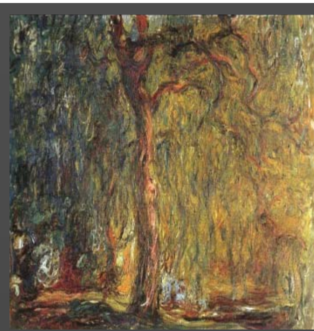
# Sparse Coding
# for
# Image and Video Understanding



Jean Ponce
http://www.di.ens.fr/willow/
Willow team, LIENS, UMR 8548
Ecole normale supérieure, Paris

# Sparse Coding
# for
# Image and Video Understanding

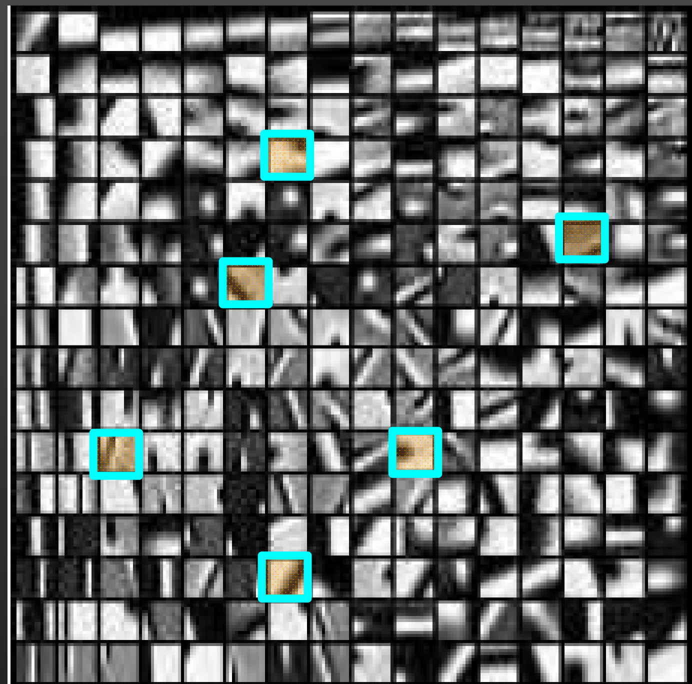
Julien Mairal and Francis Bach

Guillermo Sapiro (Minnesota) and Andrew Zisserman (Oxford)

# Sparse linear models

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

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



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

(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

$\min_{\alpha} 1/2 \, | x - D\alpha |_2^2$

$\min_{\alpha} 1/2 \, | x - D\alpha |_2^2 + \lambda \, |\alpha|_0$

$\min_{\alpha} 1/2 \, | x - D\alpha |_2^2 + \lambda \, \psi(\alpha)$

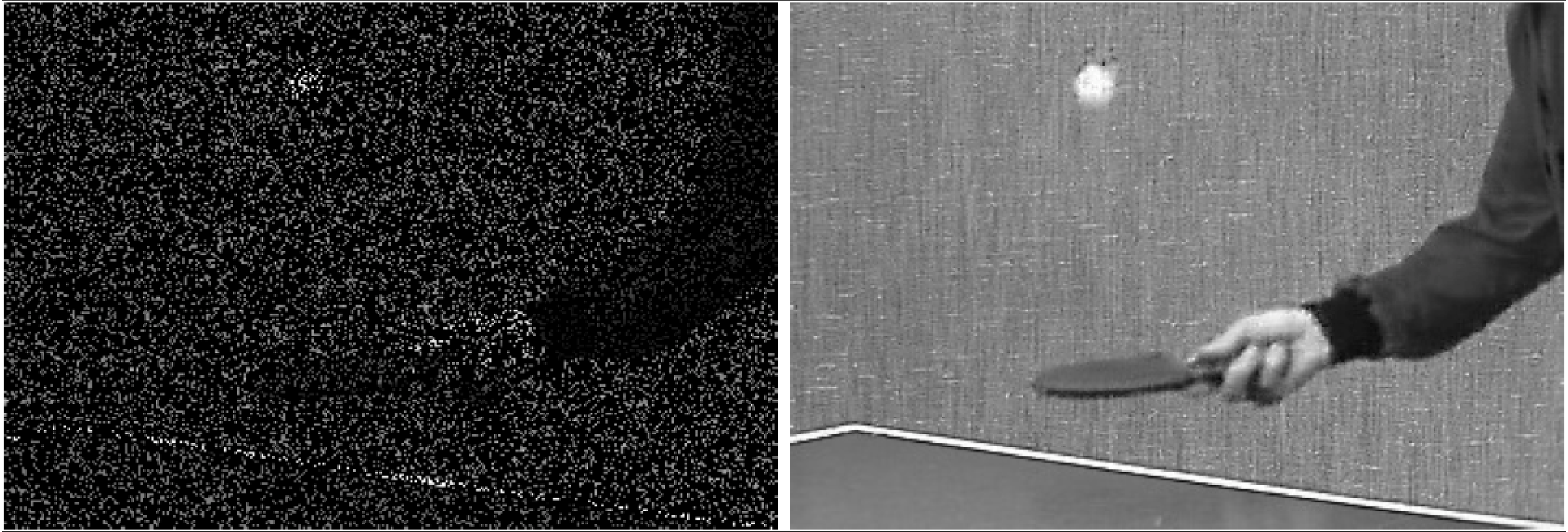
$\min_{D \in C, \alpha_1, \ldots, \alpha_n} \sum_{1 \leq i \leq n} [\, 1/2 \, | x_i - D\alpha_i |_2^2 + \lambda \, \psi(\alpha_i) \,]$

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

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

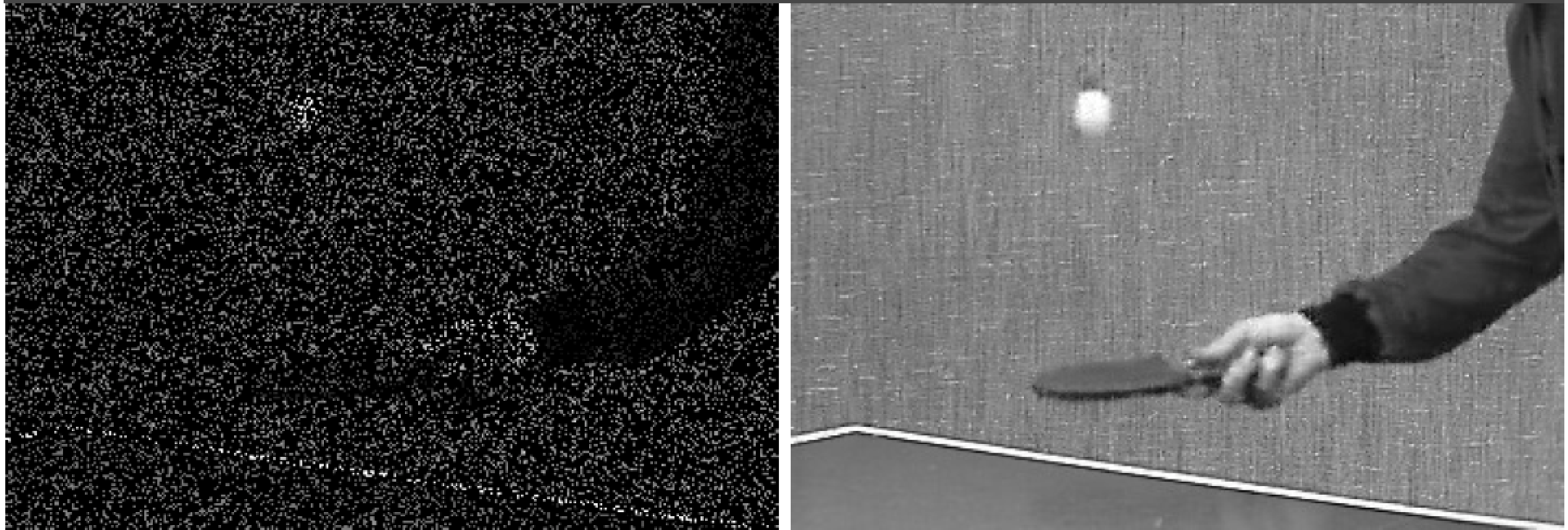
Least squares
Sparse coding
Dictionary learning
Learning for a task
Learning structures

# Video inpainting



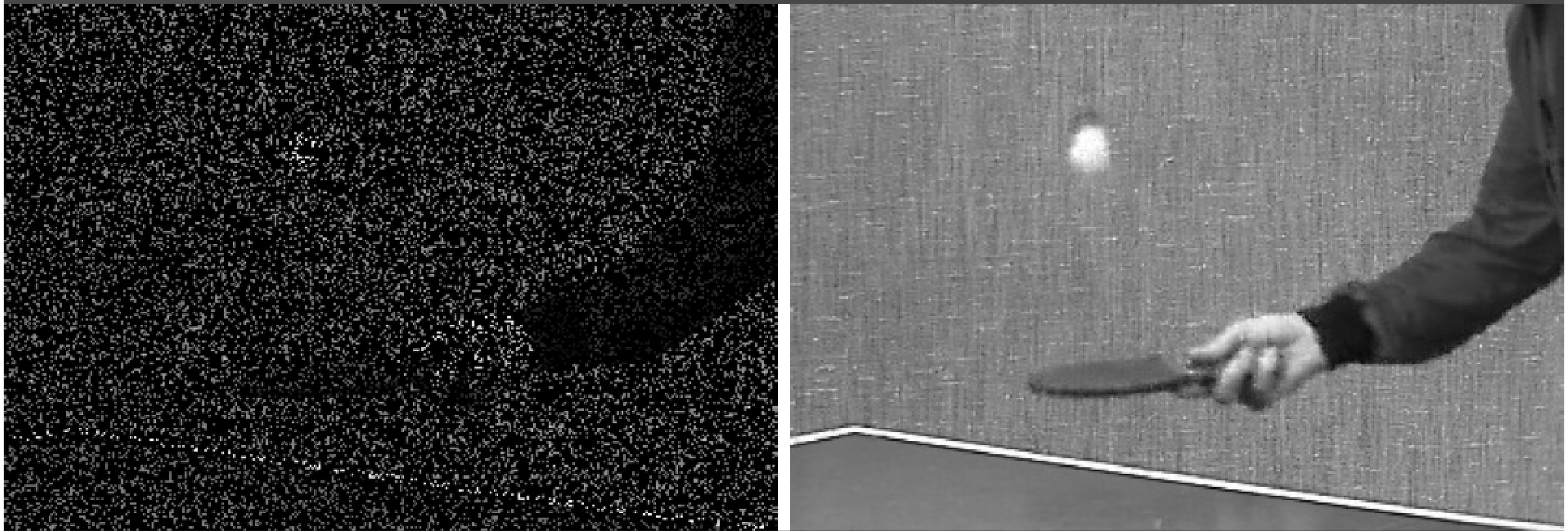
(Mairal, Sapiro and Elad, 2008)

# Video inpainting



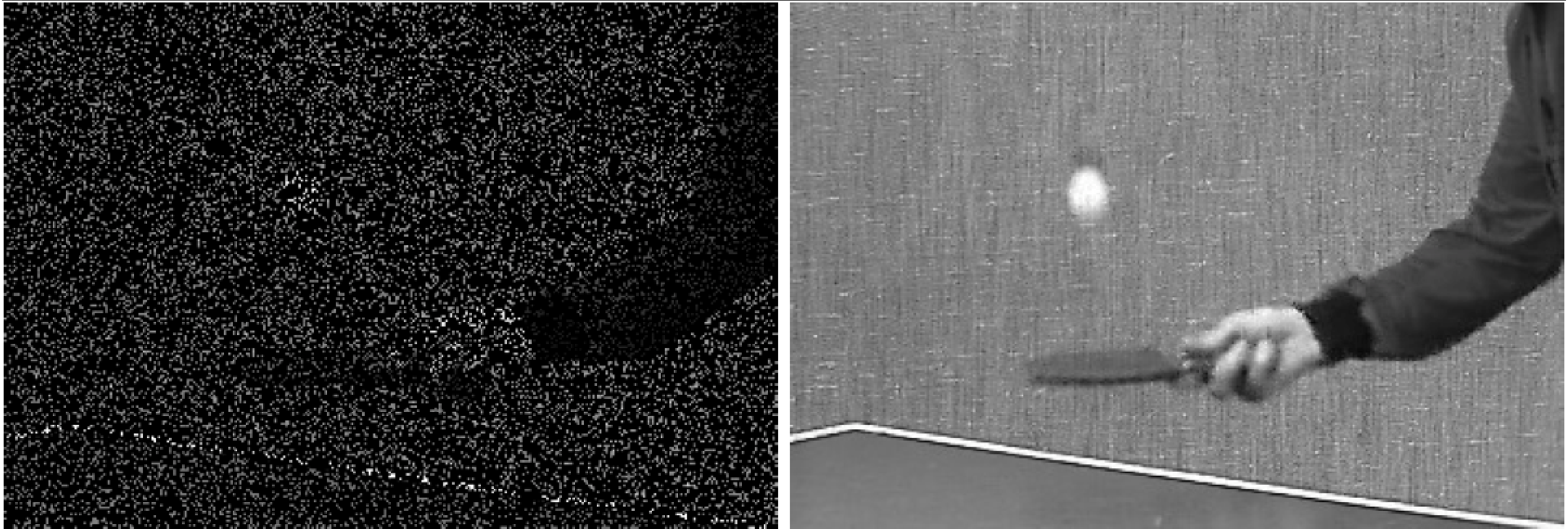
(Mairal, Sapiro and Elad, 2008)

# Video inpainting



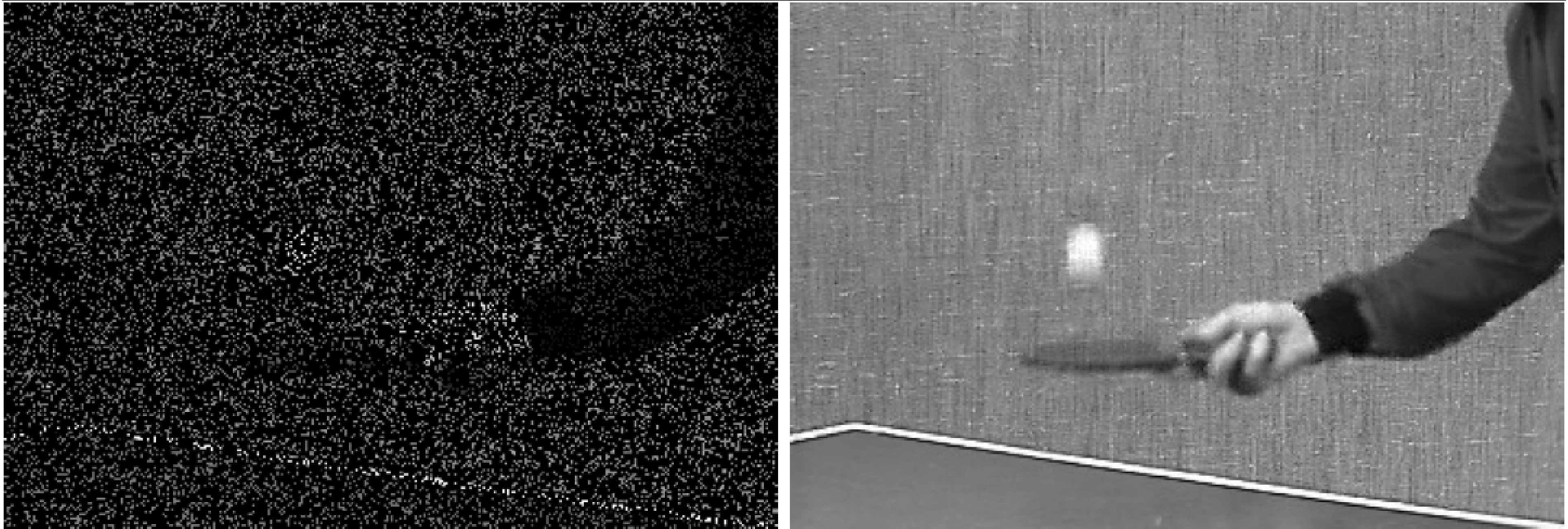
(Mairal, Sapiro and Elad, 2008)

# 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.
- Dictionary can be learned on the data you want to restore itself.
- Sparse coding is well adapted to data that admit sparse representations.
- Sparse coding is for sparse data only.
- 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:

$$\min_\alpha 1/2 \mid x - D\alpha \mid_2^2 + \lambda \mid\alpha\mid_0$$

NP-hard, greedy approximate algorithms

- The $l_1$ version:

$$\min_\alpha 1/2 \mid x - D\alpha \mid_2^2 + \lambda \mid\alpha\mid_1$$

convex, exact algorithms

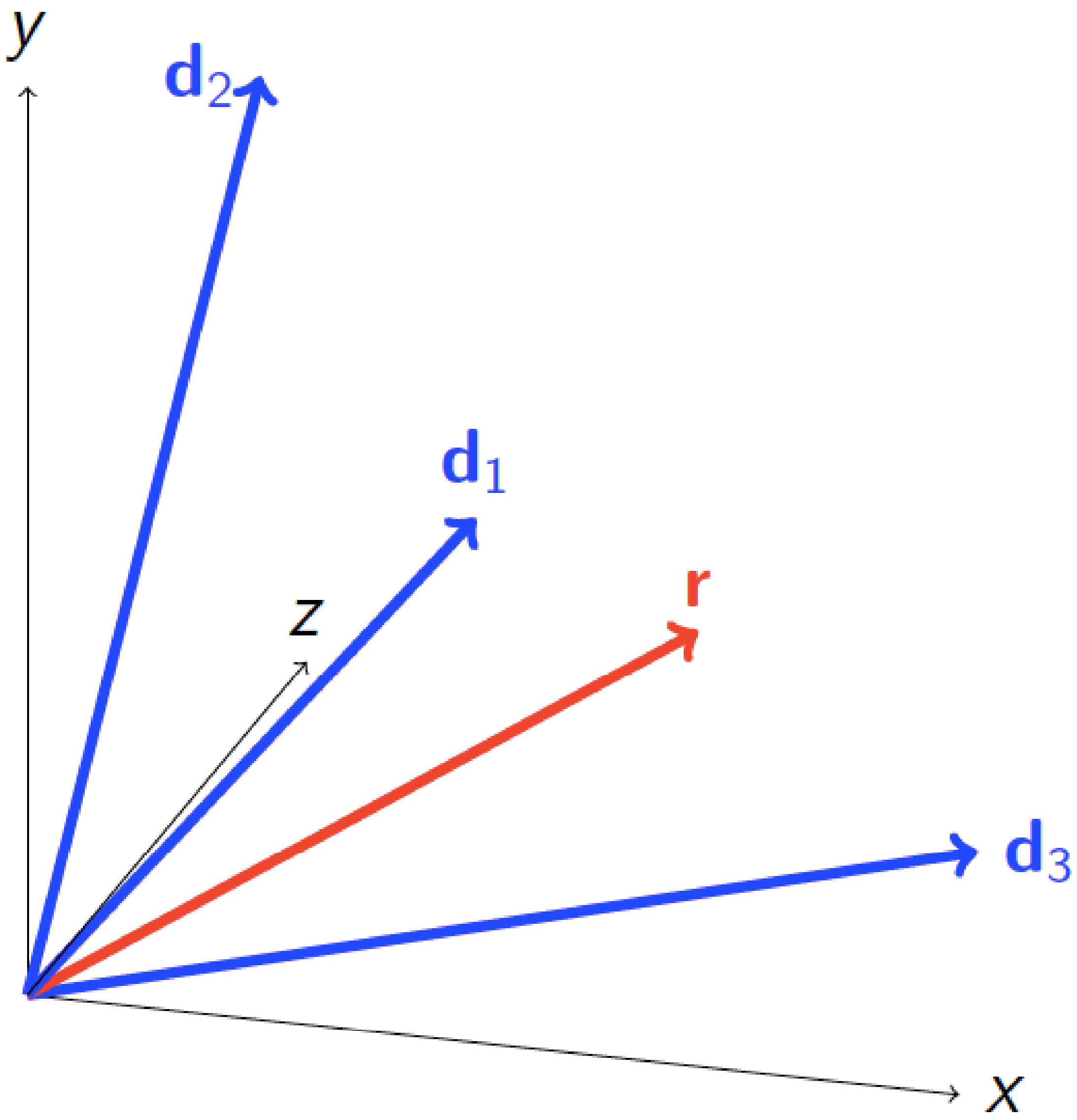
Finding your way in the sparse coding literature is not easy. The literature is vast, redundant, sometimes confusing and many papers are claiming victory.

