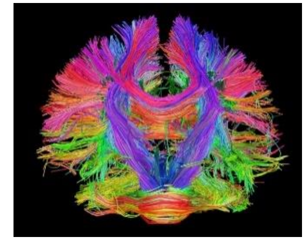


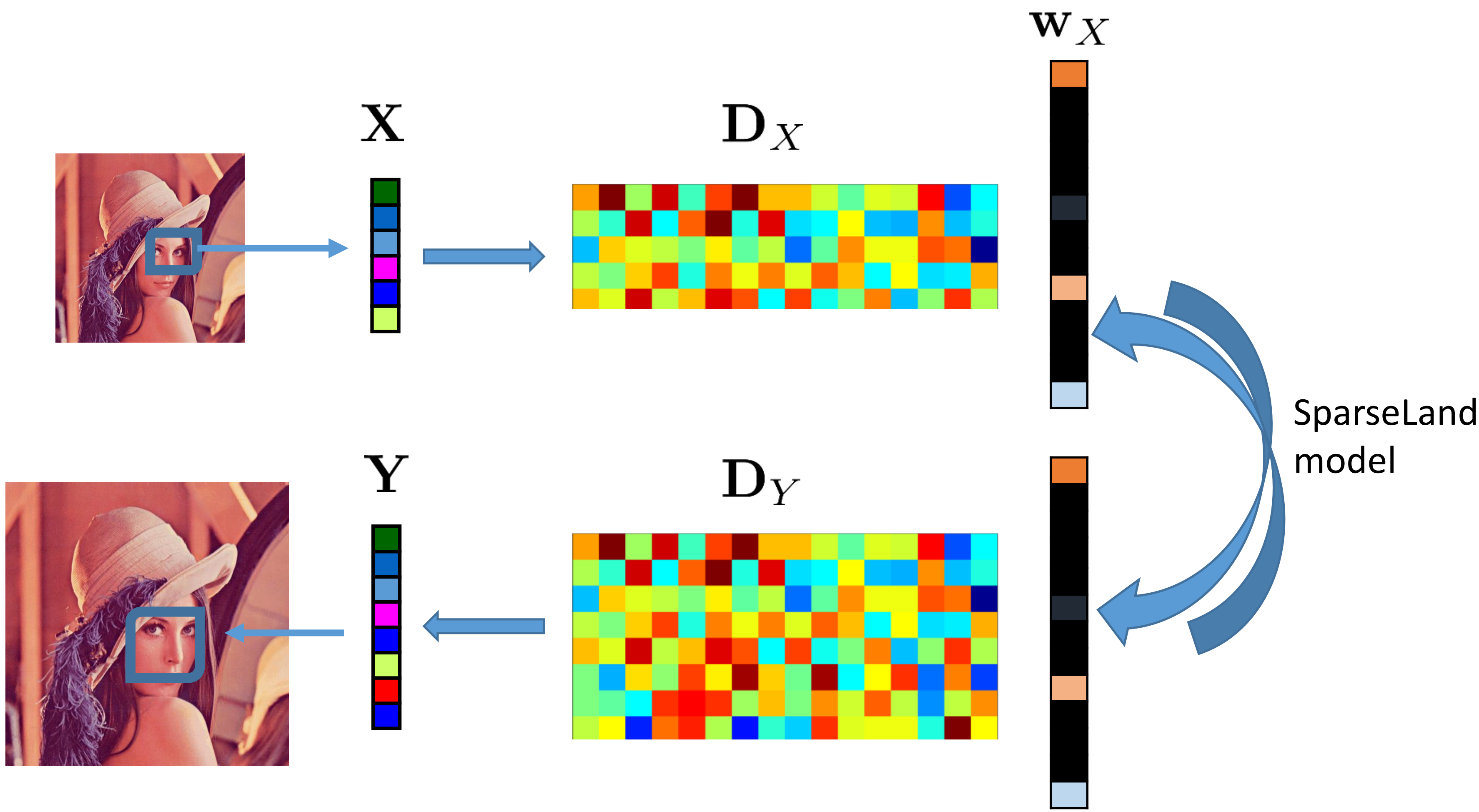

CS-570
Statistical Signal Processing

Lecture 11: Extension of sparse and low rank models

Spring Semester 2019

Grigorios Tsagkatakis

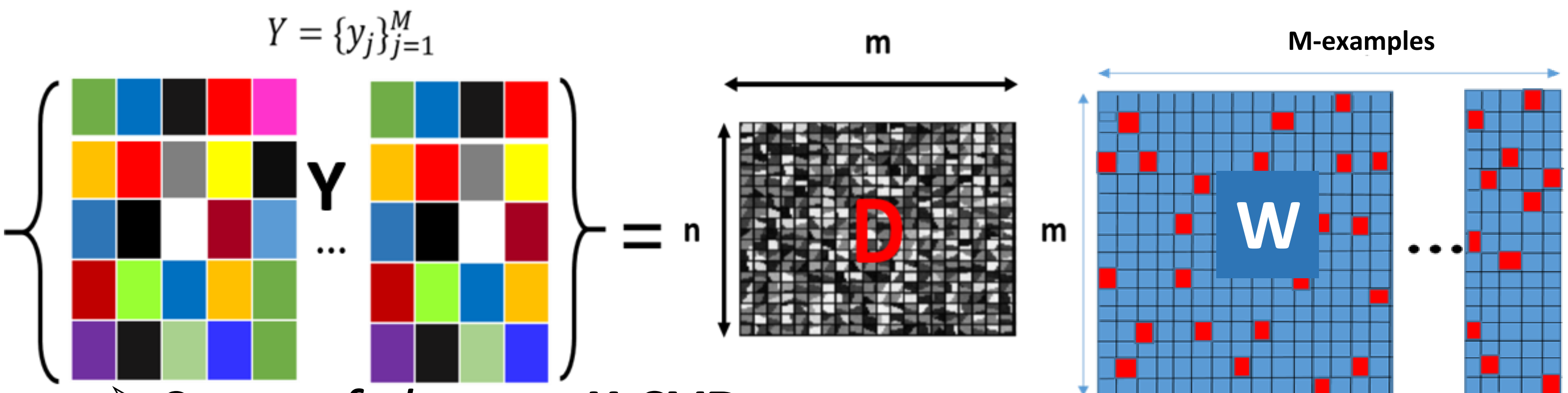
Sparsity Modeling



Dictionary Training

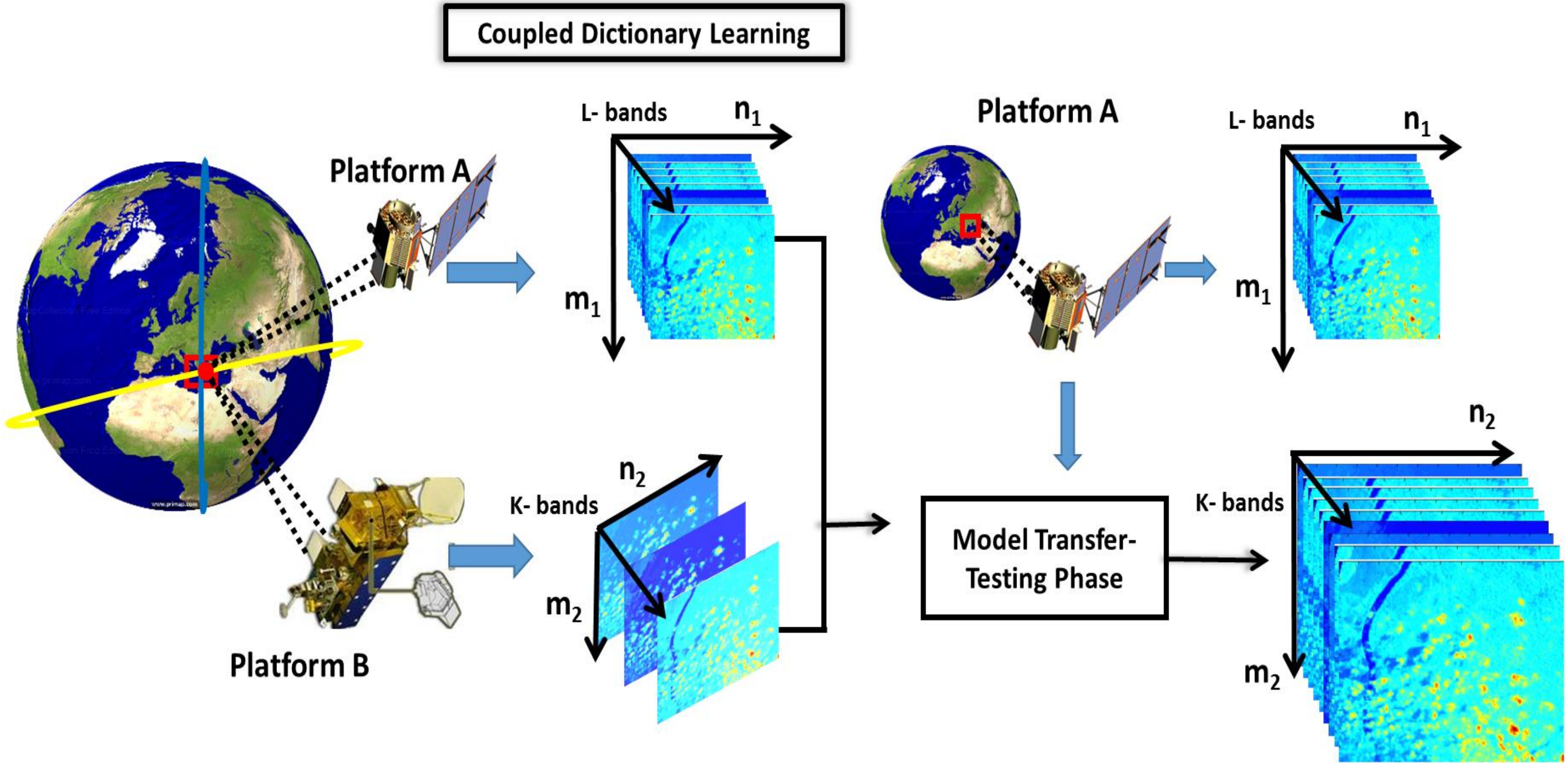
- **Optimization**

$$\min_{\mathbf{D},\mathbf{W}} \sum_{j=1}^{M} ||\mathbf{D}\mathbf{w}_j - \mathbf{y}_j||_F^2, \text{ s. t. } \forall ||\mathbf{w}_j||_1 \leq L \text{ and } ||\mathbf{D}(:,j)||_2 \leq 1$$

