Sparse Coding and Dictionary Learning for Image Analysis

Part II: Dictionary Learning for signal reconstruction

Francis Bach, Julien Mairal, Jean Ponce and Guillermo Sapiro

ICCV'09 tutorial, Kyoto, 28th September 2009

What this part is about

- The learning of compact representations of images adapted to restoration tasks.
- A fast online algorithm for learning dictionaries and factorizing matrices in general.
- Various formulations for image and video processing.

The Image Denoising Problem

$$\underbrace{\mathbf{y}}_{\text{measurements}} = \underbrace{\mathbf{x}_{orig}}_{\text{original image}} + \underbrace{\mathbf{w}}_{\text{noise}}$$

Sparse representations for image restoration

$$\underbrace{\mathbf{y}}_{\text{measurements}} = \underbrace{\mathbf{x}_{orig}}_{\text{original image}} + \underbrace{\mathbf{w}}_{\text{noise}}$$

Energy minimization problem - MAP estimation

$$E(\mathbf{x}) = \underbrace{||\mathbf{y}-\mathbf{x}||_2^2}_{\text{relation to measurements}} + \underbrace{Pr(\mathbf{x})}_{\text{prior}}$$

Some classical priors

- Smoothness $\lambda||\mathcal{L}\mathbf{x}||_2^2$
- Total variation $\lambda||\nabla\mathbf{x}||_1^2$
- Wavelet sparsity $\lambda||\mathbf{W}\mathbf{x}||_1$
- ...

Sparse representations for image restoration

Sparsity and redundancy

$$Pr(\mathbf{x}) = \lambda||\boldsymbol{\alpha}||_0 \text{ for } \mathbf{x} \approx \mathbf{D}\boldsymbol{\alpha}$$

$$\underbrace{\begin{pmatrix} \\ \mathbf{x} \\ \\ \end{pmatrix}}_{\mathbf{x}\in\mathbb{R}^m} = \underbrace{\left(\begin{array}{c|c|c|c} & & & \\ \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \\ & & & \end{array}\right)}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \boldsymbol{\alpha}[1] \\ \boldsymbol{\alpha}[2] \\ \vdots \\ \boldsymbol{\alpha}[p] \end{pmatrix}}_{\boldsymbol{\alpha}\in\mathbb{R}^p,\text{sparse}}$$

Sparse representations for image restoration

Designed dictionaries

[Haar, 1910], [Zweig, Morlet, Grossman ∼70s], [Meyer, Mallat, Daubechies, Coifman, Donoho, Candes ∼80s-today]. . . (see [Mallat, 1999])
Wavelets, Curvelets, Wedgelets, Bandlets, . . . lets

Learned dictionaries of patches

[Olshausen and Field, 1997], [Engan et al., 1999], [Lewicki and Sejnowski, 2000], [Aharon et al., 2006] , [Roth and Black, 2005], [Lee et al., 2007]

$$\min_{\boldsymbol{\alpha}_i, \mathbf{D} \in \mathcal{C}} \sum_i \underbrace{\frac{1}{2}||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2}_{\text{reconstruction}} + \underbrace{\lambda\psi(\boldsymbol{\alpha}_i)}_{\text{sparsity}}$$

- $\psi(\boldsymbol{\alpha}) = ||\boldsymbol{\alpha}||_0$ (“ℓ_0 pseudo-norm”)
- $\psi(\boldsymbol{\alpha}) = ||\boldsymbol{\alpha}||_1$ (ℓ_1 norm)

Solving the denoising problem

[Elad and Aharon, 2006]

- Extract all overlapping 8×8 patches $\mathbf{y}_i$.
- Solve a matrix factorization problem:

$$\min_{\boldsymbol{\alpha}_i, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^{n} \underbrace{\frac{1}{2}||\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2}_{\text{reconstruction}} + \underbrace{\lambda\psi(\boldsymbol{\alpha}_i)}_{\text{sparsity}},$$

with $n > 100,000$

- Average the reconstruction of each patch.

