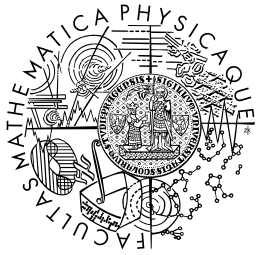

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Lada Kudláčková

Manipulating Objects through Deictic Gesture Recognition

Department of Theoretical Computer Science and Mathematical Logic

Supervisor of the master thesis: RNDr. David Obdržálek, Ph.D.

Study programme: Computer Science

Prague 2024

I declare that I carried out this master thesis on my own, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

Dedication.

Title: Manipulating Objects through Deictic Gesture Recognition

Author: Lada Kudláčková

Department: Department of Theoretical Computer Science and Mathematical Logic

Supervisor: RNDr. David Obdržálek, Ph.D., Department of Theoretical Computer Science and Mathematical Logic

Abstract: Use the most precise, shortest sentences that state what problem the thesis addresses, how it is approached, pinpoint the exact result achieved, and describe the applications and significance of the results. Highlight anything novel that was discovered or improved by the thesis. Maximum length is 200 words, but try to fit into 120. Abstracts are often used for deciding if a reviewer will be suitable for the thesis; a well-written abstract thus increases the probability of getting a reviewer who will like the thesis.

Keywords: gesture recognition, object manipulation, autonomous control

Název práce: Manipulace s objekty pomocí rozpoznávání ukazovacích gest

Autor: Lada Kudláčková

Katedra: Katedra teoretické informatiky a matematické logiky

Vedoucí bakalářské práce: RNDr. David Obdržálek, Ph.D., Katedra teoretické informatiky a matematické logiky

Abstrakt: Abstrakt práce přeložte také do češtiny.

Klíčová slova: rozpoznání gest, manipulace s objekty, autonomní řízení

Contents

Introduction

Opening statement: introducing the research field, stating the problem.
My motivation, goals and research limitation.
Overview of the thesis structure.

1 Task Analysis

1.1 Theoretical Background

1.1.1 Basic Definitions

Definition of key words: gesture recognition, object manipulation, autonomous control

1.1.2 Human-Robot Interaction (HRI)

brief description of HRI;

remote vs. proximate interactions;

roles of humans and robots in interaction: Supervisor, Operator, Mechanic, Peer, Bystander, Information consumer, Mentor (taxonomy from paper: M. A. X. Goodrich and A. C. Schultz, "Human-Robot Interaction: A Survey," Foundations and Trends R© in Human- Computer Interaction, vol. 1, no. 3, pp. 203–275, 2007.);

areas of application: industrial, search and rescue, medical, social, ...

1.2 Task description

1.2.1 Pick and Place Task

Performing 'Pick and Place' using pointing gestures:

Pick:

- Determine a object that was selected with a pointing gesture
- Navigate close to the object
- Identify the object and compute its exact coordinates
- Pick the object

Place:

- Determine a target location from a pointing gesture
- Navigate close to the location
- Place the object to the location

1.2.2 Task Specification

Requirements and restrictions:
gesture recognition based on image processing (available sensors - depth cameras), proximate robot control by single user, no interaction with other robots or humans, static indoor environment (robotic lab), safe manipulation with objects, safe navigation (obstacle avoidance without unnecessary emergency braking), ...

1.3 Goals

1.3.1 Implementation of Mobile Manipulator

Design and implement a mobile manipulator that performs 'Pick and Place' tasks according to the given requirements;

1.3.2 Comparison of deictic gestures types

Metric: the distance between the correct coordinates (of the selected object or location) and the intersection of the pointing ray and the floor.

Experiment with different ways of using deictic gestures:
- a pointing ray calculated from a pair of skeleton coordinates (head - hand, elbow - wrist, shoulder - wrist)
- pointing with or without visual feedback (pointed ray shown in rViz)

2 The state of the art

2.1 History of Gesture Recognition

Summary of gesture recognition techniques; historical development of sensors; ...

