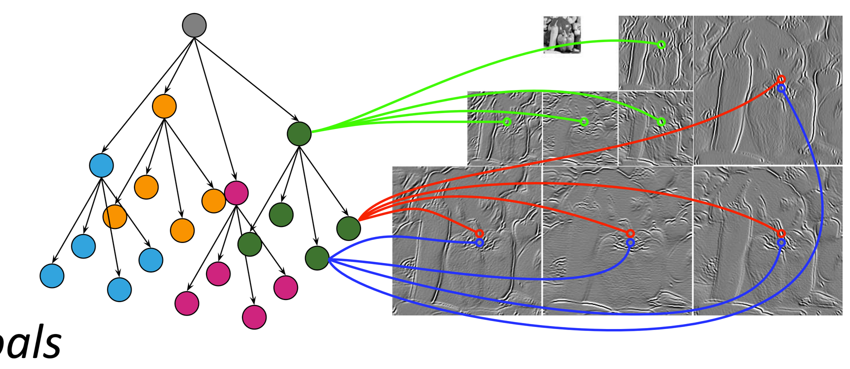
Transfer Denoising with Hierarchical Dirichlet Process Hidden Markov Trees

Jyri Kivinen, Erik Sudderth, and Michael Jordan

Helsinki University of Technology University of California, Berkeley

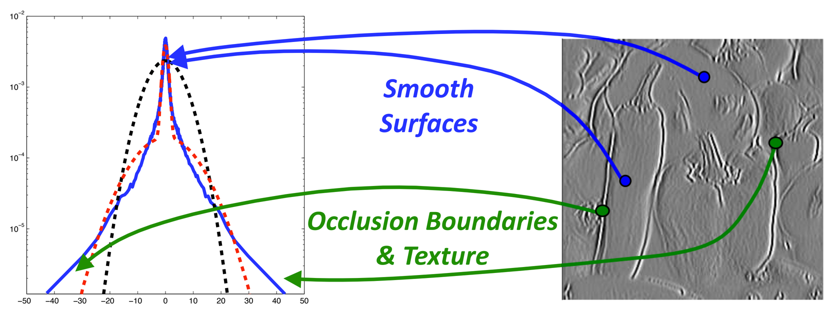
Statistical Models for Images



Goals

- Learn good statistical models for natural images
- Capture multiscale dependencies using a tree of latent variables
- Automatically adapt the number of latent states to the statistics of observed data
- Exploit availability of large image databases to develop efficient transfer denoising algorithms

Mixture Models for Heavy-Tailed Wavelet Marginals



- Extreme coefficient values resultant from edges and texture occur more frequently than with a Gaussian
- Gaussian scale mixtures provide good matches for the highly kurtotic, heavy tailed distributions

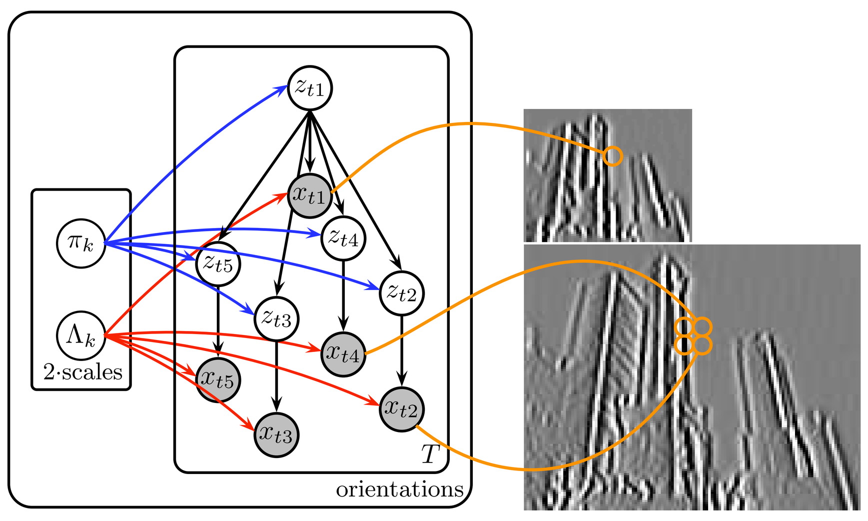
$$x_{ti} = v_{ti} u_{ti}\ ;\ v_{ti} \geq 0\ ,\ u_{ti} \sim \mathcal{N}(0, \Lambda)$$

