Nonparametric Learning for Layered Segmentation of Natural Images

Soumya Ghosh and Erik B. Sudderth
Department of Computer Science, Brown University

Segmentation of Natural Images

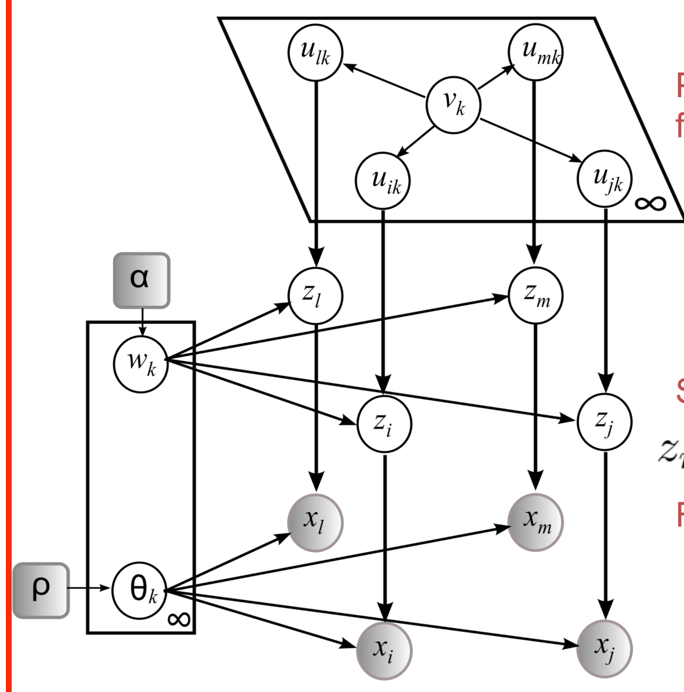
Goals: To develop robust, practical learning and inference algorithms for Bayesian nonparametric models of natural image partitions, which reliably vary segment resolution and model the uncertainty inherent in segmentation tasks.

Contributions

- *Learning:* Extending prior work on spatially dependent Pitman-Yor processes (Sudderth & Jordan, NIPS 2008), we propose an efficient low-rank covariance representation and calibrate it via a database of human segmentations.
- *Inference:* Substantially improving on conventional mean field methods, we develop a higher-order variational approximation based on expectation propagation, and robustly optimize it via stochastic search.
- *Results:* Multiple, high-quality segmentations of challenging natural images.

Modeling Image Partitions of Unknown Size

- Partitions of variable resolution are generated by depth-ordered, occluding layers.
- Layer support is modeled by a smooth continuous function, as in level set methods.



Non-Markov Gaussian Processes: Spatial Dependence

$$\mathbf{u}_k = A\mathbf{v}_k + \epsilon_k \quad \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_D)$$

