SOBITS 2024 Team Description Paper

Valentin Keith, Sim Jiahao, Kazuki Ito, Tachi Suzuki, Hirotaka Mie, Daiki Kaneko, Yuki Okuma, Ren Yamada, Yuki Mukogawa, Masashi Ogura, Hideyuki Omori, Hiroaki Kobayashi, Fumiya Ono, Koshiro Suzuki, Mai Morio, Ryota Tsutsui, Seito Suzuki, Daichi Adachi, Koei Tanabe, Hideo Shimomura, Yusaku Hanajima, Kazuaki Yuhashi, Yoko Toda, and Choi Yungwoon

Soka University of Japan, 1-236 Tangi-machi, Hachioji-shi, Tokyo 192-8577, Japan,
`https://home.soka.ac.jp/~choi/`

Abstract Breakthoughts on image and voice recognition bring with them expetations on robotics with the aim of enhancing daily work. SOBITS, with the same purpose, participates annually in local competitions such as RoboCup@Home JapanOpen and RoboCup@Home Asia-Pacific to showcase the potential of robotics in solving everyday tasks.
SOBIT PRO, an autonomous mobile manipulator, is the team's approach to tackle these tasks. Starting in 2019 with its development, accurate and stable mobility, manipulation and object recognition was achieved. In 2023, SOBIT PRO achieved outstanding score in the "GPSR" task, understanding commands and planning actions based on the order. This paper presents the technology and strategy to accomplish this task.

1 Introduction

SOBITS is a team which consists of master's and undergraduate students from the Choi Robotics Laboratory at Soka University in Japan. The team has been working on the development and research on robotics with the focus on supporting people in their daily life activities. Some of the areas encompass human-robot interaction for the stimulation in overcome a weakness, robot control in four-wheel independent steering, and robot perception in environment recognition and object manipulation.

Since 2012, SOBITS has participated in RoboCup@Home JapanOpen as competitors and technical committee members across various leagues (DSPL, OPL, S-OPL, Education). The team continuously addresses real-world challenges by upgrading their robot and skillset.

SOBIT PRO[1], the team's third developed robot, is a testament to the persistent learning journey. Earning first place in the JapanOpen OPL competition since 2021, SOBIT PRO has solidified the team's position as a leader in the field. This success, fueled by advancements in robot control and natural language understanding, paved the way for their international debut at RoboCup Bordeaux 2023. There, valuable insights were gleaned from global peers, highlighting the importance of fostering collaboration through open-source initiatives, further accelerating innovation in the field.

[1] `https://github.com/TeamSOBITS/sobit_pro`

This paper is organized as follows: Section 2 outlines the Background of the GPSR task and approach, Section 3 gives an overview of SOBIT PRO open-source hardware, Section 4 describes SOBIT PRO open-source software and applications, Section 5 provides an overall system evaluation based on 2023 results, and Section 6 discusses the conclusion and future research directions.

2 Background and SOBITS approach

In the context of the RoboCup@Home league, the General Purpose Service Robot (GPSR) task plays a pivotal role in facilitating the integration of robots into domestic environments. The necessity for GPSR arises from the demand for robots capable of interacting naturally with humans.

To enhance human-robot interaction and enable a more natural communication, the team's approach leverages a Speech Interaction System to comprehend tasks requested by humans. To address the multifaceted requirements of the GPSR task, an intelligent system encompassing object recognition, manipulation, autonomous navigation, and a human following system was developed.

The development of the SOBIT PRO focused into accomplish the objectives of the GPSR task. This robot serves as the foundational platform upon which this intelligent system is built, integrating advanced functionalities essential to success on the task execution. Through meticulous planning and a systematic approach, the main objective is to optimize the performance of SOBIT PRO when navigating diverse domestic environments, recognizing objects autonomously, and executing manipulation tasks with precision.

