Spatial distance dependent Chinese Restaurant Process for image segmentation

Anonymous Author(s)
Affiliation
Address
email

Abstract

The *distance dependent Chinese restaurant process* (ddCRP) was recently introduced to accommodate random partitions of non-exchangeable data [1]. The ddCRP clusters data in a biased way: each data point is more likely to be clustered with other data that are near it in an external sense. This paper examines ddCRP in a spatial setting with the goal of natural image segmentation. We explore the biases of the spatial ddCRP model and propose a novel hierarchical extension better suited for producing "human like" segmentations. We then study the sensitivity of the models to various distance and appearance hyperparameters and provide the first rigorous comparison of nonparametric Bayesian models in the image segmentation domain. On unsupervised image segmentation, we demonstrate that similar performance to existing nonparametric Bayesian models is possible with substantially simpler models and algorithms.

1 Introduction

The Chinese restaurant process (CRP) is a distribution on partitions of integers [2]. When used in a mixture model, it provides an alternative representation of a Bayesian nonparametric Dirichlet process mixture—the data are clustered and the number of clusters is determined via the posterior distribution.

The CRP mixture assumes that the data are exchangeable, i.e., their order does not affect the distribution over the cluster structure. This provides a computational advantage, from the perspective of approximate inference, but is often an unrealistic assumption. The *distance dependent Chinese restaurant process* (ddCRP) was recently introduced to accommodate random partitions of non-exchangeable data [1].

The ddCRP clusters data in a biased way: each data point is more likely to be clustered with other data that are near it in an external sense. For example, when clustering time series data, points that closer in time are more likely to be grouped together. In [1], the authors develop the ddCRP mixture in general and develop posterior inference algorithms based on Gibbs sampling. While they studied the ddCRP in time-series and sequential settings, ddCRP models can be used with any type of distance and external covariates.

In this paper, we study the ddCRP in a spatial setting. We use a spatial distance function between pixels in natural images and cluster them to provide an unsupervised segmentation. With the ddCRP and a spatial distance function, the pixels are clustered in a way that biases us to find connected segments.

To demonstrate the power of this approach, we develop posterior inference algorithms for segmenting images with ddCRP mixtures. Image segmentation is an extensively studied area, which we will not attempt to survey here. Influential existing methods include approaches

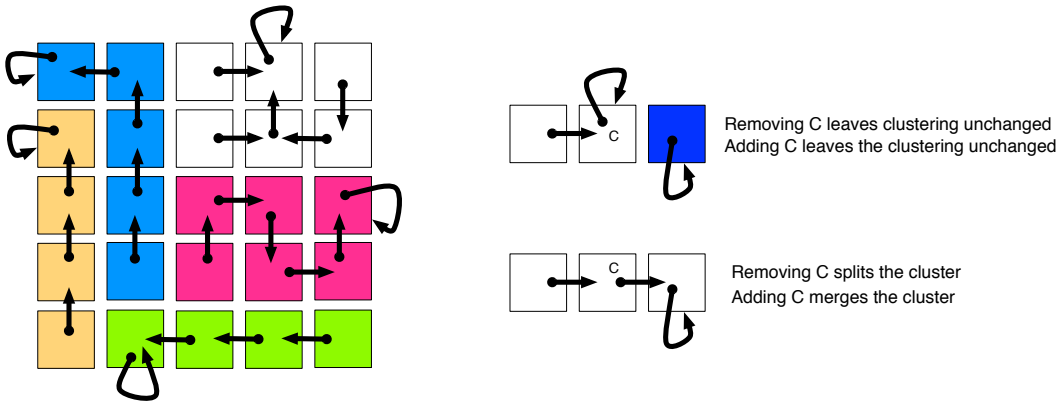

Figure 1: (Left) An illustration of the relationship between the customer assignment representation and the table assignment representation. Each square is a data point (a pixel or superpixel) and each arrow is a customer assignment. Here, the distance window is of length 1. The corresponding table assignments, i.e., the clustering of these data, is shown by the color of the data points. (Right) Intuitions behind the two cases considered by the Gibbs sampler. Consider edge C. When removed, it may leave the clustering unchanged or split a cluster. When added, it may leave the clustering unchanged or merge the cluster.

based on kernel density estimation, Markov random fields, and the normalized cut spectral clustering algorithm [3, 4]. A recurring difficulty encountered by traditional methods is the need to determine an appropriate segment resolution for each image; even among images of similar scene types, the number of observed objects can vary widely. This has usually been dealt via heuristics with poorly understood biases, or by simplifying the problem (e.g., partially specifying each image's segmentation via manual user input).

