

Non-local Sparse Models for Image Restoration

Julien Mairal¹ Francis Bach¹
Jean Ponce² Guillermo Sapiro³ Andrew Zisserman⁴

¹INRIA - WILLOW ²Ecole Normale Supérieure
³University of Minnesota ⁴Oxford University

MSR-INRIA workshop, January 25th 2010

What this talk is about

- Exploiting self-similarities in images and learned sparse representations.
- A fast online algorithm for learning dictionaries and factorizing matrices in general.
- Various formulations for image and video processing, leading to state-of-the-art results in image denoising and demosaicking.

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

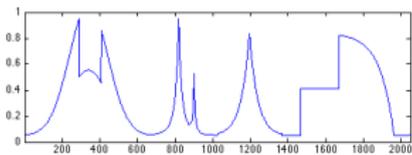
$$E(\mathbf{x}) = \underbrace{\|\mathbf{y} - \mathbf{x}\|_2^2}_{\text{data fitting term}} + \underbrace{\psi(\mathbf{x})}_{\text{relation to image model}}$$

Some classical priors

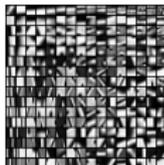
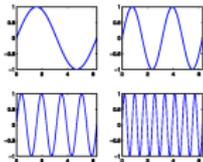
- Smoothness $\lambda \|\mathcal{L}\mathbf{x}\|_2^2$
- Total variation $\lambda \|\nabla\mathbf{x}\|_1^2$
- Wavelet sparsity $\lambda \|\mathbf{W}\mathbf{x}\|_1$
- ...

What is a Sparse Linear Model?

Let \mathbf{x} in \mathbb{R}^m be a signal.



Let $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p] \in \mathbb{R}^{m \times p}$ be a set of normalized “basis vectors”.



We call it **dictionary**.

\mathbf{D} is “adapted” to \mathbf{x} if it can represent it with a few basis vectors—that is, there exists a **sparse vector** α in \mathbb{R}^p such that $\mathbf{x} \approx \mathbf{D}\alpha$. We call α the **sparse code**.

$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x} \in \mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D} \in \mathbb{R}^{m \times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}_{\alpha \in \mathbb{R}^p, \text{ sparse}}$$

The Sparse Decomposition Problem

$$\min_{\alpha \in \mathbb{R}^p} \underbrace{\frac{1}{2} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\alpha)}_{\text{sparsity-inducing regularization}}$$

ψ induces sparsity in α . It can be

- the ℓ_0 “pseudo-norm”. $\|\alpha\|_0 \triangleq \#\{i \text{ s.t. } \alpha[i] \neq 0\}$ (NP-hard)
- the ℓ_1 norm. $\|\alpha\|_1 \triangleq \sum_{i=1}^p |\alpha[i]|$ (convex)
- ...

This is a **selection** problem.

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]

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

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

Sparse representations for image restoration

Solving the denoising problem

[Elad and Aharon, 2006]

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

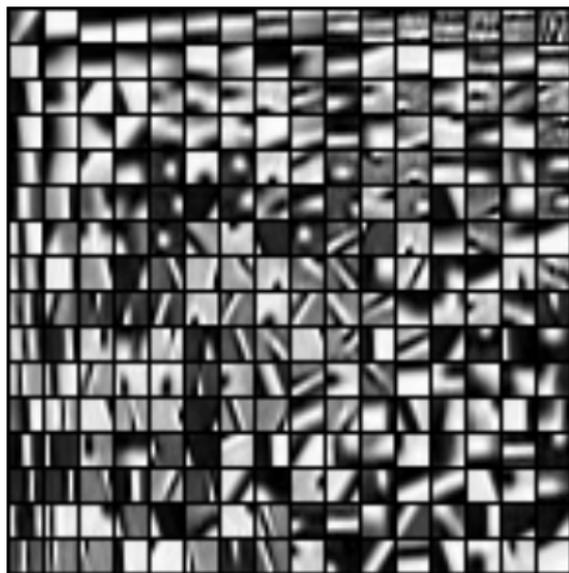
$$\min_{\alpha_i, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^n \underbrace{\frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\alpha_i\|_2^2}_{\text{reconstruction}} + \underbrace{\lambda\psi(\alpha_i)}_{\text{sparsity}},$$

with $n > 100,000$

- Average the reconstruction of each patch.

