Proposal for the Initiation of General and Military Specific Benchmarking of Robotic Convoys

Paul Maxwell  
United States Military Academy, paul.maxwell@westpoint.edu

Joshua Rykowski  
United States Military Academy

Gregory Hurlock  
United States

Follow this and additional works at: https://digitalcommons.usmalibrary.org/usma_research_papers

Part of the Electrical and Computer Engineering Commons

Recommended Citation


This Conference Proceeding is brought to you for free and open access by USMA Digital Commons. It has been accepted for inclusion in West Point Research Papers by an authorized administrator of USMA Digital Commons. For more information, please contact nicholas.oljnyk@usma.edu.
Proposal for the Initiation of General and Military Specific Benchmarking of Robotic Convoys

Paul Maxwell, Joshua Rykowski, and Gregory Hurlock
Department of Electrical Engineering and Computer Science
United States Military Academy
West Point, USA
{paul.maxwell, joshua.rykowski}@usma.edu, gregory.hurlock@us.army.mil

Abstract—This paper identifies the need for a standard method of benchmarking emerging robotic systems with a focus on military, multi-robot convoys. Benchmarking is commonly used throughout academia and industry as a method of evaluating and comparing products. In this paper we propose a generic form that these benchmarks may take in the future. Classification categories, such as, obstacle avoidance, area mapping, and convoy coherence are all possible elements of this benchmark. The goal is a standard benchmark that can be used to evaluate military multi-robot convoy systems.

Keywords—multi-robot systems, robotic benchmarks, convoys, military robotics

I. INTRODUCTION

The field of intelligent, mobile robotic systems has increased exponentially over the recent years. The Department of Defense’s (DoD) robotic roadmap [12] is driving much of the research in this field as evidenced with the research reported by the Defense Technical Information Center (DTIC) [14], [15], and [16]. This roadmap describes an end state where autonomous robots conduct a significant portion of mundane military functions to include autonomous vehicles. One of the primary motivations for this roadmap is the reduction of risk for human members of the DoD. This paper focuses on military, multi-robot convoys and therefore we will look at situations that can mitigate human risk.

In military scenarios, drivers are often fatigued and that can lead to accidents and injury. The implementation of robotic control algorithms can mitigate human error in vehicle operations. Factors such as distracted driving, impaired driving, or driver error can have fatal effects on not only the driver, but also those vehicles around him or her. Robotic control can remove that human error. An example of research in this domain would be Toyota’s semi-autonomous car that was debuted at this year’s Consumer Electronics Show [17].

Military operations require regular ground convoys to logistically support fielded units. Due to the fact that this is the lifeline of the military during a time of war these lines are often targeted. Therefore, supply operations, and convoys in general, are extremely risky to military personnel. Here, the creation of a multi-robot convoy would mitigate the risk to soldier. A multi-robot convoy would allow the operator to control or monitor the convoy from a safe distance. Should the convoy come under attack, there is no immediate threat to human life.

Robotics and, more specifically, multi-robot convoys attempt to satisfy safety, efficiency, and the advancement of the robotics field. However, as more work is completed on these types of systems, the fact that there is no central standard to compare people’s work becomes much more of a poignant issue. Many researchers have developed metrics that allow for the benchmarking of select portions of robotic systems ([9], [6], and [10]). However, each of these proposals are limited in scope and do not address a holistic benchmark. This is the primary goal of this paper; to propose and initiate the creation of an accepted standard regarding military multi-robot convoys. The creation of standard benchmarks will allow members of the field to gauge their performance and compare their results.

II. RELATED WORKS

There have been numerous strides in the fields of mobile robotics and robotic benchmarking over the past decade. In this section we discuss some of the related works in these two fields. Even with these steps in the right direction there has been no singular event to bring together mobile robotics and comprehensive benchmarking. This paper proposes the groundwork for what could become a comprehensive test of performance metrics for mobile robotics.

