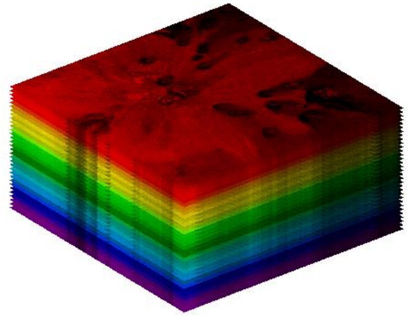


Compressed Sensing of Low Complexity High Dimensional Data

Application to Hyperspectral Imaging

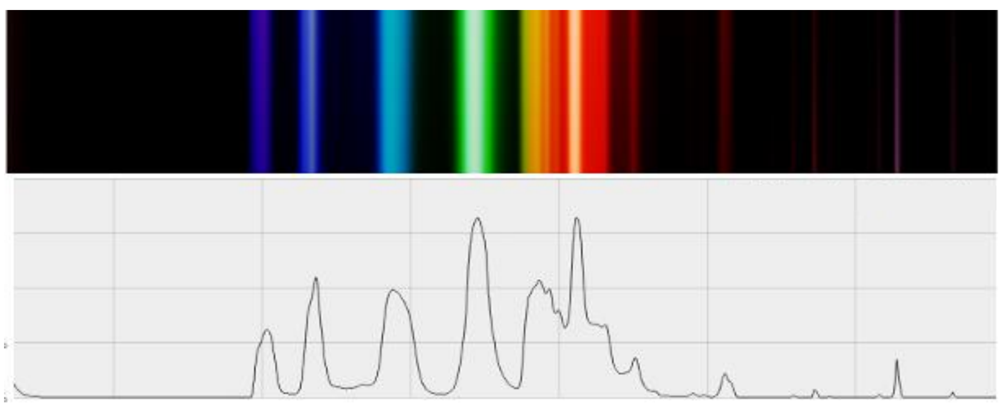
Kévin Degraux
PhD Student, ICTEAM institute
Université catholique de Louvain, Belgium

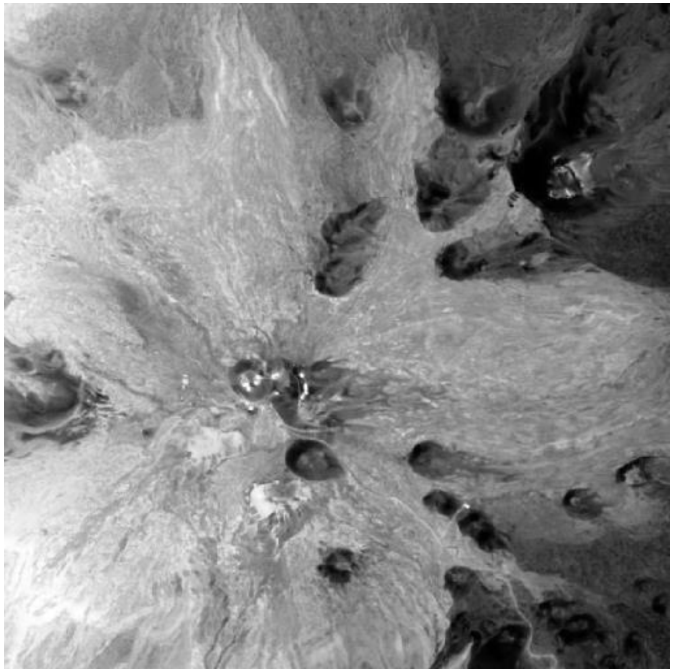
6 November, 2013

Hyperspectral imaging is the fusion of spectrometry and imaging

Spectrometry + Imaging

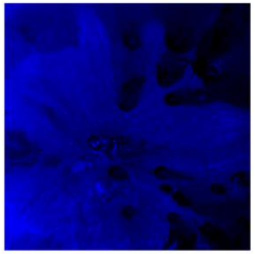
Emission (or absorption) spectrum

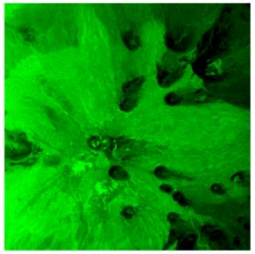


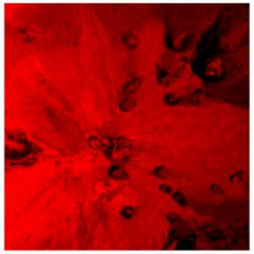


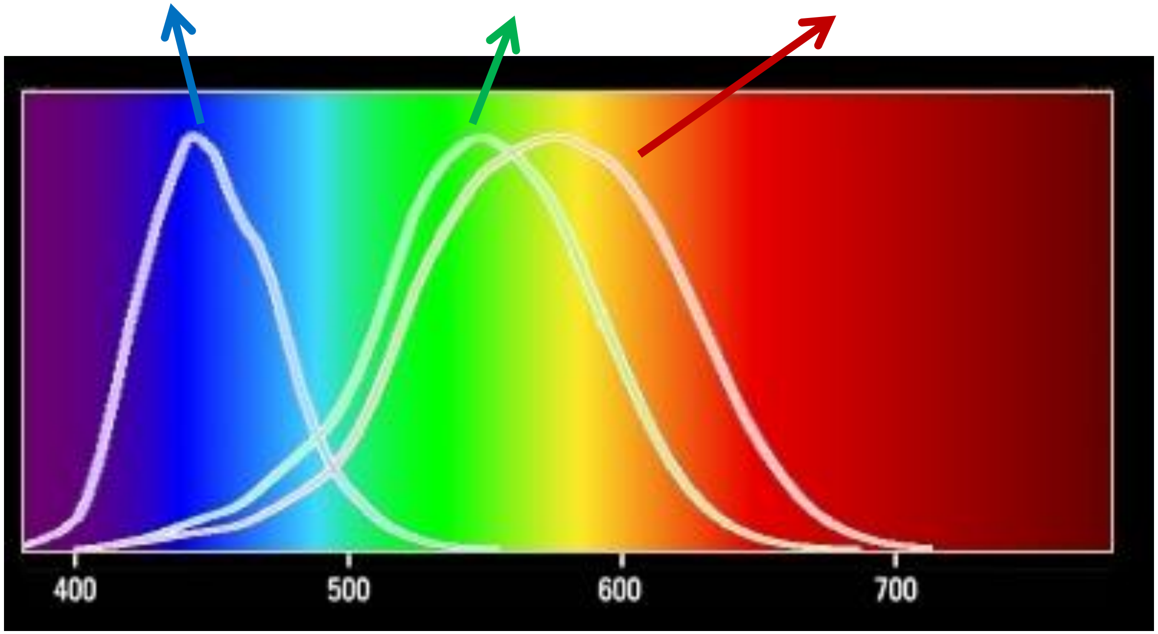
RGB imaging mimics the Human Visual System

