Object Tracking using Particle Filter in Joint Color-spatial Space

Shile Li[1], Seongyong Koo[2] and Dongheui Lee[1]

I. INTRODUCTION

With point-cloud from an RGB-D camera, a robot can distinguish different unknown objects when the objects are spatially separated by clustering in the Euclidean space. As a human moves the objects and contact them with each other, the robot has difficulty to further distinguish each object. The solution is to keep tracking the objects, thus contacted objects can still be distinguished. To track the 6 DOF object pose accurately, often recursive Bayesian Filtering is used. Especially particle filter is suitable, because it can handle non-linear motions induced by human and non-Gaussian noise. In a particle filter framework, a large amount of hypotheses need to be evaluated. The hypothesis evaluation function should reveal how close the hypothesis to the observation is, such that the pose estimate can be integrated from multiple hypotheses. The hypothesis evaluation function should be also computationally efficient due to the real-time requirement. In this work, to improve the efficiency and the accuracy of the hypothesis evaluation, we propose Joint Color-Spatial Descriptor (JCSD) that represents a probability density of a measurement point in the joint color-spatial space.

II. HYPOTHESIS EVALUATION

How to represent an object from a point-cloud such that it is discriminative for false hypothesis and robust to noise? First, we select m evaluation points that are equally distributed in the interested point-cloud space, which are the intersection points of a regular grid. Each evaluation point represents local part of an object by a point density calculated from kernel density estimation. Using a limited kernel bandwidth, the point density at each evaluation point can be efficiently calculated by accumulating each point to the nearby evaluation points. To combine color information, the Smoothed Color Ranging (SCR) technique [2] is used. SCR assigns weights to each point in the 8 color ranges (red, yellow, green, cyan, blue, purple, light gray, dark gray) with the weights $\{h_i\}_{i=1}^{8}$ and $\sum_{i=1}^{8} h_i = 1$. With the color weights, point density for each evaluation point will be calculated eight times for each color range individually. In total, JCSD consists of $8m$ point density values for m evaluation points and 8 color ranges.

For evaluation of a specific pose hypothesis, the overall hypothesis likelihood is summed up by each observation point's likelihood. For evaluation of one observation point $\mathbf{z}$'s likelihood, $\mathbf{z}$ is transformed to object coordinate based on the pose hypothesis. The true point densities on the position of the transformed observation point is estimated by interpolating nearby evaluation points. Interpolated densities are then correlated with the observation point's actual color weights, the correlation value reveals how well $\mathbf{z}$ fits the object model based on the specific hypothesis. With the joint color-spatial presentation, the influence of occluding objects is reduced, because occluding points with different color to object model have zero effect on hypothesis evaluation, as the correlation of color weights between them is zero.

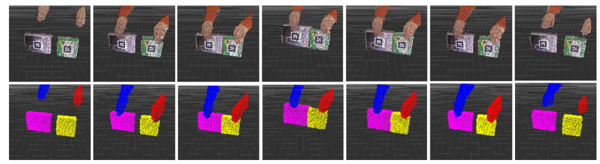

Fig. 1. Object segmentation result from our tracking method. Upper row shows the observed point-cloud from an RGB-D camera, lower row shows the segmentation result.

III. RESULTS

The proposed method is implemented with GPU programming and tested with real-scene data. The pose tracking accuracy is hard to evaluate, because model-free tracking was performed, therefore instead of evaluating the 6 DOF directly, we evaluated tracking accuracy with fitting distance. Fitting distance specifies the average distance between tracked object model points to the observation points. In Fig. 2 the better performance of our method compared to [1] is shown. The mean tracking time for one object using 100 particles takes ca. 4 ms, while [1] needs ca. 40 ms. For tracking 4 objects (Fig. 1) including online segmentation and model update, the computation time takes ca. 52 ms, which ensures the real-time capability.

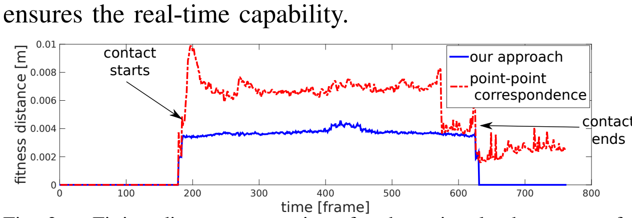

Fig. 2. Fitting distance comparison for the point-cloud sequence from Fig. 1. Blue line is from our approach, red line is from point-point correspondence approach [1]. As soon as partial occlusion occurs, our approach outperforms.

REFERENCES

[1] Changhyun Choi and Henrik I Christensen. Rgb-d object tracking: A particle filter approach on gpu. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1084–1091.

[2] Wei Wang, Lili Chen, Dongming Chen, Shile Li, and K Kuhnlenz. Fast object recognition and 6d pose estimation using viewpoint oriented color-shape histogram. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2013.

[1]Dynamic Human-Robot-Interaction for Automation Systems Group, Technische Universität München, Munich, Germany li.shile@mytum.de, dhlee@tum.de. [2]Autonomous Intelligent Systems Group, University of Bonn, Bonn, Germany koosy@ais.uni-bonn.de.