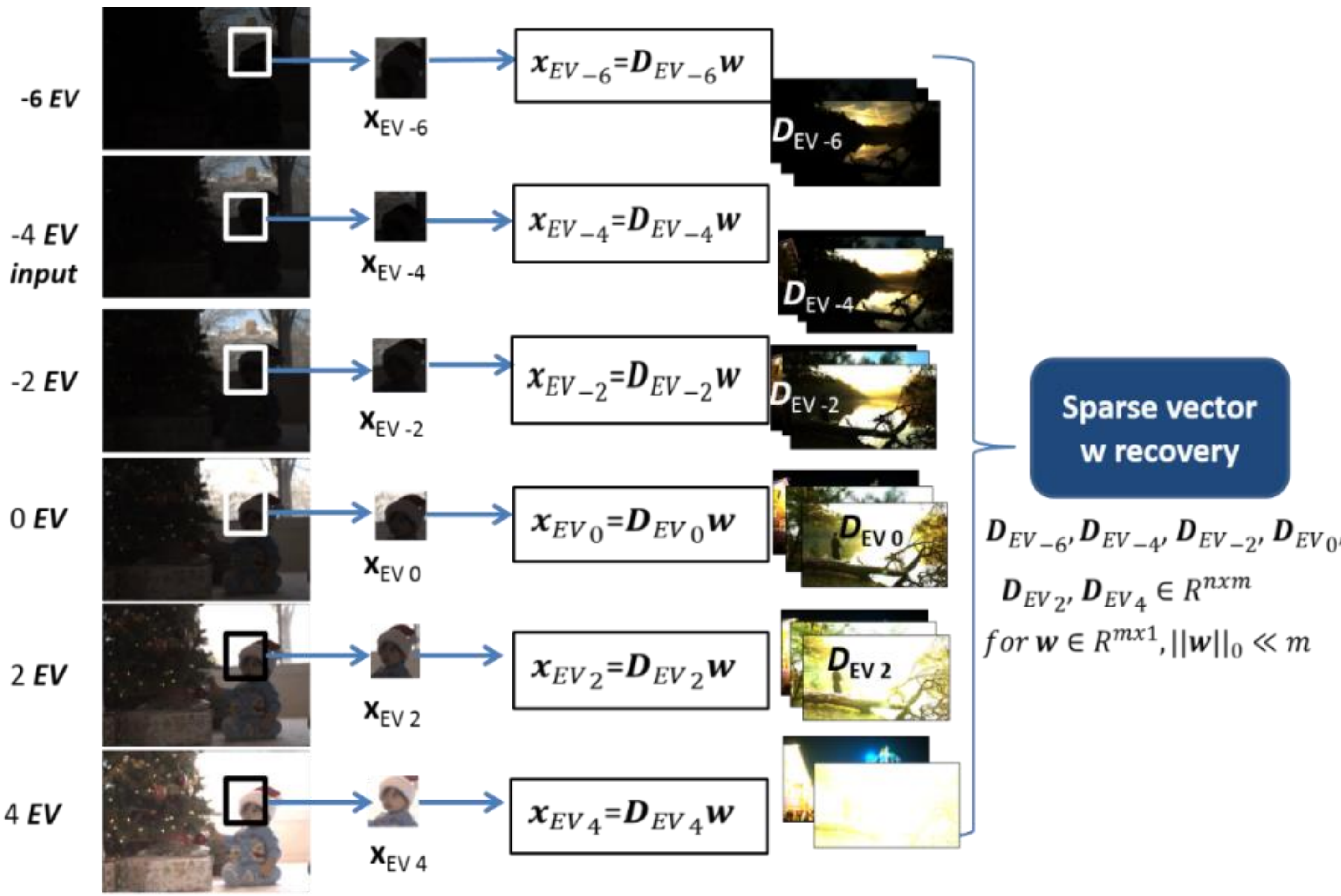


- *State-of-the-art*: **K-SVD**

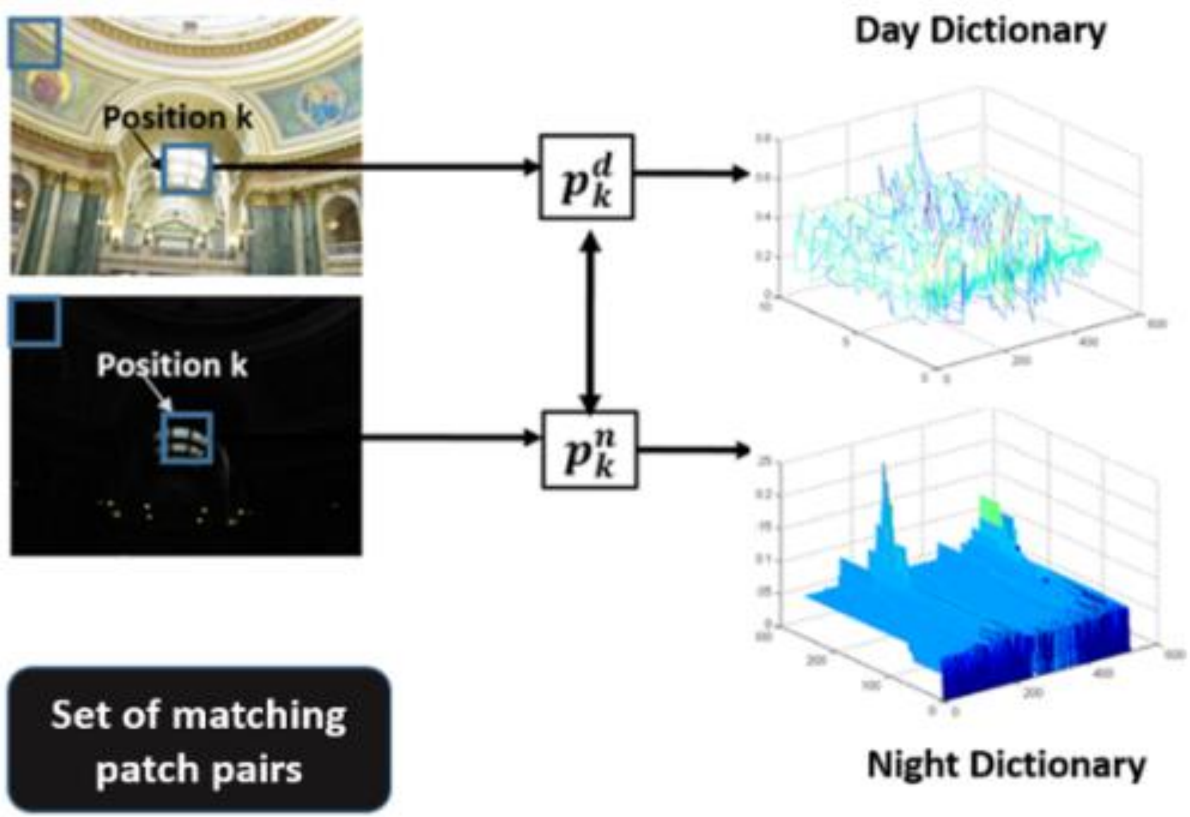
Types of problems



Types of problems



Types of problems



Coupled Dictionary Learning

- **Goal**: ***Learn jointly two dictionary matrices***: $\mathbf{D}_h$, $\mathbf{D}_\ell$

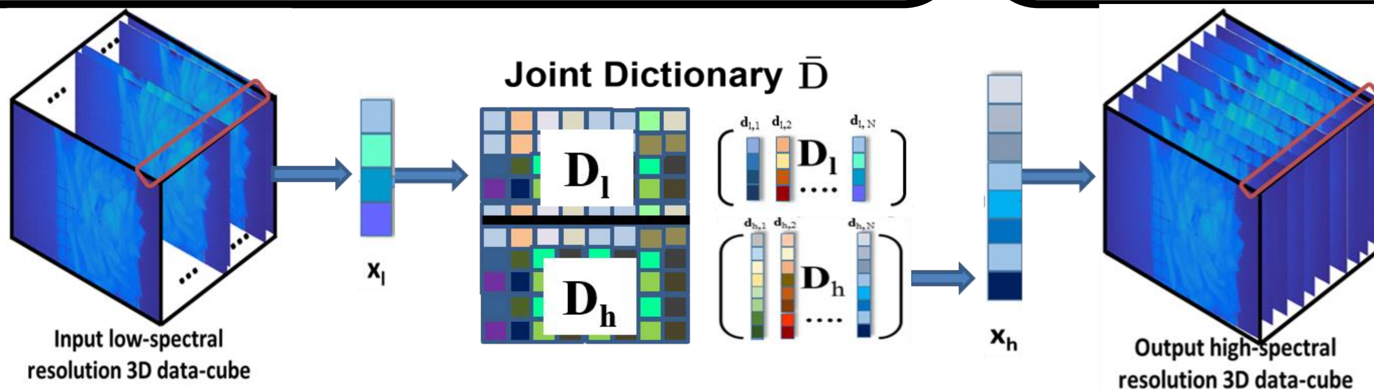
- **Concatenated feature space:**

$$\underset{\mathbf{D},\mathbf{W}}{\operatorname{argmin}} ||\bar{\mathbf{S}} - \bar{\mathbf{D}}\mathbf{W}||_F + \lambda ||\mathbf{W}||_1,$$

$$\text{s. t. } ||\bar{\mathbf{D}}(:,j)||_2^2 \leq 1, \quad j = \{1, ..., K\},$$

where

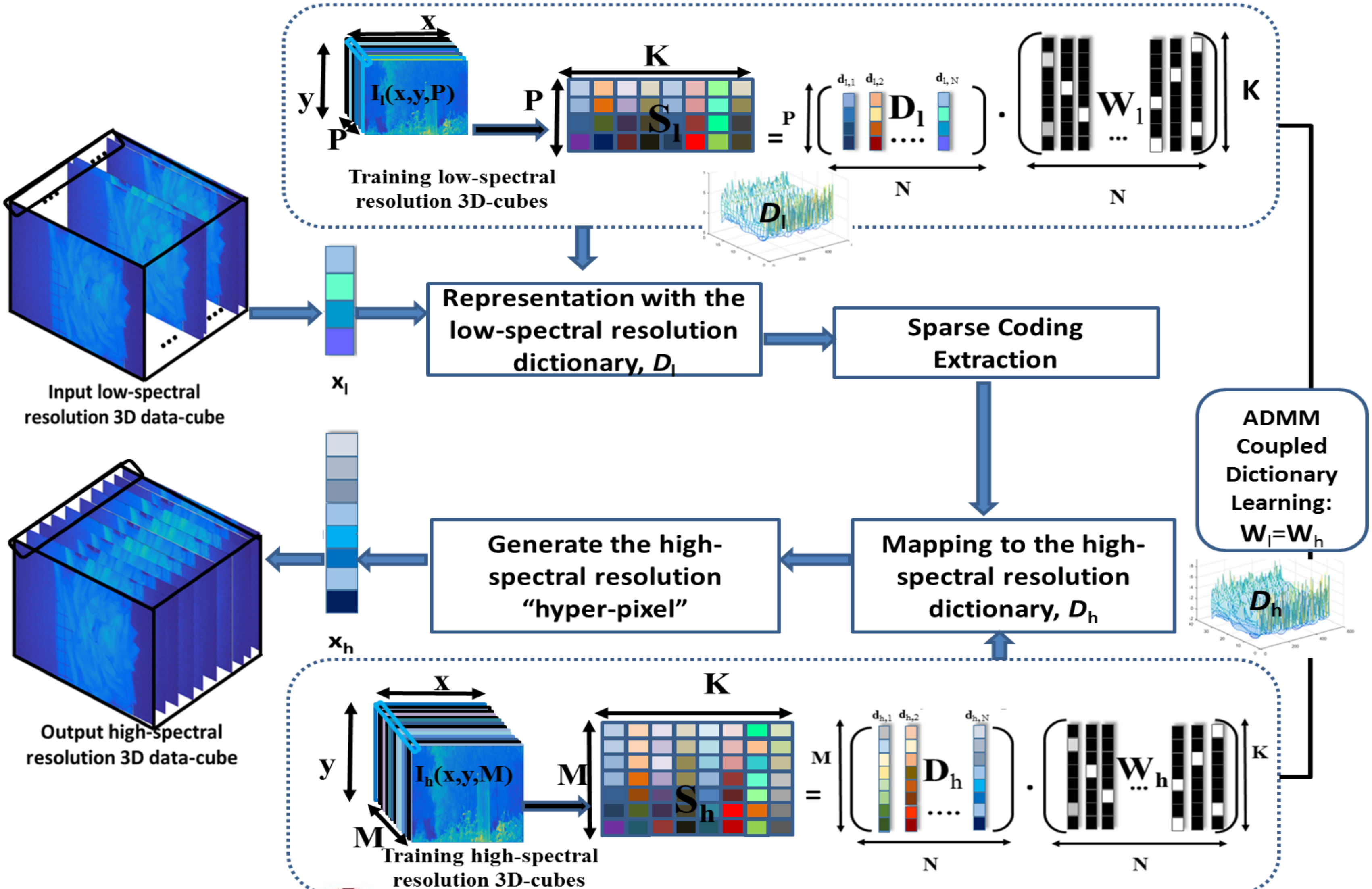
$$\bar{\mathbf{S}} = \begin{bmatrix} \mathbf{S}_h \\ \mathbf{S}_l \end{bmatrix}, \; \bar{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_h \\ \mathbf{D}_l \end{bmatrix}$$



Limitation !!

- **Sub-Optimal Coding Scheme → Individual Feature Spaces!**

Coupled Dictionary Learning (CDL)



ADMM for Coupled Dictionary Learning

- **Optimization problem:**

$$(\mathbf{D}_h, \mathbf{W}_h) = \text{argmin}||\mathbf{D}_h\mathbf{W}_h - \mathbf{S}_h||_F + \lambda_h||\mathbf{W}_h||_1$$
$$(\mathbf{D}_l, \mathbf{W}_l) = \text{argmin}||\mathbf{D}_l\mathbf{W}_l - \mathbf{S}_l||_F + \lambda_l||\mathbf{W}_l||_1,$$
$$||\mathbf{D}_h(:,j)||_2^2 \leq 1,\ ||\mathbf{D}_l(:,j)||_2^2 \leq 1,\ \text{and } \mathbf{W}_h = \mathbf{W}_l$$

- **Setting**: $\mathbf{P} = \mathbf{D}_h$ **and** $\mathbf{Q} = \mathbf{D}_\ell$

$$\min_{\mathbf{D}_h,\mathbf{W}_h,\mathbf{D}_l,\mathbf{W}_l} ||\mathbf{S}_h - \mathbf{D}_h\mathbf{W}_h||_F^2 + ||\mathbf{S}_l - \mathbf{D}_l\mathbf{W}_l||_F^2 + \lambda_l||\mathbf{Q}||_1 + \lambda_h||\mathbf{P}||_1$$
$$\text{s. t. } \mathbf{P} = \mathbf{W}_h, \mathbf{Q} = \mathbf{W}_l, \mathbf{W}_h = \mathbf{W}_l, ||\mathbf{D}_h(:,i)||_2 \leq 1, ||\mathbf{D}_l(:,i)||_2 \leq 1$$

- **Augmented Lagrangian Function:**

$$L(\mathbf{D}_h, \mathbf{D}_l, \mathbf{W}_h, \mathbf{W}_l, \mathbf{P}, \mathbf{Q}, Y_1, Y_2, Y_3) = \frac{1}{2}||\mathbf{D}_h\mathbf{W}_h - \mathbf{S}_h||_F^2 + \frac{1}{2}||\mathbf{D}_\ell\mathbf{W}_\ell - \mathbf{S}_\ell||_F^2 +$$
$$\lambda_h||\mathbf{P}||_1 + \lambda_\ell||\mathbf{Q}||_1 + < Y_1, \mathbf{P} - \mathbf{W}_h > + < Y_2, \mathbf{Q} - \mathbf{W}_\ell > + < Y_3, \mathbf{W}_h - \mathbf{W}_\ell > +$$
$$\frac{c_1}{2}||\mathbf{P} - \mathbf{W}_h||_F^2 + \frac{c_2}{2}||\mathbf{Q} - \mathbf{W}_\ell||_F^2 + \frac{c_3}{2}||\mathbf{W}_h - \mathbf{W}_\ell||_F^2$$

ADMM for Coupled Dictionary Learning

Decomposition: Sparse Coding Sub-problems

Sub-problems W_X & W_Y

$$\min_{\mathbf{W}_x} \sum_{i=1}^{N} < \Lambda_i, (\mathbf{P} - \mathbf{D}_x \mathbf{W}_x)_i > + \frac{\rho}{2} ||\mathbf{P} - \mathbf{D}_x \mathbf{W}_x||_F^2$$

$$\min_{\mathbf{W}_y} \sum_{i=1}^{N} < \Lambda_i, (\mathbf{Q} - \mathbf{D}_y \mathbf{W}_y)_i > + \frac{\rho}{2} ||\mathbf{Q} - \mathbf{D}_y \mathbf{W}_y||_F^2$$

Sub-problems P & QL

$$\min_{\mathbf{Q}} ||\mathbf{Y} - \mathbf{Q}||_F^2 + \sum_{i=1} < \Lambda_i, (\mathbf{Q} - \mathbf{D}_y \mathbf{W})_i > + \frac{\rho}{2} ||\mathbf{Q} - \mathbf{D}_y \mathbf{W}||_F^2$$

$$\min_{\mathbf{P}} ||\mathbf{X} - \mathbf{P}||_F^2 + \sum_{i=1}^{N} < \Lambda_i, (\mathbf{P} - \mathbf{D}_x \mathbf{W})_i > + \frac{\rho}{2} ||\mathbf{P} - \mathbf{D}_x \mathbf{W}||_F^2$$

FORTH
Institute of Computer Science

ADMM for Coupled Dictionary Learning

Sub-problems D_X & D_Y

$$\min_{\mathbf{D}_x} \frac{\rho}{2}||\mathbf{P} + \Lambda/\rho - \mathbf{D}_x\mathbf{W}||_F^2 \quad \text{and} \quad \min_{\mathbf{D}_y} \frac{\rho}{2}||\mathbf{Q} + \Lambda/\rho - \mathbf{D}_y\mathbf{W}||_F^2$$

In-exact solution!

$$\mathbf{E}_x = \mathbf{P} + \Lambda/\rho - \mathbf{D}_x^{(k)}\mathbf{W}$$
$$\mathbf{E}_y = \mathbf{Q} + \Lambda/\rho - \mathbf{D}_y^{(k)}\mathbf{W}$$

and

$$\phi_x = \mathbf{W}_x(j,:)\mathbf{W}_x(j,:)^T$$
$$\phi_y = \mathbf{W}_y(j,:)\mathbf{W}_y(j,:)^T$$

Dictionary update step

$$\mathbf{D}_x^{(k+1)}(:,j) = \mathbf{D}_x^{(k)}(:,j) + \mathbf{E}_x\mathbf{W}(j,:)^T/(\phi_x + \delta)$$
$$\mathbf{D}_y^{(k+1)}(:,j) = \mathbf{D}_y^{(k)}(:,j) + \mathbf{E}_y\mathbf{W}(j,:)^T/(\phi_y + \delta)$$

ADMM for CDL- Algorithm

1. **Input**: Training examples $\mathbf{S}_h$ and $\mathbf{S}_l$, numb. of iterations: **K** and step size params. **c1, c2, c3**.
2. **Initialization**:
   - **Dictionaries** → random selection of the columns of $\mathbf{S}_h$ and $\mathbf{S}_l$
   - **Lagrangian matrices** → $\mathbf{Y}_1 = \mathbf{Y}_2 = \mathbf{Y}_3 = 0$.
3. **for k = 1,...,K do**