3 large bands to get chrominance

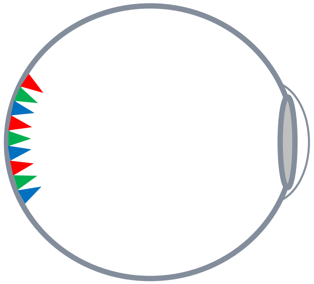








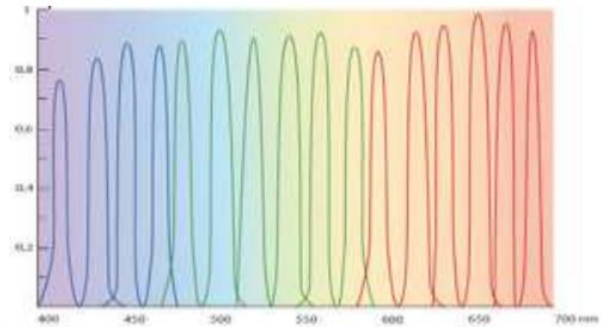
Short, Medium and Long retina cones

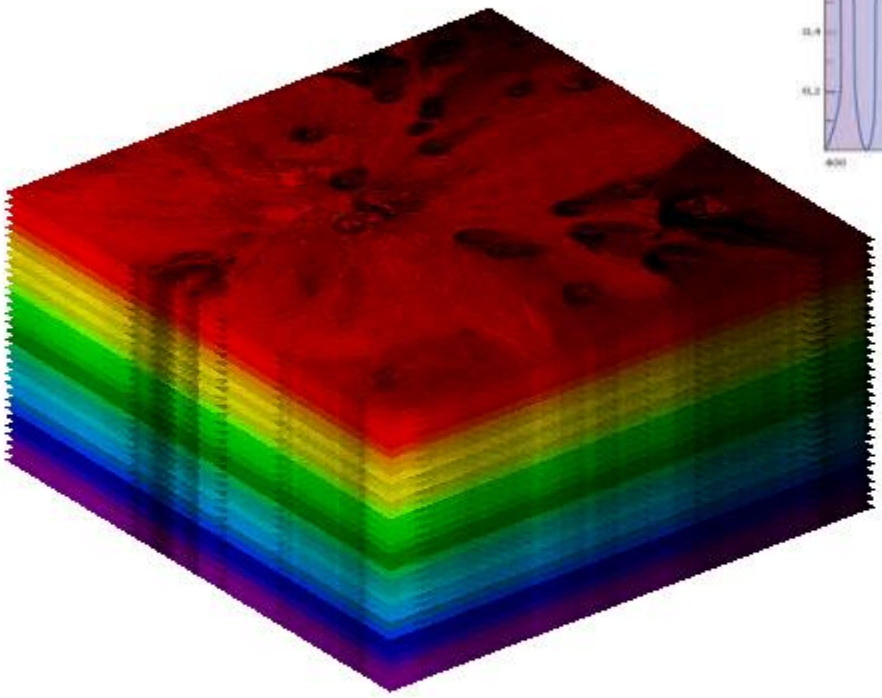


Multi/Hyper-spectral imaging goes beyond

Full spectrum (a lot of narrow bands)

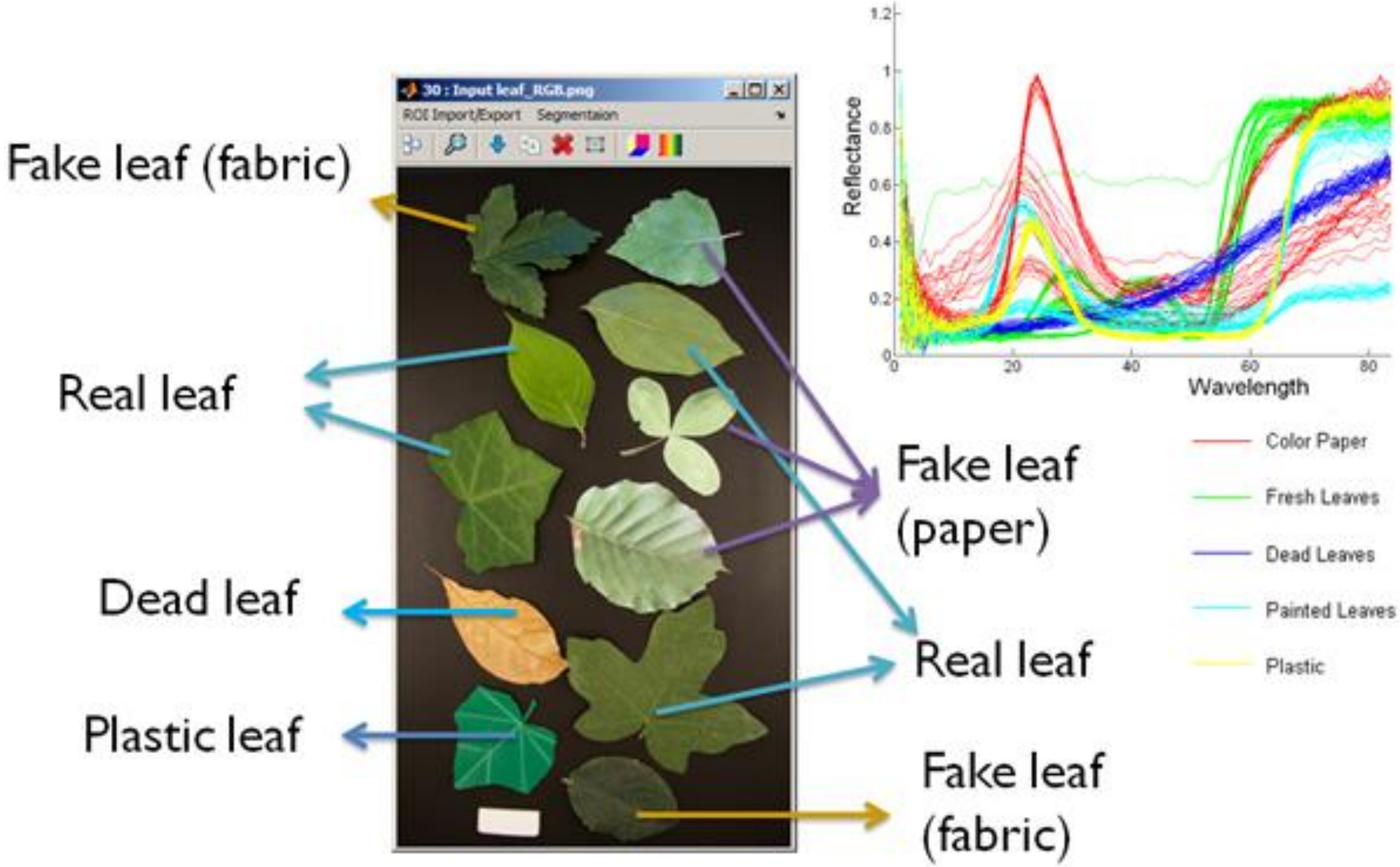
at every pixel





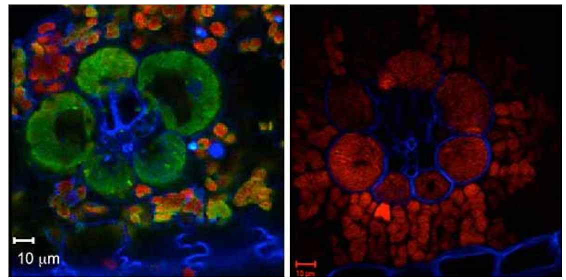
HSI mixes spatial and spectral information

Classify several materials spatially on an image thanks to their unique spectral signature (« fingerprint »).



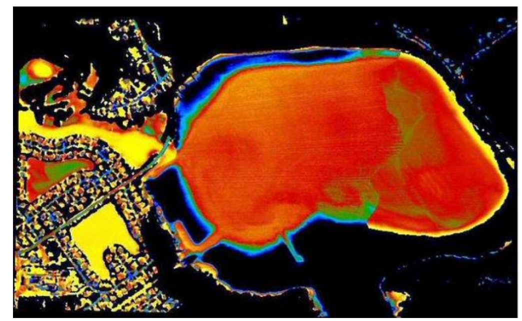
Applications

Microscopy, Spectroscopy



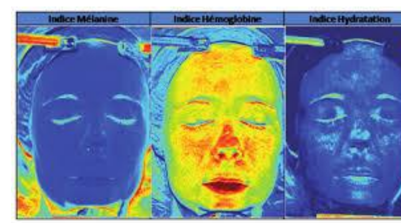
Counterfeit detection

Agriculture, Environmental Monitoring



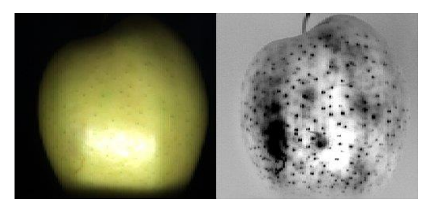
Biotechnology, Skin health, Endoscopy,...