Recently, several promising methods have been proposed based on nonparametric Bayesian methods [5, 6, 7]. In particular, an approach which couples Pitman-Yor mixture models via thresholded Gaussian processes [8] has lead to very promising initial results [5], and provides a baseline for our later experiments. (Moreover, we expand on these results to analyze the full data set of which [5] uses a subset.) We analyze 800 images of different natural scene types and show that the comparatively simpler ddCRP-based algorithms perform similarly to this work. Moreover, unlike previous nonparametric Bayesian approaches, the structure of the ddCRP allows spatial connectivity of the inferred segments to (optionally) be enforced. In some applications, this is a known property of all reasonable segmentations.

Our results demonstrate the practical utility of spatial ddCRP and hierarchical ddCRP models. We also provide the first rigorous comparison of nonparametric Bayesian models in the image segmentation domain.

2 Image Segmentation with Distance Dependent CRPs

Our goal is to develop a probabilistic method to segment images of complex scenes. Image segmentation is the problem of partitioning an image into self-similar groups of adjacent pixels. Segmentation is an important step towards other tasks in image understanding, such as object recognition, detection, or tracking.

We model images as observed collections of "superpixels" [9], which are small blocks of spatially adjacent pixels. Our goal is to associate each superpixel (x_ℓ) with a cluster z_ℓ, which form the segments of that image.

Image segmentation is thus a special kind of clustering problem where the desired solution has two properties. First, we hope to find contiguous regions of the image assigned to the same cluster. Due to physical processes such as occlusion, it may be appropriate to find segments that contain two or three contiguous image regions, but we do not want a cluster that is scattered across individual image pixels. Traditional clustering algorithms, such as

k-means or probabilistic mixture models, do not account for external information such as pixel location and are not biased towards contiguous regions. Image locations have been heuristically incorporated into Gaussian mixture models by concatenating positions with appearance features in a vector [10], but the resulting bias towards elliptical regions often produces segmentation artifacts.

Second, we would like a solution that determines the number of clusters from the image. Image segmentation algorithms are typically applied to collections of images of widely varying scenes, which are likely to require different numbers of segments. Except in certain restricted domains such as medical image analysis, it is not practical to use an algorithm that requires knowing the number of segments in advance.

In the following sections, we develop a Bayesian algorithm for image segmentation based on the distance dependent Chinese restaurant process (ddCRP) mixture model [1]. Our algorithm finds spatially contiguous segments in the image and determines an image-specific number of segments from the observed data.

2.1 Chinese restaurant process mixtures

The ddCRP mixture is an extension of the traditional Chinese restaurant process (CRP) mixture. CRP mixtures provide a clustering method that determines the number of clusters from the data—they are an alternative formulation of the Dirichlet process mixture model. The assumed generative process is described by imagining a restaurant with an infinite number of tables, each of which is endowed with a parameter for some family of data generating distributions (in our experiments, Dirichlet). Customers enter the restaurant in sequence and sit at a randomly chosen table. They sit at the previously occupied tables with probability proportional to how many customers are already sitting at each; they sit at an unoccupied table with probability proportional to a scaling parameter. After the customers have entered the restaurant, the "seating plan" provides a clustering. Finally, each customer draws an observation from a distribution determined by the parameter at the assigned table.

Conditioned on observed data, the CRP mixture provides a posterior distribution over table assignments and the parameters attached to those tables. It is a distribution over clusterings, where the number of clusters is determined by the data. Though described sequentially, the CRP mixture is an exchangeable model: the posterior distribution over partitions does not depend on the ordering of the observed data.

Theoretically, exchangeability is necessary to make the connection between CRP mixtures and Dirichlet process mixtures. Practically, exchangeability provides efficient Gibbs sampling algorithms for posterior inference. However, exchangeability is not an appropriate assumption in image segmentation problems—the locations of the image pixels are critical to providing contiguous segmentations.