$$u_{kn} \sim \mathcal{N}(0,1) \quad \epsilon_k \sim \mathcal{N}(0, \Psi)$$

Pitman-Yor prior: Segment sizes follow a *power law* distribution

$$\Pr[z_n = k] = w_k \prod_{\ell=1}^{k-1} (1 - w_\ell)$$

$$w_k \sim \mathrm{Beta}(1 - \alpha_a, \alpha_b + k\alpha_a)$$

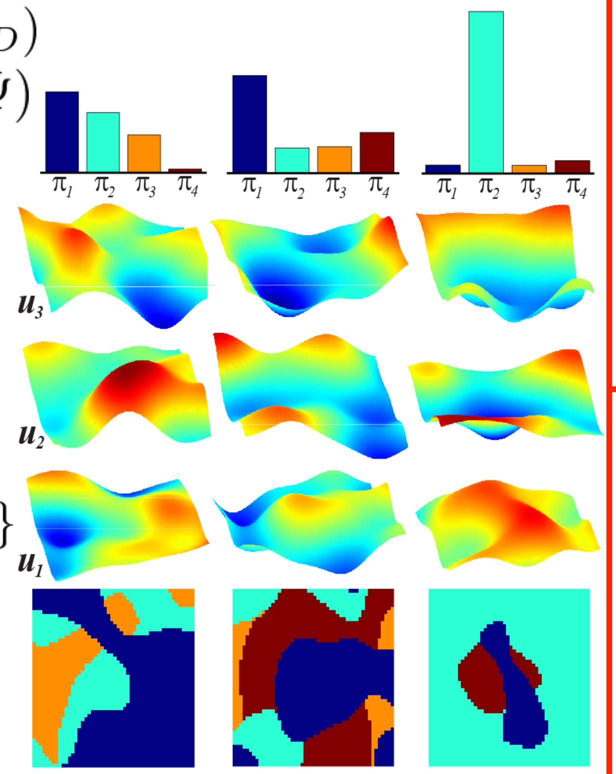
Segment Assignments: Layers

$$z_n = \min\{k \mid u_{kn} < \Phi^{-1}(w_k)\}$$

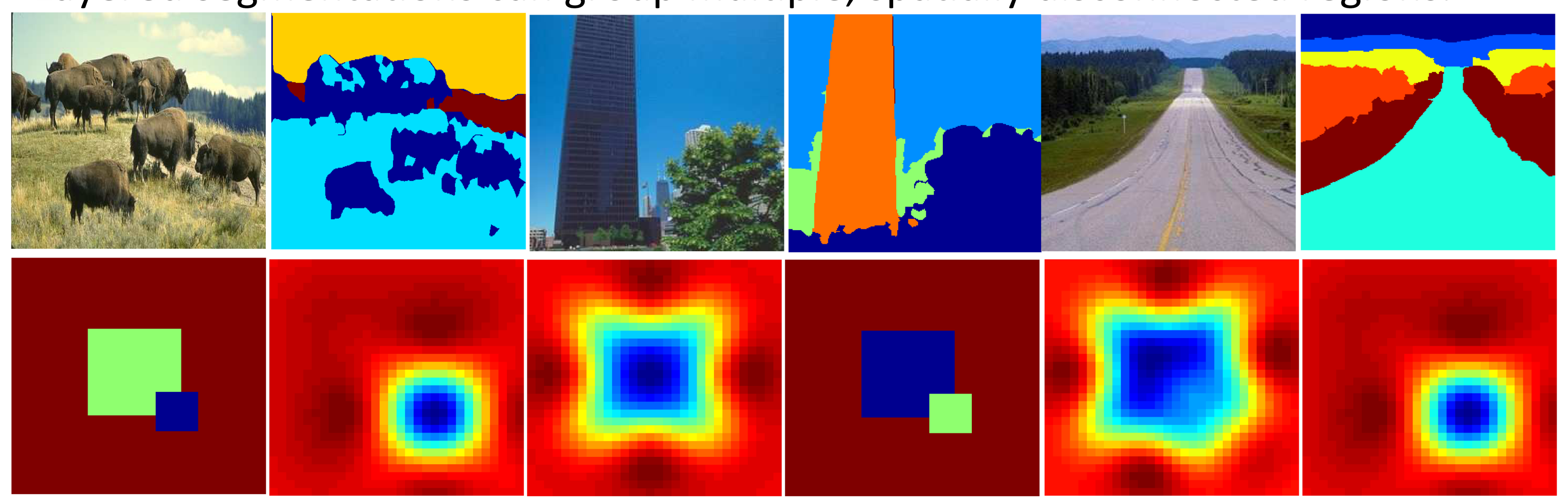
Features: Color & Texture

$$x_n^c \sim \mathrm{Mult}(\theta_{z_n}^c)$$

$$x_n^t \sim \mathrm{Mult}(\theta_{z_n}^t)$$



- Layered segmentations can group multiple, spatially disconnected regions.



More Likely | Front Layer | Middle Layer | Less Likely | Front Layer | Middle Layer

- Bias towards simply shaped regions can recover some occlusion relationships: hypotheses above assume *blue in front*, *green in middle*, *red in background*

Inference: Variational Stochastic Search

Maximization-Expectation: Marginalize parameters, maximize discrete partition.

$$p(\mathbf{z} \mid \mathbf{x}, \eta) \propto p(\mathbf{x} \mid \mathbf{z}, \rho) p(\mathbf{z} \mid \alpha, A, \Psi)$$

