Tracking Large Scale Articulated Models with Belief Propagation for Task Informed Grasping and Manipulation

Karthik Desingh[1], Jana Pavlasek[2], Cigdem Kokenoz[1], Odest Chadwicke Jenkins[1,2]

I. Introduction

In order to enable robots to perform tasks in indoor environments such as the kitchen, we require accurate pose estimation and tracking of the full scene. Such a scene is a large scale articulated model composed of multiple objects with articulations, including cabinets, drawers and appliances. To perform task informed grasping and manipulation, the full 6D pose estimate of the objects in the scene is required. The availability of depth sensors provides sensing capability in the 3D domain and motivates the extension of vision based pose estimation approaches [1]. This domain presents many technical challenges, including occlusions, sensor noise, and high computational complexity due to the high dimensional continuous pose space.

Probabilistic modeling has been widely applied to object tracking. Wuthrich et. al. [2] propose a probabilistic technique for tracking of objects being manipulated by a human or robot with known geometries using a particle filter. The particle filter models occlusions alongside the observation and process models. The framework was extended to track a manipulator end-effector [3]. In [4], Schmidt et al. introduce a general framework for tracking articulated objects with known structure using an extended Kalman filter, where the observation model employs the signed distance function. It was extended to include physics based constraints on the objects [5]. These tracking frameworks are either initialized to objects' ground truth poses or informed by joint encoder readings in the case of articulated objects. Here, we aim to develop a unified framework that performs pose estimation followed by a pose tracking stage without any initialization and using only point cloud data. Full scene pose estimation and tracking of known objects has been studied in the context of SLAM by Salas-Moreno et al. [6]. However, this work assumes objects to be static while the camera is in motion. Here, we aim to be able to work with a mobile manipulation platform where the change in the observation can be due to articulated objects as well as to the robot motion.

Nonparametric belief propagation has been effectively used in applications such as human pose estimation [1] and hand tracking [7] by modeling the graph as a particle network. In order to viably pursue NBP for robotic problems, such as scene perception, the computational efficiency of NBP methods needs to be revisited. In this regard, our previously devised Pull Message Passing Nonparametric Belief Propagation (PMPNBP) algorithm [8] is computationally efficient and has been shown to satisfy the computational needs of scene perception problems. In [8] we show promising results in pose estimation of articulated objects under severe occlusions. PMPNBP takes as input a geometrical model with articulation constraints and a 3D point cloud observation and outputs belief over the object-part poses iteratively. By factoring the state of the articulated object into its individual parts, PMPNBP is able to avoid local minima as compared to standard particle filter based approaches.

In this work, we propose to extend our previous work [8] to fully estimate and track large scale articulated models. We believe our factored approach is more suitable for large scale scenes that are partially observed due to occlusions by other objects and agents, and due to the limited field of view of an on-board depth sensor.

The proposed extension will have two stages: a pose estimation stage (global localization) followed by a pose tracking stage (local localization). The problem is formulated as a Markov Random Field (MRF) similar to the PMPNBP method, where unary and pairwise potential functions help to iteratively pass informative messages between the hidden nodes to infer the state that most likely explains a given observation. The unary potential of the graphical model in PMPNBP models how well a pose explains the observation. In the proposed extension, this unary potential function should cater to the needs of the large scale observations with varying sensor noise and occlusions, and accommodate partial and noisy observations. The pairwise potential of PMPNBP models how compatible a pair of rigid body poses are, given their articulation constraints. In the proposed extension, modeling this function specific to pose estimation and pose tracking stages would benefit both the inference and its computational needs. In this ongoing work, we plan to devise potential functions suitable for the inference and its computational needs.

II. Tracking with Belief Propagation

Given a large scale articulated model (kitchen model) O, its geometry and Unified Robot Description Format (URDF) defining its articulation, we wish to estimate the 6 DoF object pose X_s of each of its rigid parts. In the PMPNBP formulation, each articulated object is represented as a Markov Random Field (MRF) $G = (V, E)$, where the graph G has nodes V and edges E. The graph is made up of hidden variables X representing the object poses and observed variables Y representing sensor observations. The pose estimate is obtained through inference, where X_i

[1]Department of Computer Science and Engineering, University of Michigan, Ann Arbor, USA
[2]Robotics Institute, University of Michigan, Ann Arbor, USA
{kdesingh,pavlasek,ckokenoz,ocj}@umich.edu