2.2 Distance dependent CRPs

The distance dependent Chinese Restaurant Process (ddCRP) is a generalization of the Chinese restaurant process that allows for a non-exchangeable distribution on partitions [1]. Rather than represent a partition by customers assigned to tables, the ddCRP models customers linking to other customers. The seating plan is a byproduct of these links—two customers are sitting at the same table if one can reach the other by traversing the customer assignments. As in the CRP, tables are endowed with data generating parameters. Once the partition is determined, the observed data for each customer are generated by the per-table parameters.

As illustrated in Figure 1, the generative process is described in terms of customer assignments c_ℓ (as opposed to partition assignments or tables, z_ℓ). The distribution of customer assignments is

$$p\left(c_i = j \mid D, f, \alpha\right) \propto \begin{cases} f\left(d_{ij}\right) & j \neq i, \\ \alpha & j = i. \end{cases} \tag{1}$$

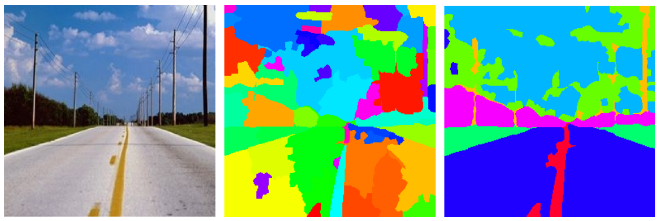

Figure 2: Segmentations produced by the ddCRP (*center*) and rddCRP (*right*) models.

Here d_{ij} is a distance between data points i and j and $f(d)$ is called the decay function. The decay function mediates how the distance between two data points affects their probability of connecting to each other, i.e., their probability of belonging to the same cluster.

Details of the ddCRP are found in [1]. We note that the traditional CRP is an instance of a ddCRP. However, in general, the ddCRP does not correspond to a model based on a random measure, like the Dirichlet process. The ddCRP is appropriate for image segmentation because it can naturally account for the spatial structure of the superpixels through its distance function. We use a spatial distance between pixels to enforce a bias towards contiguous clusters. Though the ddCRP has been previously described in general, only time-based distances are studied in [1].

Restaurants represent images, tables represent segment assignment, and customers represent superpixels. The distance between superpixels is modeled as the number of hops required to reach one superpixel from another, with hops being allowed only amongst spatially neighboring superpixels. A "window" decay function of width a, $f(d) = \mathbb{1}[d \leq a]$ is used. If $a = 1$ superpixels can only directly connect to adjacent superpixels. Note this doesn't explicitly restrict the size of segments, because any pair of pixels for which one is *reachable* from the other (i.e., in the same connected component of the customer assignment graph) are in the same image segment. Furthermore, the segments are *guaranteed* with probability one to form a spatially connected subset of the image, a property not enforced by other Bayesian nonparametric approaches [5, 6].

The full generative process for the observed features $x_{1:N}$ within an image with N superpixels is as follows:

1. For each table, sample parameters $\phi_k \sim G_0$.
2. For each customer, sample customer assignment
$$c_n \sim ddCRP\left(\alpha, f, D\right).$$
This determines the cluster assignments $z_{1:N}$, and thus the segmentation.
3. Sample observed data, $x_n \sim P(\cdot \mid \phi_{z_n})$.

The customer assignments are sampled using the spatial distance between pixels. The partition structure—derived from the customer assignments—is used to sample the observed image features. Given an image, the posterior distribution of the customer assignments induces a posterior over the cluster structure. This provides the segmentation. See Figure 1 for an illustration of the customer assignments and their derived table assignments in a segmentation setting.

The data generating distribution of the observed features studied in Section 4 is multinomial, with separate distributions for color and texture. We assume conjugate Dirichlet priors on these cluster parameters.

The "region" level model The ddCRP model, when applied to an image, produces a collection of contiguous patches (tables) homogeneous in color and texture features (Figure 2). While such segmentations are useful for various applications [10], they do not reflect he statistics of human produced segmentations, which contain larger regions [11]. We could bias our model to produce such regions by either increasing the window size a or by introducing

a hierarchy wherein the produced patches are grouped into a small number of regions. This region level model has each patch (table) associated with a region k from a set of potentially infinite regions. Each region in turn is asscoiated with an appearance model ϕ_k. The corresponding generative process is described as follows:

- For each customer, sample customer assignments $c_n \sim ddCRP(\alpha, f, D)$. This determines the table assignments $t_{1:N}$.
- For each table t, sample region assignments $k_t \sim CRP(\gamma, H)$.
- For each region, sample parameters : $\phi_k \sim G_0$.
- Finally, sample observed data $x_n \sim P(\cdot \mid \phi_{z_n})$, where $z_n = k_{t_n}$

Note, that the region level model (rddCRP) is a simple extension to the hierarchical Dirichlet process (HDP) [12], with the image partition being drawn from a ddCRP instead of a CRP. Also, unlike HDP our region parameters are not shared amongst images although it is trivial to generalize to the shared case. Figure 3 (top row) plots samples from the rddCRP and the ddCRP prior with increasing a. Observe that while the samples from rddCRP are similar to human image partitions, increasing a produces increasingly noisy partitions.

3 Inference with Gibbs Sampling

A segmentation of an observed image is found by posterior inference. The problem is to compute the conditional distribution of the latent variables—the customer assignments $c_{1:N}$—conditioned on the observed image features $x_{1:N}$, the scaling parameter α, the distances between pixels D, the window size a, and the base distribution hyperparameter λ:

$$p(c_{1:N} \mid x_{1:N}, \alpha, d, a, \lambda) = \frac{\left(\prod_{n=1}^{N} p(c_n \mid D, a, \alpha)\right) p(x_{1:N} \mid z(c_{1:N}), \lambda)}{\sum_{c_{1:N}} \left(\prod_{n=1}^{N} p(c_n \mid D, a, \alpha)\right) p(x_{1:N} \mid z(c_{1:N}), \lambda)} \tag{2}$$

where $z(c_{1:N})$ is the cluster representation that is derived from the customer representation $c_{1:N}$. Notice again that the prior term uses the customer representation to take into account distances between data points; the likelihood term uses the cluster representation.

The posterior in Equation (2) is not tractable to compute. The difficulty is in the denominator, which requires summing over a combinatorial number of terms. We appeal to Gibbs sampling, a Monte Carlo Markov chain (MCMC) method, to approximate this posterior.

The idea behind MCMC methods is to define a Markov chain over the latent variables whose stationary distribution is the posterior of interest [13]. The Gibbs sampler, when available, is a particularly simple form of MCMC. In Gibbs sampling, we define the chain by iteratively sampling each latent variable c_i conditioned on the others and the observations,

$$p(c_i \mid c_{-i}, x_{1:N}, D, \alpha, \lambda) \propto p(c_i \mid D, \alpha) p(x_{1:N} \mid z(c_{1:N}), \lambda). \tag{3}$$

The prior term is given in Equation (1). We can decompose the likelihood term as follows:

$$p(x_{1:N} \mid z(c_{1:N}), \lambda) = \prod_{k=1}^{K(c_{1:N})} p(x_{z(c_{1:N})=k} \mid z(c_{1:N}), \lambda). \tag{4}$$

We have introduced notation to more easily move from the customer representation—the primary latent variables of our model—and the cluster representation. We let $K(c_{1:N})$ be the number of unique clusters in the customer assignments, $z(c_{1:N})$ be the cluster assignments derived from the customer assignments, and $x_{z(c_{1:N})=k}$ be the collection of observations assigned to the kth cluster. We assume that the base distribution G_0 is marginalized out. This is possible when the base distribution is conjugate to the data generating distribution, e.g. Dirichlet for multinomial.

Sampling from Equation (3) happens in two stages. First, we remove the customer link c_i from the current configuration. Then, we consider the prior probability of each possible value of c_i and how it changes the likelihood term, by moving from $p(x_{1:N} \mid z(c_{-i}), \lambda)$ to $p(x_{1:N} \mid z(c_{1:N}), \lambda)$.

In the first stage, removing c_i either leaves the cluster structure intact, i.e., $z(c^{\text{old}}_{1:N}) = z(c_{-i})$, or splits the cluster assigned to data point i into two. In the second stage, randomly reassigning c_i either leaves the cluster structure intact, i.e., $z(c_{-i}) = z(c_{1:N})$, or joins the cluster assigned to data point i to another. See Figure 1 for a picture of these cases. Via these moves, the sampler explores the space of possible segmentations.

