NISTIR 5244

Recommendations for Performance Evaluation of Unmanned Ground Vehicle Technologies

Prepared by

Martin Herman
Sandor Szabo
Karl Murphy
David Coombs
Tsai-Hong Hong
Harry Scott
Nicholas Dagalakis
Kenneth Goodwin
James Albus

Sponsored by

Advanced Research Projects Agency
Software and Intelligent Systems
Technology Office

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Building 220, Room B124
Gaithersburg, MD 20899

NIST

NISTIR 5244

Recommendations for Performance Evaluation of Unmanned Ground Vehicle Technologies

Prepared by

Martin Herman
Sandor Szabo
Karl Murphy
David Coombs
Tsai-Hong Hong
Harry Scott
Nicholas Dagalakis
Kenneth Goodwin
James Albus

Sponsored by

Advanced Research Projects Agency
Software and Intelligent Systems
Technology Office

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Building 220, Room B124
Gaithersburg, MD 20899

August 1993

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Arati Prabhakar, Director

Abstract

The ARPA Unmanned Ground Vehicle (UGV) Demo II program is developing intelligent, semi-autonomous UGVs to perform cooperative tasks in militarily significant scenarios. This document provides a set of performance measures, tests, and instrumentation that can serve as the basis for a performance evaluation program under the ARPA program. These measures, tests, and instrumentation focus on those that seem feasible to do during the next three years or so, i.e., that can actually be accomplished during the course of the Demo II program. This document covers the following technologies: (1) mobility (i.e., the ability to move through the environment), (2) visual sensing for navigation and driving, (3) reconnaissance, surveillance and target acquisition, (4) radio communications, and (5) ground truth data acquisition.

Foreword

The chapters and appendices in this document were prepared by the following people.

Chapter 1............................Martin Herman
Chapter 2............................Sandor Szabo, Karl Murphy
Chapter 3............................David Coombs
Chapter 4............................Tsai Hong
Chapter 5............................Harry Scott
Chapter 6............................Nicholas Dagalakis, Ken Goodwin
Appendix I..........................Nicholas Dagalakis
Appendix II........................Nicholas Dagalakis
Appendix III.......................George Lukes, U.S. Army Topographic
Engineering Center

Equipment listings do not imply a recommendation by NIST.

Table of Contents

Chapter 1. Introduction

Performance measures and evaluation can provide quantitative methods for estimating the capabilities and limitations of unmanned ground vehicles under a variety of circumstances. This is crucial to predicting the effectiveness of UGVs under future battlefield conditions, and thus for evaluating the potential benefit of UGVs to the armed forces of the nation. Military planners must understand precisely what the performance characteristics of UGVs really are if they are to integrate UGVs into the plans and requirements documents that are needed for instituting weapons systems procurements.

The ARPA/NIST Workshop on Performance Evaluation of Unmanned Ground Vehicle Technologies, held during September 1992 in Gaithersburg, MD, has provided a first look at UGV performance evaluation issues for the Demo II program [2]. It was clear from the workshop that UGV performance evaluation is a vast field, and there are an enormous number of aspects of UGV performance that can be measured and evaluated, in a large number of different ways. It is also clear that a single program such as the ARPA UGV Demo II program can only incorporate a relatively small set of performance measures and methods.

The purpose of this document is to suggest a set of performance measures, tests, and instrumentation that can serve as the basis for a performance evaluation program under the UGV Demo II program. This set has arisen as a result of considering the variety of ideas presented during the workshop.

The measures, tests, and instrumentation presented here focus on those that seem feasible to do during the next three years or so, i.e., that can actually be accomplished during the course of the Demo II program. Some of these should be useful and applicable even to the Demo A system, which is the first of four UGV/Demo II demonstrations.

A concrete set of performance evaluation methods has the following elements:

(1) a set of tests to be performed,
(2) a purpose and set of objectives for each test,
(3) a set of parameters to be measured during each test, along with a statement of how they will be used to evaluate the performance,
(4) a description of the instrumentation required to perform the measurements,
(5) the procedure to be followed in collecting the appropriate data,
(6) a description of the analysis to be done on the data.

There are many performance tests that can be done during the various planned demos of the program, or immediately after the demo using the demo system. In order to do these tests, the instrumentation and data collection systems must be integrated into the demo system hardware and software.

Many of the approaches to performance measures and evaluation described in this document are relatively unexplored. We have attempted to provide the type of information suggested in items 1-6 above for these approaches. However, these need to be refined and made more explicit, so that each test can eventually be written up as a concrete sequence of steps. NIST therefore proposes to refine the set of tests and measures described in this document, and apply them to the Demo II program. This will involve doing experiments at NIST and other facilities to refine and make more explicit the procedures required to perform the tests. This will also involve designing or procuring instrumentation (both hardware and software) and demonstrating how they can be integrated into the UGV system. Once this is done, the actual tests and evaluations on the UGV system can be performed by NIST and/or by other members of the Demo II team. Many of the tests can potentially be applied to the Demo A system. As the tests are described throughout this document, the text indicates which are applicable to Demo A.

The U.S. Army Combat Systems Test Activity, Aberdeen Proving Ground, Maryland, has developed a test plan and performed tests on the Surrogate Teleoperated Vehicle (STV) [1]. Although many of the tests they suggest can be applied to the Demo II vehicles, they have not developed tests for the kinds of capabilities that the Demo II vehicles will have but which the STV vehicles do not have. These capabilities include autonomous and supervised autonomous functions.

1.1. Customers and Developers

Generally, there are two groups of people who are interested in performance evaluation. The first group consists of the customers, who are interested in evaluating a total product. Examples of customers for UGV systems are Army General staff, the Unmanned Ground Vehicles-Joint Program Office, the Office of the Secretary of Defense, ARPA upper management, and Congress. The customers don't particularly care how a certain capability is achieved, or which technology is used. They are primarily interested in overall performance on tasks of interest to them. Performance measures and evaluation are useful to this group of people for several reasons. First, they provide a quantitative measure of how well the system is progressing over time. Second, they provide a means to determine the potential benefit versus cost of the system. For example, if a UGV system is to be integrated into a scout platoon, the benefits and costs can be used to determine how much the UGV system will buy in terms of enhanced system performance, and how much it will cost in terms of operator work load, logistics support, communications bandwidth, etc. In a UGV system under supervisory control, performance evaluation can aid in the allocation of tasks between the human and the machine. Finally, DOD plans and requirements documents can use these measures and parameters to specify their needs.

The second group consists of the technology developers, who develop the algorithms, software, hardware, and integrated systems that constitute a UGV system. The developers want to be able to improve their systems by (1) identifying the weaknesses and strengths of the subsystems and (2) measuring whether the subsystem performance improves with certain modifications in the subsystem. Evaluation can help determine when a technique or subsystem is appropriate, characterize its reliability, precision, and limitations. It can also be used to provide focus for future research and development.

In this report, we have looked at performance evaluation approaches that would be useful for both groups, although we have put more emphasis on approaches useful for the technology developers. However, we have generally considered "high-level" measures. The reason is that we want to have a program that helps developers determine what kind of performance they need to focus on. But measures involving the details of algorithms should be left for the developers themselves to define and apply as their algorithms and software are being developed.

1.2. Overview of Document

This document provides performance measures and methods in the following UGV technology areas: (1) mobility, (2) visual sensing for navigation and driving, (3) RSTA, and (4) radio communications. In addition, methods and instrumentation for acquiring ground truth data are described.

1.2.1. Mobility

Chapter 2 of this document discusses performance evaluation of mobility operations. This deals with evaluating the ability to move through the environment. Mobility functions and tests are classified into three categories: coordinated vehicle functions, individual vehicle functions, and subsystem functions.

1.2.1.1. Subsystem Functions

Test procedures are provided for the following two subsystem functions: actuation control and (non-imaging) navigation sensing. For actuation control, the test objectives are to evaluate the actuator and control system used to manipulate the driver control devices (steering, brake, etc.). These tests measure the actuation system (actuator, linkages, gear trains, etc.) and the computer control system.

For navigation sensing, the objective is to measure the accuracy and repeatability of the on-board operational sensors in estimating vehicle poses. The estimated parameters must be compared to a reference measurement, or ground truth, which is obtained by external systems having an order of magnitude greater precision than the system under evaluation.

1.2.1.2. Individual Vehicle Functions

Test procedures are provided for the following individual vehicle functions: driving, road following, intersection driving, obstacle avoidance, and off-road driving. Many of the tests described in this area can be applied to the Demo A system, particularly tests dealing with basic driving, road following, and intersection driving.

The objective of the driving tests is to evaluate vehicle performance during basic driving operations. The desired command signals (steering, speed, or destination pose) would be generated by teleoperators, by prerecorded paths, or by use of parameterized equations. The results will provide insight into the accuracy, smoothness, and stability of the vehicle in executing motions.

The objective of the road following tests is to evaluate vehicle performance when traveling on roads with various types of road boundaries, such as gravel, grass, painted lanes, etc.

The objective of the intersection driving tests is to evaluate the response of the vehicle to various types of intersections that are encountered during road following tasks. The tests determine whether the vehicle is able to locate intersections, select the proper path through the intersection, and re-acquire the new road segment.

The objective of the obstacle avoidance tests is to evaluate the response of the vehicle to obstacles in the vehicle path: either stop or avoid. Does the vehicle stop in time and in a smooth manner? If the vehicle maneuvers around the obstacle, does it take the "best" path and do it smoothly? This chapter does not deal with obstacle detection, which is a component of perception and is covered in Chapter 3.

For off-road driving, the test objective is to evaluate the ability of the vehicle to navigate cross country without the aid of roads. The paths taken by the vehicle depend on the terrain and the location of obstacles and landmarks. The performance tests here are very similar to the tests for obstacle avoidance. In addition, stability of the vehicle and vulnerability to roll-over need to be measured.

1.2.1.3. Coordinated Vehicle Functions

Test procedures are discussed for two aspects of coordinated vehicle functions, line formation and bounded overwatch.

1.2.1.4. Test Apparatus

The chapter concludes with a description of two types of test apparatus. The first is an on-board data acquisition system used to evaluate performance with respect to the vehicle reference frame. It consists of a calibrated multicamera system, video recording system, accelerometer/inclinometer/speedometer system, a system for capturing control data, and an off-line analysis workstation.

The second test apparatus is a digital terrain workstation used to evaluate performance with respect to a global reference frame. It consists of a digital terrain database, a vehicle tracking system, and an analysis workstation.

1.2.2. Visual Sensing for Navigation and Driving

Chapter 3 discusses performance evaluation of aspects dealing with visual sensing for navigation and driving. The visual sensing functions are classified into two categories:

(1) Feature extraction and local mapping. This category considers separately obstacle detection and the local map used for vehicle driving. The local map is derived from visual sensing and contains the positions of local obstacles and features relative to one another and relative to the vehicle.
(2) Symbolic recognition. This category includes terrain classification, landmark recognition, and traffic guide recognition.

The impact of gaze control on visual perception is also considered in this chapter.

Each of the tests described in this chapter should be performed in a variety of environmental conditions, such as varying light levels, cloud cover, rain, etc. In addition, the effects of varying viewing parameters (e.g., camera baseline, vergence angle, focus, aperture, zoom) should be tested.

The vehicle's performance should be compared both qualitatively and quantitatively with human performance for the following tasks: obstacle detection, landmark recognition, self-localization, recognition of traffic guides, and gaze control. This may involve placing a human in the vehicle and recording his perceptions, or having the human viewing remotely obtained video data from cameras on the vehicle.

1.2.2.1. Feature Extraction and Local Mapping

Many of the tests dealing with feature extraction for obstacle detection and with local mapping accuracy can be applied to the Demo A system.

For obstacle detection, the test objective is to determine the detectability of classes of obstacles on- and off-road and with the vehicle stationary as well as moving. Obstacles to be detected include discrete obstacles (e.g., rocks, trees, bushes, posts, poles, potholes, puddles, structures, vehicles, animals) and extended obstacles (e.g., ditches, ridges, embankments, steep slope changes). The tests measure not only whether the obstacles can be detected, but whether the obstacle parameters (e.g., size, slope of embankment) can be accurately determined.

The test objective for local map construction is to evaluate the accuracy of the local maps. The tests measure how well the local map registers with both the ground truth map and the global map (on which the route plan is described). The local map must accurately locate features, whether or not they appear in the global map.

1.2.2.2. Symbolic Recognition

For terrain classification, the test objective is to measure the ability to identify and distinguish terrain. The tests evaluate how accurately and robustly terrain and features are classified, and how accurately and robustly the system can distinguish hazardous terrain and features.

For landmark recognition, the test objective is to determine what classes of landmarks can be recognized and how accurately and robustly the system can localize itself using them. In particular, the tests will establish how well the vehicle can localize itself relying only on visual landmark identification in case external reference systems (e.g., GPS) become unavailable.

For traffic guides, the objective is to measure the accuracy, robustness, and speed with which it can recognize traffic guides so as to take the appropriate actions.

1.2.2.3. Gaze Control

The objective of the gaze control tests is to determine the adequacy of gaze control to incorporate observations from varying gaze directions and to provide coverage of the visual space to enable proper visual perception. The accuracy of visual mapping should not suffer because of gaze control, and no region of visual space should be neglected.

1.2.3. Reconnaissance, Surveillance and Target Acquisition

Chapter 4 discusses performance evaluation of RSTA operations. RSTA operations are classified into six categories:

(1) RSTA from a stationary platform,
(2) RSTA from a moving platform,
(3) target recognition (from both stationary and moving platforms),
(4) RSTA teleoperation,
(5) target designation, and
(6) intelligent search.

Categories (1) and (2) involve RSTA operations such as target detection and acquisition, but not target recognition. The latter is covered by category (3).

RSTA evaluation must be done in a wide range of environments, including a range of weather conditions, visibility conditions, terrain types and conditions, clutter, target sizes, target types, and movements. We therefore use two measures of scene complexity -- overall scene complexity and overall target complexity. RSTA performance measures include effectiveness of the system for a given environmental condition, robustness over a range of environments, and accuracy for a given environmental condition.

For RSTA from a stationary platform, the test objectives are to assess the performance of the RSTA system in detecting stationary or moving targets within a range of 200 m to 3 km when targets are visible along a line of sight with less than 50% obscuration by the terrain. The test objectives for RSTA from a moving platform are the same, except that the platform can move at speeds up to 35 mi/h (56 km/h).

For target recognition, the test objectives are to assess the capability for recognizing targets within the range of 200 m to 3 km. The platform and/or the target may be moving or stationary.

The test objectives for RSTA teleoperation are to assess the operator's capability for inspection and verification of target recognition and tracking and to examine the effectiveness and efficiency of the operator's control of the RSTA sensors. This test can be applied to the Demo A system.

For target designation, the test objectives are to assess the performance of an operator in initiating a laser designation sequence for both stationary and moving targets.

For intelligent search, the test objectives are to assess the performance of the RSTA intelligent search function using terrain and doctrinal knowledge to sequence sensor pointing.

1.2.4. Radio Communications

Chapter 5 discusses performance evaluation of aspects dealing with radio communications. Since we are interested primarily in measuring the overall performance of the UGV system, we consider performance of particular communication links, but do not consider the detailed workings of a particular radio component. Potential communication parameters of interest include transmitted/received power, bandwidth, line-of-sight characteristics, range, data rate, message latency, error rates, effects of other communication on individual links, and communication system power consumption.

The following tests can all be applied to the Demo A system: link performance, coverage, line-of-sight characteristics, and communication activity.

The objective of the link performance test is to measure the throughput on communication links of interest.

The objective of the coverage test is measure the radio coverage provided by the communication system.

The objective of the line-of-sight test is to measure the degree to which the radio system can deal with obstructions.

The objective of the communication activity test is to measure the total volume of communicated messages and the rates of communication for various elements of the system. The rate of communicated information is an indicator of the relative vehicle autonomy in a given implementation. As progress is made toward Demo II, this measure will provide some indication of advances toward more autonomous operation.

1.2.5. Ground Truth Data Acquisition

Chapter 6 discusses the acquisition of ground truth data. Ground truth is used to evaluate landmark navigation, general driving, and RSTA. In all three cases, the ground truth involves map data and vehicle position data.

To evaluate landmark navigation and RSTA, maps covering a range on the order of 10 km are required, with resolution between 1 m and 10 m. Vehicle positions (both the UGV platforms and RSTA targets) are required at accuracies between 1 m and 10 m. To evaluate driving, maps covering a range on the order of 100 m are required, with resolution between 50 mm and 2 m. Vehicle positions are required at accuracies between 50 mm and 2 m.

Several different measuring instruments and methods are described for obtaining this ground truth. The method involving GPS can be applied to the Demo A system.