3 SOBIT PRO Open-Source Hardware

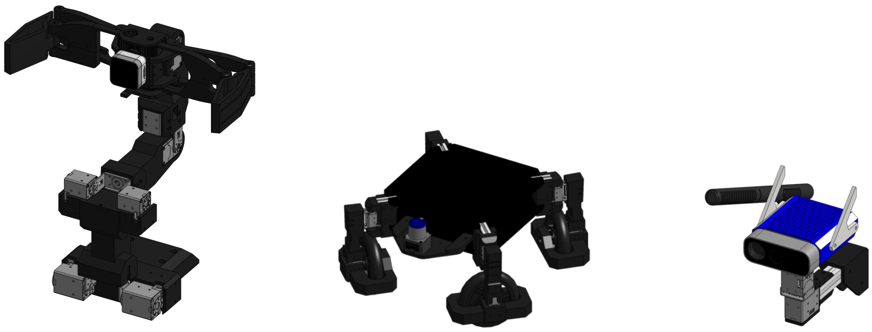
3.1 Robot Configuration

In this section, the configuration of the "SOBIT PRO" developed by SOBITS is introduced. The main objective of this robot is to maneuver through a house environment while avoiding obstacles and to grasp and transport objects to their designated locations. Additionally, robot smoothness and accurate was tested to prevent any abrupt movements that may pose a threat to both the robot structure and its environment.

3.2 Manipulator

The robot's manipulator can move in a range from 20 [cm] to 130 [cm] in height, allowing it to reach objects from the ground and high shelves. The manipulator has 5DoF, as shown in Figure 1a, and can hold objects up to 0.5 [kg].

The end of the manipulator is equipped with a hand that can grip objects up to 20 [cm] wide. A sponge rubber has been attached to the hand's surface, allowing it to grip objects of various shapes with a high coefficient of friction, ensuring the object remains securely grasped without damage.



(a) 5DoF Manipulator (b) Mobile Base (c) Pan-Tilt Structure

Figure 1: SOBIT PRO structure

3.3 Mobile Base

The movement of the robot, as shown in Figure 1b, is controlled by its 4 wheels, each of which is equipped with an independent steering mechanism. This allows the robot to move in any direction. Two actuators are used, one in the rotation axis and another in the swivel axis of each wheel. While the swivel angle of the robot is between -90 [deg] and +90 [deg], each wheel can rotate in multiple directions by rotating the actuators on these swivel axes. The rotation axis and the swivel axis are positioned so that all wheels can head in the same direction and set the same velocity vector when facing towards the destination.

3.4 Pan-Tilt Structure

The pan-tilt structure, as shown in Figure 1c, allows the camera to increase its range of vision without moving the base of the robot. The pan and tilt functions pivot the camera horizontally (pan) and vertically (tilt). The pan and tilt are used in the SOBIT PRO to enhance the robot's imaging capabilities.

4 SOBIT PRO Open-Source Software

4.1 General system for GPSR

This section describes the strategy for GPSR, illustrating how the robot interacts with humans through natural language in a domestic environment as shown in Figure 2 . SOBIT PRO is equipped with subsystems such as the Speech Interaction System, Autonomous Navigation, Human Following System, and Image Processing and Manipulation System.

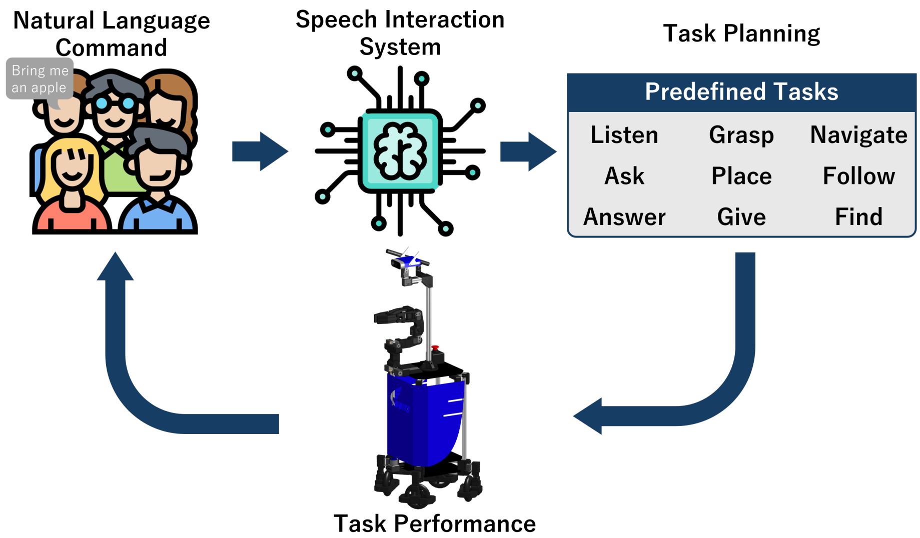


Figure 2: GPSR task flowchart

4.2 Speech Interaction System

The Speech Interaction System[1] consists of five important stages: speaker localization, sound source separation, speech recognition, command understanding and speech synthesis, as shown in Figure 3.

