Author: Rossum, A.C. van
Title: Nonparametric Bayesian methods in robotic vision
Issue date: 2021-06-03

CHAPTER 1

INTRODUCTION

Contents The thesis addresses nonparametric Bayesian methods in robotic vision. Nonparametric Bayesian models can be simultaneously employed to perform inference over the number of entities observed and over the shape or nature of these entities. This chapter introduces nonparametric Bayesian models, the research methodology based on the Bayesian methodology, the main contribution towards robotic vision, and the general organization of the thesis.

Outline The scope of this thesis is to apply nonparametric Bayesian methods to robotic vision (Section 1.1). Bayesian nonparametric models define entities together with noise in such a way that inference can be performed in an optimal manner (Section 1.2). Particular problems in robotic vision that can benefit from Bayesian nonparametric methods are formulated and detailed (Section 1.3). The research methodology is described (Section 1.4). Our main contribution is to introduce nonparametric Bayesian models in robotic vision (Section 1.5). At the end of this chapter the organization of the thesis is given (Section 1.6).

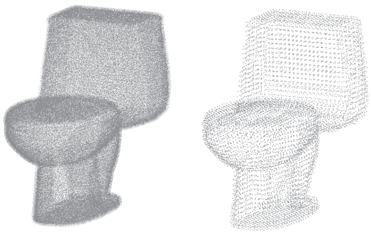
1.1 Scope of the Thesis

In the thesis, modern Bayesian nonparametric methods are used to answer long-standing questions within computer and robotic vision. The following three challenging questions are typical examples. Is there a Bayesian form of line detection rather than applying the traditional Hough transform? Which of the nonparametric Bayesian priors can be used to detect multiple features simultaneously? What are efficient inference methods for these priors?

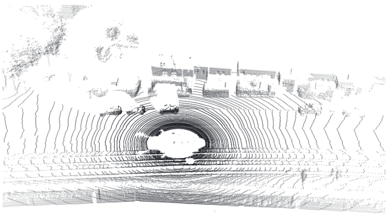
The scope of the thesis is the transfer of knowledge on Bayesian nonparametrics to well-described application domains. It will not establish a new body of work around a new family of stochastic processes. The detailed application of complex models towards robotic vision is expected to help and encourage people in entirely different application domains, such as collaborative filtering, search engine optimization, and audio processing. All these different applications do not always need dedicated algorithms, but do deserve and can exploit the same optimal general inference techniques from Bayesian nonparametrics.

1.2 Bayesian Nonparametrics

In robotic vision (computer vision and depth perception) traditionally custom-made algorithms have been developed for a given task. There are specific methods to detect corners (e.g., Förstner and Gülch, 1987; Harris and Stephens, 1988; Shi and Tomasi, 1994), to detect edges (e.g., Sobel, 1970; Canny, 1986), to detect features (e.g., Hough, 1962), and to describe features (e.g., Lowe, 1999; Dalal and Triggs, 2005; Bay et al., 2006).



(a) The PointNet40 dataset (Wu et al., 2015) has forty examples of common objects in the house of which the toilet is an example (Garcia-Garcia et al., 2016).



(b) The KITTI dataset (Geiger et al., 2012) has point clouds made by the Velodyne lidar. It is data that can be used on self-driving cars to become more aware of their surroundings. This example is from (Ioannou et al., 2012).

(c) Example of a point cloud generated by the Kinect depth sensor. This type of sensor can be used for applications indoors, e.g. autonomous cleaning robots. This particular example is a test on how a transparent sheet will show up on such a structured light sensor.

Figure 1.1: Examples of point clouds.

On the one hand, it is desirable that such sophisticated methods are generalizable to other application domains. On the other hand, it is important to take particular information about an application domain into account. The methods described in the previous paragraph are limited to their specific task. An example of limited generalizability can be found in the Hough transform. The Hough transform can be used to detect lines, but the way inference is performed in the algorithm does limit its application to basic forms of object detection. An example of limited specificity can be found in linear regression. Linear regression assumes a linear relationship between input and output variables.

