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**

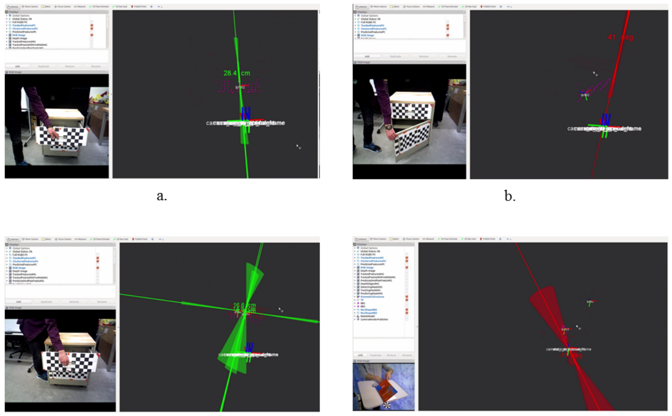
I. Introduction

An online mutli-level interactive perception algorithm is presented [9], [10]. 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.

Literature review - Jia [19], [11], [3], [12], [2], [7], [15], [11], [9], [14], [16], [17], [4], [6], [5].

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. Finally, Section IV provides the conclusions of the studied methodology and gives directions for future work.



a.

b.

c.

d.

Fig. 1. Integrated online perception applications, a. Drawer opening using RGB-D stream, b. Door opening using RGB-D stream, c. Combination of movements using RGB-D stream, d. Book opening using bag file.

II. Problem Formaulation

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 2 the kinematic structure of the methodology is depicted.

The basic process of the integrated perception method is shown in figure 3. The foundation of the method is the multi-level kinematic estimation method. The shape-based tracker is introduced to refine the results of feature-based tracker. As stated in the related work, the visual pose estimation technique is approached in two ways, which are based on shape model and texture feature respectively. Here these two methods are integrated with the process of shape reconstruction.

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 $p(x^t_k|z^t_{1:k})$. In this case the targets are features of the object in the $3D$ space, $x^f_k \in \mathbb{R}^{3m}$, where $k \in \mathbb{N}$ is the time index, f

This work was submitted on 07/04/17 and is part of the project for the class ME5524-Bayesian Robotics, offered by Prof. Tomonari Furukawa, tomonari@vt.edu

J. Guo and G. Kontoudis are with the Mechanical Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060, USA, jguo18@vt.edu, gpkont@vt.edu

Names were alphabetically ordered.

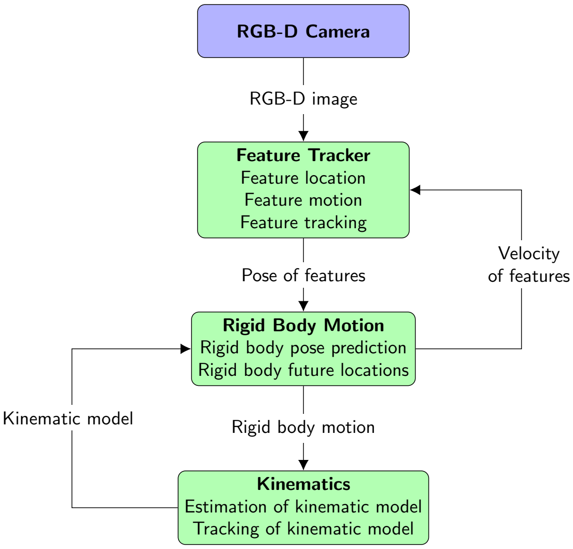


Fig. 2. Kinematic identification scheme.

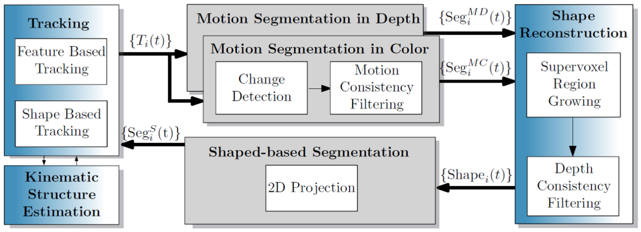

Tracking
Feature Based Tracking
Shape Based Tracking
Kinematic Structure Estimation
$\{T_i(t)\}$
Motion Segmentation in Depth
Motion Segmentation in Color
Change Detection
Motion Consistency Filtering
$\{\mathrm{Seg}_i^{MD}(t)\}$
$\{\mathrm{Seg}_i^{MC}(t)\}$
Shape Reconstruction
Supervoxel Region Growing
Depth Consistency Filtering
$\{\mathrm{Seg}_i^{S}(\mathrm{t})\}$
Shaped-based Segmentation
2D Projection
$\{\mathrm{Shape}_i(t)\}$

Fig. 3. Flow chart of integrated perception method [10].

