2020 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
AUTONOMOUS GROUND VEHICLE TECHNICAL SESSION
AUGUST 11-13, 2020 - NOVI, MICHIGAN

AUTONOMOUS CONVOY OPERATIONS IN THE ROBOTIC TECHNOLOGY KERNEL (RTK)

Nicholas Alton[1], Matthew Bries[1], Joseph Hernandez[1]

[1]Southwest Research Institute, San Antonio, TX

ABSTRACT

This paper presents two techniques for autonomous convoy operations, one based on the Ranger localization system and the other a path planning technique within the Robotic Technology Kernel called Vaquerito. The first solution, Ranger, is a high-precision localization system developed by Southwest Research Institute® (SwRI®) that uses an inexpensive downward-facing camera and a simple lighting and electronics package. It is easily integrated onto vehicle platforms of almost any size, making it ideal for heterogeneous convoys. The second solution, Vaquerito, is a human-centered path planning technique that takes a hand-drawn map of a route and matches it to the perceived environment in real time to follow a route known to the operator, but not to the vehicle.

Citation: N. Alton, M. Bries, J. Hernandez, "Autonomous Convoy Operations in the Robotic Technology Kernel (RTK)", In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 11–13, 2020.

1. Introduction

Autonomous convoy operations have the potential to transform transportation and logistics, leading to dramatically enhanced cost savings, efficiencies, and safety. Convoys are particularly compelling because a manually operated vehicle can lead an automated convoy, leveraging the perception and intelligence of a trained human operator and extending it to an arbitrary number of following vehicles. As a result, the U.S. Army has invested considerable resources in developing automated convoys on a variety of platforms.

Recent efforts include the Autonomous Mobility Appliqué System (AMAS) [1], Autonomous Ground Resupply (AGR), and the Expedient Leader-Follower (ExLF) projects.

Most of these projects have targeted homogeneous convoys for simplicity and speed of development. However, in practice, many convoys are heterogeneous, made up of several types of vehicles ranging from large Palletized Load System (PLS) trucks and Light Medium Tactical Vehicles (LMTVs) to smaller escort vehicles such as High Mobility Multipurpose Wheeled Vehicles (HMMWVs) and Mine-Resistant Ambush

Protected vehicles (MRAPs). Fielding a useful automated convoy system requires that it handle any likely combination of platforms.

In 2017, the Coalition Assured Autonomous Resupply (CAAR) [2] project set out to deliver exactly such an automated convoy system.

The specific target was a six-vehicle convoy featuring two HMMWVs, two LMTVs, two British HX60s, and a UK-built deployable unmanned aerial vehicle (UAV). These vehicles were chosen to ensure the resulting convoy system could not only control heterogeneous platforms but also combine assets from multiple forces as necessary.

Building on the success of previous programs, CAAR aimed to combine the capabilities of AMAS with the Robotic Technology Kernel (RTK) [3], the ground vehicle autonomy system of the U.S. Army Combat Capabilities Development Command Ground Vehicle Systems Center (CCDC GVSC). While AMAS had focused on providing a low-cost, low-risk, kit-based (“appliqué”) convoy solution, RTK provides a complete vehicle-agnostic autonomy suite including environmental perception and navigation. The combination of the two kickstarted the advanced automated convoy envisioned by CAAR.

The basic functionality of the combined system was first demonstrated in 2017 with two HMMWVs running RTK and four LMTVs equipped with AMAS. The RTK vehicles were first and last, running autonomous route following behaviors, with the AMAS vehicles deployed between them as convoy followers. The speeds of all of the vehicles were automatically synchronized to create a loosely integrated convoy.

After this proof-of-concept demonstration, the full integration commenced, eliminating redundant components and combining the strengths of both systems. To provide robust convoy following capabilities to RTK vehicles not equipped with the AMAS kit, SwRI developed two novel convoy following technologies: Ranger Following and Vaquerito Following.

2. Ranger Following Concept

Ranger [4] is a high-precision localization system based on a single low-cost downward-facing camera. This section presents the adaptation of this localization system into a high-precision convoy following system.