For obtaining vehicle positions for evaluation of both RSTA and landmark navigation, differential GPS can be used to obtain positions in real-time to an accuracy of 1 m. Positions can be obtained in non-real time to an accuracy of up to 0.1 m using post processing. We recommend that Demo A be utilized to evaluate the use of differential GPS for performance evaluations.

For evaluation of driving, several metrology systems are described. The Laser Tracker Interferometer (LTI) and the LIDAR Tracking System (LTS) are very expensive and so are described in Appendix I. The Manually Tracking LIDAR (MTL) consists of a LIDAR mounted on top of the telescope of a Theodolite. The telescope is then used to view and manually track a retroreflector target mounted on the UGV. In practice, there would be two or three MTLs and several retroreflectors on the UGV.

This system can also be used to develop topographic maps of the test site. The retroreflector target assembly can be mounted on a topographic mapping vehicle, which can be driven around the test site to obtain the map. An error of 10 to 100 mm/km can be expected.

The laser-based Real time Position Measurement (RtPM) system requires a local area positioning system to be established around a given site. The system itself consists of a transmitter set on a tripod, and a receiver comprised of two optical pieces on a pole. The receiver is mounted on the UGV. The position and orientation of the pole can be calculated from the positions of the optical pieces. This system appears to be promising for short range, real time position determination of UGVs.

Ground truth maps can also be generated by the U.S. Army Topographic Engineering Center (TEC), including a custom high resolution Digital Elevation Model and Digital Orthophoto. TEC can also provide real-time geolocation of vehicles using differential GPS. A memorandum outlining potential support from TEC to the Demo II program is included as Appendix III.

REFERENCES

[1] Raymond J. Resendes, "Detailed Test Plan for Technical Feasibility Test (TFT) of Surrogate Teleoperated Vehicle (STV)." TECOM Project No. 1-VC-200-UGV-001. U.S. Army Combat Systems Test Activity, Aberdeen Proving Ground, MD 21005-5059, March 1992.

[2] Martin Herman, ed., "Report of the ARPA/NIST Workshop on Performance Evaluation of Unmanned Ground Vehicle Technologies, September 16-17, 1992", NIST Internal Report NISTIR 5237, 1993.

Chapter 2. Mobility

2.1. Introduction

An essential requirement for unmanned ground vehicles is the ability to move through the environment. The question arises as to how well the vehicle performs such a task. Unfortunately, movement for tactical vehicles covers a wide spectrum, which in turn presents many challenges for those interested in testing and measuring the qualities of mobile unmanned vehicles.

In this chapter, we examine the test and evaluation of unmanned ground vehicle mobility operations. Two aspects are examined. First, what is to be measured, and second, how it is measured. Besides the traditional "what" measurements, such as speed and accuracy, one must also determine the purpose of the test. This, along with suggestions for organization, is briefly covered in the first section. Specific test procedures are presented in the following section. Finally, recommendations are made for development of test apparatus to facilitate performance evaluations.

2.2. Organization of Functions and Testing

Organization is the first step in tackling broad problems. With regard to performance evaluation, we organize in two areas: classification of mobility functions and classification of mobility tests.

2.2.1. Classification of Mobility Functions

A functional organization classifies mobility into three categories: coordinated vehicle functions, individual vehicle functions, and subsystem functions. A bottom-up approach is most often used to develop these functions and therefore a bottom-up approach will be adopted in performance evaluation.

2.2.1.1. Subsystem Functions

We assume the UGV is an actuated HMMWV, and therefore vehicle performance issues such as maximum speed, acceleration, and horsepower are provided by the manufacturer. The autonomous mobility portion of a UGV is composed of many subsystem components. Perception may be considered a subsystem of mobility, but it is significantly complex to warrant coverage in a separate chapter. A dividing line between perception and mobility is discussed later. The subsystems in which we are interested are the following:

<u>Actuation Control</u> - the components that apply mechanical forces and torques to vehicle control devices such as the steering wheel, brake, throttle, emergency brake, transmission and transfer case. Since performance specifications of actuators are provided by manufacturers, we are interested in the performances exhibited by the vehicle control devices, i.e., steering wheel, brake pedal, etc. These performance measures indicate the effectiveness of actuation, that is, how well the actuator's performance is transferred to the device. Other performance characteristics include ease of use, failure rates and response to unsafe operating events (emergency stop).

<u>Navigation Sensing (non-imaging)</u> - the components that sense vehicle pose while stationary and on the move. While vision and LADAR systems provide valuable information with respect to navigation, it falls under the domain of perception, discussed in Chapter 3. We do consider systems such as dead reckoning systems (steering angle and odometry), inertial systems (gyros and accelerometers) and external reference systems (GPS, LORAN, RFNG).

2.2.1.2. Individual Vehicle Functions

At the level above subsystems we are interested in the overall mobility performance of an individual vehicle. In a strict sense, mobility performance evaluation examines vehicle speeds, accelerations (and decelerations), smoothness and stability. But when evaluating autonomous mobility functions, perception plays a significant role. Mobility evaluation tests and perception evaluation tests will often be conducted simultaneously. In this chapter, when possible, we try to maintain separation between mobility and perception performance evaluation. Some examples of vehicle mobility functions that should be evaluated include:

Driving - the minimum set of functions that enables a vehicle to perform some trajectory. An analogy to driving is teach programming of an industrial robot without the aide of sensors such as vision. While factory robots obtain position information from encoders, vehicles utilize navigation sensors such as dead reckoning and inertial systems. In a similar fashion, the robot vehicle can be taught a path or commanded to perform a straight-line path or some other arbitrarily specified path. Repeatability, accuracy and resolution describe the performance of each robot.

Road-following - the function that maintains the vehicle on a road at appropriate speeds. We assume that the location of the road (i.e., its coordinates or curvature) is provided by a Perception system. The vehicle must smoothly track the road coordinates, often at high speeds, while not exceeding safety limits. One issue in determining performance will be in decoupling the performance of perception. For example, the perception system may accurately locate the road but the vehicle may drive poorly. Performing separate driving tests will provide insight into the stability of the vehicle controls and may help isolate causes of poor performance.

Road Intersection - the function that handles vehicle driving through various road configurations such as intersections, merges and splits. In the case of making a turn at an intersection, the Mobility system goal is to approach the intersection, determine the correct turn-off and negotiate the turn. The Mobility system must slow the vehicle and execute the correct path based on the Perception system's view of the intersection. This function is listed separate from Road-following functions since unique algorithms may be used by each.

Obstacle Avoidance - the function that enables a vehicle to avoid obstacles. The obstacles are detected by a Perception system using various sensors such as, acoustic rangefinder, microwave, vision and

LADAR. Overall performance is determined by the ability to detect true obstacles and false obstacles, the ability to select appropriate paths around an obstacle, and the smoothness of the motions.

Off Road - the function that enables a vehicle to travel across terrain without the aide of structured roads. The paths taken by the vehicle depend on the terrain, the location of obstacles and the location of landmarks (known obstacles), all of which are determined by the Perception system. Performance of the Perception system is described separately.

2.2.1.3. Coordinated Vehicle Functions

Once the performance of individual mobility control functions have been evaluated the next level, coordinated vehicle functions, can be focused upon. Coordinated vehicle functions include:

Form Formations - the function that enables groups of vehicles to travel in specific formations. Some example formations include: coil, herringbone, column, staggered, wedge, vee and line. When considering performance evaluation, of interest is the ability of groups of vehicles to maintain inter-relationships based upon position, orientation, and line-of-sight. Vehicles must balance responses to local circumstances that impact formation (such as navigating around obstacles and difficult terrain) and global circumstances (knowledge that other vehicles are experiencing difficulties).

Bounding Overwatch - the function that enables vehicles to perform coordinated motions while providing surveillance and coverage for each other. Each waypoint during a bounding overwatch is selected to enable a vehicle to perform an enemy surveillance. Thus, accuracy of vehicle pose is critical if the specified zone is to be observed.

2.2.1.4. Mobility vs. Other Vehicle Functions.

The mobility functions, or tasks, described above map to elements of the vehicle control system. In many instances, the tasks are dependent on elements that may be considered outside the responsibility of mobility (and possibly outside the scope of mobility performance evaluation). For example, radio communications may have a profound effect on some forms of mobility (coordinated vehicle functions) but are treated as outside the domain of mobility. Due to complexity, perception and high-level planning elements have also been separated from mobility, though their distinctions are less clear.

The primary benefit in seeking further decomposition is that separation of these functions helps provide greater resolution in system evaluation. This may allow developers to select the best elements when performing specific tasks or to more effectively apply manpower and resources toward improving system performance. From the user's point of view this will provide the best final product.

Mobility and Perception

We divide the line between perception and mobility as follows. Perception strives to determine the current state of the world. The representation of the world is maintained in a world model. Perception derives information such as the identity of features (road edge, obstacle, landmark), the locations of features and the dimensions of features. We judge perception on how well it performs these functions.

Mobility functions control the vehicle in accordance with a plan. They utilize perception information (derived from vision, external reference systems like GPS, maps, etc.) and generate steering and speed commands for the vehicle. In the UGV design, this occurs in two stages: expert modules produce steering and speed preferences which are then combined by an arbitrator. Some of the steering preferences (e.g., by ALVINN) are generated directly from relatively unprocessed images. In this case perception and the first stage of steering control are inextricably bound together, and the perception cannot be evaluated independent of the control preferences. Steering preferences are also generated from several other sources of information (e.g., maps). Evaluation of mobility functions (steering and speed preferences, and final arbitrated paths) are described in this chapter.

Mobility and Mission Planning

In general, mission planning specifies the desired action to perform based on the mission goals and the current state of the world. Mission planning may also be classified as a task performed by military personnel with assistance from automated mission planners. The output is a mission plan expressed in a symbolic form typically associated with a digital map. While some mobility functions are clearly specified by the mission plan ("perform zone reconnaissance in the direction of objective Bantam"), we have excluded mission planning from mobility. We feel that the human-in-the-loop will provide substantial performance evaluation of semi-automated mission planning. For mobility evaluation, we take for granted the plan for a given mission is correct.

Other forms of planning are of interest in evaluating mobility functions. For example, given an obstacle coordinate and dimension (detected by Perception), we would like to determine how well the vehicle traveled around the obstacle (i.e., path planning). Negotiating maneuvers with other vehicles is another example of mobility planning. These examples suggest a need to refine the role of planning in mobility and to separate its evaluation during performance testing.

Where is the line drawn between mission planning (involving humans) and mobility planning? One interpretation of planning is the ability to remove the operator from direct interaction with the system. At some point in a mission the operator enters the desired goal. The vehicle is then told to execute. A plan exists internally that directs the vehicle toward the goal. Events during the execution may indicate the need to change driving modes, slow down for an intersection or make a left hand turn. The performance of planning is based on the quality of the decision making and the amount of operator interaction required to achieve a goal. The baseline for evaluating mobility planning is based upon the performance of a human operator.

2.2.2. Classification of Mobility Tests

A second area for classification of mobility performance evaluation regards the types of tests. Example test types include: verification prior to integration, performance measurement, satisfaction of contractual obligations. Most of the early Demo II efforts will focus on development and integration of system components. Often a bottom-up approach is used to develop mobility functions. Individual groups or organizations develop these functions which ultimately must meet system interface and performance specifications. Verification of adherence to these specifications prior to integration is one type of testing. When a system nears completion, questions then focus on how well the contract requirements were satisfied. Many factors are considered, such as, completeness, documentation, and performance under various environmental conditions. In this report, we focus primarily on performance measurements. Performance is important to all types of tests.

Performance evaluation asks how well a component performs. Tests are conducted to determine the upper bounds of performance. Baselines exist, or are created, to compare performances. Baselines may consist of earlier test results, of results achieved by other systems, or by results achieved by humans. Baselines are often misleading. Underlying differences between systems must be clearly stated. The role of evaluators must go beyond simple statistical evaluation to seek out why performances differ.

Performance has many aspects. Specific measurement can be made of performance variables such as those described in the following section. Of equal importance to specific performance variables are the operational constraints of a mobility operation. For example, the maximum speed for road following will differ when operating with or without obstacle avoidance. This example also illustrates a primary difference between developers and users. Developers might emphasize the speed a vehicle reaches during road following, which will be much greater without obstacle avoidance. Users will deem this function of marginal use and will be more interested in speeds obtained under safe driving conditions.

Another aspect of performance evaluation for mobility functions is operational characteristics. These characteristics provide insight into system ease of use, reliability and maintenance. Some examples include: initialization time, mode transition time, failure rate, failure response, recovery time, diagnostics capability.

All tests should include evaluation of the operational constraints and characteristics.

2.3. Test Procedures

In this chapter test procedures are suggested for performance evaluation. The tests are categorized starting with mobility subsystems, then mobility functions for individual vehicles and finally mobility functions requiring coordination of multiple vehicles. Each test states the objective, various procedures to follow and methods for analyzing the data.

2.3.1. Subsystem Functions

In this section, the performance tests for subsystems responsible for vehicle mobility functions are described.

2.3.1.1. Actuation Control

Objective

Evaluate the actuator and control system used to manipulate driver control devices (steering, brake, etc.). These tests measure performance of the actuation system (actuator, linkages, gear trains, etc.) and the computer control system. These tests most likely will be conducted by developers and the results should be documented for future reference.

Test procedures

a) Develop interfaces to actuator servos' input and output sections and ability to capture data. Capability should include ability to read and time stamp data at servo loop rate.

b) Select performance variables and measurement apparatus for evaluating actuator control:

   i. Actuator position - to determine the accuracy and repeatability in moving an actuator to a desired position. Use the control system position feedback sensor since it is already in place. May want to verify accuracy of data by comparing to external measurement (using an LVDT for example).

   ii. Actuator speed - to determine speed (or time) of actuator in traversing the range of motion. Use control system position feedback sensor.

   iii. Actuator force - to determine amount of force applied to devices (e.g., brake pedal). Use control system force feedback sensor if appropriate. An external force sensor should be used to calibrate control system feedback sensor.

c) Command actuator motions that exercise full range of positions and forces. Measure performance variables.

d) Introduce unsafe operating conditions. For example, press emergency stop, interrupt power, interrupt communications. Document responses.

e) Document system characteristics.
   i. Time required to initialize system.
   ii. Failure events, frequency and recovery procedures.

Analysis:
a) Analyze actuator position/force performance. Examine errors (mean and variance) between commanded and true positions and forces. Determine causes of extreme errors.
b) Analyze control response characteristics. Plot commanded vs actual actuator positions and forces. Correlate actuator positions with events such as emergency stop activation. Measure delays of system.

2.3.1.2. Navigation sensing (non-imaging)

Objective

Evaluate the performance of the navigation sensing system. As mentioned above, navigation sensors are restricted to non-imaging sensors. The performance parameters to measure provide insight into the accuracy and repeatability of the on-board operational sensors in estimating vehicle poses. The measured parameters must be compared to an external reference, or ground truth, which is obtained by external systems having an order of magnitude greater precision than the system under test.

Test procedures

a) Develop interfaces to onboard navigation sensing modules and ability to capture data. Capability should include ability to read and time stamp data at pose update rate.
b) Select performance variables and measurement apparatus for evaluating navigation sensing:
   i. Vehicle pose - to determine resolution, accuracy and repeatability in estimating vehicle pose. Use vehicle tracking systems, such as laser tracking, differential GPS and/or overhead camera systems (with calibrated targets) to measure vehicle pose. Descriptions of laser tracking and GPS are presented in Chapter 6. A description of a possible balloon supported overhead camera system is presented in section 2.4., Conclusions and Recommendations.
   ii. Vehicle velocity - to determine accuracy in estimating vehicle velocity. Differentiate vehicle pose measurements made by tracking system.

c) Establish calibration transformation between navigation and tracking systems. For relative pose measurements from a starting point, only the initial navigation pose is required.
d) Conduct tests. Move vehicle to various positions. Capture navigation system pose estimates and correlate with tracking system pose measurements. Repeat motions for measurements of pose estimation at various velocities.

<u>Analysis</u>

a) Analyze performance of vehicle pose estimation. Transform navigation system pose estimates into tracking system coordinate frame. Examine errors (Root Means Squared, Circular Probability, Maximum) while vehicle is stationary and moving. Determine resolution of pose sensing. Determine causes of extreme errors.

2.3.2. Individual Vehicle Functions

These tests evaluate performance of individual vehicles performing various mobility functions.

