

Integrated Online Perception of Articulated Objects for Manipulation

Jia Guo

George Kontoudis

Abstract—An online interactive perception methodology for articulated objects in unstructured environment is presented. The main contribution of this methodology lies in the online solution, which utilize recursive Bayesian estimation techniques. The RGB-D algorithm consist of three sub-recursive estimation problems and each one forms a separate level of estimation. In this way, we can get uncomplicated solution at each level which feed forward and backward information to guarantee robustness, accuracy, and uncertainty elimination. The efficacy of the proposed method is verified through robot manipulation experiments.

Index Terms: Online perception, recursive Bayesian estimation, manipulation

I. INTRODUCTION

An online mutli-level interactive perception algorithm is presented [8], [9]. Grasping and manipulation in unstructured environments require knowledge of the object's shape, position, orientation, and kinematic structure. Visual perception could be a solution to successfully get the information needed for robotic manipulation. Although, it becomes non-trivial to online address this problem and estimate robust solutions. However, the allocation to sub-level algorithms which are interconnected, simplify the overall solution.

In this project we deal with online interactive perception algorithm for robotic manipulation purposes. The algorithm includes the identification of the object's shape, the recognition of its kinematic structure, and the motion tracking of its position and orientation. Moreover, the derivation of the explored object's shape, position, orientation, along with the kinematic structure is achieved. The object's position, orientation, and estimation of kinematic structure part incorporates three recursive Bayesian estimation steps. Then, the shape reconstruction of the investigated object is addressed that thrust the motion tracking.

Robotic manipulation is an emerging field for many decades. [1], [12], [4], [13], [14], [5], [6], [15], [2], [10], [16], [3], [11].

The remainder of this report is organized as follows. Section II discusses the problem formulation. The efficacy of the proposed methodology is assessed in Section III by a set of experiments. Section IV discusses the results of this project. Finally, Section V provides the conclusions of the studied methodology and gives directions for future work.

J. Guo and G. Kontoudis are with the Mechanical Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060, USA.

Names were alphabetically ordered.

II. PROBLEM FORMAUATION

In this section we formulate the proposed methodology for integrated online perception. First, the interactive perception problem tracks the motion, and estimates the kinematic structure of the object. Then, this information is provided to the tracker which performs feautre-based tracking, and shape based tracking along with information given from shape-based reconstruction module. The motion segmentation in color and in depth follows to identify fault information from the environment. Next, the object's reconstruction takes place by separating the unreliable information from the picture. Finally, shape-based segmentation is accomplished to feed the shape-based tracking. In figure 1 the kinematic structure of the methodology is depicted.

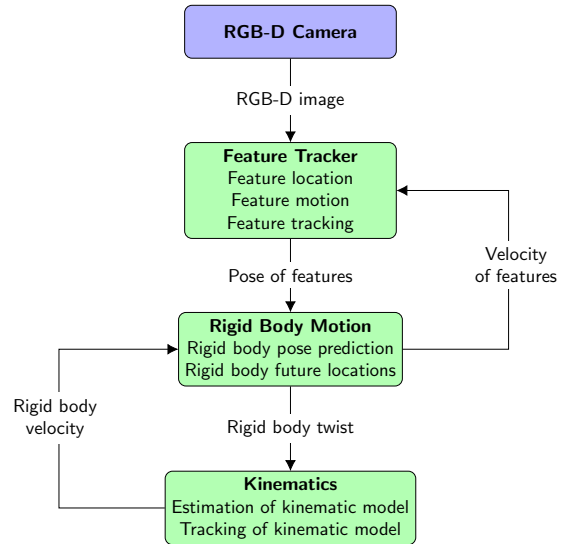


Fig. 1. Kinematic identification scheme.

A. Feature Motion Estimation

The gathered information from an RGB-D camera employed to track feature motion with recursive estimation. The recursive estimation seeks to update the belief of target $p(x_k^f | z_{1:k}^f)$. In this case the targets are features of the object in the 3D space, $x_k^f \in \mathbb{R}^{3m}$, where $k \in \mathbb{N}$ is the time index, f is the feature motion, and $m \in \mathbb{N}$ is the number of features we track. Then we get observations from the camera in the 2-D image plane, $z_k^f \in \mathbb{R}^{2m}$, where $k \in \mathbb{N}$ is the time index, f is the feature motion, and $m \in \mathbb{N}$ is the number of features we track.