   - Update $\mathbf{W}_h$ and $\mathbf{W}_l$

$$\begin{cases} \mathrm{W}_h = (\mathrm{D}_h^T \cdot \mathrm{D}_h + c_1 \cdot I + c_3 \cdot I)^{-1} \cdot (\mathrm{D}_h^T \cdot \mathrm{S}_h + Y_1 - Y_3 + c_1 \cdot \mathrm{P} + c_3 \cdot \mathrm{W}_l) \\ \mathrm{W}_l = (\mathrm{D}_l^T \cdot \mathrm{D}_l + c_2 \cdot I + c_3 \cdot I)^{-1} \cdot (\mathrm{D}_l^T \cdot \mathrm{S}_l + Y_2 + Y_3 + c_2 \cdot \mathrm{Q} + c_3 \cdot \mathrm{W}_h) \end{cases}$$

   - Update **P** and **Q** → $\mathbf{P} = S_{\lambda_h}\left(\left|\mathbf{W}_h - \frac{Y_1}{c_1}\right|\right)$ **and**: $\mathbf{Q} = S_{\lambda_l}\left(\left|\mathbf{W}_l - \frac{Y_2}{c_2}\right|\right)$

   for j = 1,...,N do

   - Update $\boldsymbol{\varphi}_h$ and $\boldsymbol{\varphi}_l$ → $\phi_h = \mathbf{W}_h(j,:) \cdot \mathbf{W}_h(j,:)^T$ **and** $\phi_l = \mathbf{W}_l(j,:) \cdot \mathbf{W}_l(j,:)^T$
   - Update $\mathbf{D}_h$ and $\mathbf{D}_l$

$$\mathbf{D}_h^{(k+1)}(:,j) = \mathbf{D}_h(:,j)^{(k)}(:,j) + \frac{\mathbf{S}_h \cdot \mathbf{W}_h(j,:)}{\phi_h + \delta}$$

$$\mathbf{D}_l^{(k+1)}(:,j) = \mathbf{D}_l(:,j)^{(k)}(:,j) + \frac{\mathbf{S}_l \cdot \mathbf{W}_l(j,:)}{\phi_l + \delta}$$

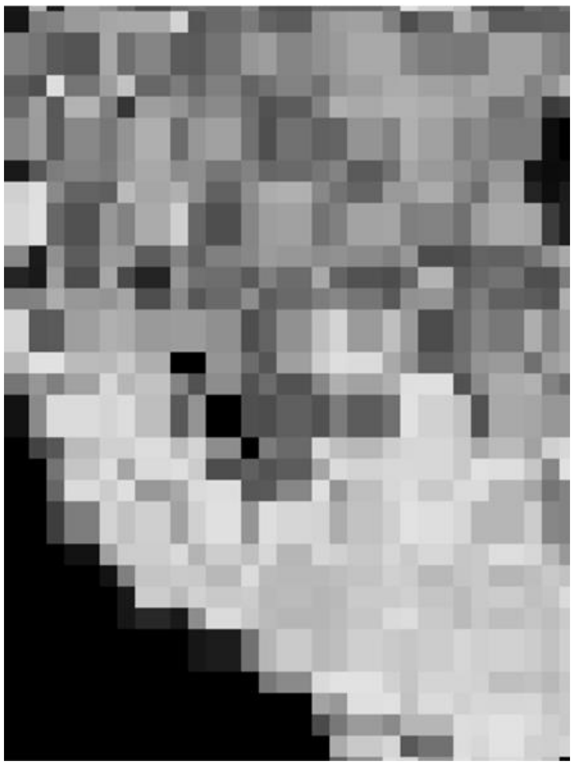
   end

   - **Normalize** $\mathbf{D}_h$ and $\mathbf{D}_l$ between [0,1]
   - **Update** Lagrange multiplier matrices Y_1, Y_2 and Y_3

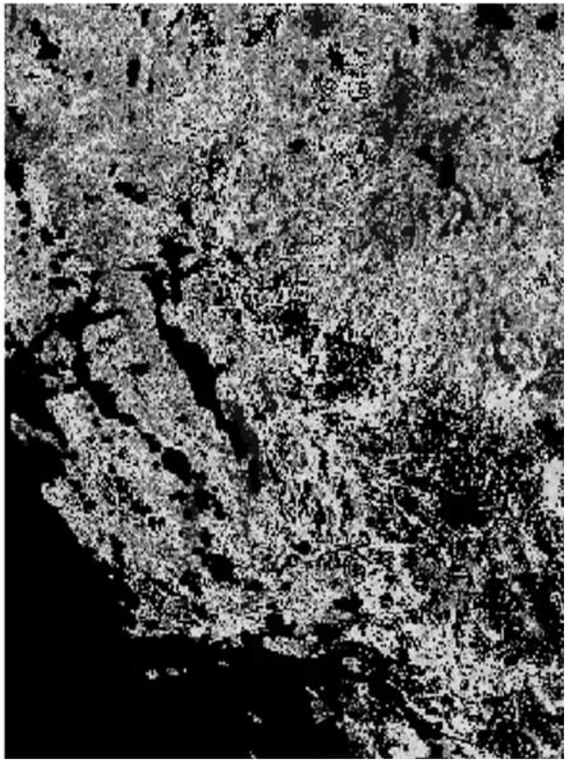
end

Experimental Results: California Region

➢ Reconstruction Performance



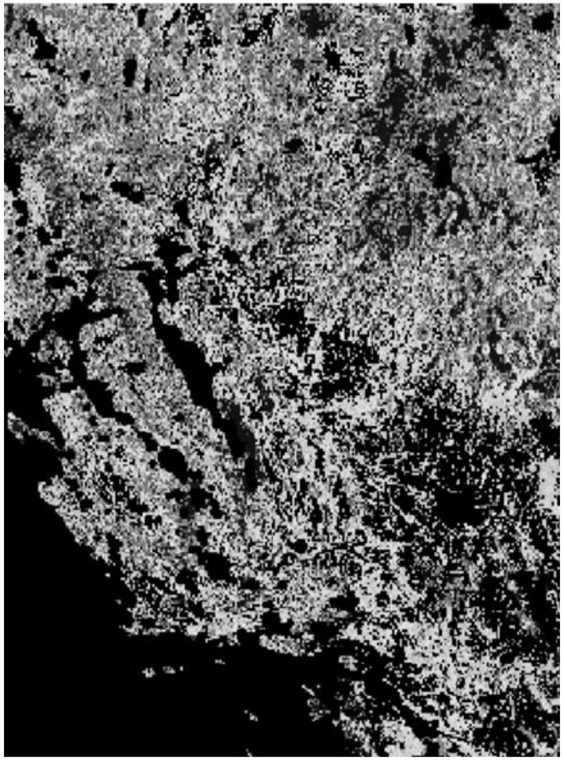
Input (Gray-scale)



Output (Active)

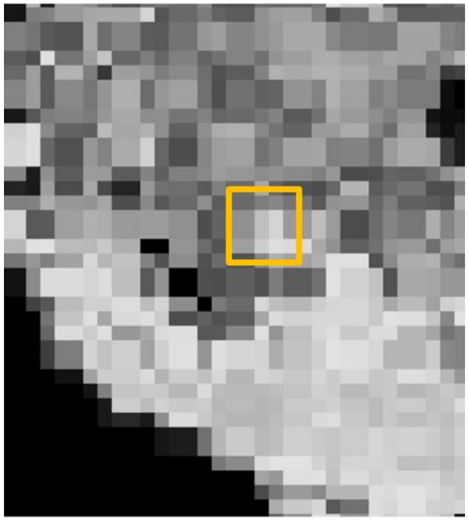
RMSE: 2.97

Absolute Difference: 1.265

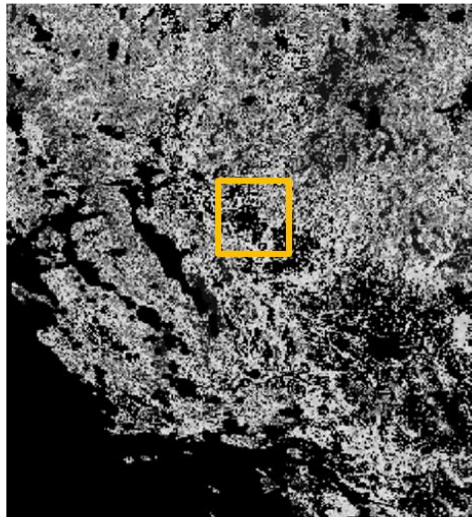


Ground Truth (Active)

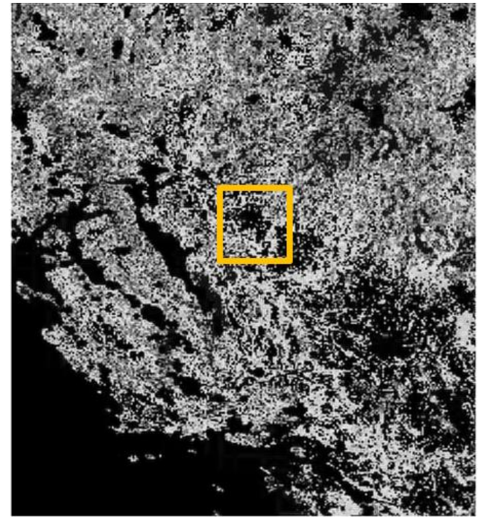
California Region: Comparison with SoA



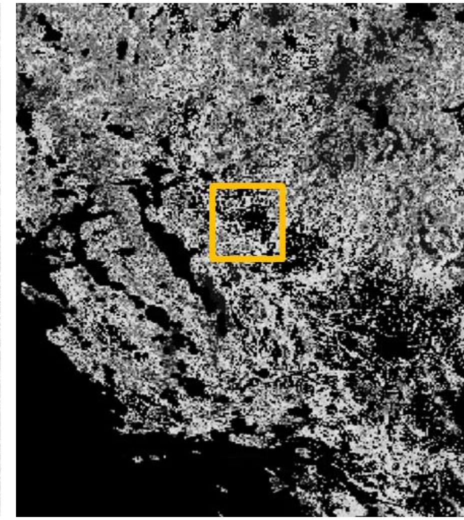
Input Passive
(California Region)



Ground Truth,
(California Region)



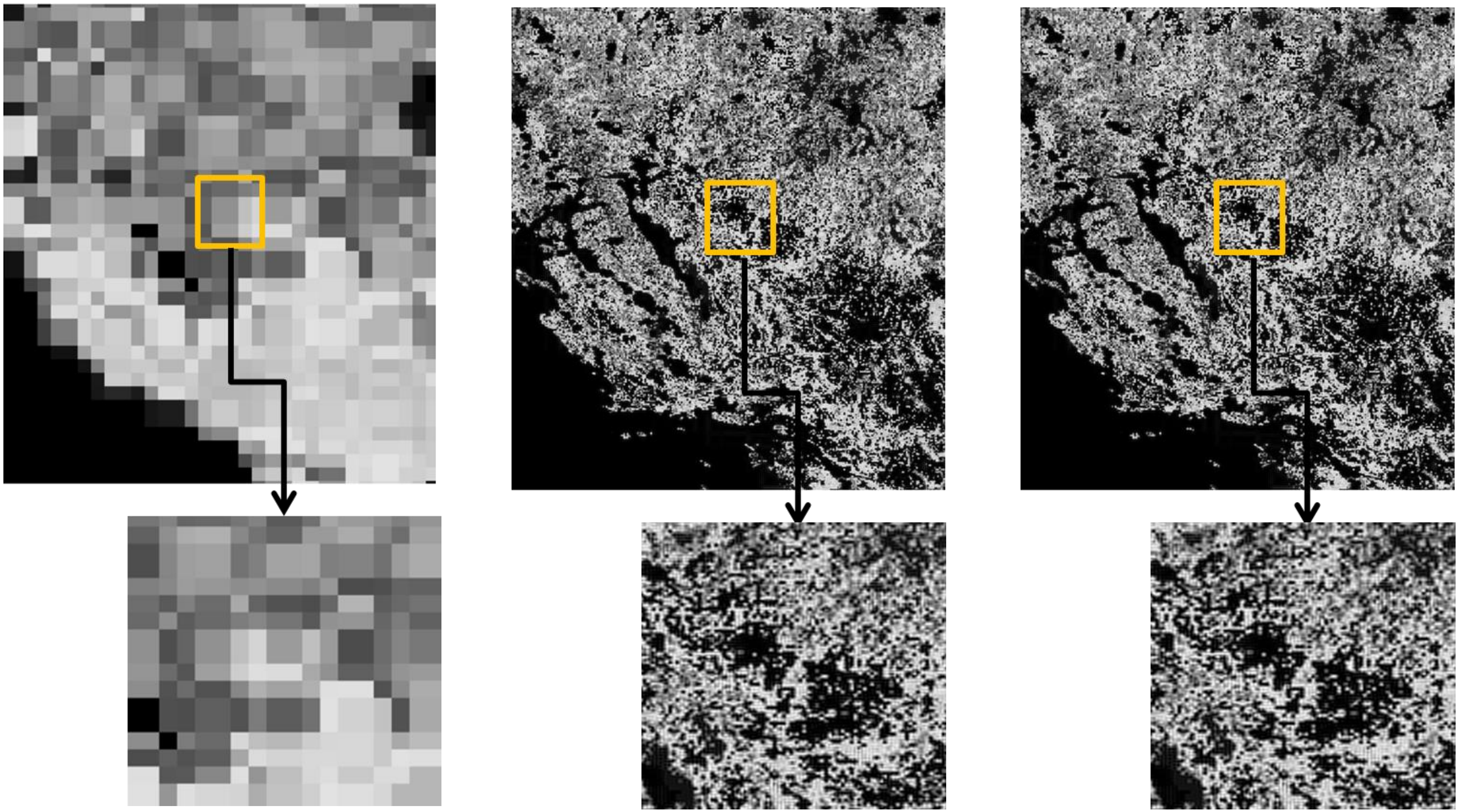
K-SVD Active Rec.
RMSE: 5.22,
SSIM: 0.95
Absolute error: 1.45



ADMM Active Rec.
RMSE: 2.97,
SSIM: 0.98
Absolute error: 1.265

➢ ***The proposed ADMM CDL algorithm outperforms the K-SVD State-of-the-art technique!***