Surveillance, Security and Defense

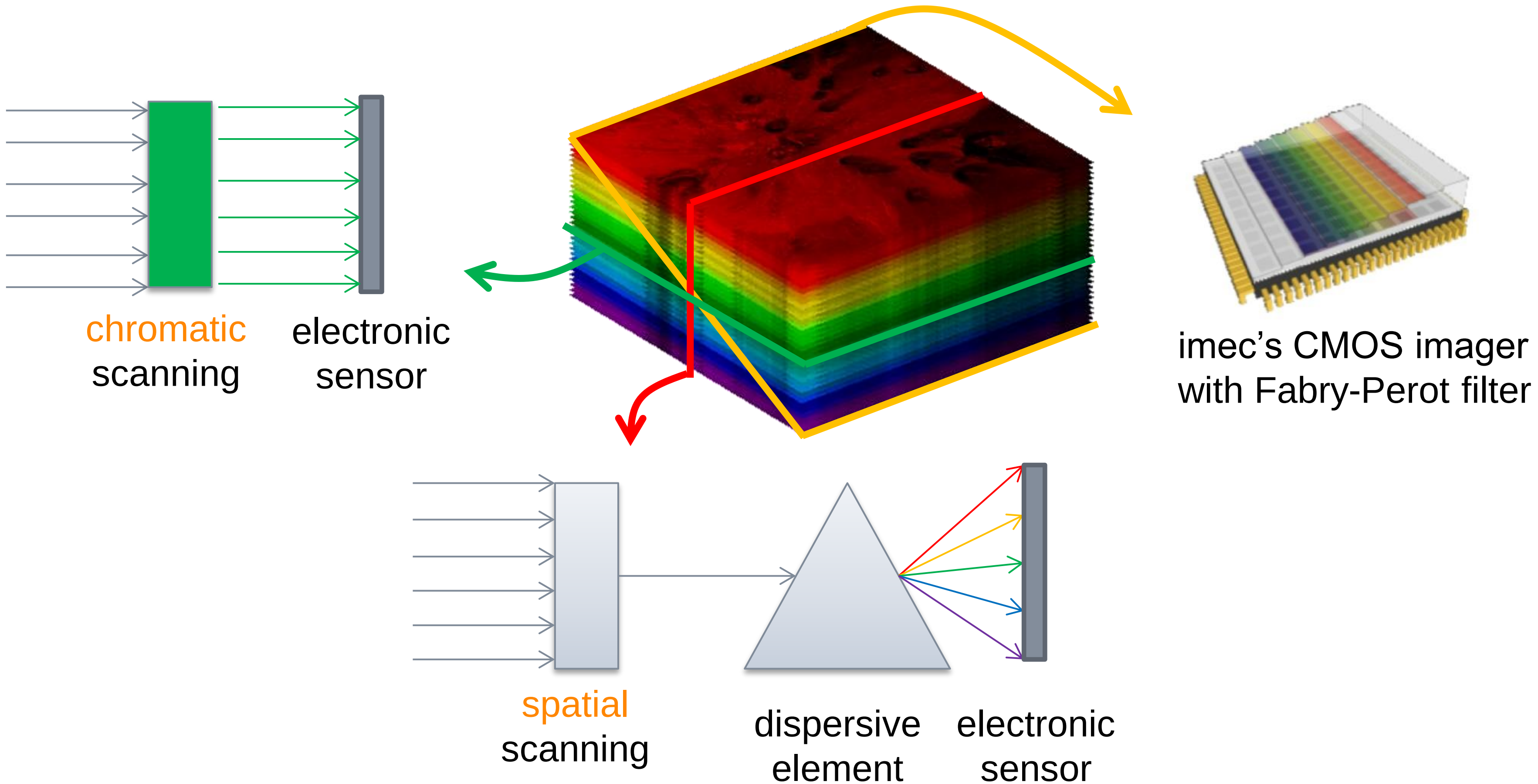


Non-contact quality control (e.g. Thin Films, Food, Pharmaceutical,...)

...



How is it usually done?



What are the issues?

Spatially very restrictive (low resolution, line scanning...)

Low speed and power consuming acquisition

Big amount of data (e.g. 200 spectral bands for every pixel) to acquire, compress, transmit and store.

Complex devices

Why "Low Complexity" HD signals?
The paradox

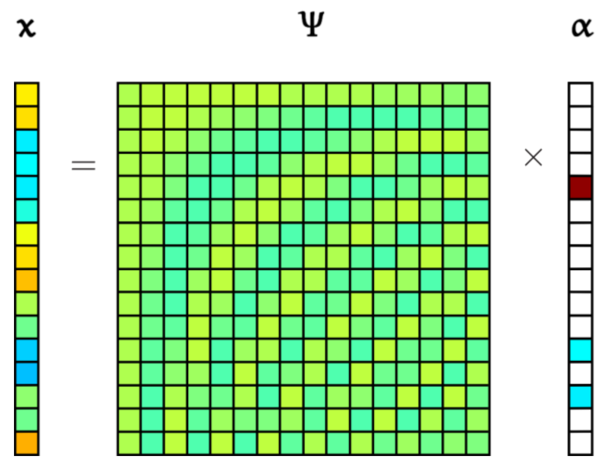
Huge amount of data

→Big effort to acquire (expensive sensors) **$** m

But high level of "redundancy"

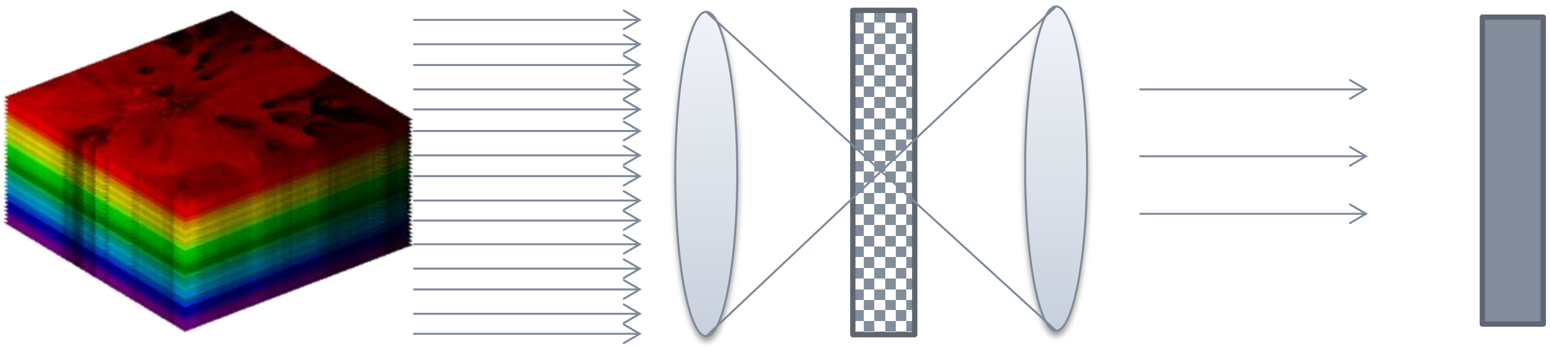
→Big effort to compress (expensive DSP) **$** m

example :
K-sparse in Ψ



HD Data is the ideal field for Compressed Sensing

Compression at the acquisition (e.g. optically).



$x \in \mathbb{R}^N$
Signal

Φ
Optical (≈ linear)
compressed sensing

$y \in \mathbb{R}^M$
measurements
(sensor)

Big gain both in the sensor and in the DSP.

HD Data is the ideal field for Compressed Sensing

Asymptotic theory → Works better at high dimensions

Classically in CS theory if K is the signal sparsity,

Compression rate

$$\frac{M}{N} \geq C \frac{K}{N} \log\left(\frac{N}{K}\right)$$

relative sparsity

For HD data K increases slower than N so,

$$\frac{K}{N} \downarrow \text{ and } \boldsymbol{\frac{M}{N}} \downarrow$$

Compressed Sensing of High Dimensional Data

How to?

Model good signal priors

Build and assess accurate signal model

Use suited sparsity basis Ψ (wavelets, DCT,...)

