Applied Bayesian Nonparametrics

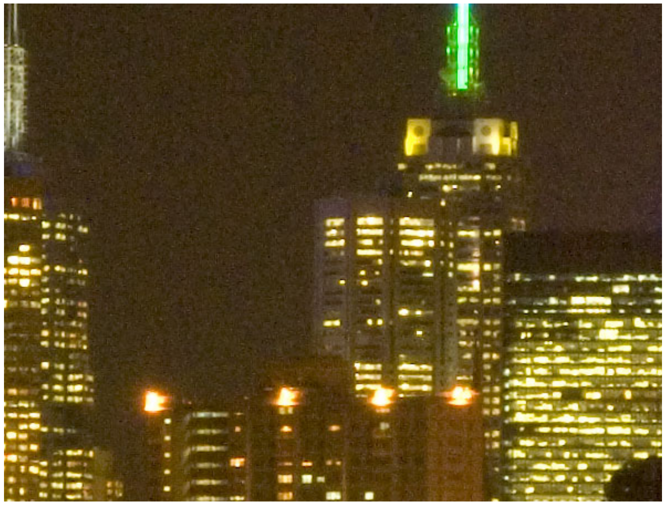
4. Infinite Hidden Markov Trees

Tutorial at CVPR 2012

Erik Sudderth
Brown University

Work by J. Kivinen, E. Sudderth, & M. Jordan
ICCV 2007: *Learning Multiscale Representations of Natural Scenes using Dirichlet Processes*
ICIP 2007: *Image Denoising with Nonparametric Hidden Markov Trees*

Low-level Image Analysis



Noise Removal

Deblurring

Inpainting & Restoration

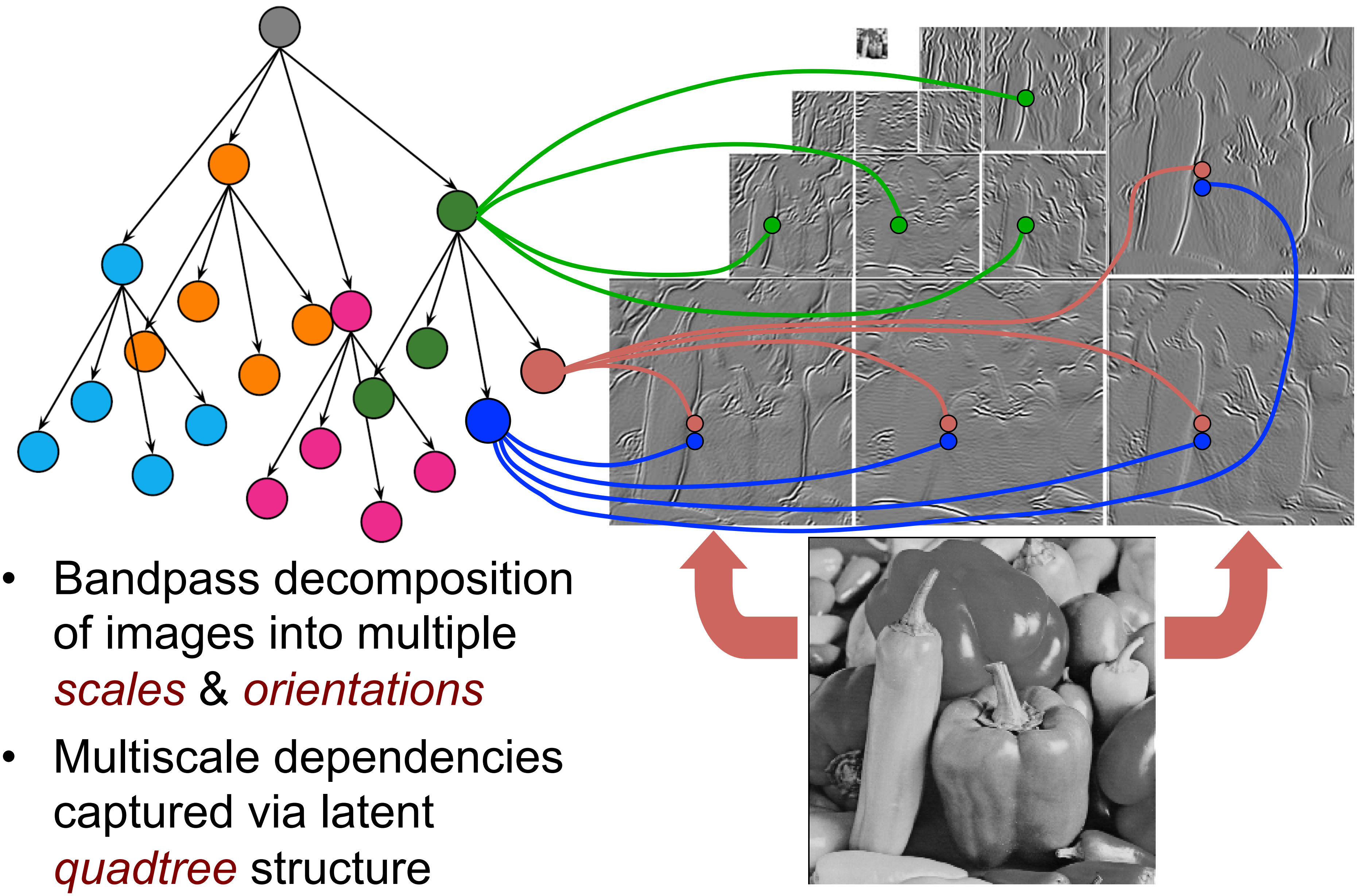
What are the statistical properties of natural images?

Natural Scene Categorization

Coast Forest Open Country Street Tall Building

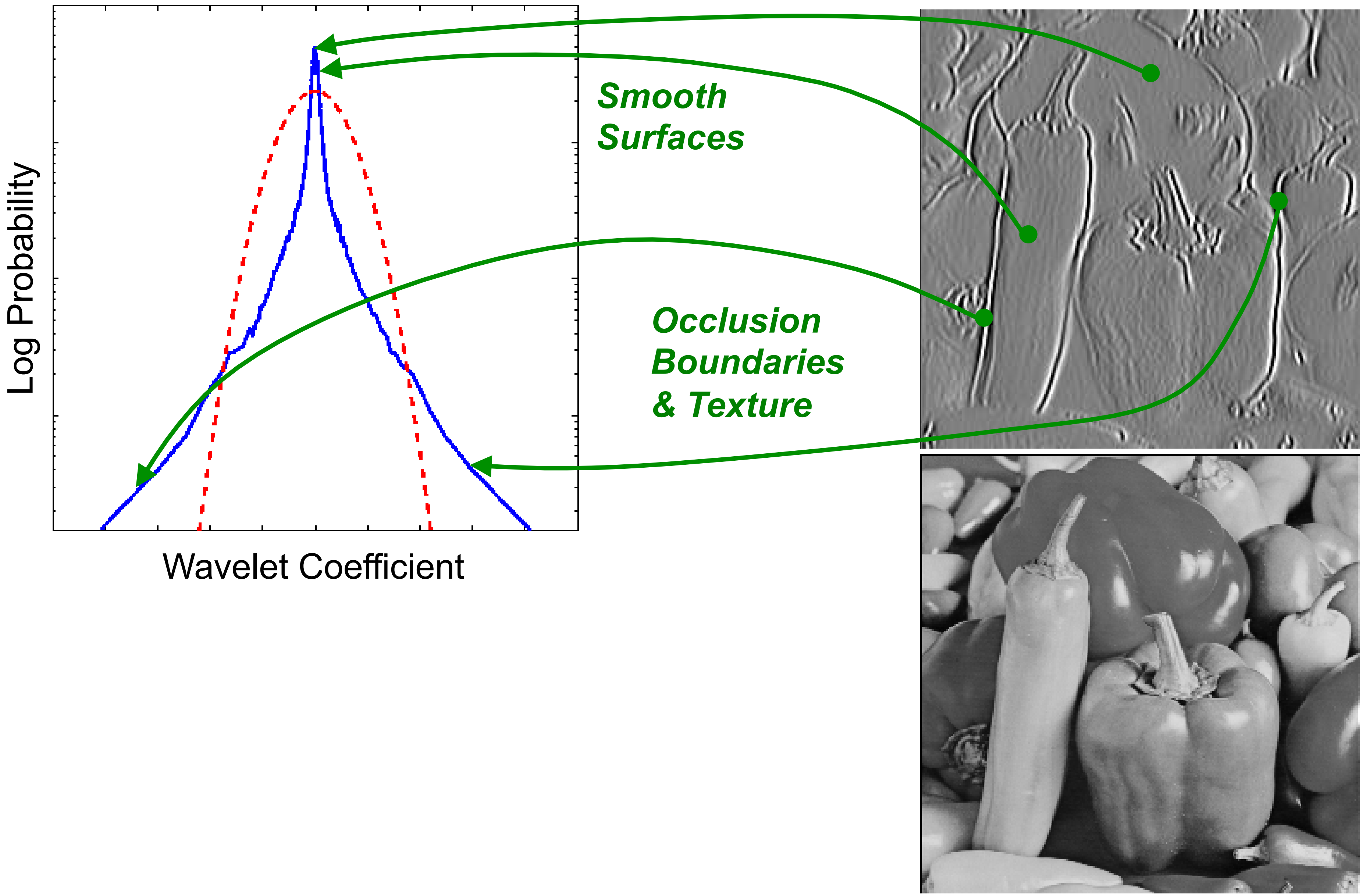
How do semantic labels affect these properties?

