In: RoboCup@Home League Team Descriptions, Bordeaux, France, July 2023.

NimbRo@Home 2023 Open Platform League Team Description

Raphael Memmesheimer, Jonas Bode, Malte Splietker, Simon Bultmann, Benedikt T. Imbusch, and Sven Behnke

Autonomous Intelligent Systems, Computer Science, Univ. of Bonn, Germany
nimbroathome@ais.uni-bonn.de
https://www.ais.uni-bonn.de/nimbro/@Home/

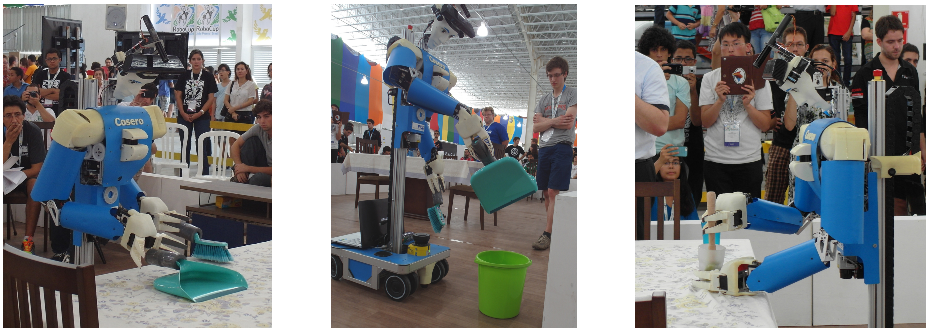
Abstract. This team description paper describes the setup, contributions and efforts of the NimbRo@Home team of the Autonomous Intelligent Systems group from the Rheinische Friedrich-Wilhelms-Universität Bonn for the intended participation at RoboCup@Home Open Platform League taking place in 2023 in Bordeaux, France. We plan to attend the competition with a PAL Robotics TIAGo++ omnidirectional, two-armed robot platform. Further, we describe our intended approaches for object pose and grasp estimation, semantic mapping and human-robot-interaction. Our software contributions can be found on" https://github.com/AIS-Bonn/.

1 Introduction

The NimbRo team has a well established track record of successful participation in various robotic competitions ranging from domains like humanoid soccer in the RoboCup AdultSize league, unstructured environments like the DARPA Grand Challenge 2016 to autonomous bin picking challenges like the Amazon Picking Challenge. Recently, in 2022 the NimbRo team won the ANA Avatar XPRIZE challenge. The team already took successfully part in the RoboCup@Home league and won three consecutive international RoboCup@Home competitions (2011 Istanbul [7], 2012 Mexico City [6], 2013 Eindhoven [5] and also won numerous RoboCup@Home German Open challenges. An excerpt of our performance during the RoboCup@Home final is given in Figure1, where we utilized a dustpan, interacted with a trash can and prepared a cocktail. We focused on two-armed manipulation and tool usage in our demonstrations. With our intended RoboCup@Home 2023 participation, we aim at reinitiating our domestic autonomous service robot activities.

We developed methods for real-time environment and object perception, 3D object pose and grasp estimation using 3D sensors such as laser scanners and RGB-D cameras. We developed efficient planning methods for navigation and object manipulation. Further, we developed approaches for improving syntactical data generation methods. Furthermore, the robots are equipped with a multimodal dialogue system.

In this paper we briefly outline the intended robotic platform, our scientific contributions and give a coarse overview of our control approaches.



(a) Grasping dustpan

(b) Putting dust from duspan into tash can

(c) Preparing a Caipirinha

Fig. 1: RoboCup@Home Final 2014 demonstration.

2 Hardware

We are elaborating different hardware setups, namely using a PAL Robotics TIAGo[1] platform as depicted in Figure2 or extending our custom build AVATAR robot [4]. The TIAGo++ robot is equipped with an omnidirectional platform on mechanum wheels, a linear liftable torso with two 7-DOF arms, a pan-tilt-unit with an RGB-D camera. We aim at incorporating additional computational resources like a NVIDIA Jetson AGX Orin development kit, and a notebook to support the processing of complex models, mainly for vision purposes. Additional sensors like 3D laser range finders and directional microphones are intended to be added to our robotic platform.

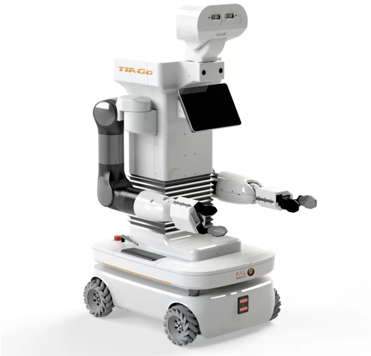


Fig. 2: TIAGO++ omnidirectional robot platform.