The speaker localization is estimated by combining the sound source coordinate by the triangulation principle from the two microphone arrays, and the person coordinate acquired by the RGB-D sensor. The sound source separation removes the noise from outside the speaker direction by using a fixed beamforming method to emphasise the speech. For speech recognition, Whisper[2] is used to convert the speech waveform into text.

Then, a novel Attention-based Seq2Seq (Sequence to Sequence) model[3] is employed in command understanding. This model encodes the input sentences into symbols, decodes the relationships between words and outputs 19 words in a specific order, such as ”task”, ”object” and ”destination”. This allows the robot to understand commands without using logical expressions. Additionally, an embedding process represents the meaning of words as vectors, with pre-trained vector data to handle not previously trained words. Finally, for speech synthesis, the open-source module Pico Text-to-Speech is used to convert text into speech waveforms and output speech files for human interaction.

4.3 Autonomous Navigation

The robot navigates autonomously using a sophisticated system adept at avoiding both moving obstacles and collisions within narrow passages. Inspired by

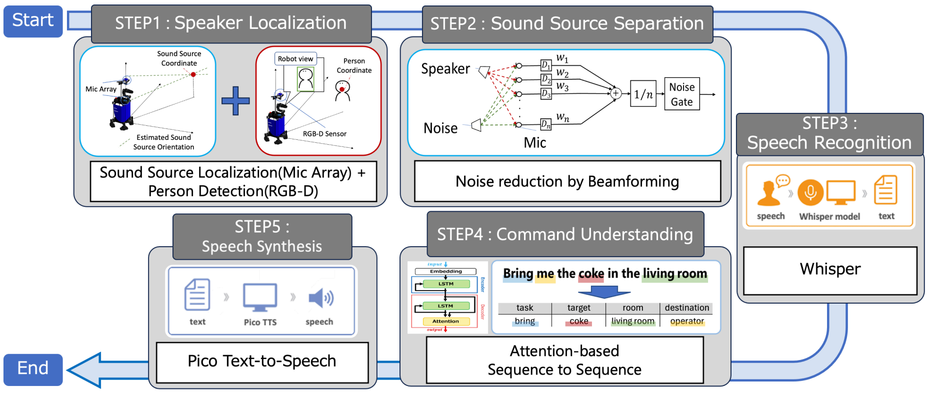


Figure 3: Speech Interaction System

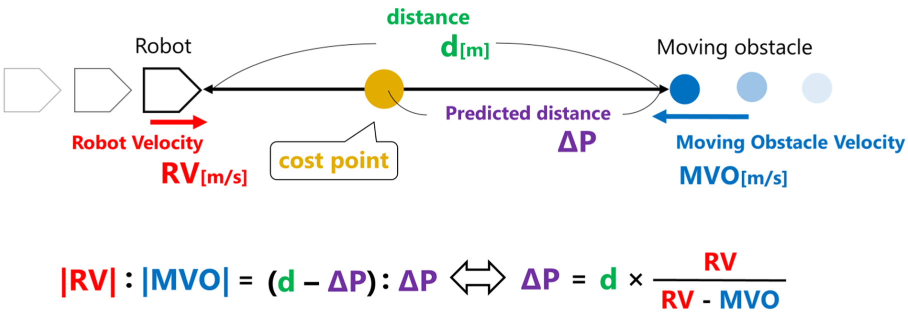
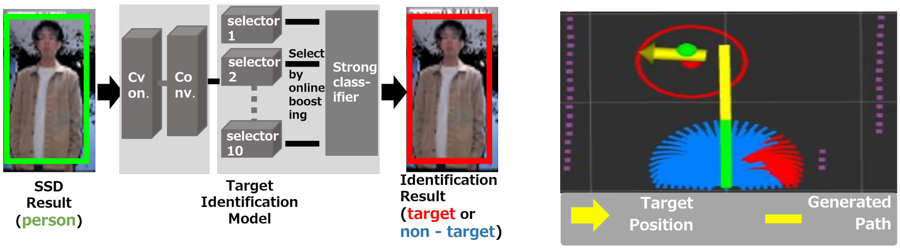


Figure 4: Moving Obstacle Avoidance System

how humans intuitively navigate in crowded spaces by predicting others' movements, the system estimates pedestrian trajectories and covers them into path planning costs. As depicted in Figure 4, the system predicts potential collision points based on the robot's current speed, the obstacle's speed, and their relative distance, assuming constant velocities. Utilizing this dynamic cost map, the robot charts a collision-free path, ensuring navigation alongside moving people.

