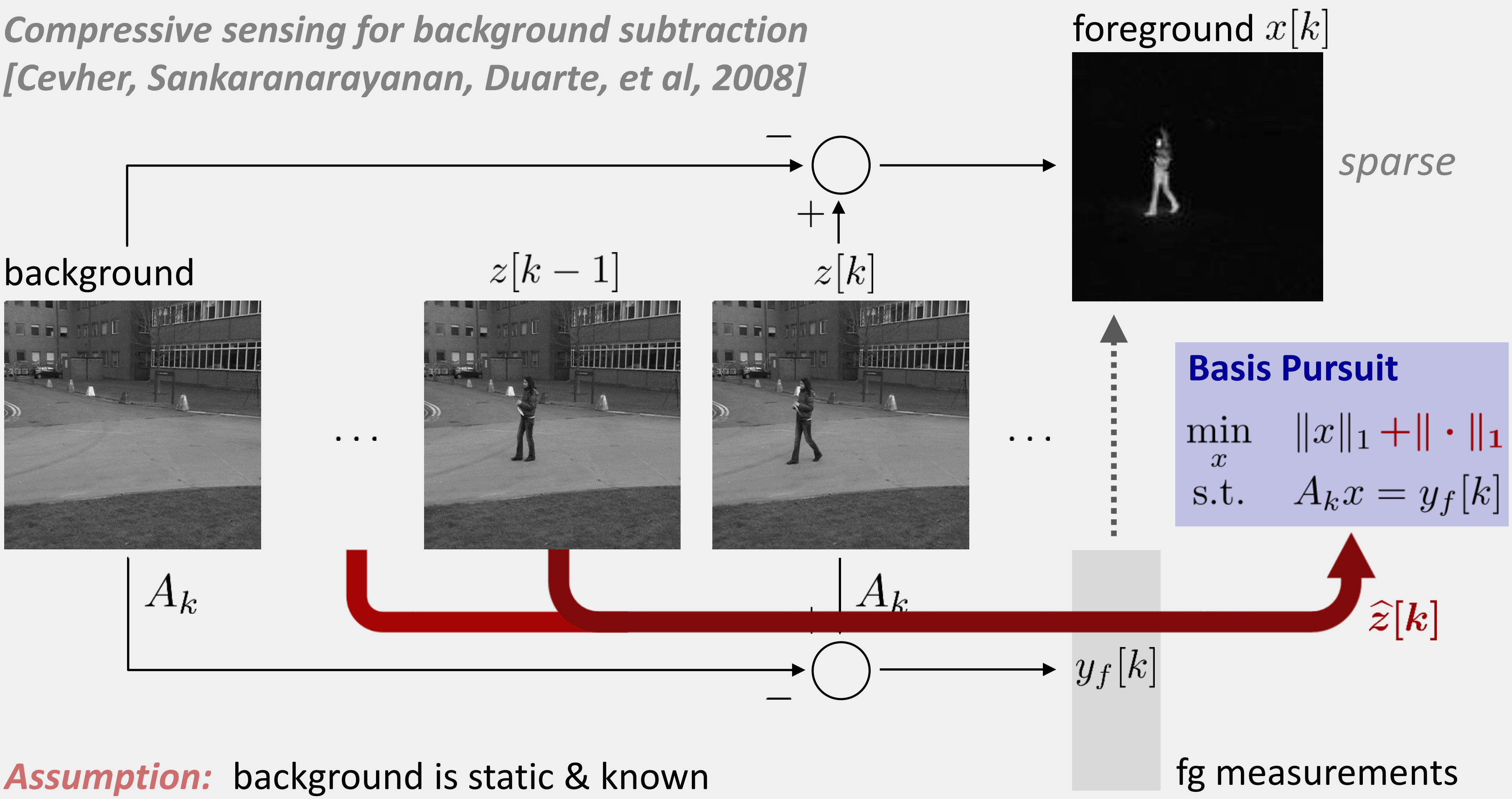
Compressive Background Subtraction



How to recover $z[k]$ *from* $y[k]$ *online ?*

How many measurements m_k *from frame* $z[k]$ *?*

Compressive sensing for background subtraction
[Cevher, Sankaranarayanan, Duarte, et al, 2008]