2.3.2.1. Driving

<u>Objective</u>

Evaluate vehicle performance during basic driving operations. The goal of these tests is to measure performance of the vehicle under limited types of driving. The desired command signals (steering, speed, or destination pose) will be generated by teleoperators, or automatically by pre-recording paths or by use of parameterized equations. The results will provide insight into the accuracy, smoothness and stability of the vehicle in executing motions. For simplicity, vision based navigation is not included in this section. This test is appropriate for the Demo A UGV. The procedures, interfaces and instrumentation for this test will serve as a basis for future testing.

<u>Test procedures</u>

a) Develop interfaces to control system elements that generate vehicle steering and speed commands. Two types of commands are recorded, the desired command (e.g., input from the teleoperator) and the actual command (e.g., steering command that minimizes error signal). Capability should include ability to read and time stamp data at servo loop rate.

b) Repeat use of interface to navigation sensing system described above.
c) Select performance variables and measurement apparatus for driving:
   i. Vehicle position - to determine accuracy in achieving commanded position. Use tracking system from navigation performance evaluation.
   ii. Vehicle velocity - to determine accuracy in maintaining commanded speed. Use tracking system from navigation performance evaluation. A tap into the speedometer cable may also provide additional data.
   iii. Vehicle inclination - to determine slope of vehicle and potential for tip over. Use onboard inclinometers.
   iv. Vehicle longitudinal acceleration and jerk - to determine smoothness of vehicle motion during start and stop operations. Use onboard three axis accelerometer system.
   v. Vehicle lateral acceleration and jerk - to determine smoothness of vehicle motion during cornering. Use onboard three axis accelerometer system.
d) Command vehicle to drive along pre-specified path and measure accuracy of vehicle trajectory.
e) Measure system delays due to computation and communication, both external and internal.
f) Measure open loop step response and frequency response to various inputs as appropriate. This would include sending open loop commands directly to the steering and speed controllers.
g) Perform driving tests along structured (e.g., slalom) course under teleoperation (high-bandwidth feedback) and waypoint driving (low-bandwidth feedback). Perform tests several times to obtain distribution of measurements and to evaluate effects of training. Also perform tests under various weather and lighting conditions.
h) Document vehicle operational characteristics such as types of failures, rate of failure, down time, recovery time, safety violations.

<u>Analysis</u>

a) Analyze vehicle position performance. Examine errors (Root Mean Squared, Circular Probability, Maximum) between estimated, commanded and true position. Determine causes of extreme errors.

b) Analyze vehicle accelerations. Process longitudinal and lateral accelerations (e.g., smooth) and plot. Determine if motions exceed allowable accelerations.
c) Analyze vehicle jerks. Derive jerk from acceleration. Process similar to acceleration.
d) Analyze vehicle stability. Examine vehicle slope, accelerations and center of gravity and determine whether vehicle exceeded safe operating envelope. Determine the effect of time delay on system stability at various speeds. Calculate stability measures such as percent overshoot, phase margin and location of the system zeros and poles.
e) Determine system timing attributes. Determine system update rates (e.g., frequency of steering commands). Correlate and determine delays between estimated, commanded and true data for steering and speed. Determine causes of large delays.

2.3.2.2. Road Following

Objective

Evaluate vehicle performance when traveling on-road with various types of lane boundaries. Initial test does not include obstacles or complex road geometry (e.g., intersections). This test is appropriate for the Demo A UGV.

Test procedures

a) Develop interface to vehicle elements that generate steering and speed preferences based on visual perception. These may be different from elements used during driving tests described above.
b) Develop interfaces to vehicle elements that visually estimate vehicle position relative to the road. These elements may by closely coupled to the elements that generate steering and speed preference, as for example the ALVINN system.
c) Select performance variables and determine measurement procedure for road following:
   i. Vehicle road tracking - to determine quality of road following algorithm. Use side looking cameras to record vehicle displacement from road boundary. The video signal is stored on a video tape for post-processing.
   ii. Road boundary type - to determine robustness of algorithm to variations in road boundary types (gravel, grass, painted lanes, etc.). Boundary type also captured by side looking camera.

iii. Vehicle speed, acceleration, jerk and stability - Use measurement system described under driving performance evaluation (speedometer, accelerometers, etc.).

d) Develop storage, time stamping and retrieval system to record vehicle control signals during test. This system should be able to correlate the side-looking cameras' video with vehicle control signals.

e) Initialize road following algorithm. Document interaction required by operator to include:
   i. Specific information entered and purpose of information.
   ii. Time required by operator to enter information.

f) Allow vehicle to automatically travel along various road surfaces. Conduct tests under various weather and lighting conditions. Perform tests several times to obtain statistical distribution.

g) Document vehicle operational characteristics such as types of failures, rate of failure, down time, safety violations.

Analysis

a) Analyze road tracking performance. Process side looking camera video. Perform edge detection on lane boundary and smooth or fit curve through data. Compute mean and variance of position of lane in camera frame. Correlate commanded vehicle steering and velocity commands with measured vehicle displacement. Compute error statistics (mean, variance). Determine causes of large errors. Correlate performance with road boundary type and lighting conditions and make subjective evaluation.

b) Analyze quality of vehicle motion. Process acceleration, jerk and stability data as described in driving test.

c) Determine system timing attributes. Same as driving test.

2.3.2.3. Intersection Driving

Objective

Evaluate response of vehicle to various types of intersections that are encountered during road following tasks. Determine if vehicle is able to locate intersection, select proper path through intersection and re-acquire new road segment. Preliminary, test and analysis procedures closely follow those described under Road Following. A forward camera should be also installed to record the forward motion of the vehicle during the test.

2.3.2.4. Obstacle Avoidance

Objective

Evaluate the response of the vehicle to obstacles in vehicle's path. Obstacle avoidance depends on accurate and robust detection of various obstacle types. The Perception system performs obstacle detection, and evaluation of its performance is described in Chapter 3. In this section we describe a test to evaluate the performance of obstacle avoidance, the ability of the vehicle to maneuver around an obstacle. A significant portion of the performance evaluation makes use of the test setup described in Road Following.

Test procedures

a) Develop interfaces to control system elements that estimate location and size of obstacles. These elements may be closely coupled to the steering and speed preference generators.

b) Select variables and determine measurement procedure for obstacle avoidance:

   i. Obstacle avoidance response - to determine response of vehicle to obstacles. Use side looking cameras and add forward looking camera. The output of the three "reference" cameras is stored on a video tape (using split screen format). The stored video will serve as a document of the test and will provide information for post-process analysis of the test. Record vehicle control and sensor signals and synchronize the data with each video frame. This correlation will be useful when examining control signals, for example, when a discrepancy shows up in a steering command, the video images may be examined to see possible causes. Other benefits of this system are discussed below and in section 2.4.

   ii. Time to detection and avoidance - to determine time required to detect obstacles and start actuation response to perform correct maneuvers (stop or avoid). Use a reference camera system and storage/time stamp system to capture first visual contact with obstacle, time of first avoidance commands from control system and time when vehicle first responds to the control signals.

   iii. Smooth response - to determine smoothness of vehicle motion as vehicle stops or maneuvers around obstacles. Use accelerometer/inclinometer system described above.

c) Drive vehicle over various courses and vary types of obstacles. Record the steering preferences and the input data during the test trials described in the section on visual sensing. Note that some runs will follow the same path, and others will be permitted to deviate. In the case of objects that are placed in the canonical path of the vehicle, the path should deviate.

<u>Analysis</u>

a) Process reference camera video. The obstacle is highlighted by the evaluator and the position relative to vehicle is computed. The ability to store 3-D trajectory of vehicle with respect to obstacle will depend on knowledge of the camera system and of the obstacle (i.e., model of obstacle). Trajectory of vehicle relative to the obstacle is evaluated and a decision is made as to the appropriateness of maneuver. Example criteria for evaluation include:
   i. Did vehicle stop for obstacle or plan appropriate path around obstacle?
   ii. Did vehicle plan path around obstacle (if clear pathway existed)?
   iii. Did the vehicle take the "best" path?
b) Analyze the stability of the preferences in neighboring viewpoints during each trial and among trials. Were the transition between viewpoints smooth?
c) Analyze response to controlled objects. Compare the differences of segments of preference traces between sets of trials that differ by the presence of each controlled object. I.e., for each controlled object determine the effect of its presence on the steering preferences for which the object location is visible.
d) Determine system timing attributes. Events selected by evaluator are correlated with measured data to determine system responses and delays during obstacle avoidance.
   i. Obstacle detection response time. Evaluator examines video sequence and marks time when obstacle is sighted. The time at which the Perception system detects obstacle is used to measure response.
   ii. Control response time. Measure timing between acknowledged obstacle and appropriate control response.
   iii Vehicle response time. Two analysis methods may be used to measure the response of the vehicle to control signals. In one method an evaluator compares the start of the vehicle motion based on shift in recorded video with the start of control signals indicating an avoidance maneuver.

In the second method, the data from accelerometers indicates actual start of the vehicle avoidance maneuver and it is correlated with control commands to determine vehicle response.

2.3.2.5. Off-Road Driving

Objective

Evaluate performance of off-road driving which is the ability of the vehicle to navigate cross country without the aide of roads. The paths taken by the vehicle depend on the terrain, the location of obstacles and the location of landmarks (known obstacles), all of which are determined by the Perception system. Performance of the Perception system is described separately. The performance tests for off-road driving are very similar to the tests for obstacle avoidance and make use of the test set-up described above. In addition, stability of the vehicle and vulnerability to rollover needs to be measured.

Other possible parameters may be measured to further evaluate performance of off-road driving. For example, the goal of a trajectory may include the criteria to optimize some parameter such as distance traveled, time, roughness (as determined by vehicle vertical accelerations and jerk), stealth (maintain defilade), fuel consumption, etc. A digital terrain map registered to ground truth helps to evaluate performance for these types of trajectories. A workstation is used to plot vehicle paths and to compare them with other paths generated by humans or more complex route generators. Such a test set-up will help to evaluate coordinated vehicle tasks and RSTA as well.

2.3.3. Coordinated Vehicle Tasks

Two coordinated vehicle tasks that have been discussed frequently within the UGV program are Line Formations and Bounded Overwatch. During various phases of these tasks, each vehicle is given a specific mobility function. Measuring the performance of coordinated vehicle tasks builds upon the test procedures used for individual vehicle evaluation. The extensions to the evaluations focus on the accuracy in obtaining the desired relationships between vehicles.

Two important performance parameters are the distance and timing intervals between vehicles. These can be measured with the tracking system and use of a common clock. Other performance criteria include the time to achieve formation, how vehicles respond to threats during a move and how the formation responds to a limited set of irregularities. Some example irregularities that may be introduced are vehicle breakdowns and loss of communications.

The digital workstation described in Off-Road Driving, with additional software, would ease performance evaluation of coordinated vehicle tasks.

2.4. Conclusion and Recommendations

In conclusion, we see that while there are many aspects of performance that may be measured, the effort can be reduced by incremental evaluation of performance and appropriate selection and development of test apparatus. A sequence of tests have been proposed that achieves incremental evaluation. Also, two types of test apparatus are recommended which would greatly enhance the ability to conduct evaluations.

The first apparatus is an on-board data acquisition system which is used to evaluate performance with respect to the vehicle reference frame. It consists of the following components:

1) A calibrated camera system (2 side and 1 forward looking) and video recording system that can document vehicle motions as perceived by an on-board observer.
2) An accelerometer/inclinometer/speedometer system for measuring vehicle dynamics.
3) A generic communications mechanism (possibly common memory based) to capture control system data (commanded steering, obstacles parameters, etc.).
4) A time stamping mechanism to correlate video, vehicle dynamics and vehicle control signals.
5) An off-line analysis workstation for viewing, processing and displaying results.

The second test apparatus is a global data acquisition system with digital terrain workstation which assists in evaluating vehicle performance in a wider reference frame. It consists of the following components:

1) A digital terrain database. The initial resolution of the database is from 1-5 meters. The database should be constructed so that higher resolution may be added based on new ground truth information, terrain attributes such as vegetation and soil condition, and man made objects such as buildings, signs, etc.
2) A vehicle tracking system. A system capable of tracking the position of multiple vehicles. The desired resolution should be approximately ten centimeters per kilometer (I in 10,000). The accuracy should be a meter per kilometer or better. The range of the system should encompass roughly a 2 km square. Targets should be capable of being tracked at speeds up to 100 km/hr (though the resolution and accuracy may degrade as a function of speed).

Technology for measurement systems of this scale is very limited. Two types of approaches are suggested in this report: laser tracking and GPS.

A second type of system pursues the approach that the best point of view is obtained from overhead. Aircraft have been proposed for aerial images but are costly and must continually compensate for motion. Balloons offer better platforms but have problems maintaining stability and the desired camera orientation. A novel approach is a six-cabled balloon-based viewing system which is based on Stewart crane research ongoing at NIST. A camera pod is suspended from a helium filled balloon tethered by six cables. The six cables, possibly made of dacron, provide a large degree of stability, and support a power and fiber optic video link to the camera pod. The balloon is allowed to rise up to a height of approximately 1 km. At this height, and with the appropriate camera and optics, we estimate it is possible to obtain one meter resolution over a 2 km sq. grid.

Special targets on the ground allow the image produced by the camera to be registered with ground truth. Software developed for aerial photography can be used to process the imagery to determine detailed information about terrain, vehicle positions and target positions. The time stamping system proposed for on-board data acquisition can be extended to correlate events observed from overhead with events during a mission.

3) An Analysis Workstation. A workstation system capable of displaying digital maps, overlaying vehicles positions, correlating events, and evaluating performance during the mission. In addition, software should be obtained that will allow quick documentation and publishing of results.

Chapter 3. Visual Sensing for Navigation and Driving

3.1. Introduction

These evaluations are intended to establish the performance levels of visual sensing capabilities. The capabilities range from obstacle detection, which is critical to successful driving of any kind, to landmark recognition, which is required to robustly follow a mapped course in the face of uncertainty.

The visual functions are classified as feature extraction and local mapping, and symbolic recognition. The impact of gaze control on visual perception is also considered. Feature extraction and local mapping considers separately obstacle detection and local mapping. Symbolic recognition includes terrain classification, landmark recognition, and traffic guide recognition.

Methodology: Evaluating the performance of visual perception requires the ability to manipulate the input in a way that has a predictable effect on the result. It is also crucial to test the system in the environment in which it is expected to perform, not only under ordinary circumstances but also under unusually stressful conditions. These evaluations center on calibrated manipulation of a test course. The location of every rock on the course cannot be mapped at a reasonable cost. However, it is possible to observe the effect of selectively placing known objects on the course where objects are expected or not.

A basic set of test procedures is designed to generate rich data for evaluating various visual sensing (and driving) capabilities in order to capitalize on the effort required to perform the tests.

Each of the tests should be performed in a variety of environmental conditions, such as varying light levels, cloud cover, rain, etc. In addition, the effect of varying viewing parameters (e.g., camera baseline, vergence angle, focus, aperture, zoom) should be tested. The sensor data, as well as the results, should be recorded (e.g., on synchronized video tapes) to aid further algorithm development. Similarly, sensor data should be collected in anticipation of each capability to aid development. Ground truth data should also be gathered if they are available. Finally, field data should be collected in conditions and environments ranging beyond those for which calibrated data are available. Performance under these conditions can be evaluated qualitatively. These procedures will not be detailed below for each evaluation, but they apply to all the test procedures.

A further procedure will be common to many of the tests of visual sensing. Tests that involve driving the vehicle over the test course in multiple trials will produce data sets that must be registered with one another and ground truth for analysis. To enable registration of these data, easily-recognized control points will be placed over the test course. These points will be mapped in the global map and ground truth maps and they will appear in the local maps or images produced during the test runs. The data of several trials can be registered with one another and ground truth by performing correspondence of the control points found in each data set. Each triple of control points defines a planar map segment and a related coordinate system. Map segments can be registered with one another by transforming one triple to another. The associated data sets can be related to one another or to ground truth data by the affine transformation that relates the corresponding coordinate systems. In the case of images (such as obstacle images) this approximation will be reasonable only for small affine transformations. (Particularly, the skew component must be small.) In these cases, the views to be corresponded must be very similar from the start. Therefore the images should be taken from nearly the same paths or locations on each trial.