Wavelet Decompositions

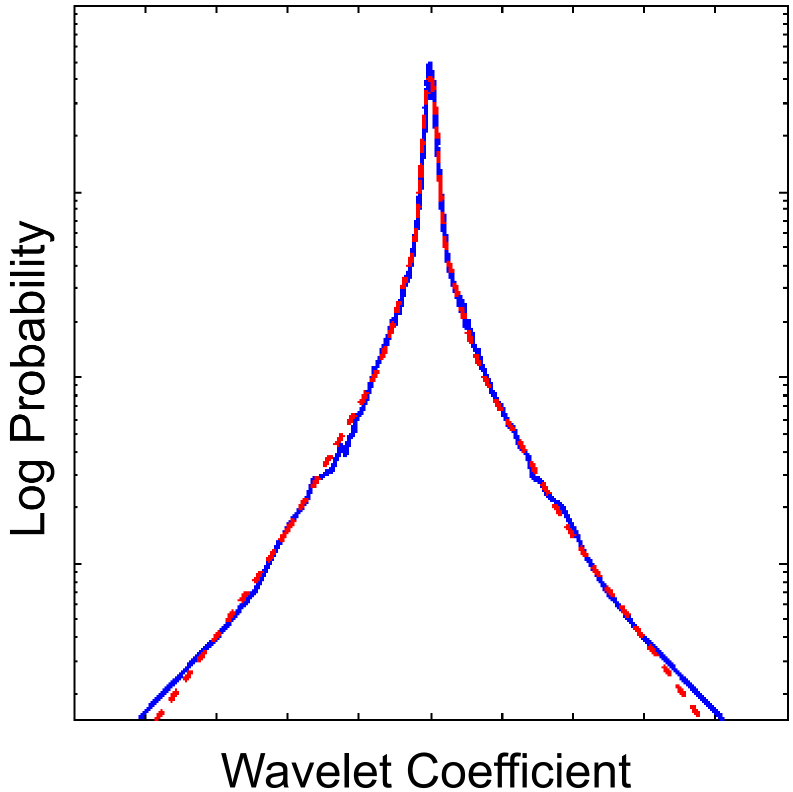


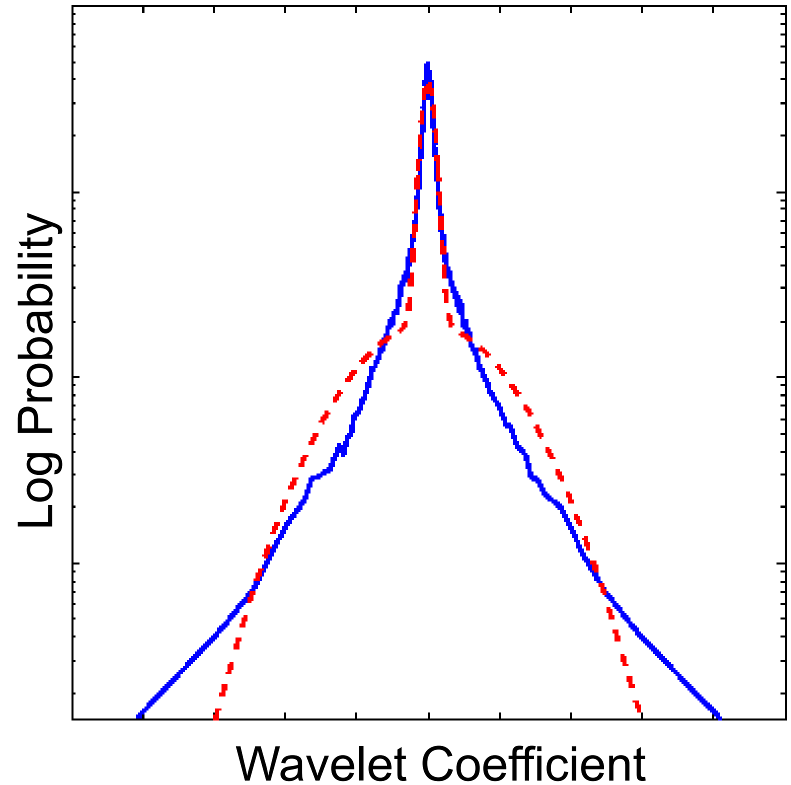
- Bandpass decomposition of images into multiple *scales* & *orientations*
- Multiscale dependencies captured via latent *quadtree* structure

Wavelets: Marginal Statistics



Gaussian Mixture Models





$$x_i = v_i u_i$$

$$v_i \geq 0 \qquad u_i \sim \mathcal{N}(0, \Lambda)$$

Gaussian Scale Mixture (GSM)

Wainwright & Simoncelli, 2000

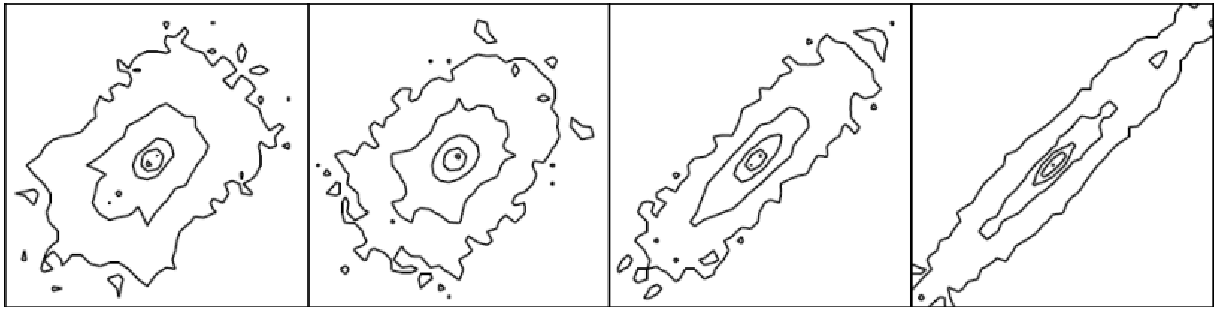
$$x_i \sim \pi \, \mathcal{N}(0, \Lambda_0) + (1-\pi)\mathcal{N}(0, \Lambda_1)$$

Binary Gaussian Mixture

Computational advantages…

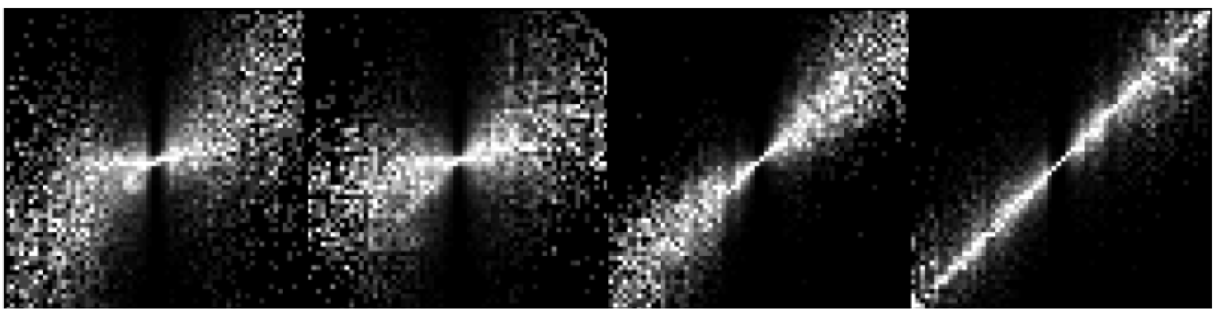
Wavelets: Joint Statistics

Pairwise Joint Histograms:

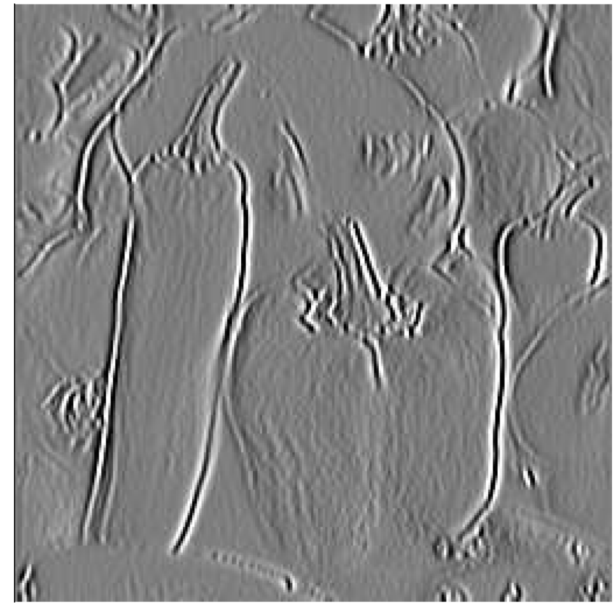


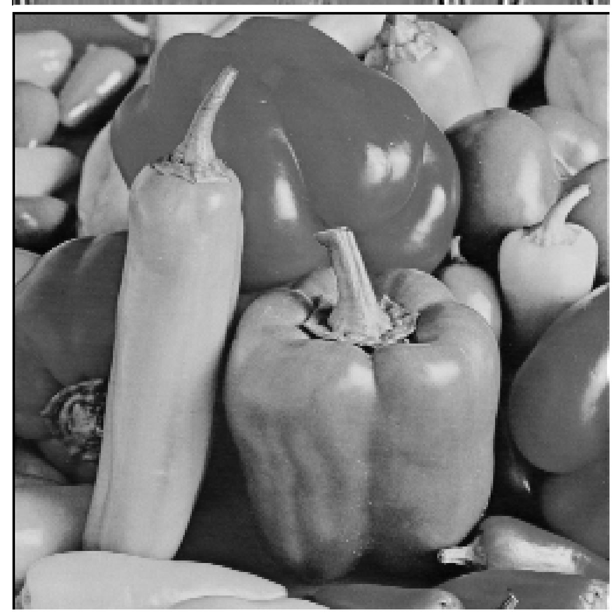
Orientation *Scale* *Vertical* *Horizontal*

