Towards Contextual Robots for Collaborative Manufacturing

José de Gea Fernández[1], Dennis Mronga[1], Martin Günther[1], Sebastian Stock[1], Nils Niemann[1], Hendrik Wiese[1], Rohit Menon[1], Elsa Kirchner[1,2], and Stefan Stiene[1]

I. Motivation

In the near future, rigid production lines will be likely complemented, if not completely replaced, with hybrid teams composed of humans and robots. Robots, in this case, that are collaborative, intelligent, being both stationary and mobile. Remarkably, hybrid teams will take over a variety of assembly operations on the same existing production lines. In this scenario, robots do not solely execute fully-automatized processes but instead become cooperative partners of the human worker, taking over non-ergonomic assembly operations and use contextual information to support selection of tools or adapt to ongoing certain tasks. Until now, companies had to decide in advance either for flexible but manual production processes or for highly-productive, repetitive automation with relatively low flexibility. In the future, there will be no need to decide whether a whole process needs to be fully automatized or not, but rather it will be possible to adapt specific tasks according to the skills of both the human and the robot.

The goal of the Hybr-iT project [1] (funded by the German Federal Ministry of Education and Research (BMBF)) is the development and testing of hybrid teams composed of humans, robots and software-based assistant systems in real industrial environments. One essential component of the project is a resource-oriented architecture (ROA) acting as a middleware for the interconnection of heterogeneous cyber-physical systems and IT environments, and the hybrid teams. Another key component, subject of this paper, is a modular and robot-agnostic control software architecture which seamlessly interacts with the ROA and establishes the information flow for the control of robots in hybrid teams. Through the interplay of the ROA and the robot architecture, existing IT environments and robots should be allowed to be interconnected, so that robots receive additional and valuable information such as semantic task descriptions for hybrid teams, semantic information about the objects, current ongoing tasks and plans, location of objects and humans, to name a few. With the use of this additional contextual information and background knowledge from the IT systems, the aim is to grant robots with more flexible robot skills (such as navigation, interaction, perception or manipulation) which allow them to quickly adapt to the ongoing situation or to changing needs.

II. Methods

The components related to the generic real-time robot control and its interconnection with other components (vision cameras, laser scanners, gesture recognition devices) as well as the initial execution of abstract task descriptions, certain level of reactivity, and the re-use of robot skills were described in a previous work [2]. The current work focuses on the use of environment and contextual information to adapt current robot motions to ongoing tasks or changing contexts. The missing features and, thus, goals for the Hybr-iT project pivot around context-dependent behaviour, intuitive programming and interaction, as well as automatic online plan generation. To endow the system with such a contextual information and adaptivity, the following components are of utmost importance:

1) Whole-Body Control (WBC). A constraint-based controller (Whole-Body Control) is used as a central control component. In WBC, tasks like "apply a contact force", "follow a trajectory" or "keep away from an obstacle" are described as numerical constraints to an optimization problem, whose solution represents the robot's control input. In this project we try to extend the concept of WBC and automatize the selection and parametrization of task constraints given a high-level (semantic) descriptions of the task. The long-term goal is to generalize constraint-based task descriptions over a variety of similar contexts and, thus, increase the autonomy of robotic systems in industrial HRI setups.
2) Perception. The goal of the perception module is especially object recognition. However, the main focus is on the capability of active search of objects which are relevant for the current context. Moreover, the use of semantic object descriptions coming from the IT systems should help to recognize objects.
3) Person tracking. This module is in charge of tracking the location of the people cooperating with the robot by using a combination of vision-based skeleton tracking algorithms and information from sensor suits (equipped with inertial measurement units) worn by one or more operators. The key is in combining different sensors to overcome limitations of the individual ones, i.e., occlusion occurring during vision-based tracking and drifts of the sensor data recorded by the suits. The gathered information will also be used for the prediction of motion of the operator, collision avoidance, intention

*This work was supported through a grant of the German Federal Ministry of Education and Research (BMBF, FKZ 01IS16026A)

[1]Robotics Innovation Center, German Research Center for Artificial Intelligence (DFKI), 28359 Bremen, Germany jose.de_gea_fernandezdfki.de

[2]Robotics Research Group, Faculty of Mathematics and Computer Science, University of Bremen, 28359 Bremen, Germany

