Non-local Sparse Models for Image Restoration

Julien Mairal\textsuperscript{1}  Francis Bach\textsuperscript{1}
Jean Ponce\textsuperscript{2}  Guillermo Sapiro\textsuperscript{3}  Andrew Zisserman\textsuperscript{4}

\textsuperscript{1}INRIA - WILLOW  \textsuperscript{2}Ecole Normale Supérieure
\textsuperscript{3}University of Minnesota  \textsuperscript{4}Oxford University

MIA’09, December 16\textsuperscript{th} 2009
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

\[ y = x_{\text{orig}} + \epsilon \]

- measurements
- original image
- noise
Sparse representations for image restoration

\[ y \text{ measurements} = x_{\text{orig}} + \varepsilon \text{ noise} \]

Energy minimization problem

\[ E(x) = \| y - x \|_2^2 + \psi(x) \]

Some classical regularizers

- Smoothness \( \lambda \| \mathcal{L} x \|_2^2 \)
- Total variation \( \lambda \| \nabla x \|_1 \)
- Wavelet sparsity \( \lambda \| W x \|_1 \)
- \( \ldots \)
What is a Sparse Linear Model?

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

Let \( \mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_p] \in \mathbb{R}^{m \times p} \) 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 \( \mathbf{\alpha} \) in \( \mathbb{R}^p \) such that \( \mathbf{x} \approx \mathbf{D} \mathbf{\alpha} \). We call \( \mathbf{\alpha} \) the sparse code.
The Sparse Decomposition Problem

\[
\min_{\alpha \in \mathbb{R}^p} \quad \frac{1}{2} \left\| x - D\alpha \right\|_2^2 + \lambda \psi(\alpha)
\]

**data fitting term**

sparsity-inducing regularization

\(\psi\) induces sparsity in \(\alpha\). 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)
- \ldots

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]

\[
\min_{\alpha_i, D \in \mathcal{C}} \sum_i \frac{1}{2} \| x_i - D \alpha_i \|_2^2 + \lambda \psi(\alpha_i)
\]

- \( \psi(\alpha) = \| \alpha \|_0 \) ("\( \ell_0 \) pseudo-norm")
- \( \psi(\alpha) = \| \alpha \|_1 \) (\( \ell_1 \) norm)
Sparse representations for image restoration

Addressing the denoising problem

[Elad and Aharon, 2006]

- Extract all overlapping $8 \times 8$ patches $y_i$.
- Solve a matrix factorization problem:

$$\min_{\alpha_i, D \in \mathcal{C}} \sum_{i=1}^{n} \frac{1}{2} ||y_i - D\alpha_i||_2^2 + \lambda \psi(\alpha_i),$$

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]
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

\[
\min_{\alpha \in \mathbb{R}^{p \times n}, D \in \mathcal{C}} \sum_{i=1}^{n} \frac{1}{2} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1
\]

\[\mathcal{C} \triangleq \{D \in \mathbb{R}^{m \times p} \text{ s.t.} \forall j = 1, \ldots, p, \|d_j\|_2 \leq 1\}.\]

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

\[\text{[Mairal et al., 2009a]: Online learning can handle potentially infinite or dynamic datasets, be much faster than batch algorithms. Try it yourself!}\]
Optimization for Dictionary Learning

\[
\min_{\alpha \in \mathbb{R}^{p \times n}} \sum_{i=1}^{n} \frac{1}{2} ||x_i - D \alpha_i||_2^2 + \lambda ||\alpha_i||_1
\]

\[C \triangleq \{D \in \mathbb{R}^{m \times p} \text{ s.t. } \forall j = 1, \ldots, p, \quad ||d_j||_2 \leq 1\}.
\]

- Classical optimization alternates between \(D\) and \(\alpha\).
- Good results, but very slow!

[Mairal et al., 2009a]: Online learning can
  - handle potentially infinite or dynamic datasets,
  - be much faster than batch algorithms.
Optimization for Dictionary Learning

\[
\min_{\alpha \in \mathbb{R}^{p \times n}, D \in \mathcal{C}} \sum_{i=1}^{n} \frac{1}{2} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1
\]

\[\mathcal{C} \triangleq \{D \in \mathbb{R}^{m \times p} \text{ s.t. } \forall j = 1, \ldots, p, \|d_j\|_2 \leq 1\}.$

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

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

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

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

The Salinas Valley is in Northern California. It is in 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 names for grasses and secret flowers. I remember where a toad may live and what time the birds ceased 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 Salinas Mountains at the east of the valley were light gray mountains full of sun and loneliness 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 brown grass base. The Santa Lucias 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 sense that 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 Salinans and the mist filtered 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 fall into the bed of the Salinas River. In the winter of wet years the streams ran full, fresh and clear, 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 green water and carried them to the sea. Then when the late spring came, the river drew up from its banks and the sand banks appeared. And in the summer the river didn't run as an able ground, some pools would be left in the deep swale places under a high bank. The pools and grasses grew thick, 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 flood 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 and shells and even pebbles...
Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph
Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