Sparse representations for image restoration

K-SVD: [Elad and Aharon, 2006]

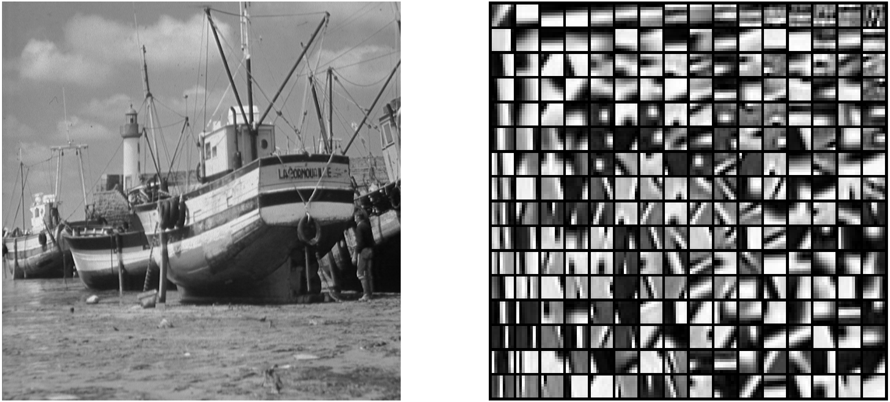


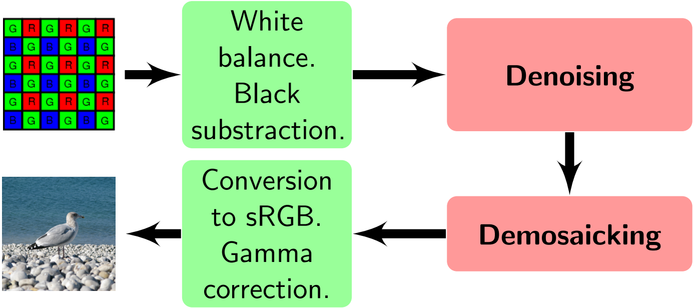
Figure: Dictionary trained on a noisy version of the image boat.

Sparse representations for image restoration

Inpainting, Demosaicking

$$\min_{\mathbf{D}\in\mathcal{C},\boldsymbol{\alpha}} \sum_i \frac{1}{2}||\boldsymbol{\beta}_i \otimes (\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i)||_2^2 + \lambda_i \psi(\boldsymbol{\alpha}_i)$$

RAW Image Processing (see our poster)

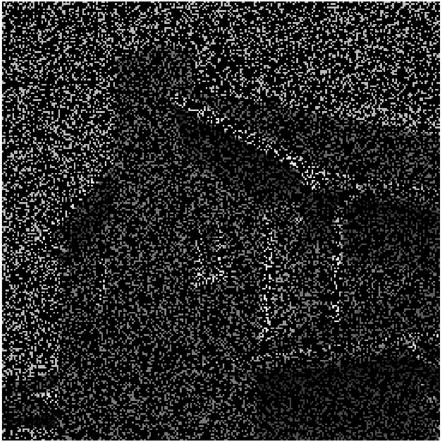


Sparse representations for image restoration

[Mairal, Bach, Ponce, Sapiro, and Zisserman, 2009c]

Sparse representations for image restoration

[Mairal, Sapiro, and Elad, 2008b]



Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008a]

Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008a]

Sparse representations for video restoration

Key ideas for video processing

[Protter and Elad, 2009]

- Using a 3D dictionary.
- Processing of many frames at the same time.
- Dictionary propagation.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008b]

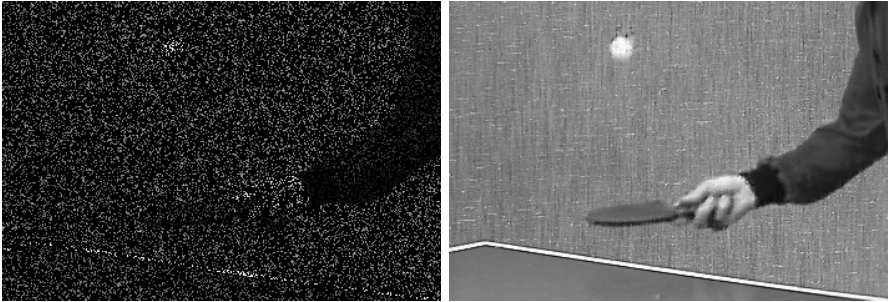


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008b]

