Sparse Coding
for
Image and Video Understanding

Jean Ponce
http://www.di.ens.fr/willow/
Willow team, DI/ENS, UMR 8548
Ecole normale supérieure, Paris
Sparse Coding for Image and Video Understanding

Julien Mairal and Francis Bach
Sparse linear models

Signal: $\mathbf{x} \in \mathbb{R}^m$

Dictionary:
$\mathcal{D} = [d_1, \ldots, d_p] \in \mathbb{R}^{m \times p}$

$x \approx \alpha_1 d_1 + \alpha_2 d_2 + \ldots + \alpha_p d_p = \mathcal{D} \alpha$, with $|\alpha|_0 \ll p$

(Olshausen and Field, 1997; Chen et al., 1999; Mallat, 1999; Elad and Aharon, 2006) (Kavukcuoglu et al., 2009; Wright et al., 2009; Yang et al., 09; Boureau et al., 2010)
Sparse coding and dictionary learning: A hierarchy of optimization problems

\[
\min_{\alpha} \frac{1}{2} |x - Da|_2^2
\]

Least squares

\[
\min_{\alpha} \frac{1}{2} |x - Da|_2^2 + \lambda |\alpha|_0
\]

Sparse coding

\[
\min_{\alpha} \frac{1}{2} |x - Da|_2^2 + \lambda \psi(\alpha)
\]

Dictionary learning

\[
\min_{D \in \mathbb{C}, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} \left[ \frac{1}{2} |x_i - Da_i|_2^2 + \lambda \psi(\alpha_i) \right]
\]

Learning for a task

\[
\min_{D \in \mathbb{C}, W, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} \left[ f(x_i, D, W, \alpha_i) + \lambda \psi(\alpha_i) \right]
\]

Learning structures

\[
\min_{D \in \mathbb{C}, W, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} \left[ f(x_i, D, W, \alpha_i) + \lambda \sum_{1 \leq k \leq q} \psi(d_k) \right]
\]
Video inpainting

(Mairal, Sapiro and Elad, 2008)
Video inpainting

(Mairal, Sapiro and Elad, 2008)
Video inpainting

(Mairal, Sapiro and Elad, 2008)
Video inpainting

(Mairal, Sapiro and Elad, 2008)
Video inpainting

(Mairal, Sapiro and Elad, 2008)
Video denoising

(Mairal, Sapiro and Elad, 2008)
Video denoising

(Mairal, Sapiro and Elad, 2008)
Video denoising

(Mairal, Sapiro and Elad, 2008)
Video denoising

(Mairal, Sapiro and Elad, 2008)
Video denoising

(Mairal, Sapiro and Elad, 2008)
Important messages

• Patch-based approaches achieve state-of-the-art results for many image processing tasks.

• A dictionary can be learned on the data of interest itself.

• Sparse coding is well adapted to data that admit sparse representations.

• Sparse coding is only adapted to those.

• It is not compressed sensing (Candes’06).
Outline

• Sparse linear models of image data
• Unsupervised dictionary learning
• Non-local sparse models for image restoration
• Learning discriminative dictionaries for image classification
• Task-driven dictionary learning and its applications
• Ongoing work
Sparse coding

- The $l_0$ version:
  \[ \min_{\alpha} \frac{1}{2} \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_0 \]
  NP-hard, greedy approximate algorithms

- The $l_1$ version:
  \[ \min_{\alpha} \frac{1}{2} \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_1 \]
  convex, exact algorithms
Finding your way in the sparse coding literature is not easy. The literature is vast, redundant, sometimes confusing and many papers are claiming victory.

The main classes of methods are:

- greedy procedures [Mallat and Zhang, 1993], [Weisberg, 1980],
- homotopy techniques [Osborne et al., 2000], [Efron et al., 2004], [Markowitz, 1956],
- soft-thresholding-based methods [Fu, 1998], [Daubechies et al., 2004], [Friedman et al., 2007], [Nesterov, 2007], [Beck and Teboulle, 2009],
- reweighted-\ell_2 procedures [Daubechies et al., 2009],
- active-set methods [Roth and Fischer, 2008].
Matching Pursuit

\[ \alpha = (0, 0, 0) \]
Matching Pursuit

\[ \alpha = (0, 0, 0) \]
Matching Pursuit

\[ \alpha = (0, 0, 0) \]
Matching Pursuit

$\alpha = (0, 0, 0.75)$
Matching Pursuit

\[ \alpha = (0, 0, 0.75) \]
Matching Pursuit

\[ \alpha = (0, 0, 0.75) \]
Matching Pursuit

\[ \alpha = (0, 0.24, 0.75) \]
Matching Pursuit

\[ \alpha = (0, 0.24, 0.75) \]
Matching Pursuit

\[ \alpha = (0, 0.24, 0.65) \]
The $l_1$ norm and sparsity
LARS (Efron et al., 2004)
Dictionary learning