California Region: Comparison with SoA

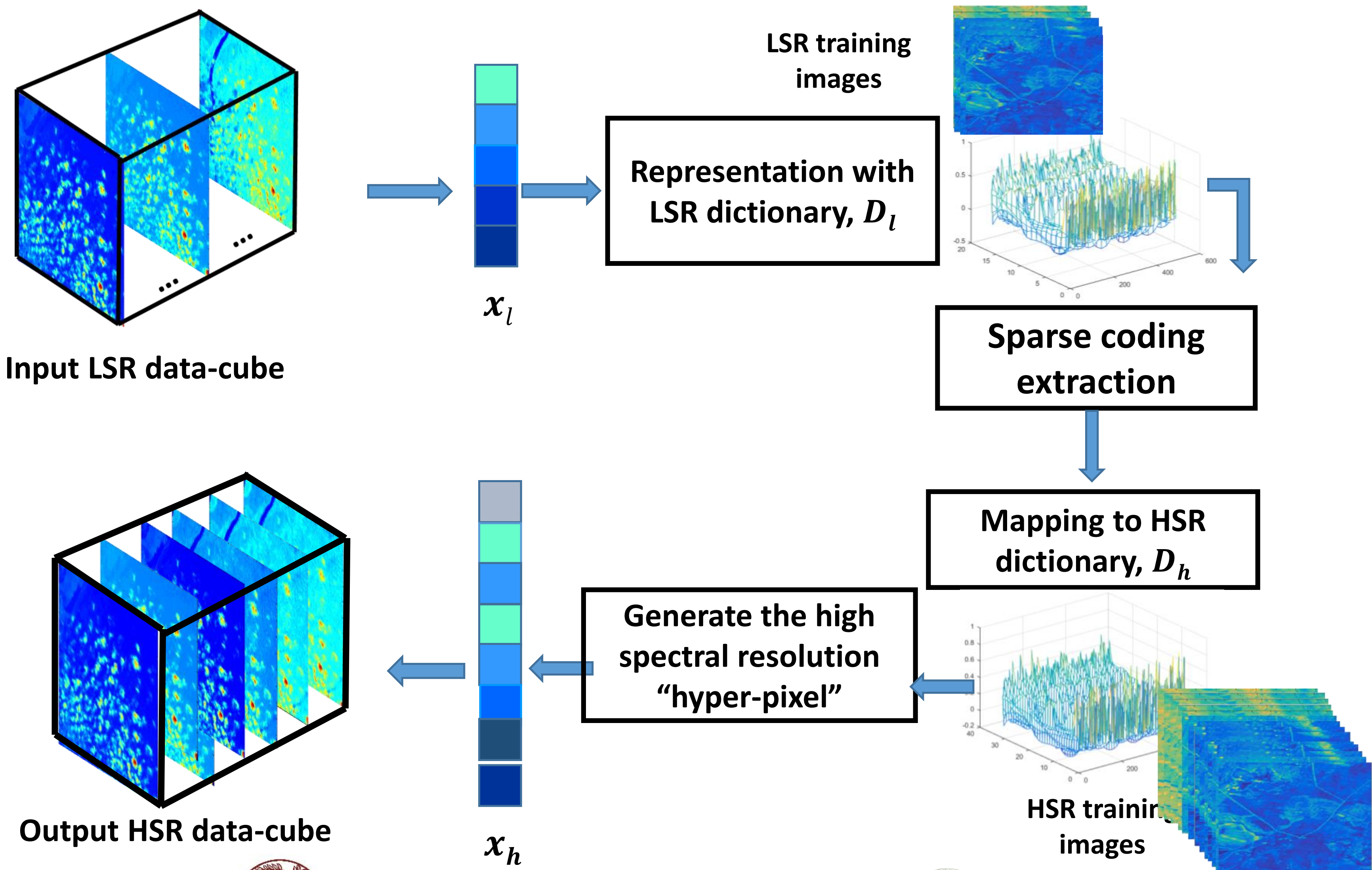


Input Passive

K-SVD Active Rec.
RMSE: 5.22

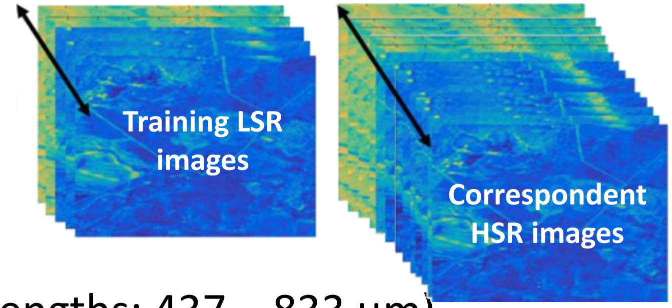
ADMM Active Rec.
RMSE: 2.97

Application in Spectral Super-resolution (SSR)



SSR Experimental Setup

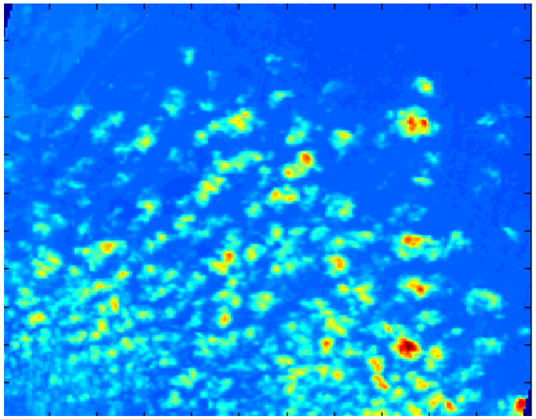
- **Dictionaries Generation**
  - 100K coupled pairs of high & low-spectral resolution hyperspectral images



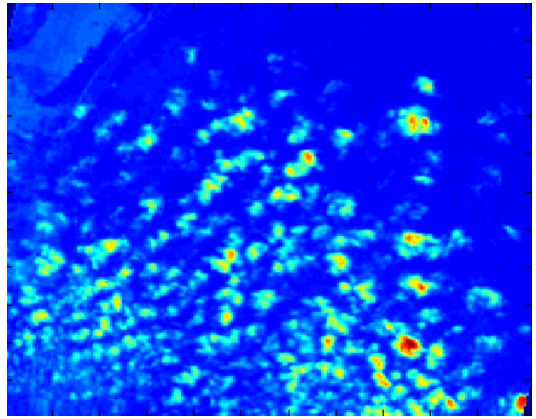
- **Hyperspectral Acquisition**
  - **Hyperion hyperspectral data**
    - ✓ **Full Spectrum:** 39 spectral bands from the VNIR region → (bands: 9-48, wavelengths: 437 – 833 μm)
    - ✓ **Down-sampling factors :** (x2), (x3)
- **Evaluation Metrics:**
  - ***Peak signal to noise ratio***

$$PSNR = 10log_{10}[L_{max}^2/MSE(x,y,\lambda)]$$

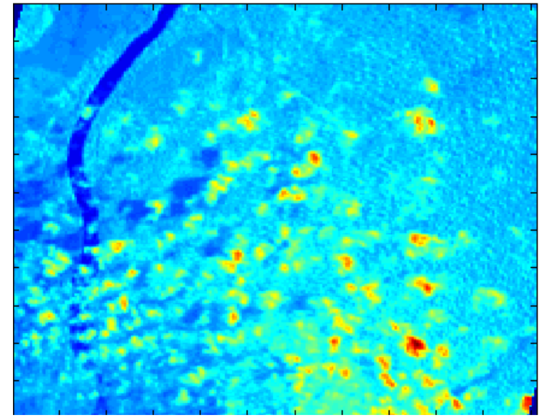
Hyperion Hawaii Scene (x2)



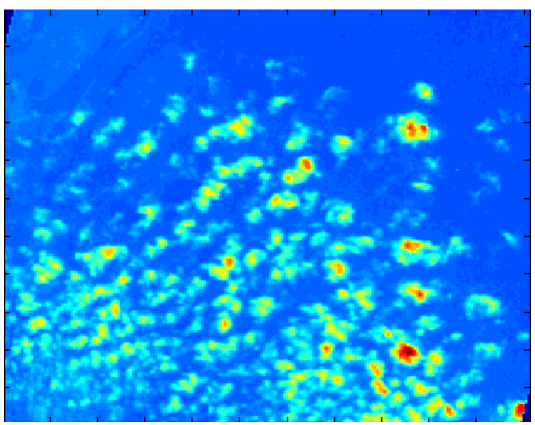
Proposed ADMM – DL
18 th Band Recovery



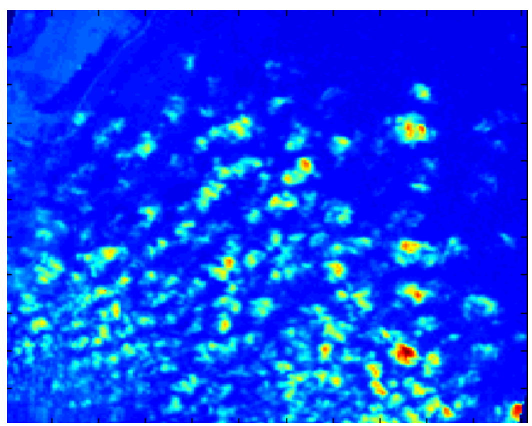
Proposed ADMM – DL
26th Band Recovery



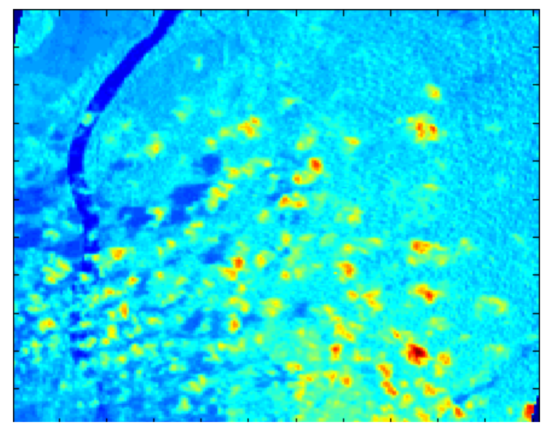
Proposed ADMM - DL
47th Band Recovery



Ground Truth 18th Band



Ground Truth 26th Band



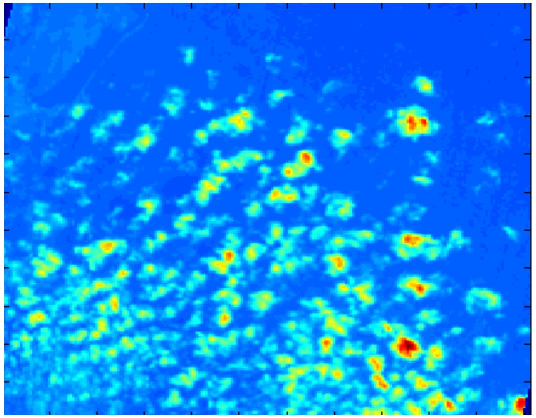
Ground Truth 47th Band

Hawaii Scene (x2)

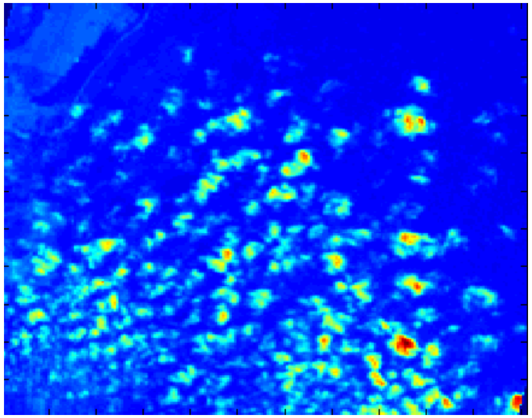
PSNR Recovery of the 3D-cube:

ADMM – DL → **49.90** dB

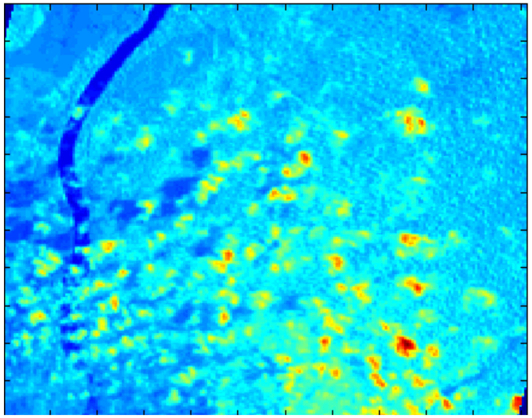
KSVD – DL → 48.62 dB



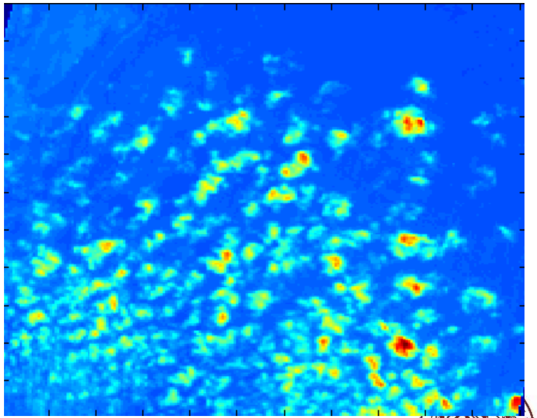
ADMM – DL, 18th Band



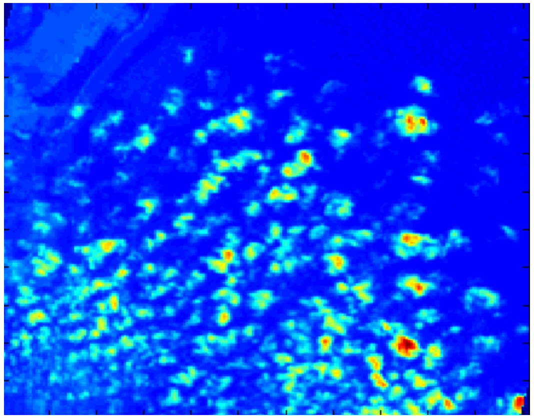
ADMM – DL, 34th Band



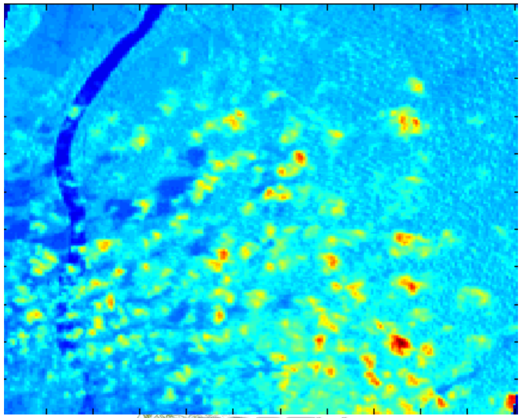
ADMM - DL , 47th Band



KSVD- DL ,18th Band

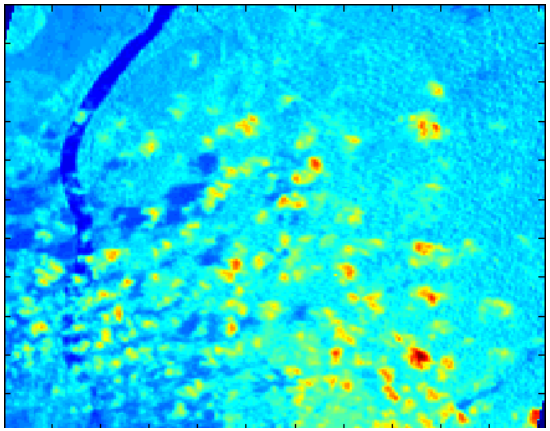


KSVD- DL ,34th Band

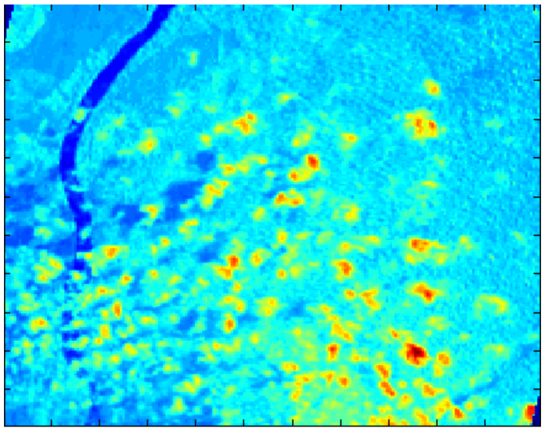


KSVD- DL ,47th Band

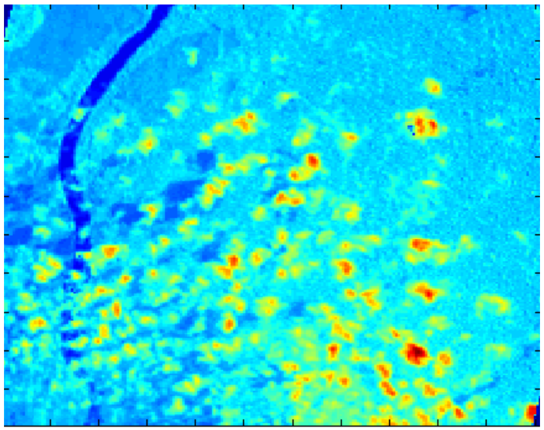
Hyperion Hawaii Scene (x3)