[1] https://pal-robotics.com/robots/tiago/

3 Perception

A research focus of the team lies in the object- and environment perception. In the following section, we present our perception-related scientific contributions.

3.1 Synthetic-to-Real Domain Adaptation using Contrastive Unpaired Translation

We developed a pipeline for obtaining semantic segmentations of real images based on solely 3D meshes in [3]. From the 3D object meshes, we generate images using a modern synthesis pipeline. We customize a state-of-the-art image-to-image translation method to adapt the synthetic images to the real domain, minimizing the domain gap in a learned manner. The translation network is trained from unpaired images, i.e. just requires an un-annotated collection of real images. Our method yields robust task performance in real settings, while not requiring any manual annotations. The pipeline is visualized in Figure 3.

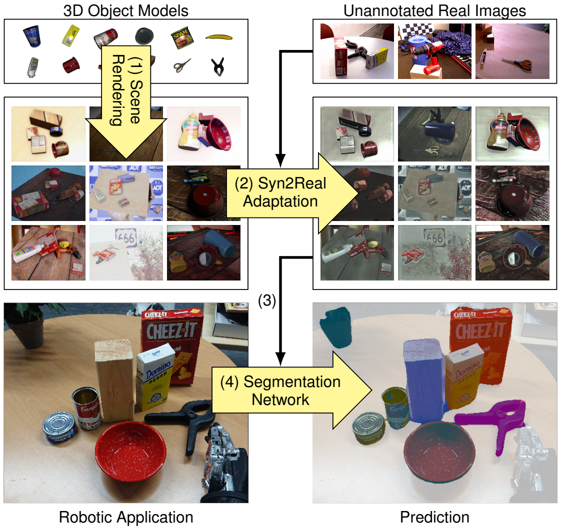


Fig. 3: We simulate and render plausible scenes from the 3D meshes (1). Our adaptation model aligns the synthetic and real image distributions more closely (2). The refined image dataset is used to train a task-specific network (3), which is applied in the target domain (4).

3.2 YOLOPose: Transformer-based Multi-Object 6D Pose Estimation using Keypoint Regression

With YOLOPose [1] (see Figure4), we presented a Transformer-based single-stage multi-object pose estimation method using keypoint regression. Our model

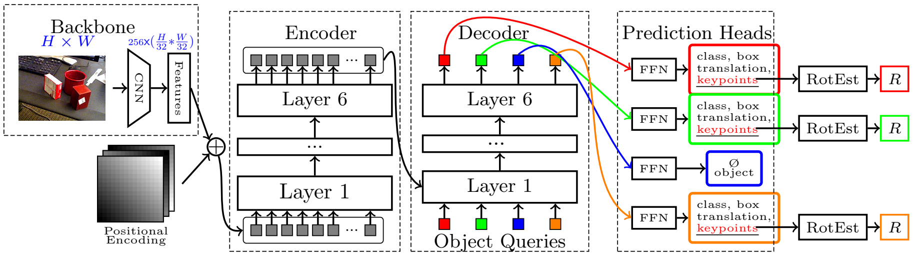
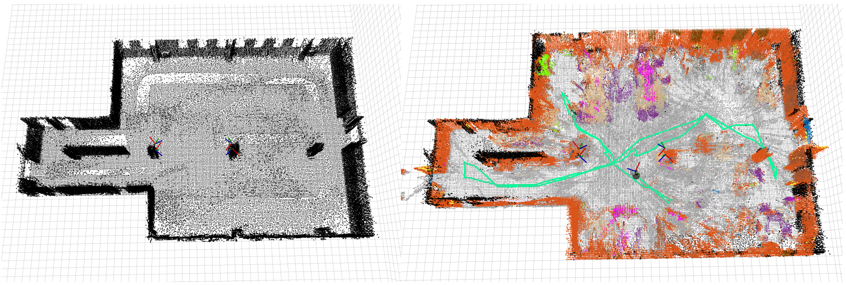


Fig. 4: YOLOPose architecture in detail. Given an RGB input image, we extract features using the standard ResNet model. The extracted features are supplemented with positional encoding and provided as input to the Transformer encoder. The encoder module consists of 6 standard encoder layers with skip connections. The output of the encoder module is provided to the decoder module along with N object queries and the decoder module also consists of 6 standard decoder layers with skip connections generating N output embeddings. The output embeddings are processed with FFNs to generate a set of N elements in parallel. Each element in the set is a tuple consisting of bounding box, class probability, translation and interpolated bounding box keypoints. A learnable orientation estimation module is employed to estimate object orientations R from the predicted keypoints.

jointly estimates bounding boxes, class labels, translation vectors, and pixel coordinates of 3D keypoints for all objects in the given input image. Employing the learnable RotEst module to estimate the object orientation from the predicted keypoints coordinate enables the model to be end-to-end differentiable.