Combine sparsity bases $\Psi = \Psi_x \otimes \Psi_y \otimes \Psi_\lambda$

Append bases $\Psi = [\Psi_1, \Psi_2, \ldots]$ → dictionaries

Learn or design application specific dictionaries

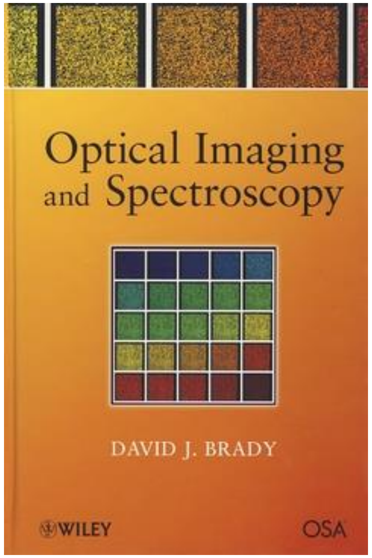
Use other low complexity priors e.g. TV, low rank (see later)

Design efficient sensing

In terms of physical implementation

➔Physically (e.g. optically) feasible

➔“Simple” efficient electronics



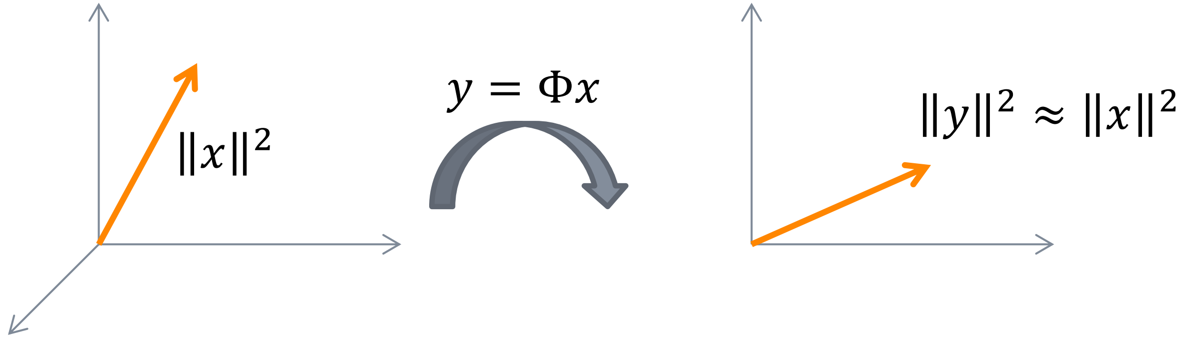
Design efficient sensing

In terms of mathematical properties

- Linear... Or not? (e.g. quantization);
- Fidel to reality;
- Restricted Isometry Property

$\exists\ 0 < \delta < 1$ such that for any “low complexity” (sparse, low rank,...) signal x,

$$(1-\delta)\|x\|^2 \le \|\Phi x\|^2 \le (1+\delta)\|x\|^2$$



Design efficient sensing

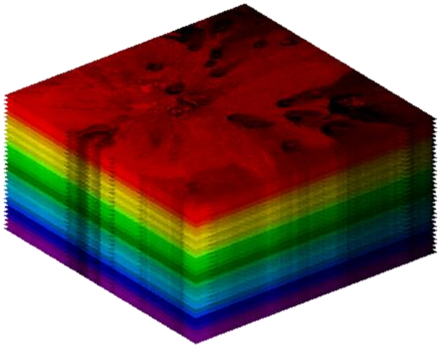
In terms of numerical algorithms efficiency

Φ = Bottleneck of reconstruction algorithms

➔Sparse, block sparse, block diagonal, binary matrices, FFT, FWT,...

➔NOT full random (e.g. Gaussian) matrices!!

For a 512x512x32 hyperspectral volume (8M voxels)
→ sensing matrix with 2^{46} entries i.e. 512 TB !



Reconstruct

Use prior to build an optimization program

assume x is sparse: low ℓ_0 “norm”

$$\arg\min_x \|x\|_0 \; s.t. \|\Phi x - y\|_2 \leq \epsilon$$

non convex ➔ hard to solve

Reconstruct

Use prior to build an optimization program

Either exact solution of convex relaxation

e.g. BPDN

$$\arg\min_x \|x\|_1 \;\; s.t. \; \|\Phi x - y\|_2 \leq \epsilon$$

convex

Or approximate solution of non convex problem

e.g. L0-LASSO

non convex

$$\approx \arg\min_x \|\Phi x - y\|_2^2 \;\; s.t. \; \|x\|_0 \leq K$$

Reconstruct

Choose the solver

see Parikh and Boyd's monograph
for a good intro

Generic slow convex optimization (proximal algorithms,...)

e.g. Douglas-Rashford, Chambolle-Pock

Dedicated non convex fast greedy methods

e.g. IHT, OMP, CoSaMP

Reconstruct

Implement: HD data is challenging

Optimized libraries (BLAS, LAPACK,…),

Randomization (power method, truncated SVD, ...)

Parallel computing

- Multi-core (e.g. OpenMP)
- GPU (CUDA, OpenCL…)
- HW accelerators (Xeon Phi,...)
- Clusters (MPI…)

…

Some examples

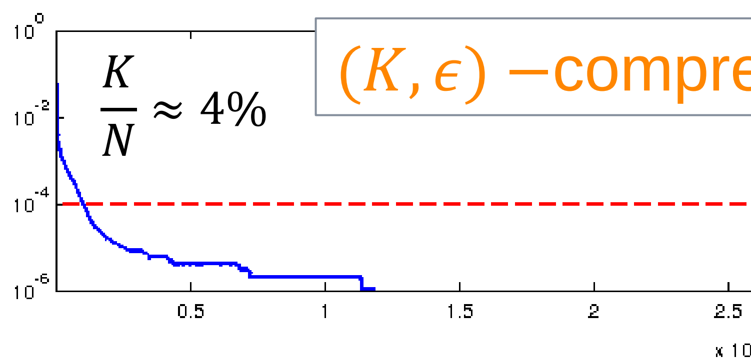
CS of a monochromatic image

512x512 Smile

Log Scale Haar DWT coefficients $\alpha/\|\alpha\|$

Sensor provides only 256x256 measurements ($\frac{M}{N} = 25\%$).

We can play with optics.



$\frac{K}{N} \approx 4\%$

(K, ϵ) −compressible

$\epsilon = 10^{-4}$

CS of a monochromatic image

Choice of the sensing

Spread Spectrum Random “Fourier” (DCT) Ensemble

$$\Phi = S\,F\,H$$

CS of a monochromatic image

Choice of the sensing

Spread Spectrum Random “Fourier” (DCT) Ensemble

$$\Phi = \underset{\substack{\text{Random} \\ \text{selection}}}{\underset{\downarrow}{S}}\ \underset{\text{DCT}}{\underset{\downarrow}{F}}\ \underset{\substack{\text{Diagonal } \pm 1 \\ \text{“spread spectrum”}}}{\underset{\downarrow}{H}}$$

CS of a monochromatic image

Choice of the sensing

Spread Spectrum Random "Fourier" (DCT) Ensemble

$$\Phi = S\,F\,H$$

Random selection

DCT

Diagonal ± 1 "spread spectrum"

Good math property if $\frac{M}{N} \geq C\,\frac{K}{N} \ln N$

CS of a monochromatic image

Choice of the sensing

Spread Spectrum Random "Fourier" (DCT) Ensemble

$$\Phi = S\,F\,H$$

Random selection

DCT

Diagonal ± 1 "spread spectrum"

Good math property if $\frac{M}{N} = 0.25 \geq C\,\frac{K}{N} \ln N \approx C\;0.49$

Small (?) constant
≈ 1 ?
< 1 ?

High ☹

CS of a monochromatic image

Choice of the sensing

Spread Spectrum Random “Fourier” (DCT) Ensemble

$$\Phi = S\,F\,H$$

Random selection — DCT — Diagonal ± 1 “spread spectrum”

Good math property if $\frac{M}{N} = 0.25 \geq C\,\frac{K}{N} \ln N \approx C\;0.49$

Small (?) constant

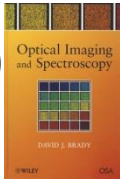
≈ 1 ?

< 1 ?

High ☹

Numerically efficient ☺

(optically feasible?)



Let’s try anyway…

CS of a monochromatic image

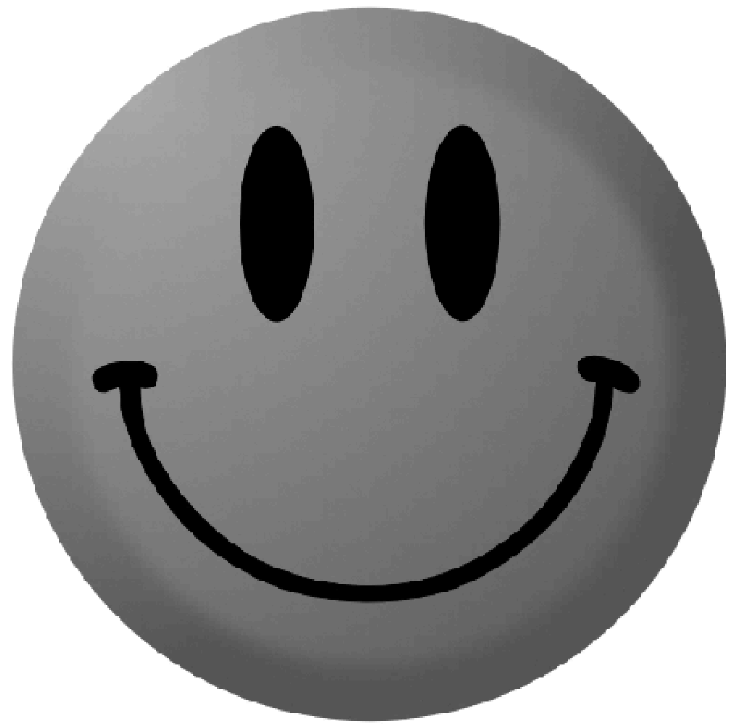
Result : BPDN solved with CP

“super-resolution”
without CS (but with BPDN)
Φ is a simple average over 4px blocks

