Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 4, 2023
A Regional Conference of the Society for Industrial and Systems Engineering

Evaluating Virtual Simulation for Autonomous Ground Vehicles

Parker Adams, Adam Bakre, Taylor Kramer, Julia Riekena, and Brandon Thompson

Department of Systems Engineering
United States Military Academy
West Point, NY 10996

Corresponding author's Email: parker27adams@gmail.com

Author Note: The Capstone Team would like to thank Mr. James Richards and the Engineer Research and Development Center Information Technology Laboratory (ITL), Institute for Systems Engineering Research (ISER), Geotechnical and Structures Laboratory (GSL) and the Mobility Systems Branch (MSB) for giving the team the opportunity to be a part of their efforts in Autonomous Ground Vehicle (AGV) research and development. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: This effort provides a potential methodology through which to assess simulation of the United States Army's Autonomous Ground Vehicle platforms. A literature review of autonomous vehicle simulation and LiDAR sensors provided metrics through which to evaluate autonomous vehicle performance in both the physical and virtual environments. A requirements traceability matrix was developed, and evaluation metrics were derived for virtual testing of autonomous vehicles from existing literature and U.S. Army doctrine. Statistical analyses showed similar behavior between the physical and virtual test results. However, further evaluation of data, including cost maps, identified discrepancies between the physical and virtual test environments that demonstrates further refinement of the autonomous vehicle simulation is needed for assessment.

Keywords: Autonomous Ground Vehicle, Simulation, Virtual Testing

1. Introduction

Autonomous Ground Vehicles (AGVs) are a critical element of the United States Army's force modernization efforts. AGVs are intended to be utilized as a force enabler, allowing soldiers to extend their sphere of influence on the battlefield without putting themselves directly in harm's way. Efforts to develop AGVs and their capabilities come from across the U.S. Army, to include Army Futures Command, the Next Generation Combat Vehicle Center (NGCVC), and the Engineering Research and Development Center (ERDC). The Army requires the ability to test the capabilities of autonomous vehicles on various types of terrain, in all weather conditions, and under multiple scenarios and mission sets. However, deliberate physical testing of AGVs is time consuming, costly, and limited to specific terrain and environmental conditions at test sites. As a result, a validated virtual test environment is needed to provide a more cost-effective and robust method of evaluating autonomous ground vehicles. ERDC is tasked with developing a simulation that accurately replicates the physical environment and provides the same sensor response to environmental characteristics and terrain features to allow the current autonomous vehicle control system, computer, and navigation software to maneuver within the virtual environment. This paper describes the researchers' efforts to develop a methodology through which to assess whether the simulation accurately replicates the physical environment and autonomous vehicle behavior.

2. Background

2.1 ERDC's Simulation Effort

One of ERDC's role in AGV development is to build a virtual simulation that can accurately replicate vehicle behavior, autonomous path planning, and physical performance. Testing through simulation is a much more cost-effective method to train and test autonomous vehicles, such that developers can contribute remotely, there is reduced risk with regards to the vehicle safety, and variables such as terrain difficulty can be varied much faster and tested in larger quantities. ERDC's ITL developed a virtual environment based on an AGV physical test site. This was done using the Virtual Autonomous Navigation Environment (VANE) and the VANE Integrated Ground Operational Robotics (VIGOR). VANE replicates the physical environment with data from Light Detection and Ranging (LiDAR) scanners, while VIGOR combines mixed-fidelity commercial-off-the-shelf tools with ERDC's vehicle-terrain interaction models and expertise to recreate platform behavior that allows the AGV to operate within the VANE simulation.