- Discrete mixtures easier to work with, reasonable denoising results even with binary mixtures:

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

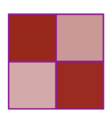
Models for Global Image Statistics

Binary Hidden Markov Trees (Crouse et. al. 1998)



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

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



$x_{ti} \longrightarrow$ *observed* wavelet coefficient

$\Lambda_k \longrightarrow$ state-specific *emission* covariances $x_{ti} \sim \mathcal{N}(0, \Lambda_{z_{ti}})$

- Wavelet coefficients marginally distributed as mixtures of two Gaussians
- Markov dependencies between hidden states capture persistence of image contours across locations and scales
- Models each scale and orientation independently

Dirichlet Process Mixtures

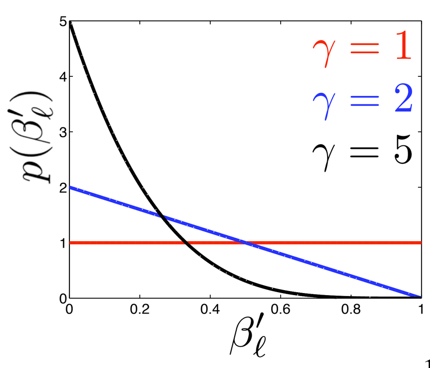
$$p(x_{ti} \mid \beta, \Lambda_1, \Lambda_2, \ldots) = \sum_{k=1}^{\infty} \beta_k \mathcal{N}(x_{ti}; 0, \Lambda_k)$$

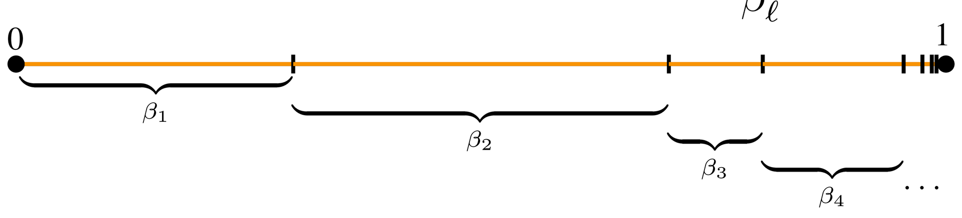
Stick-breaking prior for mixture weights controls complexity:

$$\beta_k = \beta'_k \prod_{\ell=1}^{k-1}(1-\beta'_\ell)$$

$$\beta'_\ell \sim \mathrm{Beta}(1, \gamma)$$

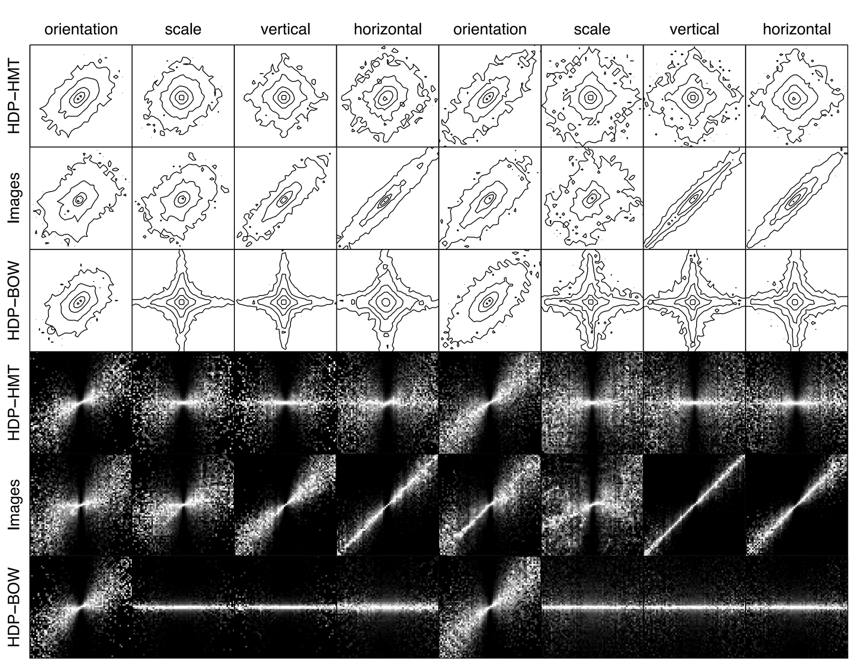
$\gamma \longrightarrow$ Concentration parameter