PSNR = 32dB

+18dB

CS
Φ is the SSRFE



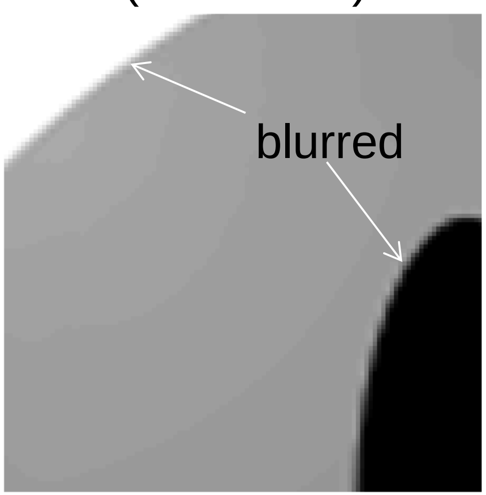
PSNR = 50dB

CS of a monochromatic image

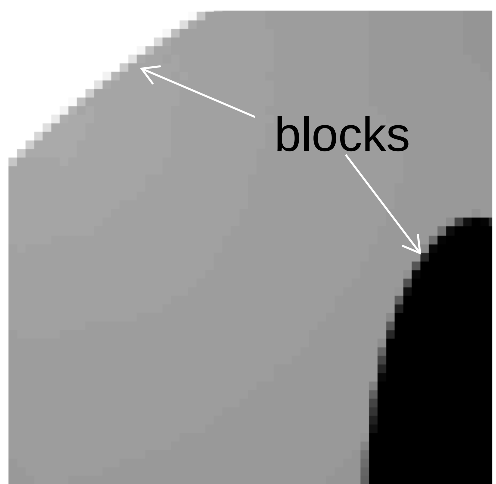
Result :

Original

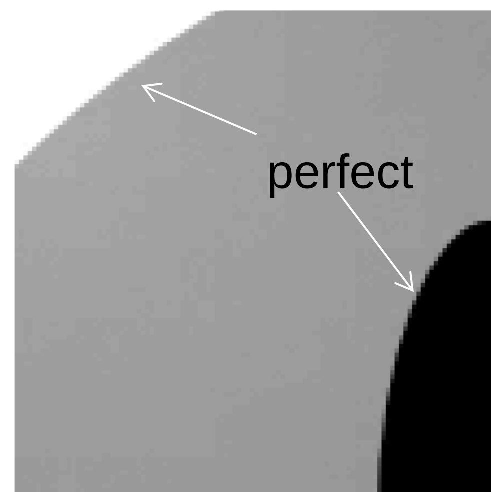
Bi-cubic interpolation
(not BPDN)



“super-resolution”
BPDN without CS



CS



CS of a monochromatic image

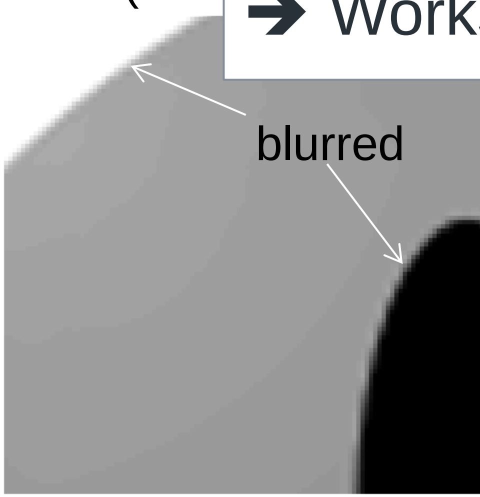
Result :

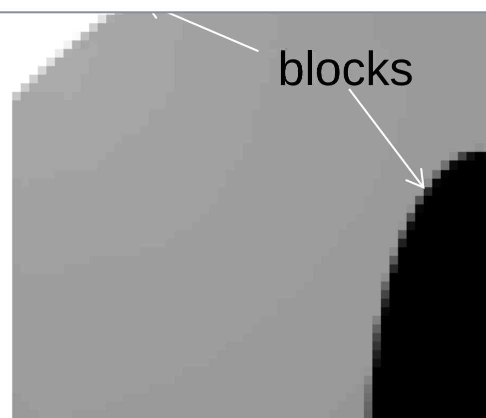
Original

This is a good small & sparse toy example.
More “natural” small (512x512) images are not sparse enough (Lena,…)

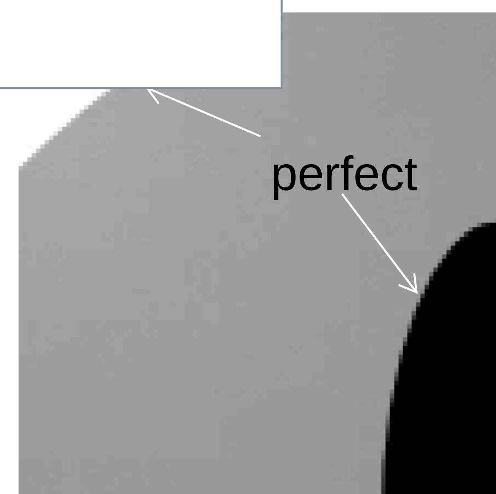
➔ Works better when $N \uparrow$ and $\frac{K}{N} \downarrow$

Bicubic i
(not





CS



Reshape

A discrete signal and in particular an image can also be mathematically treated as a matrix

$$x \in \mathbb{R}^N \rightarrow X \in \mathbb{R}^{n_1 \times n_2}$$

Low-Rank Prior

Rank and Singular Value Decomposition (SVD) of $X \in \mathbb{R}^{n_1 \times n_2}$

$$X = USV^* = \underbrace{[u_1 \quad \cdots \quad u_{n_1}]}_{\text{output unitary matrix}} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}}_{\text{singular values rank } r} \underbrace{\begin{bmatrix} v_1^* \\ \vdots \\ v_{n_2}^* \end{bmatrix}}_{\text{input unitary matrix}}$$

$$= \sum_{i=1}^{r} \sigma_i \underbrace{u_i v_i^*}_{\text{rank 1 matrices}}$$

Low rank means highly redundant (generalization of sparsity for matrices)

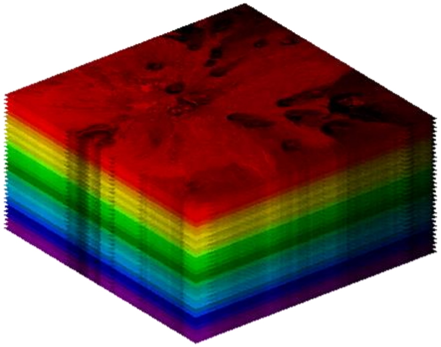
Best rank r approximation (LS) = SV hard thresholding or truncated SVD

∃ efficient algorithms e.g. Matlab `svds(X,r)` or approx. with randomization :
D. Achlioptas, F. Mcsherry, *Fast Computation of Low Rank Matrix Approximations*, 2007

