Learning Multiscale Representations of Natural Scenes using Dirichlet Processes

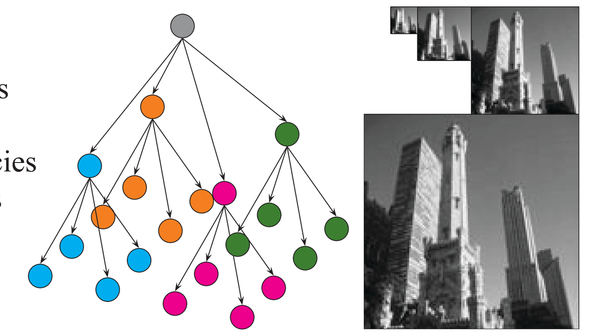
Jyri Kivinen, Erik Sudderth, and Michael Jordan

Helsinki University of Technology   University of California, Berkeley

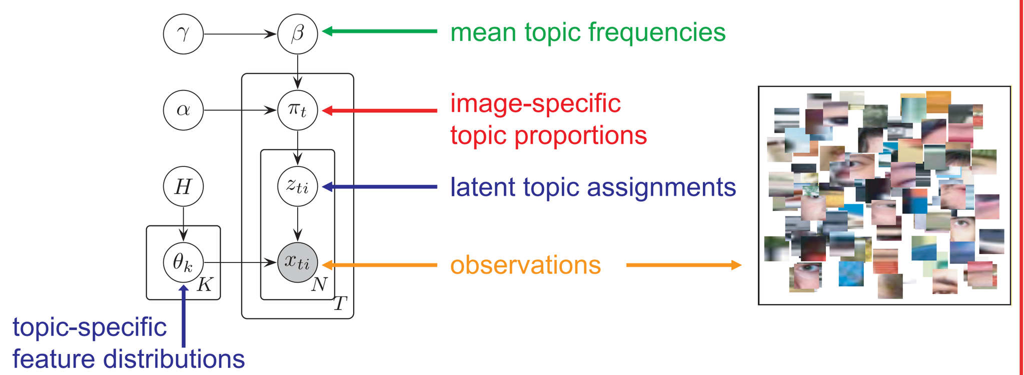
Statistical Models for Natural Scenes

Goals

- Learn global statistical models for natural scenes
- Capture multiscale dependencies using a tree of latent variables
- Automatically adapt the number of latent states to the statistics of observed data

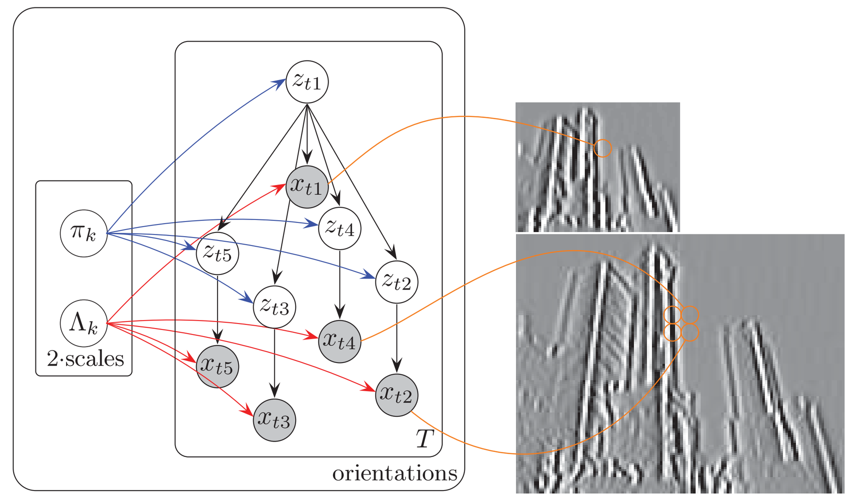


Bags-of-Features Topic Models



- Models image feature appearance with a global dictionary of visual words
- Ignores global spatial structure of visual scenes
- Previous applications to natural scene categories assumed a fixed number of latent topics *(Fei-Fei and Perona 2005, Bosch et al. 2006)*