Ground Truth 47^{th} Band



K-SVD-DL
Recovered 47^{th} Band

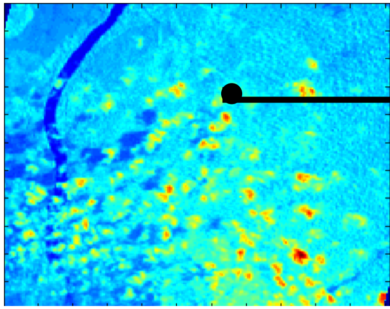


ADMM - DL
Recovered 47^{th} Band

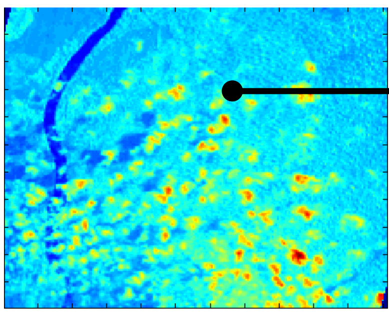
PSNR Recovery of the 3D-cube:

- **ADMM – DL** → **43.93** dB
- **KSVD – DL** → 41.62 dB

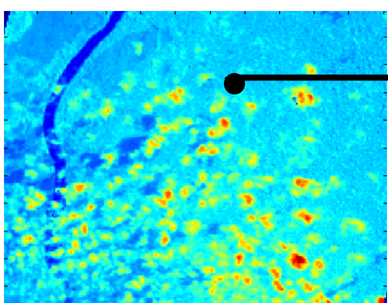
Spectral Signatures of Comparable methods- Hawaii Scene (x3)



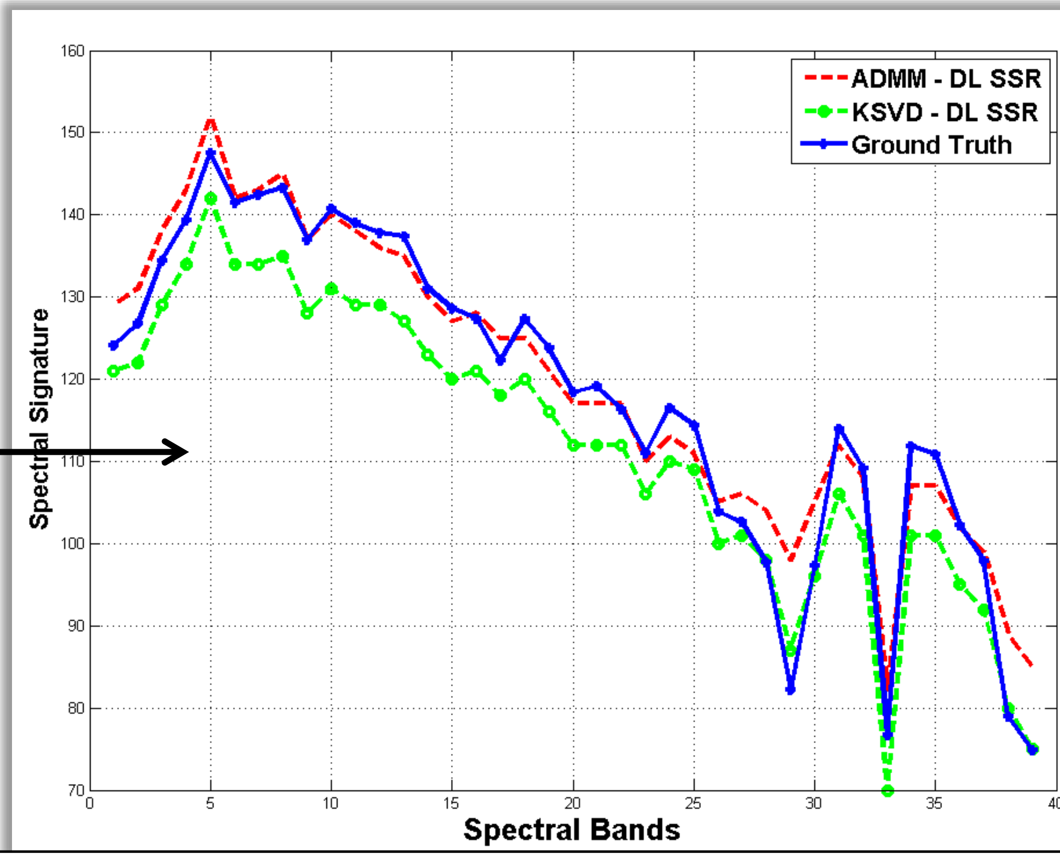
Ground Truth 47th Band



K-SVD- DL SSR 47th Band



ADMM- DL SSR 47th Band



Spectral signatures reconstruction

Semi-Blind Image De-blurring

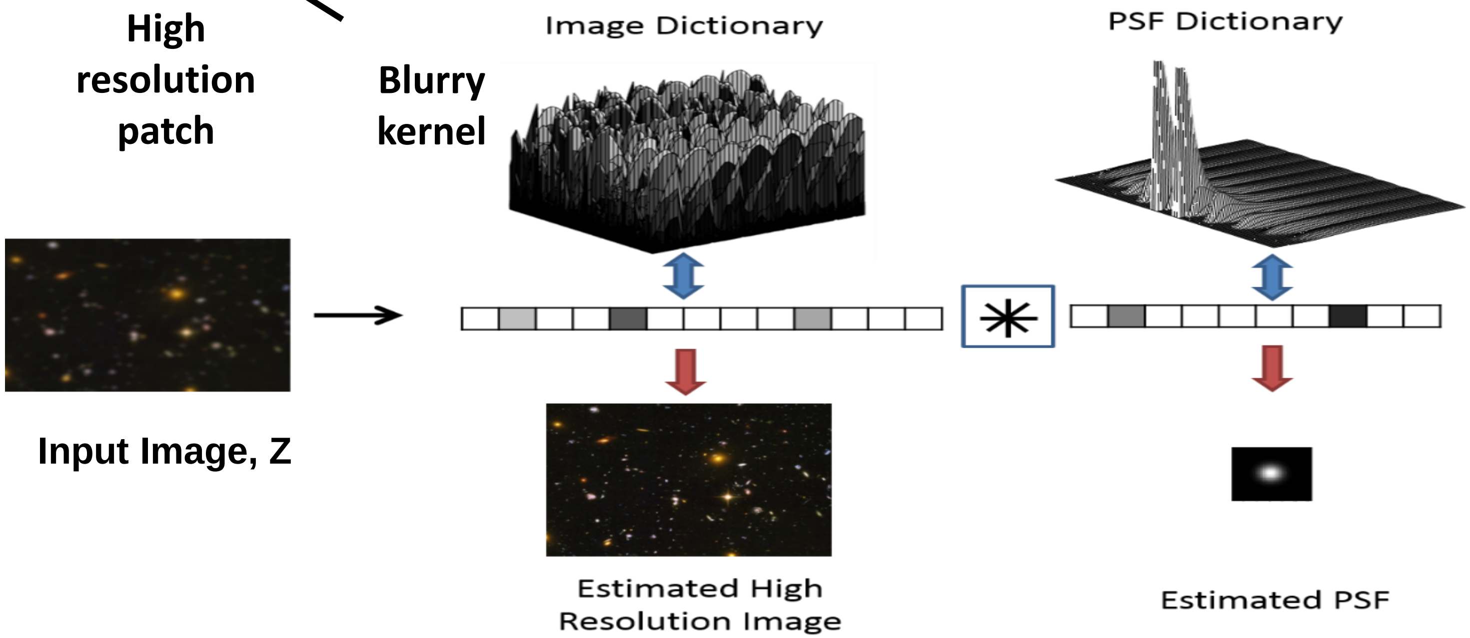
- **Assumption:** *Both the high quality image & the blurring kernel admit sparse representations in appropriate dictionaries !*

$$\mathbf{z} = \mathbf{s} * \mathbf{h} = (\mathbf{D}\mathbf{w}) * (\mathbf{B}\mathbf{k}), \quad \text{where:} ||\mathbf{w}||_0 \leq m, \quad \text{and} ||\mathbf{k}||_0 \leq p$$

Blurry patch

High resolution patch

Blurry kernel



ADMM for Semi-Blind Image De-blurring

- ***Optimization problem:***

$$\min_{\mathbf{w},\mathbf{k}} ||\mathbf{w}||_1 + ||\mathbf{k}||_1 \ \text{ subject to } \ ||\mathbf{z} - (\mathbf{D}\cdot\mathbf{w}) * (\mathbf{B}\cdot\mathbf{k})||_2^2 = 0$$

- **Setting**: $\mathbf{p} = \mathbf{w}$ **and** $\mathbf{q} = \mathbf{k}$

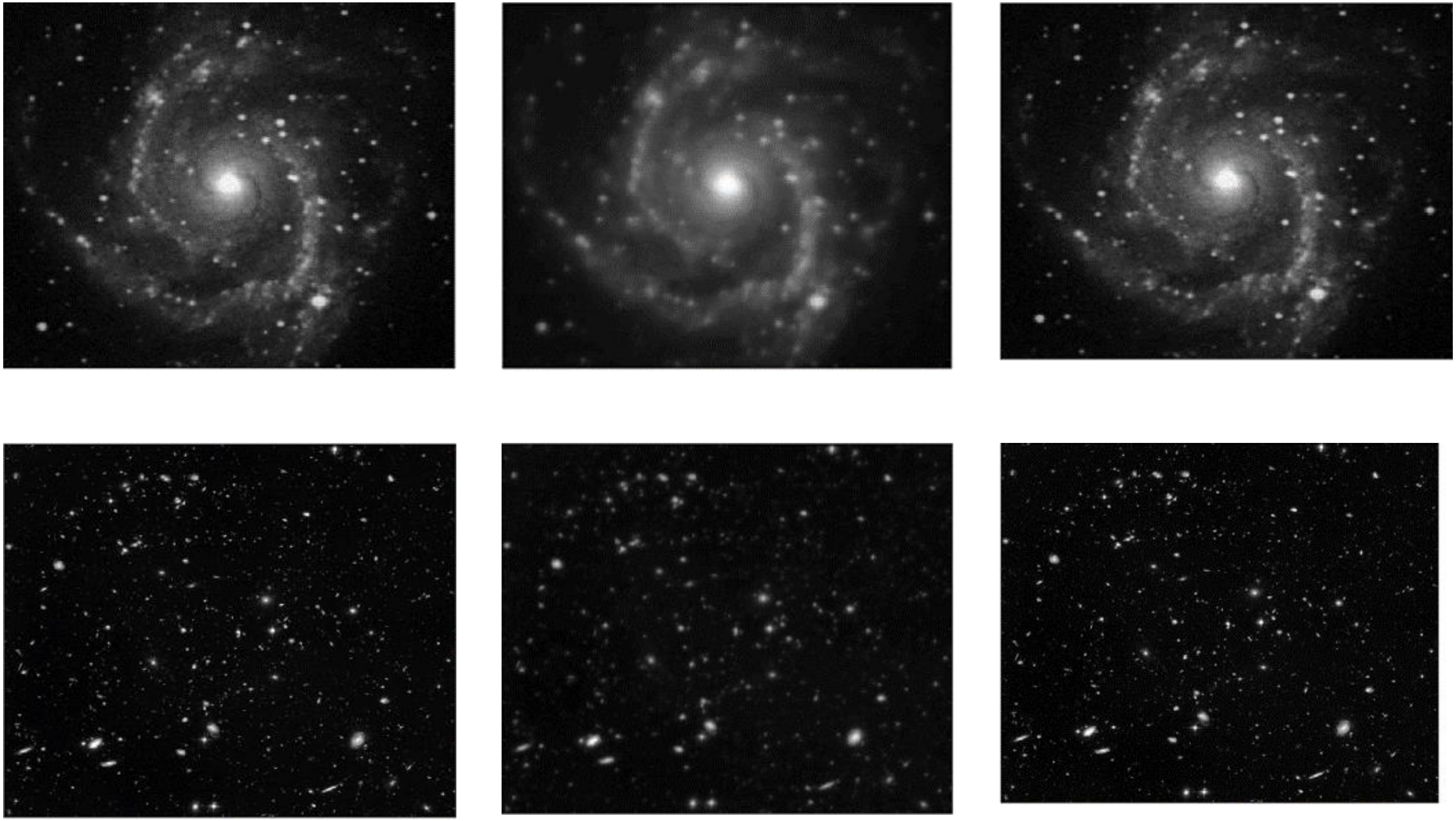
$$\min_{\mathbf{w},\mathbf{k},\mathbf{p},\mathbf{q}} ||\mathbf{z} - (\mathbf{D}\cdot\mathbf{w}) * (\mathbf{B}\cdot\mathbf{k})||_2^2 + \lambda_1||\mathbf{p}||_1 + \lambda_2||\mathbf{q}||_1 = 0, \ \text{ subject to}$$

$$\mathbf{p} - \mathbf{w} = 0, \ \mathbf{q} - \mathbf{k} = 0, \ ||\mathbf{D}(:,i)||_2 \le 1, \ ||\mathbf{B}(:,i)||_2 \le 1$$