Assumption: background is static & known

Problems

Prior frames are ignored

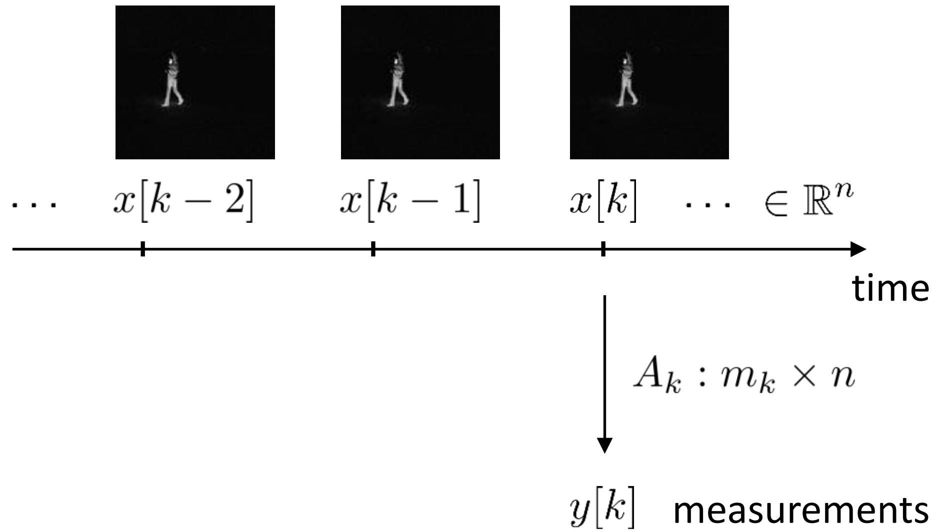
m_k : fixed; depends on foreground area

Our Approach

Estimate $z[k]$ from past frames: $\widehat{z}[k]$

Integrate $\widehat{z}[k]$ into BP via ***ℓ_1-ℓ_1 minimization***

Problem Statement



Model

sparse — arbitrary function — sparse

$$x[k] = f_k\Big(\big\{x[i]\big\}_{i=1}^{k-1}\Big) + \epsilon[k]$$

$$y[k] = A_k\, x[k]$$

Problem

Compute a minimal # of measurements m_k

Reconstruct $x[k]$ perfectly

→ *online algorithm w/ adaptive rate*

Algorithm

$$x[k] = f_k\Big(\{x[i]\}_{i=1}^{k-1}\Big) + \epsilon[k]$$

$$y[k] = A_k\, x[k]$$

Gaussian

m_k : computed at iteration $k-1$

Acquire $y[k] = A_k\, x[k]$ with $A_k : m_k \times n$

Set $w[k] = f_k\Big(\{\hat{x}[i]\}_{i=1}^{k-1}\Big)$ ***Estimate*** $x[k]$

$$\begin{aligned} \hat{x}[k] = \quad & \underset{x}{\arg\min} \quad \|x\|_1 + \big\|x - w[k]\big\|_1 \\ & \text{s.t.} \quad y[k] = A_k\, x \end{aligned}$$

L1-L1 minimization

$$\hat{\xi}_k = \big|\{i : w_i[k] \neq \hat{x}_i[k] = 0\}\big| - \big|\{i : w_i[k] = \hat{x}_i[k] \neq 0\}\big|$$

$$\hat{\overline{h}}_k = \big|\{i : \hat{x}_i[k] > 0,\ \hat{x}_i[k] > w_i[k]\} \cup \{i : \hat{x}_i[k] < 0, \hat{x}_i[k] < w_i[k]\}\big|$$