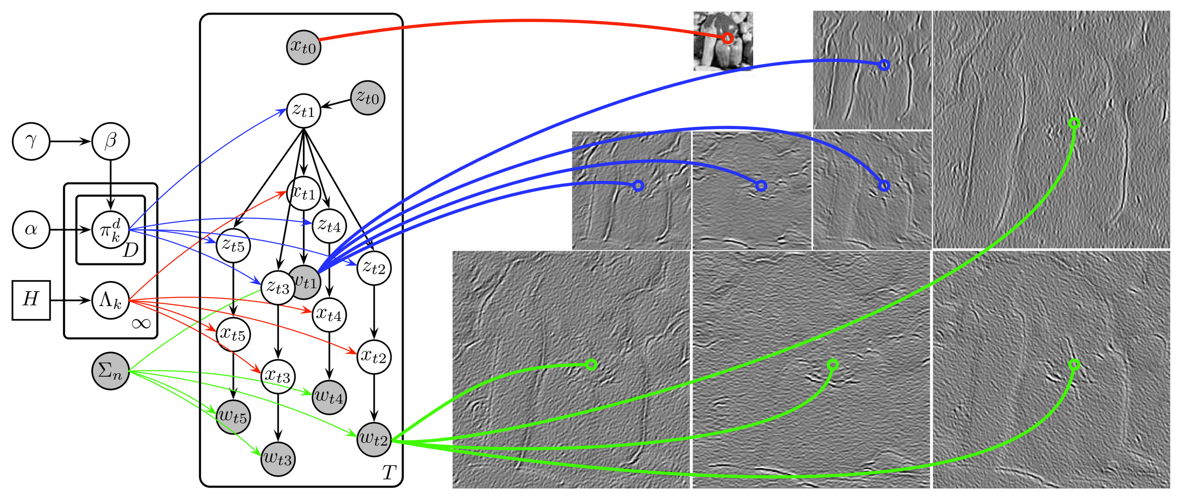


- Basis for *nonparametric* models whose complexity grows as additional data is observed
- Attractive *asymptotic guarantees*
- Leads to simple, effective *computational* methods

Pairwise Statistics of Wavelets



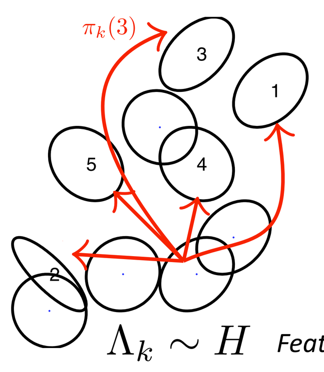
Hierarchical Dirichlet Process Hidden Markov Trees



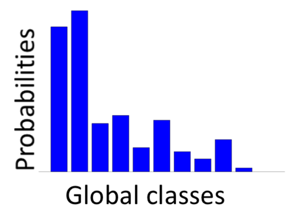
- Hidden states z_{ti} generate *vectors* of clean wavelet coefficients x_{ti} at multiple orientations
- Observations can be *corrupted* by additive zero-mean Gaussian noise of known variance
- Wavelet coefficients are marginally distributed as *infinite* Dirichlet Process (DP) mixtures
- Hierarchical Dirichlet Process (HDP) prior allows learning a potentially infinite set of *appearance patterns* from natural images

The Need for Hierarchical Dirichlet Processes (Teh et. al. 2004)

- A Hidden Markov Tree (HMT) is defined *by a set of mixture or transition distributions*, one for each value of parent state
- In our nonparametric approach, *Dirichlet Process* priors regularize an infinite state space
- The hierarchical DP ensures that a *common set* of child states are reused by multiple parent states