The main classes of methods are:

- greedy procedures [Mallat and Zhang, 1993], [Weisberg, 1980],
- homotopy [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-$\ell_2$ methods [Daubechies et al., 2009],
- active-set methods [Roth and Fischer, 2008].

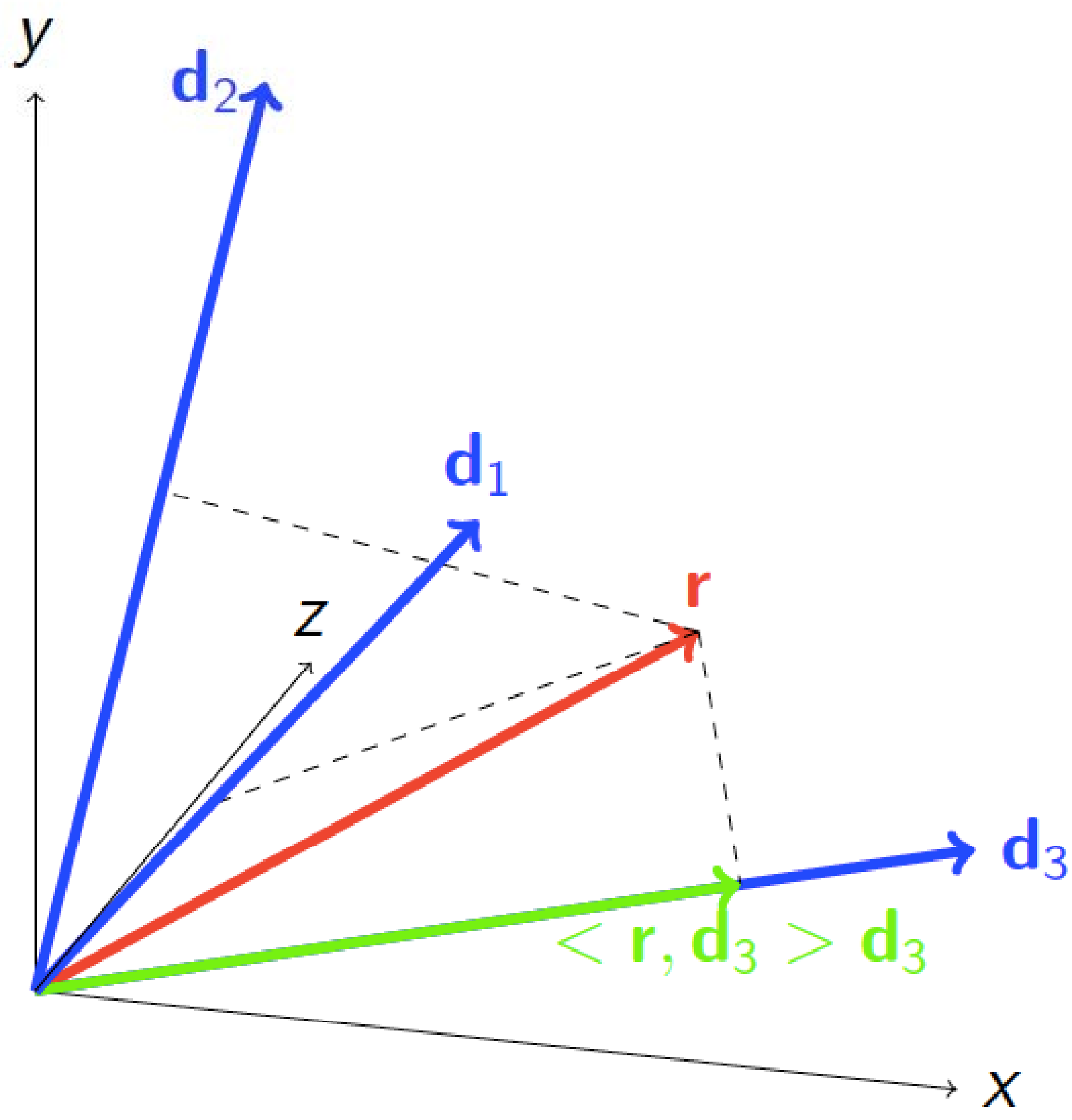
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



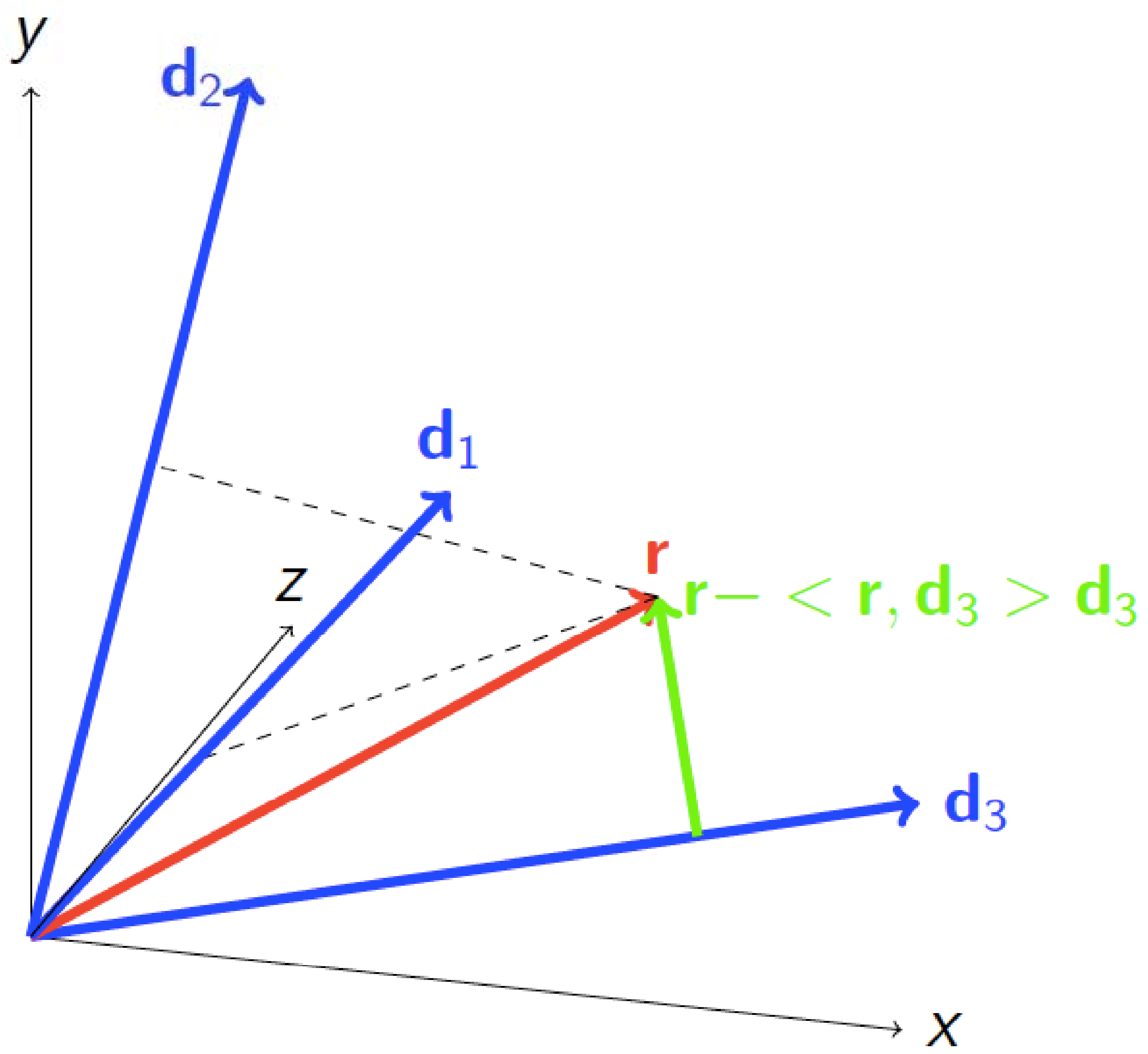
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



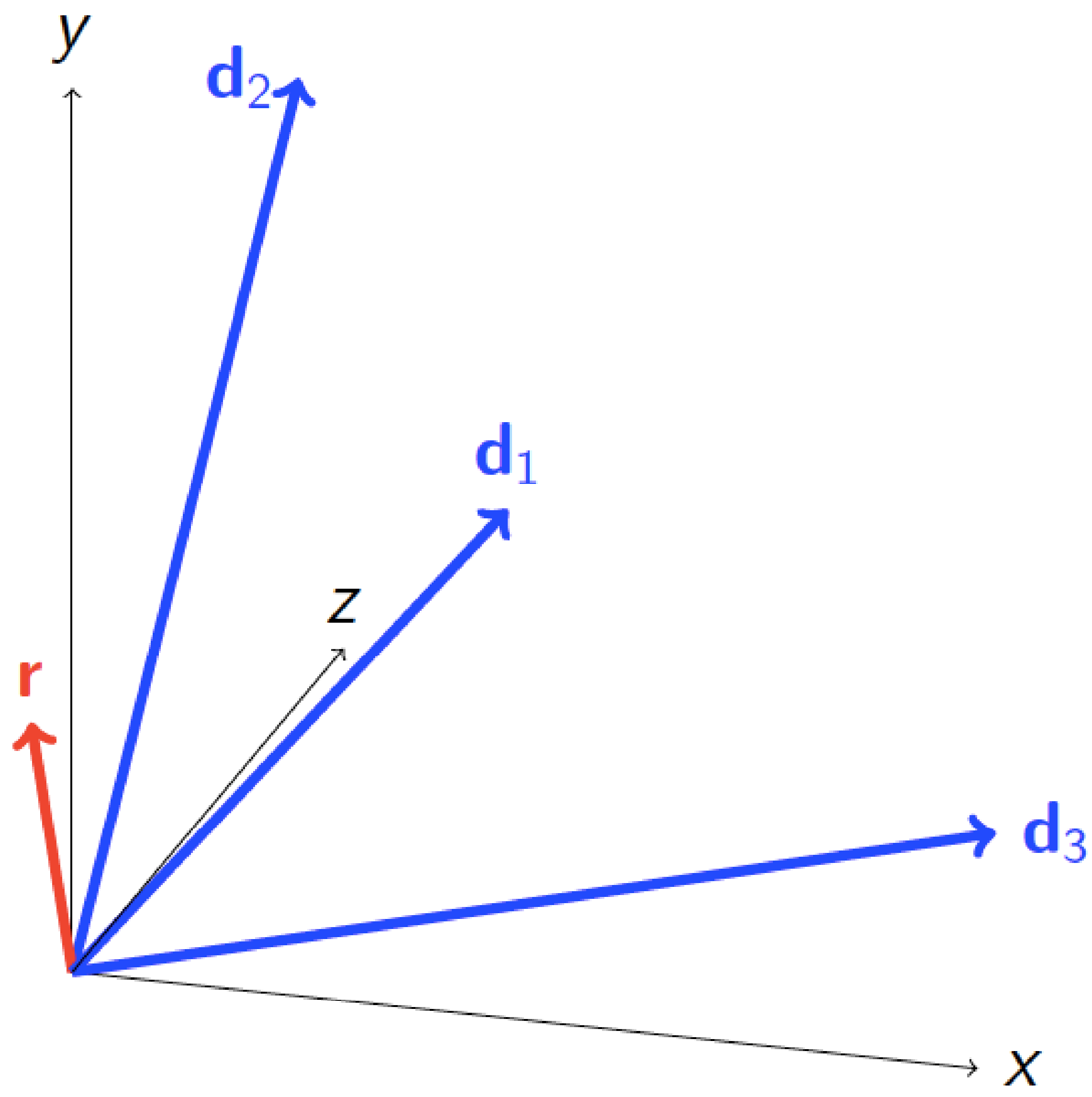
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



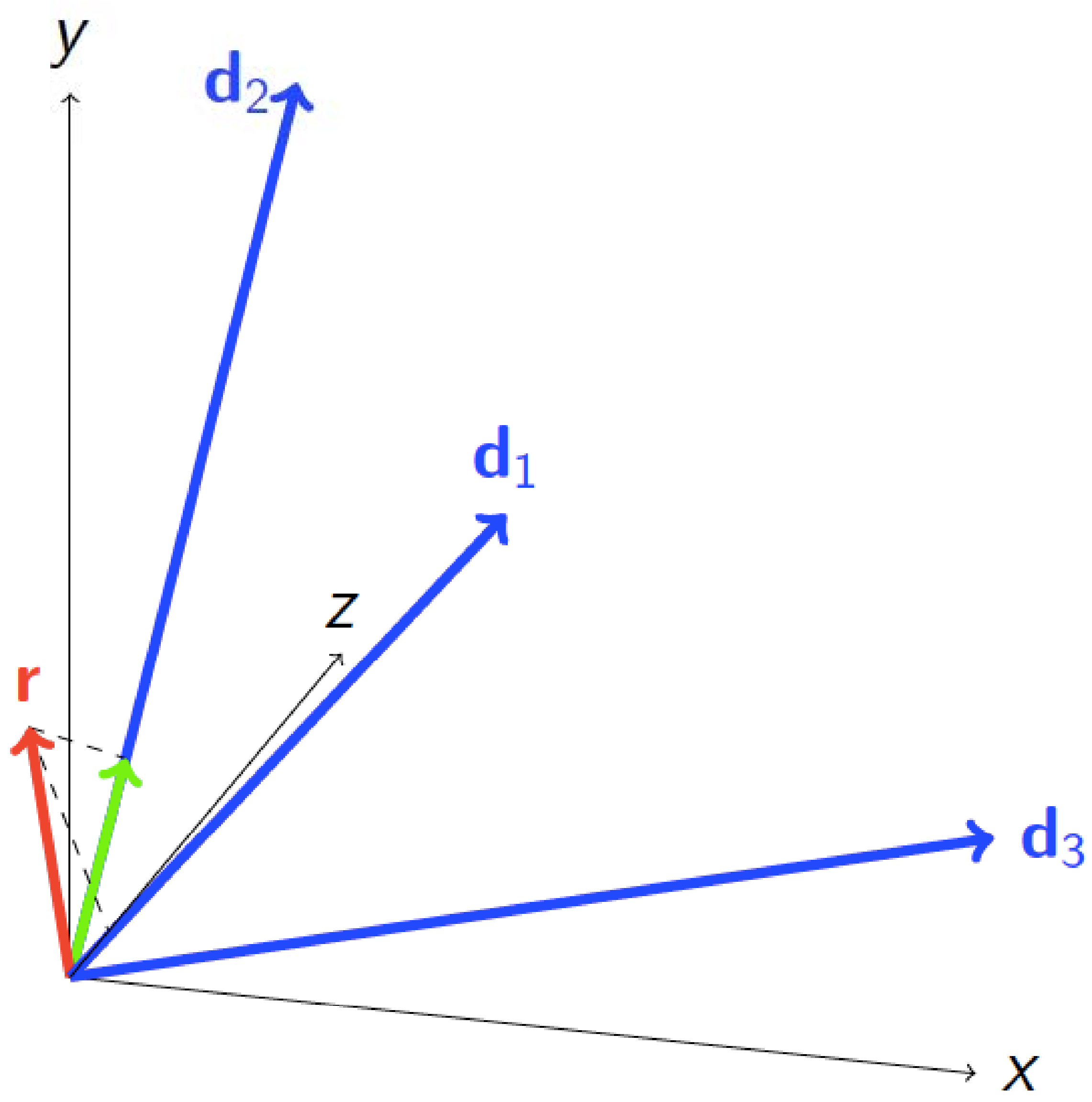
# Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0.75)$$