Sparse representations for image restoration

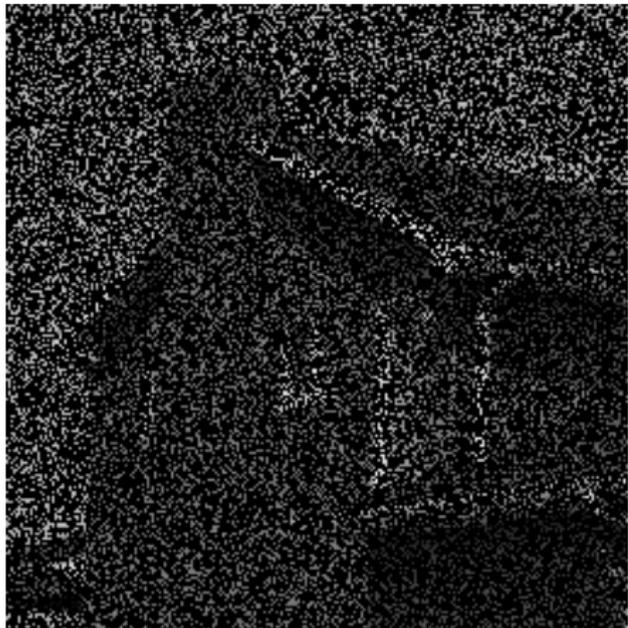
K-SVD: [Elad and Aharon, 2006]



Dictionary trained on a noisy version of the image boat.

Sparse representations for image restoration

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



Sparse representations for image restoration

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



Since 1699, when French explorers landed at the great bend of the Mississippi River and celebrated the first Mardi Gras in North America, New Orleans has brewed a fascinating melange of cultures. It was French, then Spanish, then French again, then sold to the United States. Through all these years, and even into the 1900s, others arrived from everywhere: Acadians (Cajuns), Africans, indige-

Sparse representations for image restoration

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



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

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

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

- Classical optimization alternates between \mathbf{D} and α .
- Good results, but **very slow!**

[Mairal et al., 2009a]: Online learning can

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

Optimization for Dictionary Learning

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

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

- Classical optimization alternates between \mathbf{D} and α .
- Good results, but **very slow!**

[Mairal et al., 2009a]: Online learning can

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

Try by yourself! <http://www.di.ens.fr/willow/SPAMS/>

Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph

THE SALINAS VALLEY is in Northern California. It is a long narrow swale between two ranges of mountains, and the Salinas River winds and twists up the center until it falls at last into Monterey Bay.

I remember my childhood games for grasses and secret flowers. I remember where a toad may live and what time the birds awaken in the summer and what trees and seasons smelled like-how people looked and walked and smelled even. The memory of odors is very rich.

I remember that the Gabilan Mountains to the east of the valley were light gay mountains full of sun and loveliness and a kind of invitation, so that you wanted to climb into their warm foothills almost as you want to climb into the lap of a beloved mother. They were beckoning mountains with a blown grass love. The Santa Lucia stood up against the sky to the west and kept the valley from the open sea, and they were dark and brooding unfriendly and dangerous. I always found in myself a dread of west and a love of east. Where I ever got such an idea I cannot say, unless it could be that the morning came over the peaks of the Gabilans and the night drifted back from the ridges of the Santa Lucias. It may be that the birth and death of the day had some part in my feeling about the two ranges of mountains.

From both sides of the valley little streams slipped out of the hill canyons and fell into the bed of the Salinas River. In the winter of wet years the streams ran full-freshet, and they swelled the river until sometimes it raged and boiled, bank full, and then it was a destroyer. The river tore the edges of the farm lands and washed whole acres down; it toppled barns and houses into itself, to go floating and bobbing away. It trapped cows and pigs and sheep and drowned them in its muddy brown water and carried them to the sea. Then when the late spring came, the river drew in from its edges and the sand banks appeared. And in the summer the river didn't run at all above ground. Some pools would be left in the deep swirl places under a high bank. The tules and grasses grew back, and willows straightened up with the flood debris in their upper branches. The Salinas was only a part-time river. The summer sun drove it underground. It was not a flat river at all, but it was the only one we had and so we boasted about it how dangerous it was in a wet winter and how dry it was in a dry summer. You can boast about anything if it's all you have. Maybe the less you have, the more you are required to boast.