Both generalization and specificity are formalized by a Bayesian model. A Bayesian model is general, because the problem modeled by it can be solved with general inference methods. One of such general inference methods is a Markov-Chain Monte Carlo method. A Bayesian model is also specific in that it can incorporate application-specific know-how by the definition of priors. This power of Bayesian models can be seen in many disciplines, from robotic localization (Blanco et al., 2010), and dynamical systems (Dubbeldam et al., 2011), to forensic and legal arguments on evidence (Wieten et al., 2019).

Typical problems in robotic vision will be about the recognition of several objects, multiple shapes, or objects that have multiple parts. Models that represent such objects do not have knowledge about the number of such objects, shapes, or parts. To incorporate application-specific know-how on the number of objects it is possible to define a prior that assigns a probability to this quantity. The number of objects can even be potentially infinite. The Bayesian models that define a prior on the number of objects, shapes, or parts are called nonparametric Bayesian models. This means that in contrast with conventional methods such as k-means clustering (Forgy, 1965; Lloyd, 1982) the number of objects does not need to be predefined.

1.3 Problem Statement and Research Questions

Many methods in robotics - and in particular in robotic vision - have been developed in times where computational resources were limited. Then, highly optimized algorithms have been developed, leveraging pecularities of the application domain. Recent advances in Bayesian methods, both with respect to concept development, as well as computational efficient solution strategies, now open up new ways to solve old problems (Seeger, 2000; Murphy, 2012; Huszár and Duvenaud, 2012; Gal and Ghahramani, 2016; Mandt et al., 2017). However, extending only the old methods themselves would lead to ad hoc solution strategies that will miss benefits from potential optimal and more widely applicable algorithms.

This observation leads us to the formulation of our problem statement (PS).

> **PS:** *How can robotic vision problems effectively be generalized and their structure exploited in a wider Bayesian framework?*

The problem statement is rather general. In our research, we focus on robotic vision, in the form of point cloud recognition and depth perception. In particular, we look at objects, lines, line segments, and more complex shapes.

From this problem statement we derive three research questions (RQs).

RQ 1 How can we estimate the number of objects simultaneously with the fitting of these objects?

RQ 2 How can we optimize inference over both the number of objects and fitting of those objects in the robotic vision domain?

RQ 3 How can we recognize more general 3D objects?

1.4 Research Methodology

The research methodology advocated in the thesis follows the Bayesian methodology (cf. Savage, 1972; Jaynes, 2003). So, our research methodology consists of two phases. In the first phase a Bayesian model is defined. This model exists of (1) a definition of parameters and relations between these parameters, (2) a definition of the noise, and (3) the data. In the second phase, the Bayesian method dictates all remaining unknowns, from the number of parameters to the values of the parameters. To perform Bayesian inference efficiently new methods are required if the model is complex (as is the case with robotic vision).

The Bayesian methodology aims to establish the rationale for practical questions. The following two questions are clear examples.

- If we observe a single point in an image, can we expect it to be part of a line?
- If we have two lines and we live in a world with squares, what are we able to infer?

The two questions tap into our capabilities to define models that makes our prior knowledge explicit. Moreover, if we are able to quickly assign (1) points to segments, (2) segments to lines, and (3) objects to categories, we can enrich it with all corresponding group properties without the need to have them observed for this individual.

In robotic vision we take as an example the task of line detection. Both the Hough transform (Hough, 1962) and the RANSAC method (Bolles and Fischler, 1981) do detect lines, but they do not explicitly take noise into account. We can apply a Bayesian methodology to these tasks if we extend it to perform inference over a variable number of objects. This is called nonparametric Bayesian inference (Ghosal and Van der Vaart, 2017) . The Bayesian inference method is optimal in an information-theoretic sense (Zellner, 1988), Moreover, nonparametric Bayesian models are consistent in the sense that they approach the underlying true distribution (Wasserman, 1998).

Let us write this down informally in a straightforward manner. If we have formulated a problem in the Bayesian sense, there is no better way to solve it than using Bayesian inference. Given a Bayesian model, there is no need to search for another method to infer lines in a line detection task. No variant on Hough or RANSAC will outperform the Bayesian model. If someone would find a method that seems to outperform a Bayesian method it is either (1) because the signal or noise has not been correctly modeled, or (2) because the method overfits with respect to the available data. Moreover, if approximations are used with respect to optimal Bayesian inference (either variational approximations or Markov-Chain Monte Carlo), there are theoretical guarantees on convergence (Andrieu et al., 2003). A Bayesian model is recommended also in those cases, compared to models that do not have such guarantees.