Reshape

A discrete signal and in particular an hyperspectral image can also be treated as a 3-ways tensor

$$\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$



Tensors and Low Rank Prior

Rank of a tensor : minimum r such that there exists a decomposition

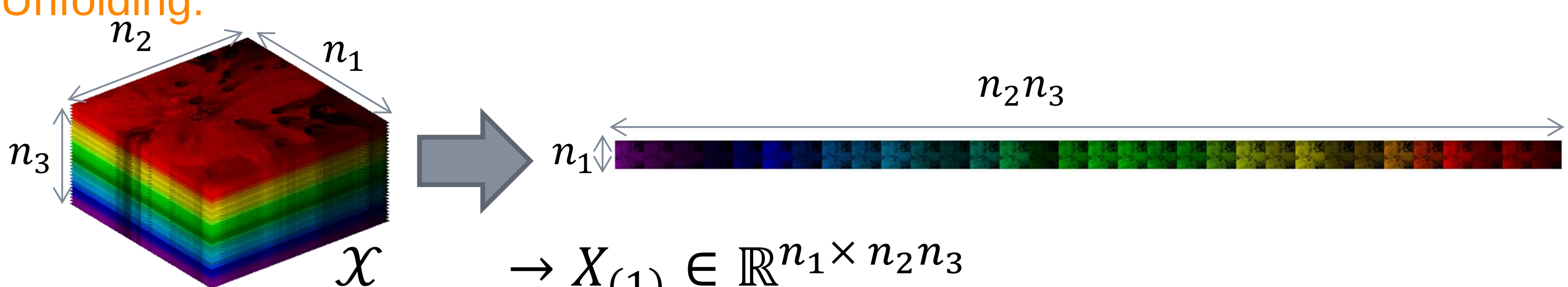
$$\mathcal{X} = \sum_{i=1}^{r} \underbrace{a_i \circ b_i \circ c_i}_{\text{rank 1 tensors}} \qquad \begin{aligned} a_i &\in \mathbb{R}^{n_1} \\ b_i &\in \mathbb{R}^{n_2} \\ c_i &\in \mathbb{R}^{n_3} \end{aligned}$$

Exterior tensor product

The rank is NP-complete (J. Håstad 1990).

Instead, we can use the “n-rank” i.e. rank of the n-unfolding matrix $X_{(n)}$

Unfolding:



$$\to X_{(1)} \in \mathbb{R}^{n_1 \times n_2 n_3}$$

$$\to X_{(2)} \in \mathbb{R}^{n_2 \times n_3 n_1}$$

$$\to X_{(3)} \in \mathbb{R}^{n_3 \times n_1 n_2}$$

Tensors and Low Rank Prior

NP-complete

We would like to solve

$$\arg\min_{\mathcal{X}} \ \mathrm{rank}(\mathcal{X}) \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

n-rank relaxation

$$\arg\min_{\mathcal{X}} \ \mathrm{rank}\big(X_{(1)}\big) + \mathrm{rank}\big(X_{(2)}\big) + \mathrm{rank}\big(X_{(3)}\big) \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

This problem is not convex and (like the ℓ_0 minimization) combinatorial.

Nuclear norm convex relaxation (analogy with ℓ_1)

$$\mathrm{rank}(X) = \#\{\sigma_i | \sigma_i > 0\} \ \rightarrow \|X\|_* = \sum_i \sigma_i$$

$$\arg\min_{\mathcal{X}} \ \big\|X_{(1)}\big\|_* + \big\|X_{(2)}\big\|_* + \big\|X_{(3)}\big\|_* \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

Tensors and Low Rank Prior

NP-complete

We would like to solve

$$\arg\min_{\mathcal{X}} \ \mathrm{rank}(\mathcal{X}) \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

n-rank relaxation

$$\arg\min_{\mathcal{X}} \ \mathrm{rank}\big(X_{(1)}\big) + \mathrm{rank}\big(X_{(2)}\big) + \mathrm{rank}\big(X_{(3)}\big) \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

This problem is not convex and (like the ℓ_0 minimization) combinatorial.

Nuclear norm convex relaxation (analogy with ℓ_1)

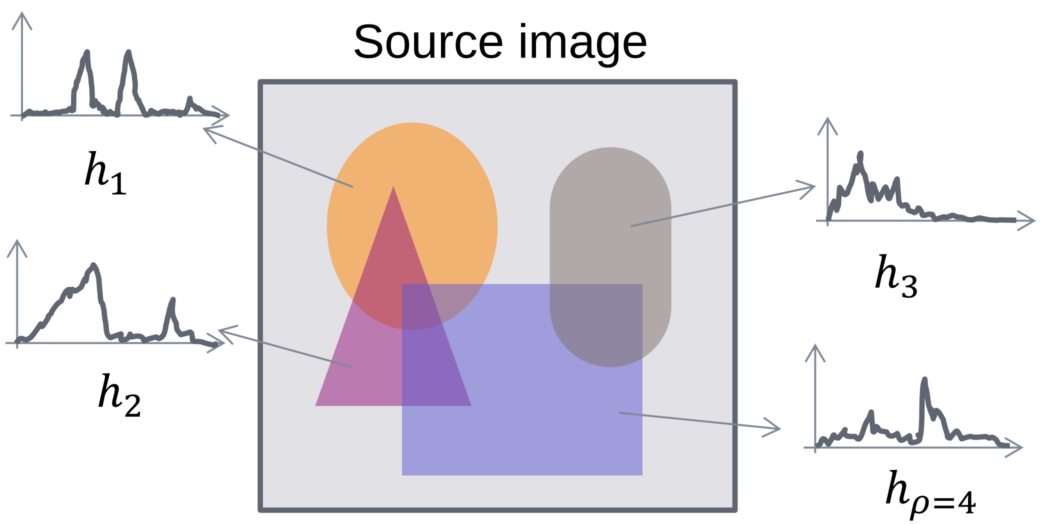
$$\mathrm{rank}(X) = \#\{\sigma_i | \sigma_i > 0\} \ \rightarrow \|X\|_* = \sum_i \sigma_i$$

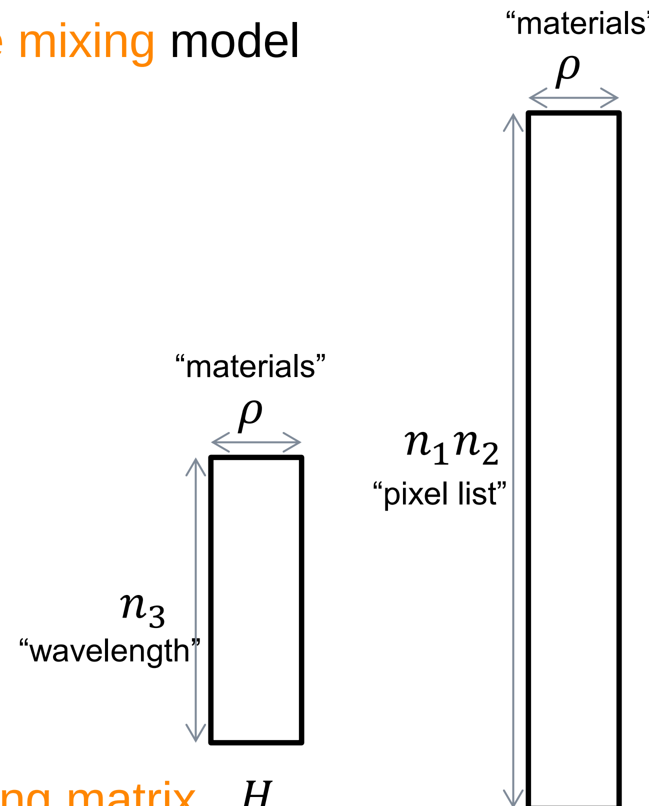
$$\arg\min_{\mathcal{X}} \ \big\|X_{(1)}\big\|_* + \big\|X_{(2)}\big\|_* + \big\|X_{(3)}\big\|_* \quad s.t. \ \ \|\Phi\mathcal{X} - y\| \leq \epsilon$$

OK … But why low rank?