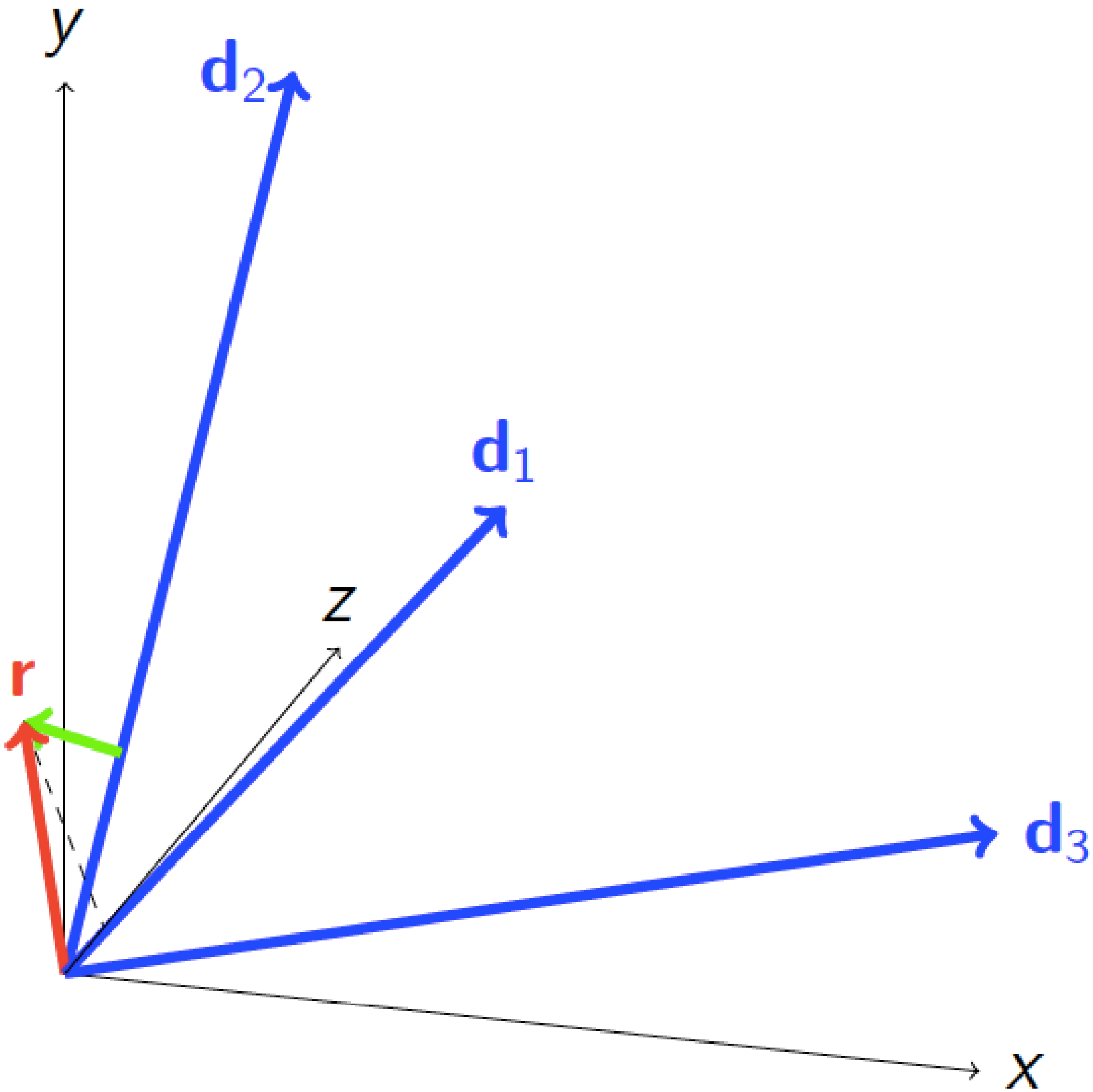
## Matching Pursuit

$\boldsymbol{\alpha} = (0, 0, 0.75)$



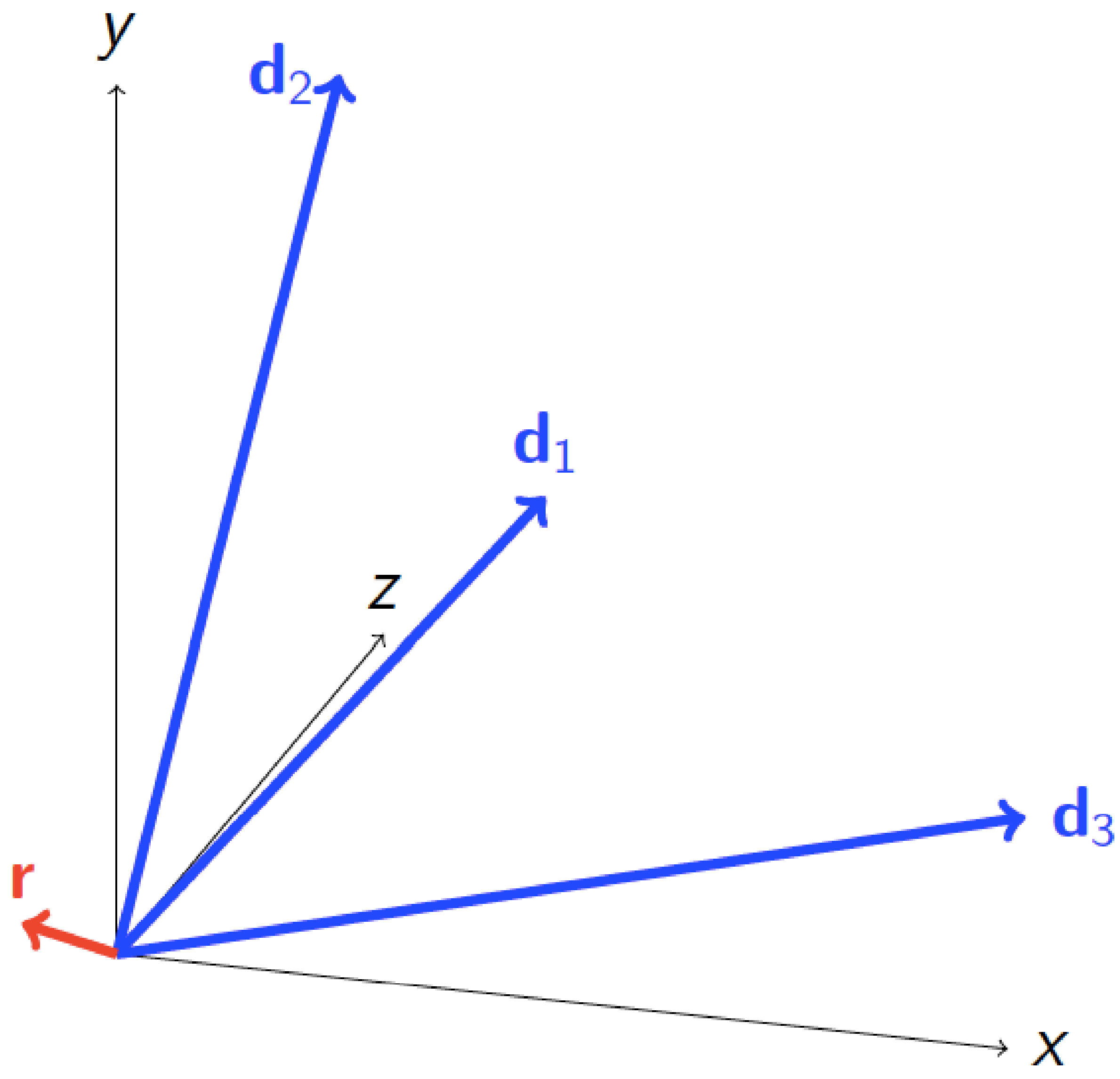
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0.75)$$



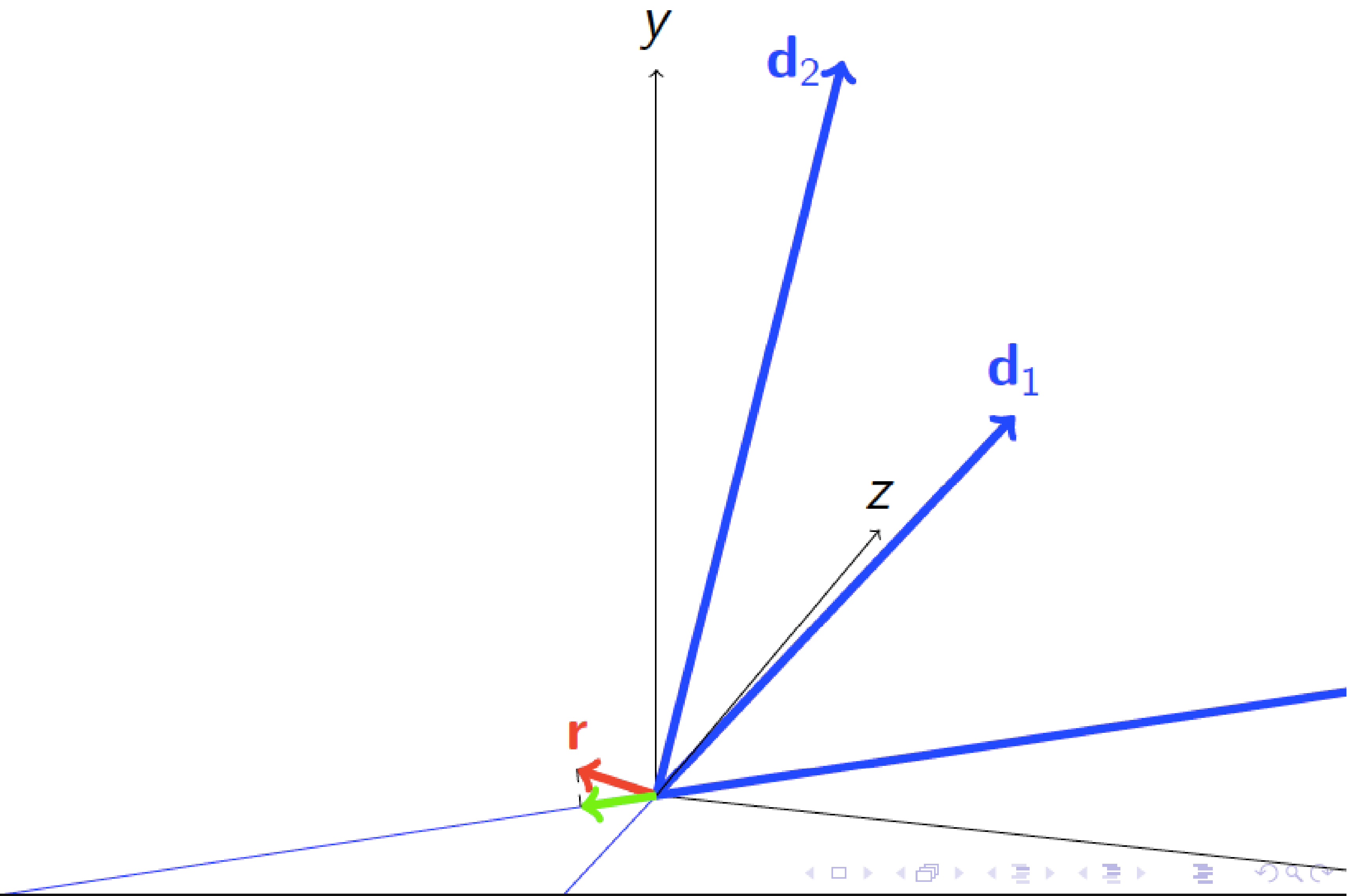
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0.24, 0.75)$$



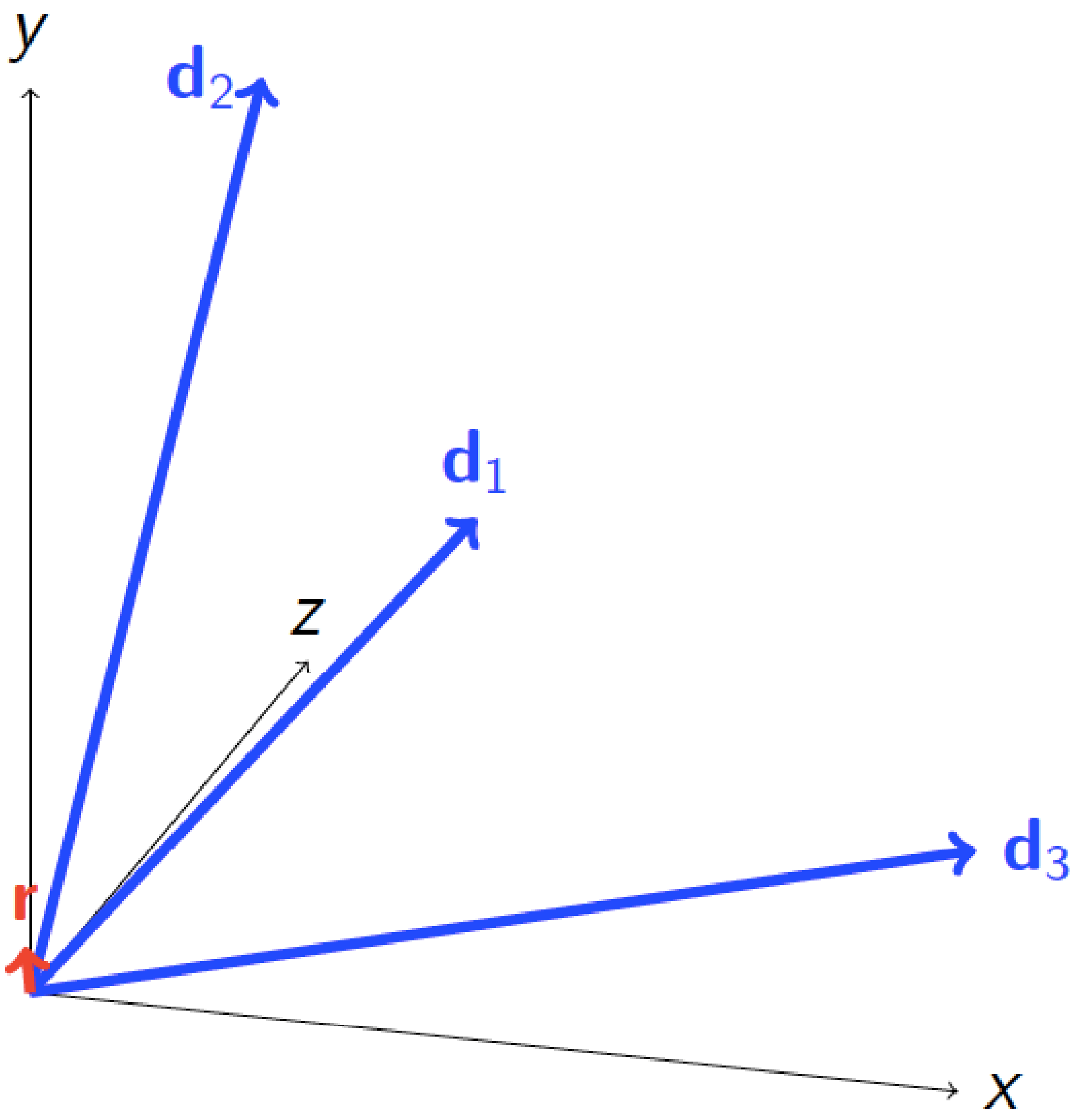
## Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0.24, 0.75)$$