2.2 Previous Testing and Efforts

Physical testing began with a representative ground vehicle that possesses the same desired maneuverability characteristics, sensors, and autonomous navigation software as the class of AGVs being evaluated. The vehicle was operated around a set of courses measuring 1-to-2 kilometers that encompassed a large swath of various types of terrain, ranging from paved roads with no vegetation to steep, forested hills. The same procedure for vehicle testing was run by ERDC in their simulation at the ERDC-ITL. Similar experimental runs were conducted in 2022. A previous research effort was tasked with analyzing this data to assess ERDC's simulation and concluded that operating the vehicle on long routes with diverse terrain introduced too much variation to confirm that the AGV's virtual performance was similar to physical tests (Cox, DeRosier, Howell, Jones, & Thompson, 2022). This made it difficult to pinpoint sources of variation and identify differences between the two testing environments. In addition, the researchers found that the experimentation was unable to provide sufficiently accurate data to produce a large enough sample size to conduct cost map analysis (Cox et al., 2022). Two updates were made to the experimental design for testing AGVs in 2023. First, the length of test runs was reduced in a concept called Operational Design Domains (ODDs). ODDs test paths range from 50-to-300 meters along distinct surface types (paved, dirt, sand, or gravel), vegetation densities (low, medium, or high), and inclines (no incline, medium incline, or steep incline). This transition attempts to isolate sources of variability that were negatively impacting analysis and validation efforts. The desired outcome of the change is to assess vehicle performance, both physical and virtual, on lower complexity ODDs first and progressively increase route difficulty. Should researchers determine that the simulation accurately emulates the physical test site, additionally complex variables such as weather could be incorporated to test how the AGV responds. Second, the representative ground vehicle that the AGV's software and hardware were integrated into was changed to a different vehicle platform. This change may significantly influence testing and assessment efforts, as the previously used vehicle was smaller and more maneuverable.

2.3 Literature Review

LiDAR, which stands for Light Detection and Ranging, is a primary method that autonomous vehicles use to scan the environment and identify obstacles. It works by emitting lasers in all directions and measuring the distance to obstacles by the reflected lasers. LiDAR technology is capable of detecting the type of obstacle by collecting the intensity of the laser (McManamon, 2019). This helps to differentiate between fatal obstacles, defined as obstacles the vehicles cannot navigate through, and non-fatal obstacles, such as bushes, that vehicles can penetrate (Deems, 2017). However, relying solely on LiDAR has its limitations. For instance, LiDAR is unable to distinguish what lies beyond an obstacle making it difficult to identify if the vehicle could breach the obstacle (Wang, et al., 2019). As a result, the simulation developed needs to be robust enough to incorporate additional sensors in future testing.

One way to visualize how LiDAR perceives an environment is through the production of cost maps. Cost maps are arrays of data that store information. In the case of robotics and autonomous navigation, cost maps typically store data about the vehicle's perception of the environment, such as obstacles detected by sensors. These cost maps can then be used for autonomous path planning or visualization of how a robot perceives its environment (Ferguson & Likhachev, 2008). The AGV's LiDAR sensors store data as "friction" values, which range on a scale of 0 to 255. Higher friction values represent obstacles, such as trees, boulders, or buildings, that may be more difficult to navigate through or around. Obstacles perceived to be impenetrable are given a value of 255. Carnegie Mellon's National Robotics Engineering Center (NREC) utilized a method known as direct pixel subtraction to compare physically and virtually produced cost maps to determine differences in the two environments (Rander & Browning, 2009). The two images are compared pixel by pixel, with the difference generating a new value for each specific pixel. The resulting cost map then shows the difference between physical and virtual obstacle perception. NREC found a notable difference between simulated vs reality costs maps occurred behind obstacles, specifically vegetation. Potential causes for discrepancies in the two costs maps were differences in the simulated vehicle pose vs true vehicle pose, as well as differences in the simulated height, size, and density of obstacles (Rander & Browning, 2009).

3. Simulation Requirements Development

Requirements documents, often categorized as Capability Design Documents (CDD), are documents that outline the operational requirements for the system. Requirements documents include metrics along with associated threshold and objective values for the system. Thresholds are the minimum values required to make the system successful, and objectives are the ideal goal values. The authors reviewed several sets of requirements documents to identify operational requirements along with threshold and objectives for the AGV. While a formalized requirements document for the current Army systems are not final, requirements documents that were similar to the one being developed were reviewed to allow for metric comparisons. Once these metrics, with their threshold and objective values, were found, a requirements traceability matrix was developed. The matrix consisted of four categories defined in Cox et al. (2022), AGV environment, sensor, autonomy, and behavior, and two subcategories, threshold and objective. Once a metric was identified that could adequately answer the requirement, it was grouped into one or more of the categories at which point the threshold and objective values were identified from the requirements documents or existing literature. Table 1 shows a sample of the requirements traceability matrix detailing the thresholds and objectives for three autonomous vehicle requirements.