$$\hat{s}_k = \big|\{i : \hat{x}[k] \neq 0\}\big|$$

parameters of $x[k]$

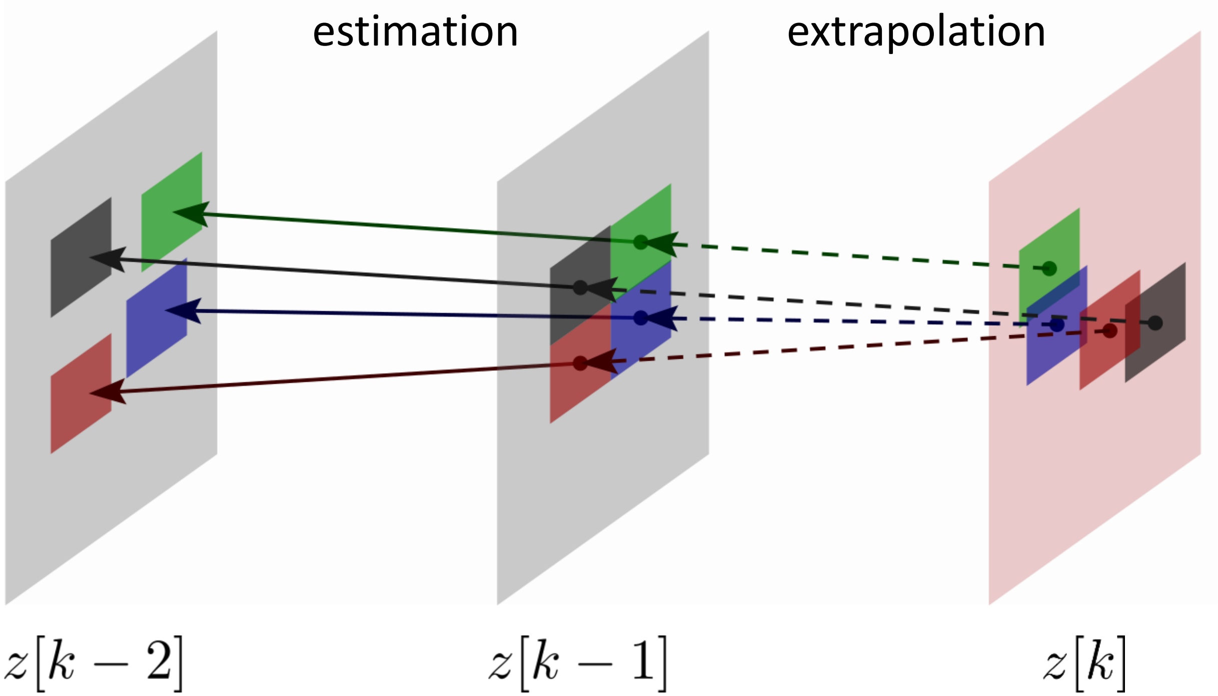
$$m_{k+1} = \left[2\hat{\overline{h}}_k \log\left(\frac{n}{\hat{s}_k + \hat{\xi}_k/2}\right) + \frac{7}{5}\left(\hat{s}_k + \frac{\hat{\xi}_k}{2}\right) + 1\right](1 + \delta_k)$$

 # measurements of $x[k+1]$

oversampling factor

$k \leftarrow k+1$ and repeat ...