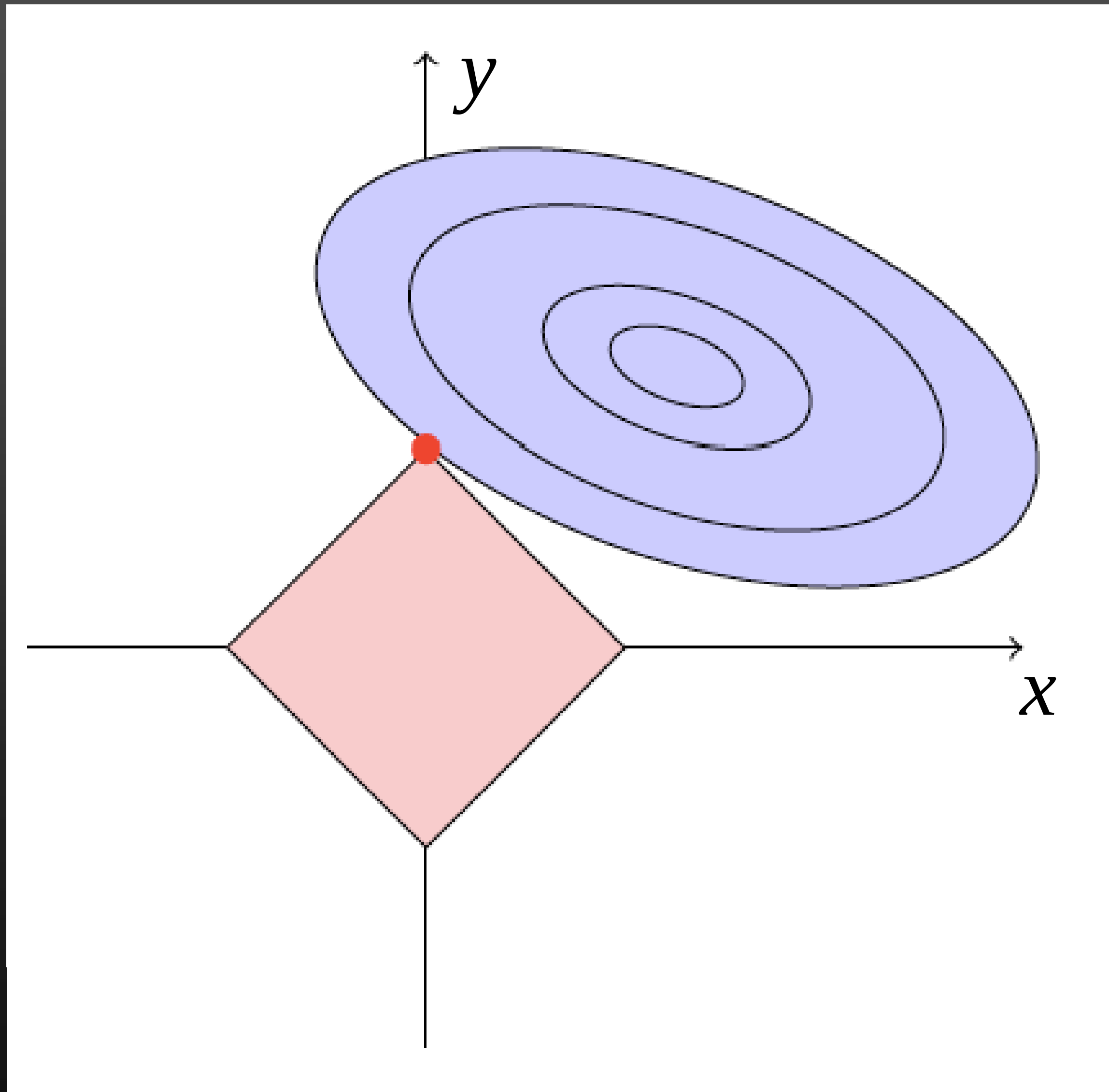


## Matching Pursuit

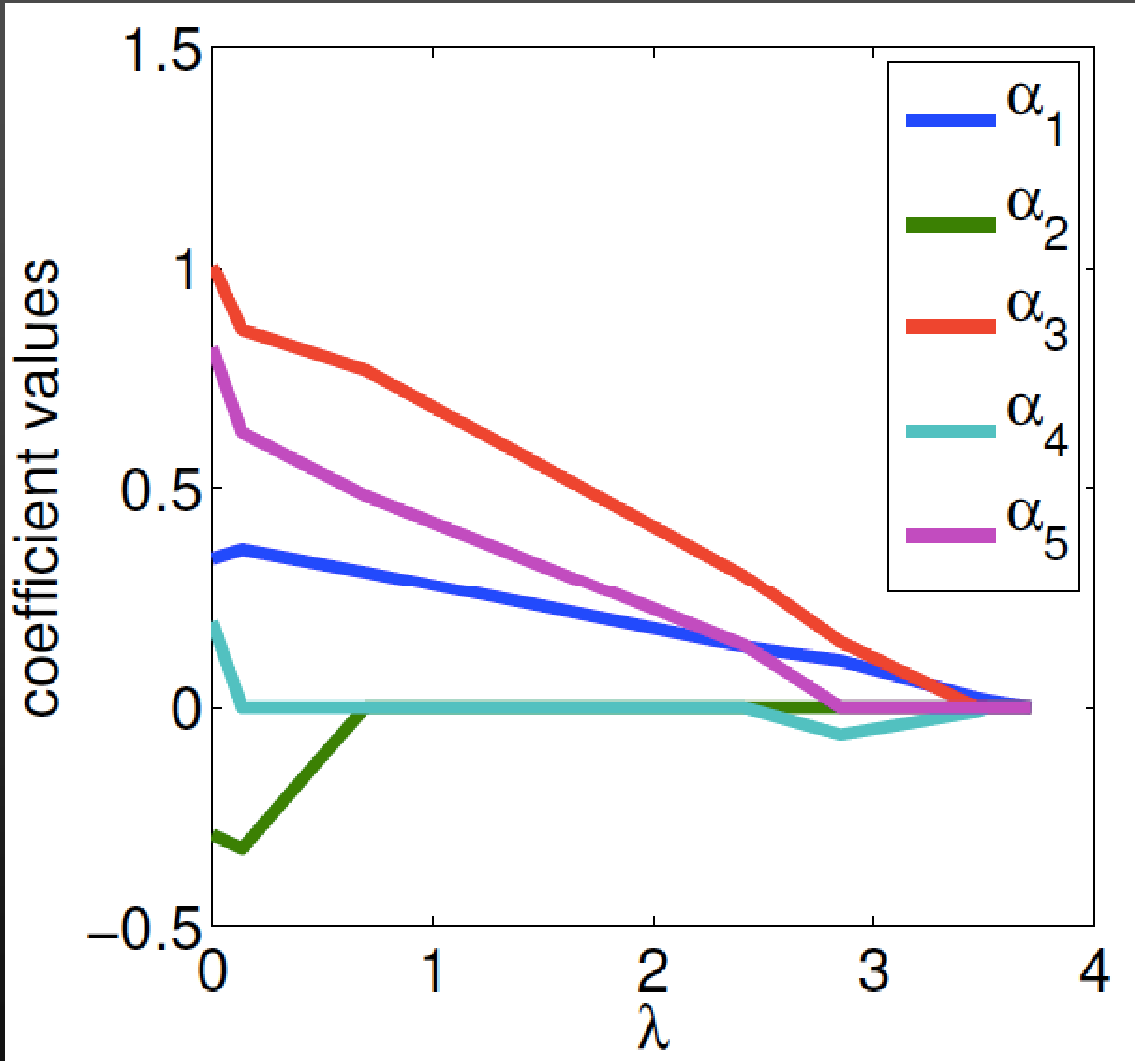
$$\boldsymbol{\alpha} = (0, 0.24, 0.65)$$



# The $l_1$ norm and sparsity



# LARS (Efron et al., 2004)



# Dictionary learning

- Given some loss function, e.g.,

$$L(x, D) = \min_{\alpha} 1/2 \mid x - D\alpha \mid_2^2 + \lambda \mid\alpha\mid_1$$

- One usually minimizes, given some data $x_i$, i = 1, ..., n, the empirical risk:

$$\min_D f_n(D) = 1/n \sum_{1 \leq i \leq n} L(x_i, D)$$

- 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 $\rightarrow$ Stochastic gradient descent)

# Online sparse matrix factorization

(Mairal, Bach, Ponce, Sapiro, ICML'09, JMLR'10)

**Problem:**

$$\min_D f(D) = \mathbb{E}_x [ L(x, D) ]$$

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

**Algorithm:**

Iteratively draw one random training sample $x_t$ and minimize the quadratic surrogate function:

$$g_t(D) = 1/t \sum_{1 \le i \le t} [ 1/2 \, | x_i - D\alpha_i |_2^2 + \lambda \, |\alpha_i|_1 ]$$

(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:**

$$\min_D f(D) = \mathbb{E}_x [ L(x, D) ]$$

$$\min_{D \in C, A} [ 1/2 \| X - DA \|_F^2 + \lambda \|A\|_1 ]$$

**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)

**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

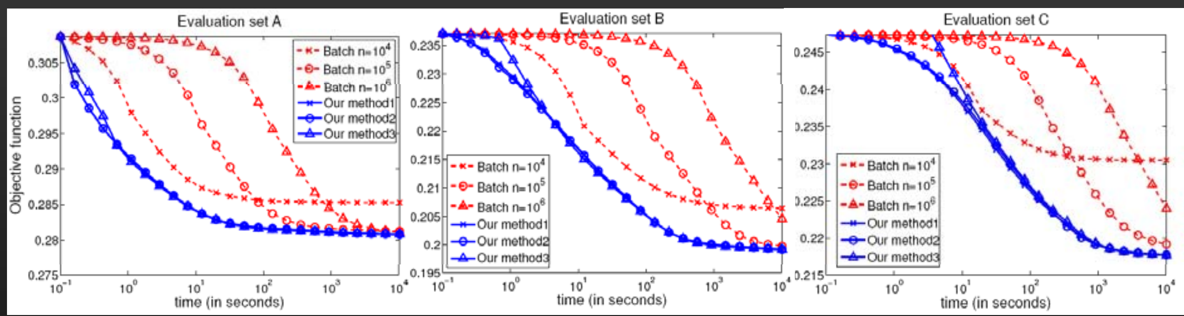
Three datasets constructed from 1,250,000 Pascal'06 patches (1,000,000 for training, 250,000 for testing):

- A: 8×8 b&w patches, 256 atoms.
- B: 12×16×3 color patches, 512 atoms.
- C: 16×16 b&w patches, 1024 atoms.

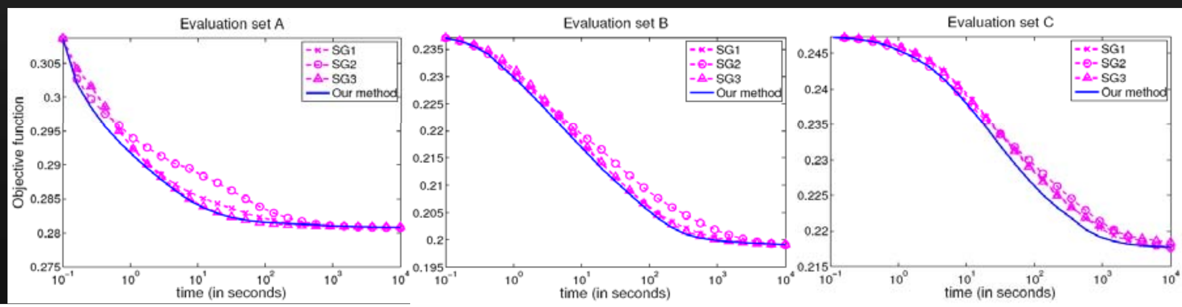
Two variants of our algorithm:

- 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

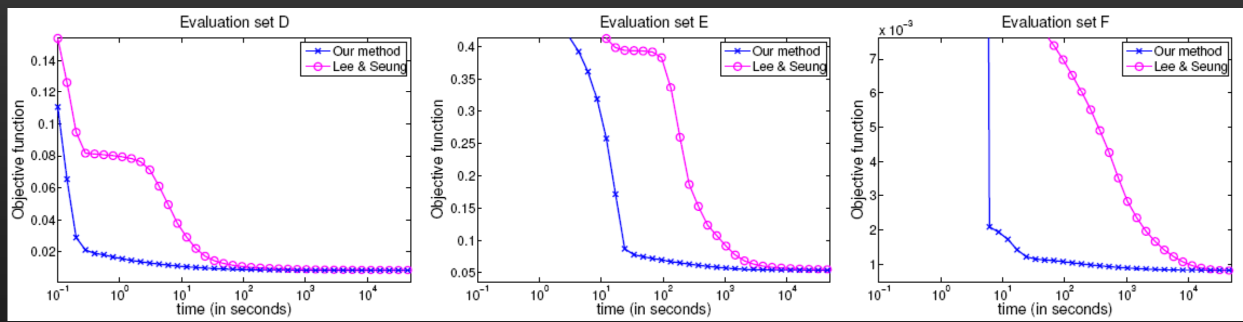
## Three datasets:

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

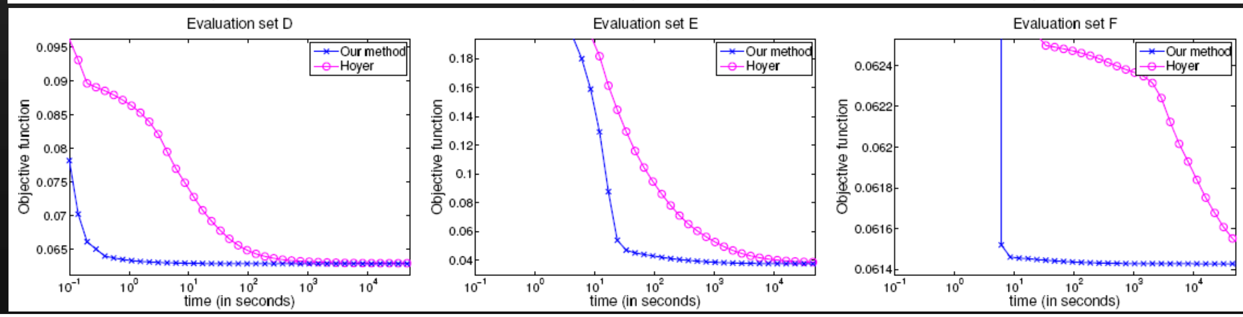
## Three implementations:

- 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

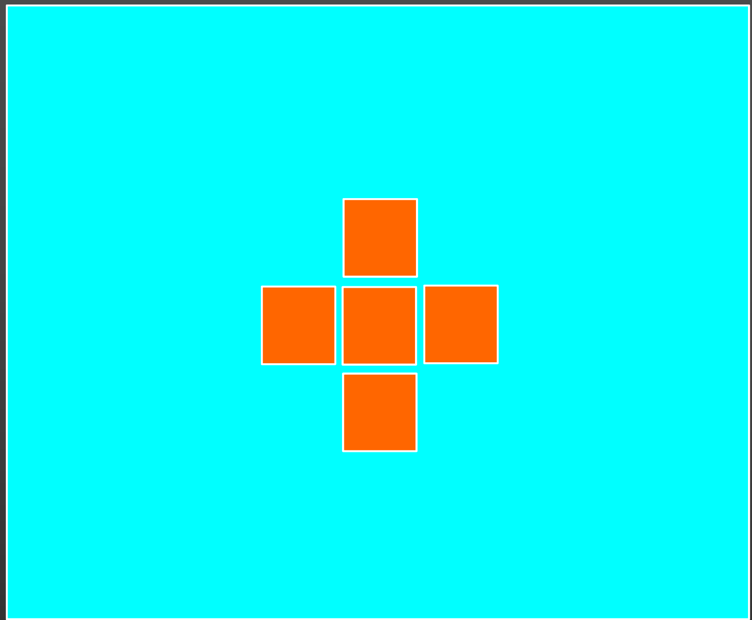


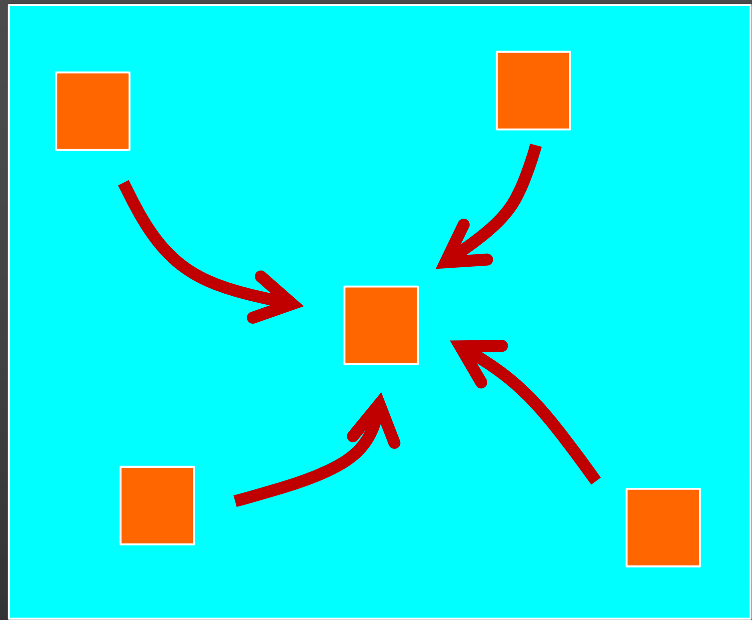

SPCA vs NNSC




Inpainting a 12MP image with a dictionary learned from 7x10⁶ patches in 500s (Mairal et al., 2009)
g mountain
alley from th
lf a dread o
orning cam

# State of the art in image denoising





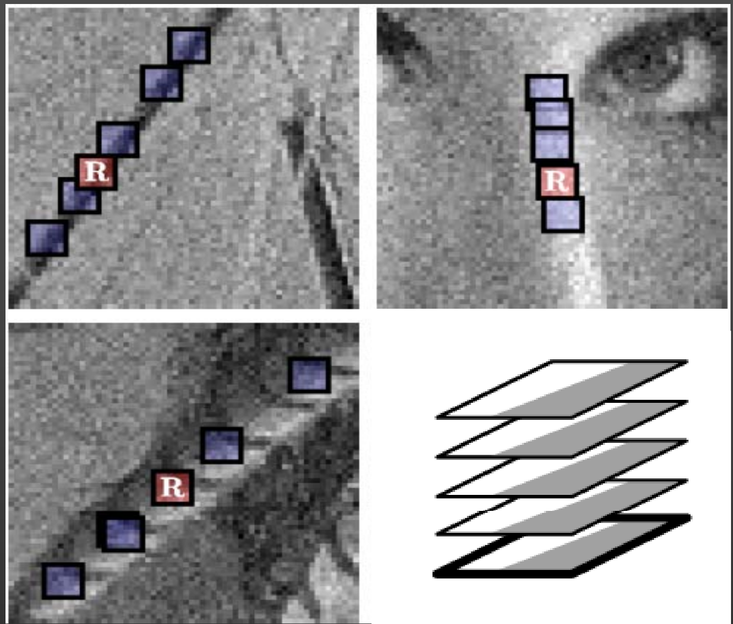
Non-local means filtering (Buades et al.'05)

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

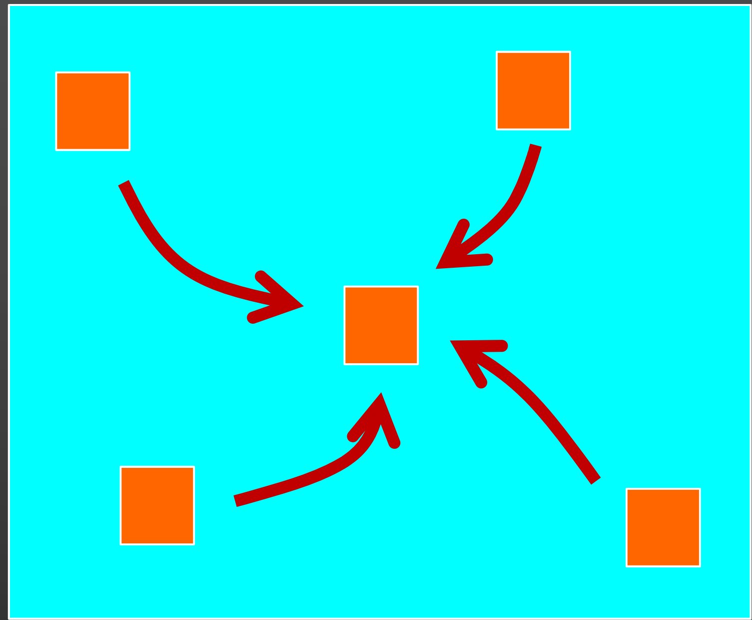
$$\min_{D \in C, \alpha_1, \ldots, \alpha_n} \sum_{1 \le i \le n} \left[ 1/2 \, | x_i - D\alpha_i |_2^2 + \lambda \, |\alpha_i|_1 \right]$$

$$x = 1/n \sum_{1 \le i \le n} R_i D \alpha_i$$

# State of the art in image denoising




BM3D (Dabov et al.'07)