2.1. Background

Many autonomous driving systems, including RTK, rely on having a precise measure of the vehicle’s position in the world.

For this purpose, global navigational satellite systems (GNSS) like the Global Positioning System (GPS) typically do not provide sufficient precision. In addition, satellite-based systems are vulnerable to jamming, geometry-related numerical issues, and terrain-based signal loss.

Mapping generally mitigates these problems: the system compares some stored representation of its surroundings to the data streaming from its sensors and estimates its position in that map, which is somehow geo-registered. However, achieving the required precision from a typical map-based approach requires high-precision 3D sensors such as lidar. Lidar sensors produce tremendous amounts of data and come at high costs. Maps created with lidar are significantly susceptible to changes in their surroundings.

Ranger takes a different, simpler approach: it just looks at the ground beneath the vehicle to build a map and estimate the vehicle’s position on it.

The complete Ranger system is described in detail in [4], but a brief synopsis is given here.

The system consists of a downward-facing camera (Figure 1), a low-cost computer, an LED-based lighting system, and an electronic control board (Figure 2). Its total cost is generally below even inexpensive lidar hardware.

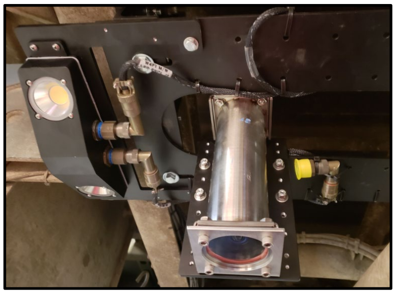

Figure 1. *Ranger downward-facing camera mounted on a vehicle.*

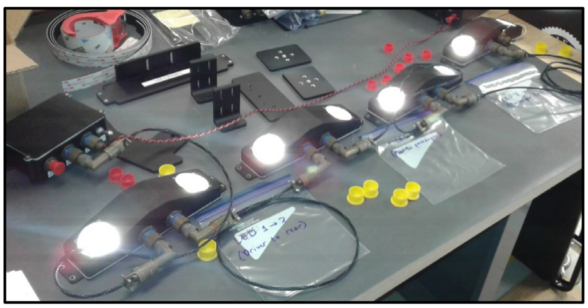

Figure 2. *Ranger lighting and electronics system on a bench.*

The camera and LEDs are synchronized by the control hardware to capture well-lit, low-blur images of the ground while traveling up to 80 mph (130 kph). These ground images appear similar to the human eye, but turn out to be surprisingly unique, allowing a single patch of ground to be reliably matched in a map of millions. Ranger's matching system is also robust to changes or occlusions in the captured images of up to 70%, making it a stable long-term solution for precise vehicle localization.

Figure 3 shows a visual representation of this matching process. The current ground image is highlighted on top of the route in the upper-left, while the matches between that image and the matching map tile are shown in green in the bottom-right. The blue boxes identify the position of the match image relative to the live image and vice-versa.

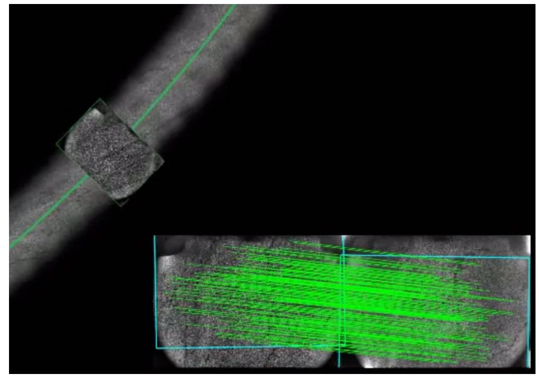

Figure 3. *Ranger map highlighting matched frame; inset: individual matches*

The Ranger image matching algorithm is composed of three steps: contrast stretching, feature detection, and feature correspondence.

Contrast stretching normalizes and enhances the captured images to enable matching across variable lighting conditions. Even though Ranger supplies its own light source, bright environmental lights like the sun create light and dark regions in images for which a standard autoexposure system cannot compensate. Ranger breaks the image into a grid and maximizes the dynamic range within each grid cell while keeping it balanced with its neighbors. The grid cells can be any size, although using smaller cells can slow the system and poorly compensate for higher-frequency changes in lighting conditions across the image.

