

sDSPL - Towards a benchmark for general-purpose task evaluation in domestic service robots

Luis Contreras ¹, Yosuke Matsusaka ², Takashi Yamamoto ², and Hiroyuki Okada ²

¹Tamagawa University

²Affiliation not available

October 30, 2023

Abstract

Although the skills required to solve isolated robotics problems are reaching amazing performances recently, we propose the evaluation of such individual solutions in fully integrated robot systems tested in real daily situations like those presented at international robotics competitions. The simulation Domestic Standard Platform League (sDSPL), which utilizes the HSR simulator developed for the World Robot Summit, surges from the necessity to standardise and spread the research on Domestic Service Robots where a series of solutions can be tested to solve a general-purpose task in a standard domestic environment; this approach has been proven successful at several international competitions, namely, the RoboCup Japan Open, the Mexican Tournament of Robotics, and the RoboCup 2021.

******This work was accepted to The 39th Annual Conference of the Robotics Society of Japan.

sDSPL - Towards a benchmark for general-purpose task evaluation in domestic service robots

oLuis Contreras^{1,*}, Yosuke Matsusaka², Takashi Yamamoto³, Hiroyuki Okada¹

Abstract—Although the skills required to solve isolated robotics problems are reaching amazing performances recently, we propose the evaluation of such individual solutions in fully integrated robot systems tested in real daily situations like those presented at international robotics competitions. The simulation Domestic Standard Platform League (sDSPL), which utilizes the HSR simulator developed for the World Robot Summit, surges from the necessity to standardise and spread the research on Domestic Service Robots where a series of solutions can be tested to solve a general-purpose task in a standard domestic environment; this approach has been proven successful at several international competitions, namely, the RoboCup Japan Open, the Mexican Tournament of Robotics, and the RoboCup 2021.

I. INTRODUCTION

The presence of service robots in domestic environments is increasing recently and, in consequence, the familiarity of users and the difficulty of the tasks that might request. Therefore, standardisation on full task performance (in contrast to individual skills' evaluation) is necessary to compare different software and hardware implementations; furthermore, complex scenarios in natural places are preferable over controlled laboratory spaces. In this context, we propose international robot competitions – and we offer a standard simulation platform to evaluate domestic tasks – as a benchmark for general-purpose service-robot performance.

The rest of this paper is organized as follows. Section II presents a review on service robots' use and performance evaluation and Section III presents the Tidy-Up task as in the World Robot Summit (WRS) and RoboCup international competitions. Section IV expands on the robot evaluation at a competition level. The paper ends outlining our conclusions.

II. RELATED WORK

In [1], the authors presented a novel navigation system for service robots by training an Attention Branch Network [[2]] to learn visual cues when navigating based on human behaviours and semantic information from their global path planning [[3]] and they tested it in a service robot in an indoors environment. While able to navigate autonomously by extracting an attention map that allows it to generate a series of behaviours to traverse the path based on visual information, dynamic information was not included in the study, as expected in real home environments.

Other applications using service robots include object manipulation; for example, in [4] a robot starts their motion planing system using previous experience in consistent situations. Similarly, in [5] and [6], the authors present a system where a service robot is able to manipulate unknown doors by receiving a single users instruction. Although those system presented high performance, the limitation of the application and the necessity of user feedback make them hard to extend those systems to general purpose service robots.

A common object manipulation approach that has been proven efficient uses the current RGBD information to perform a series of simulations to obtain the best grasping strategy, however, dealing with partial information due to view occlusions can result in non-optimal solutions; furthermore, when the task consists on manipulating dozens or even hundreds of objects, the solution becomes inefficient. Recently, a test platform for several manipulation applications has been presented in [7] where they use a simulator with a fixed robot arm on a surface and an upper RGBD camera and an eye-in-hand monocular camera, and they feature 100 different tasks providing proprioceptive and visual observations.