The floor of the Salinas Valley, between the ranges and below the foothills, is level because this valley used to be the bottom of a hundred-mile inlet from the sea. The river mouth at Moss Landing was centuries ago the entrance to this long inland water. Once, fifty miles down the valley, my father bored a well. The drill came up first with topsoil and then with gravel and then with white sea sand full of shells and even pl...

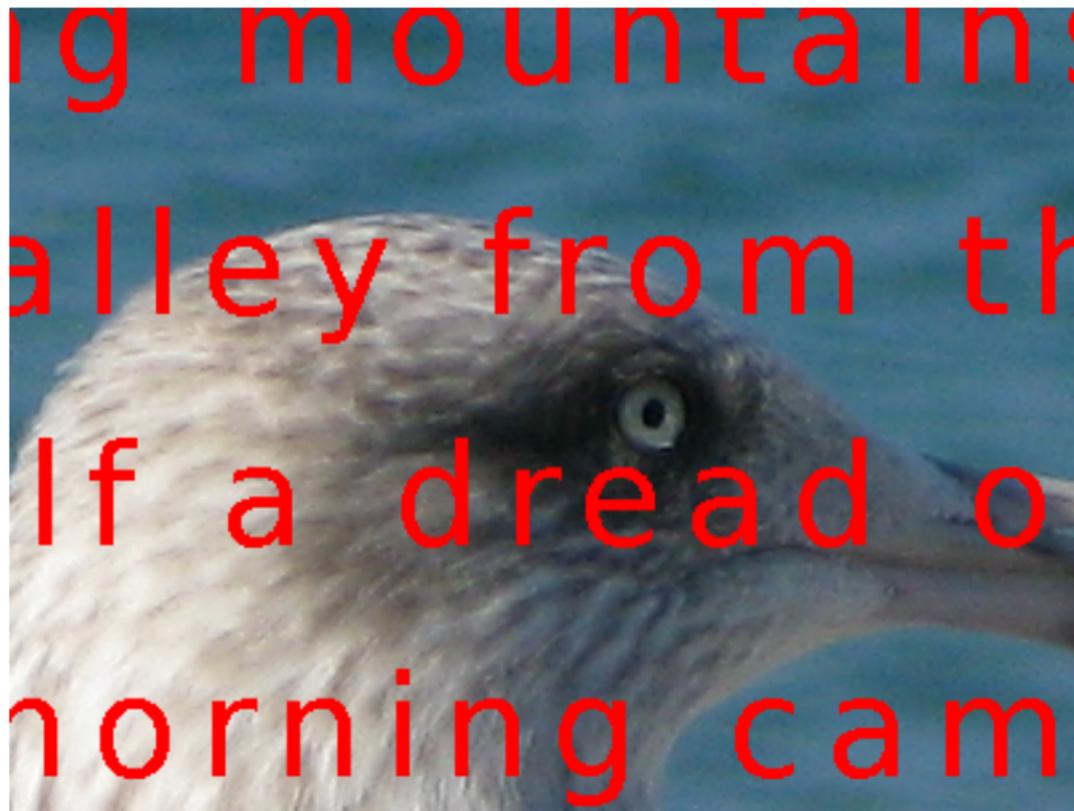
Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Optimization for Dictionary Learning

Inpainting a 12-Mpixel photograph



Exploiting Image Self-Similarities

Buades et al. [2006], Efros and Leung [1999], Dabov et al. [2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^n \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^n K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \quad (1)$$

Exploiting Image Self-Similarities

Buades et al. [2006], Efros and Leung [1999], Dabov et al. [2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^n \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^n K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \quad (1)$$

Successful application to texture synthesis: Efros and Leung [1999]

... to image denoising (**Non-Local Means**): Buades et al. [2006]

... to image demosaicking: Buades et al. [2009]

Exploiting Image Self-Similarities

Buades et al. [2006], Efros and Leung [1999], Dabov et al. [2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^n \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^n K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \quad (1)$$