The recursive estimation is based on priors of Kanade-Lucas-Tomasi (KLT) tracking algorithm [14], [7] the small motion, and the brightness constancy constraint. Small motion means that the feature points do not move far away from their previous position. Brightness constancy means that the projection of the same feature is the same at each frame. First, a feature initialization takes place from the image, based on Kanade-Tomasi corner detection. Then, a prediction of 3-D feature's location is occurred from its current location and the velocities obtained from the higher stage. Next, this prediction is employed to project this features in the image plane and use them as input to KLT algorithm.

B. Extended Kalman Filter of Rigid Bodies

In the second stage an extended Kalman filter (EKF) is utilized to estimate the motion of rigid bodies. The information needed is collected from the lower stage feature tracker and from the upper stage kinematic model estimator. The methodology deals with EKF, because it is based on non-linear Gaussian models. The stochastic motion and sensor models have the form of

$$x_k^t = f^t(x_{k-1}^t, u_{k-1}^t) + w_{k-1}^t, \quad (1)$$

$$z_{k-1}^t = h^t(x_{k-1}^t) + v_{k-1}^t, \quad (2)$$

where $k \in \mathbb{N}$ is the time index, x_k^t is the vector of states, z_{k-1}^t is the vector of observations, w_{k-1} and v_{k-1} are the zero-mean independent identically distributed Gaussian noises $w_{k-1}^t \sim \mathcal{N}(0, \Sigma_{w_{k-1}^t}) = \mathcal{N}(0, Q_{k-1})$, with $Q_{k-1} \geq 0$, and $v_{k-1}^t \sim \mathcal{N}(0, \Sigma_{v_{k-1}^t}) = \mathcal{N}(0, R_{k-1})$ with $R_{k-1} > 0$. To estimate the state, the EKF first predicts the state and the covariance as

$$\bar{x}_{k|k-1}^t = f^t(x_{k-1|k-1}^t, u_{k-1}^t), \quad (3)$$

$$\Sigma_{\bar{x}_{k|k-1}^t} = A_{k-1} \Sigma_{x_{k-1|k-1}^t} A_{k-1}^T + Q_{k-1}, \quad (4)$$

where

$$A_{k-1} = \frac{\partial f^t(x_{k-1|k-1}^t, u_{k-1}^t)}{\partial x_{k-1|k-1}^t}. \quad (5)$$

Then the Kalman gain is calculated from

$$K_k = \Sigma_{\bar{x}_{k|k-1}^t} C_k^T (C_k \Sigma_{\bar{x}_{k|k-1}^t} C_k^T + R_k)^{-1}, \quad (6)$$

where

$$C_k = \frac{\partial h^t(x_{k|k-1}^t)}{\partial x_{k|k-1}^t}. \quad (7)$$

Next, the correction is given as

$$\bar{x}_{k|k}^t = \bar{x}_{k|k-1}^t + K_k(z_k^t - h^t(\bar{x}_{k|k-1}^t)), \quad (8)$$

$$\Sigma_{\bar{x}_{k|k}^t} = (I - K_k C_k) \Sigma_{\bar{x}_{k|k-1}^t}. \quad (9)$$

More specifically the correct features from the lower stage are defined as rigid bodies. Every rigid body has its own EKF and the state composed from position and orientation, and twist, $x_k^t = [p^T \ v^T]^T$, $x_k^t \in \mathbb{R}^{12}$. The observations contain the location of each rigid body in the 3-D space, $z_k^t \in \mathbb{R}^{3m}$. Three simultaneous actions were employed to predict the state of rigid bodies. One action predicts the next state of

rigid bodies by employing an EKF as previously described. The other action examines the sudden immobilization of rigid bodies, while the last action employs data from the upper kinematic stage to predict an alternative state of rigid bodies. The observation part includes feature locations in 3-D space f_0^m and the homogeneous transformation using the position and orientation of the rigid bodies $T(p)$. The exact probabilistic sensor model of 2 becomes

$$z_k^t = h^t(x_{k-1}^t) + v_{k-1}^t = \begin{bmatrix} T(p)f_0^1 \\ \vdots \\ T(p)f_0^m \end{bmatrix} + v_{k-1}^t, \quad (10)$$

To effectively estimate the motion of a group with multiple rigid bodies the generation of valid sets is needed. After defining these sets the next step is to estimate their motion in 3-D space. For this purpose we employ EKF to predict the group motion, but in case that the error exceeds $2cm$, random sample consensus (RANSAC) is utilized. If the group consists of at least 15 features we consider it as a valid set.

