Nonparametric Learning for Layered Segmentation of Natural Images

Soumya Ghosh and Erik B. Sudderth
Department of Computer Science, Brown University, Providence, RI, USA
sghosh@cs.brown.edu, sudderth@cs.brown.edu

Abstract

We explore recently proposed Bayesian nonparametric models of image partitions, based on spatially dependent Pitman-Yor processes. These models are attractive because they adapt to images of varying complexity, successfully modeling uncertainty in the structure and scale of human segmentations of natural scenes. By developing substantially improved inference and learning algorithms, we achieve performance comparable to state-of-the-art methods. For learning, we show how the Gaussian process (GP) covariance functions underlying these models can be calibrated to accurately match the statistics of example human segmentations. For inference, we develop a stochastic search-based algorithm which is substantially less susceptible to local optima than conventional variational methods. Our approach utilizes the expectation propagation algorithm to approximately marginalize latent GPs, and a low rank covariance representation to improve computational efficiency. Experiments with two benchmark datasets show that our learning and inference innovations substantially improve segmentation accuracy. By hypothesizing multiple partitions for each image, we also take steps towards capturing the variability of human scene interpretations.

1. Introduction

Image segmentation algorithms partition images into spatially coherent, approximately homogeneous regions. Segmentations provide an important mid-level representation which can be leveraged for various vision tasks including object recognition [11], motion estimation [26], and image retrieval [4]. Despite significant research [23, 5, 7, 15, 2], segmentation remains a largely unsolved problem. One major challenge is to move beyond seeking a single "optimal" image partition, and to recognize that while there are commonalities among multiple human segmentations of the same image, there is also substantial variability [12].

Most existing segmentation algorithms are endowed with a host of tunable parameters; a particular configuration may work well on some images, and poorly on others. Often these parameters are tuned via manual experimentation, or expensive validation experiments. Noting this issue, Russell et al. [21] produced a "soup of segments" by varying the parameters of the normalized cuts algorithm, and collecting the range of observed outputs. Others have used agglomerative clustering methods to produce a nested tree of segmentations [2]. A limitation of these approaches is that they do not provide any image-specific estimate of which particular segmentations are most accurate.

In this paper, we instead pursue a Bayesian nonparametric statistical approach to modeling segmentation uncertainty. We reason about prior and posterior distributions on the space of image partitions, and thus consider segmentations of all possible resolutions. In contrast with parametric segmentation models based on finite mixtures [4, 1, 22] or Markov random fields [8], we do *not* need to pre-specify the number of segments. Our inference algorithm automatically provides calibrated estimates of the relative probabilities of segmentations with varying numbers of regions.

Because we define a consistent probabilistic model and not just a segmentation procedure, our approach is a natural building block for more sophisticated models. We improve earlier work on spatially dependent Pitman-Yor (PY) processes [25], which was motivated by the problem of jointly segmenting multiple related images. This PY model was later extended to allow prediction of semantic segment labels, given supervised annotations of objects in training images [24]. Here we focus on the problem of segmenting single images containing unknown object categories.

The model we consider is a minor variation on the dependent PY process of Sudderth and Jordan [25], which captures the power law distribution of human image segments via a stick-breaking construction, and uses Gaussian processes (GPs) to induce spatial dependence. Our first major contribution is a new posterior inference algorithm that is far less susceptible to local optima than previous mean field variational methods [25]. Our algorithm combines a discrete stochastic search, capable of making large moves in the space of image partitions, with an accurate higher-order variational approximation (based on expectation propagation [14]) to marginalize latent GPs. We improve computational efficiency via a low rank representation of the GP covariance, an innovation that could be applicable to many