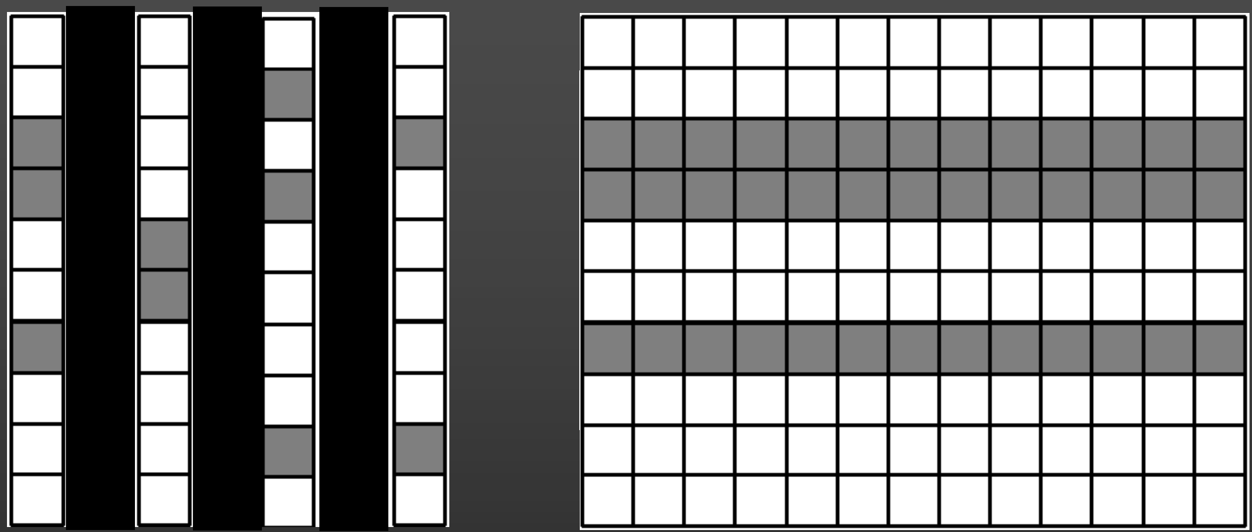
Non-local means filtering (Buades et al.'05)

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

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

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

# Non-local sparse models for image restoration (Mairal, Bach, Ponce, Sapiro, Zisserman, ICCV'09)




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

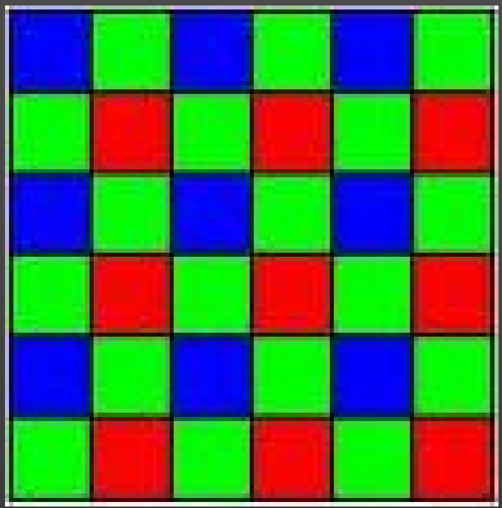
$$|A|_{p,q} = \sum_{1 \leq i \leq k} |\alpha^i|_q^p \quad (p,q) = (1,2) \text{ or } (0,\infty)$$

| σ | [23] | [25] | [12] | [8] | 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** |

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].

# Demosaicking experiments



Bayer pattern


| Im. | AP | DL | LPA | SC | LSC | LSSC |
|---|---|---|---|---|---|---|
| 1 | 37.84 | 38.46 | 40.47 | 40.84 | 40.92 | **41.36** |
| 2 | 39.64 | 40.89 | 41.36 | 41.76 | 42.03 | **42.24** |
| 3 | 41.40 | 42.66 | 43.47 | 43.15 | 43.92 | **44.24** |
| ... | ... | ... | ... | ... | ... | ... |
| 23 | 41.93 | 43.22 | 43.92 | 43.47 | 43.93 | **44.34** |
| 24 | 34.74 | 35.55 | 35.44 | 35.59 | 35.85 | **35.89** |
| **Av.** | 39.21 | 40.05 | 40.52 | 40.88 | 41.13 | **41.39** |

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)

$$\alpha^*(x,D) = \underset{\alpha}{\mathrm{Argmin}}\ |\, x - D\alpha\, |_2^2 \ \text{ s.t. } |\alpha|_0 \leq L$$

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

Orthogonal matching pursuit
(Mallat & Zhang'93, Tropp'04)

Reconstruction (MOD: Engan, Aase, Husoy'99;
K-SVD: Aharon, Elad, Bruckstein'06):

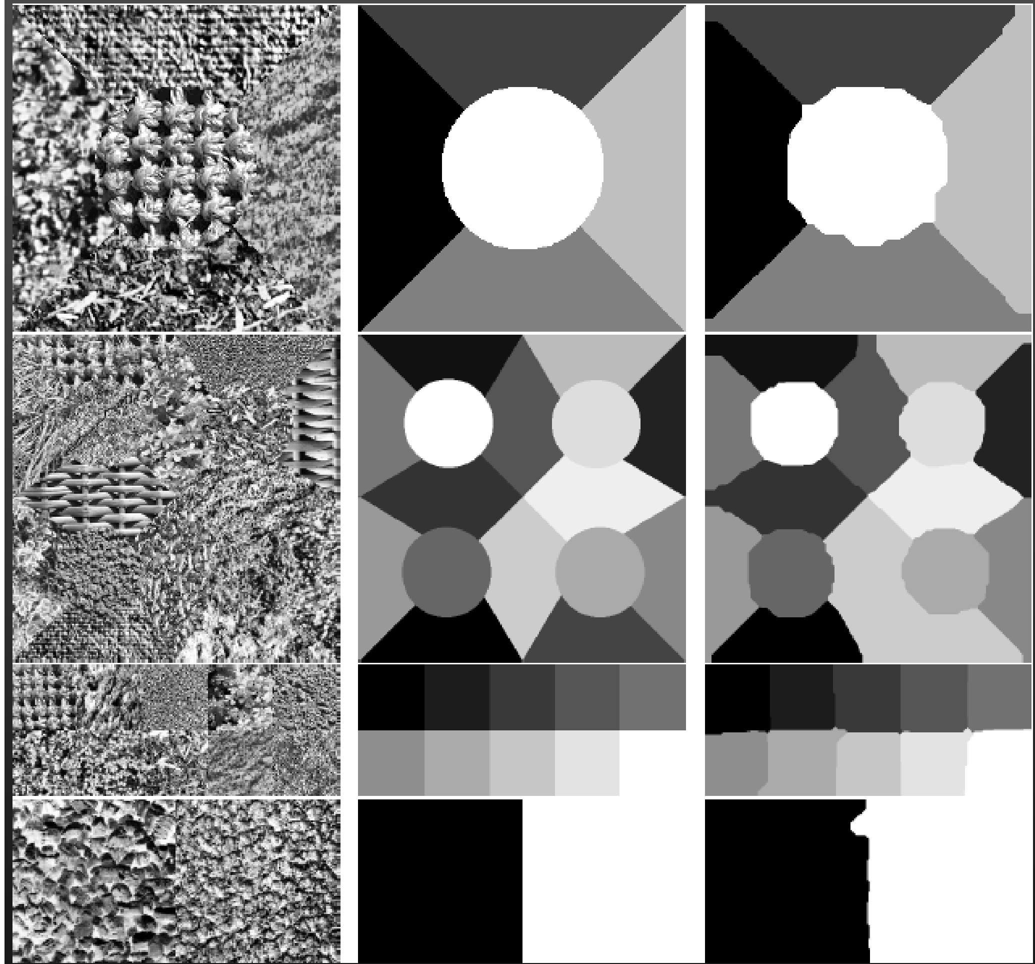
$$\min_{D} \textstyle\sum_l R^*(x_l,D)$$

Discrimination:

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

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

# Texture classification results

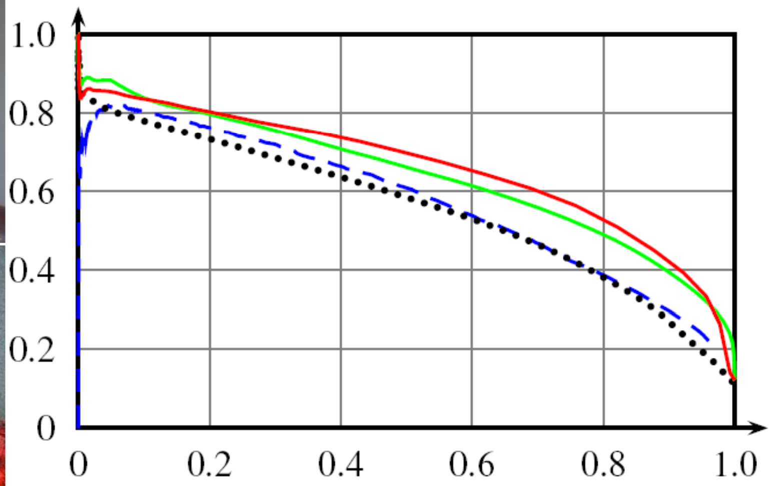


| ♯ | [28] | [17] | [34] | [16] | R1 | R2 | D1 | D2 |
|---|---|---|---|---|---|---|---|---|
| 1 | 7.2 | 6.7 | 5.5 | 3.37 | 2.22 | 1.69 | 1.89 | **1.61** |
| 2 | 18.9 | 14.3 | **7.3** | 16.05 | 24.66 | 36.5 | 16.38 | 16.42 |
| 3 | 20.6 | 10.2 | 13.2 | 13.03 | 10.20 | 5.49 | 9.11 | **4.15** |
| 4 | 16.8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

# Pixel-level classification results

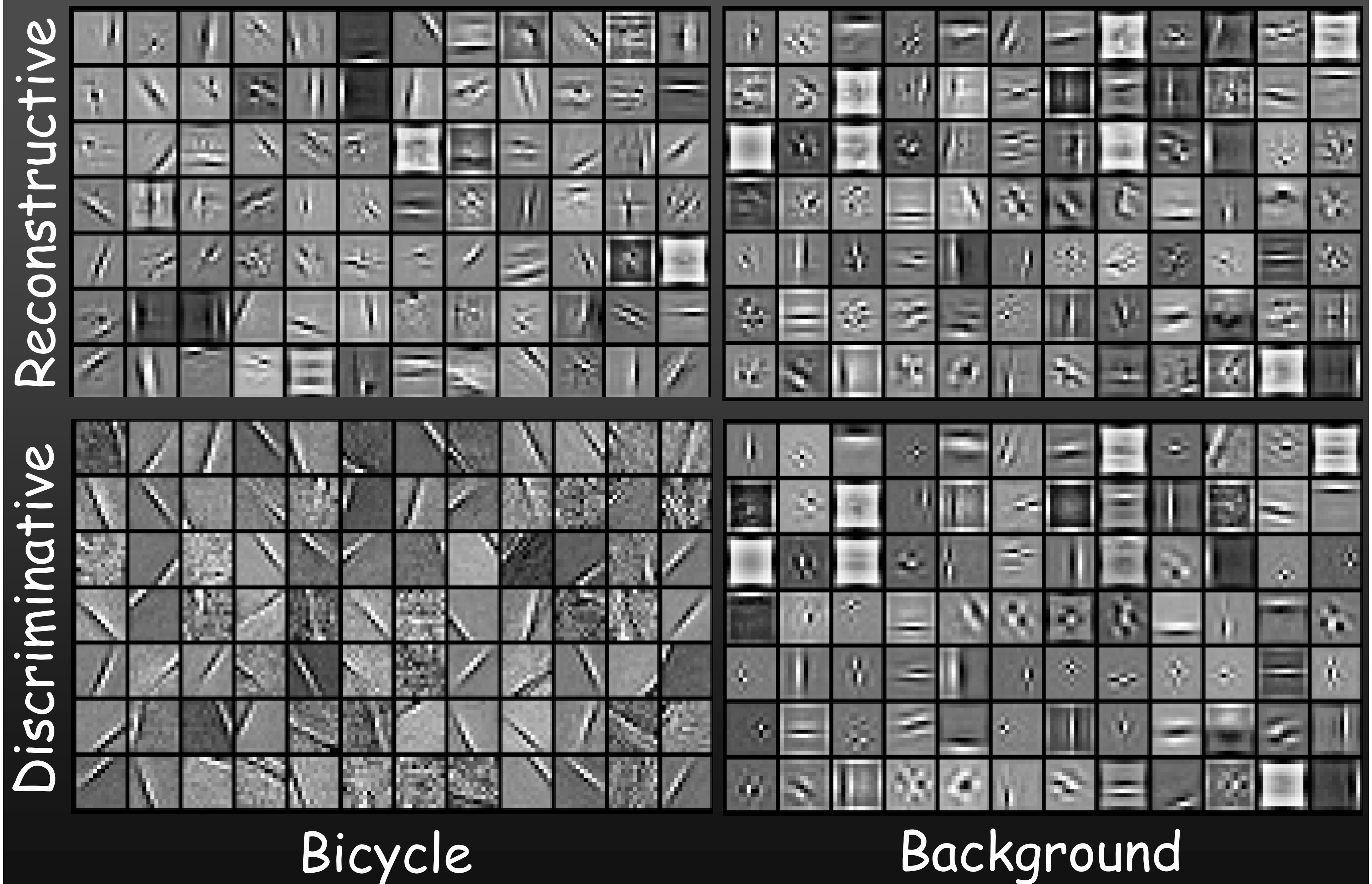
Qualitative results, Graz 02 data

Quantitative results




Comparaison with Pantofaru et al. (2006) and Tuytelaars & Schmid (2007).

# Reconstructive vs discriminative dictionaries



# Learning discriminative dictionaries with $l_1$ constraints

(Mairal, Leordeanu, Bach, Hebert, Ponce, ECCV'08)

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

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

Lasso: Convex optimization
(LARS: Efron et al.'04)

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

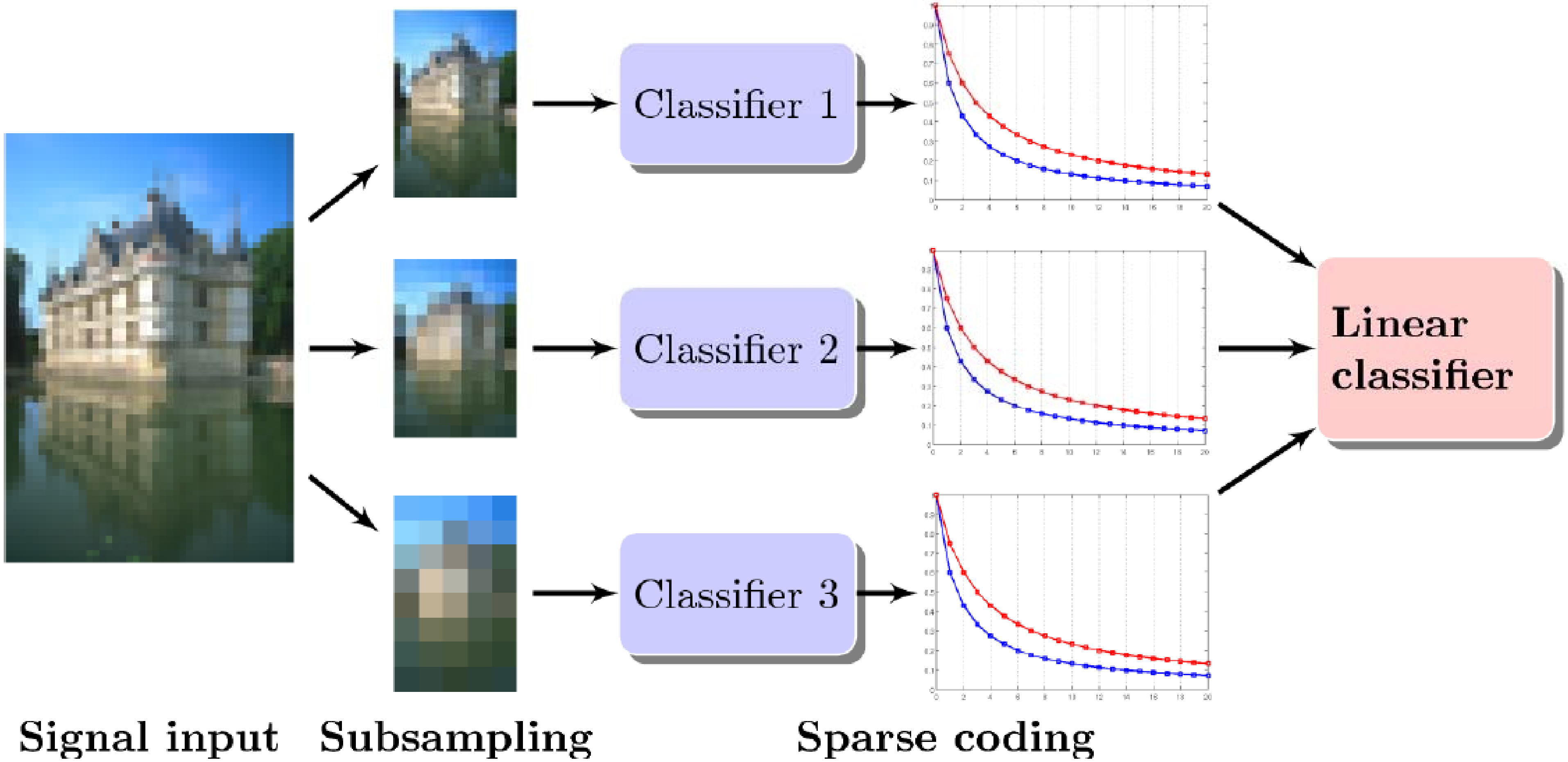
$$\min_{D} \sum_l R^*(x_l,D)$$

Discrimination:

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

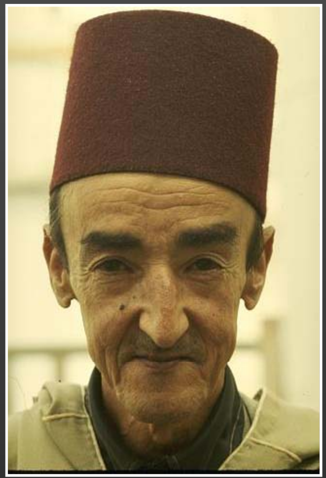
(Partial dictionary update with Newtown iterations on the dual problem; partial fast sparse coding with projected gradient descent.)