There are a number of research projects focused on mobile robotics and their development into multi-robot convoys. These projects can be categorized based on their methodologies. We will look at three different papers, each focusing on a different method of control. Michaud et al. develop a convoy of robots using Pioneer 2 mobile robots equipped with cameras in order to recognize a color associated with each individual robot [1]. This control method works well in that the robots are capable of distinguishing between what they’re following. However, where this design suffers is in its lack of localization and mapping capabilities that are robust in real world environments that the military will encounter. Because of this the robots will possibly fall short of accomplishing area mapping which will then affect any control algorithms that the system may utilize.

Santos et al. on the other hand take a completely opposite approach to convoy control [2]. The authors describe their use of Pioneer-3DX mobile robots equipped with laser range finders. In this case, the robots are fitted with hardware that
allows them to conduct precision movements. However, due to the fact that they are not equipped with any visualization hardware, the design is somewhat flawed when applied in a more complex environment. Allowing the robot to drive based on sensory data alone is dangerous because it is possible for the robot to not know exactly what it is following. The fusion of the current sensor setup with visualization hardware would assist the robot in making that distinction.

Finally, Hayes et al. take a lower level hardware approach with the creation of swarming robots [3]. This method using IR sensors is simple and diverse. However, in its application to real-world robotics and scenarios, the design falls short of a practical system. Due to the lack of accurate ranging devices or visualization hardware, a larger scale implementation of this design would not be feasible.

Thus far we have referenced research papers in the field of mobile robotics. Among these papers we have noted a few areas where these designs could improve when compared to practical application and real-world military scenarios. These points bring to light the importance of creating a standard benchmark. Madhavan et al., recognize the need for benchmarking in intelligent robotic systems [4]. In their paper they describe the importance of a benchmark and offer examples of work already done by groups such as the RoboCup Federation, NIST, and the Performance Metrics for Intelligent Systems Workshop. These groups all present some sort of test to evaluate intelligent systems. However, what they lack is comprehensiveness.

Collins et al. and Calisi et al. both offer benchmarking metrics in their papers [5], [6]. Collins’ group demonstrates how area mapping could be quantified when comparing a rendered map to the ideal solution. In their paper, Calisi et al. present a framework for the evaluation of mobile robot and vehicle control algorithms. Both of these works contribute to what could be a comprehensive set of test metrics. However, in reality these are stand-alone efforts that do not give a complete set of test metrics for mobile robotics.

A benchmark suite for path planning and trajectory following is introduced in a paper by J. Baltes [8]. This work recommends benchmark tests for these items that maintain a broad applicability across domains and are not implementation specific. The author’s proposals are a starting point for robot benchmarking but are not complete. We propose the development of a benchmark suite that covers all of the relevant functions of a military, multi-robot convoy.

III. **GENERAL BENCHMARKING OF MULTI-ROBOT CONVOYS**

It is our opinion that performance metrics for military robotic convoys can be separated into six categories: military application (milApp), area mapping (areaMap), autonomous navigation (autoNav), convoy coherence (conCoh), convoy integrity (conInt), and obstacle avoidance (obsAvoid). These categories, although beneficial to a high level design, should not be considered equal priority. The reason for this is because within these six categories, there may be alternative design possibilities that could mitigate the need for that particular aspect of the design. Therefore, as benchmark standards are created for each category, we believe that there should be different weights associated with each category. The weights of the score in this example proposal should sum to 1. Due to the dynamic environment within the robotics field the categories and scoring are open for adjustment and expansion. However, as an initial proposition, we believe that these six categories accurately represent the elements necessary to create a well functioning multi-robot convoy. Finally, in addition to these categories, we recommend there be consideration for added functionality. For example, these add-ons could be functions such as wireless monitoring, kill switches or manual overrides, or the inclusion of heterogeneous robots in order to make the system more robust. Again, these categories, as well as point ranges and weight associated with the categories, are very flexible. Equation 1 shows an example benchmark score based on the proposed categories and their associated weights.