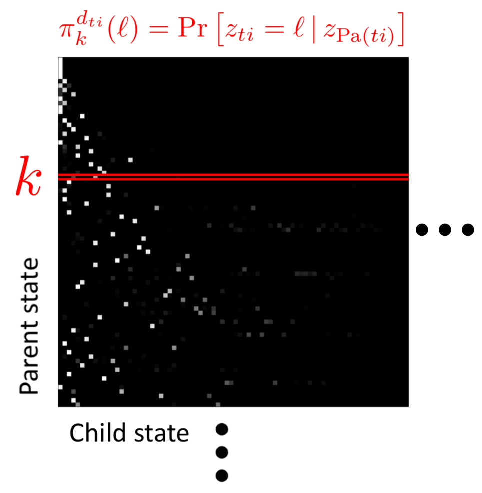


$\beta \sim \mathrm{Stick}(\gamma)$ *Average state frequencies*



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

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



$\Lambda_k \sim H$ *Feature distributions*

$z_{ti} \mid z_{\mathrm{Pa}(ti)} \sim \pi^{d_{ti}}_{z_{\mathrm{Pa}(ti)}}$ *State assignments*

$x_{ti} \mid z_{ti} \sim \mathcal{N}(0, \Lambda_{z_{ti}})$ *Observed features*

Learning with a Truncated Gibbs Sampler

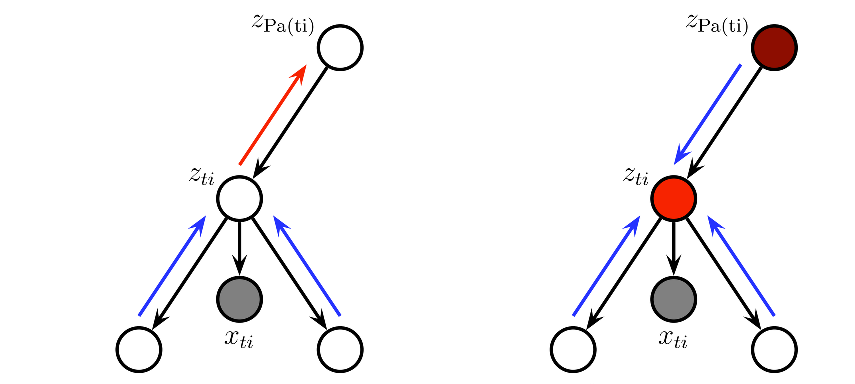
Weak limit approximations use high probability upper bounds on the number of states observed in a finite dataset:

$$\beta = (\beta_1, \ldots, \beta_K) \sim \mathrm{Dir}(\gamma/\mathrm{K}, \ldots, \gamma/\mathrm{K}) \qquad \Lambda_k \sim H$$

- Truncated model converges in distribution to $\mathrm{DP}(\gamma, \mathrm{H})$ as $K \to \infty$
- In a truncated HDP-HMT, each state-specific transition distribution is then sampled from a finite Dirichlet:

$$\pi_t = (\pi_{t1}, \ldots, \pi_{tK}) \sim \mathrm{Dir}(\alpha\beta_1, \ldots, \alpha\beta_K)$$

Sample hidden state assignments jointly using belief propagation:

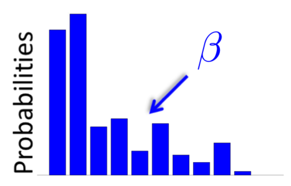


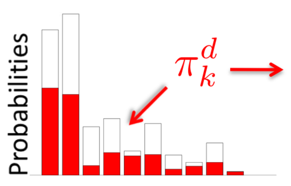
Compute messages from leaves to root *Sample assignments from root to leaves*

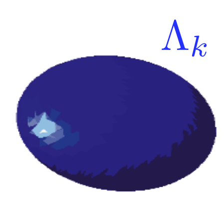
For noisy images, sample clean wavelet coefficients:

$$x_{ti} \sim \mathcal{N}\left(\left(\Lambda_{z_{ti}}^{-1} + \Sigma_n^{-1}\right)^{-1} \Sigma_n^{-1} w_{ti}, \left(\Lambda_{z_{ti}}^{-1} + \Sigma_n^{-1}\right)^{-1}\right)$$

Sample parameters:

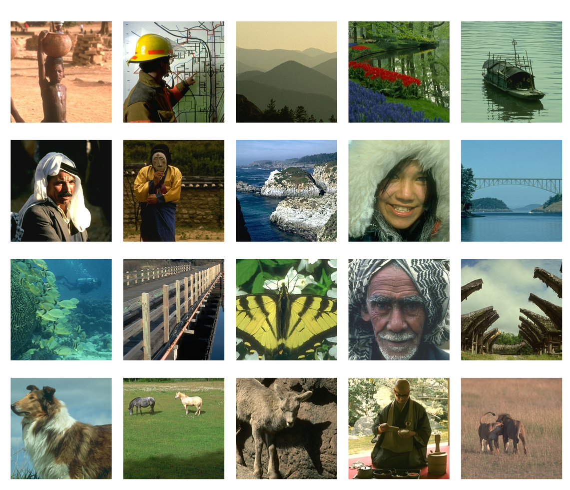






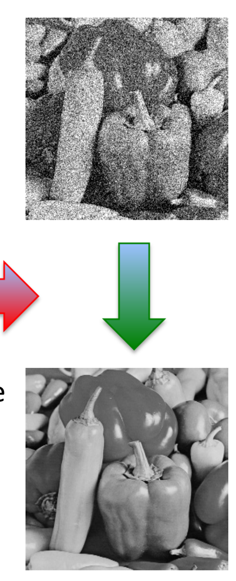
Sample transition probabilities *Sample feature distributions*

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



$$\mathbb{E}[x_{ti} \mid \mathbf{w}, \theta^{(s)}] = \sum_{k=1}^{K_s} p(z_{ti} = k \mid \mathbf{w}, \theta^{(s)})\, \mathbb{E}[x_{ti} \mid w_{ti}, \Lambda_k^{(s)}]$$

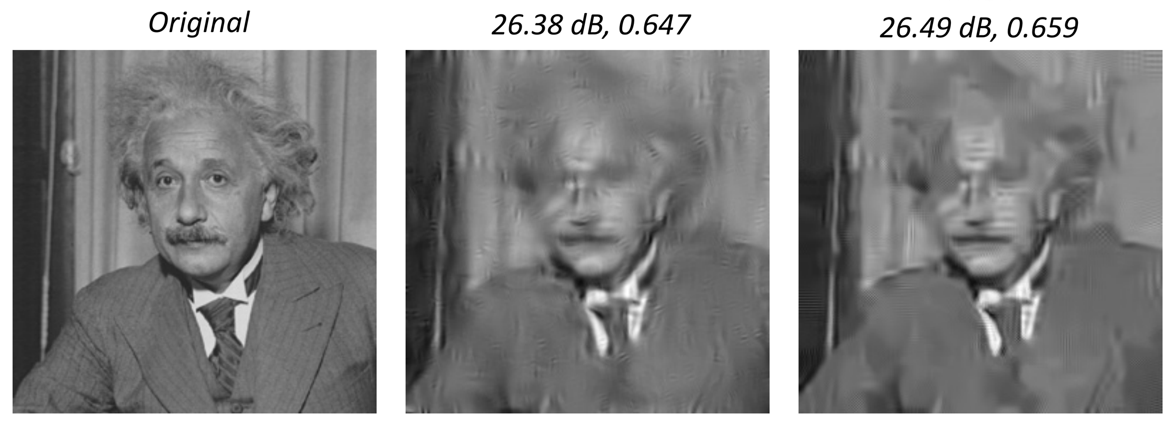
From belief propagation *Linear least-squares smoothing*