Estimating a Frame



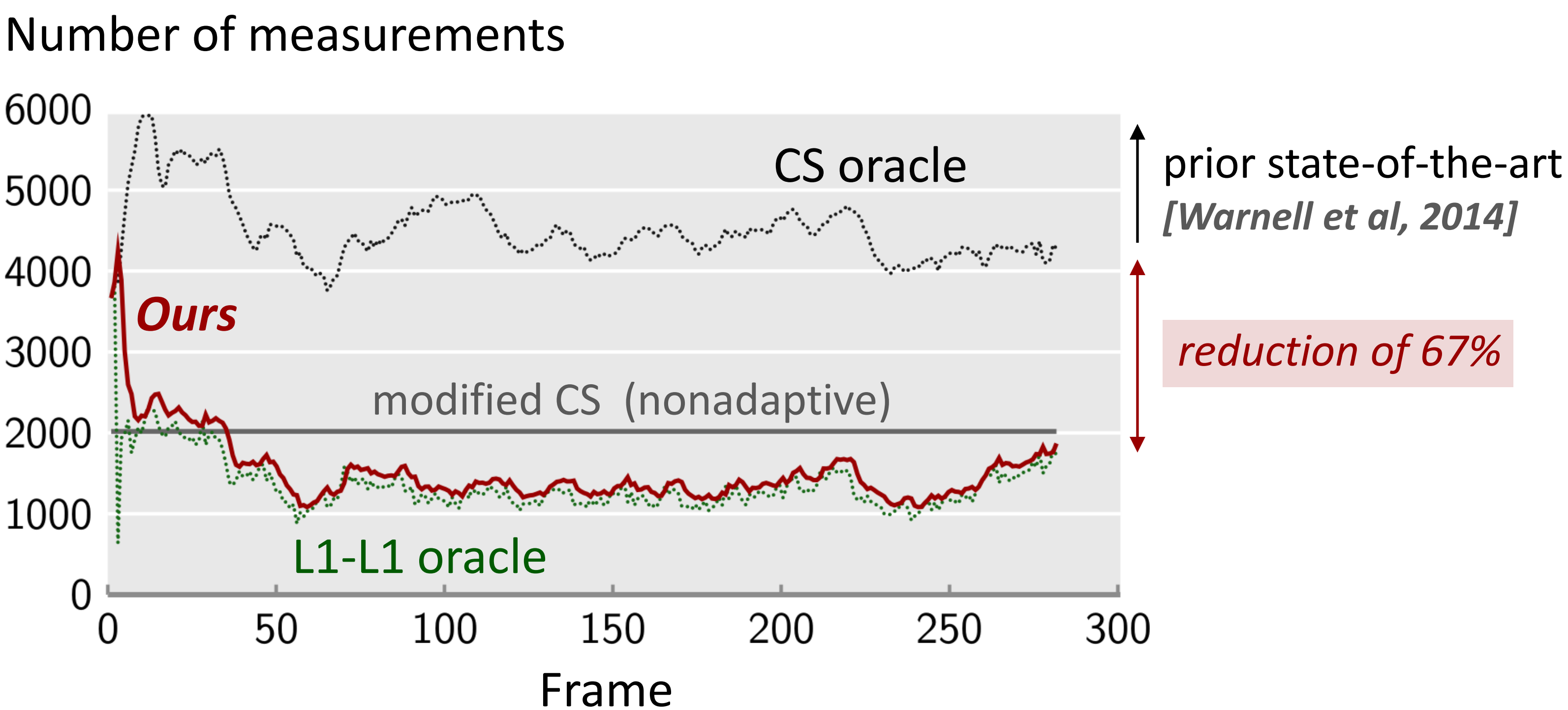
linear motion

overlap: take average; *gaps:* fill w/ average of neighbors

state-of-the-art in video coding

Experimental Results

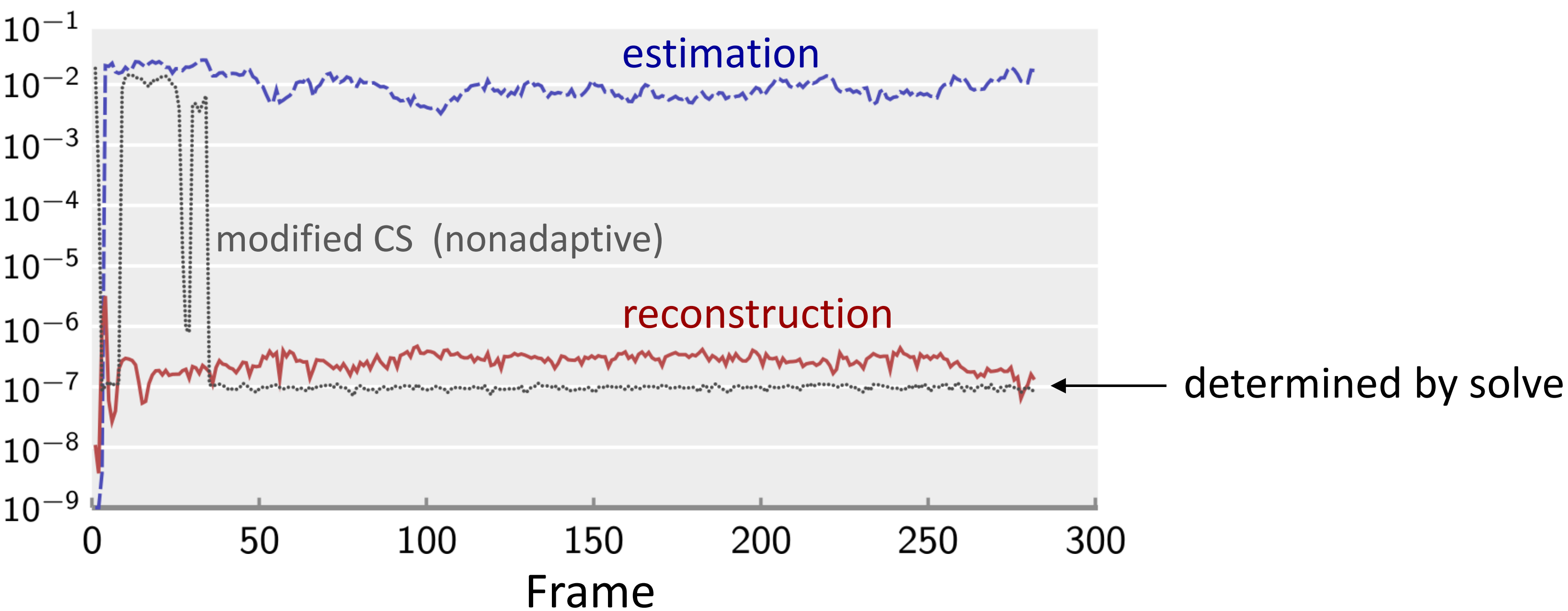
280 frames



Experimental Results

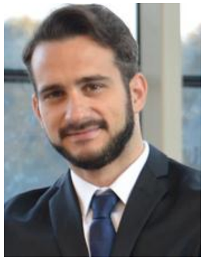
280 frames

Relative error



Outline