If a dread of morning came
big mountains
alley from the
Mairal, Bach, Ponce, Sapiro & Zisserman

Non-local Sparse Models for Image Restoration
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:

\[
\hat{x}[i] = \sum_{j=1}^{n} \frac{K_h(y_i - y_j)}{\sum_{l=1}^{n} K_h(y_i - y_l)} y[j],
\]  

(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:

\[
\hat{x}[i] = \sum_{j=1}^{n} \frac{K_h(y_i - y_j)}{\sum_{i=1}^{n} K_h(y_i - y_i)} y[j],
\]

(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:

\[ \hat{x}[i] = \sum_{j=1}^{n} \frac{K_h(y_i - y_j)}{\sum_{l=1}^{n} K_h(y_i - y_l)} y[j], \]  

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

\textbf{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**:

```
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Sparsity vs. joint sparsity**
Non-local Sparse Image Models

Sparsity vs. joint sparsity

Joint sparsity is achieved through specific regularizers such as

\[
\|A\|_{0,\infty} \triangleq \sum_{i=1}^{k} \|\alpha_i\|_0, \quad \text{(not convex, not a norm)}
\]

(2)

\[
\|A\|_{1,2} \triangleq \sum_{i=1}^{k} \|\alpha_i\|_2. \quad \text{(convex norm)}
\]
Non-local Sparse Image Models

Basic scheme for image denoising:

1. Cluster patches

\[ S_i \triangleq \{ j = 1, \ldots, n \text{ s.t. } \| y_i - y_j \|_2^2 \leq \xi \} , \]  

2. Learn a dictionary with group-sparsity regularization

\[
\begin{align*}
\min_{(A_i)^n_{i=1}, D \in C} & \sum_{i=1}^{n} \frac{\| A_i \|_{1,2}}{|S_i|} \\
\text{s.t.} & \forall i \sum_{j \in S_i} \| y_j - D \alpha_{ij} \|_2^2 \leq \epsilon_i
\end{align*}
\]  

3. Estimate the final image by averaging the representations
Basic scheme for image denoising:

1. Cluster patches

\[ S_i \triangleq \{ j = 1, \ldots, n \text{ s.t. } \| y_i - y_j \|_2^2 \leq \xi \} \]  

2. Learn a dictionary with group-sparsity regularization

\[
\min_{(A_i)_{i=1}^n, D \in \mathbb{C}} \sum_{i=1}^n \| A_i \|_{1,2} \quad \text{s.t.} \quad \forall i \sum_{j \in S_i} \| y_j - D \alpha_{ij} \|_2^2 \leq \epsilon_i
\]

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.
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)**

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>GSM</th>
<th>FOE</th>
<th>KSVD</th>
<th>BM3D</th>
<th>SC</th>
<th>LSC</th>
<th>LSSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>37.05</td>
<td>37.03</td>
<td>37.42</td>
<td>37.62</td>
<td>37.46</td>
<td>37.66</td>
<td>37.67</td>
</tr>
<tr>
<td>10</td>
<td>33.34</td>
<td>33.11</td>
<td>33.62</td>
<td>34.00</td>
<td>33.76</td>
<td>33.98</td>
<td>34.06</td>
</tr>
<tr>
<td>15</td>
<td>31.31</td>
<td>30.99</td>
<td>31.58</td>
<td>32.05</td>
<td>31.72</td>
<td>31.99</td>
<td>32.12</td>
</tr>
<tr>
<td>20</td>
<td>29.91</td>
<td>29.62</td>
<td>30.18</td>
<td>30.73</td>
<td>30.29</td>
<td>30.60</td>
<td>30.78</td>
</tr>
<tr>
<td>25</td>
<td>28.84</td>
<td>28.36</td>
<td>29.10</td>
<td>29.72</td>
<td>29.18</td>
<td>29.52</td>
<td>29.74</td>
</tr>
<tr>
<td>100</td>
<td>22.80</td>
<td>21.36</td>
<td>22.10</td>
<td>23.25</td>
<td>22.46</td>
<td>22.62</td>
<td>23.39</td>
</tr>
</tbody>
</table>

Improvement over BM3D is significant only for large values of $\sigma$. 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

Restored
Non-local Sparse Image Models
Demosaicking results, Kodak database

**Average PSNR on the Kodak dataset (24 images)**

<table>
<thead>
<tr>
<th>Im.</th>
<th>AP</th>
<th>DL</th>
<th>LPA</th>
<th>SC</th>
<th>LSC</th>
<th>LSSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av.</td>
<td>39.21</td>
<td>40.05</td>
<td>40.52</td>
<td>40.88</td>
<td>41.13</td>
<td><strong>41.39</strong></td>
</tr>
</tbody>
</table>

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 quality of results,
- and lead to state-of-the-art results for image denoising and demosaicking.
Conclusion

- Clustering of patches stabilizes the decompositions and improves the quality of results,
- 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 quality of results,
- 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 quality of results,
- 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
References I


References III


References IV