- ***Augmented Lagrangian Function:***

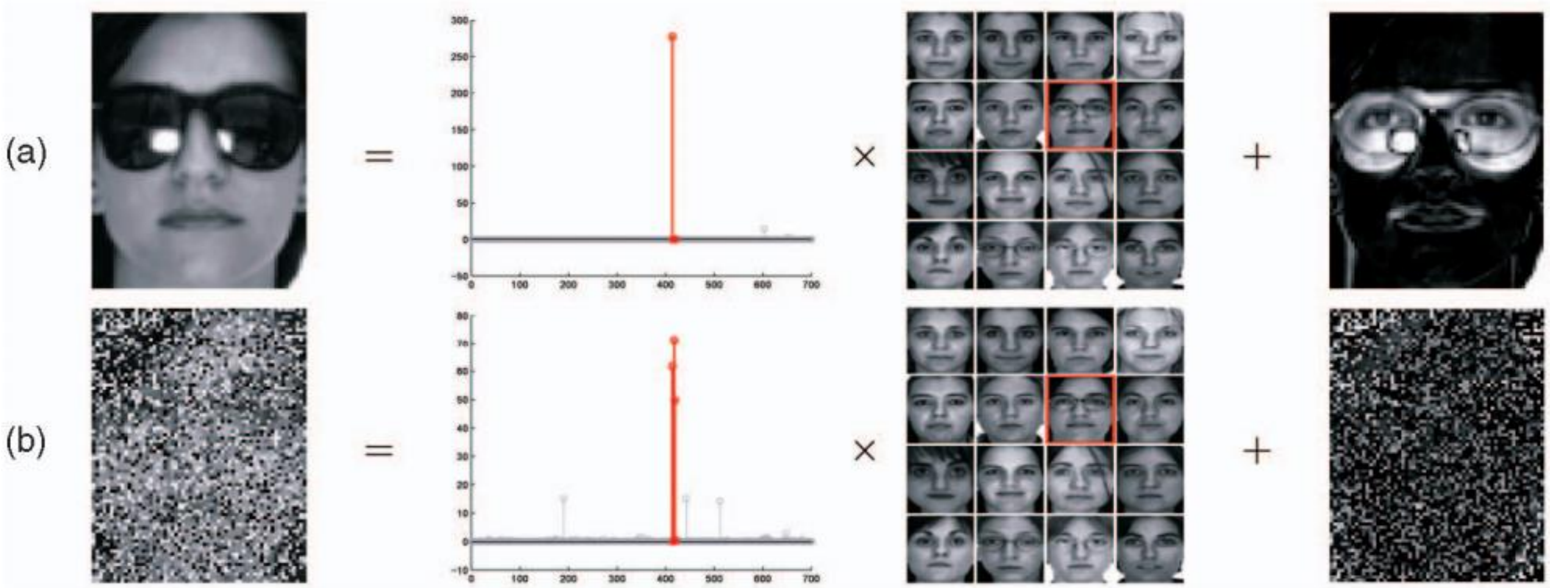
$$L(\mathbf{w},\mathbf{k},\mathbf{p},\mathbf{q},\mathbf{y}_1,\mathbf{y}_2) = ||\mathbf{z} - (\mathbf{D}\cdot\mathbf{w}) * (\mathbf{B}\cdot\mathbf{k})||_2^2 + \lambda_1||\mathbf{p}||_1 + \lambda_2||\mathbf{q}||_1 +$$

$$\mathbf{y}_1^T(\mathbf{p} - \mathbf{w}) + \mathbf{y}_2^T(\mathbf{q} - \mathbf{k}) + \frac{c_1}{2}||\mathbf{p} - \mathbf{w}||_2^2 + \frac{c_2}{2}||\mathbf{q} - \mathbf{k}||_2^2,$$

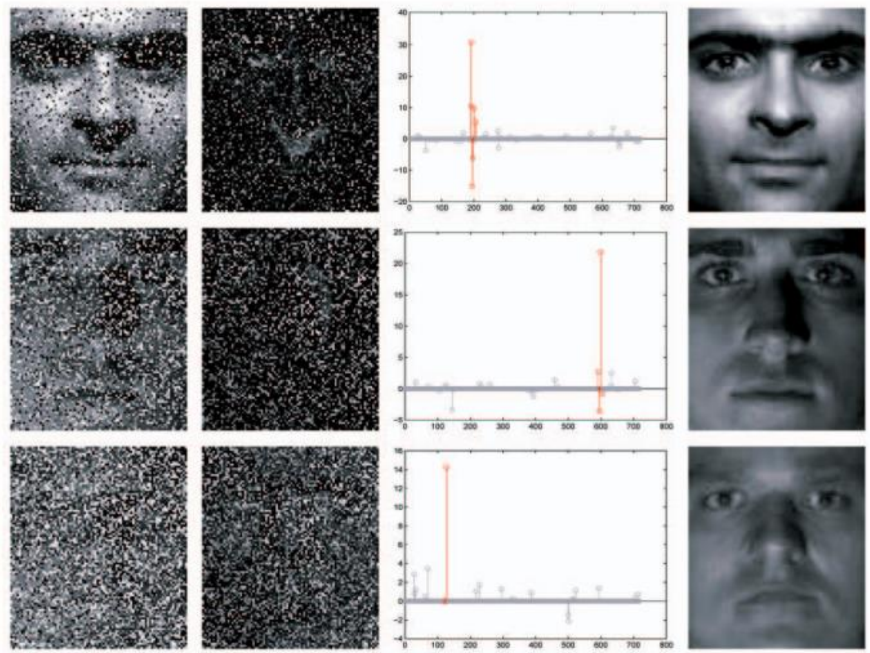


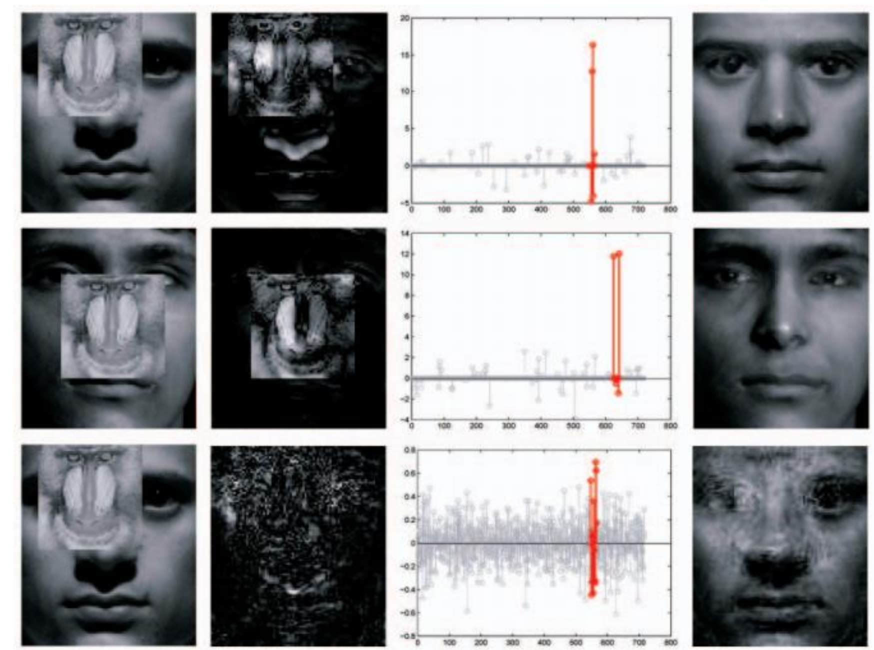
Ground truth　　　Blurred　　　Restored

Face Recognition



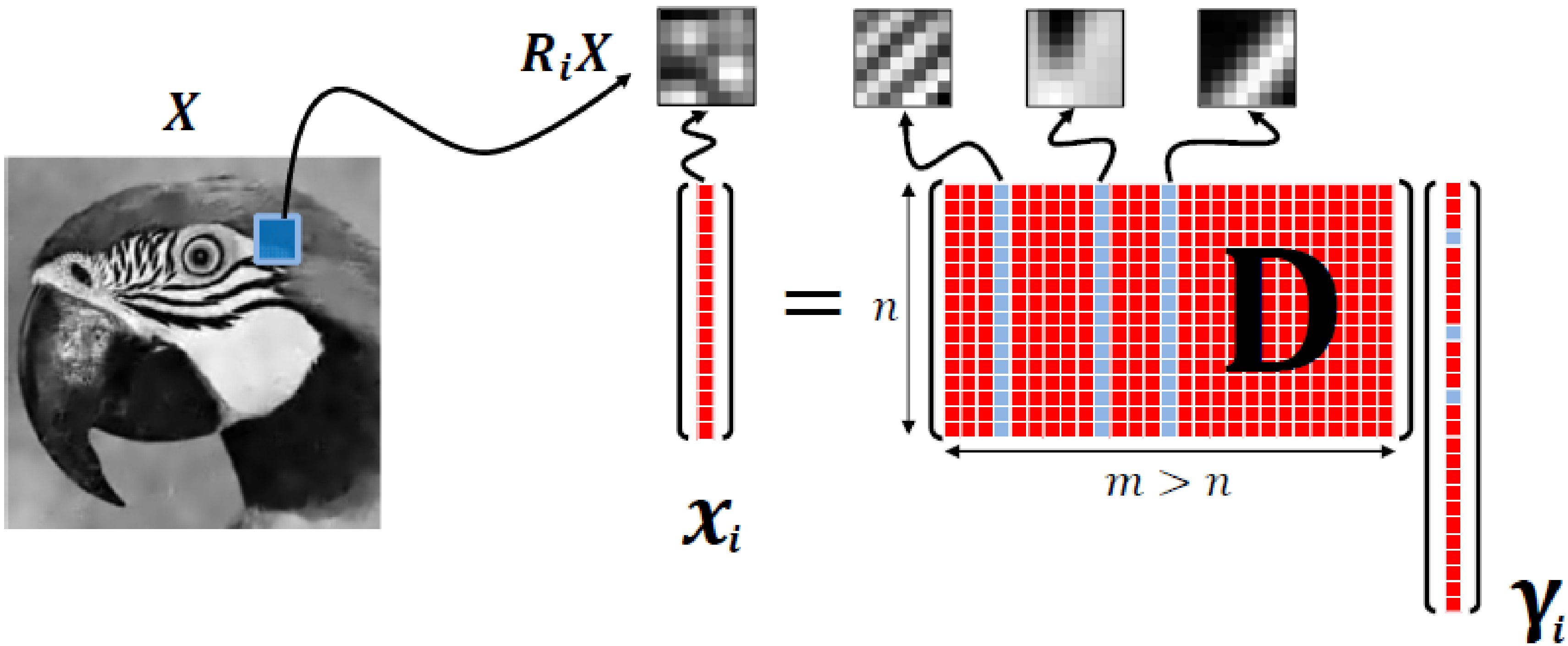
Face Recognition





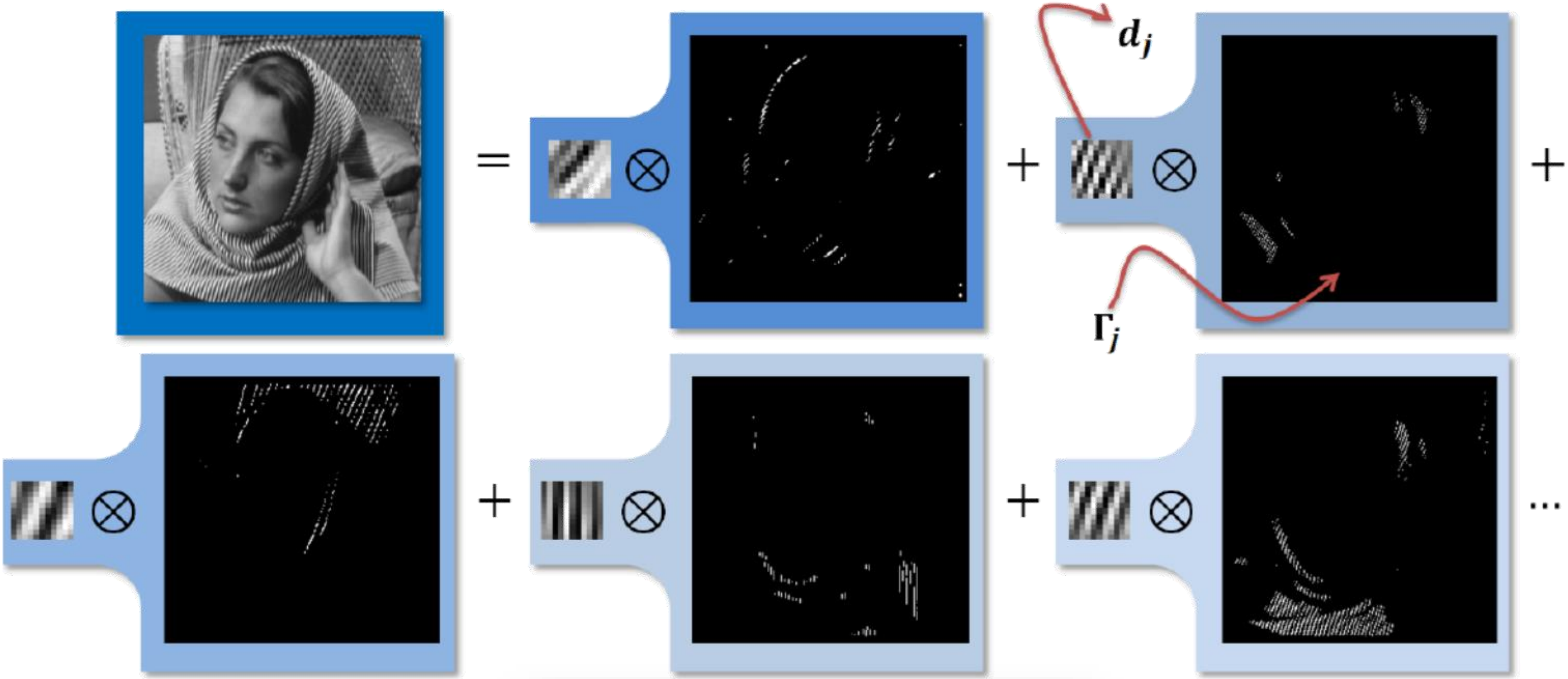
FORTH Institute of Computer Science

Classical Sparse Coding



$$(P_0): \quad \min_{\gamma} \quad \|\gamma\|_0 \quad \text{s.t.} \quad \mathbf{x}_i = \mathbf{D}\gamma_i$$

Convolutional Sparse Coding (CSC)

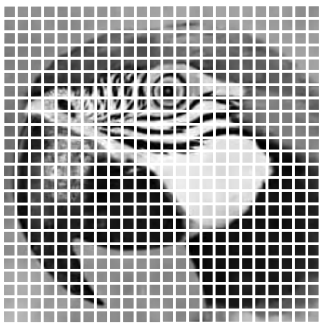


$$\mathbf{X} = \sum_{j=1}^{m} \mathbf{d}_j * \mathbf{\Gamma}_j$$

Convolutional Sparse Coding (CSC)

m filters convolved with their sparse representations

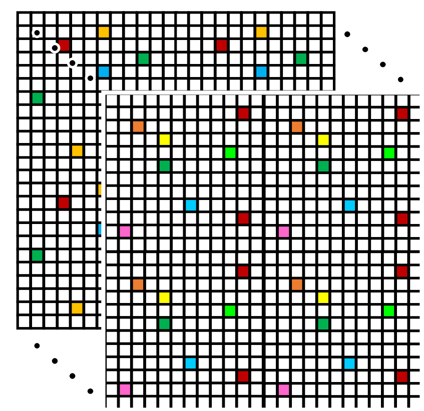
i-th feature-map: An image of the same size as $\mathbf{X}$ holding the sparse representation related to the i-filter



$$\mathbf{X} = \sum_{i=1}^{m} \mathrm{d_i} * \mathrm{z_i}$$

An image held as a column vector of length N

The j-th filter of small support n



Convolutional Sparse Coding (CSC)

- Global model with shift-invariant local prior
- Inherently no disagreement between overlapping patches
- Related to current practices (i.e., patch averaging)