Pairwise Conditional Histograms:



Orientation *Scale* *Vertical* *Horizontal*



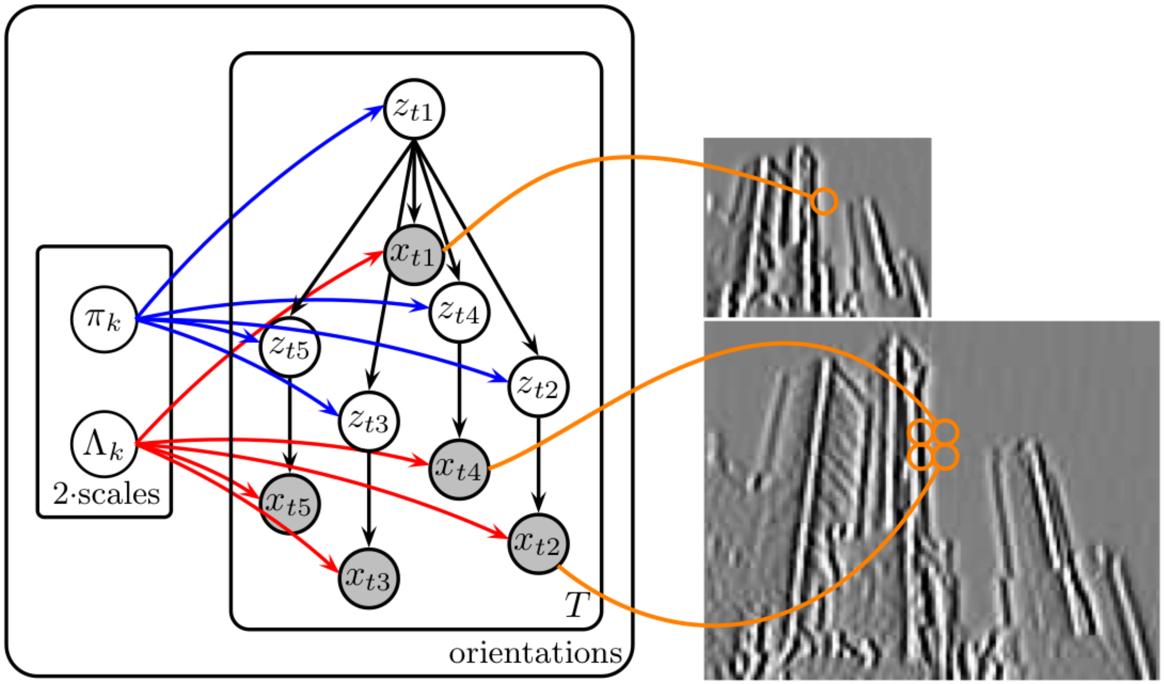


Large magnitude wavelet coefficients…

- *Persist* across multiple scales
- *Cluster* at adjacent spatial locations

Binary Hidden Markov Trees

Crouse, Nowak, & Baraniuk, 1998



π_k → state *transition* distributions
$z_{ti} \sim \pi_{z_{\mathrm{Pa}(ti)}}$

Λ_k → state-specific *emission* covariances
$x_{ti} \sim \mathcal{N}\left(0, \Lambda_{z_{ti}}\right)$

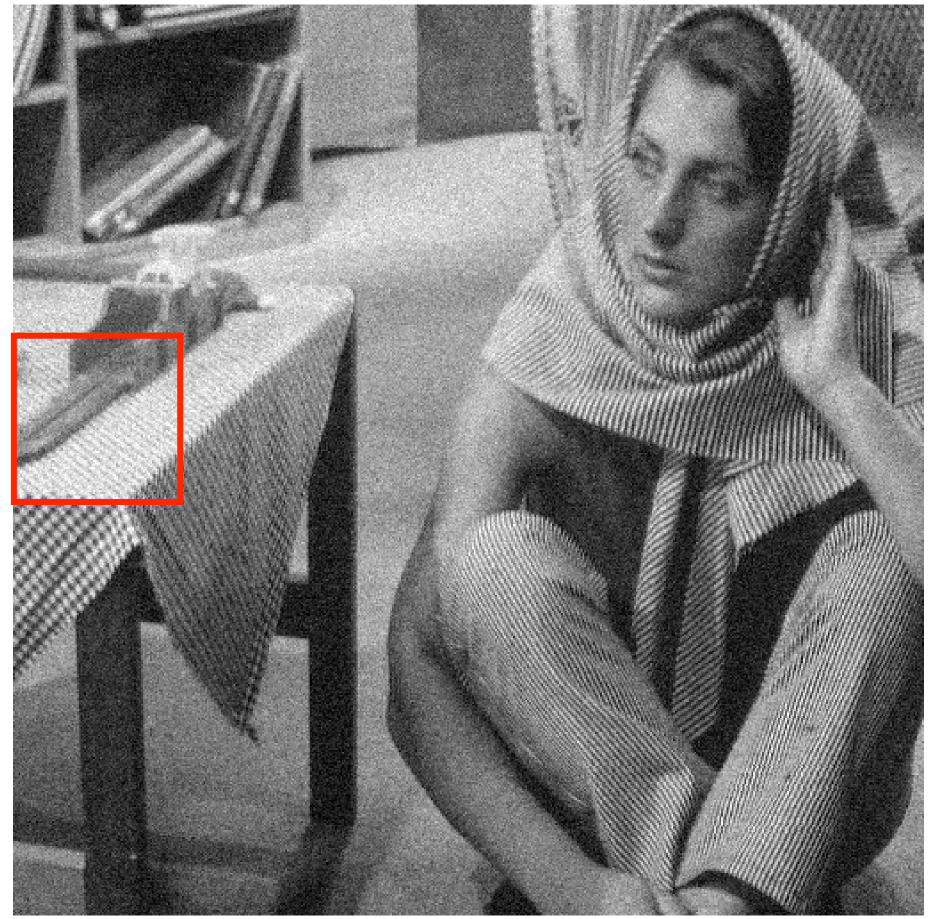
z_{ti} → hidden *state* or cluster assignment
$z_{ti} \in \{0, 1\}$

x_{ti} → *observed* wavelet coefficient

- Coefficients marginally distributed as mixtures of two Gaussians
- Markov dependencies between hidden states capture persistence of image contours across locations and scales
- Each orientation is modeled independently