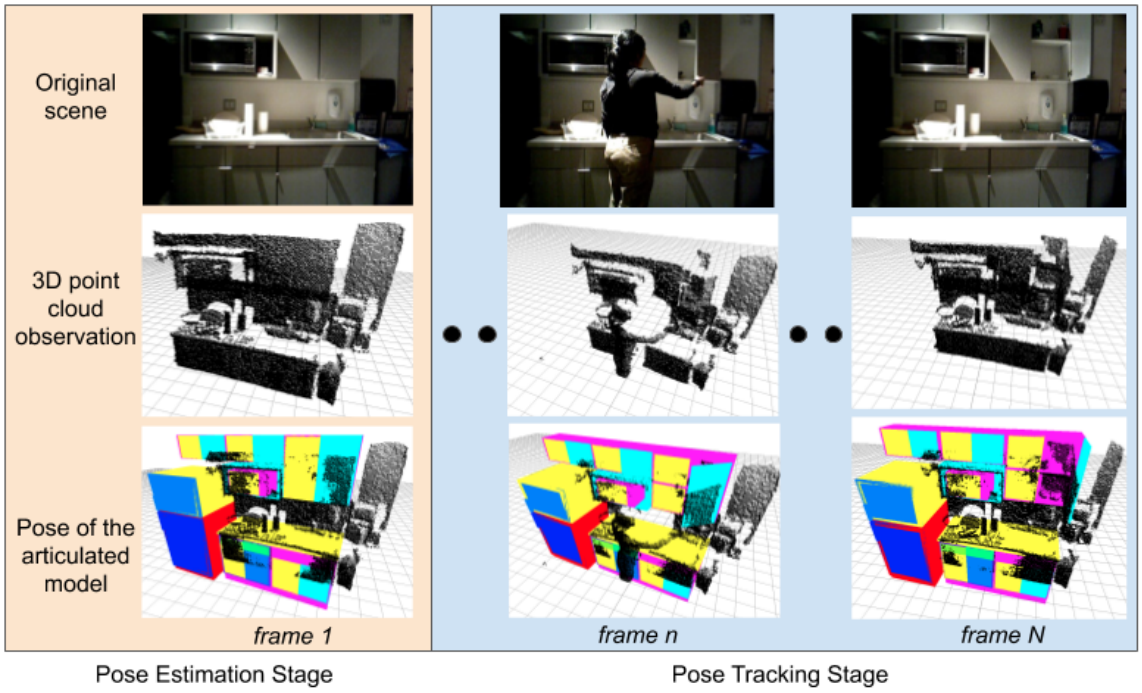

Fig. 1: Illustration of the expected behavior of the pose estimation and tracking framework for a kitchen setting. The pose estimation stage localizes the kitchen model in the 3D point cloud data. The pose tracking stage tracks the changes in the pose under occlusion (human opening a cabinet) over a stream of 3D point cloud data.

maximizes the joint probability of the observed and hidden variables:

$$p(X,Y) = \frac{1}{Z} \prod_{(i,j)\in E} \psi_{i,j}(X_i, X_j) \prod_{s\in V} \phi_i(X_i, Y_i) \qquad (1)$$

where $\psi_{i,j}(X_i, X_j)$ is the pairwise potential between nodes X_i and X_j, $\phi_i(X_i, Y_i)$ is the unary potential between the hidden node X_i and the observed node Y_i, and Z is a normalizing factor. PMPNBP performs inference through pull message passing in the graph to obtain the estimated pose $\hat{X}_i$ of each part. Details on this factorization, the PMPNBP algorithm and its benefits can be found in [8]. Motivated by [1], we develop a tracking stage as an extension to the pose estimation stage, with changing observations at every iteration.

Figure 1, illustrates the layout of this framework with a kitchen scene whose 3D geometry and articulation model are known. Other work [9] has explored learning articulation models through interactive perception, however our work assumes the model is given in order to focus on the challenges of pose estimation and tracking which arise once the model is obtained. The main challenge of both objectives is partial observation. This partial observation is mainly due to two factors: occlusion due to objects or agents that are not part of the kitchen model and the limited viewing angle of the robot's on-board sensor. In other words, some objects in the model are not present in the observations, and vice versa. This ambiguity in observation demands the notion of maintenance of distribution over possible hypotheses. Our prior work [8] shows that belief propagation is suitable for maintaining and propagating belief for a single observation. In this work, we propose to extend our efficient belief propagation method toward tracking large scale articulated models over continuous observations.

References

[1] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard, "Tracking loose-limbed people," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004, pp. 421–428.

[2] M. Wuthrich, P. Pastor, M. Kalakrishnan, J. Bohg, and S. Schaal, "Probabilistic object tracking using a range camera," in *IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 3195–3202.

[3] C. G. Cifuentes, J. Issac, M. Wüthrich, S. Schaal, and J. Bohg, "Probabilistic articulated real-time tracking for robot manipulation," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 577–584, 2016.

[4] T. Schmidt, R. A. Newcombe, and D. Fox, "DART: dense articulated real-time tracking," in *Robotics: Science and Systems X, University of California, Berkeley, USA, July 12-16, 2014*, 2014.

[5] T. Schmidt, K. Hertkorn, R. Newcombe, Z. Marton, M. Suppa, and D. Fox, "Depth-based tracking with physical constraints for robot manipulation," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 119–126.

[6] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. J. Kelly, and A. J. Davison, "SLAM++: Simultaneous localisation and mapping at the level of objects," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1352–1359.

[7] E. B. Sudderth, M. I. Mandel, W. T. Freeman, and A. S. Willsky, "Visual hand tracking using nonparametric belief propagation," in *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, 2004, pp. 189–189.

[8] K. Desingh, S. Lu, A. Opipari, and O. C. Jenkins, "Efficient nonparametric belief propagation for pose estimation and manipulation of articulated objects," *Science Robotics*, vol. 4, no. 30, 2019.

[9] R. M. Martin and O. Brock, "Online interactive perception of articulated objects with multi-level recursive estimation based on task-specific priors," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2494–2501.