The vehicle's performance will be compared both qualitatively and quantitatively with human performance of several tasks, such as obstacle detection, landmark recognition, self-localization, recognition of traffic guides, and gaze control. Ideally, the performance of a human driving the vehicle would be measured in the case of driving-related tasks. Unfortunately, we must rely on the human to report her perceptions. This would place too much cognitive burden on a driver, so a passive observer will be in the vehicle during the trials. Visual perception data must be registered with the other data. The observer will have to report the location of objects in the vehicle coordinate system. The best devices for recording these data are eye trackers mounted on a helmet that is itself tracked in vehicle coordinates. Trackers are available that register the gaze traces with images from a helmet-mounted video camera. (They are bore-sighted using half-silvered mirrors.) The subject can indicate detection by pressing a button while foveating (looking directly at) an object. Another possible indication method would involve projecting a cursor on a "heads-up" display and having the subject use it to indicate detection. The cursor could be controlled by a mouse or track-ball or similar device. The subject should be instructed to mark as many objects as possible; however, priority should be given to objects the subject judges most important. For instance, in an obstacle detection task the subject should distinguish untraversable obstacles, say, by double-marking them. The video record of the head-eye tracker would offer the advantage that the subject's view can be registered with data from the vehicle's visual systems for analysis by using the visible control points in the test course and the same software tools already required to integrate the vehicle's perception with the ground truth data.

Measuring performance in recognition tasks of features, such as landmarks and traffic guides, will require recording a verbal protocol of the subject naming each feature as she identifies it. The features located in the visual record will have to be subsequently tagged with the identifications from the protocol.

If the cost of such head and eye tracking systems is prohibitive, the subject might be able to report the locations of objects in azimuth or compass direction and elevation or range. A hand-held digital compass and range-finder or other sighting aid could be used in this approach. Some subject training would be required, accuracy would suffer, and perhaps reporting speed would suffer as well. Integrating these data with the location data recorded by the vehicle's perceptual systems would require considerable manual effort unless the devices used offer digital outputs that can be linked to the recording computer systems.

The data collected with hand-held instruments might also be complemented by the performance of a human subject remotely viewing video data from cameras on the vehicle. The cameras might be the vehicle's own stereo cameras under the control of the autonomous system, or they might be on a teleoperated head, slaved to a helmet worn by the subject. The subject should have a device, such as a light pen or mouse, for indicating obstacles and features. These data, though gathered under teleoperated or telepresence conditions rather than physical presence, would be accessible to the automated evaluation systems for comparison with vehicle performance.

Analysis: In addition to the analyses recommended for each evaluation, standard performance evaluation (such as is used during training exercises) should be used whenever it applies. For instance, the quality of the vehicle's path should be scored in part on its use of terrain-driving techniques to minimize the vehicle's exposure (visibility). The recommended analyses range from error rates to qualitative evaluations. Visual sensing can be especially sensitive to variation in viewpoint and other parameters, so the stability of the perception over a similar set of observations must be considered. This implies that the test procedures must collect data with small variations in parameter values to enable stability assessment. Even when error rates cannot be computed because calibrated data are lacking, the stability of a measure can be assessed over a short sequence of images. Similarly, it is important to establish the precision with which the size of an obstacle can be estimated so the hazard that it presents can be judged safely. It is also important to measure the latency of each component of visual perception systems, so the vehicle's reaction times (e.g., to avoid an obstacle) can be estimated. Finally, qualitative evaluation must be used to measure performance under conditions that cannot be controlled at a reasonable cost.

Applicability to the Demo A System: Feature extraction (such as obstacle detection using sonar) and local mapping will be a part of the Demo A system, and it will certainly be appropriate to evaluate the performance of these tasks. If additional sensing modalities (such as stereo vision) are expected to be employed in the future to augment these systems, calibrated sensor data should be collected during the evaluation of the Demo A systems. Calibrated real-time field data from a wide variety of scenes in a range of conditions will contribute invaluably to development efforts. Similarly, field data will help the development of robust symbolic recognition systems. These prerecorded data will enable the designers to test the landmark recognition components of landmark navigation systems. A limitation of prerecorded data though is the inability to support testing of active search strategies that involve camera pointing. Terrain classification data may be able to use the imagery used for feature extraction, so the only additional information required for a terrain classification database is the ground truth terrain map. Traffic guide data are likely best collected in the course of ordinary maneuvers, so there should be ample opportunity to collect field data at any time it is required. Gaze control evaluation is only warranted in the evaluation of systems that dynamically direct their sensors. Further, the performance of the gaze control component of such a system is implicit in the overall performance of the system; thus, if system performance is adequate, there may be no need to evaluate the gaze control component of the system. In summary, it is suggested that unique opportunities should be exploited to collect calibrated data for any system that is expected to be deployed in the near future in addition to currently functioning systems; the evaluation of the Demo A system may present such an opportunity.

3.2. Feature Extraction and Local Mapping

3.2.1. Obstacle Detection

3.2.1.1. Objective

Measure and document the performance of obstacle detection systems. Determine detectability of classes of obstacles on- and off-road and with the vehicle stationary and moving. Observe obstacle detection in a wide range of conditions. Systems are to detect obstacles and potential obstacles. Discrete obstacles, such as rocks, trees, bushes, posts, poles, potholes, puddles, structures, vehicles and animals should be detected. Extended obstacles (e.g., ditches, ridges, embankments, and steep slope changes) should also be identified. Parameters of the obstacles, such as the slope of an embankment, should be estimated.

3.2.1.2. Test procedures

1. Obstacle detectability from a stationary vehicle

With the vehicle stationed in a variety of environments, on- and off-road, place various sized obstacles at various locations. Obtain ground truth for the obstacles by surveying the locations, and record any parameters of each obstacle. Using each of the available sensing systems (e.g., daylight stereo, infrared stereo, sonar, LADAR) record obstacle detections and any parameter estimates for each detected obstacle. For an obstacle that is not easily fabricated and placed at will, position the vehicle at various distances and orientations to the obstacle.

2. Obstacle detection from a moving vehicle

The vehicle will be run repeatedly over the same path on a test course. In addition to any natural obstacles on the course, artificial obstacles will be placed on the course for some trials. Some of these object locations should be noted in the global map to enable testing of all four combinations of obstacles being present or absent where they are mapped or not. The results of multiple trials must be integrated, so there must be a means of registering the results of the runs with one another. For this purpose, easily-recognized control points can be placed along the course and views can be corresponded as described in the introduction of this chapter.

Run the vehicle over the test course several times to record baseline obstacle detection. Add objects that will not require altering the vehicle's path and rerun the course, recording obstacle maps. Make several runs with various combinations of these obstacles. Add obstacles in the vehicle's path, and make several runs with combinations of these obstacles and the other objects. In some trials in which no object is placed as an obstacle in the canonical path, the vehicle should be forced to follow the path to ensure viewpoint consistency. In other trials, the vehicle should be permitted to deviate from the canonical path. This provides richer data for stability assessment. If the vehicle is allowed to choose its path autonomously, this also enables evaluation of its response to objects near its path. Deviation is certainly to be permitted in trials in which obstacles are placed in the canonical path.

Again, for an obstacle that is not easily fabricated and placed, drive the vehicle on a variety of paths near the obstacle. Again record the sensor data. Be aware that during repeated runs off-road (e.g., in mud), the vehicle's own tracks may alter the scene.

3. Human performance

Record obstacles reported by human subjects during the trials, using appropriate methods as discussed in the introduction. The subject should be instructed to mark as many obstacles as possible in real-time; however, priority should be given to objects the subject judges untraversable and they should be distinguished, say, by double-marking.

3.2.1.3. Analyses

1. Obstacle detection

For each class of artificial object and extended natural object, collect statistics such as miss rates and false alarm rates. These will allow calculation of signal-to-noise ratios. SNRs will indicate how many "hits" are required to reliably indicate an obstacle. SNRs also indicate the smallest obstacle of a given class that can be detected at a given range.

a. miss rates (M (c,s,r))

Percentage of controlled objects that were not detected, for each class, size and range of obstacle.

b. false alarm rates (F(c,s,r))

Percentage of the samples in which an obstacle is reported when none is present. Collate according to class (if known), size and range of the illusory obstacle. The false alarm rates of controlled objects can be found automatically since ground truth is known; however, calculating the false alarm rates of unsurveyed objects would require human determination of ground truth. It may be necessary to pool the false alarm responses to estimate an overall false alarm rate.

c. signal-to-noise ratios (SNR(c,s,r))
The signal-to-noise ratio, measured in decibels, is defined for each class of obstacle, of each size, at each range, as a function of the miss and false alarm rates, M and F:

$$\mathrm{SNR}(c,s,r) = 20\log_{10}\frac{\text{signal}}{\text{noise}}\,\mathrm{dB}$$

$$= 20\log_{10}\frac{1-M}{F}\,\mathrm{dB}$$

Here, the signal is defined as the detection rate (1-M), and the noise is the false alarm rate (F).

d. stability (S(c,s,r,...))
Examine these statistics for stability in each neighborhood of object and viewing parameters. E.g.,

$$\mathrm{Stability}_{\mathrm{SNR}}(c,s,r) = \frac{1}{1+\left(\dfrac{\sum_{i=1}^{n}\left|\mathrm{SNR}(c,s,r)-\mathrm{SNR}(c\pm i,s\pm i,r\pm i)\right|}{\mathrm{SNR}(c,s,r)}\right)}$$

will have a value near one if variability is small and near zero if variability is large.

2. Comparison with human performance
Analyze the human-detected obstacle data for miss and false alarm rates and SNR as above. Compare the statistics with those of the automatic systems. Be aware that a human may tend to mark only obstacles she considers untraversable. Also compare the obstacles reported by the vehicle systems with the human data, using the human set as a ground truth measure of traversability.

3. **Parameter estimation**

If a property of an obstacle is estimated, assess the accuracy and stability of the estimate. E.g., compare the mean estimate with ground truth, and measure the variance of the estimate over a reasonable neighborhood of viewing parameters and property values.

4. **Hazard estimation**

Analyze the property estimation to determine whether resolution is sufficient to distinguish true hazards from harmless obstacles (e.g., can the size of a rock be estimated with enough resolution to allow judgement of whether it must be avoided?)

5. **Qualitative analysis**

Qualitatively judge the performance of the system. Data visualization will be critical to this task. Obstacles might be projected onto the best available map or overlaid on a stereo video display for this qualitative evaluation.

3.2.2. Local Map Accuracy

3.2.2.1. Objective

Characterize the accuracy of the local mapping used for vehicle driving. The local map may be a single map derived from visual sensing and other cues, or it may simply be a collection of such sensory maps. The local map should register well with the global map on which the route plan is described. The local map must accurately locate features whether or not they appear in the global map. For example, features smaller than the resolution of the global map will not appear in it, and it is essential that they be contained in the local map. A ground truth map will be constructed for each trial that represents the placement of controlled and surveyed objects on that trial. Each local map will be evaluated by comparison with the trial's ground truth map.

3.2.2.2. Test Procedures

1. **Control points for data registration**

Control points will be used to assess the correspondence between the maps. Control points that are easily detected will be placed in the test environment. They will be mapped in the global map and in the local map during the test runs. Make several runs using artificial features both present and absent in both mapped and unmapped locations as described for obstacle detection. The controlled objects and the manually surveyed objects (such as embankments) will be described in a ground truth map of each trial. The ground truth map will always accurately reflect the locations of the objects marked in it, even though the global map may fail to contain small features and it may describe features that are not actually present.

2. Human mapping
Record maps of features reported by human subjects, using appropriate techniques as discussed in the introduction. The subject should map as many features as possible. It is likely that the human will report only features that seem prominent. It is hoped that these features will be on a smaller scale than the global map and that they can be used to fill in details left vague in the global map and not surveyed in the ground truth map.

3.2.2.3. Analyses

1. Map registration
Measure the error in registering the local map with the ground truth and global maps. Control points will be located in each local map, and they will be corresponded with the control points in the global and ground truth maps. A triple of control points defines a planar map segment and a related coordinate system. Errors in map registration can be characterized by the affine transformation that relates the corresponding coordinate systems.

2. Feature locations
Given the transformation between the planar segments of the map, find the correspondence between features of the local vehicle-generated map and global, ground truth, and human maps. (Features in the global, ground truth or human maps will be called "mapped" and those in the local vehicle-generated map will be called "detected".) Measure placement errors for mapped and detected features. Log misses (mapped but undetected features) and log and flag potential false positives (detected but unmapped features) to be inspected visually.

Have a human evaluator classify unmapped detected features as either valid features of the environment or spurious detections, using a display of the local map projected onto the sensor images (in 3D, if it is possible). Display the validated maps graphically for visual inspection and evaluation. (E.g., display wireframe models of the maps in 3D or perspective view. Overlay the local, global and ground truth maps in different colors, to clearly show missed features and unmapped detected features. Again, if it is possible, project the maps on appropriate sensor images if it aids interpretation.) Collect statistics as described in the section above on obstacle detection:

a. miss rates
Percentage of mapped features that were not detected.

b. false positives
Percentage of "detected" features that were judged to be illusory.

c. SNRs
For each class of feature, parameterized by size and range.

d. stability
Calculate the stability of each feature class as a function of feature size, range, and any other relevant viewing parameters, as described above for obstacle detection.

e. location accuracy
Compute the accuracy of localization of detected features that can be corresponded with the global or ground truth maps. E.g., analyze the data to discover directional biases in localization, and calculate the root mean squared error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

for n error measures x_i.

3.3. Symbolic Recognition

3.3.1. Terrain Classification

3.3.1.1. Objective

Measure and document the ability of the system to identify and distinguish terrain. Terrain and features should be classified correctly, and the system should attempt to distinguish hazardous terrain and features.

3.3.1.2. Test Procedures

1. Collect data

Run the vehicle several times over the test course and record the sequences of local terrain and feature maps.

2. Ground truth

Obtain ground truth terrain and feature maps manually with the assistance of a human expert. Include the identification of hazardous features and terrain. These might be obtained by online data recording as described in the introduction or by off-line traditional mapping techniques.

3.3.1.3. Analyses

1. Terrain classification

Compare the system's classification to the ground truth classification of the human expert. Gather statistics as before:

a. correct classification rates

Percentage of classifications that match the ground truth.

b. error rates

Percentage of classifications that differ.

c. miss rates

Percentage of classifications that should have been judged hazardous according to the ground truth but were not.

d. false alarm rates

Percentage of classifications that were falsely judged to be hazardous.

e. SNRs
As a function of correct (signal) and erroneous (noise) classifications for each terrain class.

f. Stability
Estimate the stability of each terrain and feature class and of hazard judgements as a function of the SNR in neighboring observations. The stability should be parameterized at least by terrain class and range to the patch. It might also be parameterized by the size of the patch.

g. Confusions
Collate the confusions among the terrain and feature types. These are commonly represented by a matrix. Assess the danger of these confusions if they are ignored. E.g., what hazards would the vehicle attempt to drive through? Assess the operational restrictiveness of interpreting the terrain classifications conservatively (by making use of the confusion matrix). E.g., what proportion of safe paths will not be attempted because they have mistakenly been judged hazardous?

3.3.2. Landmark Recognition

3.3.2.1. Objective

Determine what classes of landmarks can be recognized, and how accurately and robustly the system can localize itself using them. In particular, establish how well the vehicle can localize itself relying only on visual landmark identification in case other external reference systems (such as GPS and LORAN) become unavailable. The system should be able to localize itself within bounds that are theoretically achievable given the locations of recognizable landmarks and the expected error in localizing a landmark.

3.3.2.2. Test Procedures

1. Landmark Recognition
Place the vehicle on a test course at a known start position and orientation, and inform the vehicle of these data. Drive

the vehicle over the course and log all recognized landmarks. Each recognition will be classified by an observer as either correct or erroneous.

2. Self-localization

The vehicle's self-localization based solely on landmark recognition should be recorded to be compared later with the true vehicle path.

3. Landmark-based navigation

a. Landmark recognition

Test the system's ability to accurately recognize visual landmarks that are used in route plans. Drive the vehicle over the test course. Record the system's ability to recognize the indicated landmarks along the route.

b. Visual route following

Test the ability to follow a course using only landmarks to locate the vehicle's position on its map (i.e., using only incremental devices such as INS and odometry with visual landmark self localization, but without other external reference systems such as GPS). Disable the external reference systems and let the vehicle attempt to follow the route plan autonomously. Record the vehicle's recognition of landmarks as it proceeds.

4. Human performance

Record human performance in landmark recognition, self-localization, and landmark-based navigation. For this task, it is probably sufficient for the subject to employ the usual manual navigation techniques and record a verbal protocol of all landmarks identified in the process. This test should be performed under two conditions: with the subject riding passively and recording navigation performance while the vehicle is driven along the controlled test path, and again with a human team navigating and driving the vehicle. Nevertheless, additional data gathered using head and eye trackers would be available for automated comparison with vehicle performance data. If the traditional tools can make digital output available for automatically recording manual navigation, that will aid comparison of vehicle performance with human performance under natural conditions.