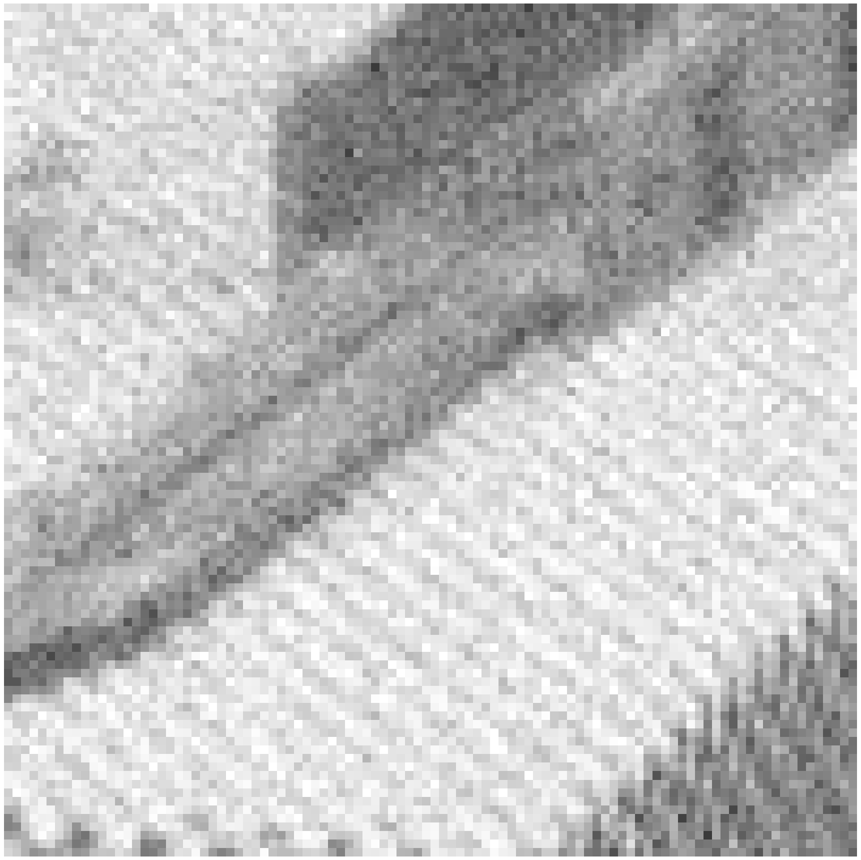
Validation : Image Denoising

Original

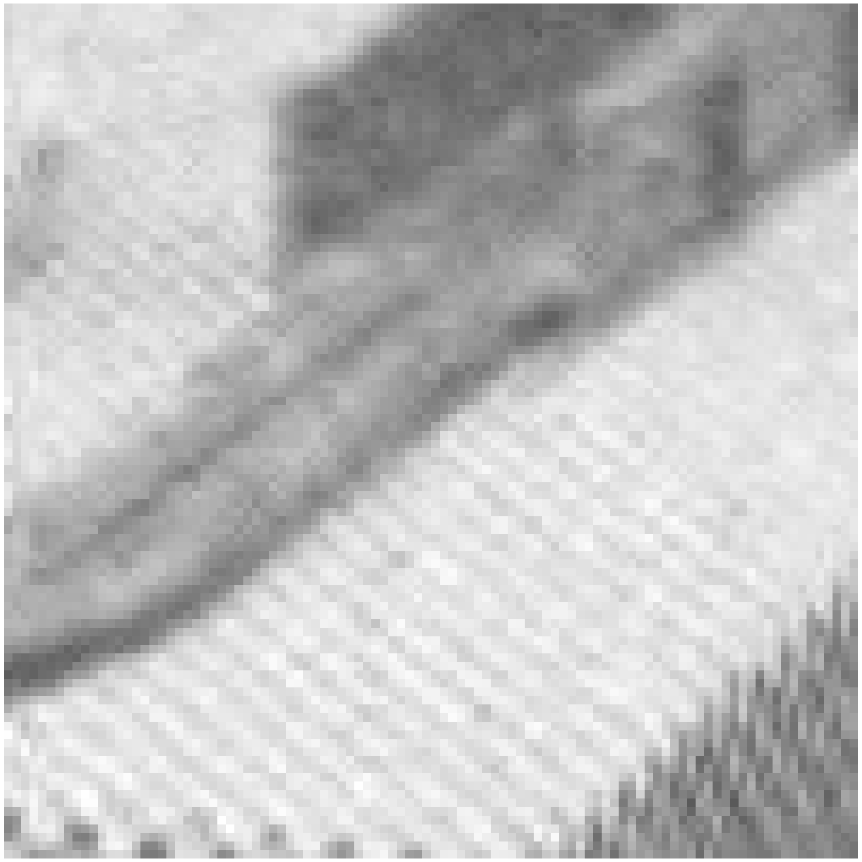


Corrupted by Additive White Gaussian Noise (PSNR = 24.61 dB)

Denoising with Binary HMTs



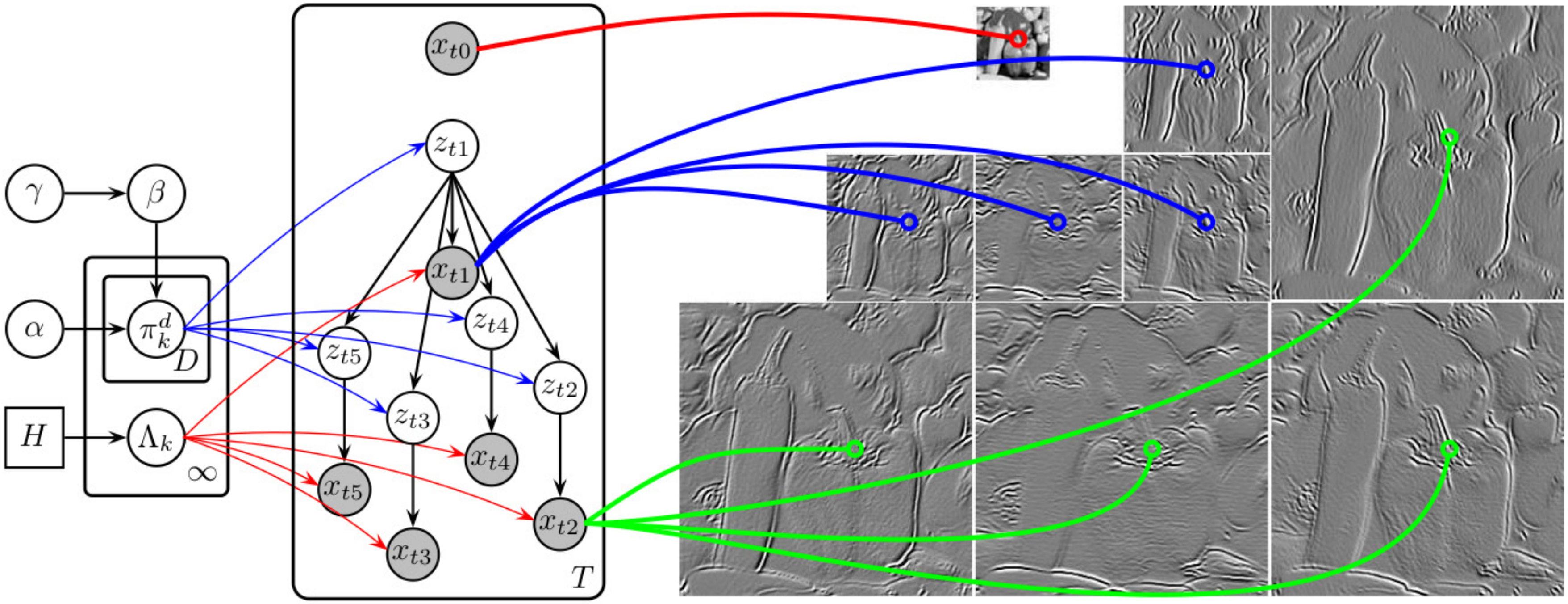
Noisy Input



Denoised (EM algorithm)

- Is two states per scale sufficient? How many is enough?
- Should states be shared the same way for all images, or for all wavelet decompositions?

Hierarchical Dirichlet Process Hidden Markov Trees



$z_{ti} \longrightarrow$ indexes *infinite* set of hidden states

$$z_{ti} \in \{1, 2, 3, \ldots\}$$

$\pi_k \longrightarrow$ infinite set of state *transition* distributions

$$z_{ti} \sim \pi_{z_{\mathrm{Pa(ti)}}}^{d_{ti}}$$

$x_{ti} \longrightarrow$ observed *vector* of wavelet coefficients

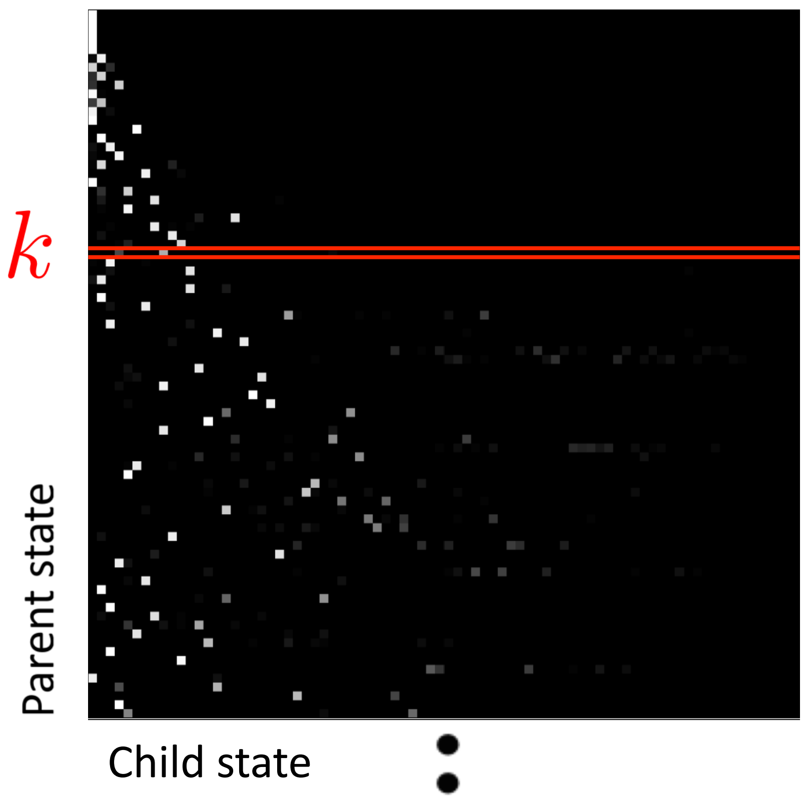
$\Lambda_k \longrightarrow$ state-specific *emission* covariances

$$x_{ti} \sim \mathcal{N}\left(0, \Lambda_{z_{ti}}\right)$$
$$\Lambda_k \sim H$$

Why a Hierarchical DP ? *(Teh et. al. 2004)*

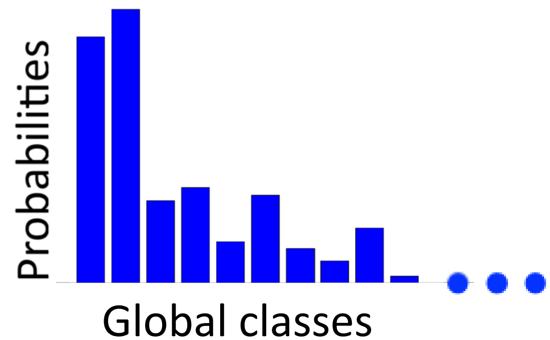
- Hierarchical DP prior allows us to learn a potentially infinite set of *appearance patterns* from natural images
- Hierarchical coupling ensures, with high probability, that a common set of *child* states are reachable from each *parent*

$$\pi_k^{d_{ti}}(\ell) = \Pr\left[z_{ti} = \ell \mid z_{\mathrm{Pa}(ti)}\right]$$