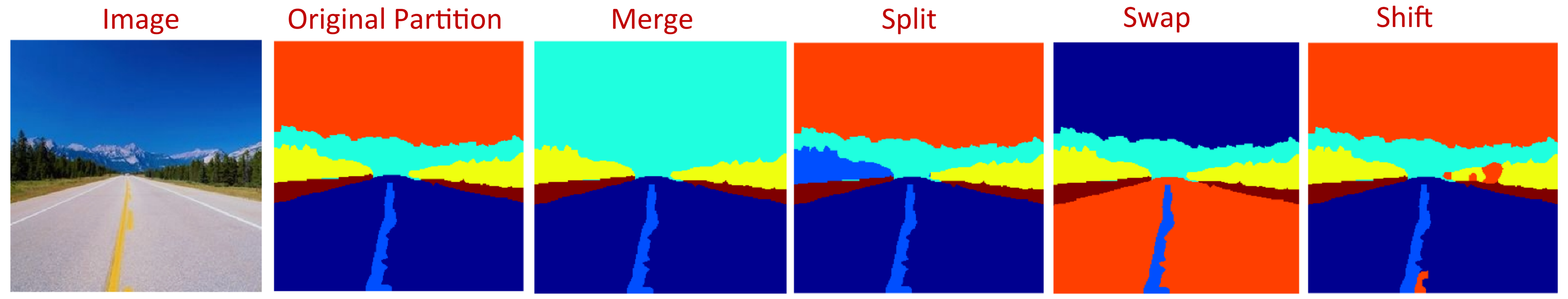
Integrate likelihood parameters analytically (Dirichlet prior conjugate to multinomial)

Marginalize layer support functions via expectation propagation (EP): approximate but very accurate

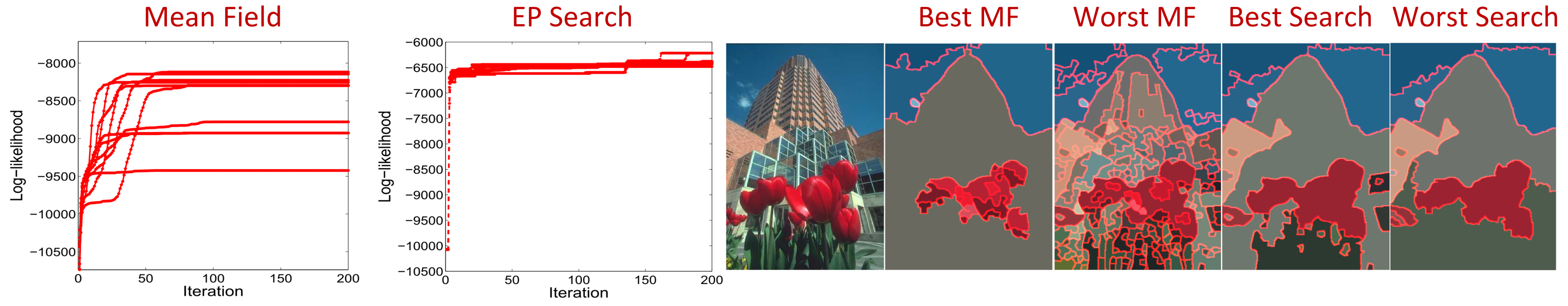
- Unlike conventional variational methods, no upper bound on the number of layers.
- EP more accurately models uncertainty, allowing selection of segmentation resolution.

Search for a Collection of Probable Partitions

- Space of partitions is explored through *local* (shift) and *global* (split, merge, swap) moves.



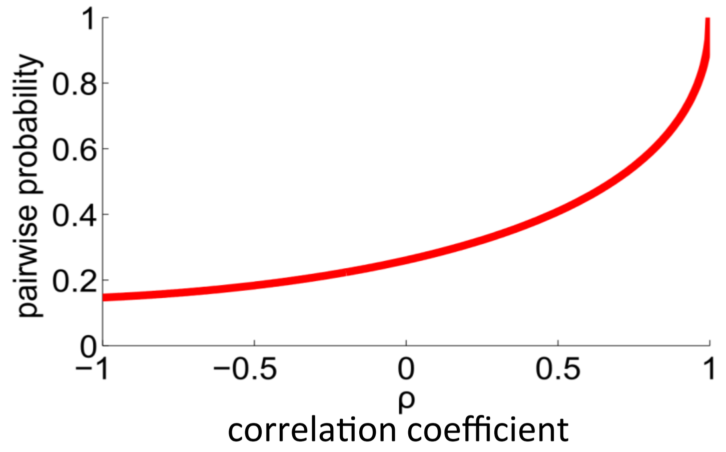
- Robust to initialization and finds better partitions than conventional local optimization.