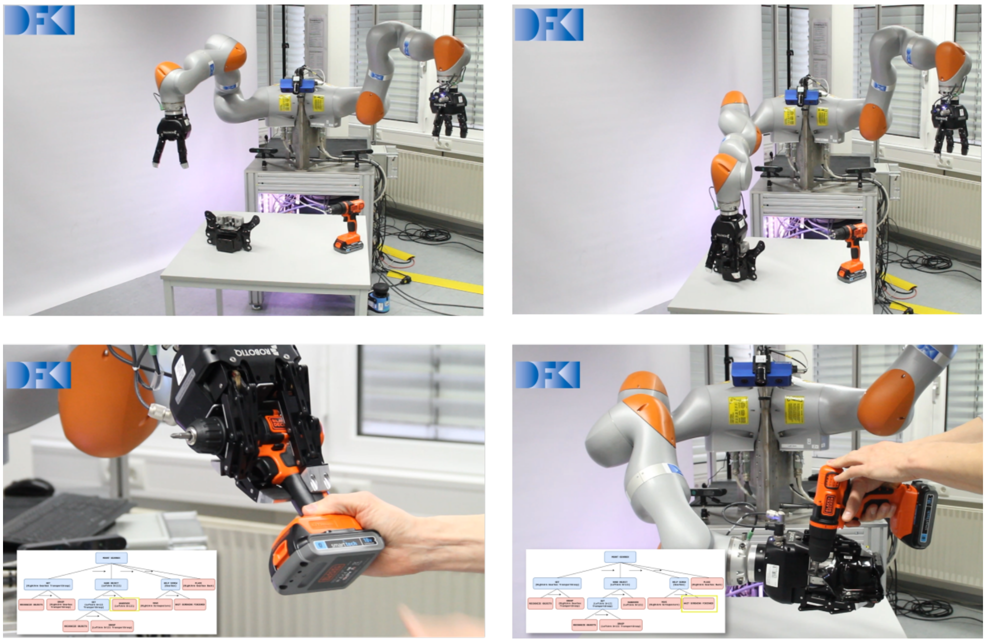


Fig. 1: Snapshots of the demonstration during the first vertical integration of the software components, in which human and robot cooperate in the assembly of a gearbox.

recognition, and recognition of certain gestures with which the operator will be able to communicate with the robot.

4) Environment representation. This component holds a representation of the objects known by the robot, with their location in space and additional information about them (such as confidence of a certain object type or optimal grasp points). Moreover, it can aid the active object search providing interesting search areas given the context, last position seen, etc. The main goal is to combine geometric with semantics: object types, location, poses, CAD models and object-centric information like context-specific grasping points. That would allow to infer and answer such questions such as ”how many round objects are 1 meter away from the screwdriver?”. The symbolic information will be used by the task planner and the geometric information by the robot execution layer.
5) Task planning and monitoring. For task planning the hierarchical hybrid task planner CHIMP[3] is used. It plans by decomposing abstract tasks in subtasks taking into account both causal and temporal constraints as well as available resources. Its rich hybrid causal and temporal representation of the plan will also be used for plan execution and monitoring.
6) Context generation. This component generates information that can be used to characterize a certain situation and, thus, can be considered relevant for adapting the interaction between robot, human and task.

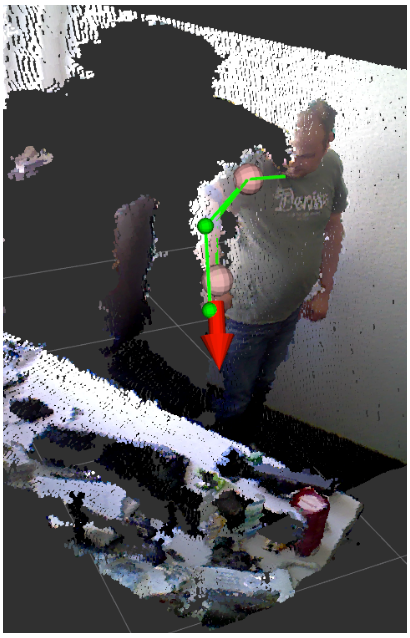
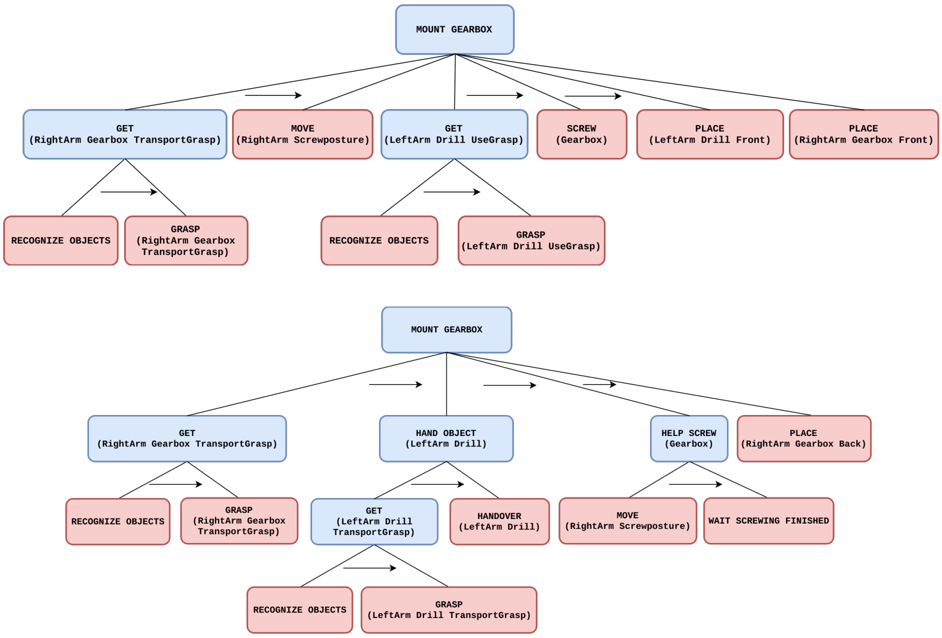