\[
\text{Benchmark Score} = (w_1 \cdot \text{milApp}) + (w_2 \cdot \text{areaMap}) + \\
(w_3 \cdot \text{conCoh}) + (w_4 \cdot \text{conInt}) + \\
(w_5 \cdot \text{autoNav}) + (w_6 \cdot \text{obsAvoid})
\]

A. **Military Environment**

The military environment standards will be developed based solely upon military specifications. These specifications include such factors as: temperature thresholds, weather resistance, impact resistance, and communication’s security. For example, military convoys must be capable of correctly functioning in harsh or unpredictable weather such as a dust storm. In a military environment, it is essential that the optical recognition function operates correctly in any adverse weather and light situation. Designs such as the one by Santos et al. [2] would not function well in the complex domain of a military environment. In this design, the authors use a simple fiducial system in order to produce formations. These graphical patterns can easily be obstructed by weather or light conditions and potentially cause faults within a convoy. Therefore, this would not be a suitable solution for a military application due to its easily compromised nature.

When developing mobile robotic systems, robust performance standards such as those needed by the military could be extended to civilian convoys as a universal standard. For example, a civilian, multi-robot convoy needs to be able to identify other robots in the convoy in all weather and light conditions. All robust robotic systems should be prepared to encounter numerous, unexpected scenarios when performing in the real-world. Therefore, we propose that the military environment metric be based solely on the already established military specification standards and the remaining benchmark categories be made increasingly stringent.

B. **Localization and Mapping**

A robot requires the ability to accurately map its surroundings and determine its position within those surroundings. To be able to create a map the robot needs to be able to estimate its own location. This is often done through the use of a robust sensor suite that may include a laser range finder. Using one of these devices allows the robot to take a precise measurement of its surroundings, and then through the use of localization algorithms, these measurements can be
made in to a detailed map. These maps are then used for navigation.

In terms of multi-robot convoys, localization and mapping is the lowest priority. Peripheral sensors such as GPS and laser range finders allow a convoy to successfully and coherently move throughout an area without actually mapping it. Additionally, if maps are considered necessary for the convoy’s operation, pre-rendered area maps can be loaded into the systems for navigation. A system capable of completing a basic convoy operation can succeed without mapping.

Evaluating the map produced by the multi-robot convoy is a difficult problem to address. One method would be to compare the map produced by the robot to an actual map of the area. However, this is only possible if current maps of the area exist and are easily accessible. In areas that the military operates in this may not be the case. A method to be able to properly assess one localization and mapping implementation against another implementation was proposed by Burgard et al. [13]. In their method of comparing these implementations they examine the poses of the robots during the robots’ data acquisition. This provides two benefits. First, this method is able to compare localization and mapping implementations that produce different types of maps. Second, this method is sensor invariant in that it does not matter whether the robot performed the task utilizing a laser range finder or a visual technique.

C. Autonomous Navigation

Autonomous navigation is closely tied to the aforementioned localization and mapping but is, more specifically, the ability of a robotic system to navigate from one location to another location without any external assistance. Often robotic systems are classified as semi-autonomous rather than fully autonomous, that is, a user has the ability to supersede the robot’s actions. This is due to the pervasive lack of trust in a robot’s ability to execute their tasks safely. As a society we simply have not developed the confidence in the systems we create to give them full autonomy in something such as navigation. With respect to multi-robot convoys, autonomous navigation is a r. In the case of a multi-robot convoy, this navigation can be left to the lead member of the convoy. The designer must decide whether they will use an intelligent leader with less capable followers (a heterogeneous solution) or a convoy where all members have similar capabilities (a homogeneous solution). The latter decision increases the capabilities of the convoy but also increases the overall cost of the convoy systems. The former option reduces the system cost but creates a single point of failure for the convoy should there be some malfunction in the lead vehicle.