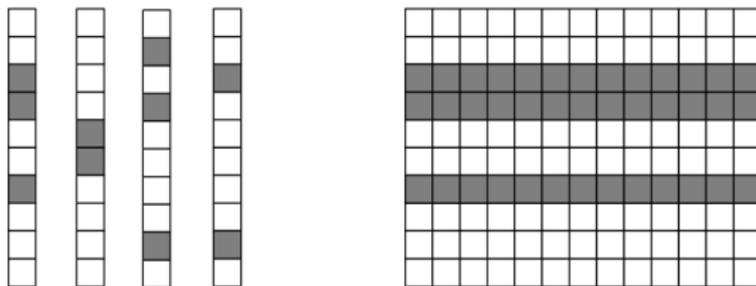
Successful application to texture synthesis: Efros and Leung [1999]
... to image denoising (**Non-Local Means**): Buades et al. [2006]
... to image demosaicking: Buades et al. [2009]

Block-Matching with 3D filtering (BM3D) Dabov et al. [2007],
Similar patches are **jointly** denoised with orthogonal wavelet thresholding
+ several (good) heuristics: \implies state-of-the-art denoising results, less artefacts, higher PSNR.

Non-local Sparse Image Models

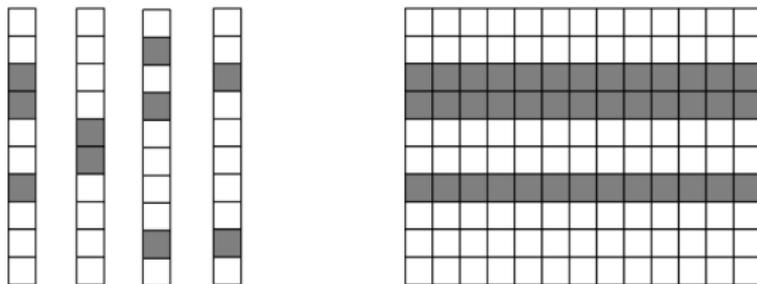
- **non-local means**: **stable** estimator. Can fail when there are no self-similarities.
- **sparse representations**: “unique” patches also admit a sparse approximation on the learned dictionary. potentially **unstable** decompositions.

Improving the stability of sparse decompositions is a current topic of research in statistics Bach [2008], Meinshausen and Buehlmann [2010].
Mairal et al. [2009b]: Similar patches should admit similar patterns:



Sparsity vs. joint sparsity

Non-local Sparse Image Models



Sparsity vs. joint sparsity

Joint sparsity is achieved through specific regularizers such as

$$\|\mathbf{A}\|_{0,\infty} \triangleq \sum_{i=1}^k \|\alpha^i\|_0, \quad (\text{not convex, not a norm})$$
$$\|\mathbf{A}\|_{1,2} \triangleq \sum_{i=1}^k \|\alpha^i\|_2. \quad (\text{convex norm})$$
(2)

Non-local Sparse Image Models

Basic scheme for image denoising:

- 1 Cluster patches

$$S_i \triangleq \{j = 1, \dots, n \text{ s.t. } \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq \xi\}, \quad (3)$$

- 2 Learn a dictionary with group-sparsity regularization

$$\min_{(\mathbf{A}_i)_{i=1}^n, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^n \frac{\|\mathbf{A}_i\|_{1,2}}{|S_i|} \text{ s.t. } \forall i \sum_{j \in S_i} \|\mathbf{y}_j - \mathbf{D}\alpha_{ij}\|_2^2 \leq \varepsilon_i \quad (4)$$

- 3 Estimate the final image by averaging the representations

Non-local Sparse Image Models

Basic scheme for image denoising:

- 1 Cluster patches

$$S_i \triangleq \{j = 1, \dots, n \text{ s.t. } \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq \xi\}, \quad (3)$$

- 2 Learn a dictionary with group-sparsity regularization

$$\min_{(\mathbf{A}_i)_{i=1}^n, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^n \frac{\|\mathbf{A}_i\|_{1,2}}{|S_i|} \text{ s.t. } \forall i \sum_{j \in S_i} \|\mathbf{y}_j - \mathbf{D}\alpha_{ij}\|_2^2 \leq \varepsilon_i \quad (4)$$

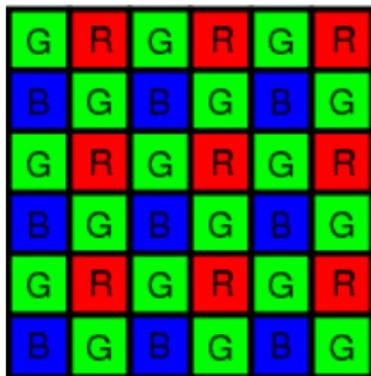
- 3 Estimate the final image by averaging the representations

Details:

- Greedy clustering (linear time) and online learning.
- Eventually use two passes.
- Use non-convex regularization for the final reconstruction.