Let ℓ and m be the indices of the tables that are joined to index k. We first remove c_i, possibly splitting a cluster. Then we sample from

$$p(c_i \mid c_{-i}, x_{1:N}, D, \alpha, \lambda) \propto \begin{cases} p(c_i \mid D, \alpha)\Gamma(x, z, \lambda) & \text{if } c_i \text{ joins } \ell \text{ and } m; \\ p(c_i \mid D, \alpha) & \text{otherwise,} \end{cases} \tag{5}$$

where

$$\Gamma(x, z, \lambda) = \frac{p(x_{z(c_{1:N})=k} \mid \lambda)}{p(x_{z(c_{1:N})=\ell} \mid \lambda)p(x_{z(c_{1:N})=m} \mid \lambda)}. \tag{6}$$

This defines a Markov chain whose stationary distribution is the posterior of the spatial ddCRP defined in Section 2. Though our presentation is slightly different, this algorithm is equivalent to the one developed for ddCRP mixtures in [1].

In rddCRP, the algorithm for sampling the customer indicators is nearly the same, but with two differences. First, when c_i is removed, it may spawn a new cluster. In that case, the region identity of the new table must be sampled from the region level CRP. Second, the likelihood term in Equation (4) depends only on the superpixels in the image assigned to the segment in question. In rddCRP, it also depends on other superpixels assigned to segments that are assigned to the same region. Finally, rddCRP also requires region assignments which are sampled from

$$p(k_t = l \mid k_{-t}, x_{1:N}, t(c_{1:N}), \gamma, \lambda) \propto \begin{cases} m_l^{-t} p(x_t \mid x_{-t}, \lambda) & \text{if } l \text{ is used;} \\ \gamma p(x_t \mid \lambda) & \text{if } l \text{ is new,} \end{cases} \tag{7}$$

where x_t is the set of customers sitting at table t, x_{-t} is the set of all customers associated with region l barring x_t, m_l^{-t} is the number of tables associated with region l barring x_t.

4 Empirical Results

We compare the performance of the ddCRP to manual segmentations of images drawn from eight natural scene categories [14]. Non-expert users segmented each image into polygonal shapes, and labeled them as distinct objects. The collection, which is available from LabelMe [11], contains a total 2,688 images.[1] We randomly select 100 images from each category.

This image collection has been previously used to analyze an image segmentation method based on spatially dependent Pitman-Yor processes [5], and we compare both methods using an identical feature set. Each image is first divided into approximately 1000 superpixels [9, 15] [2]. To this end we use the normalized cut algorithm [4].[3] We describe the texture of each superpixel via a local texton histogram [16], using band-pass filter responses quantized to 128 bins. A 120-bin HSV color histogram is also computed. Each superpixel i is summarized via these histograms x_i.

Our goal is to make a controlled comparison to alternative nonparametric Bayesian methods on a challenging task. Performance is assessed via agreement with held out human segmentations, via rand index. We also present segmentation results for qualitative evaluation, in Figures 3 and 4 .

4.1 Sensitivity to Hyperparameters

Our models are governed by the CRP concentration parameters γ and α, the appearance hyperparameter $\lambda = (\lambda_0, ...\lambda_0)$ and the window size a. γ has little impact on the segmentation