3.3.2.3. Analyses

1. Landmark recognition

Calculate landmark recognition success and robustness measures

a. correct identification rates

Percentage of landmarks correctly identified.

b. error rates

Percentage of landmarks erroneously labeled.

c. miss rates

Percentage of landmarks missed.

d. false positive rates

Percentage of landmarks "identified" that do not appear in the map.

e. SNRs

The signal-to-noise ratio of each landmark as a function of correct and erroneous identifications.

f. stability

Stability of recognition of each landmark from neighboring viewpoints.

g. recognition region

Extent (in angle and range) over which each recognized landmark was correctly identified. The result is a map region for each landmark in which it was recognized during the test.

h. landmark classification

Identify classes of landmarks that can be robustly recognized.

2. Self-localization

Compute measures of self-localization and route plan landmark identification. Route plan landmarks are landmarks given explicitly in the route plan to augment the route description.

a. error
Root mean squared error (RMSE) of self-localization relying on landmarks. Also study the errors for directional biases.

b. miss rates
Percentage of route plan landmarks missed.

c. missed transitions
Percentage of turns or other transitions missed in the course of attempting to follow the route.

d. increase in travel time
Increase in time to reach each landmark or control point along the course when relying only on landmarks for self-localization (vs. landmarks and other external reference systems such as GPS or LORAN).

3. Landmark-based navigation
Compute measures of navigation performance, such as time to reach each control point or landmark, percent of landmarks recognized along the way, deviation from the optimal path, and perhaps a qualitative assessment of goodness of the path taken.

4. Comparison with human performance
Calculate statistics of human landmark recognition and self-localization performance and compare with vehicle performance. Compare the paths taken by the vehicle under autonomous and manual navigation.

3.3.3. Traffic Guides

3.3.3.1. Objective

Document the ability to recognize traffic guides. The system should recognize traffic guides in ample time to take the appropriate action.

3.3.3.2. Test Procedures

1. Test data

Compile a suite of data of traffic guides that can be expected during ordinary vehicle movements and under extreme conditions. (These data should probably be collected on videotape or in long sequences of digital images.) Present the data to the traffic guide recognition system and log the results.

2. Human performance

Record the performance of a human subject during the taping of traffic guides, using the techniques discussed in the introduction for recording the results.

3. Field test

Verify system operation on a test course.

3.3.3.3. Analyses

1. Accuracy and stability

Analyze the results for accuracy and stability of recognition of the guides.

2. SNRs.

3. Recognition range

Determine the maximum range at which correct recognition can be expected.

4. Confusion
Compile the confusion matrix and examine it for indications of potentially dangerous problems.

5. Comparison
Compare the system's performance with human traffic guide recognition.

3.4. Gaze Control

Gaze control evaluation is only warranted in the evaluation of systems that dynamically direct their sensors. Further, the performance of the gaze control component of such a system is implicit in the overall performance of the system; thus, if system performance is adequate, there may be no need to evaluate the gaze control component of the system. In case evaluation of gaze control seems appropriate, some suggestions follow.

3.4.1. Objective

Determine the adequacy of gaze control to incorporate observations from varying gaze directions and to provide coverage of the visual space to enable proper visual perception. Accuracy of visual mapping should not suffer due to the use of steerable cameras. Information that is gathered and integrated from various visual gaze directions should have similar accuracy if it is acquired from a single steerable camera system or from a cluster of fixed cameras. Similarly, no region of visual space should be neglected. If the system lacks the ability to process visual data from the entire viewing sphere at all times, attention must be turned to each area of visual space enough to permit an adequate assessment of the vehicle's situation at all times.

3.4.2. Test procedures

1. Gaze traces
During tests of obstacle and landmark mapping, note the involvement of gaze control, tracing gaze during runs of the test course.

2. Human gaze
Instrument the driver of the vehicle in the manner described in the introduction, and record her gaze as she drives the vehicle

over the test course. Another useful record might be the gaze trace of a teleoperating driver. Have a driver teleoperate the vehicle over the test course using the same sensors and a reasonable gaze control interface. A natural interface would be a helmet-mounted stereo display system with the gaze servoed to the helmet orientation. Trace the resulting gaze of the vehicle sensors.

3.4.3. Analyses

1. **Localization accuracy**
   Compare the localization accuracy of mobile high resolution sensors with localization by fixed-gaze low resolution sensors. Determine the trade-off of visual resolution and motor resolution.

2. **Gaze coverage**
   Analyze the vehicle's gaze control to identify areas of inadequate attention. Correlate density of gaze with success of visual perception tasks (obstacle detection, terrain classification, and landmark and traffic guide recognition).

3. **Comparison**
   Compare the coverage of gaze over the test course by the automatic, manual and teleoperated systems to identify significant differences.

3.5. Acknowledgments

These evaluation recommendations have benefited from discussions with Bob Bolles of SRI, Larry Matthies of JPL, and Keith Nishihara of Teleos, in which they shared their ideas on visual perception evaluation based on their experience with evaluating the performance of stereo vision algorithms.

Chapter 4. Reconnaissance, Surveillance and Target Acquisition

4.1. Introduction

Measurement helps to frame results so that managers can understand the information a system generates. It provides quantitative values for estimating the capabilities and limitations of a system under a variety of circumstances and leads to proper appreciation of the system's accomplishments. Reconnaissance, surveillance, and target acquisition (RSTA) are challenging problems for unmanned ground vehicles(UGVs). The measurements of RSTA are even more difficult, because of the wide range of environments in which RSTA needs to be performed.

In this chapter, we suggest an engineering methodology for determining the capabilities and limitations of a system under a variety of circumstances. The first section addresses the necessity of RSTA evaluation. Section 2 presents RSTA system requirements and evaluation goals and objectives. In section 3 we provide an initial perspective on RSTA metrics specifically aimed at UGV Demo II and it's constituent demos. Finally, we present a methodology and a detailed test plan for evaluating the RSTA operational performance for UGVs.

4.2. RSTA system requirements and Evaluation Objectives

There are two categories for evaluation in RSTA, technical performance and operational performance. Surprisingly little work in the RSTA community has dealt with operational performance measurement. Most literature has concentrated on evaluating technical performance. For example, the Night Vision Lab (NVEOD), Wright Lab, and other organizations [1 ,2 ,3 ,5, 11, 12, 13, 14, 16, 17, 18, 20, 23] have been developing target detection and recognition performance evaluation methodologies. This section focuses on operational performance, which is the ability of the system to meet its operational requirements. In order to evaluate operational performance, it is first necessary to know the system operational requirements. Based on the RSTA program plan and approach developed in July 1992 for the ARPA Demo II program [10], the system operational requirements include the following functions:

- RSTA from a stationary platform: This function detects stationary or moving targets from a range of 200 m to 3 km when targets are visible along a line of sight less than 50% obscured by terrain. This function involves operations such as acquisition and detection, but not recognition.

- RSTA from a moving platform. This function performs reconnaissance, surveillance and target acquisition within a range of 200 m to 3 km when the speed of the vehicle is up to 35 m.p.h. (56 km.p.h.). This function involves operations such as acquisition and detection, but not recognition.

- Target recognition. This function recognizes targets within a range of 200 m to 3 km.

- RSTA teleoperation. This function allows the operator to visually inspect and verify detection, tracking and recognition outputs and allows the operator to point and control all sensors.

- Target designation. This function allows the operator to initiate a designation sequence with eye-safe training laser designators for stationary and moving targets.

- Intelligent search. This function uses terrain and doctrinal knowledge to sequence sensor pointing and acquire the targets in the scene.

With the description for the RSTA system requirements, we are in a position to define the objectives of RSTA performance evaluation. The success of demonstrating the completion of operational requirements depends on RSTA performance metrics. Performance evaluation can help frame results so that a variety of audiences can understand the information the system generated. It provides quantitative methods for estimating the capabilities and limitations of RSTA under a variety of circumstances. Based on the operational requirements and the audience for which they are intended, the objectives of RSTA performance evaluations include:

1. Developers
- Mark the internal progress of the RSTA system.

- Focus attention on areas of potential difficulties and opportunities.

- Provide quantitative methods for estimating the capabilities and limitations of RSTA under a variety of circumstances.

- Provide measurements for following the program progress with respect to Department of Defense (DOD) management expectations.

2. Department of Defense (DOD)
- Predict the effectiveness of RSTA under future battlefield conditions.

- Address the concern of certain DOD constituencies about the integration of RSTA with the existing organization's technology infrastructures.

4.3. Metrics for RSTA System Evaluation

In actual operation, RSTA functions may need to be performed in a wide range of environments which can affect performance. Thus, RSTA operational evaluation must include not only the evaluation of operational requirements, but also environmental complexity. The complexity of the environment includes the following variety of situations:

- dust
- exhaust
- temperature
- fire and smoke
- fog, haze, rain
- clouds and cloud shadow
- wind
- tree, bush and grass movement
- target-like clutter
- animal movement
- IR reflections
- partial target obscuration
- low target to background contrast
- wide variety of target types
- wide range of apparent target velocities and acceleration
- wide range of apparent target sizes and aspect angles
- large number of targets
- moving sensors

A measure of scene complexity should be based on metrics such as:

- weather conditions
- terrain types and conditions
- amount and types of clutter
- amount of target maneuvers (position, velocity, acceleration)
- number of targets and target type
- complexity of targets (target to background contrast, target sizes, target occlusion)

These may be combined into two measures of scene complexity: overall scene complexity and overall target complexity. These metrics can be obtained by using human judgment or automated means. Once the complexity of a scene has been determined, the

RSTA system performance on that scene can be evaluated. In order to evaluate the RSTA system performance against the operational requirements, metrics such as the following are proposed.

- Effectiveness of RSTA for a given environmental condition. This metric assesses the ability and efficiency of the system to perform given tasks.
- Robustness of RSTA. This metric assesses the system effectiveness over a range of environmental conditions.
- Accuracy of RSTA under a given environmental condition. This metric measures the extent to which the function performs correctly by obtaining the scoring of the distances/errors from ground truth.

The measurement can be presented in tabular or graphical forms.

4.4. Methodology for RSTA Technology Evaluation

This section presents an engineering approach for evaluating the RSTA operational performance for UGVs. In order to evaluate end-to-end UGV operational performance during the UGV demos, a comparability approach is a necessary tool. This approach is at a level of abstraction that is appropriate for developers/ customers to understand the capabilities and limitations of the presented system in actual operations. The comparability approach includes a comparison with human performance and a comparison with ground truth. The intent is to produce a picture of a system's quantitative and qualitative performance to customers and developers during demos and development.

4.4.1. Environmental Conditions

4.4.1.1. Evaluation objectives

The objective of evaluating environmental conditions is to evaluate the performance of RSTA systems as a function of different kinds of scenes.

4.4.1.2. Evaluation test procedure

The evaluation procedure includes measuring the scene complexity and the target complexity. These are described next.

4.4.1.2.1. Scene complexity measurement

(a) Measure weather conditions: monitor wind speed, amount of dust, exhaust, smoke, fire, fog, haze, rain, clouds, shadow, sun location and daylight/night in the sensor viewing area, which is the region in the line of sight of the sensor.
(b) Measure terrain conditions: Survey the terrain geometry (e.g., hills, rocky, flat) and terrain composition (e.g., vegetation, soil) before tests and obtain a high resolution digital map. The terrain geometry and composition in each sensor viewing area can be estimated from the map by the translation between the pointing angles of the sensors and the planar coordinates of the ground.

4.4.1.2.2. Target complexity measurement

(a) Measure target range, target aspect angle, target velocity and acceleration in each sensor viewing area.
(b) Measure the amount and types of clutter (e.g., trees, bushes, rocks) in each sensor viewing area.
(c) Obtain the number of targets and target types in each sensor viewing area.
(d) Measure target specific complexity. This measurement is correlated with the detectability of targets and deals with features of the target which are discriminated from natural clutter in the scene. Clark and Velten[3] suggested several measurements such as target standard deviation, pixels on target, perimeter squared over area, average target edge strength, and target edge standard deviation. The details on the computations for the measurements can be found in [3].

4.4.1.2.3. Evaluation analytical procedure

The above measurements for scene and target complexities are not easily obtained in complex environments. However, a relatively small number of measurements taken with some degree of precision is possible. Which measurements of environmental conditions are "critical" from the perspective of a RSTA system is dependent on the type of tasks and missions. For example, the measurements of smoke and fire are generally important in a battlefield situation. On the other hand the measurement of wind speed is not generally important in the field situation. Hence, developers can choose the proper measurements from the above to evaluate scene and target complexities.

In order to obtain overall scene (target) complexity, Gilfillan[7] and Noah et al.[13] suggested that a weighted average of the scene(target) measurements should be used to express scene (target) complexity.

4.4.2. RSTA from a Stationary Platform

4.4.2.1. Evaluation objectives

The objective is to assess the performance of RSTA in detecting stationary or moving targets within a range of 200 m to 3 km when targets are visible along a line of sight less than 50% obscured by terrain. The RSTA sensors are on a stationary platform.

4.4.2.2. Evaluation test procedure

The evaluation procedure includes the following:

1. Ground truth. Survey terrain geometry, terrain composition, positions and types of targets throughout tests. Procedures and instrumentation for obtaining ground truth are described in Chapter 6. The ground truth will be obtained before tests are performed.

2. Place targets within a range of 200 m to 3 km.

3. Record the following data obtained using the RSTA system:
   (a) time at which the following data are obtained
   (b) video tapes of target imagery. (The video information can be used for measurement verification and program debugging.)
   (c) Targets:
      (c1) positions, orientations, velocities and acceleration of detected targets
      (c2) the types, number and sizes of detected targets and presented targets
      (c3) radiometric measurements of the temperature of the target and background
   (d) Scenes:

(d1) atmospheric conditions (dust, exhaust, temperature, fire and smoke, fog, haze, rain, clouds and cloud shadow, wind)
(d2) terrain composition, description and conditions

4. Soldier performance. Each trained soldier will be seated in the driver's seat of the HMMWV while the RSTA system is operating. The soldier will perform the target detection mission as if he were in a battlefield. The soldier will record the observed information listed in item 3 above.

5. Repeat steps 1, 2, 3 and 4 for several different days and different environmental conditions.

4.4.2.3. Evaluation analytical procedure

1. Measure scene complexity (see 4.4.1.2.3.).

2. Measure target complexity (see 4.4.1.2.3.).

3. Measure function performance.
   - Measure effectiveness. The effectiveness of the function involves detection rates and false alarm rates with respect to an individual parameter of the scene or target. An example is shown in Figure 4.1.
   - Measure robustness. The robustness of the function involves calculating the detection rate and false alarm rates with respect to scene or target complexities. An example is shown in Figure 4.2.
   - Measure accuracy. The accuracy of this function involves calculating the magnitude of correctness of parameters with respect to scene complexity. The parameters for the function include location, velocity, and acceleration.
   - Correctness is determined by comparing the calculated positions/velocities/accelerations with the actual positions/velocities/accelerations of targets.
   - The recorded data in (a) and (b) can be used for measurement verification and program debugging.

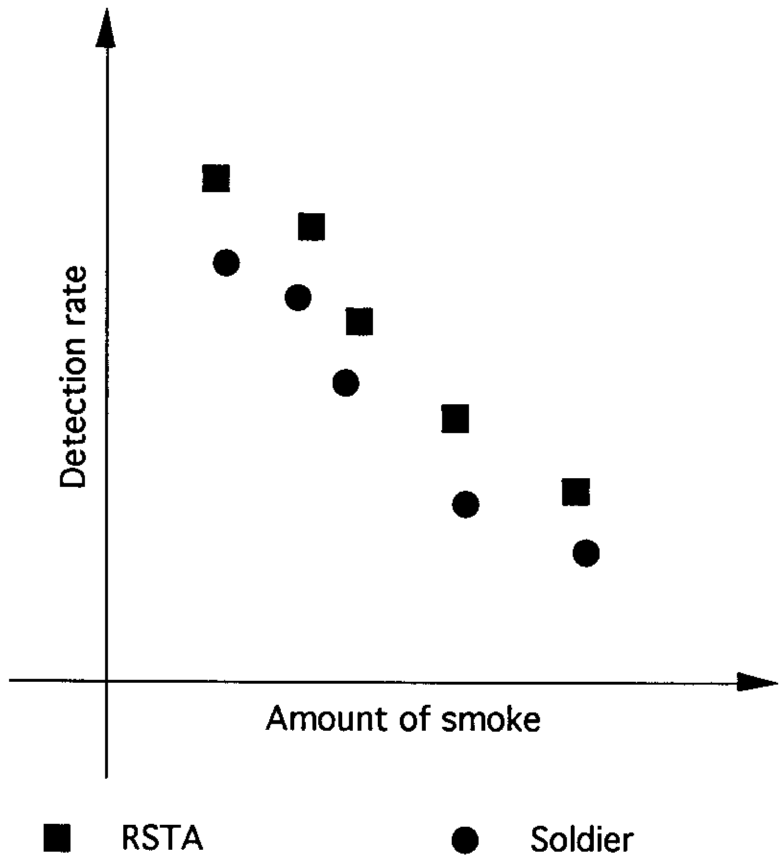

Figure 4.1. Example of effectiveness: detection rate with respect to a scene parameter (e.g. smoke).