Optimization

- ADMM,
- Proximal Gradient,
- block-Toeplitz

$$\min_{d,x} \|\mathbf{y} - \sum_{i=1}^{M} \mathbf{d}_i * \mathbf{x}_i\|_2^2 + \lambda \|\mathbf{x}_i\|_0$$

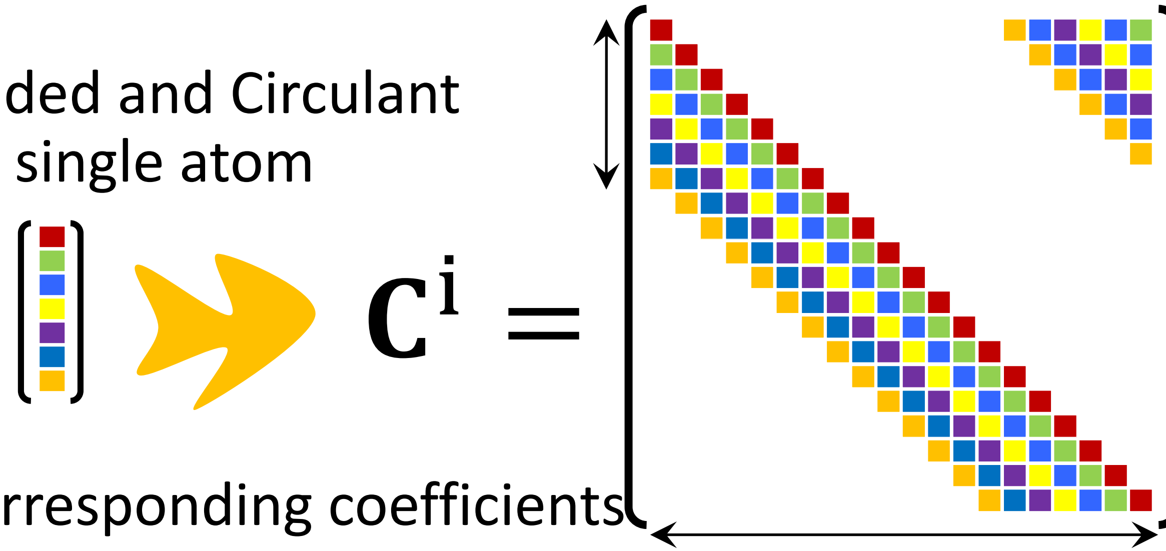
$$\text{s.t.} \quad \|\mathbf{d}_i\|_2 \leq 1 \;\; \forall i$$

CSC in Matrix Form

- Here is an alternative global sparsity-based model formulation

$$\mathbf{Y} = \sum_{\mathrm{i}=1}^{m} \mathbf{C}^{\mathrm{i}} \mathbf{X}^{\mathrm{i}} = \begin{bmatrix} \mathbf{C}^{1} & \cdots & \mathbf{C}^{m} \end{bmatrix} \begin{bmatrix} \mathbf{X}^{1} \\ \vdots \\ \mathbf{X}^{m} \end{bmatrix} = \mathbf{D}\boldsymbol{X}$$

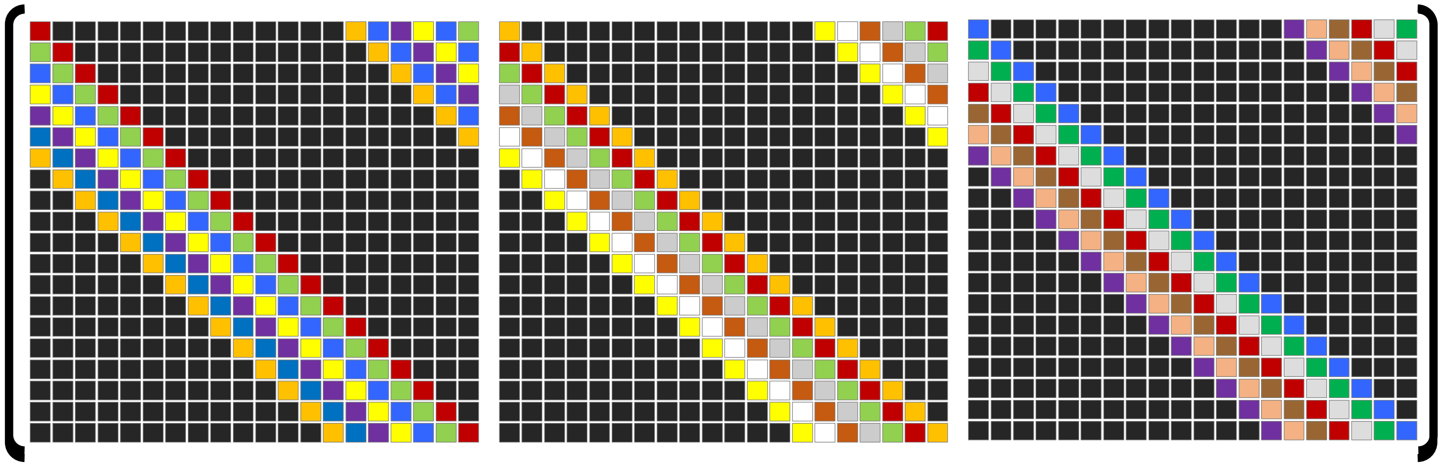
- $\mathbf{C}^{\mathrm{i}} \in \mathbb{R}^{N \times N}$ is a banded and Circulant matrix containing a single atom with all of its shifts



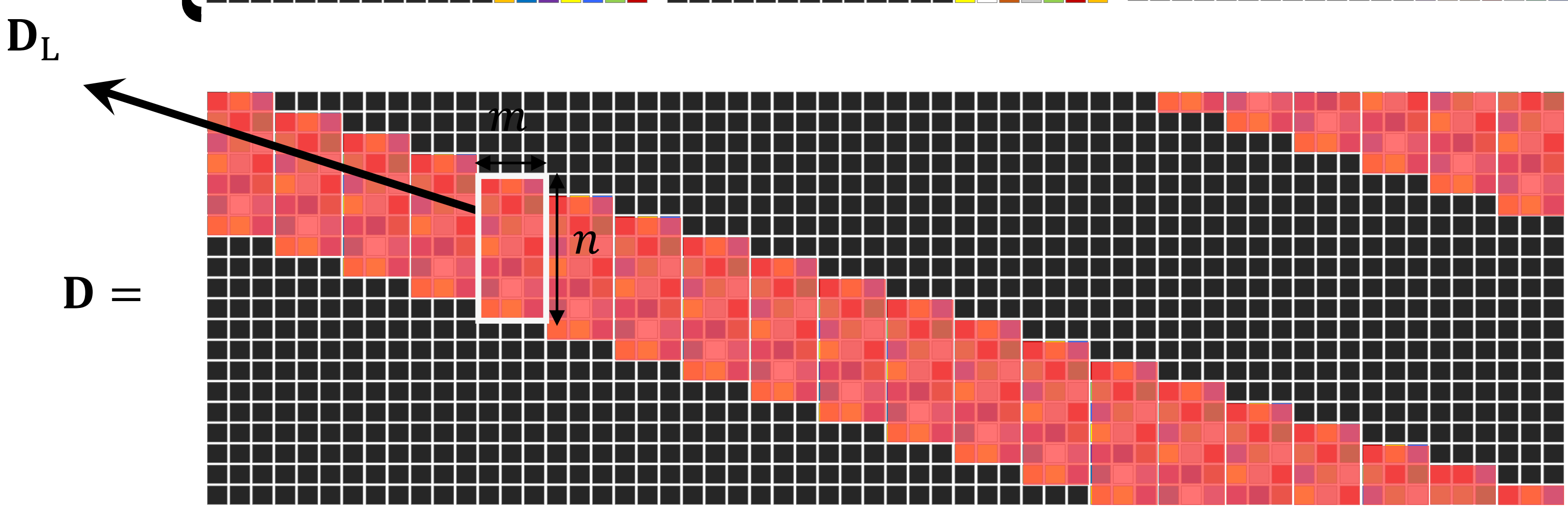
- $\mathbf{X}^{\mathrm{i}} \in \mathbb{R}^{N}$ are the corresponding coefficients ordered as column vectors

The CSC dictionary

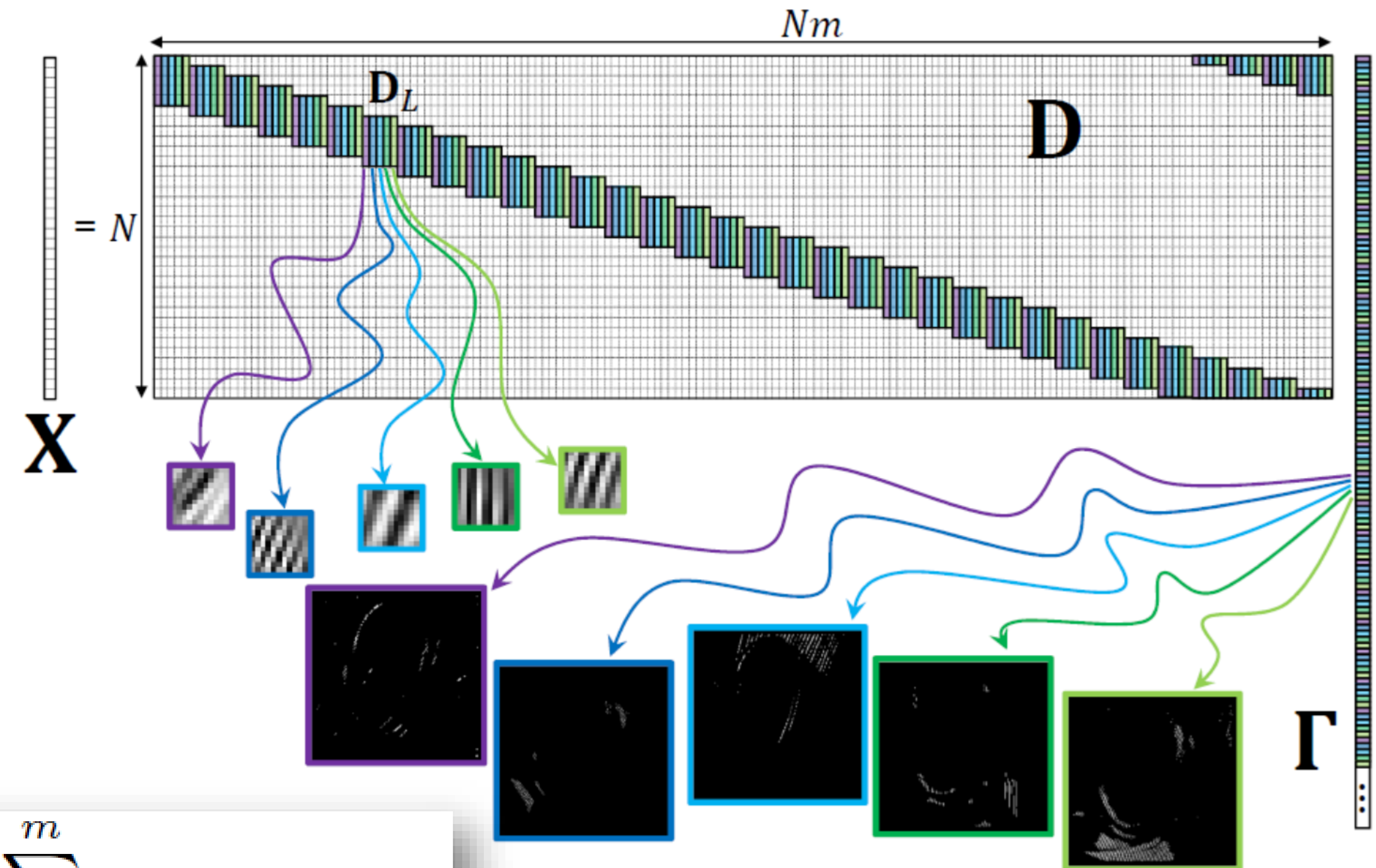
$[\mathbf{C}^1\ \mathbf{C}^2\ \mathbf{C}^3] =$



$\mathbf{D} =$



CSC mapping



$$\mathbf{X} = \sum_{j=1}^{m} \mathbf{d}_j * \mathbf{\Gamma}_j = \mathbf{D}\mathbf{\Gamma}$$

Convolutional Dictionary Learning

$$\underset{\{\mathbf{d}_m\},\{\mathbf{x}_{m,k}\}}{\arg\min} \frac{1}{2}\sum_k \left\| \sum_m \mathbf{d}_m * \mathbf{x}_{m,k} - \mathbf{s}_k \right\|_2^2 + \lambda \sum_{m,k} \|\mathbf{x}_{m,k}\|_1$$

$$\text{such that } \|\mathbf{d}_m\|_2 = 1 \; \forall m ,$$

The training images $\mathbf{s}_k$ are considered to be N dimensional vectors, where N is the number of pixels in each image, and we denote the number of filters and the number of training images by M and K respectively.