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 [18], [8] 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 form 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

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

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

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

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

$$\Sigma_{x_{k|k-1}^t} = A_{k-1} \Sigma_{x_{k-1|k-1}^t} A_{k-1}^\intercal + Q_{k-1}, \tag{4}$$

where

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

Then the Kalman gain is calculated from

$$K_k = \Sigma_{x_{k|k-1}^t} C_k^\intercal (C_k \Sigma_{x_{k|k-1}^t} C_k^\intercal + R_k)^{-1}, \tag{6}$$

where

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

Next, the correction is given as

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

$$\Sigma_{x_{k|k}^t} = (I - K_k C_k) \Sigma_{x_{k|k-1}^t}. \tag{9}$$

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, $x_k^t = [p^\intercal \;\; \nu^\intercal]^\intercal$, $x_k^t \in \mathbb{R}^{12}$. 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

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

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^t_{joint} and the observations are collected form the previous stage as a set of multiple rigid bodies $z^t_{joint} \in \mathbb{R}^6$, along with the zero-mean independent identical distributed Gaussian noise v^t_{joint}. 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

$$z^t_{pr,joint} = \begin{bmatrix} q_p \hat{o}_p \\ 0_3 \end{bmatrix} + v^t_{k-1}, \tag{11}$$

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

$$z^t_{rev,joint} = \begin{bmatrix} (-q_r \hat{o}_r) \times p_r \\ q_r \hat{o}_r \end{bmatrix} + v^t_{k-1}, \tag{12}$$

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 boy estimation has no information about rigid body relations $z^t_{rig,joint} = 0_6 + v^t_{k-1}$. In case none of the above stands then the algorithm defines unrelated rigid bodies.

D. Feature-Based Tracker

Jia

E. Motion Segmentation

Jia

F. Shape Reconstruction and Shape-Based Segmentation

Jia

III. Experiments and Results

In this section we perform experiments to validate the efficacy of this project. We intended to implement experiments on grasping, but eventually the robot arm faced some malfunctions and we could not operate it. So we worked on identifying the problem and finding a solution for the robot arm. The first part of the experiments test with bag files the functionality of the ROS [13] package OMIP, that was developed from the authors of [9], [10]. More information about OMIP can be found on the following link.

`http://wiki.ros.org/omip`

In the next part we employ a RGB-D camera, ASUS Xiton Pro live [1] to successfully identify the kinematic structure of a drawer and a door. A video of the conducted experiments can be found in the following link.

`https://youtu.be/weG94fqyQpY`

A. Online Perception Algorithm

Bag files[1], provided by the authors of the OMIP package, utilized to identify the kinematic structure of objects. The whole procedure in ROS is depicted in figure 4. This graph consists of the camera, the feature tracker, the rigid body tracker, the joint tracker, and the camera base link. First, the camera base link nodes publish messages on the topic of transformations its initial configuration. These transformation messages are subscribed by the camera manager node and the rigid body tracker node. Then, the camera manager node publishes messages to specific topics that can be efficiently subscribed by the feature tracker. This part is the RGB-D image streaming of the camera as described in section II. Next, the feature tracker node publishes messages to the state topic which are subscribed from the rigid body node, while it subscribes messages from the predicted measurement topic which are published from the rigid body tracker node. This step was described previously as the feature tracker level, where pose feature information fed the next level of rigid body motion and velocity of features were received as feedback simultaneously. Rigid body tracker node publishes messages to the state topic which are subscribed from the joint tracker node, while it subscribes messages from the predicted measurement topic that are subscribed from the joint tracker node. This sector works similar with the rigid body motion level where the next kinematic level receives information regarding the rigid body motion and provides feedback for the kinematic level. The graph architecture is similar with the kinematic identification scheme in figure 2.