[1] *labelme.csail.mit.edu/browseLabelMe/*

[2] *www.cs.sfu.ca/~mori/*

[3] *www.eecs.berkeley.edu/Research/Projects/CS/vision/*

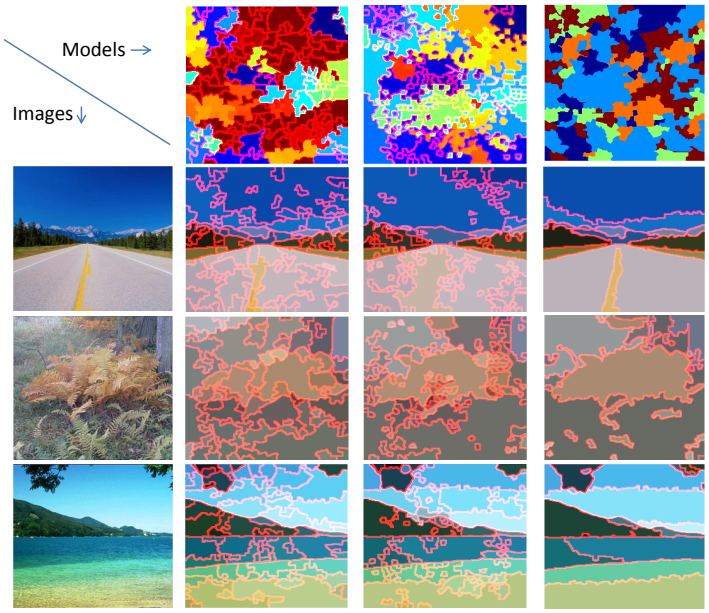

Figure 3: Segmentations produced by ddCRP and rddCRP. The columns (from left to right) display, natural images, segmentations produced by ddCRP with $a = 1$, with $a = 2$ and rddCRP with $a = 1$. The top row displays partitions sampled from the three priors. The three prior partitions have 130, 54 and 5 segments respectively.

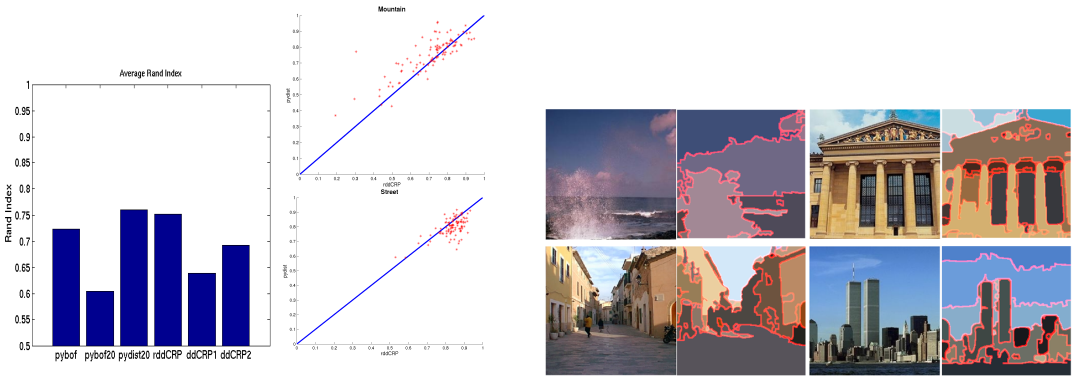

Figure 4: *Left* Average performance across the dataset, as measured by Rand index. *Center* pydist20 and rddCRP, the two top performing models, are compared via a scatter plot of Rand indexes for the *Mountain* and *Street* categories. *Right* Segmentations produced by the rddCRP model on a random subset of the dataset.

results, due to the high-dimensional and informative image features. For all our experiments we set γ to 1. α and λ_0 induce opposing biases, a small α encourages larger segments while a large λ_0 encourages larger segments. We found $\alpha = 1e-8$ and $\lambda_0 = 20$ to work well. The most influential prior parameter is a, the effect of which is visualized in figure 3. For the ddCRP model, setting $a = 1$ (*ddCRP1*) produces a set of contiguous segments. Increasing to $a = 2$ (*ddCRP2*), results in fewer segments, but the produced segments are spatially fragmented. The phenomenon is further exacerbated with larger values of a. The rddCRP model groups together segments produced by ddCRP. If the segments produced by ddCRP are poor, rddCRP has a hard time recovering meaningful partitions. Not surprisingly then, rddCRP performs best with $a = 1$.

4.2 Image Segmentation Performance

We now quantitatively measure the performance of our models. The ddCRP and the rddCRP samplers were run for 100 and 500 iterations respectively. Both samplers displayed rapid mixing and often stabilized withing the first 50 iterations. Note that similar rapid mixing has been observed in other applications of the ddCRP [1].

We also compare to two previous models [5]: a Pitman-Yor mixture model with no spatial dependence (*pybof20*), and a Pitman-Yor mixture with spatial coupling induced via thresholded Gaussian processes (*pydist20*). To control the comparison as much as possible, the Pitman-Yor models are tested with identical features and base measure β, and other hyperparameters as in [5]. We also compare to the non-spatial Pitman-Yor with $\lambda_0 = 1$, the best bag-of-feature model in our experiments (*pybof*). We employ non-hierarchical versions of the Pitman-Yor models, so that each image is analyzed independently, and perform inference via the previously developed mean field variational method.