The second step is feature detection, which is performed by traditional detectors from computer vision research. By default, Ranger uses the STAR [5] feature detector and ORB [6] feature descriptors, but any available pair may be chosen using simple parameters.

Finally, Ranger finds the map tile with the greatest correspondence to the features extracted from the live image. Beginning with the most likely map tile, defined as the one spatially closest to the most-recent position estimate, features are compared in both directions (map to live and live to map). Any feature pairs in which both elements are the best match for the other are stored, and a Random Sample Consensus (RANSAC) method is used to estimate a rigid transform between the images. If a sufficient (configurable) quantity of additional feature pairs also fit this rigid transform, a four-point homography is performed, again using RANSAC, to eliminate any spurious pairs. Finally, a 2D rigid transform is computed by least-squares regression using the remaining pairs. This transform provides a high-precision position of the vehicle relative to the map. These transforms have been shown to have a standard deviation under 2cm when compared to ground truth.

Ranger relies on accurate heading information to rapidly match tiles; when heading information is missing or inaccurate, matching quality degrades (or computational effort increases dramatically through the use of rotation-invariant features or iterative heading matching). However, in the Ranger Following application described below, the following vehicles can be reasonably assumed to have the same heading as the lead vehicle at the same point in the path. Therefore, this heading information is always available.

2.2. Ranger Following

Ranger Following is an expansion of the Ranger methodology to the convoy following application.

In a typical Ranger application, building a map requires collecting and optimizing data, followed by human-in-the-loop editing and annotation. The refined map is stored offline for use by any number of vehicles.

However, in Ranger Following, a lead vehicle builds the map in real time by collecting ground images and odometry data and extracting image features. The extracted position-linked features are sent via radio to one or more following vehicles also equipped with Ranger. The following vehicles compare real-time imagery with the incoming feature sets from the leader to compute a transform from the leader's historical path to the follower's current position. This transform allows the follower to track the leader's path with the high precision of Ranger, even in GPS-denied environments.

Ranger Following expands the set of usable surfaces from Ranger, as well, especially for unimproved and soft surfaces. A durable Ranger map requires that the road surface be relatively consistent over time; malleable surfaces like snow, sand, and grass tend to cause problems. However, a Ranger Following streaming map is created only a few seconds or minutes before it gets used. As a result, it can function even over dynamic surfaces. Ranger Following has been successfully tested on concrete, asphalt aggregate of varying age and quality, many kinds of dirt and mud, sand, gravel, and off-road paths formed by tire tracks. The only surfaces that have caused Ranger Following to fail to date are highly dimensional surfaces like tall grass and reflective surfaces like smooth ice.

In addition to the precision made available by Ranger, Ranger Following has the additional benefit that the following distance is limited by only the range of the communication link and the storage capacity of the following vehicles. No visual line-of-sight is necessary, allowing it to work even in low-visibility environments.

Furthermore, Ranger Following need not suffer from string stability issues, because each follower can operate on exactly the march unit leader's path instead of relying on the path of its immediate leader. Errors do not propagate down the convoy, and the entire convoy's path remains stable.

2.3. Following Performance

Ranger Following has been shown experimentally to provide precise convoy following as shown in Figure 4. In operational environments with heterogeneous vehicles, Ranger Following has a mean absolute cross-track error of 22 cm. Given Ranger's oft-demonstrated precision, it is hypothesized that the additional error in Ranger Following is a consequence of vehicle path following performance instead of Ranger Following system performance.

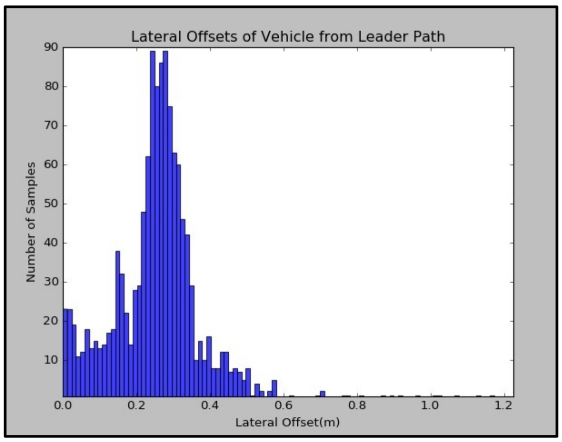

Figure 4. *Absolute lateral offset of follower vehicles using Ranger Following. The spike around 0.0m is likely due to steering jitter.*

Ranger Following has been tested at speeds up to 35 mph (56 kph), although its theoretical maximum speed is limited by only hardware capabilities (with a cap at 80 mph) and total radio bandwidth. Each Ranger tile occupies about 70 KB, and these tiles are collected, processed, and transmitted roughly every 75 cm. As a result, at 35 mph, Ranger Following requires about 1.5 Mbps per follower, assuming a standard one-to-many network topology. Mesh networks or relay transmission could significantly reduce this requirement, enabling a convoy to grow as large as necessary.

Ranger Following has been shown to be an effective and precise following technique, but there are notable failure cases. Like a normal Ranger deployment, Ranger Following requires live imagery to overlap with the map. However, at system startup of a convoy, the followers cannot have seen the leader's path yet, so another following mechanism must be used until they do.

A similar scenario occurs when the follower vehicle has significantly different steering characteristics from the leader, so the necessary path around a sharp turn must differ and the follower must leave the leader's Ranger map due to platform limitations. In most cases, the duration of this loss is short, and the follower's estimate of the leader's position remains good; however, if the duration is longer, a fallback approach must be in place.

An additional challenge arises when the Ranger hardware is placed at very different heights above ground because of different vehicle platforms. In these cases, the pixel density (pixels per meter) of the ground images can vary significantly, so feature extractors identify different features, and fewer true matches are available. Two solutions have been shown to be effective: physically lowering the Ranger camera on the taller vehicle; and adjusting the pixel density in software. Depending on the difference in height, the latter approach may substantially impact the robustness of the Ranger matching, requiring greater image overlap to identify a transform. However, it does not require physical changes to the vehicle.

Finally, large sweeping shadows caused by varying vehicle geometry can impact the contrast stretching algorithm in different ways, resulting in a different set of extracted features. Changing contrast stretching configuration can help in this scenario, but pathological cases would require a different technique (see Section 5, Future Work). These shadows can be seen in Figure 5.

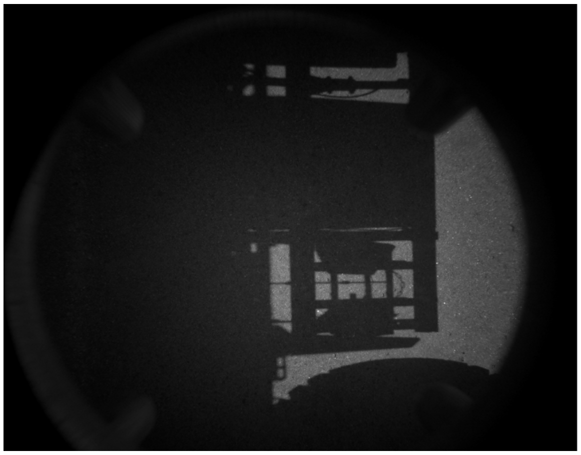

Figure 5. *Shadowing effects in a Ranger image.*

2.4. Components of Ranger Following

The Ranger Following system is composed of three key software components: the Ranger streaming nodes, the Ranger streaming localizer, and the RTK convoy navigation module called Trail Guide.

The first major components are the streaming nodes: the publisher, server, and client.

The publisher node incorporates the first half of the Ranger process, combining vehicle odometry with Ranger features extracted from camera images to provide georeferenced features for transmission to follower vehicles.

The server node takes the generated messages from the publisher and sends them over the radio interface to other vehicles. The server splits the incoming data into three data streams to manage available network bandwidth. Ranger data is transmitted over a TCP stream, other necessary data streams like path history over a second TCP stream, and high-frequency data like vehicle position updates are sent over UDP.

Finally, the client nodes collect transmitted data and republish it into the follower vehicle's ROS system to estimate the leader's path, generate a transform, and follow that path.

The second major component of Ranger Following is the Ranger Streaming Localizer (RSL). The RSL forms a pair with the Publisher streaming node, matching extracted features from live Ranger imagery to the incoming leader path data. While a typical Ranger setup would search through a map to find the best tile match, the RSL searches for matches in a streaming buffer of image features. This buffer is a sliding window of feature data organized by their proximity to the follower vehicle. After finding a match, the RSL outputs a transform from the follower vehicle to the leader vehicle's relative localization frame. Using the relative frame instead of the global frame allows Ranger Following to operate in completely GPS-denied environments.

The final major component of the Ranger Following system is Trail Guide, a multifunction node made of two primary modules, the Route Creation Module and the Speed Control Module.

The Route Creation Module constructs the leader's path using the best available data for later use either by Ranger Following or by Vaquerito Following, discussed in Section 3.

Trail Guide is not specific to Ranger Following; instead, it attempts to use whatever incoming data is available to generate the best estimate of the leader's path according to a hierarchy of data precedence. The highest-priority data is the Ranger Following path, as it is the most accurate and precise; this path is used as long as the RSL continues to match the leader's path. If the RSL fails for any reason, Trail Guide will simply pass through the path generated by the AMAS Perception Vehicle Following (PVF) system, a line-of-sight following technique. Finally, if both RSL and PVF fail, Trail Guide will output the Vaquerito Following path. Combining these inputs into a single output results in mostly seamless transitions between following methods. The RTK Path Following Controller, which is responsible for controlling the vehicle to actually track the intended path, performs some smoothing downstream to ensure switching path inputs does not cause unsafe or unwanted vehicle behavior.

The second major Trail Guide module is the Speed Control Module, which computes the optimal speed for the following vehicle. It takes the immediate leader's position as well as the distance along its path to the leader vehicle and calculates a speed that will maintain a safe gap distance at all times while also attempting to achieve a nominal convoy formation. The Speed Control Module is extremely flexible, allowing input from a variety of sources to ensure that it can still function even on vehicles that do not support a particular following technique.

The current implementation of Ranger Following uses a few additional, smaller modules, the most important of which is described in brief below.

The 2D rigid transform computed by the Ranger feature correspondence module described above has some error in the yaw estimate, causing follower vehicles to yaw back and forth. Adding an Extended Kalman Filter (EKF) smoothed out the discrepancies between the Ranger transform from the RSL and the follower vehicle's actual odometry.

The transforms output by the EKF relate the Ranger camera pose to the global map frame. However, the conventions ROS uses for storing and transmitting rigid transforms throughout a robotic system assume each system has a single parent frame, but Ranger Following actually connects two transform trees through the Ranger system. As a result, Ranger Following implements a small node to invert this component of the transform graph on follower vehicles, causing confusion for developers (but no impact on system performance).

3. Vaquerito Following

As discussed in Section 2.1, Ranger Following requires additional hardware that may not be equipped on all vehicles in a heterogeneous convoy. In addition, Ranger Following has some failure cases based on environmental conditions, system limitations, and vehicle platform capabilities that require a fallback following method. Simply following GPS breadcrumbs from a lead vehicle does not always provide the necessary accuracy or precision to follow a path appropriately. Therefore, the Vaquerito path registration technique from RTK was adapted for a convoy following application.

3.1. Vaquerito Overview

Vaquerito is a path registration technique that allows a vehicle to follow a rough route from a map or a previous recording or even a route drawn by hand. The technique assumes that GPS positioning error can be approximated as a slowly moving offset from the vehicle's actual position. Vaquerito attempts to estimate this offset over time to determine the correct placement of the route in the vehicle's actual environment. This assumption works well in practice and even successfully adapts for the warping of the target route relative to the environment, such as when the map was drawn by hand instead of recorded.