As we can see, although several application have been developed, they mostly are task-specific and highly depend on the experimental setup. It's in this context that the use of robot competitions as a benchmark for robot systems' performance has been proposed [[8]], where several skills can objectively be evaluated in human-like scenarios to mark trends and challenges in the area as in [9] and [10]. The most recent work towards this goal is Habitat 2.0 [11] where the authors present a simulated home environment where the robot can perform several tasks; however, no specific rules and regulations are provided. Furthermore, the authors of this last study express that their "experiments suggest that complex, multi-step tasks such as setting the table or taking out the trash are significantly challenging. Although we were able to train individual skills (pick, place, navigate, open drawer, etc) [...] training a single agent that is able to accomplish all such skills and chain them without cascading errors remains an open challenge."

Moreover, differences in performance between laboratory and competition setups are presented at competition scenarios where teams have few chances to test their solutions. In the particular task of service robots attending users in a restaurant, while [12] and [13] developed a system able to navigate in complex scenarios, their system was tested in laboratory setups with few participants and without dynamic obstacles. On the other hand, [14] developed a navigation

The authors are with the ¹Advance Intelligence and Robotics Research Center, Tamagawa University, ²MID Academic Promotions Inc, and the ³Toyota Motor Corporation.

*contreras.luis@lab.tamagawa.ac.jp. o Presenter

system able to work in unknown environment and with dynamic obstacles at a competition level where pedestrians occasionally interfere and interact with the robot in unexpected ways that the robot has to deal with while solving the given task.

Similarly, in the Tidy Up task, where a robot has to clean a room and place the objects in their correct locations, [15] has shown a high performance in object recognition and manipulation; however, they highly depend on external camera devices and high and external computing. On the other hand, in [16], [17], the proposed system operates autonomously in the same task and follows standard rules and regulations like those proposed in WRS and RoboCup to evaluate their performance.

III. TIDY UP

Following the rules presented at WRS - Partner Robot Challenge (Real Space), we proposed the Tidy-Up task to evaluate the robot performance. It consist on taking objects from the incorrect locations to a predetermined deposit and then providing a person within a group with some food from a shelf when requested while avoiding obstacles when navigating (<https://worldrobotsummit.org/>). It uses the YCB Object and Model set ([18]) consisting of objects commonly found in home and office environments and they vary on shape, material, size, texture, weight, etc.

Whereas promoting Smart solutions, the key performance indicator is based on a 4S philosophy: Speed, Smooth, Stable, and Safe, using a compact field that can be easily set up anywhere to allow for continuous evaluation for a variety of research activities [19]. While rewarding actions like opening the drawers, depositing objects softly and in free spaces and, in some cases, placing objects according to a specific orientation, the rulebook discourages actions such as dropping or hitting objects and furniture, false deliveries, and delivering an object in an occupied space that might prevent teams from getting full scores in a specific action.

To show the variety of solutions aim at solving the same task, we will present few different proposal in the Tidy-up task proposed by some of the participating teams – a complete overview on team approaches to solve the same task can be seen in their Poster Presentation at <https://bit.ly/3h1r2Q8>.

In [17], eR@sers Team uses a system where the robot constantly updates its belief by gathering spacial, visual, and contact information in the low-level behaviours during the object manipulation process while Hibikino-Musashi Team in [20] presents their object detection and recognition systems based on data augmentation by using a 3D scans of the objects and automatic annotations in different scales, orientations, and backgrounds.

Similarly, in [21] OIT-RITS Team presents their object detection and recognition strategy in cluttered scenes where they first segment an image using edge information and then form groups of segments with similar color information to generate an object view from similar partial views and their

corresponding 2D position in the image. Then, they feed a recognition system with the generated view of the object.

Finally, AISL-TUT Team in [22] presents a reactive system to unexpected situations. They consider grasping errors by using contact sensors in the hand and navigation errors using the laser in the base: when an obstacle is in the way to a target location, they use the upper camera to find it and perform an action to remove it and free the path. However, they do not consider changes in the environment in the manipulation process.

IV. VIRTUAL COMPETITION PERFORMANCE

As mentioned before, simulation DSPL is highly based on WRS and, therefore, some rules might not be applicable to a simulated environment, especially those limited by real physical and mechanical principles (<http://humansupportrobot.org/robocup2021-dspl-simulation/>). The sDSPL has been evolving from the WRS Real Space to a virtual setup through several local tournaments as follows.