Tensors and Low Rank Prior

Hyperspectral images are low rank : the source mixing model





ρ spectra, each with n_3 spectral samples : the mixing matrix H

$n_1 n_2$ pixels, each with a concentration of the ρ sources : the source matrix S

$$X_{(3)} = HS^T = \sum_{i=1}^{\rho} h_i s_i^T \Rightarrow \operatorname{rank}(\mathcal{X}) \leq \operatorname{rank}\left(X_{(3)}\right) \leq \rho$$

Tensors and Low Rank Prior

Rank in the spatial directions ($X_{(1)}$or $X_{(2)}$) may not be a good proxy for tensor rank and is not particularly relevant for natural images.

r = 16, PSNR = 24.3dB

SVT

K = 16384, PSNR = 33dB

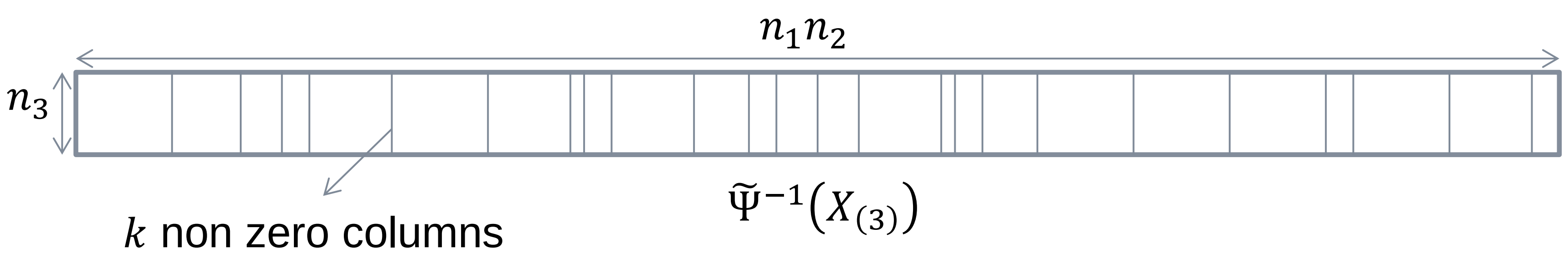
Haar wavelets

Same number of 'degrees of freedom' $K = r(n_1 + n_2)$

Low Rank and Joint Sparse

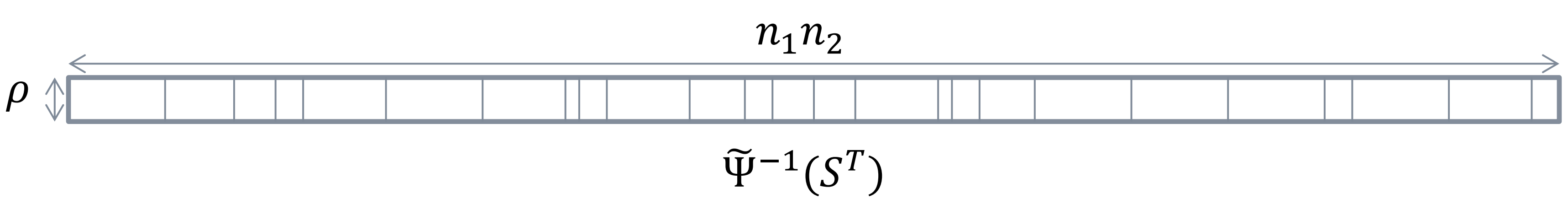
[M. Golbabee, P. Vandergheynst 2012]

Joint sparsity of the unfolded tensor $X_{(3)}$ (in sparsity basis Ψ)



Same support for wavelet transform of each spectral band
(= union of supports of each spectral band).

Low-rank and joint sparse with the source mixing model :
Source matrix is joint sparse (in the sparsity basis).



Low Rank and Joint Sparse

Some theoretical results from Golbabee & Vandergheynst :

Sparsity : $\|X_{vec}\|_0 = \#\{x_{i,j} | x_{i,j} \neq 0\}$ (number of non zero entries)

Convex relaxation $\|X_{vec}\|_1 = \sum_{i,j} |x_{i,j}|$

Convex minimization for sparse signal (BPDN)

$$\arg\min_X \|X_{vec}\|_1 \quad s.t. \quad \|\Phi(X) - y\| \leq \epsilon$$

$$M \leq O\left(K \log\left(\frac{n_1 n_2 n_3}{K}\right)\right) = O\left(k n_3 \log\left(\frac{n_1 n_2}{k}\right)\right)$$

(for nice (sub)Gaussian Φ)

NB : we note $X = X_{(3)} = (x_{i,j})$
and assume $\Psi = \mathrm{Id}$

Low Rank and Joint Sparse

Some theoretical results from Golbabee & Vandergheynst :

Joint sparsity : $\|X\|_{p,0} = \#\left\{x_{.,i} \middle| \left\|x_{.,i}\right\|_p > 0\right\}$ (number of non zero columns)

Convex relaxation $\|X\|_{2,1} = \sum_i \|x_i\|_2$

Convex minimization for joint sparse signal

$$\arg\min_X \|X\|_{2,1} \quad s.t. \quad \|\Phi(X) - y\| \leq \epsilon$$

$$M \geq O\left(k \log\left(\frac{n_1 n_2}{k}\right) + k n_3\right) \approx O(k n_3)$$

(for nice (sub)Gaussian Φ)

Low Rank and Joint Sparse

Some theoretical results from Golbabee & Vandergheynst :

Low rank: $\mathrm{rank}(X) = \#\{\sigma_i | \sigma_i > 0\}$ (number of non zero singular values)

Convex relaxation $\|X\|_* = \sum_i \sigma_i$

Convex minimization for low rank signal

$$\arg\min_X \|X\|_* \quad s.t. \quad \|\Phi(X) - y\| \leq \epsilon$$

$$M \geq O\big(r(n_1 n_2 + n_3)\big)$$

(for nice (sub)Gaussian Φ)

Low Rank and Joint Sparse

Some theoretical results from Golbabee & Vandergheynst :

Convex minimization for low rank and joint sparse signal (one example)

$$\hat{X} = \arg\min_{X} \|X\|_{2,1} + \lambda\|X\|_* \quad s.t. \;\; \|\Phi(X) - y\| \leq \epsilon$$

$$M \geq O\left(k \log\left(\frac{n_1 n_2}{k}\right) + kr + n_3 r\right)$$

(for nice (sub)Gaussian Φ)

Error bound

$$\left\|X - \hat{X}\right\|_F \leq \kappa_0' \left(\frac{\left\|X - X_{r,k}\right\|_{2,1}}{\sqrt{k}} + \frac{\left\|X - X_{r,k}\right\|_*}{\sqrt{2r}}\right) + \kappa_1' \epsilon$$

if not exactly joint-sparse

if not exactly low rank

if noisy measurements

Low Rank and Joint Sparse

Demo example 512x512x32

Block diagonal sensing matrix $\Phi = \widetilde{\Phi} \otimes \mathrm{Id}_{\mathrm{n}_3}$

Where $\widetilde{\Phi}$ is the SSRFE (applied spatially) with $\frac{m}{n_1 n_2} = 25\%$

No spectral mixing $\approx$ no dispersive optical element.

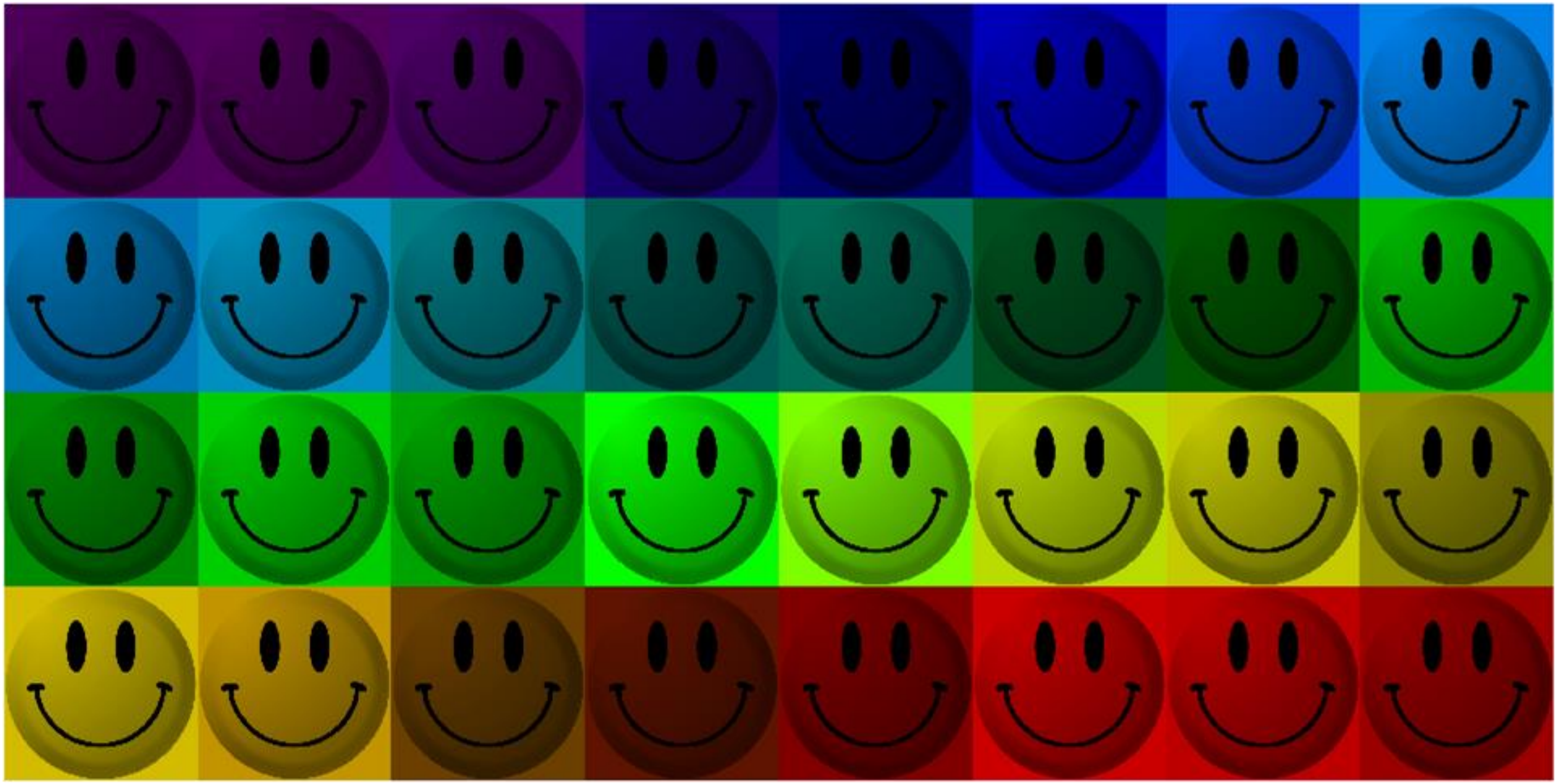
→ quite far from full Gaussian.