To estimate the correct placement of the target route in the environment, the vehicle first needs a concept of a drivable path. In RTK, two-dimensional metric maps identifying traversibility, known as costmaps, represent the environment to the system. An example of these costmaps, built up from the complete suite of sensors available to the vehicle, is shown in Figure 6. Areas of high traversibility, like a road, are shown in lighter colors.

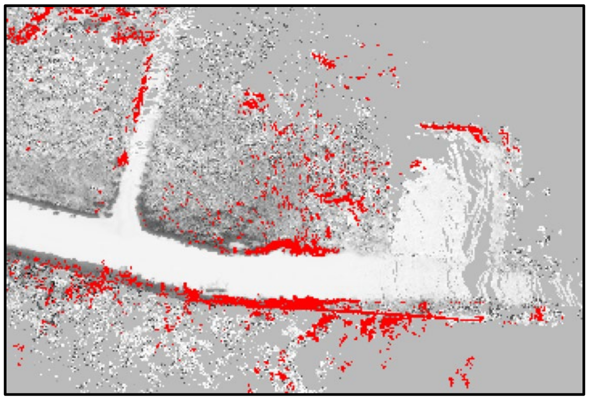

Figure 6. *Example RTK costmap. Higher traversibility is shown in lighter colors; obstacles in red.*

Once the vehicle has a concept of traversibility, the route can be placed in the environment and evaluated for fit. The target route is initially superimposed on the costmap at its nominal georeferenced position, typically identified by GPS coordinates, and the total traversal cost is estimated by averaging the cost of each costmap cell intersected by the route.

This process is repeated for a configurable variety of linear and offsets from the starting position, and the offset with the lowest cost is selected as the best placement of the route. In Figure 7, the nominal position of the target route is shown in green, and the final offset path is shown in blue. Note that the blue path occupies the center of the light corridor on the costmap identifying low traversal cost. Because the costmap changes whenever new data is collected, Vaquerito applies regularization and rate limiting to ensure the output path does not jump and result in erratic vehicle movement.

In some cases, especially around improved roads, several offsets result in a similar traversibility cost. In those cases, Vaquerito attempts to identify the center of the route by combining the top offset candidates. An ellipse is fit to the top 5% of candidates (a configurable value), and the offset is chosen along the major axis of the ellipse, effectively ensuring that the cross-track error of the new offset is equivalent to the average cross-track error of the top offset estimates.

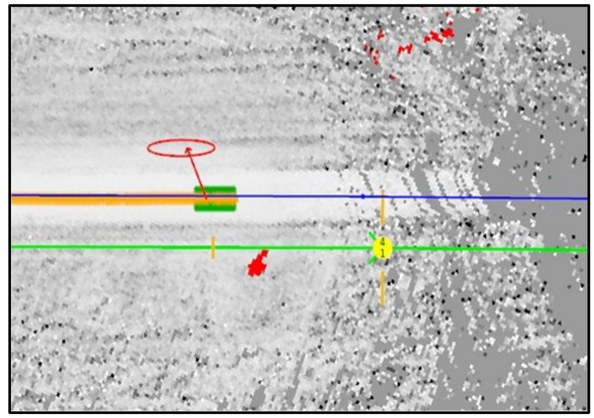

Figure 7. *An RTK Costmap of data on SwRI's test facility in San Antonio, TX. The green line represents the Vaquerito target route; the blue line represents the least cost offset.*

3.2. *Integration with Following*

Because the typical use of Vaquerito already operates on georeferenced paths, adapting it for following was straightforward. The leader's path is transformed to the follower's local coordinates and used as the target route. Then fitting the target route to the follower's costmap takes place as usual. The components of Vaquerito detailed above allow Vaquerito Following to be smoothly integrated into the RTK convoy following system as either a primary or a fallback mechanism.

3.3. *Vaquerito Following Performance*

Vaquerito deliberately offsets the leader's path to match the observed environment, so measuring its cross-track error as a metric of performance makes no sense. In fact, if the lead vehicle is following an improved or semi-improved road, but offset to either side, Vaquerito may cause the follower to center itself in the route rather than exactly follow the leader. In addition, because it relies on

GPS positioning, on paths wider than GPS error, it performs similarly to simply following GPS paths.