3.3 Visual Pose Estimation with Smart Edge Sensors and Collaborative Semantic Mapping

We developed an external camera-based mobile robot pose estimation approach for collaborative perception with smart edge sensors. Our approach allows for the initialization and correction of a mobile robot's pose from N external static cameras. Furthermore, robot observations from changing viewpoints are fused into the allocentric scene model to extend the view of the static cameras. The robot pose is estimated using a robot detection and keypoint estimation approach trained on a combination of synthetic and real data. The robot pose is then recovered by minimizing the reprojection errors in multiple views. We verify the performance of our approach with various experiments using a Toyota HSR robot in an approximately 250 square meter lab. First, we demonstrate the external visual pose estimation for initialization of the robot pose in the map, when having no or a far-off initial localization. Our approach shows to perform well for initial camera-based localization. Second, we evaluate the accuracy of our pose estimation approach using an HTC Vive tracking system as a reference.



(a) Initialization (b) After Exploration

Fig. 5: Resulting semantic map of the proposed collaborative semantic mapping approach with improved consistency by external pose refinement. The approach is initialized with an empty map (a) and a mobile smart edge sensor node explores the environment, resulting in an explored final semantic map (b).

Through the continuous correction via localization feedback from the external cameras, the pose error remains below a few centimeters. We integrate the external pose estimation approach in a collaborative semantic mapping scenario. The external localization improves the consistency of the resulting semantic map. Finally, we demonstrate long-term robustness in a highly cluttered environment (see Figure5). The laser-based robot-internal localization accumulates a high localization error over longer paths, which leads to unreachable targets or high positional errors at the target locations that could lead to collisions. Our external pose estimation can correct and compensate for the localization errors for better robustness.

3.4 Directional Volumetric Grasping Network

We developed an extension of the Volumetric Grasping Network (VGN) which uses the Directional Truncated Signed Distance Function (DTSDF) to model the grasp scene. Our preliminary results show promising gains in overall grasp success. We retrained the network with training data specialized to the HSR's gripper configuration. Moreover, we implemented a grasping pipeline to test the directional VGN on the HSR. Depth images from the head-mounted depth sensor are fused into the DTSDF, which is input into the directional VGN for grasp pose synthesis. Our recent experiments on the HSR show, that the autonomous grasping pipeline works, as demonstrated in the attached video file. In the next step, we will train the network with more data and quantify our results with extensive experiments on synthetic data and on the HSR.

4 Behaviour Control

We developed an abstraction layer for integration of novel robot behaviours based on state machines encapsulating basic behaviours and functionalities, having in mind that these are replaced by learned behaviours in the future. In the following section, we give a coarse overview of the underlying behaviour control approaches.

4.1 Mapping and Navigation

Our mapping approach is based on the approach presented by Grisetti, Stachniss, and Burgard [2] with additional layers included containing measurements from additional sensors like the RGB-D camera and encoding additional semantic information about the environment. The additional layers are incorporated to adapt navigation to avoid obstacles invisible for the laser range finder or improve social navigation by predicting human motions. Our approach enables collaborative perception for pose-refinement, semantic mapping and human-aware navigation from the robot's sensors and incorporate measurements from instrumented environments like external static cameras like described in Section 3.3. We also consider hierarchical definitions of the locations to refine the robot's pose from various position and orientations, especially with larger manipulation-centric locations like tables.

4.2 Task Execution

We employ an abstraction layer for simplifying higher-level behaviour development of complex state machines by re-using generalized sub-statemachines on various levels of complexity on a functionality and behaviour level. These statemachines are designed such that their execution and individual parameters can be mapped from natural language descriptions. In its current state the state machines are sequences of manually defined behaviour and functionality sequences, however our system designed to learn these behaviours and functionalities e.g. by demonstrations.

5 Conclusions

In this team description paper we presented the intended robot platform, scientific contributions and intended approaches for a RoboCup 2023 participation of team NimbRo in the RoboCup@Home Open Platform League. Our team previously participated successfully in the RoboCup@Home competition. With our intended RoboCup@Home participation, we are aiming at resuming our autonomous mobile domestic service robot research activities. Our most recent developments about our team can be found on our team webpage `https://www.ais.uni-bonn.de/nimbro/@Home/`.