Non-linear Sparsity Models

Generic formulation $\min \mathcal{F}(\mathbf{y}, \mathbf{s}) \;\; \text{s.t.} \;\; \mathbf{s} \in \mathcal{S}$

State-of-the-art

- Kernels $\mathbf{y} \rightarrow \phi(\mathbf{x}) \qquad \mathcal{K}(x_i, x_j) = \phi(x_i)\phi(x_j)$

$$\min \|\phi(\mathbf{y}) - \mathbf{D}\phi(\mathbf{s})\|_2 \;\; \text{s.t.} \;\; \|\mathcal{K}(\mathbf{s}_i, \mathbf{s}_j)\|_1 \leq K$$

- Quantization $\mathcal{F}(\mathbf{y}, \mathbf{s}) = \mathcal{Q}(\mathbf{y}, \mathbf{s})$
  - Quantized Orthogonal Matching Pursuit (Q-OMP)

$$\min \|\mathbf{y} - \mathcal{Q}(\mathbf{\Phi}\mathbf{D}\mathbf{x})\|_2 \;\; \text{s.t.} \;\; \|\mathbf{x}\|_0 \leq K$$

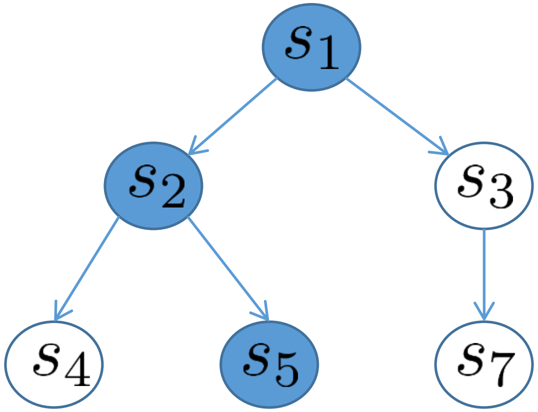
Chen, Yi, Nasser M. Nasrabadi, and Trac D. Tran. "Hyperspectral image classification via kernel sparse representation." IEEE trans. *Geoscience and Remote Sensing,* 2013.
Recovery of quantized compressed sensing measurements, G Tsagkatakis, P Tsakalides, IS&T/SPIE EI 2015

Hierarchical Sparse Coding

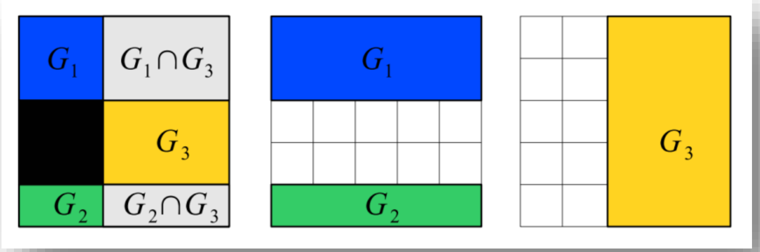
- Assume hierarchical sparse coding

$$\mathbf{y} = \mathbf{D}\mathbf{s} \text{ and } \Omega(\mathbf{s}) \leq K$$

  - $s_i$possible only if ancestor is active
  - Structure in sparsity inducing norm
  - Solution via proximal operators
- Extension to other structures



$$\Omega(\mathbf{s}) = \sum_{G \in \mathcal{G}} \Big(\sum_{i \in G} (d_i^G)^2 \|s_i\|^2 \Big)^{\frac{1}{2}}$$



Jenatton, R., Mairal, J., Obozinski, G., & Bach, F. (2011). Proximal methods for hierarchical sparse coding. *The Journal of Machine Learning Research.*

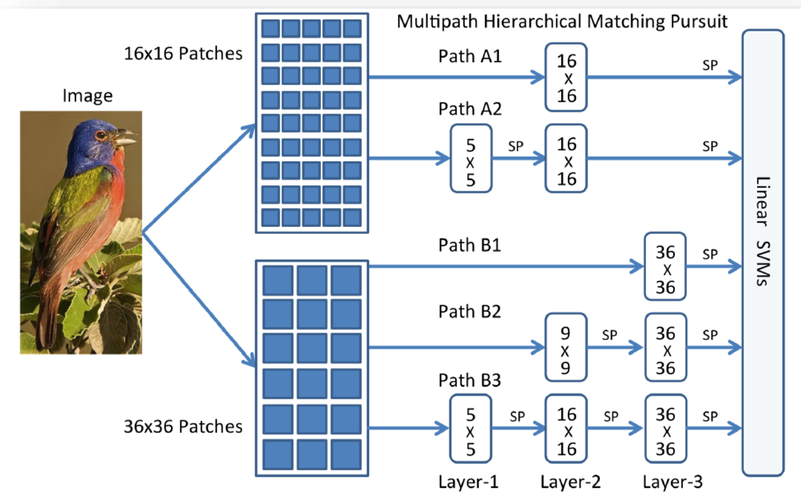
Jenatton, R., Audibert, J. Y., & Bach, F. (2011). Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research.*

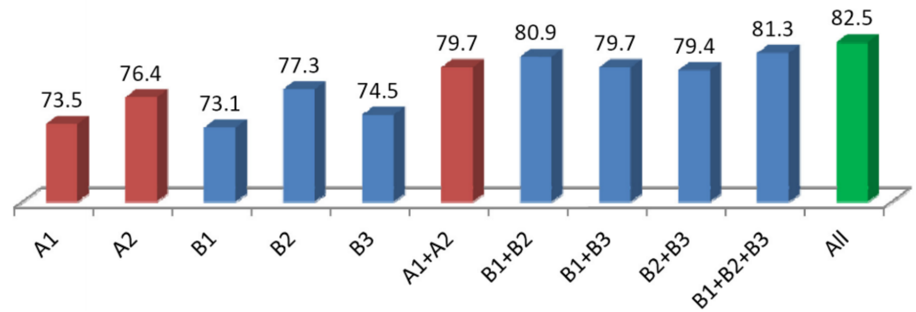
Multipath sparse coding

- Multipath Hierarchical Matching Pursuit
- Dictionary learning
  - Reconstruction error
  - Mutual Coherence

$$\min_{D,X} \|\mathbf{Y} - \mathbf{DX}\|_F + \lambda \sum_{i=1}^{M} \sum_{j=1, i\neq j}^{M} |\mathbf{d}_i^T \mathbf{d}_j|$$

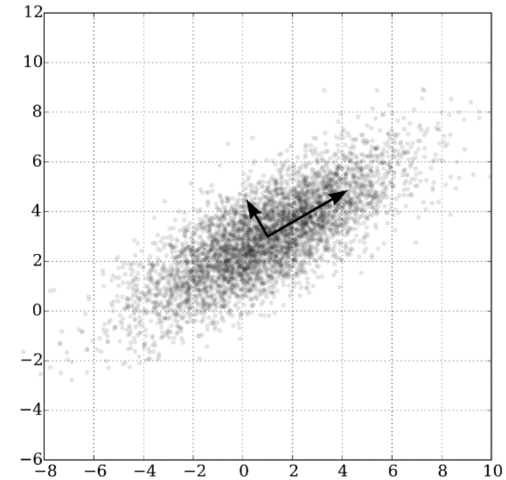
$$\text{s.t.} \quad \|\mathbf{x}_i\|_0 \leq K \qquad \|\mathbf{d}_i\|_2 = 1 \ \ \forall i$$





Bo, Liefeng, Xiaofeng Ren, and Dieter Fox. "Multipath sparse coding using hierarchical matching pursuit." *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013.

Principal Components Analysis



$$M = L + N$$

- L: low-rank (unobserved)
- N: (small) perturbation

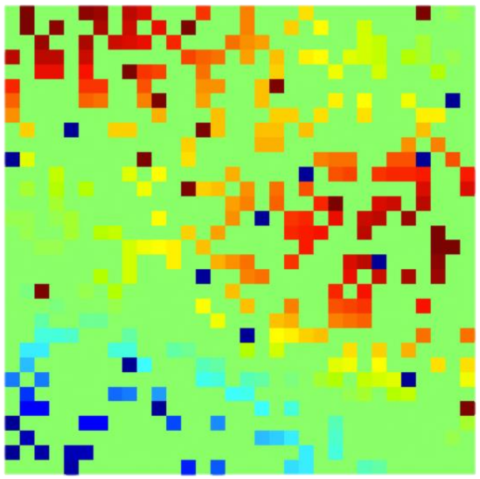
Dimensionality reduction (Schmidt 1907, Hotelling 1933)

$$\begin{array}{ll} \text{minimize} & \|M - \hat{L}\| \\ \text{subject to} & \text{rank}(\hat{L}) \leq k \end{array}$$

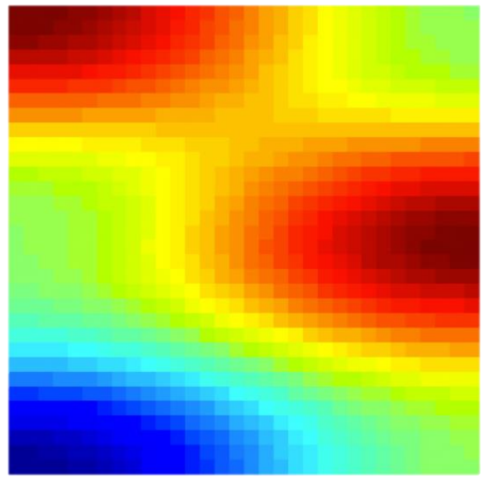
Solution given by truncated SVD

$$M = U\Sigma V^* = \sum_i \sigma_i u_i v_i^* \quad \Rightarrow \quad \hat{L} = \sum_{i \leq k} \sigma_i u_i v_i^*$$

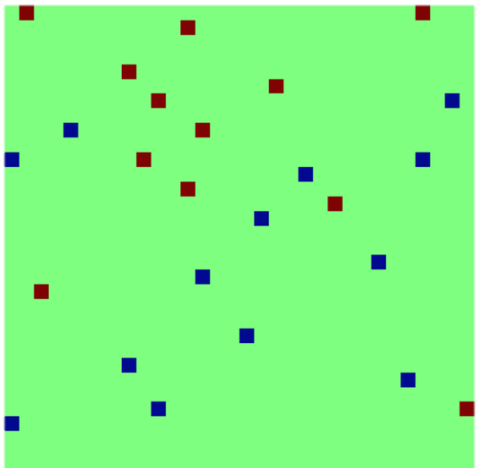
The separation problem



missing + **corrupted** entries



low rank matrix



sparse corruptions

Formulation

Seek the lowest-rank A that agrees with the data up to some sparse error:

$$\min \ \text{rank}(A) + \gamma \|E\|_0 \quad \text{subj} \quad A + E = D.$$

Not directly tractable, relax:

$$\|E\|_0 = \#\{E_{ij} \neq 0\} \quad \rightarrow \quad \|E\|_1 = \textstyle\sum_{ij} |E_{ij}|.$$

$$\text{rank}(A) = \#\{\sigma_i(A) \neq 0\} \quad \rightarrow \quad \|A\|_* = \textstyle\sum_i \sigma_i(A).$$

Convex envelope over $B_{2,2} \times B_{1,\infty}$

$$\min \ \|A\|_* + \lambda \|E\|_1 \quad \text{subj} \quad A + E = D.$$

Semidefinite program, solvable in polynomial time

Principal Components Pursuit (PCP)

$$\text{minimize}_{\mathbf{L},\mathbf{E}} \quad \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1$$
$$\text{subject to} \quad \mathbf{L} + \mathbf{E} = \mathbf{M}$$

Theorem (C., Li, Ma and Wright, 09)

- *L is $n \times n$ of rank$(L) \leq \rho_r n \, \mu^{-1} (\log n)^{-2}$*
- *E is $n \times n$, random sparsity pattern of cardinality $m \leq \rho_s n^2$*

Then with probability $1 - O(n^{-10})$, PCP with $\lambda = 1/\sqrt{n}$ is exact:

$$\hat{L} = L, \quad \hat{E} = E$$

Same conclusion for rectangular matrices with $\lambda = 1/\sqrt{\text{max dim}}$

Robust PCA

$$\text{minimize}_{\mathbf{L},\mathbf{E}} \quad \|\mathbf{L}\|_* + \lambda\|\mathbf{E}\|_1$$
$$\text{subject to} \quad \mathcal{A}(\mathbf{L}+\mathbf{E}) = \mathcal{A}(\mathbf{M})$$

- L *as before, rank*$(L) \leq \rho_0\, n\, \mu^{-1} (\log n)^{-2}$
- Ω_{obs} *random set of size* $m = 0.1n^2$ *(missing frac. is arbitrary)*
- *Each observed entry corrupted with prob.* $\tau \leq \tau_0$

Then with prob. $1 - O(n^{-10})$, *PCP with* $\lambda = 1/\sqrt{0.1n}$ *is exact:*

$$\hat{L} = L$$

Same conclusion for rectangular matrices with $\lambda = 1/\sqrt{0.1 \textit{max dim}}$

Application to video surveillance

- Sequence of 200 video frames (144×172 pixels) with a static background
- Problem: detect any activity in the foreground

Repairing vintage movies

Original D **Repaired** A

Corruptions

Frame 1

480×620 pixels

Repairing vintage movies

Original D **Repaired** A

Corruptions

Frame 2

Repairing vintage movies

Original D **Repaired** A

Corruptions

Frame 5

Aligning Bill Gates faces from the Internet

Input

Input: faces detected by a face detector (D)

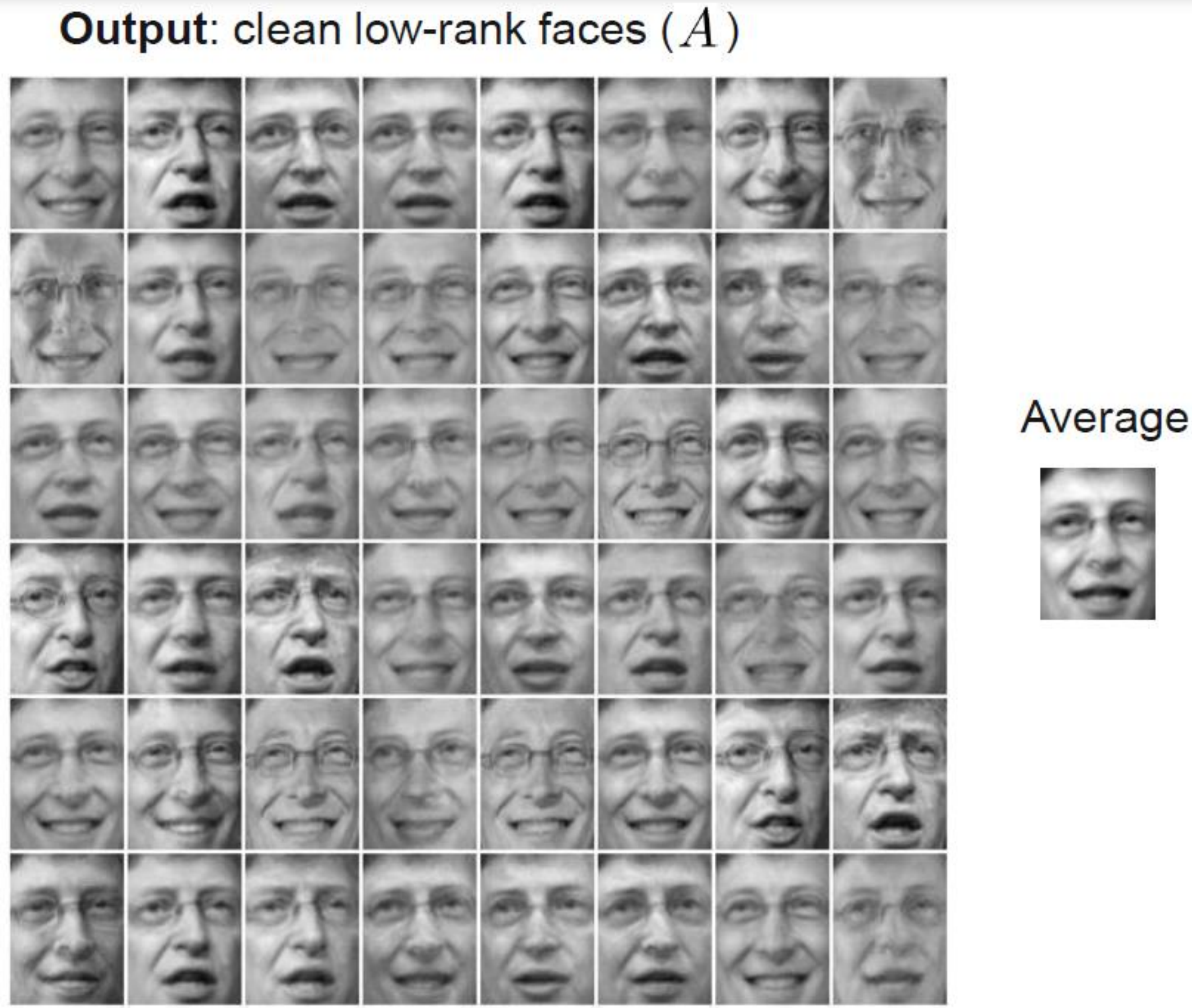
Average

Output

Output: aligned faces ($D \circ \tau$)

Average

Output: clean low-rank faces



Sparse error