Fig. 2: Screenshot of the person tracking visualization showing the skeleton tracking and the IMU sensor-based tracking of the right arm

III. Preliminary Results

Figure 2 shows a screenshot of the visualization of the person tracking system. In the image, the right arm of the operator is tracked by the vision-based skeleton tracking as well as by two motion sensors (IMUs), one at each limb. The green lines depict the tracked limbs (the links between the joints from both sources, i.e. from the visual and the motion



(a)

(b)

Fig. 3: Context-dependant plans: (a) Robot performs task by itself (no human around), (b) Robot cooperates with a human to perform the task

sensors). The joints tracked by the skeleton tracking are depicted by the bigger light pink spheres. The joints, as they are reconstructed from the IMU tracking system, are shown as small green spheres. Furthermore, the red arrow shows the pose of the right wrist joint and points along the primary axis of the operator's lower right arm. This information can be used, for instance, to coordinate a handover operation between robot and human.

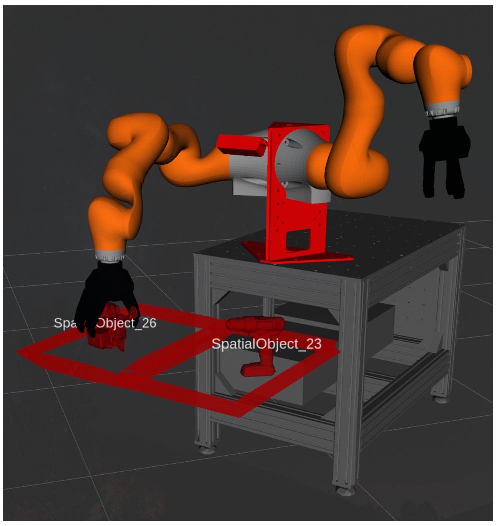


Fig. 4: Screenshot of the environment representation. Objects that have been recognized by the object detection are labeled with their unique ID.

Figure 1 shows some snapshots during a demonstration of an initial first vertical integration of the main software components for a knowledge-based and contextual robotic manipulation. The demonstration includes initial versions of the task planner (CHIMP), a semantically-annotated environment representation (shown in Figure 4), a person tracking used to perform a robot-human handover action, an event-based process monitoring and execution layer as well as a generic robot-agnostic real-time control of the robot. The integration of those components is shown in a context-dependent manipulation task: grasping and mounting a gearbox. The task is once performed with the support of the worker and once autonomously. The robot behavior and, thus, the task plan is generated online depending on the task context (in this simple case, whether the human worker is around or not, and thus, whether the task is going to be performed in cooperation with a human or not). Figure 3 shows two examples of such plans. In the first plan the robot works alone and in the second plan the robot assists the worker by handing him the screwdriver and holding the gearbox. Notice these are not two pre-defined plans and one or the other will be executed, but one or the other plan will be online generated depending on the task context and other current causal, temporal and resource constraints.

The current status shown in this demonstration for the different modules is the following: for the object recognition, a RGB camera with an Ensenso 3D stereo camera are integrated. The first one delivers color information to obtain a colored 3D point cloud. Using PCL-based segmentation and off-the-shelf object recognition, the pose of the object is estimated by matching this information with the one from the CAD models stored on the environment representation. For the environment representation, both geometric and semantic information is integrated and connected with task planner, execution layer and object recognition modules. The task planner CHIMP receives a goal task as well as symbolic information about the objects from the environment representation. The output plan is then sent to the robot execution layer. Context information is used to select grasping points for the cordless screwdriver. Depending on whether the robot is going to use the screwdriver itself (human is not around) or hand it over to a human, the grasping pose will be chosen accordingly from the ones stored together with the object in the environment representation. In the handover case, the target pose of the gripper handing over the drill is related to the pose of the right wrist of the human, detected and localized in six degrees of freedom by the person tracking component (red arrow in Figure 2).

IV. Acknowledgment

This work was supported through a grant of the German Federal Ministry of Education and Research (BMBF) with Grant Number 01IS16026A.

References

[1] "Project Hybr-iT, DFKI Website," https://robotik.dfki-bremen.de/en/research/projects/hybr-it.html, [Online; accessed 14-May-2018].

[2] de Gea Fernández J., M. D., G. M., K. T., W. M., S. M., T. M., S. S., K. E., B. V., B. T., T. J., K. T., and K. F., "Multimodal sensor-based whole-body control for humanrobot collaboration in industrial settings," *Robotics and Autonomous Systems*, vol. 94, pp. 102 – 119, 2017.

[3] S. Stock, M. Mansouri, F. Pecora, and J. Hertzberg, "Online task merging with a hierarchical hybrid task planner for mobile service robots," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 6459–6464.