Table 1. Excerpt from the AGV Requirements Traceability Matrix

Requirement	Threshold	Objective	Requirement Source
Waypoint Following	Radius of half the body length or within 2 meters	< 2 meters	Previous requirements document
Minimal Intervention	> 1 intervention	Level 5 autonomy	Society of Automotive Engineers (SAE) (SAE International, 2021)
Avoid Obstacles	Radius of half the body length or within 2 meters	< 2 meters	Previous requirements document
Stop if Rollover Conditions	Side slope = 40%	Side slope = 60%	Previous requirements document

4. Simulation Validation

4.1 Data Refinement

Researchers were given access to physical and virtual testing data from 2022, totaling 13 total iterations of vehicle testing. Given the transition to the ODD concept for 2023, the data was further refined to incorporate this control. The research team identified a portion of the 2022 testing route that overlapped one of the new ODD routes, titled Dirt Low Vegetation, and extracted the physical and virtual vehicle test data along that segment to evaluate the two for any differences. This segment of the course is approximately 150 meters along a dirt path with relatively low vegetation, representing some of the least difficult terrain the AGV would be required to navigate. In total, five iterations, two physical and three virtual, with vehicles moving in the same direction were analyzed. Figure 1 depicts all 13 vehicles tests from 2022 on the respective course. Physical paths are represented by shades of green circles, with virtual represented by shades of red triangles. Each data point represents vehicle position taken at one-second intervals along the route. Figure 2 shows the five selected test iterations along ODD Dirt Low Vegetation. The blue box in both figures represents the area from which the respective ODD resides along the original course. The yellow stars represent the start and end of the ODD. Both figures were generated by the research team utilizing ArcGIS® Pro software.

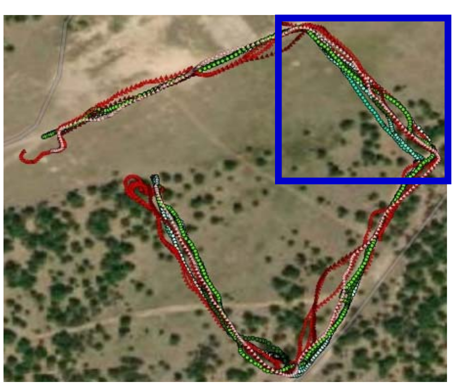


Figure 1. 13 AGV Tests from 2022 (Physical and Virtual)

Figure 2. Physical and Virtual Tests Along ODD Dirt Low Veg

ISBN: 97819384964-4-8

4.2 Metrics and Methodology

Four metrics were extracted from the five respective iterations for analysis. Total completion time is the duration in seconds for the vehicle to navigate the ODD. Distance is the total distance traveled in meters by the vehicle over each respective iteration. Speed is the average speed of the vehicle in meters per second. Mean fatal obstacles is the average number of fatal obstacles detected by the vehicle's autonomous navigation system. Fatal obstacles are those that are either impassable or would significantly impede the vehicles' path, causing it to change course. An object is considered fatal if it has a friction value of 255, where friction is a result of the sensor return and perception algorithm. These metrics were then compared between physical and virtual runs via statistical tests to determine the magnitude of variation between the two environments.

4.3 Statistical Analysis

A surface examination of the physical and virtual tests appears to show a moderate difference between the two environments. The physical iterations on average completed the ODD in 52 seconds, traveling 145.7 meters at a speed of 2.8 m/s and detecting 157 fatal obstacles. The virtual iterations completed the ODD in an average of 45 seconds, traveling 148 meters at a speed of 3.0 m/s and detecting 350 fatal obstacles. In short, the simulated vehicle appears to be capable of navigating its environment faster than the physical vehicle in the real world, even while detecting more fatal obstacles. A complete summary of each individual vehicle test of interest is displayed in Table 2.

Table 2. Table of Statistics for Autonomous Vehicle Tests along ODD Dirt Low Vegetation

		Completion Time	Distance	Speed (m/s)		Fatal Obstacles (#)	
	Run	(seconds)	(m)	Mean	Std. Dev	Mean	Std. Dev.
ODD Dirt Low Veg	Physical 1	52	147.3	2.86	0.97	206.6	91.7
	Physical 2	52	144.1	2.78	0.79	108.2	73.6
	Virtual 1	45	146.1	3.23	1.04	281.2	106.6
	Virtual 2	40	146.7	3.65	0.72	327.2	138.5
	Virtual 3	50	151.4	3.02	1.00	441.9	260.2

After reviewing descriptive statistics for the data, statistical tests were conducted to determine whether the apparent difference was statistically significant. Due to the small sample size available, the data could not be fit to a distribution and non-parametric statistical tests were utilized to conduct analysis. Given that the virtual environment should emulate the physical, the null hypothesis was that no difference exists between the physical and virtual completion time, distance traveled, speed, or fatal obstacles. A Mood's Median Test was conducted with an alpha of 0.05, and results are shown in Table 3.

