Roomba Path-Finding Data Curation and Preperation

James Shastid, Bhumir Patel, David Wilkey, Joshua Howard

I. MOTIVATION

As technology continues to improve, the demand for innovative products to assist with routine household tasks is steadily increasing. Robotic vacuum cleaners are one of the most common products within this category today, with an estimated 14% of American households owning this type of device [1]. Despite their increasing accessibility and demand, many robotic vacuum models are plagued by various performance issues that can potentially hinder this growing industry. Perhaps the most common complaint amongst owners of these devices is navigational challenges due to obstacles such as electrical cords, floor thresholds, and toys/small objects [2]. To address this critical issue, our project will focus on the collection and analysis of navigational data to identify insights and trends that could then be used to improve overall device performance and consumer satisfaction.

II. DATA COLLECTION

There are three ways we intend to collect data from the Roomba.

A. Custom Hardware

The initial approach involves utilizing custom hardware consisting of an Arduino and an accelerometer to wirelessly gather data concerning the Roomba's relative position within the environment. A transceiver with a fifty-meter range will be employed to facilitate wireless communication with a secondary Arduino and to record the positional data. This data will subsequently be transformed into a grid-based system to streamline processing. Given that the Roomba is intended to operate exclusively on a single level, a two-dimensional grid will suffice. Additional hardware components may be incorporated as necessary to acquire more extensive and higher-quality data. These Improvements could include sonar to detect walls, or an upgraded accelerometer to collect more axes of data. This all will depend on the difficulties encountered during the development and testing.

B. Video Footage

The second potential approach involves employing one or more cameras to physically monitor the Roomba within the real environment. This method would enhance our understanding of the actual surroundings encountered by the Roomba. However, it presents certain disadvantages. It necessitates substantial computational resources and significantly increases the time required. Additionally, it would involve prolonged camera recording sessions. Due to these constraints, this approach remains conceptual and may not be realized.

C. Outsourced Data

The third source of data derives from the application of digital archaeology, involving an in-depth scan of various database sources to identify publicly accessible data, which is then integrated with our own data prior to the pre-processing phase

III. TASK ALLOCATION

Our 4-person team structure allocates responsibilities across hardware, software, testing, and analysis domains. All team members will participate in weekly progress meetings, contribute to the final presentation, and review all project deliverables.

A. Member 1 - Hardware Integration Lead:

Design and implement sensor mounting solution, wire Arduino board with accelerometer and transceivers, test wireless communication reliability, troubleshoot hardware issues during testing

B. Member 2 - Software & Data Acquisition Lead:

Develop Arduino firmware for sensor data collection, implement wireless data reception and logging system, create real-time monitoring dashboard, ensure data synchronization and quality

C. Member 3 - Testing & Experimentation Lead:

Design obstacle course configurations, conduct all Roomba test runs, maintain testing logs, capture validation video footage, ensure experimental protocol consistency

D. Member 4 - Analysis & Documentation Lead:

Process and clean sensor data, perform statistical analysis of collision events and navigation patterns, create visualizations, compile customer review data, and lead report writing

IV. PRIMARY DELIVERABLES

A. Instrumented Roomba System

A functional prototype demonstrating integration of external sensors with a commercial robotic vacuum, including documented hardware design, Arduino code, and a wireless data acquisition system suitable for replication in future projects.

B. Navigational Data Repository

A comprehensive dataset containing time-series accelerometer readings, gyroscope measurements, and annotated collision events from multiple test runs across diverse obstacle scenarios. Data will be provided in CSV format with accompanying metadata describing test conditions.

C. Technical Analysis Report

A detailed report presenting quantitative analysis of Roomba navigation performance, including collision frequency and severity by obstacle type, acceleration signatures characteristic of different failure modes, and statistical comparison of performance across floor surfaces and obstacle densities. The report will feature data visualizations, including time-series plots of collision events, heat maps of high-impact zones, and comparative bar charts.

D. Design Recommendations

Evidence-based suggestions for navigation algorithm improvements based on observed sensor patterns, such as threshold adjustments for obstacle detection, recommendations for sensor placement optimization, and strategies for reducing repeat collisions with persistent obstacles.

V. Timeline

A. October 17

The initial hardware has been completely assembled, and the firmware has been fully developed to facilitate the commencement of the initial data collection phase.

B. November 7

Data has been collected in four distinct environments with various obstacle configurations. Following this milestone, we will discuss the potential utilization of alternative data collection methods, the consideration of more advanced hardware, and the exploration of other potential trajectories for the project's development.

C. November 21

Complete all data collection activities. Initiate the preprocessing and data preparation stages. Additionally, we will commence the examination of the collected data and, if time allows, implement a reinforcement learning model to identify more effective navigation strategies for the Roomba.

D. Before Presentation

Ensure that all data are comprehensively prepared, visualizations are developed to demonstrate the value of the collected data, a presentation is prepared, and, ideally, a developed model trained on the data that can perform improved pathfinding.

VI. Expected Outcomes

We anticipate compiling a dataset of grid-based navigation information gathered from Roombas operating in multiple rooms with diverse obstacle configurations. Additionally, detailed room information will be documented alongside the data in formats conducive to analysis. We also plan to include visual representations illustrating the inefficiencies in the current Roomba pathfinding algorithm, as well as other relevant metrics collected throughout the project. Ultimately, our goal is to develop a trained model using this data that will demonstrate superior pathfinding capabilities compared to the existing Roomba model.

References

[1] B. Elad, "Smart Robot Vacuum Cleaners Statistics By Total Manufacturing Unit, Market Size and Facts," *Coolest Gadgets*, Feb. 19, 2025. [Online]. Available: https://coolest-gadgets.com/smart-robot-vacuum-cleaners-statistics/ [Accessed: Oct. 2, 2025].

[2] F. G., "7 Major Drawbacks of Robot Vacuums You Need to Know," *Family Goody*, Apr. 23, 2025. [Online]. Available: https://familygoody.com/what-are-the-disadvantages-of-robot-vacuum-cleaner/ [Accessed: Oct. 2, 2025].