Models for Global Image Statistics



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

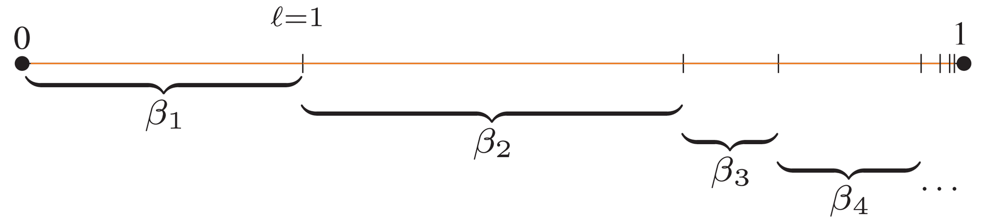
- Wavelet coefficients marginally distributed as mixtures of two Gaussians
- Markov dependencies between hidden states capture persistence of image contours across locations and scales
- Models each scale and orientation independently

Dirichlet Process Mixtures

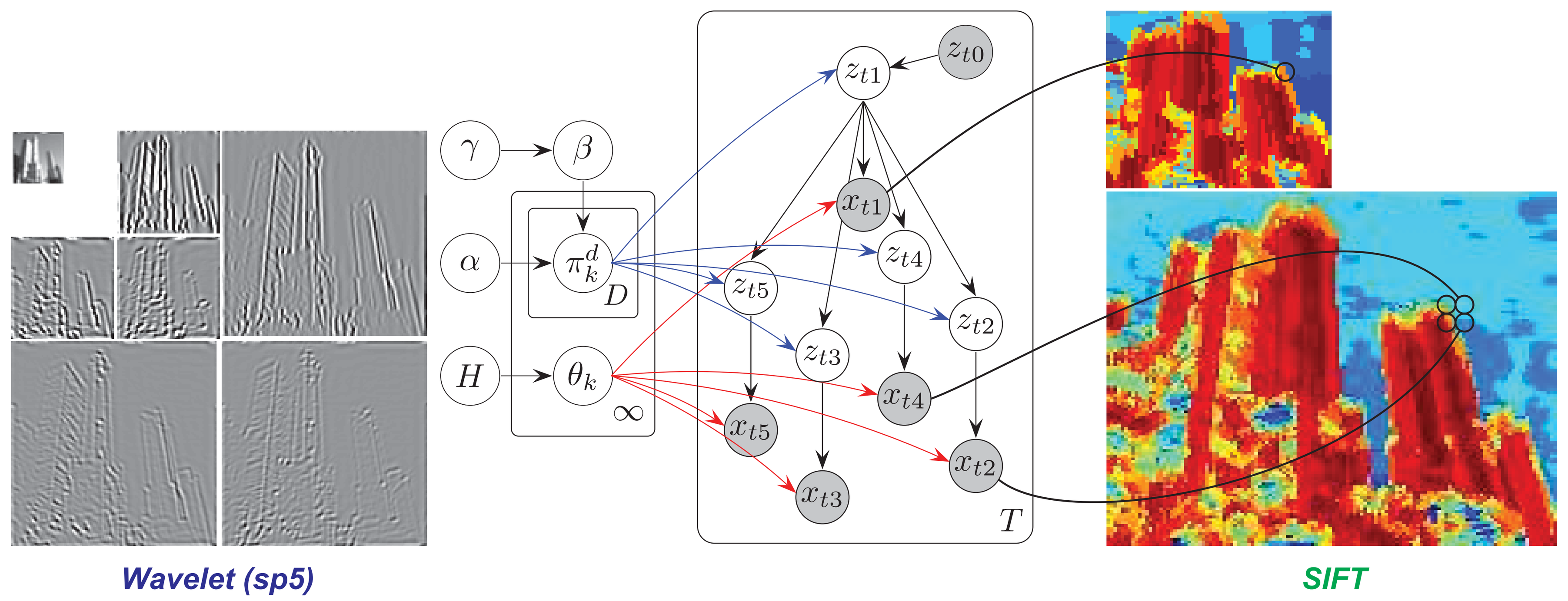
$$p(x_{ti} \mid \beta, \theta_1, \theta_2, \ldots) = \sum_{k=1}^{\infty} \beta_k f(x_{ti} \mid \theta_k)$$