$$\beta \sim \mathrm{Stick}(\gamma)$$

Average state frequencies



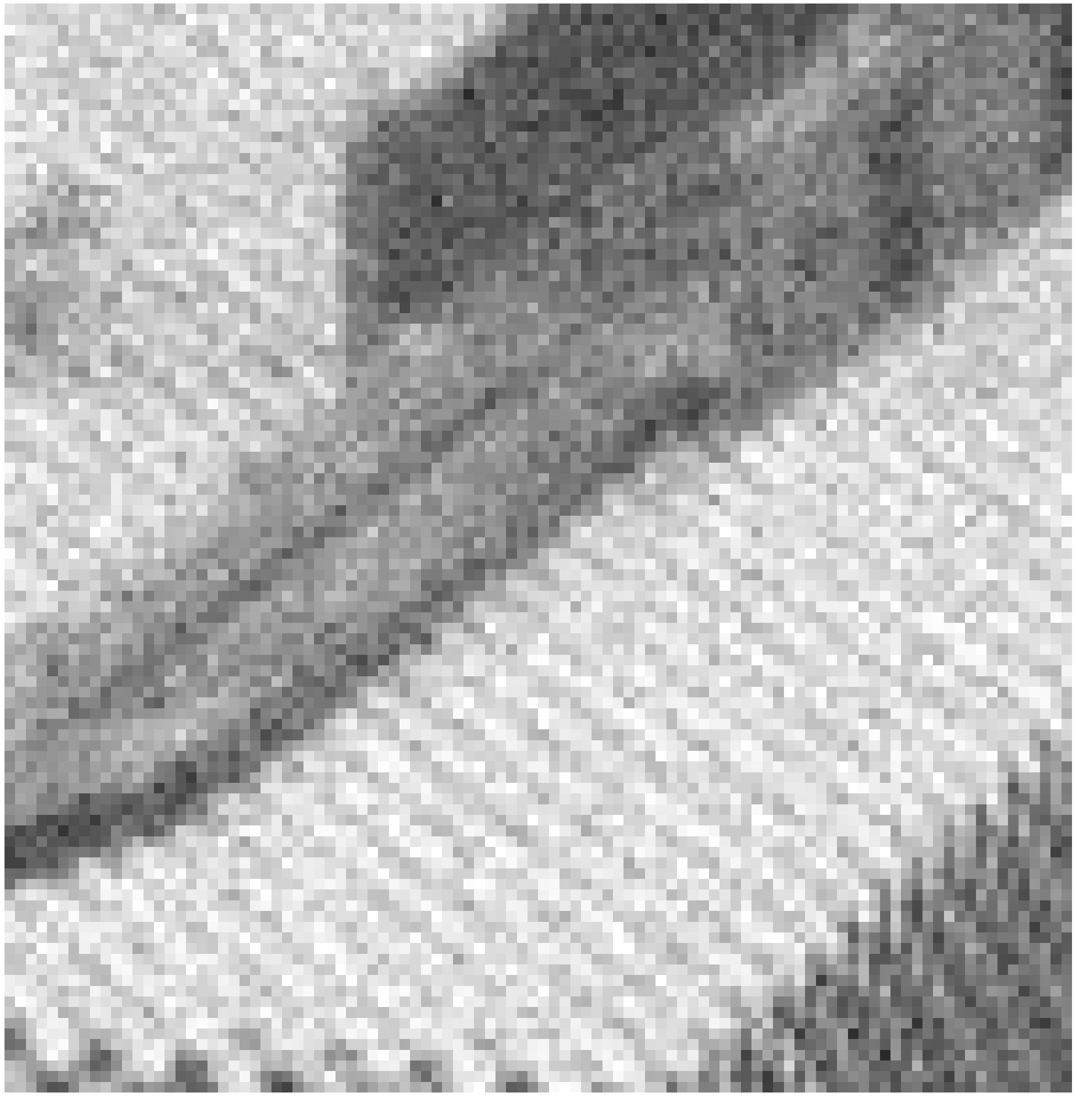
$$\pi_k^d \sim \mathrm{DP}(\alpha, \beta)$$

Transition distributions

$$\mathbb{E}\left[\pi_k^d\right] = \beta$$

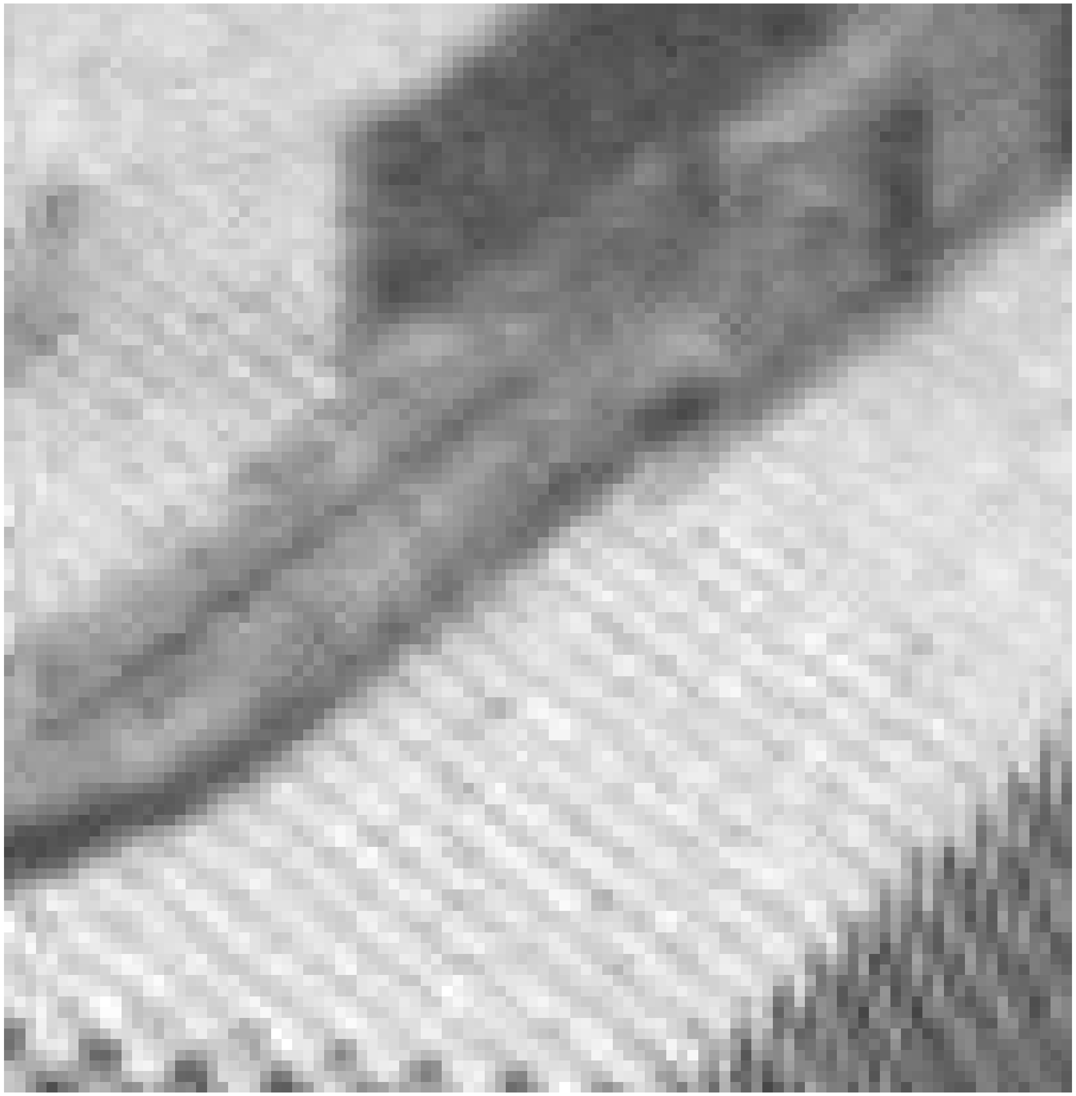
$\alpha \longrightarrow$ *Sparsity & variability of transition distributions*