The ability for the convoy to navigate from point to point is a critical piece of the operation. However, navigation does not receive the highest level of priority due to the possibility of manual override or control (e.g., expert driver in the vehicle or controlling the vehicle from a remote location). System designs exist in multi-robot convoys where there is a lead vehicle that is controlled manually. In this case, autonomous navigation would not be necessary, only the ability for other vehicles to accurately follow the lead vehicle. More advanced multi-robot convoys may have autonomous navigation implemented. Those systems that are able to incorporate both accurate localization and autonomous navigation will achieve the most precise level of movement and therefore the highest score with respect to a benchmark standard.

We have determined that evaluation of the ability of a multi-robot convoy to navigate could depend on two subcategories: completion of a pre-defined course and positional accuracy with respect to each waypoint in the course. A standard set of various test courses of increasing difficulty and point value could be created. These courses should contain multiple waypoints over varying terrain types that are a predetermined minimum distance apart. The positional error could be based on the straight line distance of the lead vehicle to each destination waypoint. To encourage timely course completion, the points in that category will be awarded based upon the course completion time for a convoy. Therefore a team that completes the course in less time would receive more points than a slower team. Setting a time standard will encourage teams to create an efficient system but at the same time will also discourage them from simply going for the fastest time possible and sacrificing accuracy. A possible equation for this metric is:

$$Nav_{score} = \text{Completion}_{score} - (\omega * \text{Pos}_{err}) + \beta_t$$

(2)

In this equation, $\omega$ is a weighting factor for positional error, $\text{Pos}_{err}$ is the positional error for waypoint $i$, and $\beta_t$ represents the bonus points for faster completion times.

D. Convoy Coherence

Convoy coherence refers to the accurate positioning of each vehicle in the convoy with respect to the other convoy members. We determined that this would be an important metric to include because of its real-world application. Military vehicle convoys must maintain formations that dictate lateral separation and the angular orientation of neighboring vehicles. For example, a convoy may be required to maintain an echelon-left formation with 100 meters lateral separation between vehicles.

Numerous research works such as [11] have demonstrated methods of achieving convoy coherence. However, these algorithms rely on methods that simplify the control problem but may not be feasible for field environments (e.g., the use of fiducial recognition panels). To obtain our goal, a coherence benchmark needs to measure vehicles in a field environment.

This performance metric is not given highest priority simply because the position of the individual vehicle within the convoy may need to change during an operation. For example, the user may want the convoy to be able to switch from a linear formation to a wedge formation. Additionally, the convoy may have a swarm technique implemented in which case there is no specific formation. These designs should still be evaluated based on position and orientation to the lead robot.

As Baltes [8] suggests, convoy coherence should be assessed based on the amount of error present in the position of members of the convoy. This could be determined in two ways: distance error in vehicle lateral separation and the angular error between vehicles. Linear distance serves as the
first evaluation tool for convoy coherence. Each vehicle present within the convoy, for example, may have a set separation distance of 50 meters. Using a given sampling rate, this metric could calculate the average separation error for a test run conducted upon a set of pre-determined courses resulting in a score for the system. The angle between the vehicles is the second metric for convoy coherence. This allows the convoy to be evaluated when dealing with various formations. The test might use the average angular error for a given test course based on a pre-determined formation. Figure 1 shows an example of how these errors might be assessed. Here, $d_{\text{actual}}$ and $\theta_{\text{actual}}$ are the original set separation and angle standards for the follower robots. $d_{\text{error}}$ and $\theta_{\text{error}}$ are the errors associated with the separation and formation angle by which the score for convoy coherence will be determined. The errors should be the summation of both the errors throughout the course, but also the summation of the errors of each subsequent follower robot. This error quantity could then be subtracted from a maximum score for a perfect trial run on the given course to obtain the evaluated system’s score. Equation 3 shows what the total error may look like for separation error. In this equation $m$ is the number of sample points and $n$ is the number of follower vehicles in the convoy.

$$AvgError = \sum_{f=0}^{n} \sum_{f=0}^{m} d_{\text{error}}$$

With respect to swarming formations, error within the formation will be a sum of all errors present between each swarm robot and the lead. The straight line separation error will remain the same as with regular convoy systems. The orientation error will be based on where the swarm robots initially position themselves with respect to the lead. For example, if one of the robots were to initially position itself at a 30° angle from the lead, the orientation error would be based off of that initial angle. The test run that these errors are evaluated on would be on a predetermined course.