References

[1] Arash Amini, Arul Selvam Periyasamy, and Sven Behnke. "YOLOPose: Transformer-based Multi-Object 6D Pose Estimation using Keypoint Regression". In: *CoRR* abs/2205.02536 (2022). DOI: `10.48550/arXiv.2205.02536`. arXiv: `2205.02536`. URL: `https://doi.org/10.48550/arXiv.2205.02536`.

[2] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. "Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters". In: *IEEE Trans. Robotics* 23.1 (2007), pp. 34–46. DOI: `10.1109/TRO.2006.889486`. URL: `https://doi.org/10.1109/TRO.2006.889486`.

[3] Benedikt T. Imbusch, Max Schwarz, and Sven Behnke. "Synthetic-to-Real Domain Adaptation using Contrastive Unpaired Translation". In: *18th IEEE International Conference on Automation Science and Engineering, CASE 2022, Mexico City, Mexico, August 20-24, 2022*. IEEE, 2022, pp. 595–602. DOI: `10.1109/CASE49997.2022.9926640`. URL: `https://doi.org/10.1109/CASE49997.2022.9926640`.

[4] Max Schwarz et al. "NimbRo Avatar: Interactive immersive telepresence with force-feedback telemanipulation". In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, pp. 5312–5319.

[5] Jörg Stückler et al. "Increasing Flexibility of Mobile Manipulation and Intuitive Human-Robot Interaction in RoboCup@Home". In: *RoboCup 2013: Robot World Cup XVII [papers from the 17th Annual RoboCup International Symposium, Eindhoven, The Netherlands, July 1, 2013]*. Ed. by Sven Behnke et al. Vol. 8371. Lecture Notes in Computer Science. Springer, 2013, pp. 135–146. DOI: `10.1007/978-3-662-44468-9\_13`. URL: `https://doi.org/10.1007/978-3-662-44468-9%5C_13`.

[6] Jörg Stückler et al. "NimbRo@Home: Winning Team of the RoboCup@Home Competition 2012". In: *RoboCup 2012: Robot Soccer World Cup XVI [papers from the 16th Annual RoboCup International Symposium, Mexico City, Mexico, June 18-24, 2012]*. Ed. by Xiaoping Chen et al. Vol. 7500. Lecture Notes in Computer Science. Springer, 2012, pp. 94–105. DOI: `10.1007/978-3-642-39250-4\_10`. URL: `https://doi.org/10.1007/978-3-642-39250-4%5C_10`.

[7] Jörg Stückler et al. "Towards Robust Mobility, Flexible Object Manipulation, and Intuitive Multimodal Interaction for Domestic Service Robots". In: *RoboCup 2011: Robot Soccer World Cup XV [papers from the 15th Annual RoboCup International Symposium, Istanbul, Turkey, July 2011]*. Ed. by Thomas Röfer et al. Vol. 7416. Lecture Notes in Computer Science. Springer, 2011, pp. 51–62. DOI: `10.1007/978-3-642-32060-6\_5`. URL: `https://doi.org/10.1007/978-3-642-32060-6%5C_5`.

Name of team : NimbRo@Home
Member : Raphael Memmesheimer, Malte Splieker, Michael Schreiber, Christian Lenz, Benedikt T. Imbusch, Jonas Bode, Sven Behnke (TBC)
Contact information : memmesheimer@ais.uni-bonn.de
Website : `https://www.ais.uni-bonn.de/nimbro/@Home/`
Hardware :

- PAL Robotics TIAGo++ Omni Edition

Software :

- ROS
- OpenCV
- PCL
- PyTorch
- Custom software:
  - 3D Semantic Scene Perception using Distributed Smart Edge Sensors `https://github.com/AIS-Bonn/SmartEdgeSensor3DScenePerception`
  - Online Marker-free Extrinsic Camera Calibration using Person Keypoint Detections `https://github.com/AIS-Bonn/ExtrCamCalib_PersonKeypoints`
  - Directional TSDF InfiniTAM `https://github.com/AIS-Bonn/DirectionalTSDF`)
  - Real-Time Multi-View 3D Human Pose Estimation using Semantic Feedback to Smart Edge Sensors `https://github.com/AIS-Bonn/SmartEdgeSensor3DHumanPose`
  - ROS transport for high-latency, low-quality networks `https://github.com/AIS-Bonn/nimbro_network`
- More open-source releases can be found here: `https://github.com/AIS-Bonn`