Denoising: Input



24.61 dB

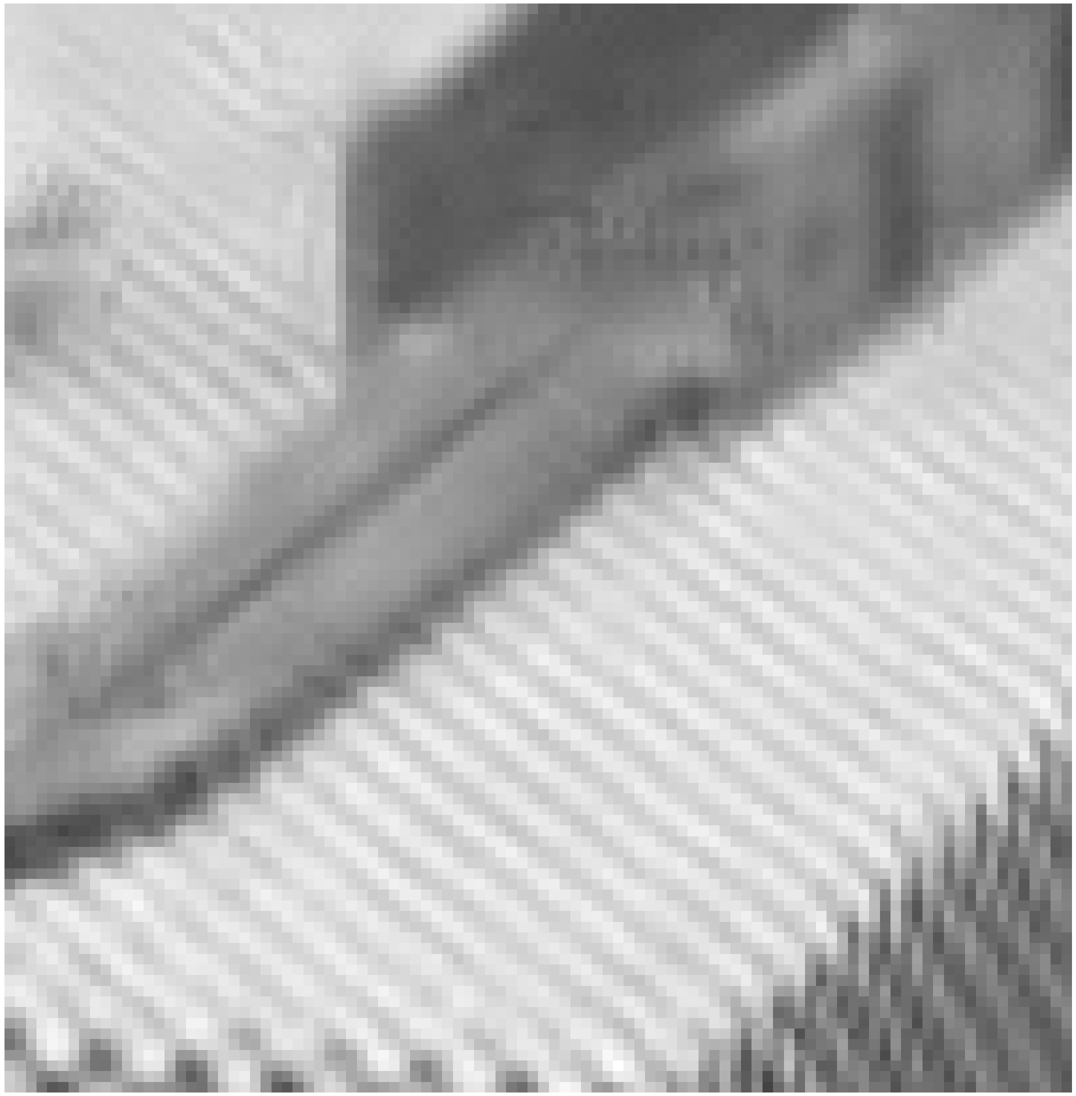
Denoising: Binary HMT



29.35 dB

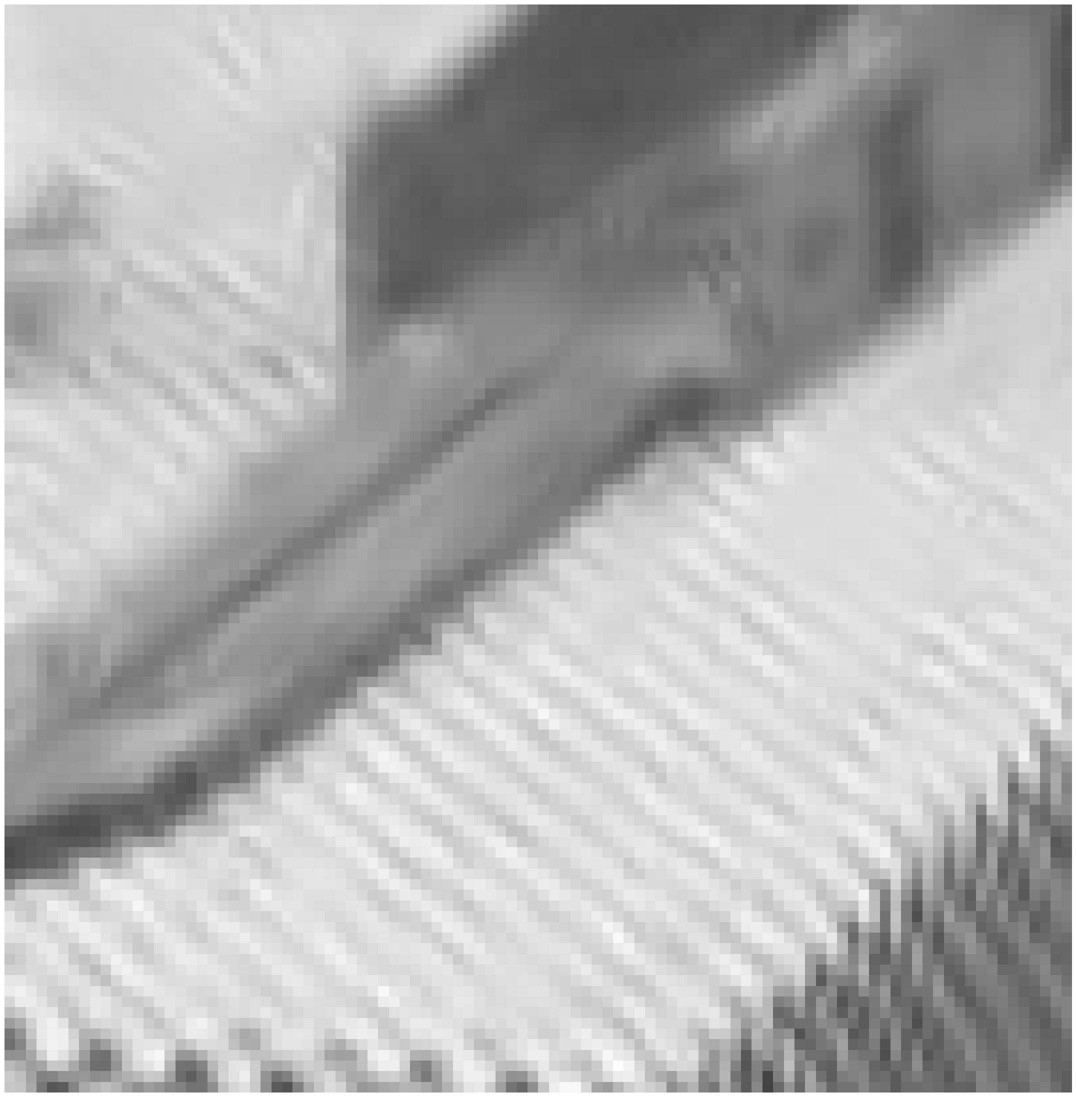
Crouse, Nowak, & Baraniuk, 1998

Denoising: HDP-HMT



32.10 dB

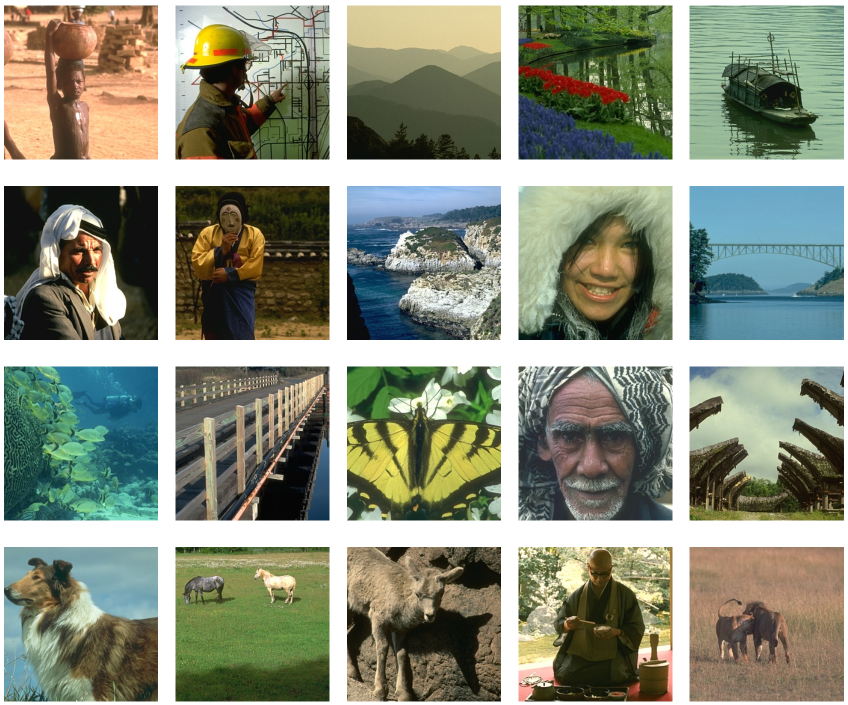
Denoising: Local GSM



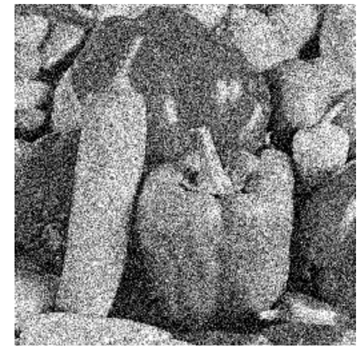
31.84 dB

Portilla et. al., 2003

Estimating Clean Images



Empirical Bayesian approach estimates model parameters from the noisy image



Transfer denoising approach reuses multiscale hidden state patterns of clean images for making robust predictions

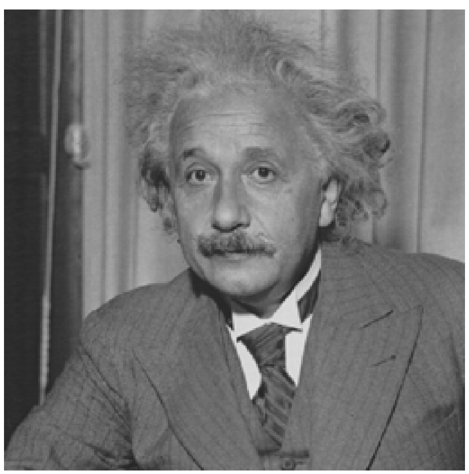
Denoising Einstein

Noisy
10.60 dB, 0.057

HDP-HMT
(Emp. Bayes)
25.64 dB, 0.564

HDP-HMT
(Transfer)
26.80 dB, 0.664

Original



BLS-GSM
26.38 dB, 0.647

BM3D
26.49 dB, 0.659

Natural Scene Denoising

Noisy
8.14 dB, 0.033

HDP-HMT
(Emp. Bayes)
24.24 dB, 0.519

HDP-HMT
(Transfer)
26.50 dB, 0.794

Original

BLS-GSM
25.59 dB, 0.726

BM3D
25.74 dB, 0.751

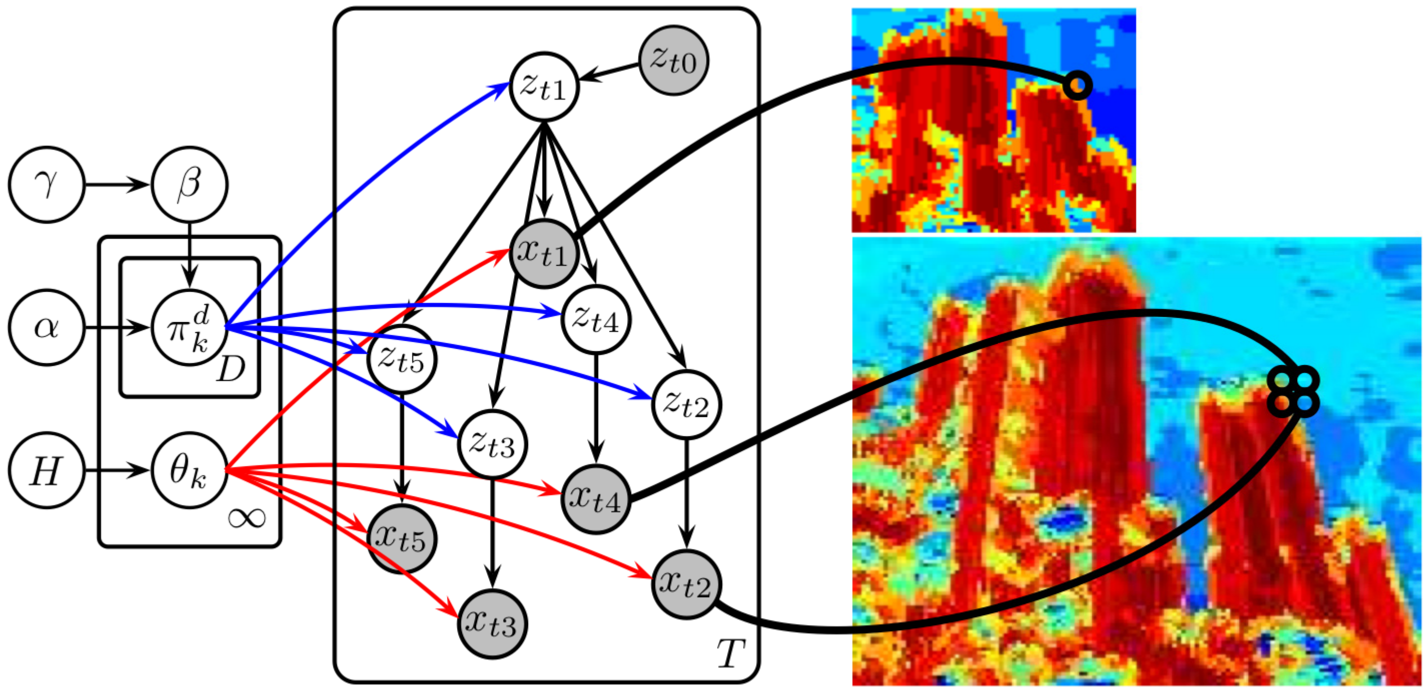
Natural Scene Categorization

Coast **Forest** **Open Country** **Street** **Tall Building**

Goals:

- Visually *recognize* natural scene categories
- Accurately model the statistics of *natural scene categories*

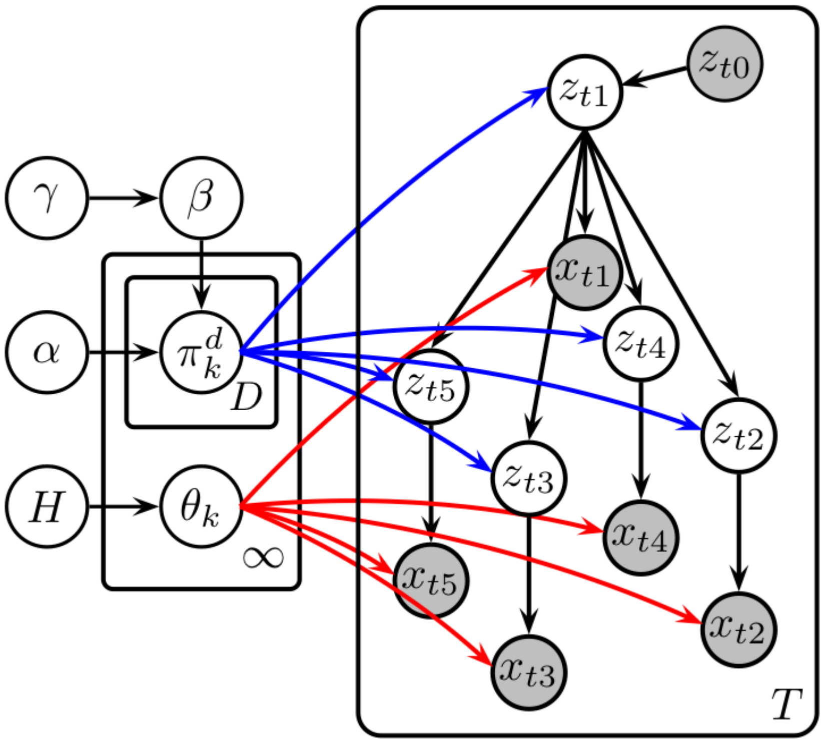
HDP-HMT Scene Model



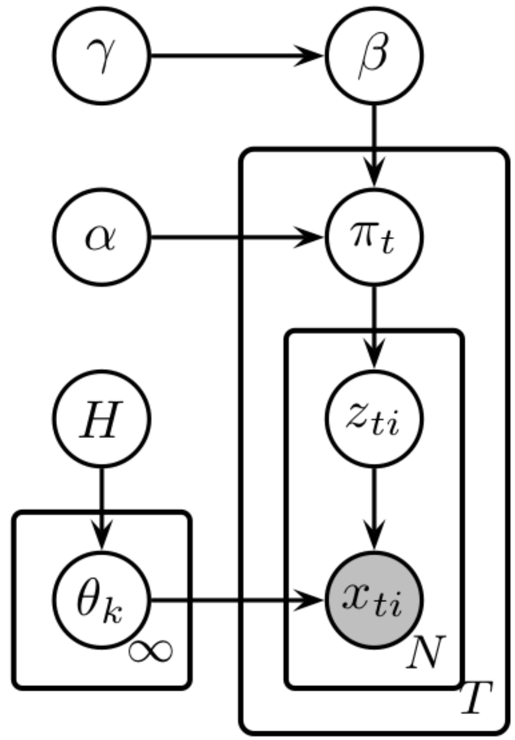
- Hidden states z_{ti} generate vectors of clean wavelet coefficients x_{ti} at multiple orientations, or dense multiscale SIFT descriptors