- Compressed Sensing with Prior Information
- Application: Video Background Subtraction
- *X-ray Image Separation*
- Conclusions



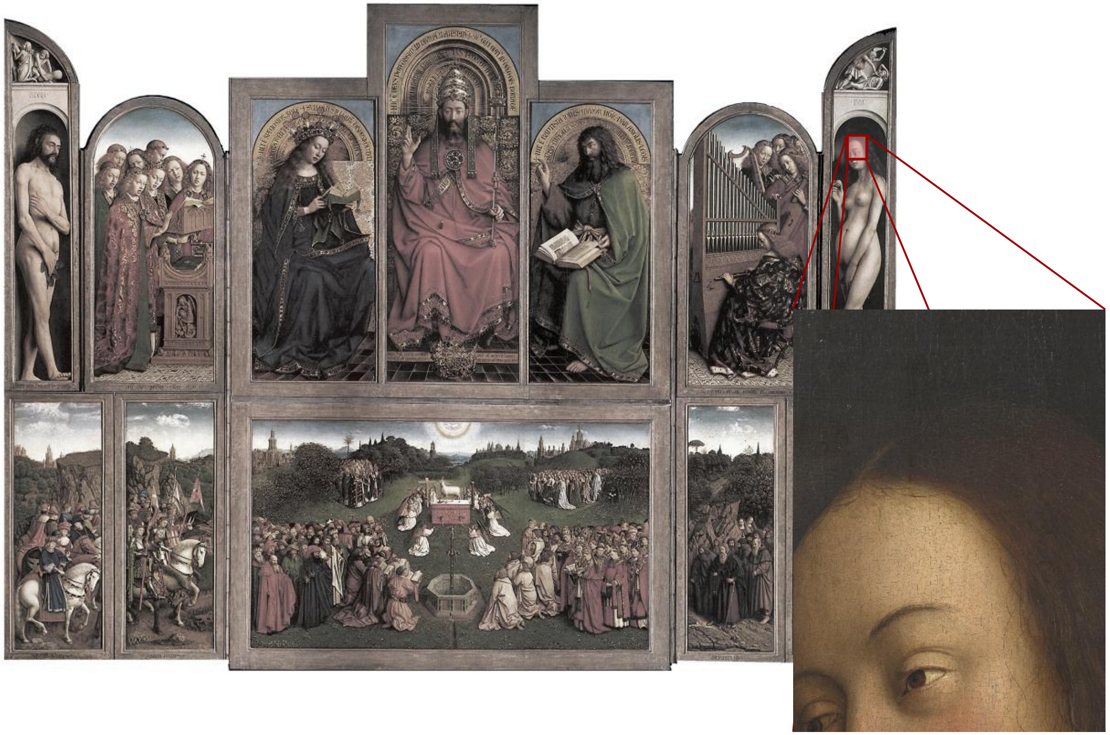
N Deligiannis
VUB-Belgium

M Rodrigues
UCL

B Cornelis
VUB-Belgium

I Daubechies
Duke-USA

Motivation: X-Ray of Ghent Altarpiece



Mixed X-Ray

Can we use the visual images to separate the x-rays?