- Given some loss function, e.g.,
  \[ L(x, D) = \min_{\alpha} \frac{1}{2} |x - D\alpha|^2 + \lambda |\alpha|_1 \]

- One usually minimizes, given some data \(x_i, i = 1, ..., n\), the empirical risk:
  \[ \min_D f_n(D) = \frac{1}{n} \sum_{1 \leq i \leq n} L(x_i, D) \]

- But, one would really like to minimize the expected one, that is:
  \[ \min_D f(D) = \mathbb{E}_x [ L(x, D) ] \]

(Bottou & Bousquet'08 → Stochastic gradient descent)
Online sparse matrix factorization
(Mairal, Bach, Ponce, Sapiro, ICML’09, JMLR’10)

Problem:
\[
\min_D f(D) = E_x [ L(x, D) ]
\]
\[
\min_{D \in \mathcal{C}, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} [ 1/2 | x_i - D\alpha_i |_2^2 + \lambda |\alpha_i|_1 ]
\]

Algorithm:
Iteratively draw one random training sample \(x_t\) and minimize the quadratic surrogate function:
\[
g_t(D) = 1/t \sum_{1 \leq i \leq t} [ 1/2 | x_i - D\alpha_i |_2^2 + \lambda |\alpha_i|_1 ]
\]

(Lars/Lasso for sparse coding, block-coordinate descent with warm restarts for dictionary updates, mini-batch extensions, etc.)
Online sparse matrix factorization  
(Mairal, Bach, Ponce, Sapiro, ICML’09, JMLR’10)

Problem:  
\[ \min_{D} f(D) = E_{x}[L(x, D)] \]

\[ \min_{D \in C, A} \left[ \frac{1}{2} \| X - DA \|_{F}^{2} + \lambda \| A \|_{1} \right] \]

Algorithm:  
Iteratively draw one random training sample \( x_t \) and minimize the quadratic surrogate function:  
\[ g_t(D) = \frac{1}{t} \sum_{1 \leq i \leq t} \left[ \frac{1}{2} \| x_i - D \alpha_i \|_{2}^{2} + \lambda \| \alpha_i \|_{1} \right] \]

(Lars/Lasso for sparse coding, block-coordinate descent with warm restarts for dictionary updates, mini-batch extensions, etc.)
Online sparse matrix factorization
(Mairal, Bach, Ponce, Sapiro, ICML’09, JMLR’10)

Proposition:
Under mild assumptions, $D_t$ converges with probability one to a stationary point of the dictionary learning problem.

Proof: Convergence of empirical processes (van der Vaart’98) and, a la (Bottou’98), convergence of quasi martingales (Fisk’65).

Extensions:
• Non negative matrix factorization (Lee & Seung’01)
• Non negative sparse coding (Hoyer’02)
• Sparse principal component analysis (Jolliffe et al.’03; Zou et al.’06; Zass & Shashua’07; d’Aspremont et al.’08; Witten et al.’09)
Performance evaluation

Three datasets constructed from 1,250,000 Pascal'06 patches (1,000,000 for training, 250,000 for testing):
- **A**: 8×8 b&w patches, 256 atoms.
- **B**: 12×16×3 color patches, 512 atoms.
- **C**: 16×16 b&w patches, 1024 atoms.

Two variants of our algorithm:
- Online version with different choices of parameters.
- Batch version on different subsets of training data.

**Online vs batch**

**Online vs stochastic gradient descent**
Sparse PCA: Adding sparsity on the atoms

Three datasets:
- D: 2429 19×19 images from MIT-CBCL #1.
- E: 2414 192×168 images from extended Yale B.
- F: 100,000 16×16 patches from Pascal VOC'06.