However, experiments on narrow trails and paths, especially those that are merely tire tracks through the environment, show that Vaquerito Following tracks the lead vehicle's path quite closely.

4. RTK Integration

One of the major tasks of the CAAR project was integrating the behaviors described in this paper into RTK so they were applicable not only to CAAR vehicles but to any RTK platform. Prior to this work, RTK's only convoy following behavior was a thin wrapper around the AMAS functionality and required AMAS-equipped vehicles. However, contemporaneously with this research, RTK gained a flexible behavior framework that allowed new behaviors to be loaded at runtime instead of compile time. A behavior module was created to use the output of the Trail Guide described above, enabling all RTK vehicles to use these following techniques.

5. Future Work

Though testing the combined convoy following system described in this paper was successful with a five-vehicle convoy, several challenges were encountered and mitigated; a full solution would benefit future systems.

First, the contrast stretching difficulty described above has occasionally been mitigated in practice by changing the mounting position of the camera on the underbody of the vehicle. However, this mitigation is not always practical or feasible, depending on the vehicle platform. Changing the grid used for contrast stretching helps, but at a great cost in complexity and time. It is possible that doing this may also worsen feature detection performance since there would be a greater number of discontinuities caused by grid edges. Segmenting light and dark areas of an image based on arbitrary shapes before applying contrast stretching, instead of using a uniform grid over the whole image, is hypothesized to resolve observed issues with lighting and shadows. Utilizing cameras and image formats with higher dynamic range could also help improve feature detection in these areas.

Second, the pixel density difficulty described above was mitigated in the same way; however, the existing software solutions come at the cost of reduced robustness to occlusion and change, making the Ranger Following system overall less consistent during operation. An alternate solution would be to use significantly higher-resolution cameras or different optics to enable the pixel densities to be cleanly matched in software.

Finally, the existing following techniques remain entirely independent, even building in redundant copies of algorithms like the EKF. Though this decision keeps the systems independent, it also increases the computational cost of the system and creates challenges to graceful fallback from one system to another. Extracting and sharing common components would resolve both issues.

6. Conclusion

Because of its potential to be a force multiplier for logistics, autonomous convoy operations continue to be an important focus area of the Army's autonomous vehicle development plans. The work described here and performed as part of the CAAR program brings more of the developed capabilities to bear in a single system. As a result, all RTK-equipped autonomous vehicles now have access to a precise and effective convoy behavior encompassing a wide range of vehicle platforms. The individual following methodologies combined produce a system that functions better and degrades more gracefully than any one methodology would on its own.

7. References

[1] Bernard Theisen, Autonomus Mobility Applique System (AMAS) JCTD, Jun 23, 2011, https://apps.dtic.mil/sti/pdfs/ADA550916.pdf Powerpoint Presentation.

[2] D. Pirozzo, J.P. Hecker, A. Dickinson, T. Schulteis, J. Ratowski, and B. Theisen, "Integration of the Autonomous Mobility Appliqué System into the Robotic Technology Kernel", In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), NDIA, Novi, MI, Aug. 13-15, 2019.

[3] Dr. Robert Kania; Mr. Phil Frederick; Mr. William Pritchett; Mr. Brian Wood; Mr. Chris Mentzer; Dr. Elliot Johnson "Dismounted Soldier Autonomy Tools (DSAT) - From Conception to Deployment", in Proceedings of the Ground Vehicle Systens and Technology Symposium (GVSETS), NDIA, Novi, MI, Aug. 12-14, 2014.

[4] K. Kozak and M. Alban, "Ranger: A ground-facing camera-based localization system for ground vehicles," 2016 IEEE/ION Position, Location and Navigation Symposium (PLANS), Savannah, GA, 2016, pp.170-178,doi: 10.1109/PLANS.2016.7479698.

[5] Agrawal M., Konolige K., Blas M.R. (2008) CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. In: Forsyth D., Torr P., Zisserman A. (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol 5305. Springer, Berlin, Heidelberg.

[6] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *2011 International Conference on Computer Vision*, Barcelona, 2011, pp. 2564-2571, doi: 10.1109/ICCV.2011.6126544.