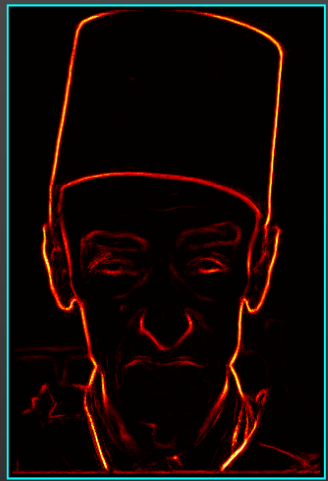
# Patch classification with learned dictionaries



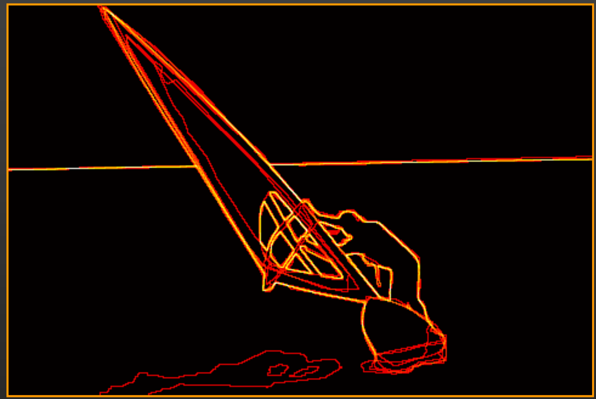
# Edge detection results

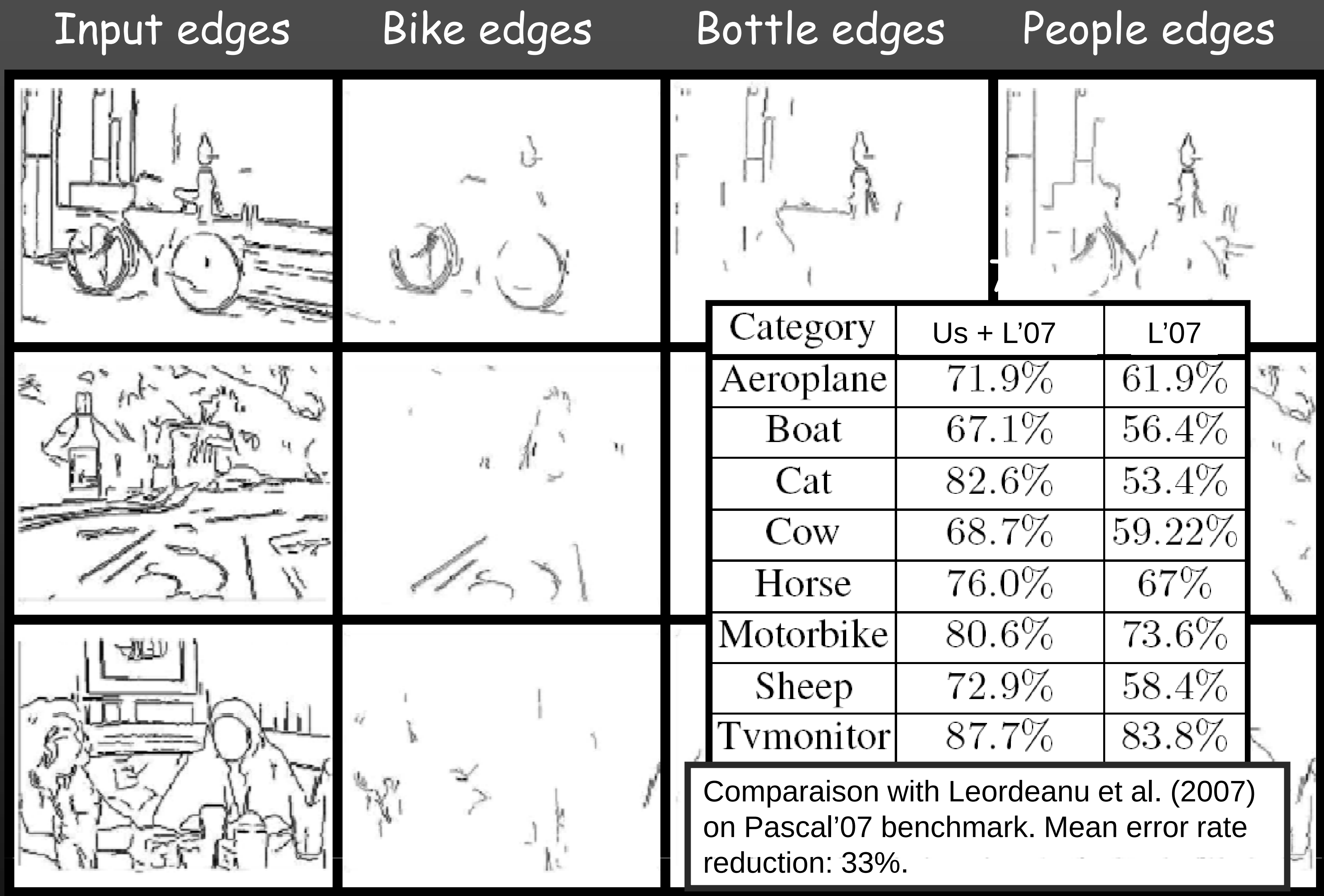
Quantitative results on the Berkeley segmentation dataset and benchmark (Martin et al., ICCV'01)





| Rank | Score | Algorithm |
|---|---|---|
| 0 | 0.79 | Human labeling |
| 1 | 0.70 | (Maire et al., 2008) |
| 2 | 0.67 | (Aerbelaez, 2006) |
| 3 | 0.66 | (Dollar et al., 2006) |
| 3 | 0.66 | Us – no post-processing |
| 4 | 0.65 | (Martin et al., 2001) |
| 5 | 0.57 | Color gradient |
| 6 | 0.43 | Random |






| Category | Us + L'07 | L'07 |
|---|---|---|
| Aeroplane | 71.9% | 61.9% |
| Boat | 67.1% | 56.4% |
| Cat | 82.6% | 53.4% |
| Cow | 68.7% | 59.22% |
| Horse | 76.0% | 67% |
| Motorbike | 80.6% | 73.6% |
| Sheep | 72.9% | 58.4% |
| Tvmonitor | 87.7% | 83.8% |

Comparaison with Leordeanu et al. (2007) on Pascal'07 benchmark. Mean error rate reduction: 33%.

# Task-driven dictionary learning

(Mairal, Bach, Ponce, PAMI'10, in press)

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

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

(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.

Authentic

Fake

(Mairal, Bach, Ponce, 2010)

Data courtesy of James Hughes & Daniel Rockmore

Authentic

Fake

Fake

Data courtesy of James Hughes & Daniel Rockmore

Authentic

Fake

Authentic

Data courtesy of James Hughes & Daniel Rockmore

# A common architecture for image classification

| | | | |
|---|---|---|---|
| Filtering ↓ | ↓ | ↓ |
| | SIFT at keypoints | dense gradients | dense SIFT |
| Coding ↓ | ↓ | ↓ |
| | vector quantization | vector quantization | sparse coding |
| Pooling ↓ | ↓ | ↓ |
| | whole image, mean | coarse grid, mean | 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)

| Method | Caltech-101, 30 training examples | | 15 Scenes, 100 training examples | |
|---|---|---|---|---|
| | Average Pool | Max Pool | Average Pool | Max Pool |
| | Results with basic features, SIFT extracted each 8 pixels | | | |
| Hard quantization, linear kernel | $51.4 \pm 0.9$ [256] | $64.3 \pm 0.9$ [256] | $73.9 \pm 0.9$ [1024] | $80.1 \pm 0.6$ [1024] |
| Hard quantization, intersection kernel | $64.2 \pm 1.0$ [256] **(1)** | $64.3 \pm 0.9$ [256] | $80.8 \pm 0.4$ [256] **(1)** | $80.1 \pm 0.6$ [1024] |
| Soft quantization, linear kernel | $57.9 \pm 1.5$ [1024] | $69.0 \pm 0.8$ [256] | $75.6 \pm 0.5$ [1024] | $81.4 \pm 0.6$ [1024] |
| Soft quantization, intersection kernel | $66.1 \pm 1.2$ [512] **(2)** | $70.6 \pm 1.0$ [1024] | $81.2 \pm 0.4$ [1024] **(2)** | $83.0 \pm 0.7$ [1024] |
| Sparse codes, linear kernel | $61.3 \pm 1.3$ [1024] | $71.5 \pm 1.1$ [1024] **(3)** | $76.9 \pm 0.6$ [1024] | $83.1 \pm 0.6$ [1024] **(3)** |
| Sparse codes, intersection kernel | $70.3 \pm 1.3$ [1024] | $71.8 \pm 1.0$ [1024] **(4)** | $83.2 \pm 0.4$ [1024] | $84.1 \pm 0.5$ [1024] **(4)** |

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

Scenes, supervised dictionary learning

| | Unsup | Discr [1024] | Unsup | Discr [2048] |
|---|---|---|---|---|
| Linear | $83.6 \pm 0.4$ | $84.9 \pm 0.3$ | $84.2 \pm 0.3$ | $85.6 \pm 0.2$ |
| Intersect | $84.3 \pm 0.5$ | $84.7 \pm 0.4$ | $84.6 \pm 0.4$ | $85.1 \pm 0.5$ |

# Learning dictionaries for image classification
(Boureau, LeCun, Bach, Ponce, CVPR'10)

| Method | Caltech-101, 30 training examples | | 15 Scenes, 100 training examples | |
|---|---|---|---|---|
| | Average Pool | Max Pool | Average Pool | Max Pool |
| | Results with basic features, SIFT extracted each 8 pixels | | | |
| Hard quantization, linear kernel | $51.4 \pm 0.9$ [256] | $64.3 \pm 0.9$ [256] | $73.9 \pm 0.9$ [1024] | $80.1 \pm 0.6$ [1024] |
| Hard quantization, intersection kernel | $64.2 \pm 1.0$ [256] **(1)** | $64.3 \pm 0.9$ [256] | $80.8 \pm 0.4$ [256] **(1)** | $80.1 \pm 0.6$ [1024] |
| Soft quantization, linear kernel | $57.9 \pm 1.5$ [1024] | $69.0 \pm 0.8$ [256] | $75.6 \pm 0.5$ [1024] | $81.4 \pm 0.6$ [1024] |
| Soft quantization, intersection kernel | $66.1 \pm 1.2$ [512] **(2)** | $70.6 \pm 1.0$ [1024] | $81.2 \pm 0.4$ [1024] **(2)** | $83.0 \pm 0.7$ [1024] |
| Sparse codes, linear kernel | $61.3 \pm 1.3$ [1024] | $71.5 \pm 1.1$ [1024] **(3)** | $76.9 \pm 0.6$ [1024] | $83.1 \pm 0.6$ [1024] **(3)** |
| Sparse codes, intersection kernel | $70.3 \pm 1.3$ [1024] | $71.8 \pm 1.0$ [1024] **(4)** | $83.2 \pm 0.4$ [1024] | $84.1 \pm 0.5$ [1024] **(4)** |

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

Scenes, supervised dictionary learning

| | Unsup | Discr[1024] | Unsup | Discr[2048] |
|---|---|---|---|---|
| Linear | $83.6 \pm 0.4$ | $84.9 \pm 0.3$ | $84.2 \pm 0.3$ | $85.6 \pm 0.2$ |
| Intersect | $84.3 \pm 0.5$ | $84.7 \pm 0.4$ | $84.6 \pm 0.4$ | $85.1 \pm 0.5$ |

# Non-blind deblurring (Couzinie-Devy, Mairal, Bach, Ponce, 2010)

| | *Cameraman* | | | | | | *Lena* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR input image | 20.76 | 22.35 | 22.29 | 24.7 | 25.53 | 23.44 | 25.84 | 27.57 | 27.35 | 29.00 | 30.74 | 28.97 |
| Richardson-Lucy [20] | 4.47 | 5.53 | 3.58 | 0.49 | 1.21 | 1.04 | 4.80 | 5.29 | 2.71 | 0.02 | 0.26 | 0.53 |
| Sparse gradient [14] | 7.73 | 6.89 | 4.78 | 2.24 | 2.64 | 2.70 | 7.02 | 2.83 | 5.44 | 4.06 | 3.30 | 3.33 |
| SA-DCT [9] | 8.55 | 8.11 | 6.33 | 3.37 | - | - | 7.79 | 7.55 | 6.10 | 4.49 | 3.56 | 3.46 |
| Dabov et al. [3] | 8.34 | 8.19 | 6.40 | 3.34 | - | - | 7.97 | 7.95 | 6.53 | 4.81 | 4.03 | 3.91 |
| Linear | 2.92 | 7.11 | 5.68 | 3.08 | 3.39 | 2.81 | 2.91 | 7.11 | 5.58 | 4.56 | 3.82 | 3.48 |
| Linear + Dictionary | 4.49 | 8.11 | 6.61 | 3.38 | 3.71 | 3.29 | 4.41 | 8.10 | 6.78 | 5.06 | 4.35 | 4.00 |

# Non-blind deblurring (Couzinie-Devy, Mairal, Bach, Ponce, 2010)

BUT
on anisotropic
kernels
(Levin et al., 2009)

| Kernel | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| (Levin et al., 2008) | 7.92 | **8.20** | **7.58** | **12.29** |
| Ours | **8.47** | 7.85 | 7.54 | 5.93 |
| Kernel | 5 | 6 | 7 | 8 |
| (Levin et al., 2008) | 9.52 | **13.02** | **12.94** | **11.38** |
| Ours | **9.91** | 8.43 | 8.51 | 7.20 |

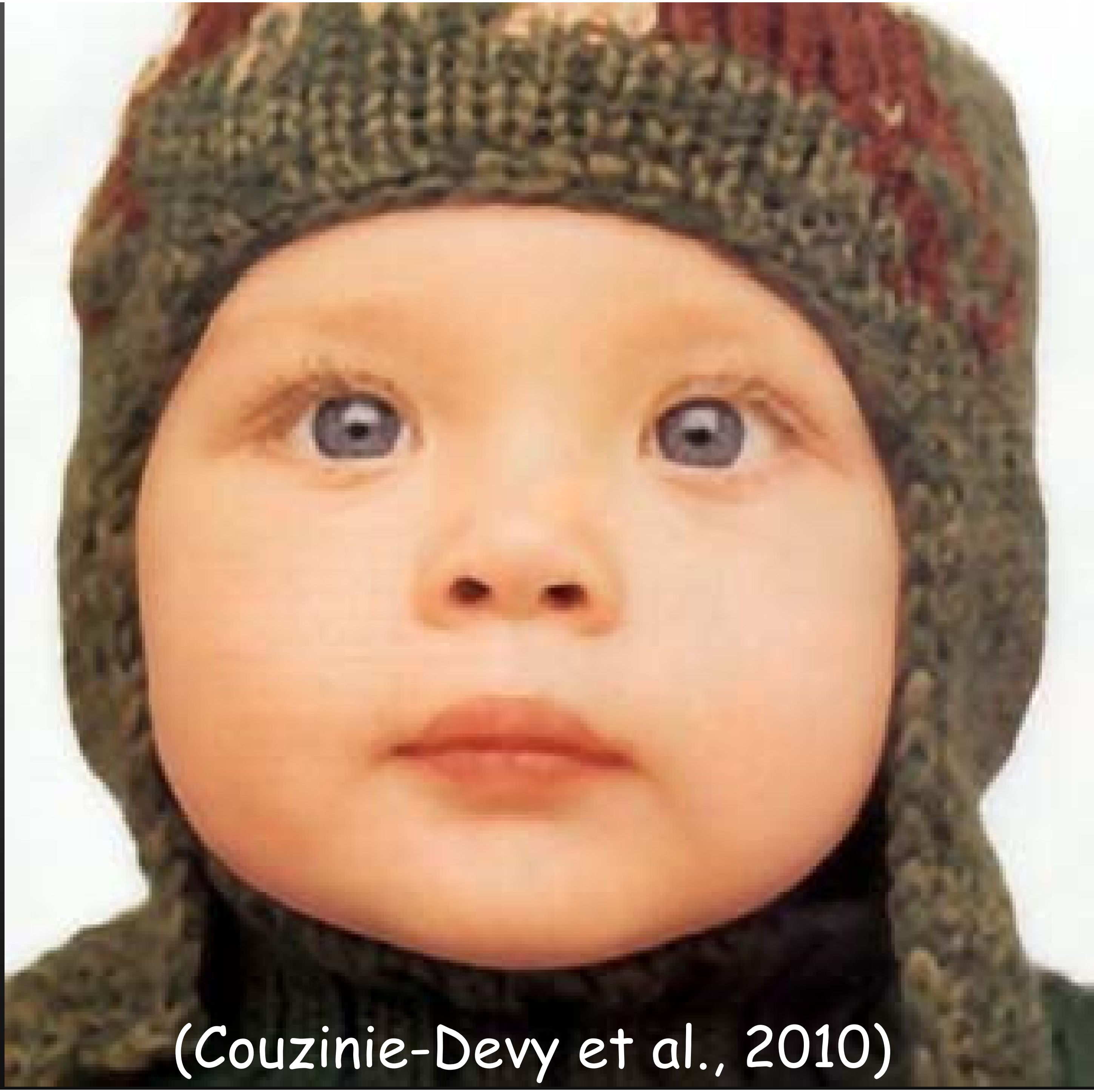
# Digital zoom (Couzinie-Devy, Mairal, Bach, Ponce, 2010)

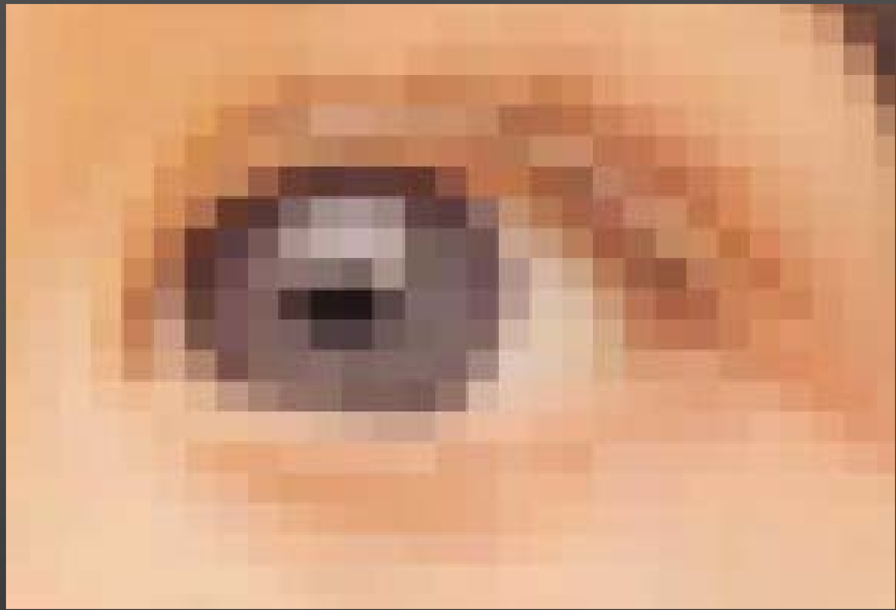
| | Cubic spline | Yang et al.'09 | Yu et al.'10 | Ours |
|---|---|---|---|---|
| Lena | 31.60 | 30.64 | 33.78 | **34.76** |
| Girl | 30.62 | 30.43 | 31.82 | **34.27** |
| Flower | 37.02 | 35.96 | 39.06 | **40.07** |


Digital Zoom


(Fattal, 2007)


(Glasner et al., 2009)


(Couzinie-Devy et al., 2010)

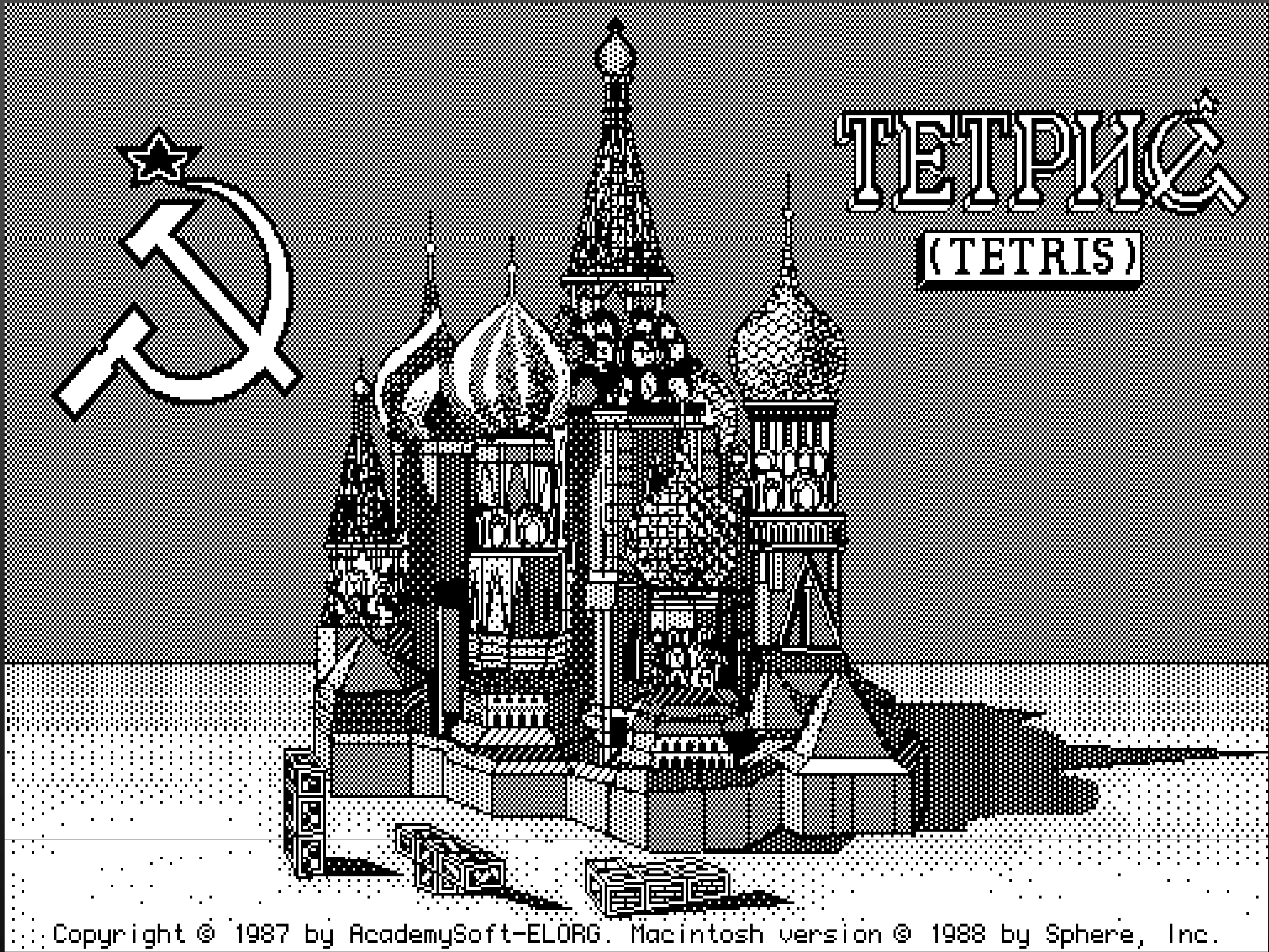


(Fattal, 2007)

(Glasner et al., 2009)

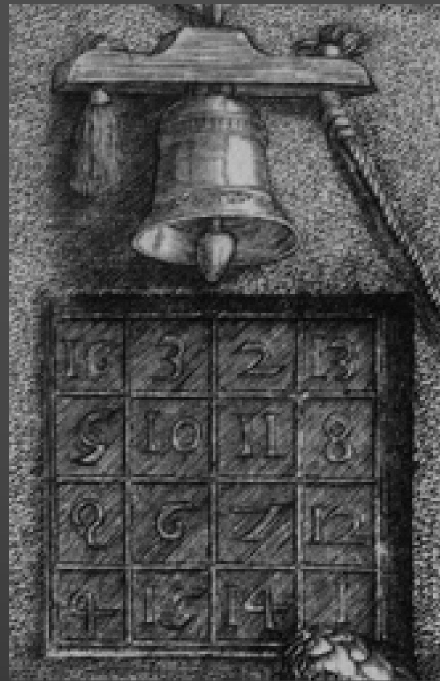
(Couzinie-Devy et al., 2010)

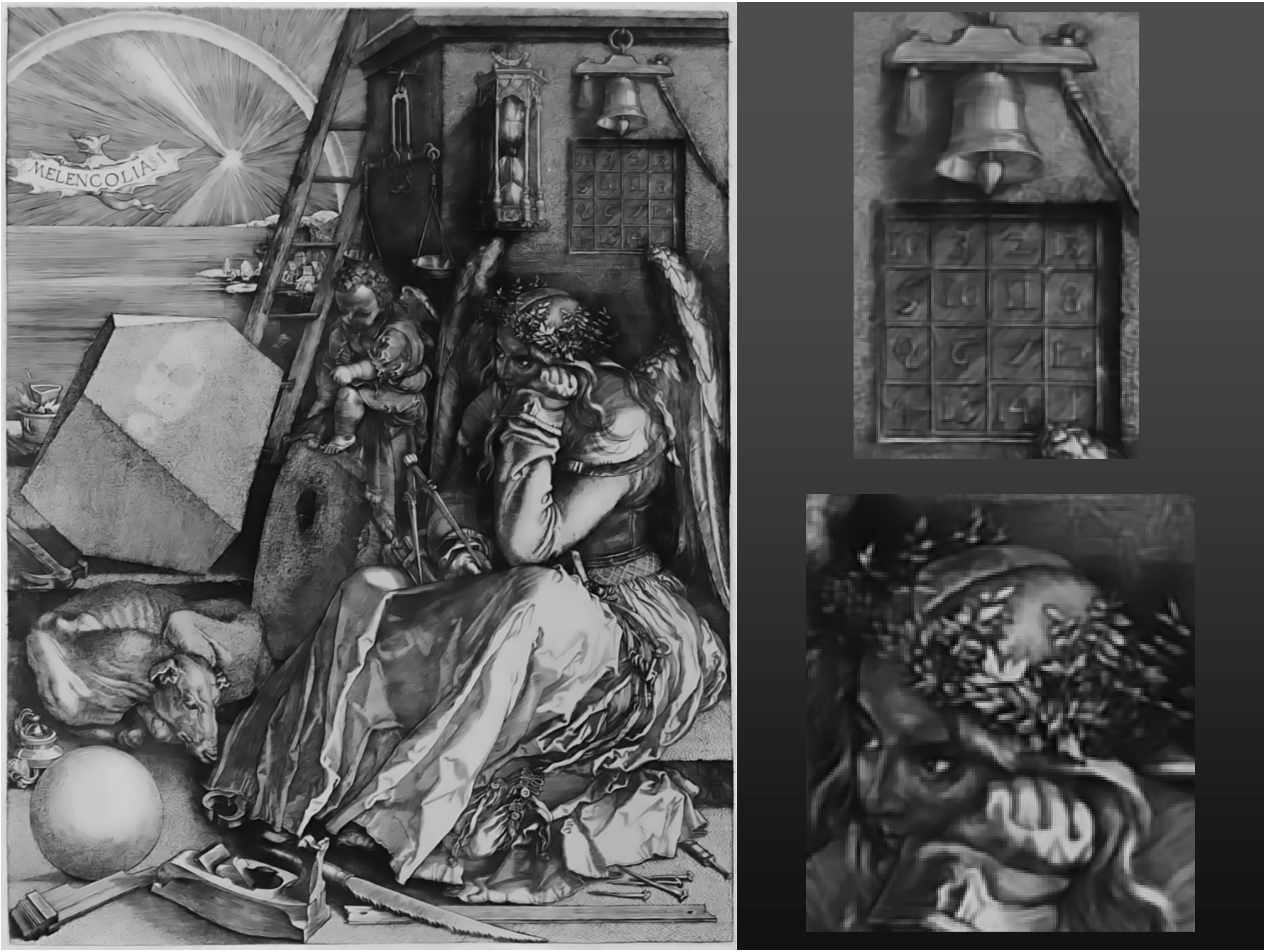
# Inverse halftoning
(Mairal, Bach, Ponce, 2010)



# Inverse halftoning
(Mairal, Bach, Ponce, 2010)


MELENCOLIA I


16 3 2 13
5 10 11 8
9 6 7 12
4 15 14 1

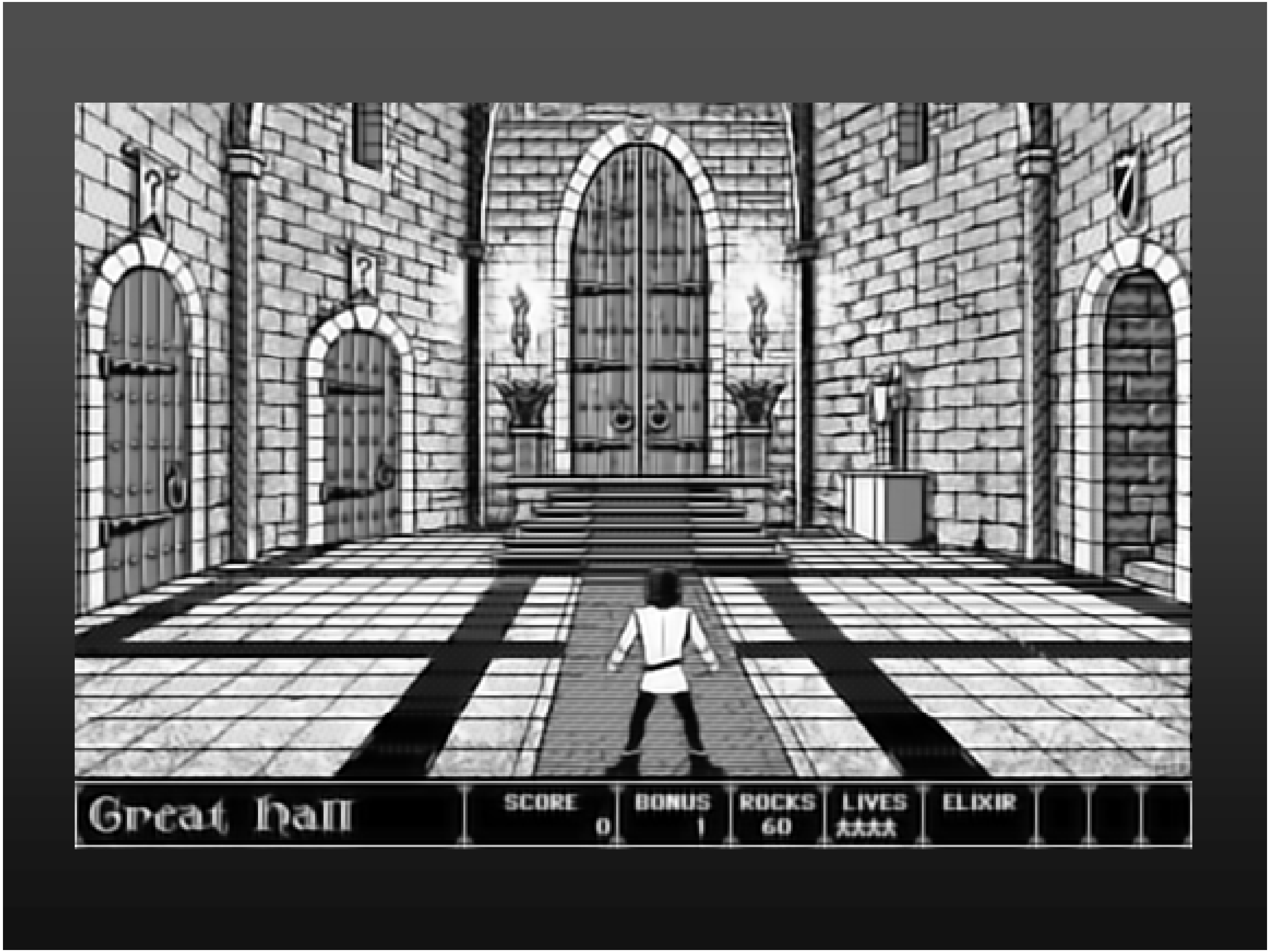

MELENCOLIA I

| | Validation set | | | | Test set | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Image | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| FIHT2 | 30.8 | 25.3 | 25.8 | 31.4 | 24.5 | 28.6 | 29.5 | 28.2 | 29.3 | 26.0 | 25.2 | 24.7 |
| WInHD | 31.2 | 26.9 | 26.8 | 31.9 | 25.7 | 29.2 | 29.4 | 28.7 | 29.4 | 28.1 | 25.6 | 26.4 |
| LPA-ICI | 31.4 | 27.7 | 26.5 | 32.5 | 25.6 | 29.7 | 30.0 | 29.2 | 30.1 | 28.3 | 26.0 | 27.2 |
| SA-DCT | 32.4 | 28.6 | 27.8 | 33.0 | 27.0 | 30.1 | 30.2 | 29.8 | 30.3 | 28.5 | 26.2 | 27.6 |
| Ours | 33.0 | 29.6 | 28.1 | 33.0 | 26.6 | 30.2 | 30.5 | 29.9 | 30.4 | 29.0 | 26.2 | 28.0 |

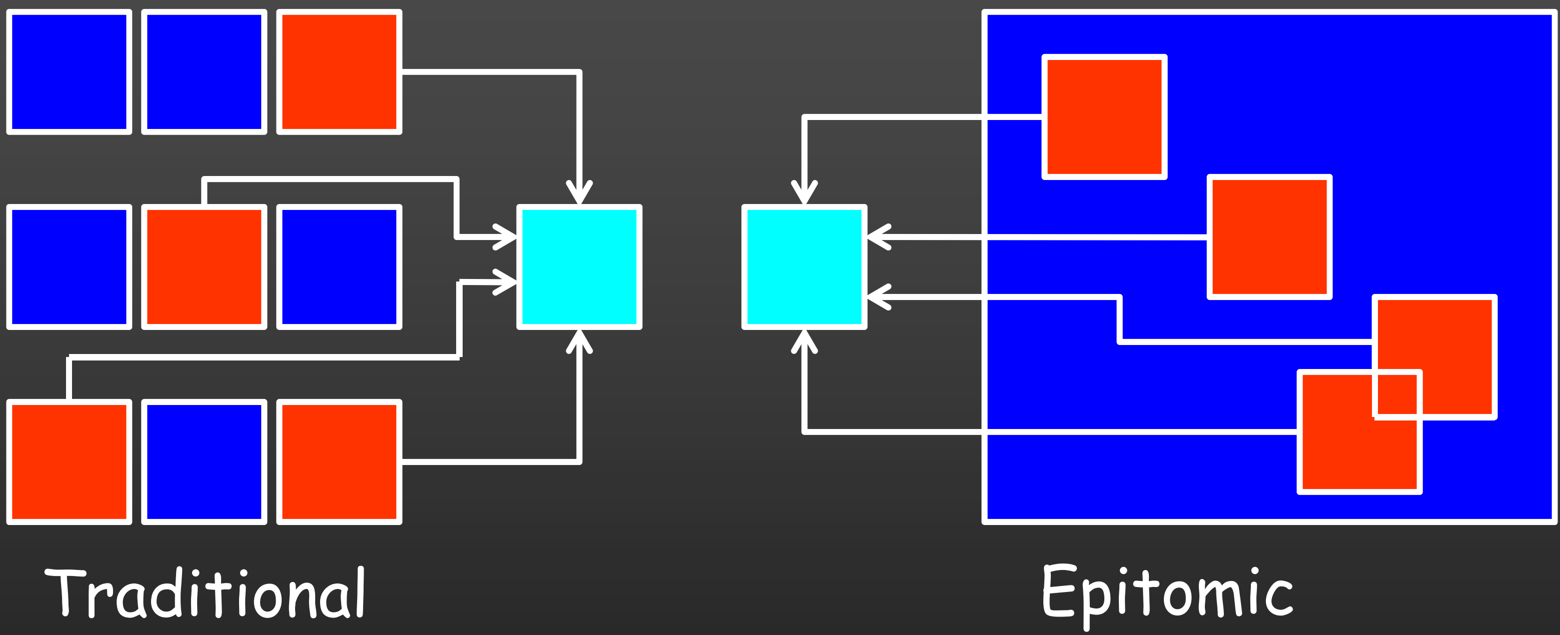
PSNR comparison between our method and Kite et al.'00 [FIHT2]; Neelamini et al.'09 [WInHD]; Foi et al.'04 [LPA-ICI]; and Dabov et al.'06 [SA-DCT].


Great hall
SCORE
0
BONUS
1
ROCKS
60
LIVES
ELIXIR


Great hall
SCORE
0
BONUS
1
ROCKS
60
LIVES
ELIXIR

# Epitomic dictionaries

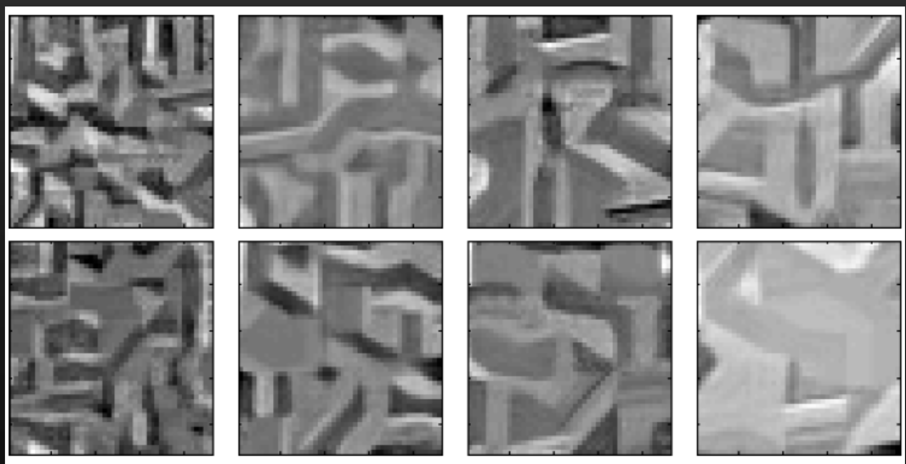
(Benoit, Mairal, Bach, Ponce, CVPR'10)




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

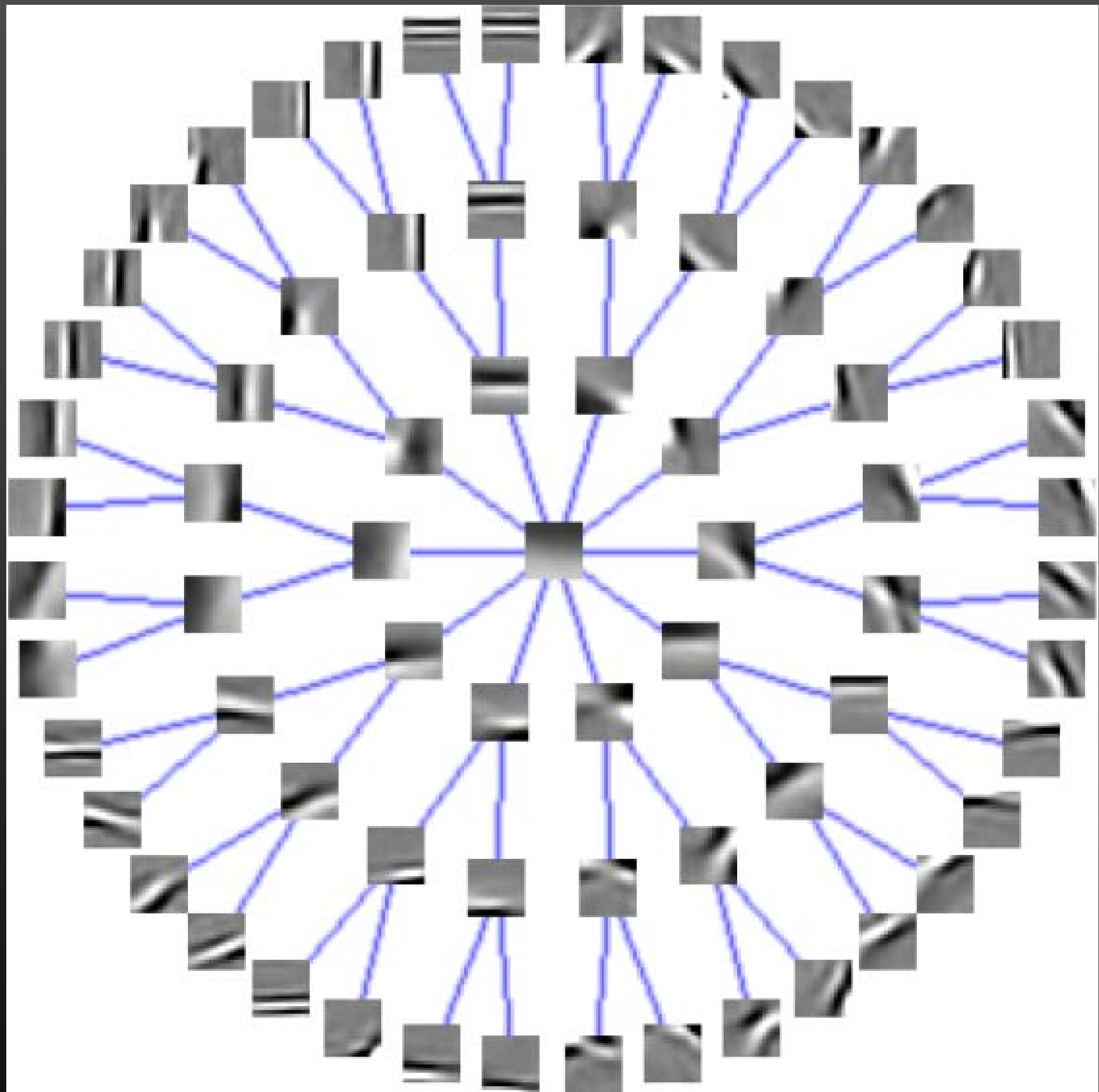


## Denoising experiment

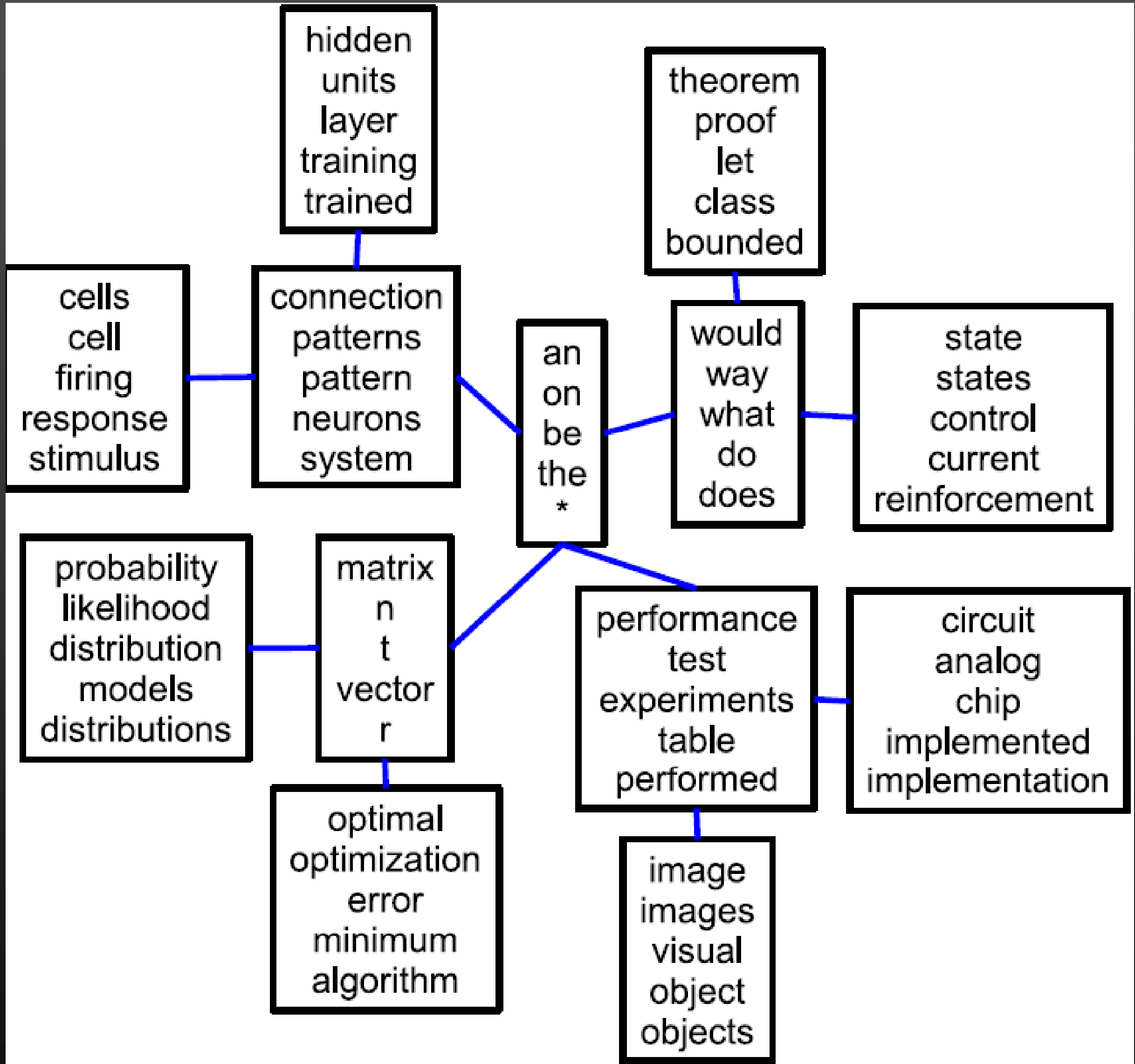
| Image | $\sigma$ | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| house | 2E | 35.89 | **34.33** | **33.25** | **32.03** |
| | 1E | 35.89 | 34.31 | 33.07 | 31.90 |
| | ISD | **36.05** | 34.25 | 32.72 | 31.76 |
| | DL | 35.63 | 33.43 | 32.01 | 30.77 |
| barbara | 2E | 34.07 | 33.91 | 30.43 | **29.24** |
| | 1E | 33.99 | 31.83 | 30.35 | 29.15 |
| | ISD | **34.21** | **32.22** | **30.71** | 29.22 |
| | DL | 34.00 | 31.71 | 30.20 | 28.94 |
| lena | 2E | **35.44** | 33.62 | 32.27 | **31.37** |
| | 1E | 35.41 | **33.67** | **32.35** | 31.34 |
| | ISD | 35.42 | 33.64 | 32.25 | 31.09 |
| | DL | 35.17 | 33.23 | 31.73 | 30.64 |
| boat | 2E | **33.66** | 31.72 | 30.33 | **29.33** |
| | 1E | 33.62 | 31.70 | 30.36 | 29.30 |
| | ISD | 33.64 | **31.79** | **30.41** | 28.45 |
| | DL | 33.49 | 31.50 | 29.99 | 28.91 |
| peppers | 2E | **34.46** | **32.37** | **30.93** | 29.70 |
| | 1E | 34.37 | 32.33 | 30.89 | **29.79** |
| | ISD | 34.23 | 32.30 | 30.69 | 29.44 |
| | DL | 33.92 | 31.76 | 30.20 | 29.03 |

ISD = (Aharon & Elad'08)
DL=flat dict. learning

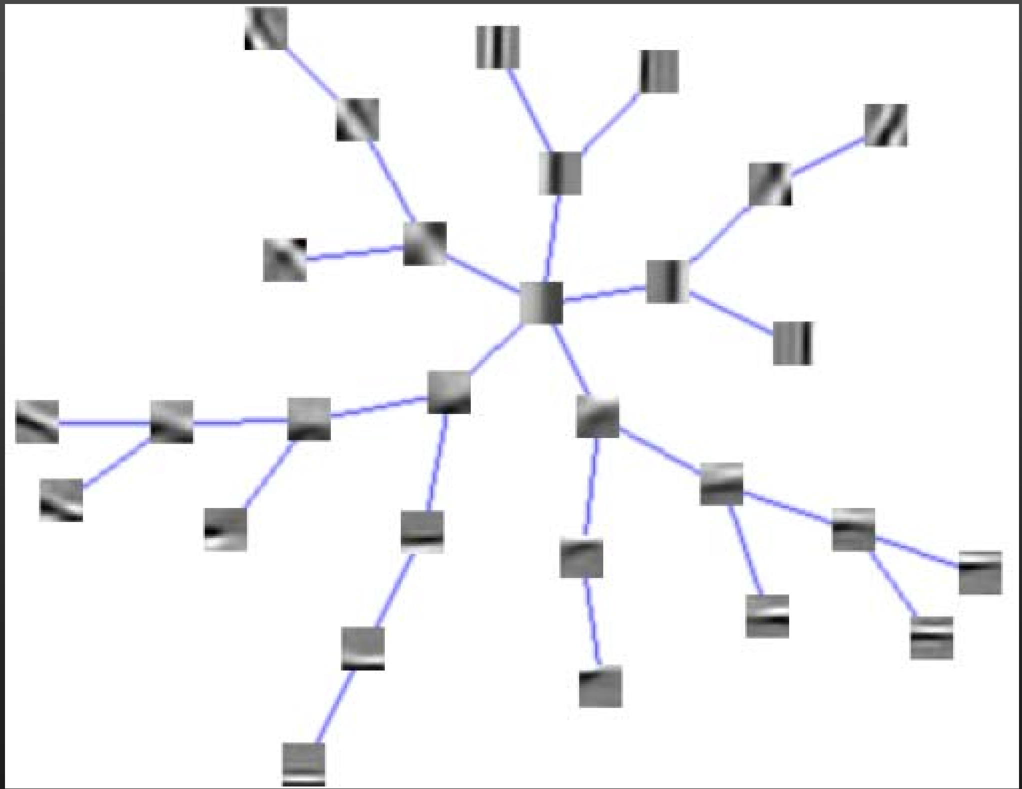
# Proximal methods for sparse hierarchical dictionary learning
(Jenatton, Mairal, Obozinski, Bach, ICML'10)

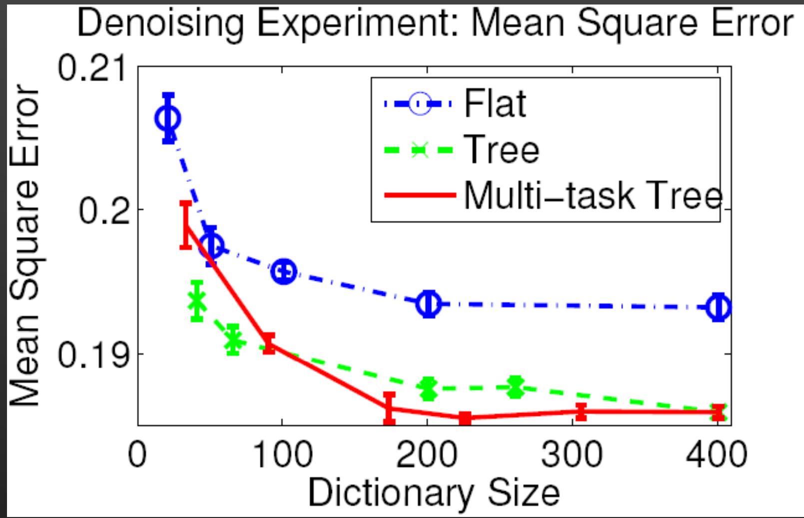


# Proximal methods for sparse hierarchical dictionary learning
(Jenatton, Mairal, Obozinski, Bach, ICML'10)



# 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 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. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

D. P. Bertsekas. *Nonlinear programming*. Athena Scientific Belmont, Mass, 1999.

D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

D. Blei, T. Griffiths, and M. Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2): 1–30, 2010.

J.F. Bonnans, J.C. Gilbert, C. Lemarechal, and C.A. Sagastizabal. *Numerical optimization: theoretical and practical aspects*. Springer-Verlag New York Inc, 2006.

# References II


J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization: Theory and examples*. Springer, 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.

Y-L. Boureau, F. Bach, Y. Lecun, and J. Ponce. Learning mid-level features for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

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.

E. Candes. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, volume 3, 2006.

S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1999.

# References III


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.

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57: 1413–1457, 2004.

I. Daubechies, R. DeVore, M. Fornasier, and S. Gunturk. Iteratively re-weighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math*, 2009.

D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of statistics*, 32(2):407–499, 2004.

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 IV


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.

J. Friedman, T. Hastie, H. Hölfling, and R. Tibshirani. Pathwise coordinate optimization. *Annals of statistics*, 1(2):302–332, 2007.

W. J. Fu. Penalized regressions: The bridge versus the Lasso. *Journal of computational and graphical statistics*, 7:397–416, 1998.

R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In *Proceedings of the Twenty-third Conference on Uncertainty in Artificial Intelligence*, 2007.

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.

# References V

K. Huang and S. Aviyente. Sparse representation for signal classification. In *Advances in Neural Information Processing Systems*, Vancouver, Canada, December 2006.

J. M. Hugues, D. J. Graham, and D. N. Rockmore. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proceedings of the National Academy of Science, TODO USA*, 107(4):1279–1283, 2009.

R. Jenatton, J-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. Technical report, 2009. preprint arXiv:0904.3523v1.

R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2010a.

R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. Technical report, 2010b. submitted, arXiv:1009.3139.

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, 2001.

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 VI


M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

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

J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008a.

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

J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce. Discriminative sparse image models for class-specific edge detection and image interpretation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008c.

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 2008d.

# References VII


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.

S. Mallat and Z. Zhang. Matching pursuit in a time-frequency dictionary. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

H. M. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, 3:111–133, 1956.

N. Meinshausen and P. Buehlmann. Stability selection. *Journal of the Royal Statistical Society, Series B*, 2010. to appear.

Y. Nesterov. A method for solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math. Dokl.*, 27:372–376, 1983.

Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, CORE, 2007.

# References VIII


J. Nocedal and SJ Wright. *Numerical Optimization*. Springer: New York, 2006. 2nd Edition.

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. R. Osborne, B. Presnell, and B. A. Turlach. On the Lasso and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–37, 2000.

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.

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.

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.

# References IX


V. Roth and B. Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. Collaborative hierarchical sparse modeling. Technical report, 2010a. Preprint arXiv:1003.0400v1.

P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. Collaborative hierarchical sparse modeling. Technical report, 2010b. Preprint arXiv:1003.0400v1.

R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.

S. Weisberg. *Applied Linear Regression*. Wiley, New York, 1980.

J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2005.

J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

# References X

J. Yang, K. Yu, , and T. Huang. Supervised translation-invariant sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, 68:49–67, 2006.

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.