Denoising Standard Test Images

Denoising Einstein and Hill

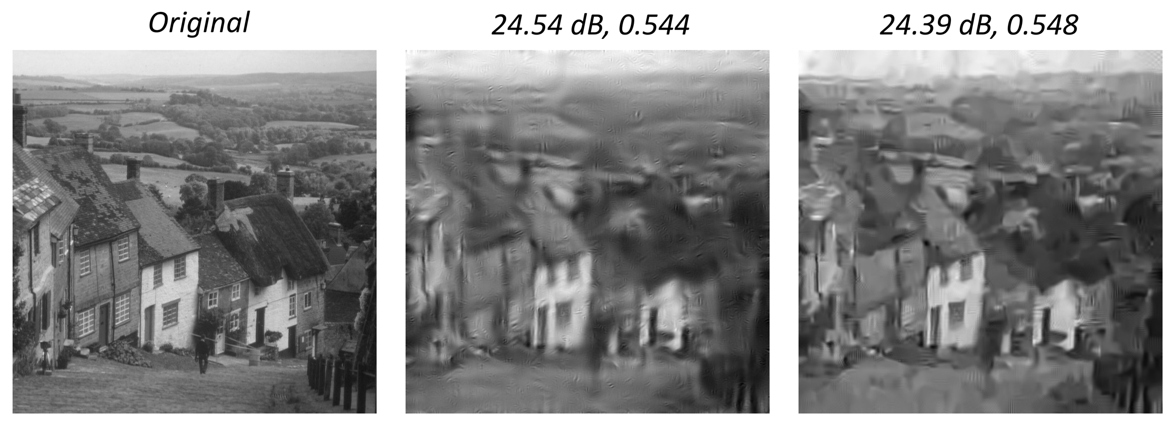
Noisy	HDP-HMT (Emp. Bayes)	HDP-HMT (Transfer)
10.60 dB, 0.057	25.07 dB, 0.520	26.80 dB, 0.664

Original	BLS-GSM	BM3D
	26.38 dB, 0.647	26.49 dB, 0.659



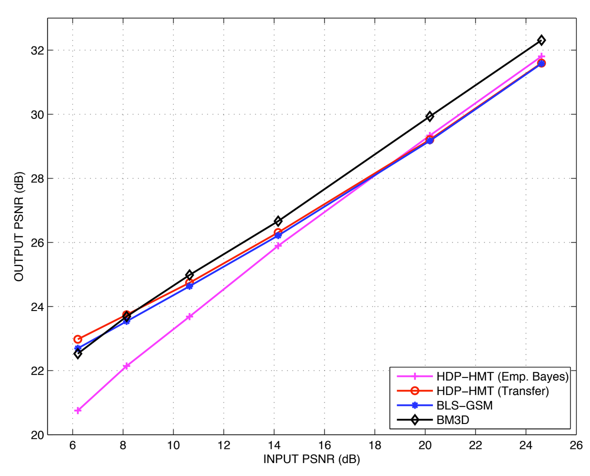
Noisy	HDP-HMT (Emp. Bayes)	HDP-HMT (Transfer)
8.12 dB, 0.038	23.05 dB, 0.414	24.74 dB, 0.568

Original	BLS-GSM	BM3D
	24.54 dB, 0.544	24.39 dB, 0.548

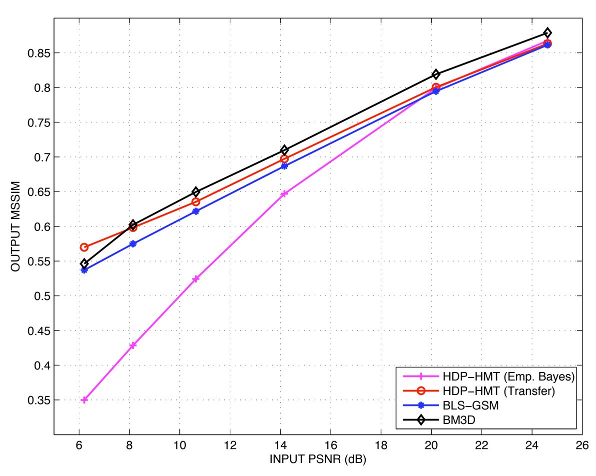


Average Denoising Performance

Peak signal-to-noise ratio (PSNR)

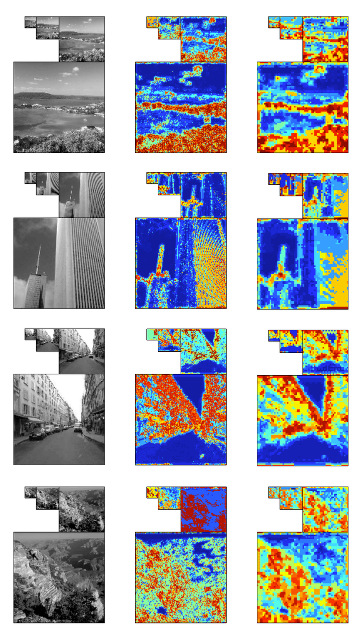


Mean structural similarity index (SSIM)