	False Alarm		Detection	
	Rates	Deviation	Rates	Deviation
Soldier				
RSTA				

scene complexity = 1.0

Figure 4.2. Example of robustness: false alarm rate and its standard deviation and detection rate and its standard deviation with respect to scene complexity. (The rating of scene complexity may be assigned a value between 0 and 1; 0 = very simple, 1.0 = very complex.)

4.4.3. RSTA from a Moving Platform

4.4.3.1. Evaluation objectives

The objective is to assess the capability of reconnaissance, surveillance and target acquisition within a range of 200 m to 3 km when the speed of the vehicle is up to 35 mi/hr (56 km/hr).

4.4.3.2. Evaluation procedure

The evaluation procedures and evaluation are the same as in section 4.4.2.

4.4.4. Target recognition

4.4.4.1. Evaluation objectives

The objective is to assess the capability of recognizing targets within the range of 200 m to 3 km.

4.4.4.2. Evaluation procedure

The evaluation procedure includes the following:

1. Ground truth. (See 4.4.2.2)

2. Use of a wide variety of targets within a range of 200 m to 3 Km.

3. Record the following data obtained using RSTA system:
   - (a) time at which the following data are obtained
   - (b) video tapes of target imagery
   - (c) Targets:
     - (c1) positions, orientations, velocities and acceleration of detected targets in UTM Coordinate system
     - (c2) the types, number and sizes of detected targets and presented targets
     - (c3) radiometric measurements of the temperature of the target and background
     - (c4) amount of obscuration (tree, bush, animal grass movement, target-like clutter)
   - (d) scenes:

(d1) atmospheric conditions (dust, exhaust, temperature, fire and smoke, fog, haze, rain, clouds and cloud shadow, wind)
(d2) terrain composition, description and conditions

4. Repeat steps 1, 2 and 3 for several different days and different environmental conditions.

4.4.4.3. Evaluation analytical procedure

1. Measure scene complexity (see 4.2.1.2.3.)

2. Measure target complexity (see 4.2.1.2.3.)

3. Measure function performance

   - Measure effectiveness. The effectiveness of this function involves calculating the classification rate with respect to an individual parameter of the scene or target.
   - Measure robustness. The robustness of this function involves calculating the classification rate with respect to scene or target complexities.
   - Measure accuracy. The accuracy of this function involves calculating the magnitude of correctness of parameters. The parameters for this function include the correctness of target type/size. This is determined by the magnitude of the similarity between the classified type/size and actual type/size of the target.
   - The recorded data in (a) and (b) can be used for measurement verification and program debugging.

4.4.5. RSTA Teleoperation.

This section could be applied to Demo A.

4.4.5.1. Evaluation objectives

The objectives are to assess the operator's capability for inspection and verification of target recognition and tracking and to examine the effectiveness and efficiency of operator-sensor control.

4.4.5.2. Evaluation procedure

The evaluation procedure includes the following:

1. Ground truth. (See 4.4.2.2)

2. Use of targets within a range of 200 m to 3 km.

3. Each operator inspects the RSTA sensor output and performs target search from the OWS. During the overall RSTA teleoperation testing, the following data are recorded:
   (a) time at which the following data are obtained
   (b) video tapes from the operator view and sensor view. (This information can be used for measurement verification and program debugging.)
   (c) operator tracking error history. (The error is defined by the distance between the actual target position and the display position indicated by the operator.)
   (d) sensor field of view.
   (e) targets:
      (e1) positions, orientations, velocities and acceleration of detected targets.
      (e2) the types, number and sizes of detected targets and presented targets
      (e3) radiometric measurements of the temperature of the target and background
   (f) scenes:
      (f1) atmospheric conditions (dust, exhaust, temperature, fire and smoke, fog, haze, rain, clouds and cloud shadow, wind)
      (f2) terrain composition, description and conditions

4. Repeat steps 1, 2 and 3 for several different days and different environmental conditions.

4.4.5.3. Evaluation analytical procedure

1. Measure scene complexity (see 4.2.1.2.3.).

2. Measure target complexity (see 4.2.1.2.3.).

3. Measure function performance.

- Measure effectiveness. The effectiveness of the function involves detection rates, false alarm rates, operator tracking error or sensor field of view with respect to an individual parameter of the scene or target.
- Measure robustness. The robustness of this function involves the detection rates, false alarm rates, operator tracking error or sensor field of view with respect to scene or target complexities.
- Measure accuracy. The accuracy of this function involves calculating the magnitude of correctness of parameters. The parameters for the function involve the number of targets, target location, velocity and acceleration.
- The recorded data in (a) and (b) can be used for measurement verification and program debugging.

4.4.6. Target Designation.

4.4.6.1. Evaluation objectives

The objective is to assess the performance of an operator to initiate a designation sequence with eye-safe laser designators for stationary and moving targets. Our main measure will involve the angular error between the ray from the laser designator to the actual target position and the ray representing the laser beam of the designator. (See Figure 4.3.)

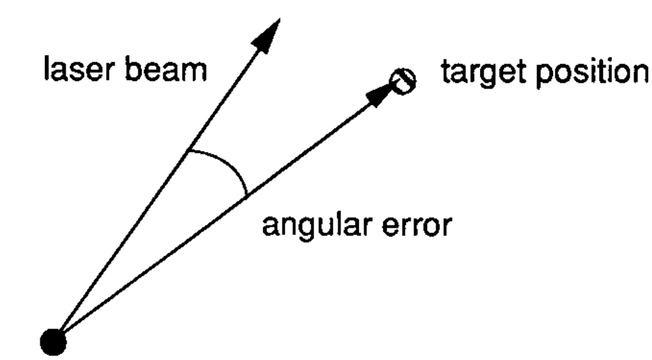

Figure 4.3. Angular error

The evaluation procedure includes the following:

1. Ground truth. (See 4.4.2.2)

2. Use of targets within a range of 200 m to 3 km. The speed of moving targets will be within a range of 0 mi/hr to 35 mi/hr (or 0 km/hr to 56 km/hr)

3. Angular error. (See above.)

4. Record the following data obtained during the target designating testing:
   - (a) time at which the following data are obtained.
   - (b) video tapes of target imagery.
   - (c) targets
     - (c1) positions, orientations, velocities and acceleration of detected targets.
     - (c2) the types, number and sizes of detected targets and presented targets
     - (c3) radiometric measurements of the temperature of the target and background
   - (d) scenes
     - (d1) atmospheric conditions (dust, exhaust, temperature, fire and smoke, fog, haze, rain, clouds and cloud shadow, wind)
     - (d2) terrain composition, description and conditions
   - (e) Operator data:
     - (e1) angular error.
     - (e2) operator training times to use the target designator.

5. Repeat steps 1, 2, 3 and 4 for several different days under different environmental conditions.

4.4.6.2. Evaluation analytical procedure

1. Measure scene complexity. (see 4.2.1.2.3.)

2. Measure target complexity. (see 4.2.1.2.3.)

3. Measure the function performance.
   - Measure effectiveness. The effectiveness of the function involves hit rates and miss rates with respect to scene or target complexities or operator training time.
   - Measure robustness. The robustness of this function involves hit rates and miss rates with respect to scene or target complexities.

- Measure accuracy. The accuracy of this function involves calculating the magnitude of the angular errors with respect to target complexity or scene complexity.
- The recorded data in (a) and (b) can be used for measurement verification and program debugging.

4.4.7. Intelligent Search.

4.4.7.1. Evaluation objectives

The objective is to assess the performance of the RSTA intelligent search function using terrain and doctrinal knowledge to sequence sensor pointing.

4.4.7.2. Evaluation procedure

The evaluation procedure includes the following:

1. Ground truth. (See 4.4.2.2)

2. Use of targets within a range of 200 m to 3 km.

3. Record the following data obtained using the RSTA system:
   (a) time at which the following data are obtained
   (b) video tapes of target imagery.
   (c) targets:
      (c1) positions, orientations, velocities and acceleration of detected targets.
      (c2) the types, number and sizes of detected targets and presented targets
      (c3) radiometric measurements of the temperature of the target and background
   (d) scenes:
      (d1) atmospheric conditions (dust, exhaust, temperature, fire and smoke, fog, haze, rain, clouds and cloud shadow, wind)
      (d2) terrain composition, description and conditions
   (e) area covered.
   (f) search time required.

4. Repeat steps 1, 2 and 3 but with several different days and different environmental conditions.

4.4.7.3. Evaluation analytical procedure

1. Measure scene complexity (see 4.2.1.2.3.).

2. Measure target complexity (see 4.2.1.2.3.).

3. Measure the function performance.

   - Measure effectiveness. The effectiveness of the function involves target search time, detection rates and false alarm rates with respect to an individual parameter of the scene or target.
   - Measure robustness. The robustness of this function involves target search time, detection rates and false alarm rates with respect to an parameter in scene or target complexities.
   - Measure accuracy. The accuracy of this function involves calculating the magnitude of correctness of parameters. The parameters for the function include target locations, velocities and accelerations.
   - The recorded data in (a) and (b) can be used for measurement verification and program debugging.

4.5. References

[1] B. Bhanu, Evaluation of automatic target recognition algorithms, Proceedings of SIP, Vol. 435, pp. 18-27, 1983.

[2] B. Bah and T. L. Jones, Image understanding research for automatic target recognition, Proceeding of DARPA image understanding workshop, pp. 249-54, 1992.

[3] L. G. Clark, V. J. Velten, Image characterization for automatic target recognition algorithms, Proceeding of SIP, Vol. 1310, Signal and Image processing system performance evaluation, pp. 68-79., 1990

[4] T. M. Conte and W. W. Hwu, Benchmark characterization, IEEE Computer, pp. 48-55, Jan., 1991.

[5] D. Duncan, Unified target modeling, validation and ATR performance evaluation, Proceeding of SIP, Vol. 1310, Signal and Image processing system performance evaluation, pp. 174-177., 1990

[6] R. Dutta and M. A. Snyder, Robustness of correspondence-based structure from motion, Proceedings of the International Conference on Computer Vision, pp. 106-110, 1990.
[7] L. Gilfillan, Image understanding evaluation metrics and methods, final report, 1992.
[8] M. Lahart, T. Jones and F. Shields, Trends and capabilities of ATR algorithm technology for ground vehicle, 1988.
[9] A. J. Heller and J.L. Mundy, Benchmark evaluation of a model-based object recognition system, Proceedings of DARPA image understanding workshop, pp. 727-741, 1990.
[10] S. Hennessy, RSTA program plan and approach, 7th DARPA UGV demo II workshop, Denver, Colorado, July, 1992.
[11] E. Lithopoulos, E. Bender, J. Hepburn, Algorithm evaluation for an infrared target detection test bed, Proceedings of SIP Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 58-65, 1990.
[12] H. Nasr and M. Bazakos, Automated evaluation and adaptation of automatic target recognition systems, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 108-119, 1990.
[13] P. V. Noah, M. A. Noah, J. W. Schroeder and J. A. Chernick, Background characterization techniques for target detection using scene metrics and pattern recognition, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 2-15, 1990.
[14] R. Resendes, Detailed test plan technical feasibility test of the surrogate teleoperated vehicle, March 1992.
[15] E. Riseman (coordinator) UGV perception document, 1992.
[16] F. Sadjadi and M. Bazakos, A perspective on ATR evaluation technology, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 2-15, 1990.
[17] F. Sadjadi, Experimental design methodology: the scientific tool target recognition systems, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 108-120, 1990.
[18] R. C. Vogt, Role of performance evaluation in automated image algorithm generation, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 2120-134, 1990.

[19] J. D. Wald and D. B. Krig, ATR: problems and possibilities for the IU community, Proceedings of DARPA Image Understanding Workshops, pp. 255-264,1992.

[20] C. P. Walters, Removing the ATR performance evaluation bottleneck the CONVEO facility, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 16-31, 1990.

[21] C. Weems, E. Riseman, A. Hanson and A. Rosenfeld, An integrated image understanding benchmark of 2.5 D'Mobile'," in Proceeding of DARPA image understanding workshop, pp. 111-126, 1988.

[22] R. P. Weicker, An overview of common benchmarks, IEEE computer, pp. 65-75, Dec., 1990.

[23] O. Young and M. Cook, Analytical modeling for automatic target recognition performance evaluation, Proceedings of SIP, Vol. 1310, Signal and Image processing Systems Performance Evaluation, pp. 162-173, 1990.

[24] Promising directions in active vision, The NSF active vision workshop, 1991.

Chapter 5. Radio Communications

5.1. Introduction

There are communications issues related to virtually every component of a complex system such as the semi-autonomous vehicles of interest here. This section narrows the scope of potential measurements to those useful for this program and presents measurable parameters and characteristics of communications systems that can be used to assess performance. The identification of useful performance measures is applicable through the Demo II milestone. The suggestions made for the initial performance measurement work focus on collection of data in the earliest phase (Demo A) of the program.

5.1.1. Scope of Measurements

It is important to keep in mind that any communications performance measure made within this effort should aid in determining the overall performance of the UGV system. This means that the communication parameters measured should be those that affect the performance of a semiautonomous vehicle in executing the mission scenarios. Differences in parameters among communications systems that are unlikely to affect overall performance are of much less interest. For this reason, detailed measurements of the workings of a particular radio component are not within the scope of interest, but the performance of a particular communication link is.

5.1.2. Assumptions

Demo II plans call for executing all required RF communications via the AN/PRC-118 low cost packet radio. In the early phases of development however, such as in Demo A, it is expected that this radio will not be used for all purposes, and that significant communications will be accomplished over developmental radios. It is assumed here that there is no interest in discussions of performance measures that involve the developmental radios, and thus there is no discussion of parameters that may apply only to, for instance, the developmental analog video radios. It is assumed that regardless of the role of the packet radio system in Demo A, it will be installed and operational, and capable of being used in communications measurements.

5.1.3. Types of Communication

Several types of communication occur in this system. There is, of course, RF communication to and from the remote vehicle. Over the course of development toward Demo II, this will take on different forms. Initially, this communication will be handled through a mix of developmental radio components and a packet radio system to provide data, video and development support. By the Demo II milestone, only the packet radio system is to be used for all data and video imagery.

At the ends of the RF link described above are the vehicle and the operator work station. In each will be modules dedicated to communications that perform such functions as handling the protocols, queuing messages, handling communications errors and retransmissions and the like. In addition, local communications within these subsystems will be handled by backplane communications and local networks. All of these will influence the performance of an end-to-end exchange of information. These communication components are particularly important because it is within these various software modules that performance data can be collected in the executing system. Both the performance of the modules themselves and the performance of the links of interest can be measured through collection of appropriate data at these points. In order to make this type of information available for the purpose of performance measures, it is useful for the developers to consider providing appropriate interfaces (for example, common memory) in their system designs.

5.2. Potential Communication Parameters

This section describes a number of communication system parameters and explains why they would be useful in determining the performance of the semi-autonomous vehicle system.

5.2.1. Transmitted/Received Power

For a communication link to be functional, the receiver must receive sufficient transmitted power. A number of factors can influence the amount of power received. Aside from the transmitters' power, there are such considerations as terrain characteristics, weather, obstructions, antenna types, locations and heights and cable losses. The specific impact of these varies with the operating frequency of the radio link.

5.2.2. Bandwidth

It is important to determine the amount of bandwidth used by the radio system. A vehicle system that is capable of the same functions using less bandwidth will be more desirable.

5.2.3. Line-Of-Sight Characteristics

Clearly the most desirable communication would not require line-of-sight. However, in order to handle high data rates in a reasonable amount of spectrum space, information is transmitted at carrier frequencies high enough to introduce serious line-of-sight characteristics. The developmental radios will be very much line-of-sight, but this is not an important issue. Of more interest are the line-of-sight characteristics of the low cost packet radio system. Since this operates at about 1800 MHz, it will exhibit line-of-sight characteristics. Since this radio is to support all communication in Demo II, an early measurement of these characteristics will prove useful.

5.2.4. Range

A measure of the radio range over a flat test site can provide a useful baseline for comparing performance after some change is made to the radio system, or to confirm the health of the radio system at a later time. It will not give a good estimate of the range to be expected at a different site with hills, obstructions or significant interference, but it can be used to pinpoint degradation of the radio link. An early test should be performed to determine that suitable coverage is provided to all required areas of the Demo A site.

