Paper Review
DART: Dense Articulated Real-Time Tracking

Alvin Sun

October 27, 2021

1 Paper Summary

This paper proposed a framework, DART, for tracking generic objects in real-time. This work represents objects as a tree of kinematic constraints between rigid bodies with pre-defined geometry, which are locally modeled as SDF (signed distance function). The tracking of the object is done by optimizing for the kinematic states (such as joint angles and camera poses) between those rigid bodies. The tracking algorithm is derived in a Kalman Filter like fashion, with a process update and a measurement correction. Given the model of the object to be tracked as well as some initial guess on the kinematic states, the process model gets to predict what the SDF should look like. Then a measurement SDF, that is computed based on a depth image, is used to make a correction through optimizing a symmetric objective function. The performance of DART is evaluated on a variety of tasks. The results are comparable or even better than many methods that are specialized for tracking particular types of objects. The proposed tracking algorithm is also trivially parallelizable, which means that it can easily be implemented GPU to achieve real-time performance.

2 What I Learned

1. Lie algebra can be really useful when dealing with optimization on 3D transformations and 3D rotations. The locally linear structure has really nice properties for calculus to be applicable in such constrained manifold.

2. The negative information (i.e. empty spaces) in SDFs can also be utilized in the optimization problem for finding the most likely kinematic states.

3 Opinions

3.1 Up Votes

- I really like how they formulate the measurement correction as optimizing for a symmetric function. The objective function can analytically describe how likely a kinematic state is considering both the predicted SDF and the measured SDF, which both contains come level of noises.

- I also like the simple formulation for this tracker, which makes the computation for each pixel independent of each other, and thus can be parallelized easily.

3.2 Down Votes

One thing I think this paper might fall short on is probably using a constant position motion model for the process update. The whole purpose for tracking is when the scene is dynamic, either the object or the camera is moving, or both. A constant velocity motion model will make more sense in the context of dynamic tracking.

4 Evaluations

The purpose of this paper is to obtain a framework that can track generic objects with depth camera sensors in real-time. This is a solid objective as most of the previous work in the tracking land is developed around specialized domains such as human-body, hands, simple geometries etc. Few has proposed framework that is capable of tracking generic objects. DART does deliver the goal well that it achieves comparable state-of-the-art performance in previously specialized domains, which makes the algorithm much more appealing to be extendable to more diverse settings.

The overall quality of this paper is exceptional in my opinion, as their methods are really generalizable to all kinds of settings. However, one important assumption for this method to work is that the rigid parts of the objects have known 3D geometry. For objects such as hand and human body that has relatively known scale, this could be easy to obtain. However, for other commonly seen objects in real-life, as simple as cylindrical ones like mugs and cups, we will need to get a model per objects due to the scale difference. This could be quite cumbersome if we want to track multiple objects or unknown objects specified in the image space. So this assumption could be the biggest limiting factor for this approach to become a anyone-can-use tracker for all settings. Other then that, one assumption made in the derivation is the independence between all pixels. This is unrealistic and makes the whole theory a little less sound then the rest of the derivation. Nonetheless, DART is a great paper that delivers the objective well by achieving state-of-the-art performance across multiple domains.

5 Questions

1. How is inverse Hessian (line 4 of the algorithm) related to the covariance of the uncertainty?

2. Does it method still work if the object model is not scaled correctly? (i.e. slightly smaller or bigger than the actual object).