A well known problem with nonparametric Bayesian models is the curse of dimensionality. Compared to maximum likelihood methods or other non-probabilistic methods that do not take noise into account at all, the nonparametric Bayesian models require significant computational resources. Our research methodology first establishes the correct models, even if solving them seems computationally infeasible. Subsequently, our approach is to develop approximations using more efficient samplers while theoretical guarantees on convergence are preserved.

As described before, the Bayesian inference method is optimal in an information-theoretic sense (Zellner, 1988). In this thesis we take the optimality of the Bayesian method as a given. Our research methodology is to perform experiments to study the efficacy of inference methods for the proposed models. We will restrict the scope of the thesis to the (subjective) priors and noise models we propose for particular models. We will not study alternative noise models and priors.

1.5 Main Contribution

Our contribution to robotic vision can be subdivided into three parts that correspond with the three research questions.

The first part addresses the problem of inference about objects from a nonparametric Bayesian perspective. Contemporary methods in robotic vision do not allow for astute statements about their performance. In practice, this means that when using computer vision to detect cells under a microscope, someone cannot be confident about the number of detected cells. An autonomous cleaning robot in a supermarket cannot be confident about the aisle it is driving into. To be able to properly take into account models and uncertainty simultaneously, Bayesian models have found mainstream adoption. State-of-the-art Bayesian methods that reason about the number of objects alongside object models are a recent object of study (cf. Ferguson, 1973; Hjort, 1990; Lijoi and Prünster, 2010; Joho et al., 2011). The thesis applies such nonparametric Bayesian models towards the applications of robotic vision and depth perception. Models such as the infinite line model and the infinite line segment model are introduced.

The second part addresses the problem of high-dimensional data. To efficiently sample more complex geometric structures, new MCMC (Markov-Chain Monte Carlo, Section 2.2.4) methods are required. The thesis introduces such an MCMC sampler, namely a new Split-Merge sampler, and applies it to complex geometric structures.

The third part addresses more complex robotic vision problems, in the form of object recognition of point clouds in 3D. It combines nonparametric Bayesian inference with models from deep learning.

1.6 Organization of the Thesis

Chapter 1 (this chapter) introduces the problem of contemporary methods in computer vision and depth perception. Due to the fact that these methods are not optimal by construction, it is hard to articulate how they perform. The need for a Bayesian methodology is sketched briefly. The problem statement and three research questions are formulated. Moreover, the research methodology is described and the organization of the thesis is outlined.

Chapter 2 describes (1) probability theory using measure theory, (2) random measures known as random processes of which five are described as nonparametric Bayesian models, and (3) six inference methods that infer model parameters of such nonparametric Bayesian models given the data. It is followed by a discussion that indicates which parts will be most useful for chapters 3 and 4.

Chapter 3 examines a first nonparametric Bayesian model, i.e., the infinite line model. The infinite line model represents a countably infinite set of lines. Gibbs sampling is used to perform simultaneous inference over (1) the number of lines and (2) line parameter values such as slope and intercept.

Chapter 4 examines a second nonparametric Bayesian model, i.e., the infinite line segment model. The infinite line segment model represents a countably infinite set of line segments. A split-merge MCMC sampling method is used to perform simultaneous inference over (1) the number of line segments and (2) line segment parameter values such as slope, intercept, and segment size. Chapters 2 to 4 answer the first research question.

Chapter 5 investigates a new MCMC method, the Triadic Split-Merge sampler. It is tailored to clustering problems and accelerates inference of the models in Chapters 3 and 4. This chapter answers the second research question.

Chapter 6 examines more complex objects, like cubes and multiple cubes in a 3D space. It employs deep learning methods, in particular an autoencoder on point clouds, to perform inference on this type of data. This chapter answers the third research question.

Chapter 7 discusses the relevance of the developed models and inference methods. The answers to the research questions are discussed. Then the problem statement is answered and conclusions are formulated. Finally, recommendations are given and future research is envisaged.