Unsupervised Scene Analysis Using Semiparametric Bayesian Models

Dominik Joho, Gian Diego Tipaldi, Nikolas Engelhard, Cyrill Stachniss, and Wolfram Burgard

Department of Computer Science, University of Freiburg, Germany
{joho,tipaldi,engelhar,stachnis,burgard}@informatik.uni-freiburg.de

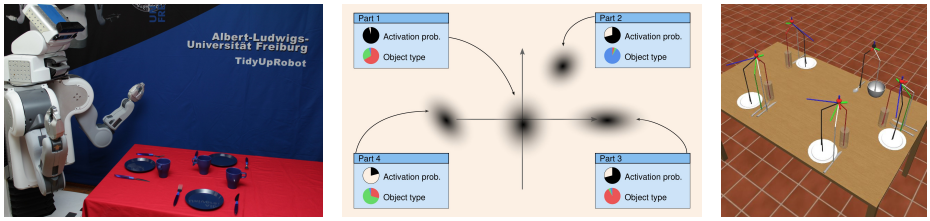
Abstract. Robots operating in domestic environments need to deal with a variety of different objects. Usually, these objects are not placed randomly. For example, objects on a breakfast table such as plates and knives typically occur in recurrent configurations. We propose a novel hierarchical generative model to infer the latent groups of objects in a scene. The proposed model is a combination of Dirichlet processes and beta processes, which allows for a rigorous probabilistic treatment of the unknown dimensionality of the parameter space. We address a set of different tasks in scene understanding ranging from unsupervised scene segmentation to generation of a new scene and completion of a partially specified scene. We use Markov chain Monte Carlo (MCMC) techniques for inference and present experiments with simulated as well as real-world data obtained from a Kinect RGB-D camera.

Keywords: scene analysis, Bayesian nonparametrics, Dirichlet process, beta process, unsupervised learning, hierarchical models

1 Introduction

Imagine a person laying a breakfast table and the person gets interrupted so that it cannot continue with the breakfast preparation. A service robot (Fig. 1) should be able to proceed laying the table without receiving specific instructions. It faces several challenges: how to infer the total number of covers, how to infer which objects are missing, and how should the missing object be arranged. For this, the robot should not require any user-specific pre-programmed model but should ground its decision based on the breakfast tables it had seen in the past.

We address the problem of scene understanding given a set of unlabeled training scenes and frame it as an unsupervised learning problem. The *key contribution* of our approach is the definition of a novel hierarchical semiparametric Bayesian model to represent the scene structure in terms of object groups. In our model, these object groups are called *meta-objects* and a meta-object is defined as a collection of parts. We assume that each scene contains an unknown number of latent meta-objects. In the breakfast table scenario, a place cover can be seen as a meta-object of a certain type that, for example, generates the observable objects plate, knife, and mug. We do not assume that all instances of the same



(a) Observed scene. (b) Meta-object type. (c) Parsed scene.

Fig. 1. (a) A scene typically contains several objects and the task is to infer the latent meta-object instances. A meta-object *instance* is a constellation of observable objects and corresponds to a draw from a meta-object *type*. **(b)** A meta-object *type* is a collection of parts, each consisting of a Gaussian distribution, a multinomial distribution over the observable object types, and a binary activation probability. The number of parts and their parametrization are initially unknown. **(c)** A parsed scene includes the meta-object poses and types and the associations of the objects to meta-object parts.

meta-object are identical. Instances differ in the sense that some parts may be missing and that the individual parts may not be arranged in the same way.

When specifying a generative model for our problem, we have the difficulty that the dimensionality of the model is part of the learning problem. This means, that besides learning the parameters of the model, like the pose of a meta-object, we additionally need to infer the *number* of meta-objects, parts, etc. Model selection approaches would be intractable in our case due to the huge number of possible models we would need to consider. We therefore follow another approach, motivated by recent developments in the field of hierarchical nonparametric Bayesian models [1,2,3] based on the Dirichlet process and the beta process. These models are able to adjust their dimensionality according to the given data, thereby sidestepping the need to select among several finite-dimensional model alternatives. Using Markov chain Monte Carlo (MCMC) techniques to sample from the posterior distribution of the model given the training scenes, we are able to parse these scenes and learn a generative model that can be used to parse new scenes or complete partial scenes. Though we consider table-top scenes as an application scenario, our model is general and could be applied to other scenarios as well.

References

1. T. L. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems (NIPS)*. 2006.
2. E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Describing Visual Scenes Using Transformed Objects and Parts. *Int. Journal of Computer Vision*, 77(1–3):291–330, 2008.
3. Y. W. Teh and M. I. Jordan. Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press, 2010.