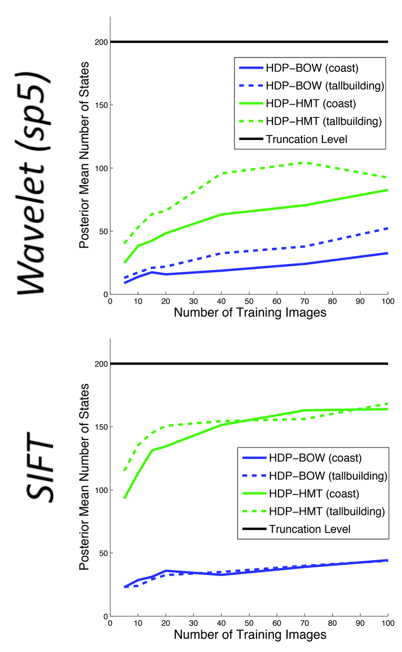


Natural Scene Analysis

MAP Assignments (sp5)



Number of States



Categorizing Natural Scenes

Wavelet (sp5) — HDP-BOF [70.5 %]

coast	58.1	0.4	21.5	0.0	0.8	18.8	0.4	0.0
forest	0.0	86.0	0.0	0.4	4.8	2.6	6.1	0.0
highway	5.6	0.0	77.5	1.9	4.4	8.8	1.9	0.0
inside city	0.0	1.9	6.2	74.0	1.0	1.9	7.2	7.7
mountain	1.1	7.3	9.1	0.4	56.9	15.3	9.5	0.4
open country	6.1	5.5	14.8	0.3	8.4	63.5	1.3	0.0
street	0.0	0.0	5.7	8.9	7.3	1.0	75.0	2.1
tall building	0.4	1.2	3.1	8.6	0.4	0.8	12.9	72.7

SIFT — HDP-BOF [79.8 %]

coast	79.2	0.8	6.5	0.0	0.8	12.3	0.4	0.0
forest	0.0	92.1	0.0	0.0	4.8	0.9	2.2	0.0
highway	7.5	0.0	76.9	1.9	5.0	5.0	3.8	0.0
inside city	0.0	0.5	3.4	85.1	0.0	1.0	5.3	4.8
mountain	1.1	3.6	3.6	0.0	74.1	10.6	6.2	0.7
open country	11.0	3.5	5.8	0.6	7.1	71.3	0.6	0.0
street	0.0	0.0	5.2	9.9	2.6	1.6	75.5	5.2
tall building	0.0	0.4	0.0	8.2	0.0	0.4	7.0	84.0

Wavelet (sp5) — HDP-HMT [79.6 %]

coast	71.5	0.4	13.1	0.0	3.5	11.5	0.0	0.0
forest	0.0	83.8	0.0	0.4	11.0	2.6	1.8	0.4
highway	1.2	0.0	89.4	1.2	3.8	3.8	0.6	0.0
inside city	0.0	0.5	5.8	82.7	1.4	0.5	3.8	5.3
mountain	1.1	3.3	6.9	0.4	76.3	9.9	1.8	0.4
open country	8.7	1.6	12.6	0.0	8.4	68.4	0.3	0.0
street	0.0	0.0	6.2	5.7	5.7	0.0	81.8	0.5
tall building	0.0	0.0	0.0	10.2	2.0	0.4	4.7	82.8

SIFT — HDP-HMT [86.5 %]

coast	86.5	0.8	2.3	0.0	0.8	9.6	0.0	0.0
forest	0.0	91.2	0.0	0.0	5.7	1.8	1.3	0.0
highway	5.6	0.0	80.0	2.5	3.8	3.8	4.4	0.0
inside city	0.0	0.5	1.0	92.3	0.0	0.5	3.4	2.4
mountain	0.4	1.8	2.2	0.0	86.9	5.5	1.1	2.2
open country	10.3	1.6	3.5	0.0	4.8	79.0	0.6	0.0
street	0.0	0.0	1.6	6.8	1.6	0.5	87.0	2.6
tall building	0.0	0.4	0.0	7.4	0.4	0.0	3.1	88.7