The performance summary is presented in Fig. 4. Not surprisingly, rddCRP outscores both versions of the ddCRP model, in terms of Rand index. Nevertheless, the patchy ddCRP1 segmentations are interesting for applications where segmentation is an intermediate step rather than the final goal. The bag of features model with $\lambda_0 = 20$ performs poorly, *pybof* with optimized λ performs reasonably but still falls short of the region level spatial models.

The spatial Pitman-Yor model and rddCRP perform similarly, with the Pitman-Yor outperforming rddCRP on some categories and vice versa. The scatter plots in Fig. 4 provide insights into when one model outperforms the other. Here, we have plotted the Rand indexes of images from the mountain and street categories. For the street images rddCRP is better, while for images containing mountains spatial PY is superior. In general, street scenes contain more objects, many of which are small. For the spatial PY model to be successful on such scenes it would require it's Gaussian Process functions to change rapidly over the image. For an apples to apples comparison with hddCRP, we tested a version of the spatial PY model employing a covariance functions dependent only on spatial distance. Such covariances are not flexible enough to allow rapid changes in the GP functions. Further performance improvements were demonstrated in [5] via a conditionally specified covariance, which depends on detected image boundaries. Similar conditional specification of the ddCRP distance function is also possible, and is a promising direction for future research.

Finally, we note that the ddCRP (and rddCRP) models proposed here are far simpler than the spatial Pitman-Yor model, both in terms of model specification and inference. The ddCRP based models only require pairwise superpixel distances to be specified, as opposed to a positive semi definite covariance matrix over all superpixels required by the spatial Pitman-Yor model. Furthermore, its thresholded Gaussian process leads to a complex likelihood function, for which inference is a significant challenge. In contrast, ddCRP inference is carried out through a straightforward sampling algorithm,[4] and thus might provide a simpler foundation for building richer models of visual scenes.

5 Discussion

We study the properties of spatial distance dependent Chinese restaurant processes and apply them to the problem of image segmentation. We show that the spatial ddCRP model is particularly well suited for segmenting an image into a collection of contiguous patches. Unlike previous Bayesian nonparametric models, it can produce segmentations which guarantee spatially connected segments. To go from patches to human like segmentations, we develop a hierarchical distance dependent Chinese restaurant process model. This hierarchical model is demonstrated to achieve performance similar to state-of-the-art nonparametric Bayesian segmentation models while using a simpler model and a substantially simpler inference algorithm.

[4] In our Matlab implementations, the core ddCRP code was less than half as long as the corresponding PY code. For the ddCRP, the computation time was 1 minute per iteration and about 20 minutes per image (but most images converged in 2 minutes). The PY code, which is based on variational approximations, took 12 minutes per image.

References

[1] D. M. Blei and P. I. Frazier. Distant dependent chinese restaurant process. *arXiv:0910.1022v1*, 2009.

[2] J. Pitman. *Combinatorial Stochastic Processes*. Lecture Notes for St. Flour Summer School. Springer-Verlag, New York, NY, 2002.

[3] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. PAMI*, 22(8):888–905, 2000.

[4] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. *CVPR*, 2:54–61, 2003.

[5] E. B. Sudderth and M. I. Jordan. Shared segmentation of natural scenes using dependent pitman-yor processes. *NIPS 22*, 2008.

[6] P. Orbanz and J. M. Buhmann. Smooth image segmentation by nonparametric Bayesian inference. In *ECCV*, volume 1, pages 444–457, 2006.

[7] Lan Du, Lu Ren, David Dunson, and Lawrence Carin. A bayesian model for simultaneous image clustering, annotation and object segmentation. In *NIPS 22*, pages 486–494. 2009.

[8] J. A. Duan, M. Guindani, and A. E. Gelfand. Generalized spatial Dirichlet process models. *Biometrika*, 94(4):809–825, 2007.

[9] X. Ren and J. Malik. Learning a classification model for segmentation. *ICCV*, 2003.

[10] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *PAMI*, 24(8):1026–1038, August 2002.

[11] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database web-based tool for image annotation. *IJCV*, 77:157–173, 2008.

[12] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet process. *Journal of American Statistical Association*, 25(2):1566 – 1581, 2006.

[13] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag, New York, NY, 2004.

[14] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145 – 175, 2001.

[15] G. Mori. Guiding model search using segmentation. *ICCV*, 2005.

[16] D. R. Martin, C.C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. PAMI*, 26(5):530–549, 2004.