Context Enhanced Multi-Object Tracker for Human Robot Collaboration

Akkaladevi Sharath Chandra[a,c]
[a]Department of Robotics and Assistive Systems, Profactor GmbH
Im Stadtgut A2, Steyr, 4407 Austria
sakkal@profactor.at

Matthias Plasch[a]; Christian Eitzinger[b]
[b]Machine Vision Department, Profactor GmbH
Im Stadtgut A2, Steyr, 4407 Austria
{mplasch, ceitzi}@profactor.at

Bernhard Rinner[c]
[c]Institute of Networked and Embedded Systems, Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
bernhard.rinner@aau.at

ABSTRACT

The demanding nature of human robot collaboration (HRC) in terms of robustness and efficiency requirements make object tracking a tough challenge. Instead of treating the area of HRC as an adversary, the idea is to exploit the abundance of context information available in a human robot collaboration scenario to enhance tracking. In this work, a multi object tracking system that uses context information and which is capable of tracking the 3D pose of multiple objects using RGBD data is presented. In order to showcase the importance of context and the enhancement possible for an object tracker when integrated into a cognitive architecture, several experiments are performed and evaluated. This approach is one of the first to apply and evaluate the concept of multiple object tracking in 3D to a human robot collaborative assembly process.

Keywords

Multi-Object Tracking; Context Awareness; Human Robot Collaboration.

1. INTRODUCTION

Assembly tasks in industrial HRC often involve manipulation of several objects and demand coordination between the human and the robot. An example of robotic manipulation in an assembly process is as shown in Figure 1. Recognizing and localizing objects of interest and tracking them to allow a smooth flow of interaction between human and the robot are of vital importance. The object localization and tracking needs to be done in 3D (position + orientation) to facilitate object manipulation by the robot in a dynamic environment that involves human users. Considering multiple objects and real-time computational requirements (especially in industries), tracking objects in 3D becomes even more complicated. However, such industrial settings themselves hold with in an abundance of information that can be exploited to enhance object tracking. One such typical setting is an assembly process that deals with assembling certain objects in a particular fashion. In such cases, the types of objects present depend on the assembly process and the number of objects change as the assembly process progresses. Hence the idea is to use this additional context information to reinforce the tracking system to efficiently verify the tracking and prune false positives (FPs). As a result, a context enhanced multi object tracking system can aid not only in tracking objects in an industrial HRC but also be a vital part of the task verification process. There are not many approaches in literature that deal with multiple object tracking using RGBD data that consider context information in an industrial scenario. The authors in [1] integrate temporal and spatial contextual information to help predict and track human effectors with an active camera in a humanoid robot interaction scenario. Analogously, the authors in [2] show how learning context to support tracking improves robot performance in various tasks. Most HRC approaches with task driven goals only consider object localization for manipulation (often with simple structure) and thereby do not consider the problems associated with dynamic environments tracking and manipulation of real world objects [3].

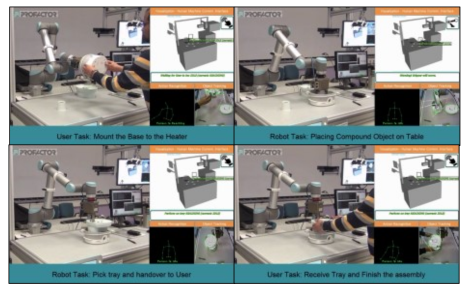

Figure 1: A human robot collaborative assembly process

Unlike other HRC approaches [3], we propose a context enhanced multi-object tracking system (CEMOT)-that is capable of tracking multiple real-world objects in a goal driven assembly process. This work is built upon the multiple object tracking approach (SMOT) proposed in [4]. The CEMOT system is also enabled with capabilities to handle dynamic situations in an industrial HRC scenario with the help of context information. The required context for CEMOT system is provided through the tightly integrated cognitive architecture which helps both the CEMOT system to successfully track objects and the cognitive architecture to verify individual steps in the assembly process. An additional advantage of the CEMOT system is that it can also handle multiple instances of multiple objects, where it exploits the concept of object anchoring to maintain the trajectories of all objects and keep their identities consistent from frame to frame. However, the topic of object anchoring is out of scope and is not in the focus of this work. The remaining part of the paper is organized as follows: In Section 2 the CEMOT system and its constituent parts and data flow are described. The experimental setup and the evaluation of the proposed CEMOT system are carried out in Section 3 followed by some concluding remarks.