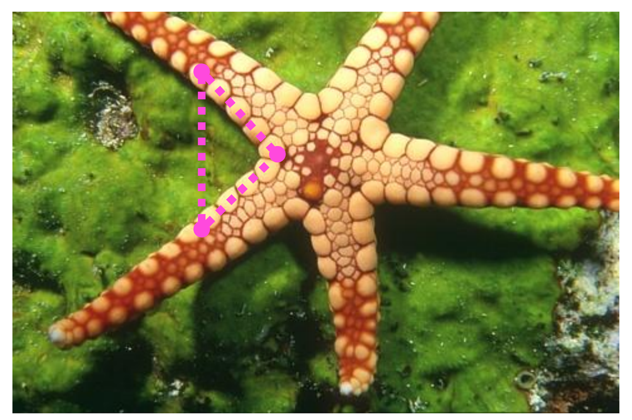


Learning: Berkeley Segmentation Dataset

- Appearance hyperparameters: Maximize training partition marginal likelihood
- Layer size hyperparameters: From Chinese restaurant process form of Pitman-Yor prior
- Covariance kernel $\Sigma = AA^T + \Psi$ is learned by

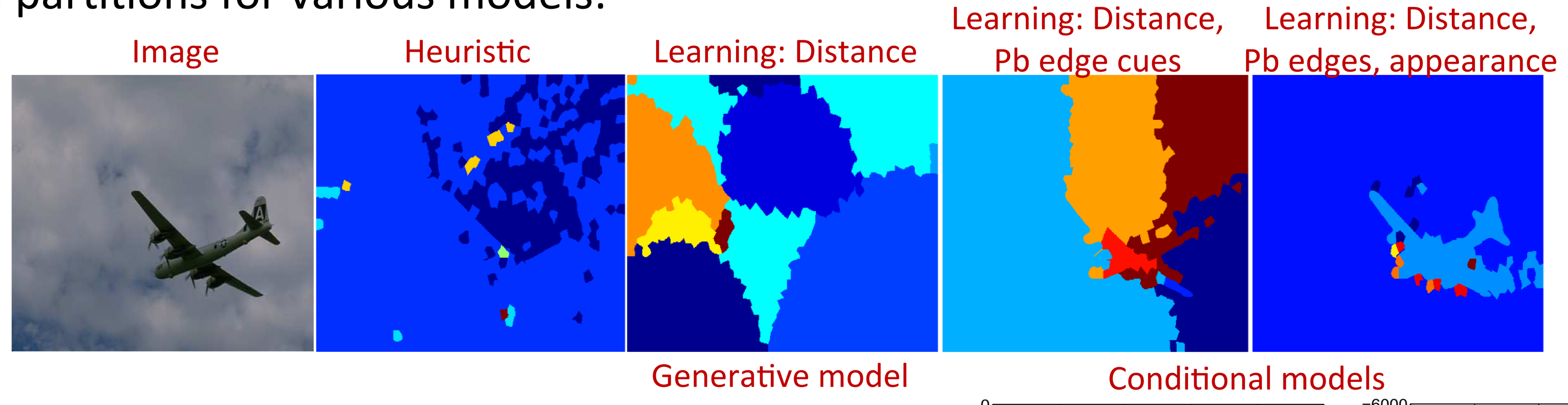
1. *Binary classification:* What is the probability that each pair of superpixels lies in the same segment? Learn via regularized logistic regression model.
2. *Co-occurrence to covariance:* There is an injective mapping between GP correlation and the probability that two superpixels are in the same segment.
3. *Guaranteeing a valid, compact model:* The pseudo-covariance constructed by considering each superpixel pair independently may not be positive definite. We use a projected gradient method to find the closest low-dimensional, valid covariance matrix.
Lowering dimension trades accuracy for speed.