C. Extended Kalman Filter of Kinematic Model

An estimation and tracking of the kinematic model is achieved in the third stage. The information needed is gathered from the lower stage rigid body estimator. The kinematic model can be related either as a prismatic joint, or as a revolute joint, or as a rigid connection, or unrelated. Each relation is represented by a different state x_{joint}^t and the observations are collected from the previous stage as a set of multiple rigid bodies $z_{joint}^t \in \mathbb{R}^6$, along with the zero-mean independent identical distributed Gaussian noise v_{joint}^t . Since, the type of joint varies the utilization of different observation model was imposed to apply EKF.

For the prismatic joint estimation the state includes the axis orientation, the joint displacement $q_p \in \mathbb{R}$, and the joint velocity. Joint velocities update by employing the previous stage twist vector. For the position and orientation prediction of the rigid bodies relation, the following observation model was utilized

$$z_{pr,joint}^t = \begin{bmatrix} q_p \hat{o}_p \\ 0_3 \end{bmatrix} + v_{k-1}^t, \quad (11)$$

where $\hat{o}_p \in \mathbb{R}^3$ is the orientation of the axis.

For the revolute joint estimation the state includes the axis orientation, the joint revolution $q_r \in \mathbb{R}$, and the joint velocity. Joint velocities update by employing the previous stage twist vector. For the position and orientation prediction of the rigid bodies relation, the following observation model was utilized

$$z_{rev,joint}^t = \begin{bmatrix} (-q_r \hat{o}_r) \times p_r \\ q_r \hat{o}_r \end{bmatrix} + v_{k-1}^t, \quad (12)$$

where $\hat{o}_p \in \mathbb{R}^3$ is the orientation of the axis, and $p_r \in \mathbb{R}^3$ is a point on the axis of rotation.

The observation for rigid body estimation has no information about rigid body relations $z_{rig,joint}^t = 0_6 + v_{k-1}^t$. In case none of the above stands then the algorithm defines unrelated rigid bodies.

III. EXPERIMENTS

IV. RESULTS

V. CONCLUSIONS

REFERENCES

- [1] L. Chang, J. Smith, and D. Fox, "Interactive singulation of objects from a pile," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3875–3882.
- [2] C. Choi and H. I. Christensen, "Rgb-d object tracking: A particle filter approach on gpu," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 1084–1091.
- [3] E. Herbst, P. Henry, and D. Fox, "Toward online 3-d object segmentation and mapping," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3193–3200.
- [4] D. Katz, A. Orthey, and O. Brock, "Interactive perception of articulated objects," in *Experimental Robotics*. Springer, 2014, pp. 301–315.
- [5] D. Katz, Y. Pyuro, and O. Brock, "Learning to manipulate articulated objects in unstructured environments using a grounded relational representation," *Robotics: Science and Systems IV*, p. 254, 2009.
- [6] M. Krainin, P. Henry, X. Ren, and D. Fox, "Manipulator and object tracking for in-hand 3d object modeling," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1311–1327, 2011.
- [7] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," 1981.
- [8] R. M. Martin and O. Brock, "Online interactive perception of articulated objects with multi-level recursive estimation based on task-specific priors," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 2494–2501.
- [9] R. Martín-Martín, S. Höfer, and O. Brock, "An integrated approach to visual perception of articulated objects," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 5091–5097.
- [10] A. Mishra, Y. Aloimonos, and C. Fermuller, "Active segmentation for robotics," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 3133–3139.
- [11] F. Pomerleau, S. Magnenat, F. Colas, M. Liu, and R. Siegwart, "Tracking a depth camera: Parameter exploration for fast icp," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 3824–3829.
- [12] J. Sturm, K. Konolige, C. Stachniss, and W. Burgard, "Vision-based detection for learning articulation models of cabinet doors and drawers in household environments," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 362–368.
- [13] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.
- [14] C. Tomasi and T. Kanade, "Detection and tracking of point features," 1991.
- [15] M. Wüthrich, P. Pastor, M. Kalakrishnan, J. Bohg, and S. Schaal, "Probabilistic object tracking using a range camera," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 3195–3202.
- [16] C. Yuheng Ren, V. Prisacariu, D. Murray, and I. Reid, "Star3d: Simultaneous tracking and reconstruction of 3d objects using rgb-d data," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1561–1568.