Non-local Sparse Image Models

Demosaicking

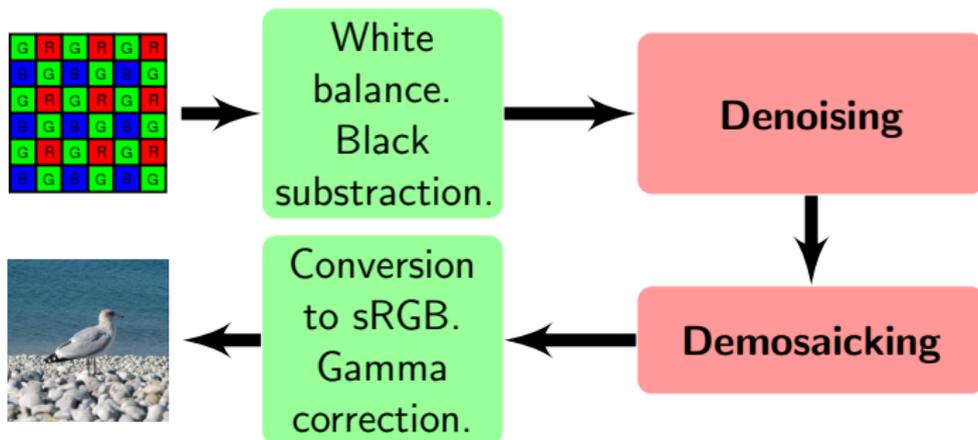


Key components for image demosaicking:

- 1 introduce a binary mask in the formulation.
- 2 Learn the dictionary on a database of clean images.
- 3 Eventually relearn the dictionary on a first estimate of the reconstructed image.

Non-local Sparse Image Models

RAW Image Processing



Since the dictionary **adapts** to the input data, this scheme is not limited to natural images!

Non-local Sparse Image Models

Denoising results, synthetic noise

Average PSNR on 10 standard images (higher is better)

σ	GSM	FOE	KSVD	BM3D	SC	LSC	LSSC
5	37.05	37.03	37.42	37.62	37.46	37.66	37.67
10	33.34	33.11	33.62	34.00	33.76	33.98	34.06
15	31.31	30.99	31.58	32.05	31.72	31.99	32.12
20	29.91	29.62	30.18	30.73	30.29	30.60	30.78
25	28.84	28.36	29.10	29.72	29.18	29.52	29.74
50	25.66	24.36	25.61	26.38	25.83	26.18	26.57
100	22.80	21.36	22.10	23.25	22.46	22.62	23.39

Improvement over BM3D is significant only for large values of σ . The comparison is made with GSM (Gaussian Scale Mixture) Portilla et al. [2003], FOE (Field of Experts) Roth and Black [2005], KSVD Elad and Aharon [2006] and BM3D Dabov et al. [2007].