2.2 Localization and navigation with deictic gestures

These are some (not all) examples of what I want to mention here:

Deictic gestures for multi-robot systems

Paper:
B. Gromov, L. M. Gambardella and G. A. Di Caro, "Wearable multi-modal interface for human multi-robot interaction," 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Lausanne, Switzerland, 2016, pp. 240-245, doi: 10.1109/SSRR.2016.7784305.

Use of the pointing gesture for localization

Paper:
B. Gromov, L. Gambardella, and A. Giusti. Robot Identification and Localization with Pointing Gestures. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 3921–3928 https://people.idsia.ch/ gromov/repository/gromov2018robot.pdf

3D Motion planning with pointing gestures

Paper:
B. Gromov, J. Guzzi, L. Gambardella, and A. Giusti. Intuitive 3D Control of a Quadrotor in User Proximity with Pointing Gestures. IEEE International Conference on Robotics and Automation (ICRA), 2020 https://people.idsia.ch/ gromov/repository/gromov2020intuitive.pdf

2.3 Interpretation of gestures

Papers:

Chaudhary, A (2018). Robust Hand Gesture Recognition for Robotic Hand Control. Springer. ISBN 978-981-10-4798-5 https://doi.org/10.1007/978-981-10-4798-5

Alikhani, M., Khalid, B., Shome, R., Mitash, C., Bekris, K.E., Stone, M. (2020). That and There: Judging the Intent of Pointing Actions with Robotic

Arms. AAAI.https://ojs.aaai.org//index.php/AAAI/article/view/6601

2.4 Object detection with pointing gestures and speech recognition

Li-Heng Lin, Yuchen Cui, Yilun Hao, Fei Xia, Dorsa Sadigh (2023). Gesture-Informed Robot Assistance via Foundation Models. https://arxiv.org/abs/2309.02721

A. Ekrekli, A. Angleraud, G. Sharma, R. Pieters (2023). Co-speech gestures for human-robot collaboration. https://arxiv.org/abs/2311.18285

3 Gesture based robot control

3.1 Pointing gesture

3.1.1 Types

Different types:
a pointing ray calculated from a pair of skeleton coordinates (head - hand, elbow - wrist, shoulder - wrist, ...)

3.1.2 Pointing ray

Calculation - intersection with the floor plane.

3.2 Confirming gesture

Some other gesture is required to confirm the pointing gesture.
The 'Hand Grip' gesture is not well recognized by the Astra camera, so I choose 'Hand Raise'.

3.3 Algorithm overview

How to select the object and the target location.

4 Design of Robotic System

4.1 Robot Model

4.1.1 Mobile Robot Manipulator

Overview of main components: mobile base, robotic arm with gripper.

4.1.2 Sensors

Depth camera for body tracking and objects detection.
LIDAR for navigation and objects detection.

4.1.3 Separated Vision System

I decided to use a vision system that is separate from the mobile robotic manipulator due to the limited distance for body tracking and object detection with the ORBBEC Astra camera;
otherwise objects would only be detected at a small distance from the robot.

4.2 Hardware

4.2.1 Mobile vehicle

Neobotix MP-500

Basic description of HW...

Laser Scanner SICK

4.2.2 Manipulator

Robotic arm UR5

Gripper

4.2.3 Visual system

ORBBEC Astra camera

4.2.4 Computers and network

Lenovo (connected via Ethernet to Neobotix).
ACER (connected to camera).

WiFi connection between computers.

4.3 System Software

Ubuntu 20.04.
Robot Operating System, Noetic.

4.4 Control by gestures

Pointing gestures to select the object and target location.
Confirmation gesture, ...
Brief overview of the whole 'Pick and Place' process .

5 Implementation

5.1 Vision system

5.1.1 ORBBEC Astra camera

Installation

Why I choose ORBBEC Astra camera over Kinect ONE (v2):
difficult installation of tools and libraries for a ROS Interface to the Kinect One (dependencies on ROS Hydro/Indigo distribution, no available packages for ROS Noetic).