… versus baseline HDP-BOF

HDP-HMT



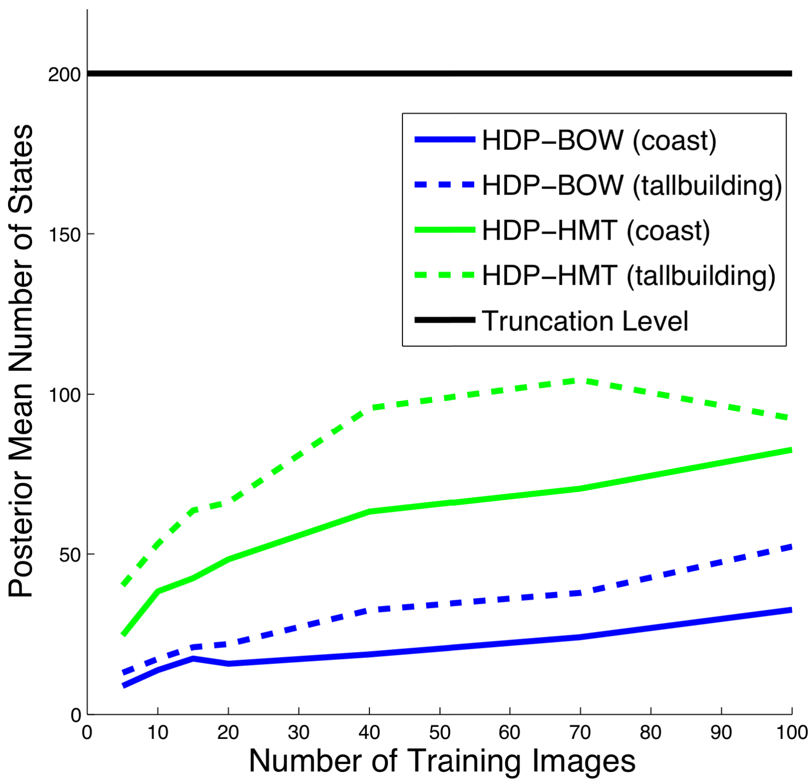
HDP-BOF



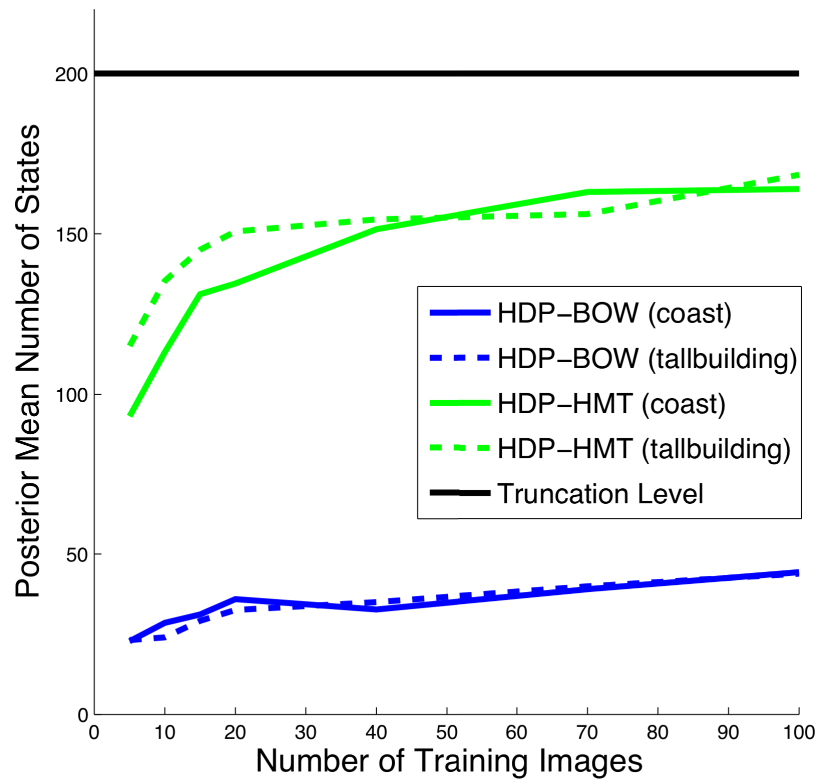
Nonparametric Bayesian extension of LDA scene models (Fei-Fei & Perona, 2005) which ignore spatial locations of locally extracted image features

Number of States

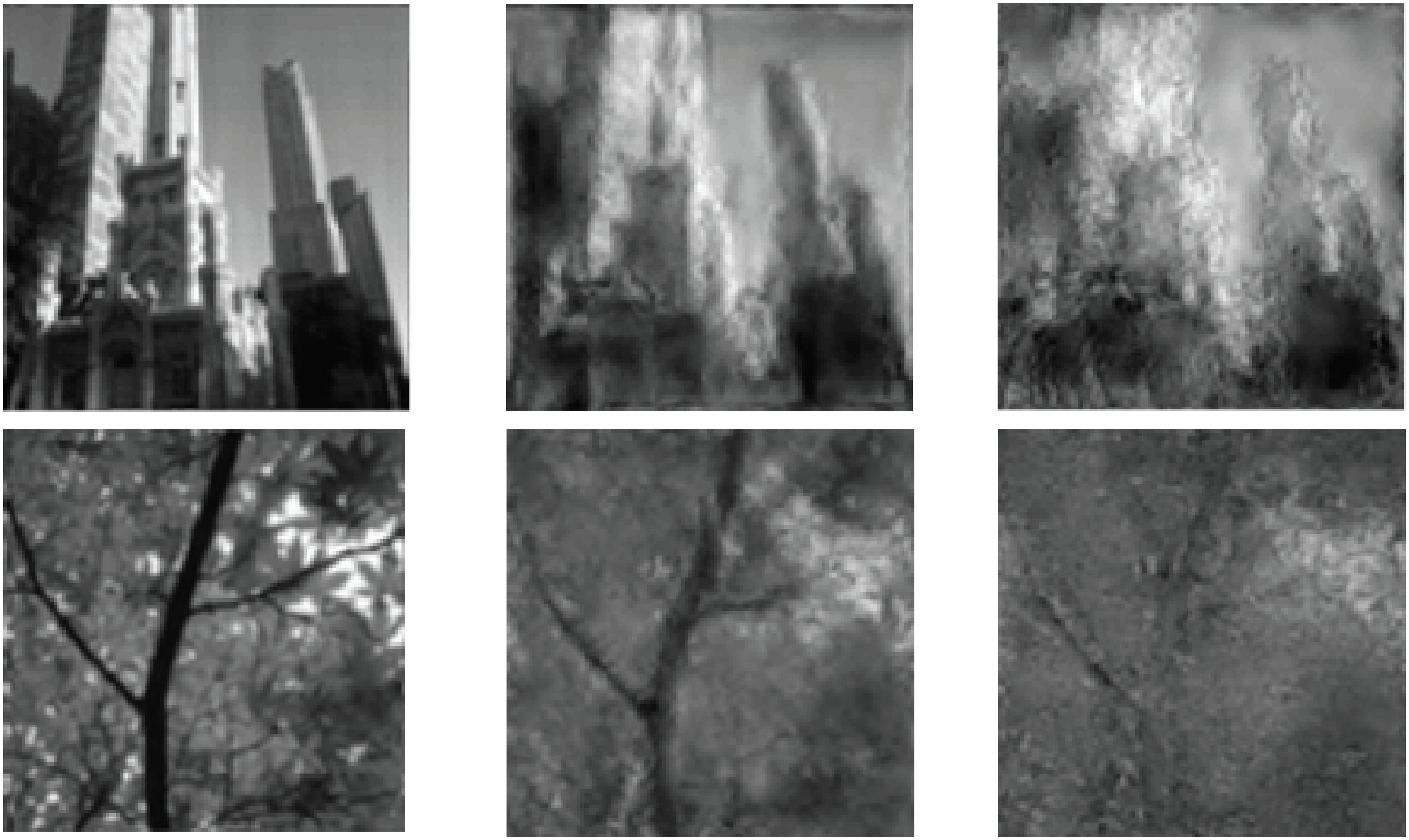
Wavelet (sp5)



SIFT



Samples given MAP states



Input Image

HDP Hidden Markov Tree

HDP Bag of Features

Categorizing Natural Scenes

Wavelet (sfp7)

coast	77.5	0.6	10.0	0.0	0.6	10.6	0.6	0.0
forest	0.0	91.4	0.0	0.0	5.5	0.8	2.3	0.0
highway	3.3	0.0	75.0	0.0	10.0	10.0	1.7	0.0
inside city	0.9	0.9	2.8	77.8	0.0	3.7	9.3	4.6
mountain	0.6	13.8	4.6	0.6	63.2	9.2	8.0	0.0
open country	8.6	10.0	3.3	0.5	11.0	61.9	4.8	0.0
street	0.0	1.1	5.4	2.2	7.6	0.0	81.5	2.2
tall building	0.0	0.0	2.6	13.5	0.6	0.6	8.3	74.4

HDP-BOF [75.3 %]

coast	90.6	0.6	4.4	0.0	1.9	2.5	0.0	0.0
forest	0.0	85.2	0.8	0.0	8.6	3.1	2.3	0.0
highway	8.3	0.0	80.0	0.0	6.7	0.0	3.3	1.7
inside city	1.9	0.9	5.6	75.0	0.9	0.9	10.2	4.6
mountain	1.7	1.1	2.9	0.0	91.4	1.1	1.7	0.0
open country	18.1	4.3	3.3	1.0	13.3	59.5	0.5	0.0
street	0.0	0.0	8.7	1.1	7.6	0.0	81.5	1.1
tall building	0.0	0.0	1.9	12.2	0.0	0.0	3.8	82.1

HDP-HMT [80.7 %]

SIFT

90.0	0.6	1.2	0.0	1.9	6.2	0.0	0.0	coast
0.0	87.5	0.0	0.0	7.8	4.7	0.0	0.0	forest
6.7	0.0	80.0	1.7	1.7	5.0	5.0	0.0	highway
0.0	0.0	1.9	87.0	0.0	0.0	9.3	1.9	inside city
1.1	0.6	0.6	0.0	90.2	5.7	0.6	1.1	mountain
11.0	1.9	1.0	0.0	5.7	80.0	0.5	0.0	open country
0.0	0.0	4.3	2.2	2.2	0.0	91.3	0.0	street
0.0	0.0	0.0	9.0	0.6	0.0	4.5	85.9	tall building

HDP-BOF [82.4 %]

86.2	1.2	4.4	0.0	0.6	7.5	0.0	0.0	coast
0.0	91.4	0.0	0.0	4.7	3.1	0.8	0.0	forest
6.7	0.0	75.0	1.7	3.3	6.7	6.7	0.0	highway
0.0	0.9	3.7	82.4	0.0	0.9	10.2	1.9	inside city
0.6	4.0	3.4	0.0	81.0	8.0	2.3	0.6	mountain
11.0	5.2	2.9	0.0	7.6	72.9	0.5	0.0	open country
0.0	0.0	6.5	2.2	1.1	0.0	89.1	1.1	street
0.0	0.0	0.6	7.1	1.3	0.0	10.3	80.8	tall building

HDP-HMT [86.5 %]