other models with high-dimensional Gaussian variables.

Our second major contribution is a procedure for learning the various model hyperparameters, including image-dependent GP covariance functions, from example human segmentations. Using training images from the Berkeley segmentation dataset [12], we calibrate our model, and then evaluate its accuracy in segmenting various images of natural scenes [12, 16]. Our results show significant improvements over prior work with PY process models [25], and demonstrate segmentations that are both qualitatively and quantitatively competitive with state-of-the-art methods.

2. Nonparametric Bayesian Segmentation

We have two primary requirements of any segmentation model – a) it should adapt to image complexity and automatically select the appropriate number of segments and b) it should encourage spatial neighbors to cluster together. Furthermore, human segmentations of natural scenes consist of segments of widely varying sizes. It has been observed that histograms over segment areas [12] and contour lengths [19] are well explained by power law distributions. Thus a third requirement is to model this power-law behavior. In this section, we first describe our image representation and then review increasingly sophisticated models which satisfy these requirements. Finally, in Sec. 2.4, we propose a novel low-rank model which improves computational efficiency while retaining the above desiderata .

2.1. Image Representation

Each image is dicided into roughly 1,000 *superpixels* [20] using the normalized cuts spectral clustering algorithm [23]. The color of each superpixel is described using a histogram of HSV color values with $W_c = 120$ bins. We choose a non-regular quantization to more coarsely group low saturation values. Similarly, the texture of each superpixel is modeled via a local $W_t = 128$ bin texton histogram [13], using quantized band-pass filter responses. Superpixel n is then represented by histograms $x_n = (x_n^t, x_n^c)$ indicating its texture x_n^t and color x_n^c.

2.2. Pitman-Yor Mixture Models

Pitman-Yor mixture models extend traditional finite mixture models by defining a Pitman-Yor (PY) process [17] prior over the distribution of mixture components. The distributions sampled from a PY process are countably infinite discrete distributions which place mass on infinitely many mixture components. Furthermore, these discrete distributions follow a power law distribution and previous work [25] has shown that they model the distribution over human segment sizes well. There are various ways of formally defining the PY process, here we consider the stick breaking representation. Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \ldots)$, $\sum_{k=1}^{\infty} \pi_k = 1$, denote an infinite *partition* of a unit area region (in our case, an image). The Pitman-Yor process defines a prior distribution on this partition via the following *stick-breaking* construction:

$$\pi_k = w_k \prod_{\ell=1}^{k-1} (1 - w_\ell) = w_k \Big(1 - \sum_{\ell=1}^{k-1} \pi_\ell \Big) \qquad (1)$$
$$w_k \sim \mathrm{Beta}(1 - \alpha_a, \alpha_b + k\alpha_a)$$

This distribution, denoted by $\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha_a, \alpha_b)$, is defined by two hyperparameters (the discount and the concentration parameters) satisfying $0 \leq \alpha_a < 1$, $\alpha_b > -\alpha_a$. It can be shown that $\mathbb{E}[\pi_k] \propto k^{-1/\alpha_a}$, thus exhibiting the aforementioned power law distribution.

For image segmentation, each index k is associated with a different segment or region with its own appearance models $\theta_k = (\theta_k^t, \theta_k^c)$ parameterized by multinomial distributions on the W_t texture and W_c color bins, respectively. Each superpixel n then independently selects a region $z_n \sim \mathrm{Mult}(\boldsymbol{\pi})$, and a set of quantized color and texture responses according to

$$p(x_n^t, x_n^c \mid z_n, \boldsymbol{\theta}) = \mathrm{Mult}(x_n^t \mid \theta_{z_n}^t, M_n)\, \mathrm{Mult}(x_n^c \mid \theta_{z_n}^c, M_n) \qquad (2)$$

The multinomial distributions themselves are drawn from a symmetric Dirichlet prior with hyper-paramter ρ. Note that conditioned on the region assignment z_n, the color and texture features for each of the M_n pixels within superpixel n are sampled independently. The appearance feature channels provide weak cues for grouping superpixels into regions. Since, the model doesn't enforce any spatial neighborhood cues, we refer to it as the "bag of features" (*BOF*) model.

2.3. Spatially Dependent PY Mixtures

Next, we review the approach of Sudderth and Jordan [25] which extends the BOF model with spatial grouping cues. The model combines the BOF model with ideas from layered models of image sequences [28], and level set representations for segment boundaries [6].

We begin by elucidating the analogy between PY processes and layered image models. Consider the PY stick-breaking representation of Eq. (1). If we sample a random variable z_n such that $z_n \sim \mathrm{Mult}(\boldsymbol{\pi})$ where $\pi_k = w_k \prod_{\ell=1}^{k-1}(1 - w_\ell)$, it immediately follows that $w_k = \mathbb{P}[z_n = k \mid z_n \neq k-1, \ldots, 1]$. The stick-breaking proportion w_k is thus the *conditional* probability of choosing segment k, given that segments with indexes $\ell < k$ have been rejected. If we further interpret the ordered PY segments $\{k = 1, \ldots \infty\}$ as a sequence of layers, z_n can be sampled by proceeding through the layers in order, flipping biased coins (with probabilities w_k) until a layer is chosen. Given this, the probability of assignment to subsequent layers is zero; they are effectively *occluded* by the chosen "foreground" layer.

The spatially dependent Pitman-Yor process of [25] pre-