Three implementations:
- Hoyer's Matlab implementation of NNMF (Lee & Seung'01).
- Hoyer's Matlab implementation of NNSC (Hoyer'02).
- Our C++/Matlab implementation of SPCA (elastic net on D).

SPCA vs NNMF

SPCA vs NNSC
Inpainting a 12MP image with a dictionary learned from 7x10^6 patches in 500s (Mailal et al., 2009)
State of the art in image denoising

Dictionary learning for denoising (Elad & Aharon’06; Mairal, Elad & Sapiro’08)

\[
\min_{D \in \mathbb{C}, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} \left[ \frac{1}{2} \| x_i - D \alpha_i \|_2^2 + \lambda \| \alpha_i \|_1 \right]
\]

\[
x = \frac{1}{n} \sum_{1 \leq i \leq n} R_i D \alpha_i
\]
State of the art in image denoising

Dictionary learning for denoising (Elad & Aharon’06; Mairal, Elad & Sapiro’08)

$$\min_{D \in C, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} \left[ \frac{1}{2} | x_i - D \alpha_i |_2^2 + \lambda | \alpha_i |_1 \right]$$

$$x = \frac{1}{n} \sum_{1 \leq i \leq n} R_i D \alpha_i$$

BM3D (Dabov et al.’07)

Non-local means filtering (Buades et al.’05)
Non-local sparse models for image restoration (Mairal, Bach, Ponce, Sapiro, Zisserman, ICCV’09)

\[
\min_{D \in \mathcal{C}, A_1, \ldots, A_n} \sum_i \left[ \sum_{j \in S_i} \frac{1}{2} \| x_j - D \alpha_{ij} \|_F^2 \right] + \lambda \| A_i \|_{p,q}
\]

\[
\| A \|_{p,q} = \sum_{1 \leq i \leq k} \| \alpha^i \|_q^p \quad (p,q) = (1,2) \text{ or } (0,\infty)
\]
PSNR comparison between our method (LSSC) and Portilla et al.'03 [23]; Roth & Black'05 [25]; Elad& Aharon'06 [12]; and Dabov et al.'07 [8].

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>[23]</th>
<th>[25]</th>
<th>[12]</th>
<th>[8]</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>
Demosaicking experiments

<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>1</td>
<td>37.84</td>
<td>38.46</td>
<td>40.47</td>
<td>40.84</td>
<td>40.92</td>
<td>41.36</td>
</tr>
<tr>
<td>2</td>
<td>39.64</td>
<td>40.89</td>
<td>41.36</td>
<td>41.76</td>
<td>42.03</td>
<td>42.24</td>
</tr>
<tr>
<td>3</td>
<td>41.40</td>
<td>42.66</td>
<td>43.47</td>
<td>43.15</td>
<td>43.92</td>
<td>44.24</td>
</tr>
<tr>
<td>23</td>
<td>41.93</td>
<td>43.22</td>
<td>43.92</td>
<td>43.47</td>
<td>43.93</td>
<td>44.34</td>
</tr>
<tr>
<td>24</td>
<td>34.74</td>
<td>35.55</td>
<td>35.44</td>
<td>35.59</td>
<td>35.85</td>
<td>35.89</td>
</tr>
<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>41.39</td>
</tr>
</tbody>
</table>

PSNR comparison between our method (LSSC) and Gunturk et al.’02 [AP]; Zhang & Wu’05 [DL]; and Paliy et al.’07 [LPA] on the Kodak PhotoCD data.
Real noise (Canon Powershot G9, 1600 ISO)
Learning discriminative dictionaries with $l_0$ constraints

(Mairal, Bach, Ponce, Sapiro, Zisserman, CVPR’08)

\[
\alpha^*(x,D) = \text{Argmin}_{\alpha} | x - D\alpha |_2^2 \quad \text{s.t.} \quad |\alpha|_0 \leq L
\]

\[
R^*(x,D) = | x - D\alpha^* |_2^2
\]


\[
\min_{D} \sum_{l} R^*(x_l,D)
\]

Discriminative approach:

\[
\min_{D_1,...,D_n} \sum_{i,l} C_i^\lambda [R^*(x_l,D_1),...,R^*(x_l,D_n)] + \lambda \gamma R^*(x_l,D_i)
\]

(Both MOD and K-SVD versions with truncated Newton iterations.)
Texture classification results

<table>
<thead>
<tr>
<th>j</th>
<th>[28]</th>
<th>[17]</th>
<th>[34]</th>
<th>[16]</th>
<th>R1</th>
<th>R2</th>
<th>D1</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.2</td>
<td>6.7</td>
<td>5.5</td>
<td>3.37</td>
<td>2.22</td>
<td>1.69</td>
<td>1.89</td>
<td>1.61</td>
</tr>
<tr>
<td>2</td>
<td>18.9</td>
<td>14.3</td>
<td><strong>7.3</strong></td>
<td>16.05</td>
<td>24.66</td>
<td>36.5</td>
<td>16.38</td>
<td>16.42</td>
</tr>
<tr>
<td>3</td>
<td>20.6</td>
<td>10.2</td>
<td>13.2</td>
<td>13.03</td>
<td>10.20</td>
<td>5.49</td>
<td>9.11</td>
<td>4.15</td>
</tr>
<tr>
<td>4</td>
<td>16.8</td>
<td>9.1</td>
<td>5.6</td>
<td>6.62</td>
<td>6.66</td>
<td>4.60</td>
<td>3.79</td>
<td>3.67</td>
</tr>
<tr>
<td>5</td>
<td>17.2</td>
<td>8.0</td>
<td>10.5</td>
<td>8.15</td>
<td>5.26</td>
<td><strong>4.32</strong></td>
<td>5.10</td>
<td>4.58</td>
</tr>
<tr>
<td>6</td>
<td>34.7</td>
<td>15.3</td>
<td>17.1</td>
<td>18.66</td>
<td>16.88</td>
<td>15.50</td>
<td>12.91</td>
<td>9.04</td>
</tr>
<tr>
<td>7</td>
<td>41.7</td>
<td>20.7</td>
<td>17.2</td>
<td>21.67</td>
<td>19.32</td>
<td>21.89</td>
<td>11.44</td>
<td>8.80</td>
</tr>
<tr>
<td>8</td>
<td>32.3</td>
<td>18.1</td>
<td>18.9</td>
<td>21.96</td>
<td>13.27</td>
<td>11.80</td>
<td>14.77</td>
<td>2.24</td>
</tr>
<tr>
<td>9</td>
<td>27.8</td>
<td>21.4</td>
<td>21.4</td>
<td>9.61</td>
<td>18.85</td>
<td>21.88</td>
<td>10.12</td>
<td>2.04</td>
</tr>
<tr>
<td>10</td>
<td>0.7</td>
<td>0.4</td>
<td>NA</td>
<td>0.36</td>
<td>0.35</td>
<td><strong>0.17</strong></td>
<td>0.20</td>
<td><strong>0.17</strong></td>
</tr>
<tr>
<td>11</td>
<td>0.2</td>
<td>0.8</td>
<td>NA</td>
<td>1.33</td>
<td>0.58</td>
<td>0.73</td>
<td>0.41</td>
<td>0.60</td>
</tr>
<tr>
<td>12</td>
<td>2.5</td>
<td>5.3</td>
<td>NA</td>
<td>1.14</td>
<td>1.36</td>
<td><strong>0.37</strong></td>
<td>1.97</td>
<td>0.78</td>
</tr>
<tr>
<td>Av</td>
<td>18.4</td>
<td>10.9</td>
<td>NA</td>
<td>10.16</td>
<td>9.97</td>
<td>10.41</td>
<td>7.34</td>
<td><strong>4.50</strong></td>
</tr>
</tbody>
</table>
Pixel-level classification results

Qualitative results, Graz 02 data


Quantitative results
Reconstructive vs discriminative dictionaries
Learning discriminative dictionaries with $l_1$ constraints

$\alpha^*(x,D) = \text{Argmin}_{\alpha} |x - D\alpha|_2^2 \text{ s.t. } |\alpha|_1 \leq L$

$Lasso$: Convex optimization

(LARS: Efron et al.'04)

$R^*(x,D) = |x - D\alpha^*|_2^2$

Reconstruction (Lee, Battle, Rajat, Ng'07):

$\min_D \sum_l R^*(x_l,D)$

Discriminative approach:

$\min_{D_1,\ldots,D_n} \sum_{i,l} C_i^\lambda [R^*(x_l,D_1),\ldots,R^*(x_l,D_n)] + \lambda \gamma R^*(x_l,D_i)$

(Partial dictionary update with Newton iterations on the dual problem; partial fast sparse coding with projected gradient descent.)
Patch classification with learned dictionaries

Signal input -> Subsampling -> Sparse coding -> Classifier 1
Classifier 2 -> Linear classifier
Classifier 3
Edge detection results

Quantitative results on the Berkeley segmentation dataset and benchmark (Martin et al., ICCV'01)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Score</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.79</td>
<td>Human labeling</td>
</tr>
<tr>
<td>1</td>
<td>0.70</td>
<td>(Maire et al., 2008)</td>
</tr>
<tr>
<td>2</td>
<td>0.67</td>
<td>(Aerbelaez, 2006)</td>
</tr>
<tr>
<td>3</td>
<td>0.66</td>
<td>(Dollar et al., 2006)</td>
</tr>
<tr>
<td>3</td>
<td>0.66</td>
<td>Us – no post-processing</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>(Martin et al., 2001)</td>
</tr>
<tr>
<td>5</td>
<td>0.57</td>
<td>Color gradient</td>
</tr>
<tr>
<td>6</td>
<td>0.43</td>
<td>Random</td>
</tr>
<tr>
<td>Category</td>
<td>Us + L'07</td>
<td>L'07</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>Aeroplane</td>
<td>71.9%</td>
<td>61.9%</td>
</tr>
<tr>
<td>Boat</td>
<td>67.1%</td>
<td>56.4%</td>
</tr>
<tr>
<td>Cat</td>
<td>82.6%</td>
<td>53.4%</td>
</tr>
<tr>
<td>Cow</td>
<td>68.7%</td>
<td>59.22%</td>
</tr>
<tr>
<td>Horse</td>
<td>76.0%</td>
<td>67%</td>
</tr>
<tr>
<td>Motorbike</td>
<td>80.6%</td>
<td>73.6%</td>
</tr>
<tr>
<td>Sheep</td>
<td>72.9%</td>
<td>58.4%</td>
</tr>
<tr>
<td>Tvmonitor</td>
<td>87.7%</td>
<td>83.8%</td>
</tr>
</tbody>
</table>

Comparaison with Leordeanu et al. (2007) on Pascal’07 benchmark. Mean error rate reduction: 33%.
Task-driven dictionary learning
(Mairal, Bach, Ponce, PAMI'12)

\[
\min_{W,D} f(W,D) = E_{x,y} [L(y, W, \alpha^*(x, D))] + \nu |W|_F^2
\]

with \(\alpha^*(x,D) = \text{Argmin}_{\alpha} |x - D\alpha|_2^2 + \lambda |\alpha|_1 + \mu |\alpha|_2^2\)

(Mairal et al.'08; Bradley & Bagnell’09; Boureau et al.’10; Yang et al.’10)

- **Applications:** Regression, classification.
- **Extensions:** Learning linear transforms of the input data, semi-supervised learning.
- **Proposition:** Under mild assumptions, \(f\) is differentiable, and its gradient can be written in closed form as an expectation.
- **Algorithm:** Stochastic gradient descent.
Data courtesy of James Hughes & Daniel Rockmore

(Mairal, Bach, Ponce, 2010)
Fake Data courtesy of James Hughes & Daniel Rockmore
Authentic

Fake

Data courtesy of James Hughes & Daniel Rockmore
A common architecture for image classification

Filtering
- SIFT at keypoints

Coding
- vector quantization

Pooling
- whole image, mean

Deep layers
- dense gradients
- vector quantization
- spatial pyramid, max

Idea: Replace k-means by sparse coding (Yang et al., CVPR’09; Boureau et al., CVPR’10, ICML’10; Yang et al., CVPR’10).
# Learning dictionaries for image classification

(Boureau, LeCun, Bach, Ponce, CVPR’10)

<table>
<thead>
<tr>
<th>Method</th>
<th>Caltech-101, 30 training examples</th>
<th>15 Scenes, 100 training examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Pool</td>
<td>Max Pool</td>
</tr>
<tr>
<td></td>
<td>73.9 ± 0.9 [1024]</td>
<td>80.1 ± 0.6 [1024]</td>
</tr>
<tr>
<td>Hard quantization, linear kernel</td>
<td>51.4 ± 0.9 [256]</td>
<td>64.3 ± 0.9 [256]</td>
</tr>
<tr>
<td>Hard quantization, intersection kernel</td>
<td>64.2 ± 1.0 [256] (1)</td>
<td>64.3 ± 0.9 [256]</td>
</tr>
<tr>
<td>Soft quantization, linear kernel</td>
<td>57.9 ± 1.5 [1024]</td>
<td>69.0 ± 0.8 [256]</td>
</tr>
<tr>
<td>Soft quantization, intersection kernel</td>
<td>66.1 ± 1.2 [512] (2)</td>
<td>70.6 ± 1.0 [1024]</td>
</tr>
<tr>
<td>Sparse codes, linear kernel</td>
<td>61.3 ± 1.3 [1024]</td>
<td>71.5 ± 1.1 [1024] (3)</td>
</tr>
<tr>
<td>Sparse codes, intersection kernel</td>
<td>70.3 ± 1.3 [1024]</td>
<td>71.8 ± 1.0 [1024] (4)</td>
</tr>
</tbody>
</table>

Results with basic features, SIFT extracted each 8 pixels

<table>
<thead>
<tr>
<th>Single - feature</th>
<th>Method</th>
<th>Caltech 15 tr.</th>
<th>Caltech 30 tr.</th>
<th>Scenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boiman et al. [3]</td>
<td>Nearest neighbor + spatial correspondence</td>
<td>65.0 ± 1.1</td>
<td>70.4</td>
<td>-</td>
</tr>
<tr>
<td>Jain et al. [9]</td>
<td>Fast image search for learned metrics</td>
<td>61.0</td>
<td>69.6</td>
<td>-</td>
</tr>
<tr>
<td>Lazebnik et al. [12]</td>
<td>(1) SP + hard quantization + kernel SVM</td>
<td>56.4</td>
<td>64.4 ± 0.8</td>
<td>81.4 ± 0.5</td>
</tr>
<tr>
<td>van Gemert et al. [27]</td>
<td>(2) SP + soft quantization + kernel SVM</td>
<td>–</td>
<td>64.1 ± 1.2</td>
<td>76.7 ± 0.4</td>
</tr>
<tr>
<td>Yang et al. [31]</td>
<td>(3) SP + sparse codes + max pooling + linear SVM</td>
<td>67.0 ± 0.5</td>
<td>73.2 ± 0.5</td>
<td>80.3 ± 0.9</td>
</tr>
<tr>
<td>Yang et al. [31]</td>
<td>(4) SP + sparse codes + max pooling + kernel SVM</td>
<td>60.4 ± 1.0</td>
<td>–</td>
<td>77.7 ± 0.7</td>
</tr>
<tr>
<td>Zhang et al. [32]</td>
<td>kNN-SVM</td>
<td>59.1 ± 0.6</td>
<td>66.2 ± 0.5</td>
<td>-</td>
</tr>
<tr>
<td>Zhou et al. [33]</td>
<td>SP + Gaussian mixture</td>
<td>–</td>
<td>–</td>
<td>84.1 ± 0.5</td>
</tr>
</tbody>
</table>

Scenes, supervised dictionary learning

<table>
<thead>
<tr>
<th>Method</th>
<th>Unsup</th>
<th>Discr[1024]</th>
<th>Unsup</th>
<th>Discr[2048]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>83.6 ± 0.4</td>
<td>84.9 ± 0.3</td>
<td>84.2 ± 0.3</td>
<td>85.6 ± 0.2</td>
</tr>
<tr>
<td>Intersect</td>
<td>84.3 ± 0.5</td>
<td>84.7 ± 0.4</td>
<td>84.6 ± 0.4</td>
<td>85.1 ± 0.5</td>
</tr>
</tbody>
</table>
Learning dictionaries for image classification (Boureau, LeCun, Bach, Ponce, CVPR’10)

<table>
<thead>
<tr>
<th>Method</th>
<th>Caltech-101, 30 training examples</th>
<th>15 Scenes, 100 training examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Pool</td>
<td>Max Pool</td>
</tr>
<tr>
<td></td>
<td>51.4 ± 0.9 [256]</td>
<td>73.9 ± 0.9 [1024]</td>
</tr>
<tr>
<td>Hard quantization, linear kernel</td>
<td>64.2 ± 1.0 [256] (1)</td>
<td>80.1 ± 0.6 [1024]</td>
</tr>
<tr>
<td>Hard quantization, intersection kernel</td>
<td>64.3 ± 0.9 [256]</td>
<td>80.8 ± 0.4 [256] (1)</td>
</tr>
<tr>
<td></td>
<td>80.1 ± 0.6 [1024]</td>
<td></td>
</tr>
<tr>
<td>Soft quantization, linear kernel</td>
<td>57.9 ± 1.5 [1024]</td>
<td>75.6 ± 0.5 [1024]</td>
</tr>
<tr>
<td>Soft quantization, intersection kernel</td>
<td>69.0 ± 0.8 [256]</td>
<td>81.4 ± 0.6 [1024]</td>
</tr>
<tr>
<td></td>
<td>81.2 ± 0.4 [1024] (2)</td>
<td>83.0 ± 0.7 [1024]</td>
</tr>
<tr>
<td>Sparse codes, linear kernel</td>
<td>61.3 ± 1.3 [1024]</td>
<td>76.9 ± 0.6 [1024]</td>
</tr>
<tr>
<td>Sparse codes, intersection kernel</td>
<td>71.5 ± 1.1 [1024] (3)</td>
<td>83.1 ± 0.6 [1024] (3)</td>
</tr>
<tr>
<td></td>
<td>70.3 ± 1.3 [1024]</td>
<td>83.2 ± 0.4 [1024]</td>
</tr>
<tr>
<td></td>
<td>71.8 ± 1.0 [1024] (4)</td>
<td>84.1 ± 0.5 [1024] (4)</td>
</tr>
</tbody>
</table>

Yang et al. (2009) have won the 2009 Pascal VOC challenge with this type of technique.

Scenes, supervised dictionary learning

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Unsup</th>
<th>Discr[1024]</th>
<th>Unsup</th>
<th>Discr[2048]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>83.6 ± 0.4</td>
<td>84.9 ± 0.3</td>
<td>84.2 ± 0.3</td>
<td>85.6 ± 0.2</td>
</tr>
<tr>
<td>Intersect</td>
<td>84.3 ± 0.5</td>
<td>84.7 ± 0.4</td>
<td>84.6 ± 0.4</td>
<td>85.1 ± 0.5</td>
</tr>
</tbody>
</table>
Non-blind deblurring *(Couzinie-Devy, Mairal, Bach, Ponce, 2011)*

<table>
<thead>
<tr>
<th>Method</th>
<th>Cameraman PSNR</th>
<th>Lena PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>20.76</strong></td>
<td><strong>25.84</strong></td>
</tr>
<tr>
<td>PSNR input image</td>
<td>22.35</td>
<td>27.57</td>
</tr>
<tr>
<td>Richardson-Lucy [23]</td>
<td>22.29</td>
<td>27.35</td>
</tr>
<tr>
<td>Sparse gradient [15]</td>
<td>24.7</td>
<td>29.00</td>
</tr>
<tr>
<td>SA-DCT [10]</td>
<td><strong>25.53</strong></td>
<td><strong>30.74</strong></td>
</tr>
<tr>
<td>BM3D [4]</td>
<td><strong>23.44</strong></td>
<td><strong>28.97</strong></td>
</tr>
<tr>
<td>Linear</td>
<td><strong>23.83</strong></td>
<td><strong>7.97</strong></td>
</tr>
<tr>
<td>Linear + Dictionary</td>
<td><strong>3.83</strong></td>
<td><strong>7.95</strong></td>
</tr>
</tbody>
</table>

*Table showing PSNR values for different methods on Cameraman and Lena images.*
Non-blind deblurring *(Couzinie-Devy, Mairal, Bach, Ponce, 2011)*

Anisotropic (motion blur) kernels *(Levin et al., 2009)*

<table>
<thead>
<tr>
<th>Kernel</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td><strong>10.67</strong></td>
<td><strong>7.17</strong></td>
<td><strong>9.02</strong></td>
<td><strong>6.63</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kernel</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse gradient [15]</td>
<td>8.64</td>
<td>9.18</td>
<td><strong>11.15</strong></td>
<td><strong>10.24</strong></td>
</tr>
<tr>
<td>Ours</td>
<td><strong>10.52</strong></td>
<td><strong>10.03</strong></td>
<td>9.64</td>
<td>7.75</td>
</tr>
</tbody>
</table>
Image courtesy of J.-L. Starck
Digital zoom (Couzinie-Devy, Mairal, Bach, Ponce, 2011)