In [23], a simulator that uses the Toyota's Human Support Robot (HSR) ([24]) and that fits the current WRS's rulebook was developed; this system was tested at the RoboCup Japan Open 2020 (<https://www.robocup.or.jp/japanopen2020b-en/>) and the results can be seen at <https://bit.ly/3hqOigF>. Then, the next step was adapting the WRS rules to a RoboCup competition format, divided in several tasks and stages, while the proposed simulator system could still be used; this format was first tested at the Mexican Tournament of Robotics 2021 (<https://www.femexrobotica.org/tmr2021/en/portfolio-item/robocup-standard-platform-league/>) where the results can be seen at <https://bit.ly/3qQfwvI>.

A. RoboCup Worldwide 2021

Finally, to meet the necessities given the current global situation, this system was tested in a worldwide scenario that allowed teams to participate remotely. The format in the RoboCup competition (<https://athome.robocup.org/home-virtual-2021/>) consisted in two stages:

- Stage I
 - Clean Up (5 min)
 - Go and Get It (5 min)
- Stage II
 - Clean Up (15 min)

The proposed platform has the flexibility to be used with several robot models that enables it to be used in several leagues. In particular, the RoboCup was divided in three leagues:

- Domestic Standard Platform League (DSPL)
- Social Standard Platform League (SSPL)
- Open Platform League (OPL)

In the DSPL and SSPL, teams use the HSR as standard robot platform to solve a given task, as shown in Figure 1.

On the other hand, in OPL, teams are allowed to modify the provided architecture or using their own; to illustrate this,



Fig. 1: Standard robot model (HSR) in the arena; the robot has to clean the first room and or provide a food item to the user that requests it.

an example template using the robot TIAGo [25] has been provided in the simulator; we can observe in Figure 2 two different OPL robot models.

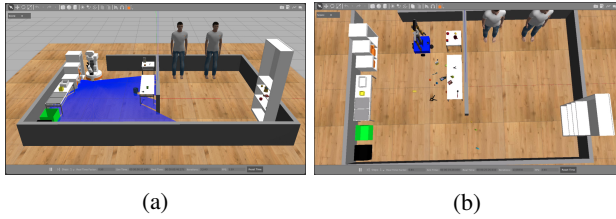


Fig. 2: Two different OPL robot models. a) Robot TIAGo [25], an off-the-shelf service robot, and b) robot Justina (as in [26]), a custom service robot for research.

A template using the HSR can be found at <https://github.com/devrt/robocup-at-home-2021-challenge> and, for the TIAGo robot, at <https://github.com/devrt/robocup-at-home-2021-opl-challenge>. A complete overview of robot models provided by OPL teams can be seen at <https://bit.ly/3qF617f>.

Using either a standard robot (DSPL or SSPL) or a custom (OPL) robot model brings different challenges; while a standard platform allows a research group to focus on the algorithmic part and consistently compare different proposals in the same setup, an open platform allows them to overcome physical limitations of a fixed model. Table I shows the performance in the different leagues to solve the same problem. We can observe that using a standard platform permits teams to focus on the tasks at hand allowing them to obtain the highest performance; however, once the limitations of the system are reached, it also allows all teams to imagine new solutions to improve the platform. An overview of the research output from different groups in all categories can be seen in the RoboCup's Open Demonstrations at (<https://bit.ly/3AjV0fT>) while all participating teams' results in the different leagues can be seen at <https://athome.robocup.org/rc2021/>.

B. Technical Challenge

In addition to a competition format, but still promoting a full task performance as a baseline for robot evaluation, we propose a technical challenge by completing the Tidy-Up task as in the World Robot Summit for real robots. It's a 20 minutes test where Task 1 (Clean Up, 15 min) and then Task 2 (Go and Get It, 5 min) evaluated as a single continuous test. The score is the one that the system provides (i.e. no manual inspection to add or remove any point is performed); a mean or median performance after several runs is considered to account for any errors and inconsistencies in the simulator and the scoring system.

During the competition, teams are invited to register their repositories to participate and all tests are run in the cloud (i.e. all commands are provided in a Dockerfile without any further user intervention); however, for research, this can be run locally and users can report their results after several runs with the same or different seeds that should be provided in order to be able to replicate the experiments.

The command to locally randomise the objects in the arena, depending on the robot used, is:

```
$ sim launch hsrb_wrs_gazebo_launch
wrs_practice0_easy_tmc.launch
use_oss_stacks:=true
fast_physics:=true
highrtf:=true seed:=10
per_category:=6
obstacles:=3
```

```
$ sim launch
tiago_roboocup2021_opl_gazebo_launch
practice0.launch seed:=10
per_category:=6
obstacles:=5
```

Table II shows the best results per task (Clean Up and Go and Get It) as well as the best Technical Challenge team among all leagues. It can be observed that the use of a standard platform, like the HSR, allows teams to focus on solving the task at hand without limiting innovation.

In addition to robot competitions, that are very limited through the year and can be hard for a single research group to participate in all of them to test their latest developments, we propose automatic scoring systems as service robot benchmarks; this approach might allow teams to consistently measure their progress and objectively compare their results with other research groups that, in consequence, will stimulate the research in full service robots and not only in the individual skills that integrate them. The constant improvement of the simulators and the increasing scope and difficulty of the tasks to solve should consider a deep analysis of real robot competitions as the state-of-the-art for this kind of systems while real robot competitions should consider the results on these benchmarks to broaden their scope, as well.

V. CONCLUSIONS

We propose a simulation Domestic Standard Platform League (sDSPL) as a benchmark to evaluate the performance of service robots while executing a task (in contrast to

TABLE I: Scores.

	Place	Team Name	Stage I		Stage II	Final Score
			Clean Up	Go and Get It	Clean Up	
DSPL	1	<i>Tidyboy</i>	90/150	250	390	790
	2	<i>Hibikino</i>	110/130	250	330	710
	3	<i>Austin Villa</i>	40/0	250	80	370
SSPL	1	<i>UChile</i>	30	250	20	300
	2	<i>LyonTech</i>	-	100/100/40	110	210
	3	<i>RoboBreizh</i>	10	150/100	20	180
OPL	1	<i>homer</i>	30	100/140	100	270
	2	<i>CATIE</i>	0	175/130	40	215
	3	<i>RTLions</i>	0/0	100	40	140

TABLE II: Awards.

Awards		
Technical Challenge	Tidyboy	DSPL
Best Test Performance (Clean Up)	Tidyboy	DSPL
Best Test Performance (Go, Get It!)	Hibikino	DSPL
Open Challenge	Hibikino	DSPL
Best Poster	Serious CC	OPL

benchmarks focused on evaluating individual skills). We believe that a standard system can boost research in two main aspects: first, using the same robot and setup, a research group can consistently test several baselines to solve a given task; second, using a standard robot architecture to optimally solve a given task, a research group can evaluate the impact of any hardware modification or, even, they can assess their own custom robot models.