HRI '17 Companion, March 06-09, 2017, Vienna, Austria
ACM 978-1-4503-4885-0/17/03.
http://dx.doi.org/10.1145/3029798.3038406

2. THE CEMOT SYSTEM

When applied to a HRC assembly scenario, the standalone object tracker system (SMOT) faces difficulties as during the execution depending on the assembly step, objects can be completely occluded for prolonged durations, some objects disappear and reappear in the scene that result in false positives and difficulty in verification of tracked results. To overcome this problem, the SMOT system is extended to configurable context enhanced multi object tracking (CEMOT) system. The communication relations between the CEMOT system and other components of the cognitive architecture are described in the CEMOT framework as depicted in Figure 2. More details on the cognitive architecture can be found in [5]. Within the CEMOT architecture, the SMOT system requests depth data from the sensor and listens to configuration updates sent by the context translator. SMOT performs object localization and tracking continuously according to the current configuration. As a standalone system SMOT has no additional information (apart from the static initialization). In the CEMOT system the tracker is provided with the following configuration information (called configuration messages) from the perception reasoning (PR) system: a) type of objects in the scene b) no. of instances of each object type c) previous known location of the object and d) what is the current mode (track, localize, forget, ignore) of each object. The PR system given the trigger from the task state reasoning (TSR), contacts the context translator. Given the knowledge of the AP, current task state, the PR can easily provide the context translator with the required input for the configuration message. The context translator prepares this configuration message and then forwards it to SMOT. The SMOT system uses this information to verify locally the tracked results and filters out false positives.

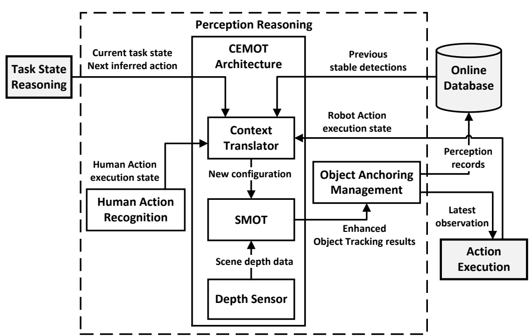

Figure 2: CEMOT framework: The SMOT enhanced with task context information

3. EVALUATION

A total of 20 experiments were conducted, with 10 execution trials for each the SMOT and the CEMOT system. For both trial series the objects considered in the evaluation use case, were arranged similarly in order to provide similar conditions for both approaches. For the execution of the assembly process the following condition has to be met in every trial: The TSR in cooperation with PR has to verify each state Si, before the next action is triggered towards reaching the goal. In the case the system is not able to verify its current state, the assembly process execution is cancelled and the trial is unsuccessful. The evaluation focuses on the accuracy of the object localization approaches SMOT and CEMOT of localizing the objects. In order to provide a measure of performance, a collection of characteristic values, referred in binary classification and machine learning, were chosen and inquired during the experiments. Within the context of object localization, true positives (TP) refer to reported detections at positions, which reflect the real situation. A false positive (FP) is a reported detection at a position where no corresponding object instance is located in the work space. False negatives (FN) refer to existing objects in the real world, but the system was not able to locate them properly. Finally, true negatives (TN) refer to object instances, which are not present in the work space and their absence is correctly confirmed. Based on these values an accuracy measure can be defined as described in the equation below:

$$Accuracy := \frac{TP + TN}{TP + TN + FP + FN}$$

The evaluation of both approaches SMOT and CEMOT is given in Table 1. Both approaches show similar performance in the initial state S_0 (all objects are present). However, in states where certain objects are excluded from tracking, the SMOT detects an increasing number of FPs as it cannot be re-configured dynamically based on the given context. It has to be pointed out that only the CEMOT system is capable of confirming the existence of TNs, thus leading to increased accuracy values, when object instances are required to be excluded from tracking.

Table 1: Evaluation of SMOT Vs CEMOT system

S_i	SMOT/CEMOT						Legend
						Acc.	**Acc**: Accuracy **N**: Number of trials **TP**: True positives **FP**: False positives **FN**: False negatives **TN**: True negatives
	N	TP	FP	FN	TN		
S_0	10/10	37/38	3/2	3/2	0/0	0.86/0.9	
S_1	7/8	20/22	7/2	1/2	0/8	0.7/0.8	
S_2	6/6	12/12	12/0	0/0	0/12	0.5/1	
S_3	6/6	6/6	18/0	0/0	0/18	0.25/1	
S_4	6/6	0/0	24/0	0/0	0/24	0/1	

4. ACKNOWLEDGMENTS

This research is funded by the projects KoMoProd (BMVIT, Austria) and CompleteMe (FFG, 849441)

5. REFERENCES

[1] Lemaignan, S., et al. 2016 Artificial cognition for social human–robot interaction: An implementation. *AI* (2016)

[2] Pandey, A., K., Muhammad A., and Rachid A. 2013. Towards a task-aware proactive sociable robot based on multi-state perspective-taking. *Int. J. of Soc. Robot*., 5,2,.

[3] Ognibene, D., et.al. 2013. Contextual action recognition and target localization with an active allocation of attention on a humanoid robot. Bioinspiration & biomimetics

[4] Akkaladevi, S., et.al., 2016. Tracking multiple rigid symmetric and non-symmetric objects in real-time using depth data. 2016 *(ICRA)*. 5644-5649.

[5] Akkaladevi, S. C. et.al., 2016. Human Robot Collaboration to Reach a Common Goal in an Assembly Process. ECAI