Installation of ORBBEC SDK for Linux and dependencies (OpenNI2, libsfml-dev, ...).

Package ros_astra_camera:
https://github.com/orbbec/ros_astra_camera
OpenNI2 ROS wrapper for Orbbec 3D cameras.

Separation from mobile robot manipulator system

Limited distance for body tracking and correct object detection with the ORBBEC Astra camera (details).

Connection to mobile robot manipulator via SSH.

5.1.2 Object detection

Package pcl_object_detection:
https://github.com/shinselrobots/pcl_object_detection
ROS node for detecting objects on a flat surface, using Point Cloud Library (with the ros_astra_camera package).

Modification of this package:
added limitation of size and position of objects (detection frame) to avoid false detection;
added custom messages, rViz markers, ...;

Subscriber of ros_astra_camera topic (point clouds) and selected object topic. Publisher of detected objects topics (coordinates, ...) and point cloud topics (clusters, planes, ...) for rViz.

5.1.3 Gesture detection

Packages: astra_body_tracker:
https://github.com/shinselrobots/astra_body_tracker
Publisher of ROS topic for body tracking information (from the ORBBEC SDK).

pointing_gesture:
modified astra_body_tracker package to get skeleton data;
added code to detect gestures, rViz markers, ...

Publisher of pointing_gesture topic (as geometry_msgs).

5.1.4 Stream switching

Problem: skeleton data was not provided in the ORBBEC SDK (without license), I needed to switch between data streams (using custom ROS messages): launch ros_astra_camera driver and pcl_object_detection node, when the object detection is complete, stop the stream and run the ORBBEC SDK with pointing_gesture package to get body tracking.

5.1.5 Object selection and target location

task_control_node (will be renamed):
Subscriber to object detection and pointing gesture topics;
provides calculations of pointing ray intersection and selection of object.
Sends data to mobile manipulator PC over SSH (coordinates of objects and target location, info about selected object).

5.2 Navigation of Autonomous Vehicle

5.2.1 Installation

Neobotix:
Packages: https://github.com/neobotix/
ros-noetic-amcl, ros-noetic-map-server, ros-noetic-move-base, ...

5.2.2 Map of Environment

Mapping procedure, selecting the map for navigation, visualization with RViz...

5.2.3 Navigation to Goal

Goal definition, movement (path, obstacle avoidance, ...).

5.3 Object Manipulation

5.3.1 Installation

Universal Robots:

Packages:
Universal_Robots_ROS_Driver https://github.com/UniversalRobots/Universal_Robots_ROS_

Universal_Robots_Client_Library
https://github.com/UniversalRobots/Universal_Robots_Client_Library

ur5_moveit_config
https://github.com/ros-industrial/universal_robot/tree/noetic-devel/ur5_moveit_config

5.3.2 Mobile Manipulator URDF

URDF for Neobotix, UR5 and gripper.

5.3.3 MoveIt Setup Assistant

How to create config and set up arm positions.
How to set up arm limits.
Simulation in rViz.

5.3.4 Code

ur_robot_driver;
ROS.urp;
move_it_planning;
trajectory commands;

5.3.5 Objects coordinates

approximate coordinates of objects obtained from the vision system;
robot navigates to objects;
exact objects coordinates from LIDAR (lidar_scan topic subscriber).

6 Experiments

Experiments descriptions:

Experiments with different ways of using deictic gestures:

- a pointing ray calculated from a pair of skeleton coordinates (head - hand, elbow - wrist, shoulder - wrist)
- pointing with or without visual feedback (pointed ray shown in rViz)

Experiments measurements:

...

result evaluation; what went wrong; future work, possible improvements

Conclusion

Results of experiments - summary.
Which gestures are well recognised by Astra camera;
most accurate pointing gestures - compare results with related work.

Suggestions for improvement.

7 Appendix

List of Figures

List of Tables

List of Abbreviations

A Attachments

A.1 First Attachment