Table 3. Mood's Median Test Results Comparing Medians for Physical and Virtual Tests

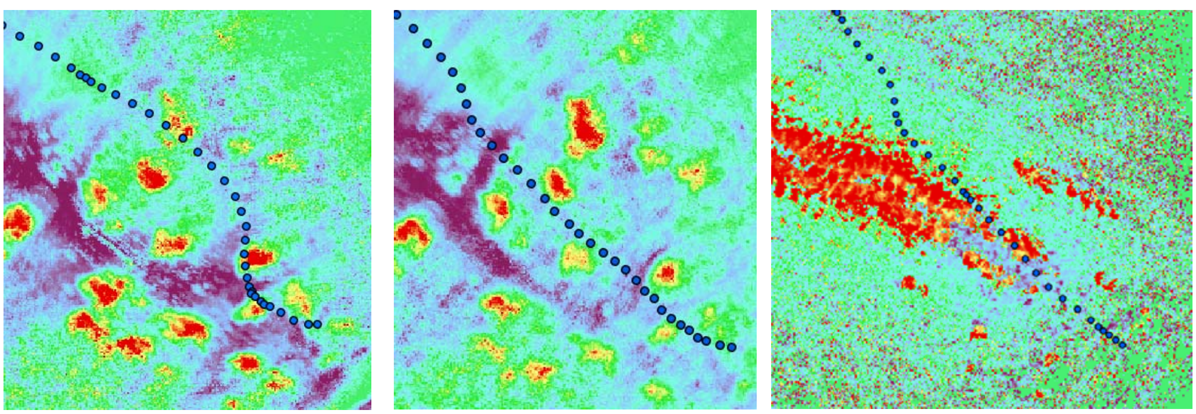
	Mood's Median Test Results			
	Completion Time	Distance	Speed	Fatal Obstacles
X^2	9.21	1.63	4.82	4.85
p-value	0.01	0.44	0.09	0.09

The median completion time for Physical Runs (52 seconds) was statistically significantly higher than the median completion time for Virtual Runs (45 seconds); $X^2(d.f{=}2, N{=}5) = 9.21$, $p{=}0.01$, with the vehicle completing the ODD in less time in the virtual environment than in the physical. No statistical difference was found for the other metrics. This could be interpreted as failing to reject the null hypothesis and concluding that autonomy performs similarly in both the physical and virtual environments. An alternate conclusion would be that the sample size was too small to identify significant variation and that there are issues between the virtual and physical environments. A cursory look at this small data set appears to suggest that the vehicle detects more fatal obstacles and moves faster in the virtual environment than in the physical. This could suggest that the autonomous navigation component of the AGV is exploiting errors in the simulation to complete the course faster, demonstrating that the virtual environment lacks the fidelity to accurately predict performance in the physical environment. However, due to the small sample size this conclusion is not supported by the statistical analyses. A larger data set would be required to confirm or deny this hypothesis. The research team recommended focusing future testing efforts on conducting more test iterations on this ODD, both physically and in simulation, to gather a more representative data set that reflects the AGV's performance in the two environments.

Because statistical analysis can only confirm or deny the presence of variation and not necessarily the reason behind it, the researchers began to look for ways in which to visualize how the vehicle perceives its environment. The methodology through which to do so was decided to be development of cost maps from LiDAR data. The AGV's autonomy stack perceives its environments and makes decisions based on the friction data collected from forward facing LiDAR sensors. The autonomous navigation components of the AGV then read this friction data and determines a path forward with the lowest cost, or the path with the least amount of friction. By developing cost maps, researchers would be able to visualize how the AGV is perceiving obstacles and if variation in perception is influencing the vehicles path.

4.4 Cost Map Analysis

To develop cost maps for analysis, perceived friction values were extracted at one-second intervals along the vehicles path on ODD Dirt Low Veg. The data was extracted into text files, and American Standard Code for Information Interchange (ASCII) administrative data was incorporated to prepare the file for raster conversion. For this study, the cost map is 400 by 400 pixels with a pixel size of 0.3 meters and covers an area of 120 meters by 120 meters. Additionally, coordinate data was included to specify the location of the lower left corner of each cost map. After the raw cost map data was properly formatted, the individual cost maps were created in ArcGIS® Pro utilizing the ASCII to Raster conversion tool. Because each cost map represents a single second over the course of a run and sensor pose may differ according to vehicle orientation, the cost maps were then combined using the Cell Statistics function to determine the average friction between the multiple images. Completed mosaic cost maps from Physical Tests 1 and 2, and Virtual Test 1 depicting the main vegetation patch along the ODD are displayed in Figure 3. Low-cost areas or "paths of least resistance" are depicted in purple, while high-cost obstacles are depicted in red.