5.2.5. Data Rate

The rate at which information can move through the system can be a limiting factor on performance of the overall system. Various tests can be performed to measure the ability of the radios to move information. Addition of various protocols and security mechanisms reduce the data throughput available to the control processes. Current measurements of the throughput of the packet radios show a significant reduction from the maximum RF rate of the radios due to this overhead. Work is continuing to improve the performance, and measurements of performance should continue to quantify the performance in the application. A good way to measure this performance is with instrumented code at each end that measures messages ping-ponging back and forth across a link.

5.2.6. Message Activity

From the point of view of the overall vehicle system performance, an important metric is the volume of message traffic that is occurring. That is, how much communication is required to accomplish the mission scenario? Differences in the levels of autonomy of the control systems will result in the need to transmit different amounts of data and different numbers of messages. Way point driving will, for example, require a greater exchange of data with the operator control station than will road following. The least amount of data transmission required is desirable, both because it is more stealthy, and because it decreases demands on frequency spectrum allocation. It also makes it likely that simpler and cheaper radio systems could be used.

One way to decrease the message activity is to provide longer duration, more complex tasks to the vehicle, and to require less frequent interaction between the vehicle and operator station to accomplish tasks. In order to accomplish the same mission under these conditions, the vehicle must exhibit greater autonomy through a higher level of on-board intelligence. Measuring the message activity required by a system to accomplish a mission, can provide a measure of the vehicle autonomy. This will hold true for both robotic and manned vehicles. One can imagine a very long mission for a highly autonomous vehicle that is initiated by a single command and monitored by an occasional status report if the mission is proceeding successfully. A great deal of communications is no doubt occurring among the lower level subsystems on the vehicle, but these systems are controlled by higher level subsystems on the vehicle rather than requiring interaction from an operator.

The communications protocol used will also have an affect on the message activity. Is a message collision detection mechanism used that may require a message transmission to be attempted several times before it is successfully transmitted? What sort of positive and negative acknowledgements are required by the protocol? Does each message require acknowledgement? Are retransmissions requested and performed automatically or does the control process have the ability to ignore bad messages if desired? These characteristics and others can affect the amount of message activity.

Instrumenting the system to collect information on numbers of transmissions performed and the amount of data exchanged will provide a measure of how autonomous the vehicle system is. Further downstream, it will be possible to compare the amount of communication necessary with the robotic vehicle with that required for the manned vehicles taking part in the same mission.

5.2.7. Message Latency

In the packet radio system, the total time required to pass information to another system is composed of both the actual transmission time, the time spent handling the message at the sending and receiving nodes and time spent contending for access to the network. These produce message latency. Time may be spent, for example, packetizing the incoming message to be sent, adding a CRC and keying up the transmitter. Once this is done, the message may be transmitted at the available rate of the radio. This latency is incurred on each transmission. So, if many short messages are transmitted, the resultant throughput will be less than the maximum transmission rate of the radio system.

In addition, the latency adds to the age of the data transmitted. In systems requiring real-time teleoperation, excessive latency can lead to instability. The Demo II semi-autonomous vehicles attempt to move away from teleoperation and toward greater autonomy. This reduces the impact of communication system induced latencies. However, both RSTA and the way-point navigation mode require a certain timeliness of information, and therefore it is important to isolate and measure the latencies present. Instrumenting the communications code to keep a time-stamped record of transmissions and receptions will be useful in obtaining this information.

5.2.8. Error Rates

If the application software handles transmission errors or if the packet radio system provides error rate data it should be possible to collect this data during normal system operation. Error rates higher than an experimentally determined baseline can indicate problems with the radios, antennas, cabling or even the appearance of new interference. High baseline error rates may indicate that a particular radio system is not well suited to this application or site. Collecting this information allows isolating an overall drop in system performance to the radio system.

5.2.9. Effects of Other Communication on Individual Links

Communication that works well in a one vehicle scenario may degrade quickly when other communicating systems are introduced. It is useful to monitor the performance of the communications system as other communicating nodes are added in a controlled fashion. The effects are sometimes hard to predict.

5.2.10. Communication System Power Consumption

Any mobile system has a finite power budget. Accurate information on the power requirements of all systems contributes to better design of succeeding generations. The communication system power requirements should be measured to provide this data. Excessive power requirements will reduce the mission range and reduce available space for mission packages.

5.3. Other Communications System Characteristics

The parameters described previously are measurable and provide a great deal of information about the performance of the communications system. However, in determining the value of a particular communications system for Demo II, there are other important criteria that should be considered. These are not directly measurable, but should be considered in the evaluation of the system. Since performance measures are the main thrust of this document, these are only briefly introduced here.

5.3.1. Protocol

Many aspects of the performance of the communications system are determined by the protocol used. Protocols with a large amount of overhead will cause data throughput to be reduced. The way in which messages of different priorities are handled is determined by the protocol. One must ensure that the most urgent messages are handled in a way that results in them being received in a time that is acceptable for the operation of the vehicle systems. Another important aspect is how corrupted messages are handled. Are retransmissions handled automatically at some low level of protocol that is beyond the control of the control code? Are messages buffered if they can't immediately be sent? These may not be desirable techniques. What happens when a link to a vehicle is lost? A further consideration is the level of standardization of the protocol. Does it use communication standards that promote interchangeability of components or is it a specially coded set of modules?

5.3.2. Number of Users

The communication system should be examined to determine the number of users it can support. This could translate into the number of subsystems or vehicles that can be supported. It may be limited by protocol, addressing scheme, bandwidth, some form of channel allocation or even by performance degradation with additional nodes.

5.3.3. Stealthiness

The degree of stealthiness will be the result of a number of factors related to communications. How much power is transmitted? Which components transmit most frequently -- the vehicle or the operator work station? How frequently are transmissions necessary and how long do they last? How directional is the transmission?

5.3.4. Scaleability to Actual Use

Given a successful implementation of a prototype system of one or some small number of vehicles, is it possible to scale up the implementation to many more vehicles or whole systems or are there obvious impediments to this?

5.3.4.1. Frequency Spectrum Allocation

Are the frequencies used in the prototypes available for actual implementations? If not, will the system work at frequencies that are available for actual systems?

5.3.4.2. Bandwidth Needed

How much bandwidth is needed at the frequencies approved for use. Is there room to add additional vehicles or systems or is all the bandwidth used by a single implementation? Does the communication system require an amount of bandwidth at a frequency that would never realistically be available for actual use?

5.3.5. Video Capabilities

How will the communication system be used to handle video information? Will it support the mission scenario? How much delay is imposed on video information by the communications system?

5.3.6. Redundancy, Reliability

How rugged is the communications systems? Are backup links available if a primary control link fails? Can the switch to a backup be done on the fly?

5.4. Measurement Descriptions

As described above, numerous factors influence the performance of a communications system. A set of measurement examples is presented here which help determine the performance in areas that are very important even at the earliest phase of implementation. These are all performance measures that should be appropriate for the Demo A system.

Four measurements are presented that help quantify radio coverage, communications throughput and error rates.

5.4.1. Link Performance

5.4.1.1. Objective

This measurement quantifies the throughput on communications links of interest.

5.4.1.2. Procedure
With the vehicle located in good line-of-sight position, measure the time for a message to be sent from the operator workstation and replied to by the remote end. A test that can be run under software control that collects the data automatically is the most convenient. No processing should be done on the remote end other than the required communications related functions.

5.4.1.3. Analysis
A spectrum analyzer can be used both to examine the exchange of message packets and to ensure that no interfering transmissions are affecting the measurements.

Instrumented code capable of sending and receiving test messages and timing the periods between transmissions and receptions will provide data for the analysis. Average and worst case message times are of interest.

5.4.2. Coverage
5.4.2.1. Objective
This measurement seeks to provide information on the radio coverage provided by the communication system.

5.4.2.2. Procedure
Perform measurements at all locations of interest on the Demo A course to determine degree of coverage by the radio system. The results can be used to compare radio systems, to compare differences due to operating frequencies, to compare antennas and antenna positions and orientations and to identify any maximum range problems.

Monitoring transmissions received from a roving vehicle on a spectrum analyzer can be useful as a first step. More detailed data, such as increases in error rates, can be obtained through instrumenting code at each end of the link.

5.4.2.3. Analysis
Data will be provided by instrumented code capable of sending and receiving test messages and monitoring error rates. An error rate map of the site can be produced. Test operator can obtain qualitative

impressions of performance during test by observing areas of poor performance.

5.4.3. Line-of-Sight Characteristics

5.4.3.1. Objective

These measurements seek to identify the degree to which the radio system can deal with obstructions.

5.4.3.2. Procedure

The procedure here is similar to that used for the Coverage Measurements. However, while it is assumed that the actual vehicle paths used in Demo A will be served by line-of-sight communication paths, in this test, attempts are made at communication in areas with obstructed line-of-sight. The roving vehicle should be driven over hills, behind trees, etc., and the performance should be monitored.

Monitoring transmissions received from the roving vehicle on a spectrum analyzer can again be a useful first step. More detailed data, such as increases in error rates, can be obtained through instrumenting code at each end of the link.

5.4.3.3. Analysis

Data will be provided by instrumented code capable of sending and receiving test messages and monitoring error rates. An error rate map of the site can be produced. Test operators can obtain qualitative impressions of performance during tests by observing areas of poor performance. The effects of particular types of obstructions (hills, foliage, other vehicles, buildings, etc.) can be noted from the data.

5.4.4. Communication Activity

5.4.4.1. Objective

This measurement seeks to obtain information on the total volume of communicated messages and the rates of communication for various elements of the system. It is suggested that the rate of communicated information is an indicator of the relative vehicle autonomy in a given implementation. As progress is made toward Demo II, this can be a useful measure, and may be useful in comparisons with manned systems.

5.4.4.2. Procedure

Instrument each transmitting entity (vehicle, operator workstation) to record and time-stamp the occurrence and size of each transmission. A running average of data bytes/sec transmitted may also be maintained and available for inspection during operation.

5.4.4.3. Analysis

Normal control and communications code enhanced with logging and time-stamping capabilities will provide the data for analysis. The communications activity for specific portions of the scenario can be evaluated. Communication intensive operations can be identified.

5.5. Summary

Communications parameters appropriate for performance measures of Demo A through Demo II systems have been presented. Examples of how to obtain information on some of these in the context of Demo A have also been suggested. There is work currently being performed which will provide additional information about the Demo II communication system, and will aid in the development of further performance measures. At this time, however, reports of this work are not yet available, but soon should be. One such effort is the field testing of the AN/PRC-118 packet radio that has been performed by CECOM as part of the Survivable Adaptive Systems Advanced Technology Demonstration. Documentation of this testing, performed at Fort Louis, is underway. A study that compares a number of radio systems, including the AN/PRC-118, has been performed under contract (SBIR Contract No. DAAE07-92-C-R042) to TACOM. A document describing this effort is soon to be completed. Finally, a market survey of related radio equipment is currently being performed by CECOM, and will result in a document sometime later.

Chapter 6. Ground Truth Data Acquisition

6.1. Introduction

To evaluate the performance of UGV's performing RSTA missions, it will be necessary to provide ground truth for comparison with the observations of the vehicle sensors. Ground truth will be needed for both the location of each vehicle and the objects, obstacles, and targets it designates. Thus there is a requirement to have an independent measurement capability to measure the geolocation of each of the moving and/or static vehicles at all times during the demonstrations.

To convert measurements or observations in the vehicles' coordinate frames to world coordinate frames and then compare to ground truth, it will be necessary to have a high resolution, detailed map of the test site. If a UGV scout reports a target at a particular azimuth, elevation, and range, moving on a bearing at a measured velocity, it will be necessary to do coordinate transformations to convert these data to locations on a suitable scaled map of the test range in world coordinates and compare the reported locations with ground truth provided by an independent metrology system. In other areas of unmanned vehicle development, such as cruise missiles and underwater vehicles, test ranges have long been established. For UGV demonstrations, existing test ranges have not been completely developed and must be supplemented with portable field metrology equipment. Given the nature of the current UGV research program, with its emphasis on different sensor systems and evaluation algorithms, a highly detailed map is desirable. However, the cost of developing such a map for a large area such as the test site at Fort Hood would be prohibitively expensive. Hence a compromise must be reached between the detail provided, or degree of map resolution and the areas for which it is developed.

Thus we recommend utilization of both large scale and small scale metrology capabilities in the field demonstrations, with development of appropriate scale maps, and all data indexed to a common digital terrain map and time clock. The field metrology systems can be used to verify the position and velocity of moving targets over kilometer scale ranges and to map stationary objects for use in landmark navigation and RSTA. For driving, greater accuracy of measurement and higher resolution maps are needed but only for a limited area, perhaps a 100 meter square area.

6.1.1. Assumptions

In this document, we make the following assumptions:

Demo A and B will be conducted on the 2 km track at the Martin Marietta-Denver site. Performance evaluation focus will be on navigating a hairpin turn, traversing paved and unpaved roads, and performing transitions between paved and unpaved roads.

Demos C and II will be conducted at Fort Hood, Texas in the Blackwell / Lone Mountains. Demo II will involve four unmanned vehicles, which will perform a RSTA mission.

The vehicles will navigate with the aid of digital terrain data, on-board stereo vision systems, LADAR, inertial gyros, and wheel revolution counting.

Differential GPS could be used by the vehicles for both operations and for performance evaluation purposes.

6.2. Real Time Geolocation (RSTA, Landmark Navigation)

6.2.1. Objective

For evaluation of landmark navigation and for evaluation of RSTA, including fixed and moving targets, establish a real time metrology capability with a resolution of one (1) meter for ranges of several kilometers. This evaluation deals only with the location of targets and does not include accuracy of target recognition.

6.2.2. Test Procedures

A variety of real time measurement techniques have been considered, including Loran and real time stereo photogrammetry. However, considering the sites involved, differential GPS clearly seems to be the preferred technology.

We recommend that the UGV Demo A in June, 1993 be utilized as an opportunity to evaluate the feasibility and demonstrate the capability of using GPS in a differential mode to provide time-tagged object positions for performance evaluations of unmanned ground vehicles. At the Martin Marietta - Denver test site, we should attempt to position the vehicle to an accuracy of one (1) meter in real time, with a potential for one tenth (0.1) meter in three dimensions with post processing, based on positional time-tags to within one (1) microsecond.

6.2.3. Analyses

1. Conduct post processing to provide high accuracy, time-tagged positions for the vehicle for the duration of the demonstration.
2. Prepare a report describing data collection and analysis.
3. Develop a preliminary design for full integration of GPS performance evaluation into UGV demonstrations B, C, and II.
4. Develop a proposal for future demos.

Preliminary estimates by the Topographic Engineering Center are that costs for tracking one vehicle during Demo A would be $97,000, including hardware procurement. Experience at Demo A will provide a better basis for refining future requirements and costs for the future demonstrations. For planning purposes for subsequent demos, preliminary cost estimates are $100,000 for one vehicle and an additional $50,000 for each additional vehicle.

6.3. Real Time Geolocation (Driving Evaluation)

6.3.1. Objective

For driving, the UGV must use its lower level control system to sense and utilize aids to navigation such as landmarks but avoid obstacles, including not only rocks and trees, but ditches, holes, mud, water, and steep slopes. Evaluation of how well the UGV control system avoids obstacles could range from such crude measures as counting the number of times the vehicle gets stuck to comparing the world model it develops with extremely detailed maps of the terrain.

As discussed in Chapter 3, it is possible to observe the effect of placing known objects on the course. These can be placed at known locations or their locations can be measured after placement.

In general, it would seem desirable to establish a field metrology capability for critical areas of the test track, perhaps a 100 meter area or mission corridor, with a resolution of 50 mm.

6.3.2. Test Procedures

6.3.2.1. Macro metrology

Macro metrology systems are useful for measuring the UGV position and for the mapping of the test site. There are several metrology systems that can be used to measure the position of an object located a significant distance away. In this section a brief review of those metrology systems that can track a moving object, like the UGV, from a few meters to several km away will be presented.

The most accurate of these metrology systems are the "Laser Tracker Interferometer" (LTI) and the less accurate "LIDAR Tracking System" (LTS), which have been used as robot and manufacturing testing and calibration tools for several years. These metrology systems can provide automatic tracking of a retroreflector target attached to the moving object and measure its position and orientation. The commercially available LTI and LTS systems have a maximum range of only 30 to 60 m and are very expensive. To increase the range to that required for the field testing of the UGV, significant modifications of the LTI and LTS systems would be necessary. These modifications would significantly add to their already considerable cost and delivery time. For these reasons, these systems are described in Appendix I and no further discussion will be included in this section.

We have also considered what lessons might be learned from past metrology and standardization activity in robot metrology that would be relevant for UGV evaluation. A discussion of this is included in Appendix II.