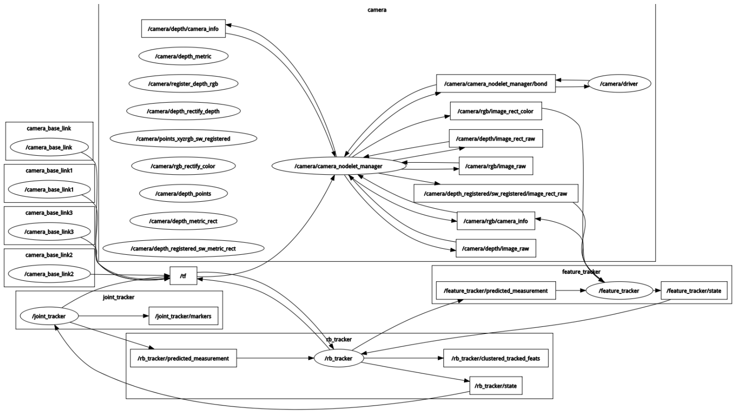

camera
/camera/depth/camera_info
/camera/depth_metric
/camera/register_depth_rgb
/camera/depth_rectify_depth
/camera/points_xyzrgb_sw_registered
/camera/rgb_rectify_color
/camera/depth_points
/camera/depth_metric_rect
/camera/depth_registered_sw_metric_rect
/camera/camera_nodelet_manager
/camera/camera_nodelet_manager/bond
/camera/driver
/camera/rgb/image_rect_color
/camera/depth/image_rect_raw
/camera/rgb/image_raw
/camera/depth_registered/sw_registered/image_rect_raw
/camera/rgb/camera_info
/camera/depth/image_raw
camera_base_link
/camera_base_link
camera_base_link1
/camera_base_link1
camera_base_link3
/camera_base_link3
camera_base_link2
/camera_base_link2
/tf
joint_tracker
/joint_tracker
/joint_tracker/markers
feature_tracker
/feature_tracker/predicted_measurement
/feature_tracker
/feature_tracker/state
rb_tracker
/rb_tracker/predicted_measurement
/rb_tracker
/rb_tracker/clustered_tracked_feats
/rb_tracker/state

Fig. 4. Online perception algorithm (OMIP) in ROS using rqt_graph.

B. Experiment Configuration and Results

Jia

[1] A bag in ROS is a predefined ROS message data format file.

1) Book Demo File Test: Jia

2) Drawer Identification Test: Jia

3) Door Identification Test: Jia

4) Mixed Motion Identification Test: Jia

C. Grasping Application

We intended to implement a third set of grasping experiments with a robot arm. Although, we had already became familiar with the robot arm platform in simulations we could not work on it, because it was malfunctioned. More specifically, the problem was that the robot left for more than 3 months on the operating mode and as a result the batteries were completely discharged. Therefore, we identify the problem and we decided to continue working on implementing the online perception technique with an online RGB-D stream from a camera.

IV. Conclusions

In this project we first study an online perception methodology for kinematic structure identification of articulated objects using a single RGB-D camera. The proposed technique included three sub-level recursive estimation models that made the algorithm efficient enough to operate online. Extended Kalman filter was used as a recursive Bayesian estimation technique because the motion and sensors models were non-linear. Next, a shape-based tracker employed to refine the outcomes of the feature based tracker. Another recursive, estimation technique that utilized was Kanade-Lucas-Tomasi algorithm for the initial tracking part. Robot arm malfunction and time restriction could not let us work on the implementation of the online perception algorithm with grasping experiments. As a result we implement the perception with bag files and we conducted experiments with an online RGB-D camera. All the experiments resulted the successful identification of the kinematic structure for both the prismatic and the revolute joints. The online perception technique efficiently recognize the kinematic structure of some articulated objects, but needs further development to make it compatible with our robotic systems. The efficacy of the perception method depends mostly on the feature tracking, so we needed to add AR tags even with the shape reconstruction extension.

As a future work we keen on continuing the software development in order to make the online perception technique compatible with our robots. Moreover, we suggest the utilization of contemporary tracking algorithm for the feature tracking part. Unscented Kalman filter might be an alternative solution for some objects because of its ability to handle heavily non-linear models. Implementation of this combinatorial system might be useful for future research purposes in grasping and manipulation.

References

[1] ASUS, "Rgb and depth sensor," 2017. [Online]. Available: https://www.asus.com/us/3D-Sensor/Xtion_PRO_LIVE/

[2] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 577–586, 2002.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.

[4] X. Huang, I. Walker, and S. Birchfield, "Occlusion-aware reconstruction and manipulation of 3d articulated objects," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1365–1371.

[5] D. Katz, M. Kazemi, J. A. Bagnell, and A. Stentz, "Interactive segmentation, tracking, and kinematic modeling of unknown 3d articulated objects," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 5003–5010.

[6] D. Katz, A. Orthey, and O. Brock, "Interactive perception of articulated objects," in *Experimental Robotics*. Springer, 2014, pp. 301–315.

[7] J. Kenney, T. Buckley, and O. Brock, "Interactive segmentation for manipulation in unstructured environments," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 1377–1382.

[8] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," 1981.

[9] 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.

[10] 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.

[11] P. Ochs, J. Malik, and T. Brox, "Segmentation of moving objects by long term video analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 6, pp. 1187–1200, 2014.

[12] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation-supervoxels for point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2027–2034.

[13] ROS, "Open-source, meta-operating system for robots," 2017. [Online]. Available: http://www.ros.org/

[14] D. Ross, D. Tarlow, and R. Zemel, "Unsupervised learning of skeletons from motion," *Computer Vision–ECCV 2008*, pp. 560–573, 2008.

[15] J. Stückler and S. Behnke, "Efficient dense 3d rigid-body motion segmentation in rgb-d video." in *BMVC*, 2013.

[16] J. Sturm, V. Pradeep, C. Stachniss, C. Plagemann, K. Konolige, and W. Burgard, "Learning kinematic models for articulated objects," 2009.

[17] 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.

[18] C. Tomasi and T. Kanade, "Detection and tracking of point features," 1991.

[19] 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.