<table>
<thead>
<tr>
<th></th>
<th>Cubic spline</th>
<th>Yang et al. [28]</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>31.91</td>
<td>32.13 / 33.06</td>
<td>33.31</td>
</tr>
<tr>
<td>Girl</td>
<td>31.44</td>
<td>31.48 / 31.93</td>
<td>32.00</td>
</tr>
<tr>
<td>Flower</td>
<td>38.48</td>
<td>38.69 / 39.59</td>
<td>39.92</td>
</tr>
</tbody>
</table>
(Glasner et al., 2009)
(Couzinie-Devy et al., 2011)
Learning to estimate and remove non-uniform image blur

(Couzinie-Devy et al., 2012)
Inverse halftoning
(Mairal, Bach, Ponce, 2010)
Inverse halftoning
(Mairal, Bach, Ponce, 2010)
PSNR comparison between our method and Kite et al.'00 [FIHT2]; Neelamini et al.'09 [WInHD]; Foi et al.'04 [LPA-ICI]; and Dabov et al.'06 [SA-DCT].
Epitomic dictionaries
(Benoit, Mairal, Bach, Ponce, CVPR’10)

Epitomes: (Jojic, Frey, Kannan, 2003)
Related ideas: (Aharon & Elad, 2007; Hyvarinen & Hoyer, 2001; Kavukcuoglu et al., 2009; Zeiler et al., 2010)
Pairs of epitomes obtained for different patch sizes

Denoising experiment

<table>
<thead>
<tr>
<th>Image</th>
<th>σ</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>house</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>35.89</td>
<td>34.33</td>
<td>33.25</td>
<td>32.03</td>
<td></td>
</tr>
<tr>
<td>ISD</td>
<td>36.05</td>
<td>34.25</td>
<td>32.72</td>
<td>31.76</td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>35.63</td>
<td>33.43</td>
<td>32.01</td>
<td>30.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barbara</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>34.07</td>
<td>33.91</td>
<td>30.43</td>
<td>29.24</td>
<td></td>
</tr>
<tr>
<td>ISD</td>
<td>34.21</td>
<td>32.22</td>
<td>30.71</td>
<td>29.22</td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>34.00</td>
<td>31.71</td>
<td>30.20</td>
<td>28.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lena</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>35.44</td>
<td>33.62</td>
<td>32.27</td>
<td>31.37</td>
<td></td>
</tr>
<tr>
<td>ISD</td>
<td>35.42</td>
<td>33.64</td>
<td>32.25</td>
<td>31.09</td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>35.17</td>
<td>33.23</td>
<td>31.73</td>
<td>30.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>33.66</td>
<td>31.72</td>
<td>30.33</td>
<td>29.33</td>
<td></td>
</tr>
<tr>
<td>ISD</td>
<td>33.64</td>
<td>31.79</td>
<td>30.41</td>
<td>28.45</td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>33.49</td>
<td>31.50</td>
<td>29.99</td>
<td>28.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peppers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>34.46</td>
<td>32.37</td>
<td>30.93</td>
<td>29.70</td>
<td></td>
</tr>
<tr>
<td>ISD</td>
<td>34.37</td>
<td>32.33</td>
<td>30.89</td>
<td>29.79</td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>33.92</td>
<td>31.76</td>
<td>30.20</td>
<td>29.03</td>
<td></td>
</tr>
</tbody>
</table>

ISD = (Aharon & Elad’08)
DL=flat dict. learning
Proximal methods for sparse hierarchical dictionary learning
(Jenatton, Mairal, Obozinski, Bach, ICML’10)
Proximal methods for sparse hierarchical dictionary learning (Jenatton, Mairal, Obozinski, Bach, ICML'10)
Network flow algorithms for structured sparsity
(Mairal, Jenatton, Obozinski, Bach, NIPS’11)
SPArse Modeling software (SPAMS)

http://www.di.ens.fr/willow/SPAMS/

Tutorials on sparse coding and dictionary learning for image analysis

ICCV’09: www.di.ens.fr/~mairal/tutorial_iccv09/
NIPS’09: www.di.ens.fr/~fbach/nips2009tutorial/
CVPR’10: www.di.ens.fr/~mairal/tutorial_cvpr2010/


References II


References VI


References VII


References VIII


References X


