Shared Segmentation of Natural Scenes Using Dependent Pitman-Yor Processes

Erik B. Sudderth and Michael I. Jordan
Electrical Engineering & Computer Science, University of California, Berkeley
sudderth@cs.berkeley.edu, jordan@cs.berkeley.edu

Abstract

We develop a statistical framework for the simultaneous, unsupervised segmentation and discovery of visual object categories from image databases. Examining a large set of manually segmented scenes, we show that object frequencies and segment sizes both follow power law distributions, which are well modeled by the Pitman–Yor (PY) process. This nonparametric prior distribution leads to learning algorithms which discover an unknown set of objects, and segmentation methods which automatically adapt their resolution to each image. Generalizing previous applications of PY processes, we use Gaussian processes to discover spatially contiguous segments which respect image boundaries. Using a novel family of variational approximations, our approach produces segmentations which compare favorably to state-of-the-art methods, while simultaneously discovering categories shared among natural scenes.

1 Introduction

Images of natural environments contain a rich diversity of spatial structure at both coarse and fine scales. We would like to build systems which can automatically *discover* the visual categories (e.g., foliage, mountains, buildings, oceans) which compose such scenes. Because the "objects" of interest lack rigid forms, they are poorly suited to traditional, fixed aspect detectors. In simple cases, topic models can be used to cluster local textural elements, coarsely representing categories via a bag of visual features [1, 2]. However, spatial structure plays a crucial role in general scene interpretation [3], particularly when few labeled training examples are available.

One approach to modeling additional spatial dependence begins by precomputing one, or several, *segmentations* of each input image [4–6]. However, low-level grouping cues are often ambiguous, and fixed partitions may improperly split or merge objects. Markov random fields (MRFs) have been used to segment images into one of several known object classes [7, 8], but these approaches require manual segmentations to train category-specific appearance models. In this paper, we instead develop a statistical framework for the *unsupervised* discovery and segmentation of visual object categories. We approach this problem by considering sets of images depicting related natural scenes (see Fig. 1(a)). Using color and texture cues, our method simultaneously groups dense features into spatially coherent segments, and refines these partitions using shared appearance models. This extends the *cosegmentation* framework [9], which matches two views of a single object instance, to simultaneously segment multiple object categories across a large image database. Some recent work has pursued similar goals [6, 10], but robust object discovery remains an open challenge.

Our models are based on the *Pitman–Yor* (PY) process [11], a nonparametric Bayesian prior on infinite partitions. This generalization of the *Dirichlet process* (DP) leads to heavier-tailed, power law distributions for the frequencies of observed objects or topics. Using a large database of manual scene segmentations, Sec. 2 demonstrates that PY priors closely match the true distributions of natural segment sizes, and frequencies with which object categories are observed. Generalizing the hierarchical DP [12], Sec. 3 then describes a *hierarchical Pitman–Yor* (HPY) mixture model which shares "bag of features" appearance models among related scenes. Importantly, this approach coherently models uncertainty in the *number* of object categories and instances.