Generated (k, ϵ) –joint-compressible low rank HS image (from smile).

With $r = 6$ and $\frac{k}{n_1 n_2} = 4\%$ for $\epsilon = 10^{-4}$

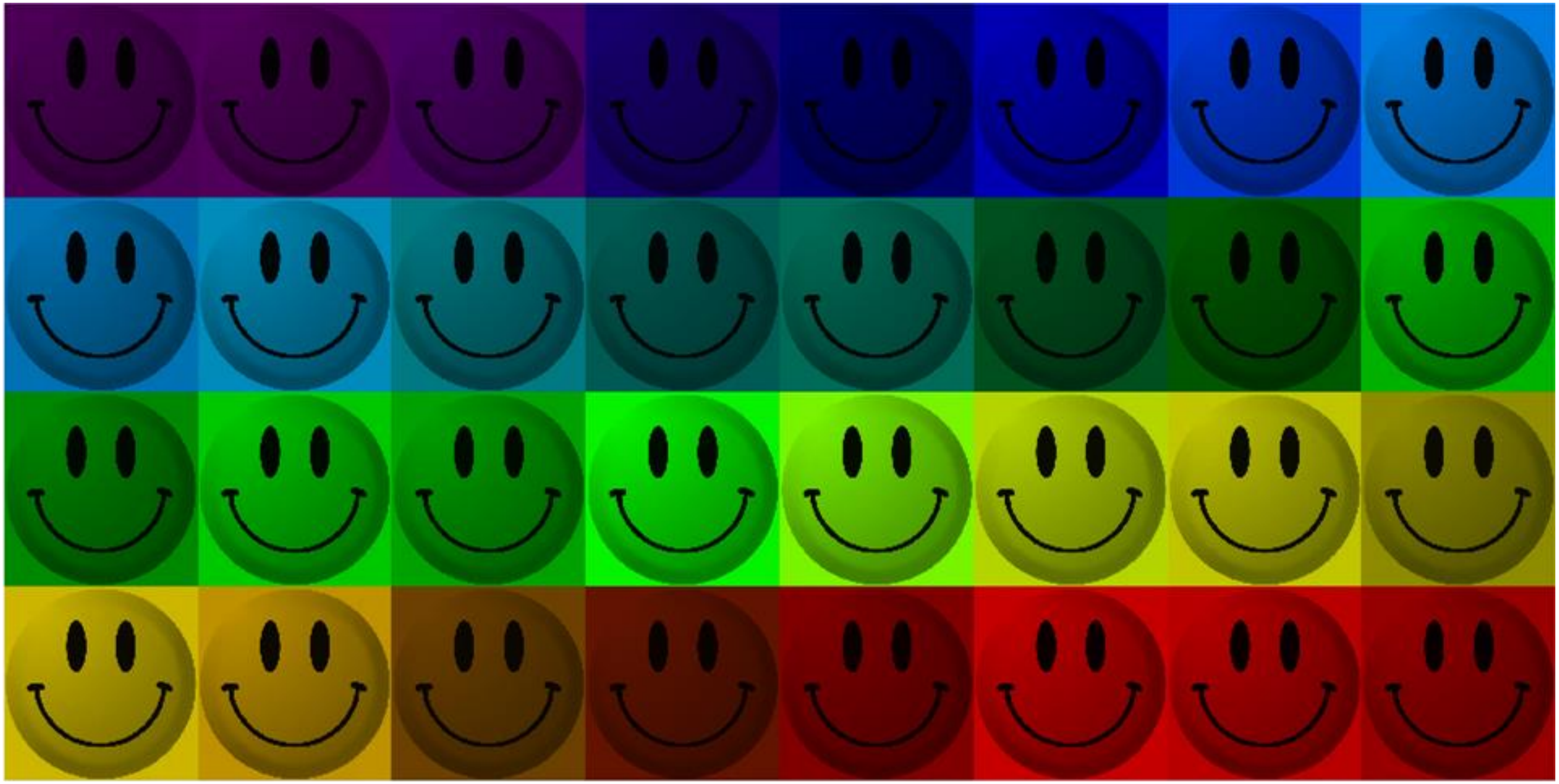
Low Rank and Joint Sparse

Demo example 512x512x32 (target)



Low Rank and Joint Sparse

Demo example 512x512x32 (reconstruction)



PSNR = 43,5dB

Source Separation

[M. Golbabee, S. Arberet, P. Vandergheinst 2012]

There exist extensive spectral databases with spectra associated to a lot of materials.

We can use a subset of these databases as a dictionary for sparsity prior.

$$X = SH^T$$

unknown concentration Image $\in \mathbb{R}^{n_1 n_2 \times \rho}$

known spectral database $\in \mathbb{R}^{n_3 \times \rho}$

Source Separation

[M. Golbabee, S. Arberet, P. Vandergheinst 2012]

Modified reconstruction model:

$$\underset{\in \mathbb{R}^M}{y} = \Phi X_{vec} = \Phi(SH^T)_{vec} = \underbrace{\Phi\left(\mathrm{H} \otimes \mathrm{Id}_{n_1 n_2}\right)}_{\Phi' \in \mathbb{R}^{M \times \rho n_1 n_2}} S_{vec}$$

$$\arg\min_X \|\mathrm{S}_{vec}\|_1 \quad s.t. \;\; \|\Phi' \mathrm{S}_{vec} - y\| \leq \epsilon$$

Source Separation

[M. Golbabee, S. Arberet, P. Vandergheinst 2012]

When $\Phi = \widetilde{\Phi} \otimes \mathrm{Id}_{n_3}$ (no spectral mixing) we can write

$\Phi \in \mathbb{R}^{M \times n_1 n_2}$, $\widetilde{\Phi} \in \mathbb{R}^{m \times n_1 n_2}$

$(M = m\, n_3)$

$$Y = \widetilde{\Phi} X$$

$Y \in \mathbb{R}^{m \times n_3}$

Decorrelation before reconstruction (dimensionality reduction)

$$Y^* = Y\left(H^\dagger\right)^T = \widetilde{\Phi} S$$

$Y^* \in \mathbb{R}^{m \times \rho}$

No dependency in H and smaller dimension.
Much easier to handle in reconstruction algorithms!

Thank You !

References

- N. Parikh and S. Boyd, *Proximal Algorithms*, 2013
- D. Brady, *Optical Imaging and Spectroscopy*, 2009
- D. Achlioptas, F. Mcsherry, *Fast Computation of Low Rank Matrix Approximations*, 2007
- J. Håstad, *Tensor rank is NP-complete*, 1990
- M. Golbabee, P. Vandergheynst, *Compressed Sensing of Simultaneous Low-Rank and Joint-Sparse Matrices*, 2012
- M. Golbabee, S. Arberet, P. Vandergheynst, *Compressive Source Separation: Theory and Methods for Hyperspectral Imaging*, 2012

For additional references, feel free to ask me at kevin.degraux@uclouvain.be.