Stick-breaking prior for mixture weights controls complexity:

$$\beta_k = \beta'_k \prod_{\ell=1}^{k-1} (1 - \beta'_\ell) \qquad \beta'_\ell \sim \mathrm{Beta}(1, \gamma)$$



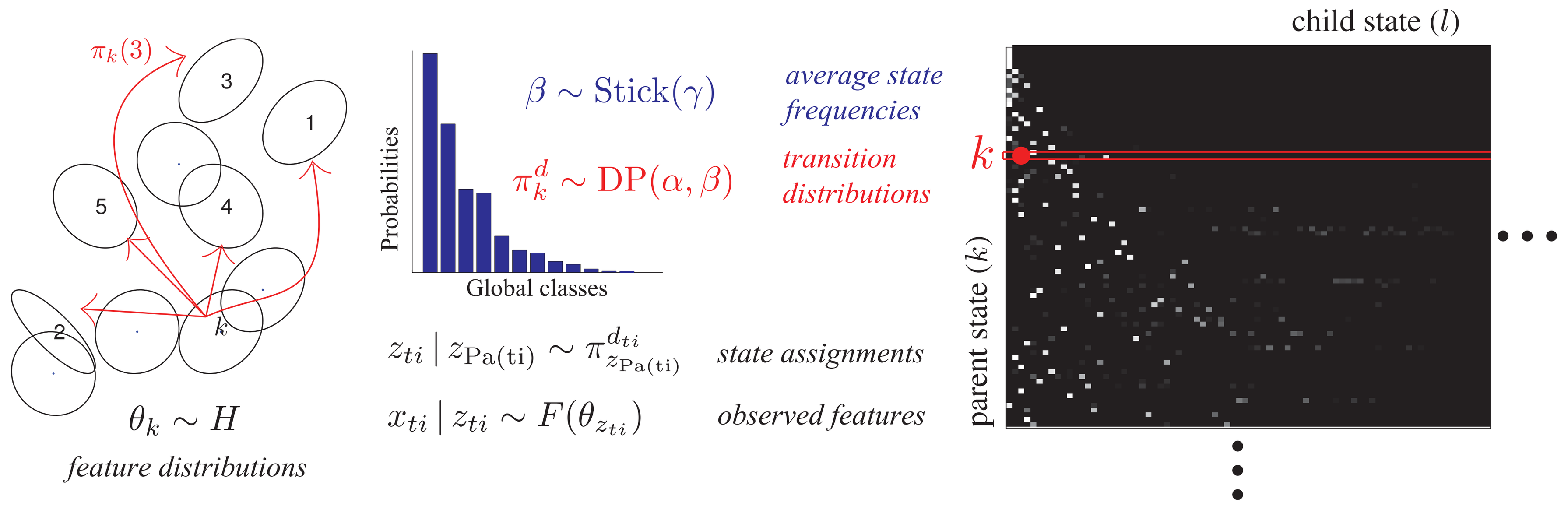
Hierarchical Dirichlet Process Hidden Markov Trees



- Hidden states z_{ti} generate vectors of *wavelet coefficients* x_{ti} at multiple orientations, or *SIFT descriptors*
- Observations are marginally distributed as *infinite* Dirichlet Process (DP) mixtures
- Hierarchical Dirichlet Process (HDP) prior allows learning of 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*