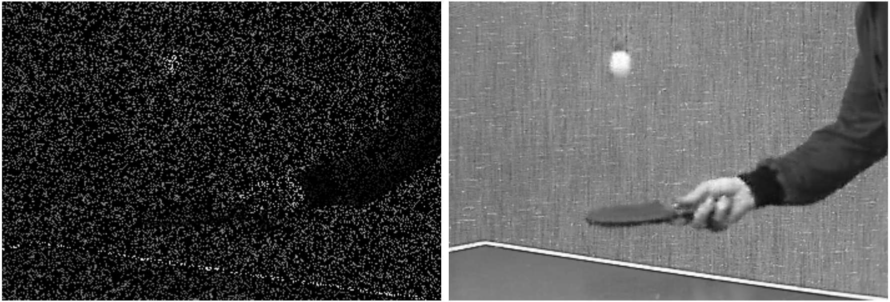


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008b]

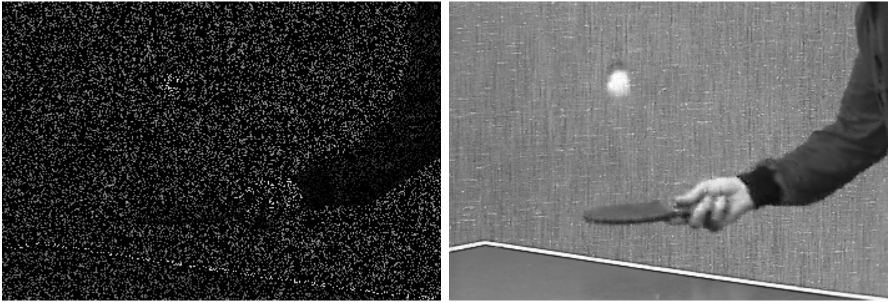


Figure: Inpainting results.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008b]

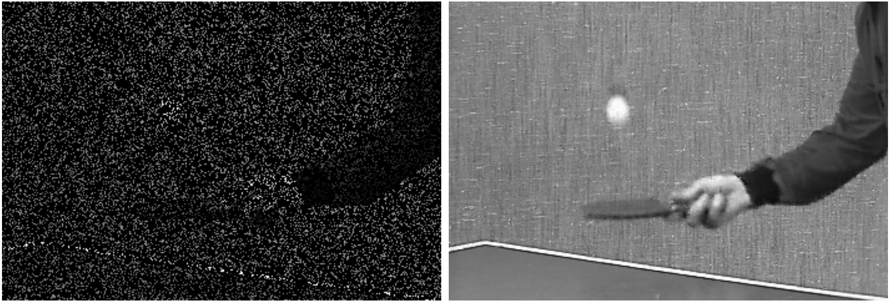


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008b]

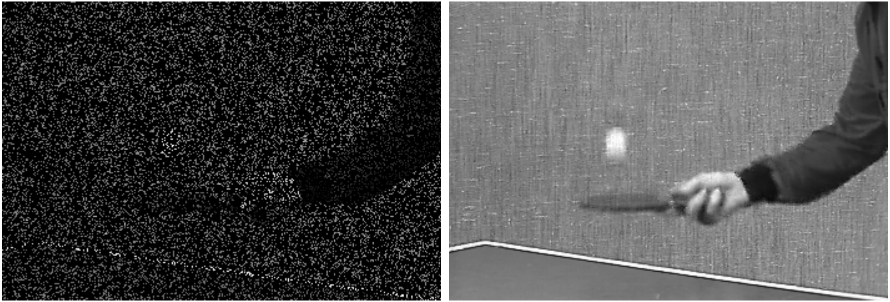


Figure: Inpainting results.

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008b]

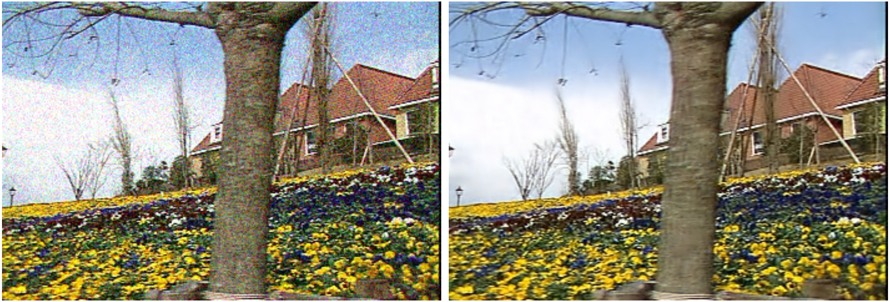


Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008b]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008b]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration
Color video denoising, [Mairal, Sapiro, and Elad, 2008b]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008b]

Figure: Denoising results. $\sigma = 25$

Optimization for Dictionary Learning

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathcal{C}}} \sum_{i=1}^{n} \frac{1}{2}||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda||\boldsymbol{\alpha}_i||_1$$

$$\mathcal{C} \triangleq \{\mathbf{D} \in \mathbb{R}^{m\times p} \ \text{ s.t. } \ \forall j = 1,\dots,p, \ \ ||\mathbf{d}_j||_2 \leq 1\}.$$

- Classical optimization alternates between $\mathbf{D}$ and $\boldsymbol{\alpha}$.
- Good results, but very slow!

Optimization for Dictionary Learning

[Mairal, Bach, Ponce, and Sapiro, 2009a]

Classical formulation of dictionary learning

$$\min_{\mathbf{D}\in\mathcal{C}} f_n(\mathbf{D}) = \min_{\mathbf{D}\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} l(\mathbf{x}_i, \mathbf{D}),$$

where

$$l(\mathbf{x}, \mathbf{D}) \triangleq \min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \frac{1}{2}||\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1.$$

Which formulation are we interested in?

$$\min_{\mathbf{D}\in\mathcal{C}} \left\{ f(\mathbf{D}) = \mathbb{E}_x[l(\mathbf{x}, \mathbf{D})] \approx \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} l(\mathbf{x}_i, \mathbf{D}) \right\}$$

[Bottou and Bousquet, 2008]: Online learning can

- handle potentially infinite or dynamic datasets,
- be dramatically faster than batch algorithms.

Optimization for Dictionary Learning

Require: $\mathbf{D}_0 \in \mathbb{R}^{m \times p}$ (initial dictionary); $\lambda \in \mathbb{R}$

$\mathbf{A}_0 = 0$, $\mathbf{B}_0 = 0$.

for t=1,...,T **do**

Draw $\mathbf{x}_t$

Sparse Coding

$$\boldsymbol{\alpha}_t \leftarrow \operatorname*{arg\,min}_{\boldsymbol{\alpha} \in \mathbb{R}^p} \frac{1}{2} ||\mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1,$$

Aggregate sufficient statistics

$\mathbf{A}_t \leftarrow \mathbf{A}_{t-1} + \boldsymbol{\alpha}_t \boldsymbol{\alpha}_t^T$, $\mathbf{B}_t \leftarrow \mathbf{B}_{t-1} + \mathbf{x}_t \boldsymbol{\alpha}_t^T$

Dictionary Update (block-coordinate descent)

$$\mathbf{D}_t \leftarrow \operatorname*{arg\,min}_{\mathbf{D} \in \mathcal{C}} \frac{1}{t} \sum_{i=1}^{t} \left(\frac{1}{2} ||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda ||\boldsymbol{\alpha}_i||_1 \right).$$

end for

Optimization for Dictionary Learning

Which guarantees do we have?

Under a few reasonable assumptions,

- we build a surrogate function $\hat{f}_t$ of the expected cost f verifying

$$\lim_{t \to +\infty} \hat{f}_t(\mathbf{D}_t) - f(\mathbf{D}_t) = 0,$$

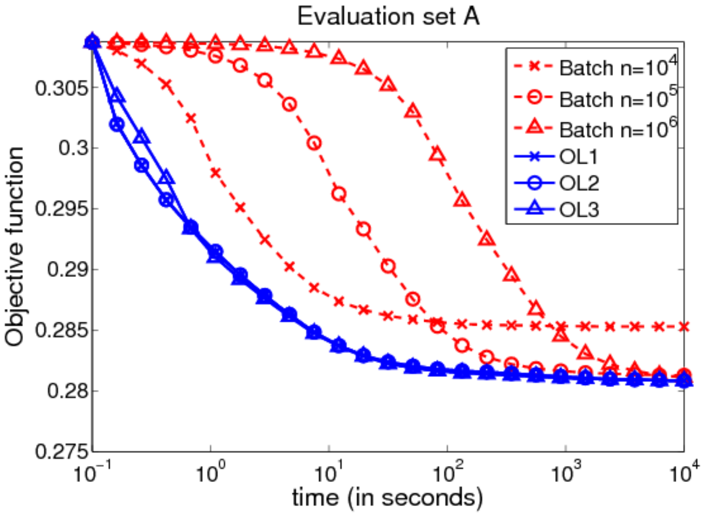
- $\mathbf{D}_t$ is asymptotically close to a stationary point.

Extensions (all implemented in SPAMS)

- non-negative matrix decompositions.
- sparse PCA (sparse dictionaries).
- fused-lasso regularizations (piecewise constant dictionaries)

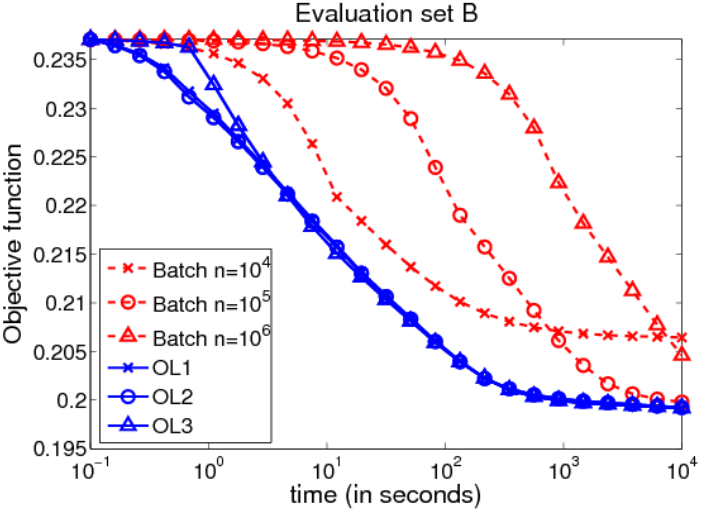
Optimization for Dictionary Learning

Experimental results, batch vs online



$m = 8 \times 8$, $p = 256$

Optimization for Dictionary Learning
Experimental results, batch vs online



$$m = 12 \times 12 \times 3, \; p = 512$$

Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph

Extension to NMF and sparse PCA

[Mairal, Bach, Ponce, and Sapiro, 2009b]

NMF extension

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathcal{C}}} \sum_{i=1}^{n} \frac{1}{2}||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 \ \text{ s.t. } \ \boldsymbol{\alpha}_i \geq 0, \quad \mathbf{D} \geq 0.$$

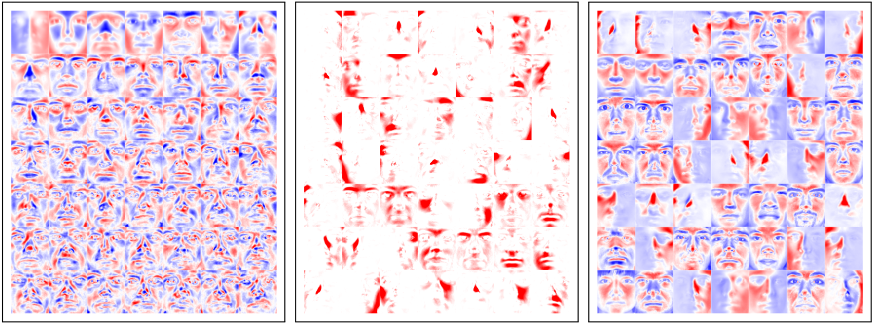
SPCA extension

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathcal{C}'}} \sum_{i=1}^{n} \frac{1}{2}||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda||\boldsymbol{\alpha}_1||_1$$

$$\mathcal{C}' \triangleq \{\mathbf{D} \in \mathbb{R}^{m\times p} \ \text{ s.t. } \ \forall j \ \ ||\mathbf{d}_j||_2^2 + \gamma||\mathbf{d}_j||_1 \leq 1\}.$$

Extension to NMF and sparse PCA

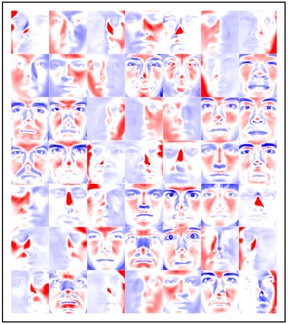
Faces: Extended Yale Database B



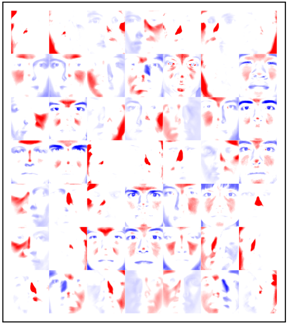
(a) PCA (b) NNMF (c) DL

Extension to NMF and sparse PCA

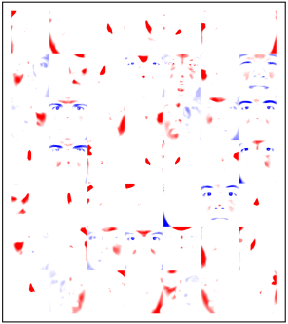
Faces: Extended Yale Database B



(d) SPCA, $\tau = 70\%$



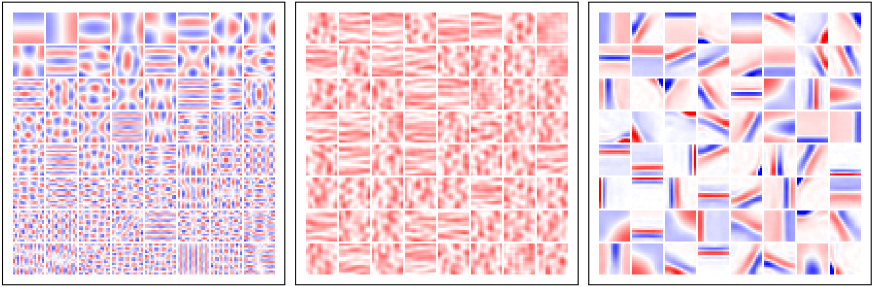
(e) SPCA, $\tau = 30\%$



(f) SPCA, $\tau = 10\%$

Extension to NMF and sparse PCA

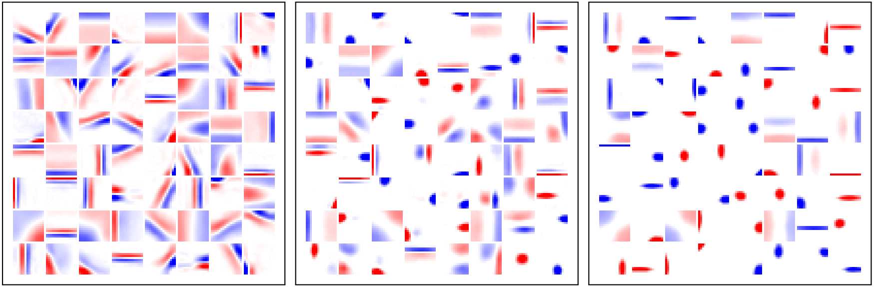
Natural Patches



(a) PCA (b) NNMF (c) DL

Extension to NMF and sparse PCA

Natural Patches



(d) SPCA, $\tau = 70\%$ (e) SPCA, $\tau = 30\%$ (f) SPCA, $\tau = 10\%$

Summary of this part

- The dictionary learning framework leads to state-of-the-art results for many image ...
- ...and video processing tasks.
- Online learning techniques are well-suited for this problem and allows training sets with millions of patches.

References I

M. Aharon, M. Elad, and A. M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.

L. Bottou and O. Bousquet. The trade-offs of large scale learning. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 161–168. MIT Press, Cambridge, MA, 2008.

M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 54(12): 3736–3745, December 2006.

K. Engan, S. O. Aase, and J. H. Husoy. Frame based signal compression using method of optimal directions (MOD). In *Proceedings of the 1999 IEEE International Symposium on Circuits Systems*, volume 4, 1999.

A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.

H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 801–808. MIT Press, Cambridge, MA, 2007.

References II

M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000.

J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, January 2008a.

J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modelling and Simulation*, 7(1): 214–241, April 2008b.

J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009a.

J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *ArXiv:0908.0050v1*, 2009b. submitted.

J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009c.

S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, New York, September 1999.

References III

B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

M. Protter and M. Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing*, 18(1):27–36, 2009.

S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.