### E. Convoy Integrity

Convoy integrity is one of the more important aspects of the military convoy performance metrics. Convoy integrity is defined as the convoy’s ability to maintain formation and continue movement towards its goal. This is an essential element due to the variability of combat environments (e.g., night, day, obscured). During both prototype testing and real world testing, unforeseen factors that affect the entire convoy and cause individual members to fail will occur. Therefore it is essential that the convoy be able to continue despite these detrimental factors that could include things such as weather or human error. Convoy integrity could be assessed using two dimensions: the recognition of the target vehicle (i.e., the vehicle being followed), and the ability of the convoy to recover from convoy member failure.

It is critical that the vehicles in the convoy be able to distinguish the vehicle they are following from their environment. For example, it is likely that these multi-robot convoys will operate in high traffic environments such as highways or urban areas. Robots will need to be able to distinguish their convoy leader from the numerous other vehicles that are moving and changing position around them. Additionally, these follower systems need to be able to distinguish the lead vehicle in various conditions. Weather conditions such as rain, sand storm, or darkness can have a catastrophic effect on the ability of the follower to distinguish the leader.

Researchers such as Guo et al. [7] have developed methods for identifying vehicles in a variety of conditions. Ideally, convoy elements should be able to recognize vehicles in its convoy using sensors such as forward looking infrared (FLIR) that operate well in the harsh conditions often found in a combat zone. The ability of a convoy member to recognize other convoy members can be tested using a suite of vehicle images such as the vehicle in Figure 2. These images can be grouped into sets based on factors such as, the intensity of the image, the percentage of the vehicle that is visible, and the angle at which the vehicle is seen. The robot can be evaluated on the number of images correctly identified for a selected test set.
The convoy must also be capable of recovering from the failure of a member within the convoy. Despite a breakdown or the destruction of a member of the convoy, the rest of the vehicles must be able to reorganize and continue their mission. It is possible that under real-world conditions the entire convoy may halt and wait for the downed vehicle to be recovered or fixed. However, this recoverability test is specifically to measure the convoy’s capability to continue and complete the given mission. The test could be designed to award a set point value for recovering from a disabled vehicle within the formation. The point value could be Go/No-Go or scaled based on the time required to successfully recover from the fault. We recommend that teams have the option of demonstrating multiple instances of recovery (under the condition that there are at least three robots in the convoy) in order to measure the robustness of their system. Figure 3 shows an example scenario consisting of three robots: a lead and two followers. In this case, the second vehicle is disabled and the third is expected to be able to continue to maneuver around the disabled vehicle and continue following.

F. Obstacle Avoidance

Obstacle avoidance is an essential function of multi-robot convoys and mobile robotics in general. Much of the functionality of obstacle avoidance can be accomplished through local sensing. When determining the best course of navigation, the robot can use a rendered map to avoid known obstacles. This may be an acceptable practice; however, this would only be acceptable under ideal conditions. In the real world we encounter a dynamic environment and not all obstacles along a route can be pre-determined. Therefore, it is critical that the robot be given a way to be aware of its immediate environment in real-time. Without the basic ability to avoid obstacles, the convoy is likely to experience some degree of failure due to the dynamics of its environment.

We recommend that obstacle avoidance should be given the highest priority among the convoy performance metrics. This is because many of the other metrics contain alternative solutions or they are not mission essential. However, within the convoy, every vehicle must be capable of avoiding an obstacle. One could say that a work-around for obstacle avoidance is to manually take control of that vehicle. However, this approach while possible is not a realistic solution to the problem. Each vehicle in the convoy needs the capability to act independently with respect to obstacle avoidance and be able to navigate around a given object and continue with the leader.