Approach: Coupled Dictionary Learning

Training step

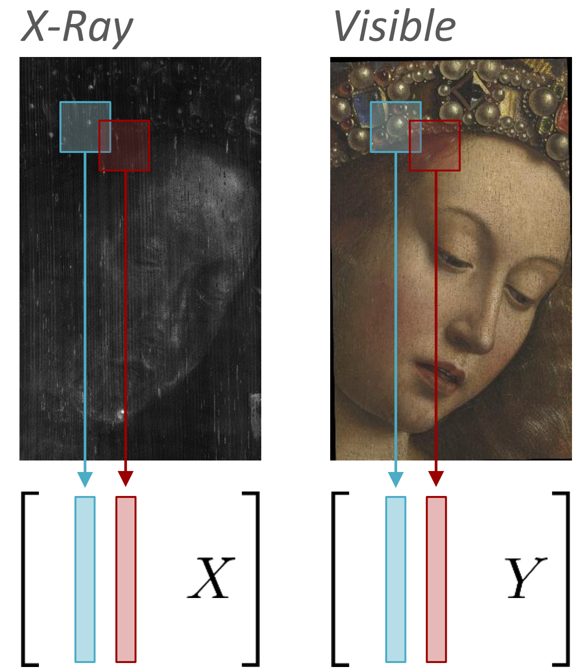
coupling

$$Y = \Psi^c Z$$
$$X = \Phi^c Z + \Phi V$$

w/ sparse columns

learn dictionaries by alternating minimization

$$\underset{\substack{\Psi^c, \Phi^c, \Phi \\ Z, V}}{\text{minimize}} \quad \|Y - \Psi^c Z\|_F^2 + \|X - \Phi^c Z - \Phi V\|_F^2$$
$$\text{subject to} \quad \text{card}(Z_i) \leq s_z\,, \quad i = 1, \ldots, T$$
$$\text{card}(V_i) \leq s_v\,, \quad i = 1, \ldots, T$$



Demixing step

mixed x-ray

$x = \Phi^c(z_1 + z_2) + 2\Phi v$

visual front

$y_1 = \Psi^c z_1$

visual back

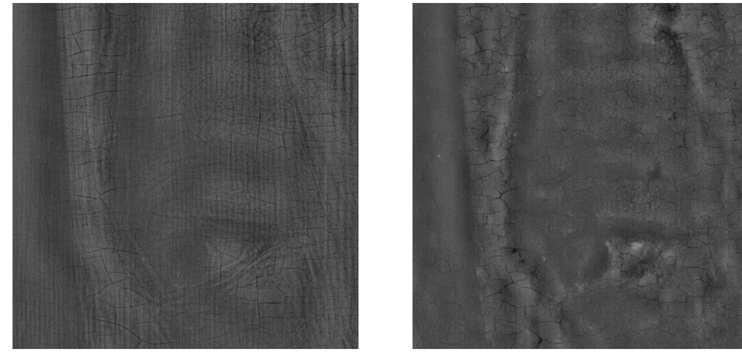
$y_2 = \Psi^c z_2$

$$\underset{z_1, z_1, v}{\text{minimize}} \quad \|z_1\|_1 + \|z_2\|_1 + \|v\|_1$$
$$\text{subject to} \quad x = \Phi^c(z_1 + z_2) + 2\Phi v$$
$$y_1 = \Psi^c z_1$$
$$y_2 = \Psi^c z_2$$