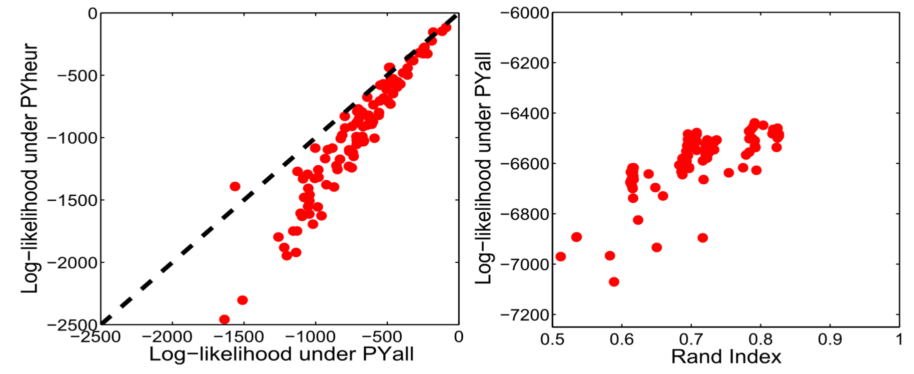


Validation of Learned Image Partition Prior

- Baseline: Heuristically specified GP covariances from Sudderth & Jordan, NIPS 2008
- Sampled partitions for various models:

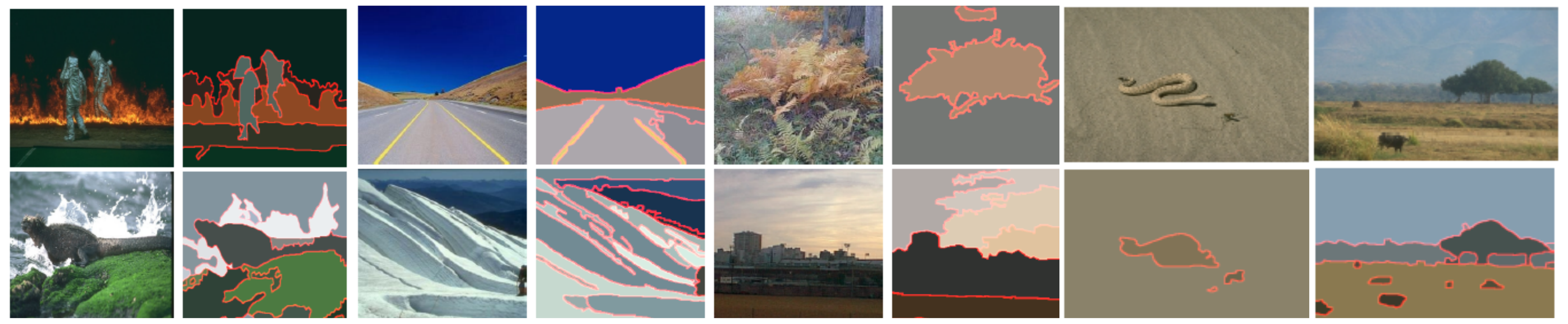


- For test images, learned models nearly always assign higher probabilities to human segmentations
- Partition probability has good positive correlation with segment quality measures like the Rand index



Results: Segmentation of Natural Images

- *BSDS300:* Berkeley Segmentation Dataset, 100 test images, 200 training images
- *LabelME:* 30 test images from each of Oliva & Torralba's eight natural scene categories
- Most probable partitions for several test images:



	BSDS300							LabelMe	
	Ncuts	MS	FH	gPb	PYheur	PYdist	PYall	gPb	PYall
PRI	0.73	0.77	0.77	0.80	0.60	0.69	0.76	0.74	0.73
segCover	0.40	0.48	0.53	0.58	0.45	0.50	0.54	0.54	0.55

Ncuts – Normalized Cuts, Shi & Malik ,PAMI00. *MS* – Mean Shift, Comaniciu & Meer, PAMI02.
FH – Graph-Based Segmentation, Felzenszwalb & Huttenlocher, IJCV04. *gPb* – Ultrametric contours from generalized Pb, Arbelaez et al. CVPR09.

- Multiple segmentations from posterior search, most probable on left: