Quantitative performance is summarized in Figure 4. The 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; with optimized $\lambda_0 = 1$ it is better, but still inferior to the best spatial models.

In general, the spatial PY and rddCRP perform similarly. The scatter plots in Fig. 4, which show Rand indexes for each image from the mountain and street categories, provide insights into when one model outperforms the other. 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, and thus disfavored by the smooth Gaussian processes underlying the PY model. To most fairly compare priors, we have tested a version of the spatial PY model employing a covariance function that depends only on spatial distance. Further performance improvements were demonstrated in [10] via a conditionally specified covariance, which depends on detected image boundaries. Similar conditional specification of the ddCRP distance function is a promising direction for future research.

Finally, we note that the ddCRP (and rddCRP) models proposed here are far simpler than the spatial PY model, both in terms of model specification and inference. The ddCRP models only require pairwise superpixel distances to be specified, as opposed to the positive definite covariance function required by the spatial PY model. Furthermore, the PY model's usage of thresholded Gaussian processes 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,[5] and thus may provide a simpler foundation for building rich models of visual scenes.

5 Discussion

We have studied the properties of spatial distance dependent Chinese restaurant processes, and applied them to the problem of image segmentation. We showed 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 with guaranteed spatial connectivity. To go from patches to coarser, human-like segmentations, we developed a hierarchical region-based ddCRP. This hierarchical model achieves performance similar to state-of-the-art nonparametric Bayesian segmentation algorithms, using a simpler model and a substantially simpler inference algorithm.

References

[1] D. M. Blei and P. I. Frazier. Distant dependent chinese restaurant processes. *Journal of Machine Learning Research*, 12:2461–2488, August 2011.

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

[3] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6(6):721–741, November 1984.

[4] Richard Socher, Andrew Maas, and Christopher D. Manning. Spectral chinese restaurant processes: Nonparametric clustering based on similarities. In *Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

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

[6] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, pages 603–619, 2002.

[7] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 309–314, 2004.

[5] 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 convergence typically happened after only a few iterations. The PY code, which is based on variational approximations, took 12 minutes per image.