Furthermore, the system tackles the challenge of navigating narrow passages like doorways, where turning could lead to wall collisions. First, a Narrow Passage Detection establishes a search range based on the robot's width around each planned path. Then, the entrance and exit of such passages are identified as potential pinch points. Second, a Waypoint Insertion is conducted so that the robot move through the centre of the passage, effectively avoiding wall collisions.



(a) Person Detection and Target Identification (b) Motion Control

Figure 5: Person Following Robot

4.4 Human Following System, SOBIT Follower

The human following system[4], *SOBIT Follower*[2], consists of three function as shown in Figure 5: person detection, target identification, and motion control.

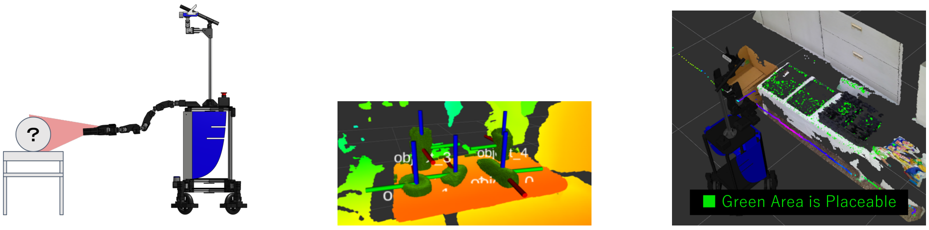
This system allows the robot to follow the target person by combining three functions using a RGB-D sensor and a 2D LiDAR. The person detection utilizes Single Shot MultiBox Detector(SSD)[5] with a RGB-D sensor and Distance Robust SPatial Attention and Autoregressive Model(DR-SPAAM)[6] with a 2D LiDAR to detect people. The target identification uses an online target identification method[7] to learn and identify the target in real-time. For motion control, the system integrates a movement control method using a virtual spring model[8] and a path generation method using the Dynamic Window Approach (DWA) to control the target in accordance with the target human's movements.

4.5 Image Processing and Manipulation System

The image processing system consists of two parts: object recognition and object position detection. For object recognition, *You Only Look Once (YOLO)[9] algorithm*[3]is used to estimate the region of an object in a 2D RGB image and classify it into a pre-learned category. To detect the object's position, Point Cloud Library (PCL) converts depth camera data into a point cloud and extract the centre point of the detected object through euclidean clustering. To increase the system's adaptability to different environments, brightness, rotation and scaling is applied to the pre-trained images using OpenCV. Figure 6 represents how the category of the object is matched with 3D coordinates using PCL.

The object manipulation system consists of Standing Position Estimation, Grasping, and Settlement Position Estimation as shown in Figure 6c. Firstly, the Standing Position Estimation enables the robot to move to the proper position in front of the object to be grasped while avoiding obstacles in its surroundings. Next, in the grasping process, the Inverse Kinematics of the robot arm are calculated when given the coordinates of the object. Finally, once the object is

---
[2] `https://github.com/TeamSOBITS/sobit_follower`
[3] `https://github.com/TeamSOBITS/yolov5_ros`



(a) Object Recognition (b) Object Detection (c) Placeable Position

Figure 6: Image Processing

grasped and transported to the target location, the Settlement Position Estimation detects planar regions obtained from PCL and calculates the placeable area where to place the object down while avoiding collisions with other objects.

5 Overall System Evaluation

This section analyzes the performance of this system with SOBIT PRO in the *GPSR task*[4]at the RoboCup@Home JapanOpen 2023. In this task robots respond to a sequence of natural language commands involving object manipulation, navigation, and information retrieval.

In the competition, the team was presented the following two scenarios:

- *Go to the living room, find Sam at the bin b and guide him to the chair b*
- *Go to the dining room, find Hunter at the chair b and guide her to the bin a*

In the first trial, the system successfully understood the command and located Sam at bin b, demonstrating effective speech interaction and action planning. However, inaccurate robot localization hampered guiding Sam to chair b, resulting in a score of 27 points.

The second trial saw flawless execution, with the robot understanding the command and completing the actions sequentially. This earned a score of 34 points, bringing the total to 61 points.

By achieving the highest GPSR score, SOBITS secured first place at the RoboCup@Home JapanOpen 2023 OPL. This local success, showcasing the system's capabilities in natural language understanding and task execution, instills confidence for the upcoming RoboCup Eindhoven 2024.