$z_{ti} \mid z_{\mathrm{Pa}(ti)} \sim \pi_{z_{\mathrm{Pa}(ti)}}^{d_{ti}}$ *state assignments*

$\theta_k \sim H$ *feature distributions*

$x_{ti} \mid z_{ti} \sim F(\theta_{z_{ti}})$ *observed features*

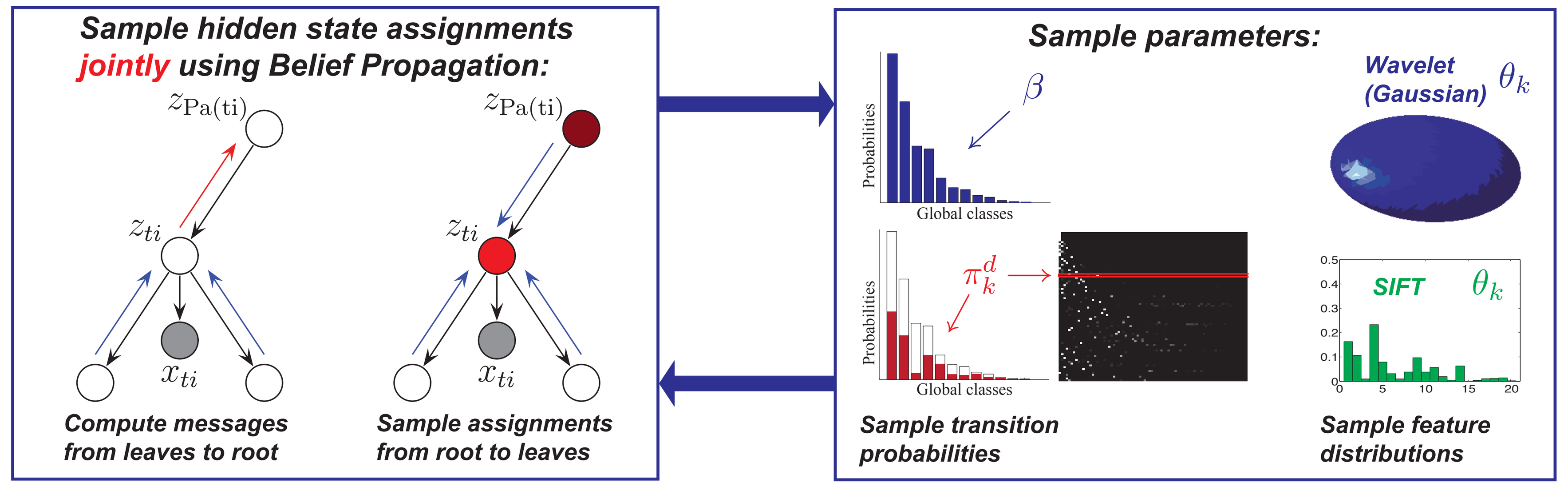
Learning with a Truncated Gibbs Sampler

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

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

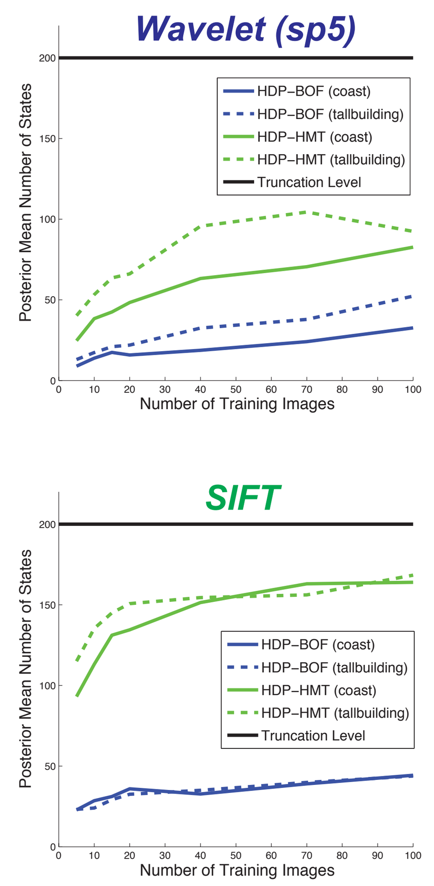
- Truncated model converges in distribution to $\mathrm{DP}(\gamma, H)$ as $K \to \infty$
- In a truncated HDP-HMT, each state-specific transition distribution is then sampled from a finite Dirichlet:

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

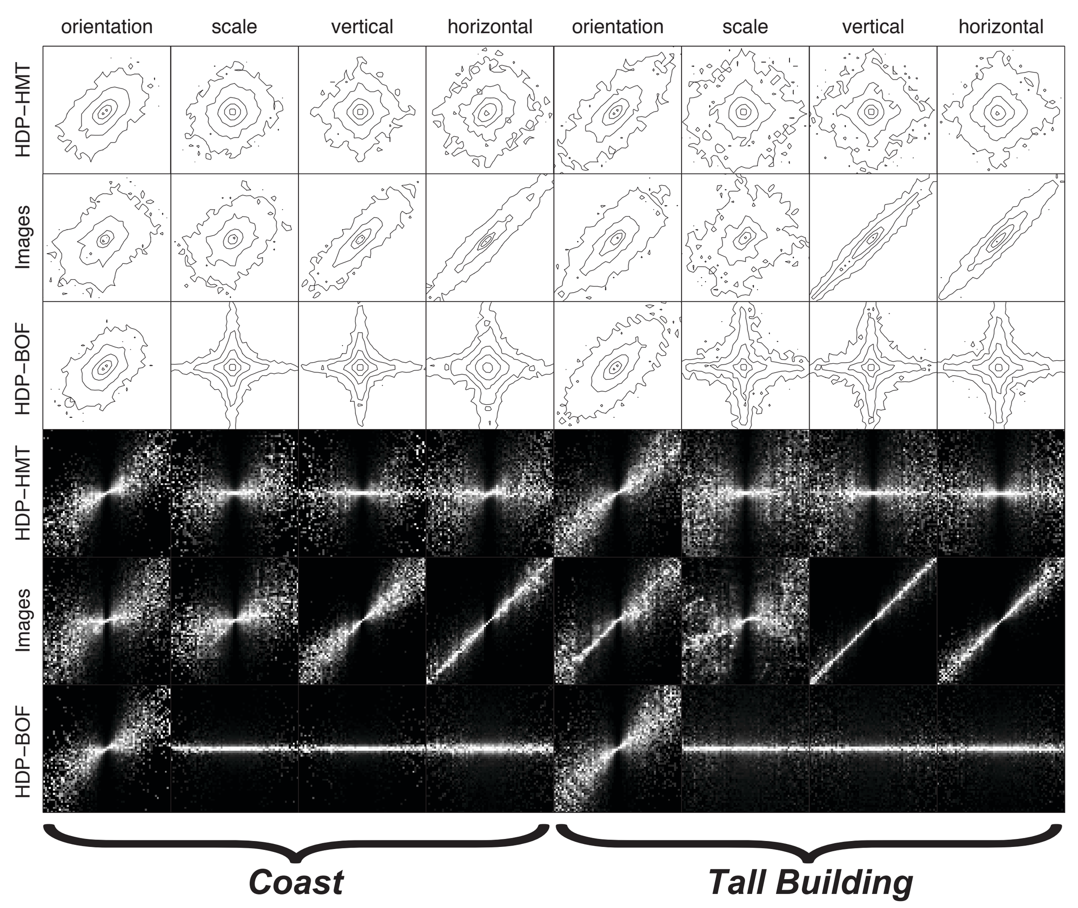


Visualizing Learned Spatial Relationships

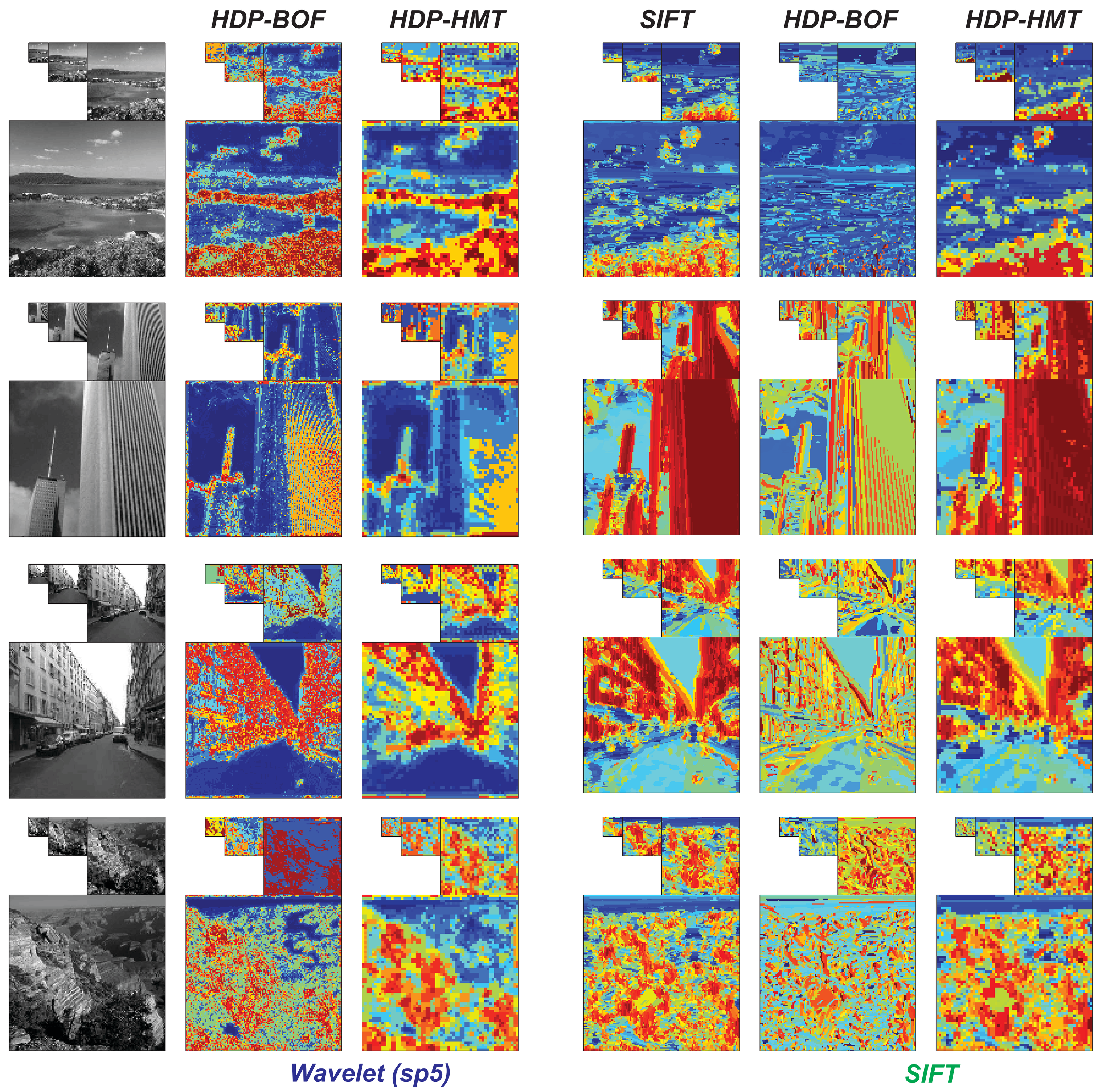
Number of States



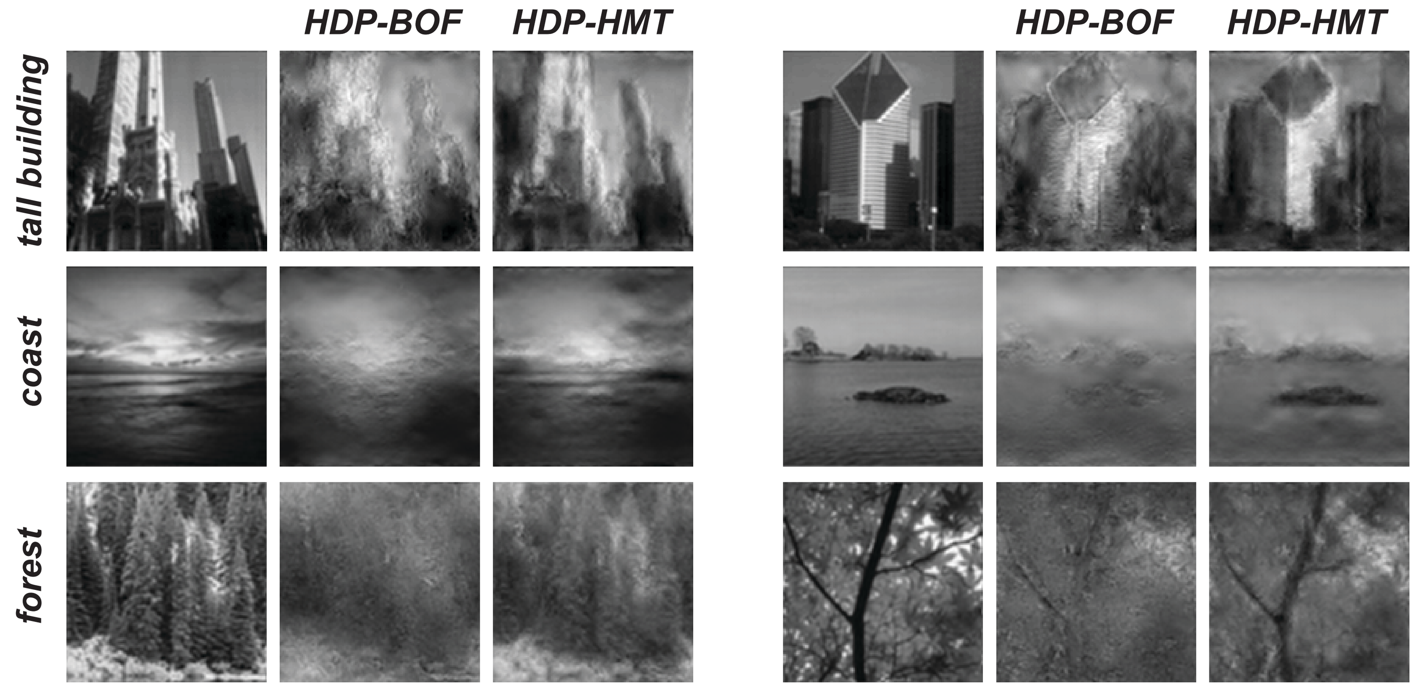
Wavelet Coefficient Histograms



MAP Assignments of Features to States



Sampling Scenes Given MAP Assignments



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 [78.4%]

coast	73.5	0.8	9.2	0.0	1.2	14.6	0.8	0.0
forest	0.0	93.0	0.0	0.0	3.9	0.9	2.2	0.0
highway	7.5	0.0	76.2	3.1	5.0	4.4	3.8	0.0
inside city	0.0	0.5	3.8	84.1	0.0	1.0	4.8	5.8
mountain	1.5	4.4	3.6	0.0	72.3	11.3	6.2	0.7
open country	8.7	4.2	6.8	1.0	5.5	73.5	0.3	0.0
street	0.0	0.0	6.2	10.4	2.1	1.0	74.0	6.2
tall building	0.0	0.4	0.0	7.4	0.4	0.0	11.3	80.5