Our proposal for quantifying obstacle avoidance performance is similar to that of the autonomous navigation metric. We believe that a repository of standardized obstacle test courses should be created. Again, each of these courses would have varying levels of difficulty and point values. While navigating the course, for each obstacle that the robot impacts or cannot circumvent, a point value would be deducted from the score. These courses will also have a time standard associated with them and more points would be awarded to convoys that complete the course in a faster time. Figure 4 shows two examples of possible obstacle avoidance courses. From one course to the next, there is an increasing level of difficulty and therefore an increased point value.

IV. EXAMPLE BENCHMARK SCORE

Table 1 is an example scoring of developed multi-robot convoy systems based on our proposed benchmark model. We have included a suggested weight for each category. In this case, no design received points for military environment because the convoy systems evaluated were not designed for military application. The convoys were awarded points in convoy integrity either based on their theoretical performance or based on the hardware described or the actual demonstration of integrity concepts. The convoy coherence was similarly evaluated however in most cases the authors provided evidence of what errors their designs experienced. Because no design actually attempted to conduct area mapping, points were awarded based on possible add-on capabilities, in short, whether or not the system had a robust sensor suite that could perform localization and mapping. Each paper received points for obstacle avoidance because they each contained a sensor that enabled them avoidance capabilities, such as, IR or sonar sensors. In this case, we
assumed that each design navigated a 500 point map (from a standardized set of maps mentioned earlier in the paper). The autonomous navigation was again based on whether or not the system had localization and mapping capability. We awarded the design points assuming that with an accurate device it would be able to render its surroundings and navigate through them. Additionally we assumed that because the laser range finder is precise, that the robot would be accurate in its navigation. Finally each design was given extra consideration based on features such as kill switches and wireless state monitors. This initial framework leaves plenty of room for improvement and expansion. However we believe that it provides a solid framework for future works.

Figure 4. Examples of obstacle avoidance courses with increasing level of difficulty. Teams would receive a set number of points for completing course (a) and a larger number for completing the more complex course (b).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Military Environment</td>
<td>0.22</td>
<td>500</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Convoy Integrity</td>
<td>0.167</td>
<td>500</td>
<td>330</td>
<td>0.0</td>
<td>200.0</td>
</tr>
<tr>
<td>Convoy Coherence</td>
<td>0.167</td>
<td>500</td>
<td>10.0</td>
<td>80.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Area Mapping</td>
<td>0.11</td>
<td>500</td>
<td>0.0</td>
<td>250.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Obstacle Avoidance</td>
<td>0.22</td>
<td>500</td>
<td>50.0</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>Autonomous Navigation</td>
<td>0.11</td>
<td>500</td>
<td>0.0</td>
<td>80.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Raw Total</td>
<td>3000</td>
<td>390.0</td>
<td>360.0</td>
<td>260.0</td>
<td></td>
</tr>
<tr>
<td>Weighted Total (Benchmark Score)</td>
<td>500</td>
<td>67.8</td>
<td>60.7</td>
<td>40.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Example scores for robotic designs including both raw scores and total scores after weights are applied.

V. CONCLUSIONS AND FUTURE WORK

In conclusion, we believe that the benchmarking of military, multi-robot convoys is an essential step in the development of more advanced systems that will meet the DoD’s goals contained in its robotic roadmap. Although there have been numerous proposals for individual benchmarks, there is a lack of a comprehensive evaluation tool. More specifically, there are no proposed benchmarks for military multi-robot systems. These individual proposals do not evaluate the full complexity needed in a robotic system navigating a real-world environment. The creation of a comprehensive set of tests will provide the robotic community with not only a guideline on what to put into a system, but also with a goal to work towards. Many people within the community have made great strides in the field of mobile robotics, however providing those researchers and manufacturers with a defined goal will increase the productivity and efficiency of future designs.