REFERENCES

- [1] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branchの導入による深層強化学習を用いたロボットナビゲーションの視覚的説明," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [2] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," 2019.
- [3] T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Scene context-aware rapidly-exploring random trees for global path planning," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2019.
- [4] R. Zhu, K. Nagahama, K. Takeshita, and K. Yamazaki, "Swept volumeの検索に基づくリーチング動作のオンライン生成," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [5] K. Nagahama, K. Takeshita, H. Yaguchi, K. Yamazaki, T. Yamamoto, and M. Inaba, "A learning method for a daily assistive robot for opening and closing doors based on simple instructions," in *IEEE 14th International Conference on Automation Science and Engineering*, 2018.
- [6] K. Nagahama, K. Takeshita, H. Yaguchi, K. Yamazaki, T. Yamamoto, and M. Inaba, "Estimating door shape and manipulation model for daily assistive robots based on the integration of visual and touch information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.
- [7] S. James, Z. Ma, D. R. Arrojo, and A. J. Davison, "Rlbench: The robot learning benchmark and learning environment," 2019.
- [8] J. Savage, M. Negrete, M. Matamoros, J. Cruz, R. Lagunas, J. Marquez, J. Marentes, and F. Villaseor, "The Role of Robotics Competitions for the Development of Service Robots," in *IJCAI, Workshop on "Autonomous Mobile Service Robots"*, 2016.
- [9] M. Matamoros, V. Seib, R. Memmesheimer, and D. Paulus, "Robocup@home: Summarizing achievements in over eleven years of competition," in *IEEE International Conference on Autonomous Robot Systems and Competitions*, 2018.
- [10] M. Matamoros, V. Seib, and D. Paulus, "Trends, challenges and adopted strategies in robocup@home," in *IEEE International Conference on Autonomous Robot Systems and Competitions*, 2019.
- [11] A. Szot, A. Clegg, E. Undersander, E. Wijmans, Y. Zhao, J. Turner, N. Maestre, M. Mukadam, D. S. Chaplot, O. Maksymets, A. Gokaslan, V. Vondrus, S. Dharur, F. Meier, W. Galuba, A. Chang, Z. Kira, V. Koltun, J. Malik, M. Savva, and D. Batra, "Habitat 2.0: Training home assistants to rearrange their habitat," *CoRR*, vol. abs/2106.14405, 2021.
- [12] Y. Kawasaki, A. Yoroze, and M. Takahashi, "Autonomous navigation using multimodal potential field to initiate interaction with multiple people," *IEEE/RSJ IROS*, 2018.
- [13] Y. Kawasaki, A. Yoroze, and M. Takahashi, "Hybrid a* アルゴリズムによる円滑な複数タスク実行のための経路計画," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [14] M. Negrete, J. Savage, and L. Contreras, "A Motion-Planning System for a Domestic Service Robot," *SPIIRAS Proceedings*, vol. 60, no. 5, pp. 5–38, 2018.
- [15] J. Hatori and K. Terada, "人間の生活空間で動くロボットのためのビジョン・言語処理技術," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [16] L. Contreras, H. Yokoyama, and H. Okada, "Visual feedback for active robot-object interaction," *The 36th Annual Conference of the Robotics Society of Japan*, 2018.
- [17] L. Contreras, H. Yokoyama, and H. Okada, "Multimodal feedback for active perception in service robots," *IEEE/RSJ IROS Workshops*, 2018.
- [18] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set," *IEEE Robotics Automation Magazine*, vol. 22, no. 3, pp. 36–52, 2015.
- [19] T. Yamamoto, H. Yaguchi, S. Kato, and H. Okada, "Evaluation of impression difference of a domestic mobile manipulator with autonomous and/or remote control in fetch-and-carry tasks," *Advanced Robotics*, vol. 34, no. 20, pp. 1291–1308, 2020.
- [20] Y. Abe, Y. Ishida, T. Ono, and H. Tamukou, "物体の3dスキャンによるホームサービスロボット向け学習データセット作製の高速化," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [21] N. Nakanishi and T. Nakayama, "一部が隠れているバラ積みオブジェクトの認識率の向上," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [22] S. Matsuzaki, Y. Miake, Y. Liu, K. Mano, L. Villamar-Gomez, C. Nakano, K. Ishihara, and J. Miura, "Robocup japan open 2019 @home dsplにおけるチームaisl-tutの取り組み," *The 37th Annual Conference of the Robotics Society of Japan*, 2019.
- [23] Y. Matsusaka, L. Contreras, H. Okada, Y. Iwanaga, and T. Yamamoto, "公平性と民主性を両立させた robocup 用競技シミュレーションシステムの開発," *The 39th Annual Conference of the Robotics Society of Japan*, 2021.
- [24] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, "Development of human support robot as the research platform of a domestic mobile manipulator," *ROBOMECH Journal*, vol. 6, 2019.
- [25] J. Pages, L. Marchionni, and F. Ferro, "Tiago: the modular robot that adapts to different research needs," *IEEE*, 2016.
- [26] J. Savage, D. Rosenbluth, M. Matamoros, M. Negrete, L. Contreras, J. Cruz, R. Martell, H. Estrada, and H. Okada, "Semantic reasoning in service robots using expert systems," *Robotics and Autonomous Systems*, vol. 114, 2019.