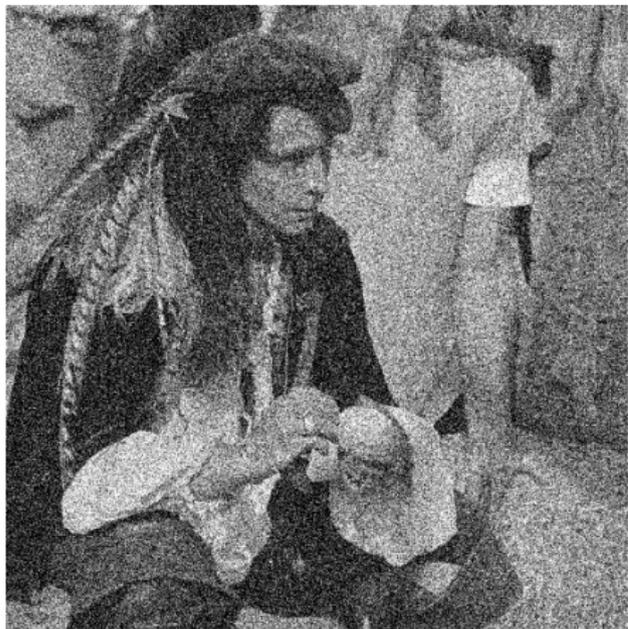
Non-local Sparse Image Models

Denoising results, synthetic noise



Non-local Sparse Image Models

Denoising results, synthetic noise



Non-local Sparse Image Models

Demosaicking results, Kodak database

Average PSNR on the Kodak dataset (24 images)

Im.	AP	DL	LPA	SC	LSC	LSSC
Av.	39.21	40.05	40.52	40.88	41.13	41.39

The comparison is made with AP (Alternative Projections) Gunturk et al. [2002], DL Zhang and Wu [2005] and LPA Paliy et al. [2007] (best known result on this database).

Non-local Sparse Image Models

Demosaicking results, Kodak database

More importantly than a PSNR improvement:



Regular sparsity on the left, Joint-sparsity on the right

Conclusion

- Clustering of patches stabilizes the decompositions and improves the results quality,
- and lead to state-of-the-art results for image denoising and demosaicking.

Conclusion

- Clustering of patches stabilizes the decompositions and improves the results quality,
- and lead to state-of-the-art results for image denoising and demosaicking.

Not the end of the story

- download the paper for preliminary raw image processing results.
- other applications coming (deblurring, superresolution)
- structured sparsity: Jenatton et al. [2009] ...
- task-driven dictionaries ...

Conclusion

- Clustering of patches stabilizes the decompositions and improves the results quality,
- and lead to state-of-the-art results for image denoising and demosaicking.

Not the end of the story

- download the paper for preliminary raw image processing results.
- other applications coming (deblurring, superresolution)
- structured sparsity: Jenatton et al. [2009] ...
- task-driven dictionaries ...

Tutorial on Sparse Coding available at

http://www.di.ens.fr/~mairal/tutorial_iccv09/

Conclusion

- Clustering of patches stabilizes the decompositions and improves the results quality,
- and lead to state-of-the-art results for image denoising and demosaicking.

Not the end of the story

- download the paper for preliminary raw image processing results.
- other applications coming (deblurring, superresolution)
- structured sparsity: Jenatton et al. [2009] ...
- task-driven dictionaries ...

Tutorial on Sparse Coding available at

http://www.di.ens.fr/~mairal/tutorial_iccv09/

Software for learning dictionaries with efficient sparse solvers

<http://www.di.ens.fr/willow/SPAMS/>. Image processing functions and group-sparsity solvers coming soon.

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.
- F. Bach. Bolasso: model consistent lasso estimation through the bootstrap. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.
- A. Buades, B. Coll, and J.M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modelling and Simulation*, 4(2):490–530, 2006.
- A. Buades, B. Coll, J.-M. Morel, and C Sbert. Self-similarity driven color demosaicking. *IEEE Transactions on Image Processing*, 18(6):1192–1202, 2009.
- K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.
- A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1999.

References II

- 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.
- BK Gunturk, Y. Altunbasak, and RM Mersereau. Color plane interpolation using alternating projections. *IEEE Transactions on Image Processing*, 11(9):997–1013, 2002.
- A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.
- R. Jenatton, J-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. Technical report, 2009. preprint arXiv:0904.3523v1.
- 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.

References III

- 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, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009b.
- S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, New York, September 1999.
- N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society, Series B*, 2010. to appear.
- 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.
- D. Paliy, V. Katkovnik, R. Bilcu, S. Alenius, and K. Egiazarian. Spatially adaptive color filter array interpolation for noiseless and noisy data. *Intern. J. of Imaging Sys. and Tech.*, 17(3), 2007.

References IV

- J. Portilla, V. Strela, MJ Wainwright, and EP Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.
- 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.
- L. Zhang and X. Wu. Color demosaicking via directional linear minimum mean square-error estimation. *IEEE Transactions on Image Processing*, 14(12): 2167–2178, 2005.