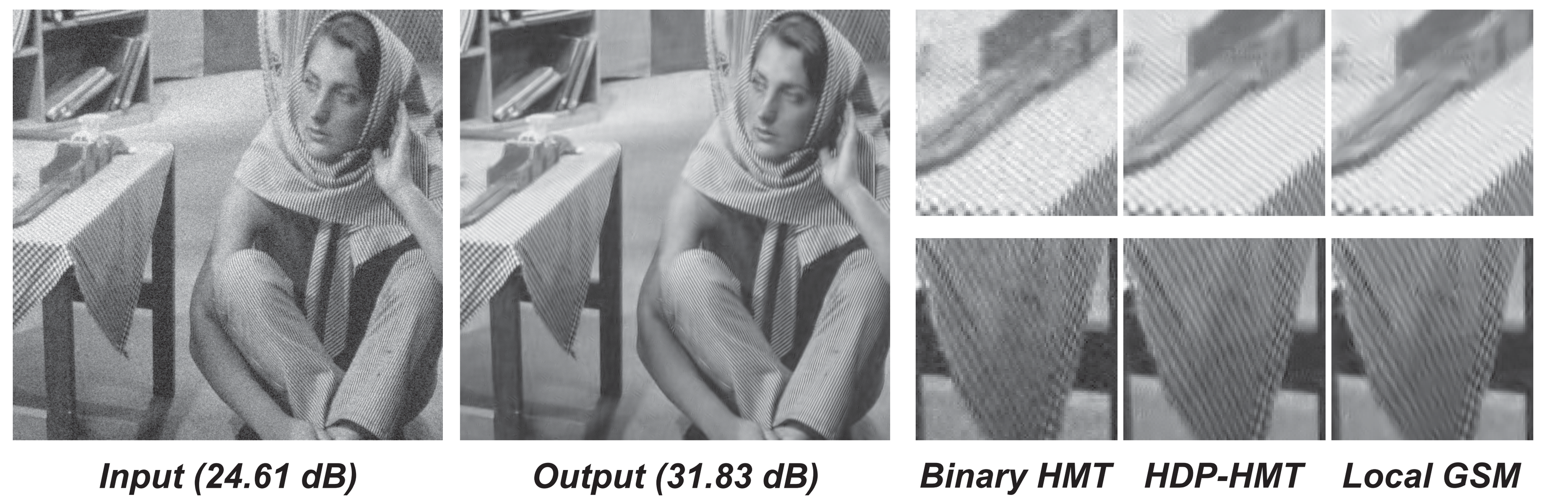
Wavelet (sp5) — HDP-HMT [79.6%]

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 [84.5%]

coast	75.8	0.4	6.2	0.0	3.1	14.6	0.0	0.0
forest	0.0	82.5	0.0	0.0	7.9	7.0	1.8	0.9
highway	3.1	0.0	86.9	1.2	1.9	5.0	1.2	0.6
inside city	0.0	0.0	3.4	88.9	0.0	0.0	1.4	6.2
mountain	1.1	1.1	0.7	0.7	88.3	6.2	1.1	0.7
open country	9.7	1.3	2.9	0.0	6.8	78.4	1.0	0.0
street	0.0	0.0	2.6	2.6	2.1	0.0	91.1	1.6
tall building	0.0	0.0	0.0	9.4	1.2	0.0	5.5	84.0

Wavelet-based Image Denoising



Input (24.61 dB)   *Output (31.83 dB)*   *Binary HMT*   *HDP-HMT*   *Local GSM*

J. Kivinen, E. Sudderth, and M. Jordan, Image Denoising with Nonparametric Hidden Markov Trees
IEEE International Conference on Image Processing, 2007