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

CHAPTER 7

DISCUSSION AND CONCLUSIONS

Autonomous systems, cars, robots, rely on depth information. Point clouds, points in a 3D space, are data streams that these systems have to encode into objects, infer object type, count, and in the end use to orient themselves in the world and act accordingly.

Bayesian methods have the advantage of - given a particular prior and likelihood - optimally deduct the posterior. In computer vision the Hough transform and other traditional methods have been used many times. Now, with the advent of more modern machines it becomes possible to implement full Bayesian methods. The Bayesian methods deployed do not just infer a particular type of line, but multiple of them. Moreover, the number of objects is not known in advance.

7.1 Research Questions

Below we summarize the answer to the three research questions, (RQs). We use RA1 to indicate the answer to RQ1. Let us start to reiterate the first research question.

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

We have introduced two nonparametric Bayesian models in Chapter 3 and Chapter 4.

RA 1 We can estimate the number of objects simultaneously with fitting the objects for lines as well as segments with nonparametric Bayesian methods. For lines Gibbs sampling is sufficient. For line segments, due to nonconjugacy of prior and likelihood, Gibbs sampling with auxiliary variables has to be used.

This answers the first research question.

The second research question concerns optimization for the domain of robotic vision.

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

The point clouds in the robotic vision domain have spatial properties. One of those properties is that objects can intersect. In Chapter 5 we introduced a new MCMC method, the Triadic Split-Merge sampler.

> **RA 2** Inference method in the robotic vision domain can be improved through the use of a triadic split-merge sampler that takes into account object intersection.

This answers the second research question.

The third research question asks about generalization.

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

Generalization to 3D, as well as considering more complex objects than lines or line segments, requires a great deal of sampling time. By introducing data-driven priors, inference can be accelerated, as shown in Chapter 6.

> **RA 3** More general 3D objects can be recognized by using data-driven priors. These priors can be generated by deep learning methods. The use of complex, data-driven priors allows us to to perform inference over more general 3D objects such as cubes, compared to simple lines or line segments.

This answers the third research question.

7.2 Problem Statement

The research questions support the answer to the general problem statement.

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

The answer to the problem statement is a combination of general approaches, which have been made possible to (1) advances in nonparametric Bayesian models as well as (2) advances in MCMC inference methods and (3) modern data-driven deep learning methods.

Answer: *Robotic vision problems that concern the recognition of multiple objects of which the number is unknown in advance can be modeled by nonparametric Bayesian models. The inference methods - even though there is no conjugacy - can benefit from spatial characteristics and can be further accelerated by using data-driven priors.*

7.3 Limitations

Inference methods for nonparametric Bayesian models can converge quite slowly. The triadic sampler as postulated in this thesis does accelerate inference for a particular type of robotic vision problem, but it takes too much time for more complex data objects such as 3D cubes. Standard reconstruction loss functions in deep learning methods such as autoencoders fail to take into account properties of the robotic vision domain: e.g. translation invariance and the occurrence of multiple objects. This can be mitigated by adjusted loss functions as proposed in this thesis. For robotic vision problems with multiple types of complex 3D objects, the current model is not sufficient. The autoencoder would represent multiple objects simultaneously. For a correct handling of this challenge, a more complex loss function has to be imposed on the autoencoder.

7.4 Recommendations

We recommend an in-depth study of a variety of reconstruction loss functions for autoencoders on datasets that have the spatial properties inherent to the robotic vision domain. The recommendation is to not only address spatial invariance or object copies as studied in this thesis. There are other properties, such as rotation invariance, scale invariance, and object overlap which would require our attention. There are also dynamic properties, such as temporal occlusions, temporal consistency, light conditions, and gravitational constraints, might will need to be captured by priors to perform Bayesian inference. Those priors (1) can be postulated by the domain expert as is regularly done in the Bayesian methodology or they (2) can be obtained in a data-driven manner as is done in the deep learning literature.