6 Discussion and Conclusion

This paper provided a detailed account of the comprehensive design and development journey of SOBIT PRO encompassing both hardware and software

[4] `https://github.com/RoboCupAtHomeJP/Rule2023/`

aspects. Notably, team's hardware innovation involves the implementation of a four-independent steering mechanism, enabling a omnidirectional movement. On the software front, open-source projects was the strategy taken by the team. Through the utilization of already published open-source projects, development process was not only streamlined, but also, allowed the team to contribute to the community through *SOBITS official GitHub organization*[5].

Through collaborative endeavours, the team aims to foster a culture of innovation, knowledge exchange, and skill development within the robotics field.

Taking into account the outlined objectives, the team is committed to sustaining contributions to the robotics community through open-source initiatives while further refining SOBIT PRO for the RoboCup Eindhoven 2024.

References

1. Hirotaka Mie and Yongwoon Choi. Development of a command sentence understanding system using speaker localization in noisy environments for task execution by life support robots. In *ISSJ2023*, 2023.
2. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision, 2022.
3. Tsurue Takumi and Choi Yongwoon. Instruction understanding using attention-based seq2seq in robot task execution. *Proceedings of the National Conference of the Japanese Society for Artificial Intelligence*, JSAI2021, 2021.
4. Taichi Suzuki and Yongwoon Choi. Development of a target identification function for a person-following robot using a 2d lidar and a rgb-d sensor on a pan/tilt/rotate mechanism. In *SI2023*, pages 2664–2667, 2023.
5. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. *SSD: Single Shot MultiBox Detector*, page 21–37. Springer International Publishing, 2016.
6. Dan Jia, Alexander Hermans, and Bastian Leibe. Dr-spaam: A spatial-attention and auto-regressive model for person detection in 2d range data, 2020.
7. Kenji Koide, Jun Miura, and Emanuele Menegatti. Monocular person tracking and identification with on-line deep feature selection for person following robots. *Robotics and Autonomous Systems*, 124:103348, 2020.
8. K. Morioka, Joo-Ho Lee, and H. Hashimoto. Human-following mobile robot in a distributed intelligent sensor network. *IEEE Transactions on Industrial Electronics*, 51(1):229–237, 2004.
9. Ultralytics. YOLOv5: A state-of-the-art real-time object detection system. `https://docs.ultralytics.com`, 2021. Accessed: 02-02-2024.

[5] `https://github.com/TeamSOBITS`

Robot SOBIT PRO Hardware Description

SOBIT PRO is a life-support autonomous robot developed by SOBITS (Choi Laboratory). Hardware Specifications are as follows:

- Base: 4-steering mobile base, vel: 0.7[m/s].
- Torso: keeps the main computer safe inside.
- Manipulator: Mounted on torso. 5DoF and parallel jaw gripper. Payload: 0.5[kg].
- Head: Mounted on a 0.5[m]-long pole. Pan-tilt structure for wide camera vision range.
- Dimensions: height: 1.5[m], width: 0.5[m], depth: 0.5[m].
- Robot weight: 10[kg].

Also SOBIT PRO incorporates the following devices:

- 2× Makita BL1860B 18[V] 6.0[Ah] Battery.
- LiDAR Hokuyo UST-20LX.
- Azure Kinect (Placed on the pan-tilt structure).
- Realsense D405 (Placed on the manipulator).
- 12× Dynamixel XM430-W350-R (Robot arm and mobile base).
- 2× Dynamixel XM540-W270-R (Robot arm).
- 4× Dynamixel XM430-W210-R (Mobile base).

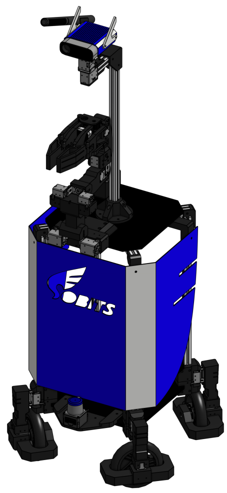


Figure 7: SOBIT PRO

Robot's Software Description

The following software is used:

- Platform: Robot Operating System (ROS).
- Navigation: A combination of AMCL, move_base and DWA.
- Face recognition: Single Shot Multibox Detector.
- Speech recognition: Whisper.
- Command recognition: Attention-based Seq2Seq Model.
- Speech generation: Pico Text-to-Speech.
- Object recognition: YOLO and PCL.
- Actuator control: Dynamixel SDK Position Controller.

External Devices

SOBIT PRO relies on the following external hardware:

- Speaker.
- Unidirectional microphone and microphone array.
- DELL G15 laptop.