A) Physical Test 1 B) Physical Test 2 C) Virtual Test 1

Figure 3. Mosaic Cost Maps Depicting Vegetation Cluster Along ODD Dirt Low Veg

An initial visual analysis of the mosaic cost maps presents intriguing insight. The vehicle does not appear to always take the lowest cost route and instead appears to bisect obstacles. After discussing with the ERDC-ITL team, this may be a result of the AGV's path planning behavior. Even though there may be a path that provides less resistance, the quickest way to travel between two points is a straight line. Even though an obstacle may interfere with the vehicle's most direct route, if it can maintain a safe distance from an obstacle by "hugging' it, it is quicker to hug the impediment than to completely travel around.

When comparing the cost maps of the two physical iterations, an explanation for the resulting path deviation can be identified. In Physical Test 1, the AGV identified a break in the vegetation that it deemed navigable and bisected it as a result. In Physical Test 2, the same break is not identified, and the vehicle takes a different path. While the variation in obstacle detection resulted in an altered path, the vehicle still navigated to the end of the ODD in both scenarios. If the vehicle can demonstrate repeated successful autonomous navigation to a predetermined location, minor variations in its path may not be important. It is important to consider the relatively small scale of this test however, as these variations may become more significant as the AGV is required to navigate a longer course along more complex terrain.

Analyzing the virtual cost map, further explanation for the discrepancy between physical and virtual environments becomes apparent. While the vehicle can successfully navigate the ODD to the same degree of success as in physical testing, there may exist differences between the simulation environment and simulated LiDAR sensors that are causing the AGV to detect more friction and more fatal obstacles. Similar to the results from the physical tests, while this may not prove to be an issue on a small scale, the variation may become more significant on a larger scale. If this is not deemed to be an issue upon further testing of this ODD, the degree of variation should be monitored as testing is expanded to more complex ODDs.

Analysis through direct pixel subtraction methodology may provide insight into the reason for variation in the AGV's ability to detect obstacles in the physical and virtual environments. However, the research team was unable to perform this analysis due to the lack of fidelity in the virtual mosaic cost maps. The team recommends it as a possible analysis technique going forward, should the simulation fidelity be improved.

5. Conclusion

Although statistical analysis of summary data for physical and virtual, or simulated, vehicle performance does not suggest that there is a significant difference between the AGV's performance at the test site and in the virtual simulation, the researchers suggest conducting larger scale testing on ODD Dirt Low Veg to confirm this conclusion based on discrepancies found in the descriptive statistics and cost map analysis. The metrics investigated provide a general understanding of vehicle performance in the two environments, which is supplemented by visual analysis of cost maps. ERDC has decided to adopt the research teams' cost map visualization methodology to continue to evaluate the performance of the virtual simulation. Once holistic analysis of cost maps from the respective ODD is finished, further conclusions may be drawn about the validity of the virtual simulation environment.

6. References

Cox, J., DeRosier, S., Howell, A., Jones, J., and Thompson, B. (2022). Developing Requirements to Validate Autonomous Ground Vehicle Simulations. *Proceedings of the Annual General Donald R. Keith Memorial Conference* (pp. 127-132). West Point, New York: Society for Industrial and Systems Engineering.

Deems, J. S. (2017). Lidar measurement of snow depth: a review. *Journal of Glaciology*, 467-479.

Ferguson, D., and Likhachev, M. (2008). Efficiently Using Cost Maps For Planning Complex Manuevers. *Proceedings of International Conference on Robotics and Automation Workshop on Planning with Cost Maps.* Philadelphia: University of Pennsylvania.

McManamon, P. F. (2019). History of LiDAR. In P. F. McManamon, *LiDAR Technologies and Systems.* Bellingham: SPIE.

Rander, P. W., and Browning, B. (2009). *Empirical Evaluation of the Virtual Autonomous Navigation Environment.* Pittsburgh: NREC.

SAE International. (2021). Taxonomy and Definitions for Terms Related to On-Road Automated Driving Systems

Wang, Y., Chao, W.-L., Garg, D., Hariharan, B., Campbell, M., and Weinberger, K. Q. (2019). Psuedo-LiDAR from Visual Depth Estimation: Bridgin the Gap in 3D Object Detection for Autonomous Driving. *IEEE/CVF Conference on Computer Vision and Patter Recognition* (pp. 8437-8445). IEEE.

ISBN: 97819384964-4-8