Results

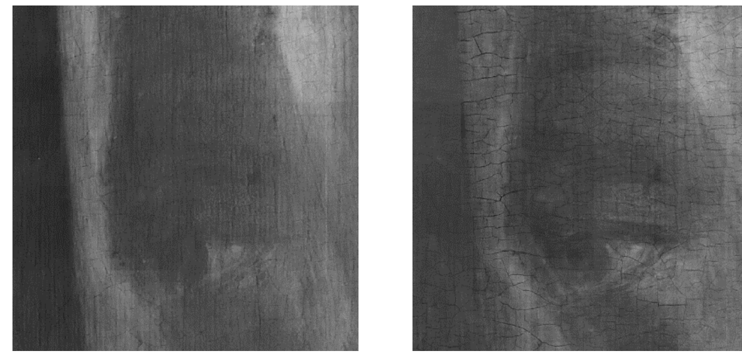
reconstructed x-rays

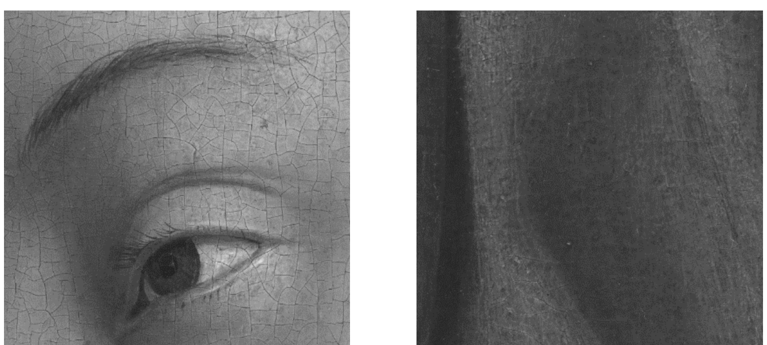
MCA ***[Bobin et al, 07']***



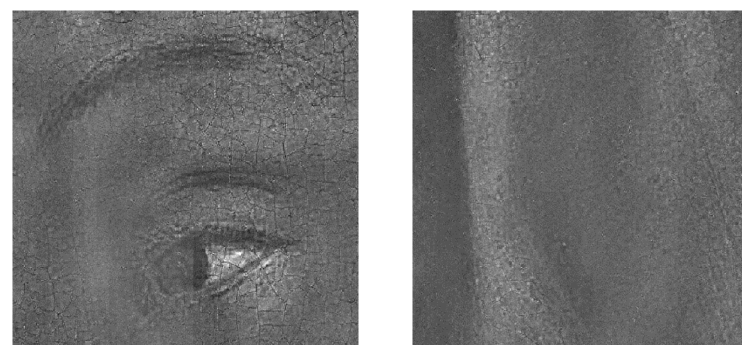
mixed x-ray

multiscale MCA w/KSVD





Ours



visuals in grayscale

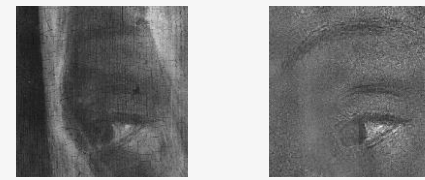
Summary / Conclusions

$$X = \Phi^c Z + \Phi V$$
$$Y = \Psi^c Z$$

data *dictionaries* *sparse columns*

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \Phi^c & \Phi \\ \Psi^c & 0 \end{bmatrix} \begin{bmatrix} Z \\ V \end{bmatrix}$$

X-ray separation



sparse

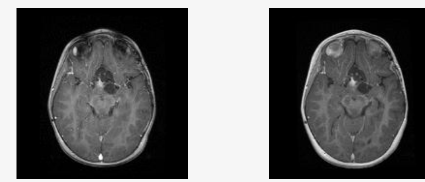
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & I \\ A & 0 \end{bmatrix} \begin{bmatrix} z \\ v \end{bmatrix}$$

sparse

$x = z + v$ prior information

$y = Az$ measurements

Reconstruction w/ PI



observations

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \ldots$$

multi-modal features

Low-rank model

Better models?

Guarantees?

Scalable algorithms?

Applications

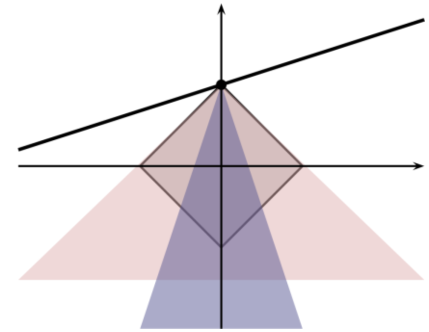
medical imaging (MRI + PET + ECG)

super-resolution (depth + visual)

SAR + microwave imaging

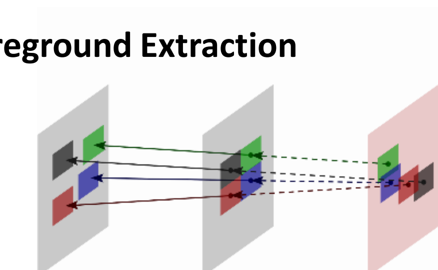
robotics (laser + sonar)

References



J. F. C. Mota, N. Deligiannis, M. R. D. Rodrigues
Compressed Sensing with Prior Information: Optimal Strategies, Geometries, and Bounds
IEEE Transactions on Information Theory, Vol 63, No 7, 2017

J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, M. R. D. Rodrigues
Adaptive-Rate Reconstruction of Time-Varying Signals with Application in Compressive Foreground Extraction
IEEE Transactions on Signal Processing, Vol 64, No 14, 2016



N. Deligiannis, J. F. C. Mota, B. Cornelis, M. R. D. Rodrigues, I. Daubechies
Multi-Modal Dictionary Learning For Image Separation With Application in Art Investigation
IEEE Transactions on Image Processing, Vol 26, No 2, 2017