6.3.2.2. Manually Tracking LIDAR

These are timed-pulse or amplitude-modulated infrared laser light electronic distance-measuring instruments. They measure the time needed for a pulse of infrared laser light to travel from the instrument to the target and back, or the phase shift between the returning and outgoing laser light. To reduce the effect of noise the pulse is sent hundreds or even thousands of times and the total time between the first pulse sent and the last received is measured and divided by the total number of pulses. This can be repeated from a few times per second to 100 times per second. For a higher sampling rate less averaging takes place, but faster moving objects can be tracked.

A LIDAR with an automatic target tracking system is described in Appendix I; here, a less expensive and more readily available manual system is considered. In this case the LIDAR is mounted on top of the telescope of a Theodolite, like that shown in Figure 6.1. The telescope can be used to view and manually track a moving target mounted on the UGV. Due to divergence, at a distance of a few kilometers the laser beam diameter will be of the order of one meter. Thus even if the center-line of the beam is not aimed directly on to the target, enough laser energy should be reflected back to allow the measurement of the distance.

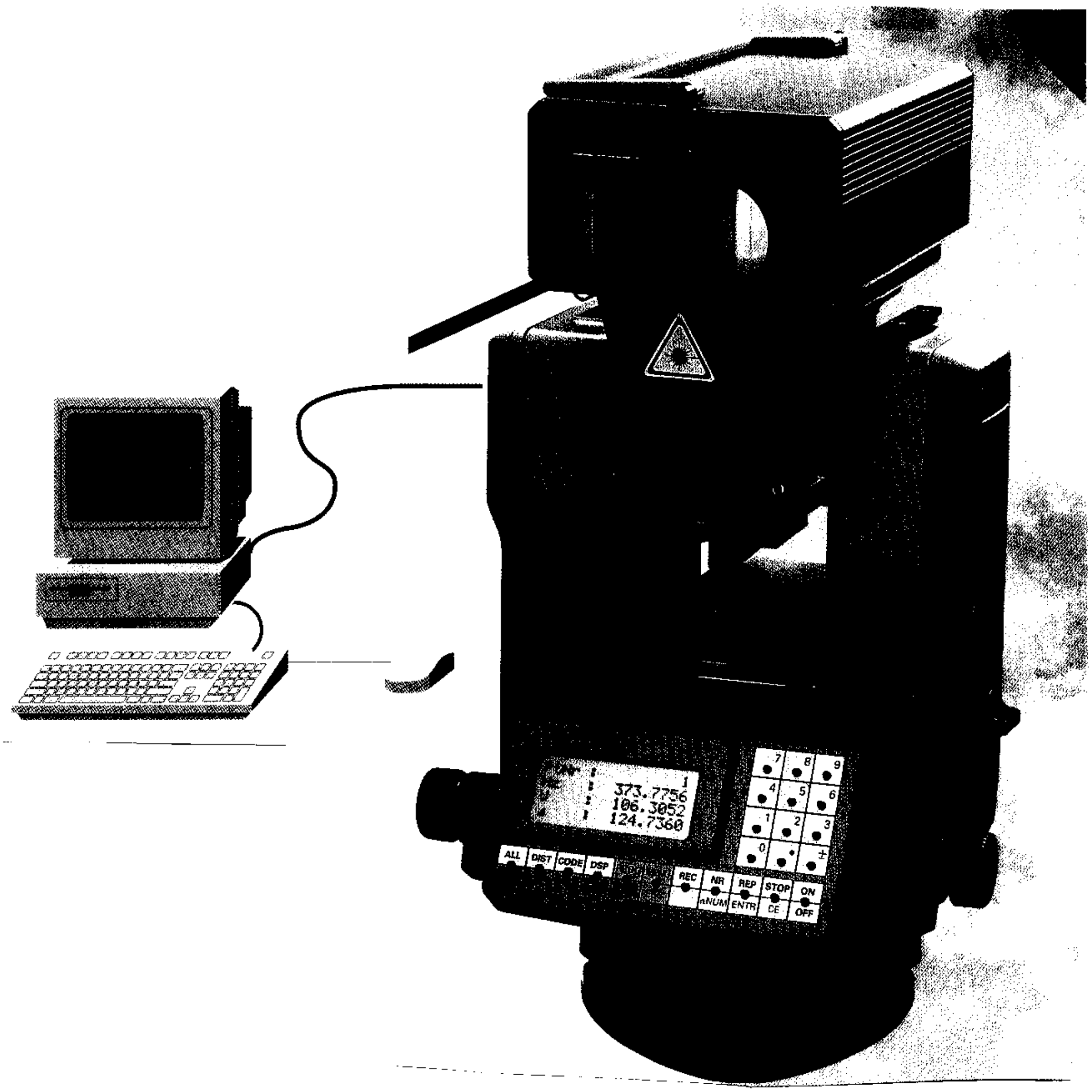

Figure 6.1. LIDAR instrument mounted on a Theodolite telescope

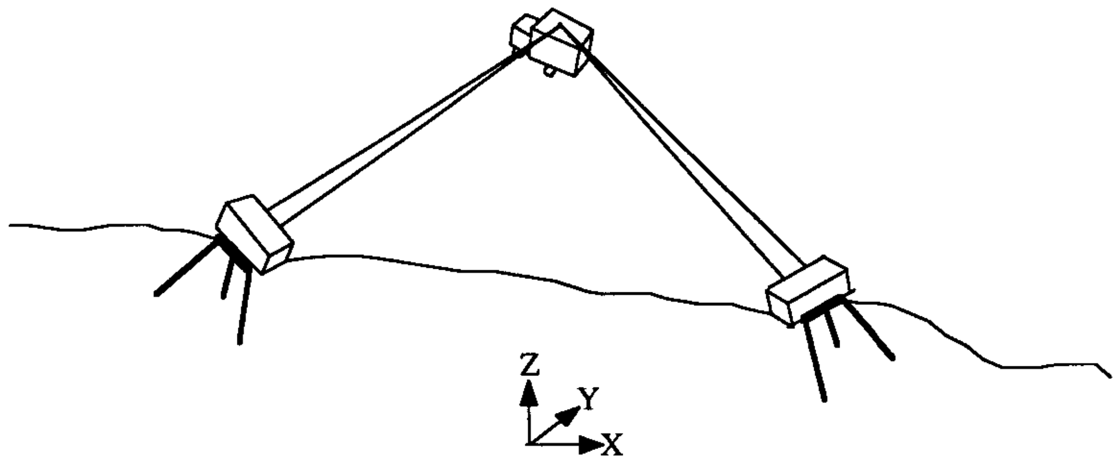

Figure 6.2. A manually tracking dual LIDAR system.

If the UGV target is stationary or moving at a low speed it is possible to use a single Manually Tracking LIDAR (MTL) to determine its position. Besides the distance information the Theodolite azimuth and elevation angles can be used to specify the spherical coordinates of the target. For higher speeds, though, the Theodolite telescope will not be able to follow the high frequency motions of the UGV target although the LIDAR will continue to measure distance information. In that case two or three MTLs will have to be used. Figure 6.2 shows the case where two MTLs are used. In this case it is assumed that a good digital map of the test site is available. This would make possible the estimation of the X, Y, and Z coordinates of the UGV target, with respect to the test site reference frame, based on the measurement of its distance from the two MTL sites. It is of course assumed that the coordinates of these sites with respect to the same reference frame are known. If the map accuracy is not very good then a third MTL system will have to be used. The MTL sites will have to be selected such that the UGV target is visible throughout the test. The MTLs will be controlled by one or more Personal Computers (PCs). The PCs would receive the test timing pulses, trigger the MTLs to measure the distance to the UGV target and then save the results together with the pulse timing information.

Because of the wide angle changes of the direction of the UGV, its target will have to be designed to either track the laser beam or to consist of several wide angle retroreflectors like cat-eyes. These targets have a ± 60 degrees acceptance angle. Thus three of them mounted at 120 degrees from each other should be able to reflect the laser beam back from most orientations of the UGV. Large diameter cat-eye targets should be used. If it is found that at large distances from the MTL sites not enough laser light radiation is reflected back, then more rows of cat-eyes can be added to the target assembly. Commercially available LIDARs have a range of up to 14 km (8 miles). The target assembly must be mounted on the top of the UGV, and its relative position to the coordinate frame of the UGV must be known.

The MTL can be used to track the position of the UGV and it can also be used to develop topographic maps of the test site. The target assembly will be portable and can be mounted on the UGV or a topographic mapping vehicle. The topographic mapping vehicle could be a motorized all terrain vehicle or other appropriate mobile platform. The most simple map could be that of the outline of special test track features like roads, lakes, obstacles, etc. A grid map of the whole test site could also be created. The test site would be divided into a certain number of rectangles, which lie on a horizontal reference plane, defined by the MTL controller as specified by the operator. The topographic mapping vehicle will be guided to visit each rectangle and its vertical axis distance from the horizontal reference plane will be recorded. Using various interpolation techniques the data will be combined to create three dimensional stereo maps of the test site.

The sampling rate of commercially available LIDAR instruments can range from 3 to 100 samples per second. Their accuracy depends on the number of distance measurements per sample. An accuracy error of 10 to 100 mm/km can probably be expected. The repeatability error depends on the random component of the measurement. The main source of this error is the turbulence in the air between the MTL and the target. It should depend on the weather conditions and is very difficult to predict without experimental data.

The cost of a good LIDAR is approximately $20 K. The cost of a Theodolite is approximately $10 K and that of a cat-eye $5 to $10 K.

Thus the cost of two MTL systems with one PC controller for each system should be approximately $90 K. Delivery should not take more than a few months. Thus, for those evaluations requiring accuracy error of less than 100 mm, the MTL systems would be advantageous.

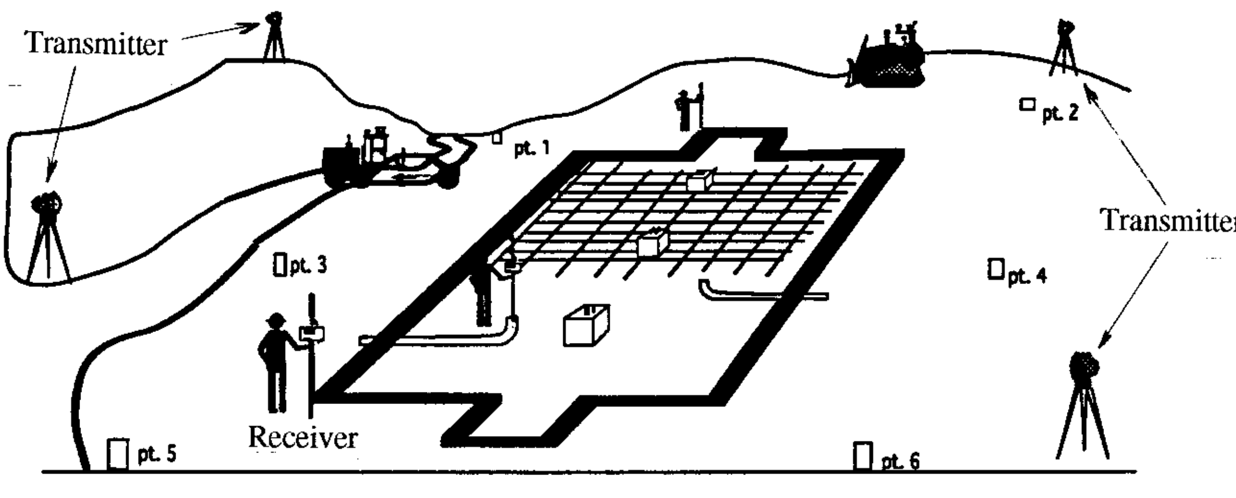

Figure 6.3. Work site with RtPM

6.3.2.3. Real-Time Position Measurement System

The Real time Position Measurement (RtPM) system was developed at the Virginia Polytechnic Institute for the real time mapping of construction sites. A consortium of users, researchers, hardware developers, owners and funding agencies was organized by the Civil Engineering Research Foundation. Further development of a commercial prototype is proceeding with CPAR funding by the Army Corps of Engineers by Spatial Positioning Systems, Inc. A prototype has already been demonstrated for three-dimensional position measurement of points, vehicle tracking, and for the quick determination of large surface contours.

To allow for RtPM implementation, a local area positioning system is established around a given site (see Figure 6.3). This can be a set of transmitters that are set up around the site or a control line as a base for kinematic GPS. If transmitters are used, the energy transmitted can be RF, microwave, or lasers. There are two primary components to a laser based RtPM system: a transmitter and a receiver. The transmitter is set on a tripod with a front face such that laser light can be scattered about the site. A minimum of two transmitters would be needed. The receiver is comprised of a computer and screen, two optical pieces on a pole housing batteries, and a data entry and retrieval system. The two optical pieces are in line, such that the position and orientation of the pole can be calculated from the positions of the optical pieces.

The laser based system depicted in Figure 6.3 requires two transmitters and one receiver at a minimum. Additional transmitters allow for redundancy and also help receivers maintain a direct view to a minimum of two transmitters. Once the transmitters are set up, any number of receivers can calculate position information using line of sight to any two transmitters. The laser system would work as follows:

1. Transmitters are set up at random points with the front face generally aimed at the site. An alternative setup would be to locate transmitters at known points with known orientation.

2. To calibrate the system a set of preestablished points is needed. The location of these points is entered into the computer on the receiver. The receiver, in turn, is set on three of the known points. The receiver then calculates the position and orientation of the transmitters. A check of the three points used for calibrating the system of transmitters is done by touching a fourth point.

If the receiver is mounted on the UGV, its position and orientation about two axes can be estimated in real time. The sampling rate is 5 Hz. The range of the current RtPM prototype is only 50 m with centimeter accuracy, but according to the developers, future versions will have a larger range. The future cost of commercial units is estimated to be about $60,000.

The RtRM appears to be very promising as a technology for short range, real time metrology of UGVs. However, it is still under development and is not yet on the market. It should be given additional evaluation when a commercial model becomes available.

6.4. Ground Truth Map Generation (RSTA, Landmark Navigation)

6.4.1. Objective

Support UGV performance evaluation of off-road navigation and RSTA with high-resolution digital terrain data for the UGV Denver testsite.

6.4.2. Test Procedure

The Army Topographic Engineering Center (TEC) currently has responsibility for digital terrain data for the ARPA Autonomous Navigation Program and the DOD Unmanned Ground Vehicle (UGV) program. Various digital terrain products have already been generated and distributed to the UGV research and development community to support UGV mission planning and operations. These include Controlled Stereo Imagery, custom (5 meter post to post spacing) Digital Elevation Model (DEM), Digital Orthophoto (1 meter), and Controlled Stereo Imagery.

6.4.3. Analyses

The currently available terrain data products appear to be adequate for performance evaluation of landmark navigation and RSTA functions. For the UGV mission corridor, it is possible to obtain way point and landmark coordinates via point-positioning from CSI to less than one meter.

6.5. Ground Truth Map Generation (Driving Evaluation)

6.5.1. Objective

Support UGV performance evaluation of driving by providing high-resolution digital terrain data for a limited area at the UGV Denver test site.

6.5.2. Test Procedure

We recommend that the TEC generate a custom high resolution digital elevation model (DEM) and Digital Orthophoto for the RSTA field of view at the UGV Denver test site from Controlled Stereo Imagery (CSI). This higher resolution map would reduce the DEM post to post spacing form 5 meters to 1 meter. With point-positioning from CSI, it would be possible to provide the coordinates of critical objects and features.

These evaluation recommendations are based in large part upon consultation with the U.S. Army Topographic Engineering Center at Fort Belvoir, Virginia. A copy of a memorandum from George Lukes, Chief of the Autonomous Technologies Division, which outlines the potential support that the TEC could provide is included as Appendix III.

6.5.3. Analyses

The TEC would provide a report of its findings and activities together with recommendations for future demonstrations. A preliminary cost estimate for Demo A and B is $18,000.

Appendix I. Macro Metrology Systems

This appendix describes macro metrology systems for measuring the UGV position and orientation and for mapping the test site.

Laser Tracking Interferometer (LTI) and LIDAR Tracking Systems (LTS) are the most accurate systems available for macro metrology. However, they are expensive and limited in range. Thus they do not appear well suited for evaluation of field demo performance of UGVs. However, they could be used in the laboratory and for short range evaluations in the field.

I.1. Laser Tracker Interferometer System

The Laser Tracker Interferometer (LTI) was invented and developed at the National Institute of Standards and Technology (NIST) in the mid eighties. It is currently a commercial product and several units have been purchased by universities and industry. The LTI can remotely measure, with high accuracy, repeatability and resolution, the three dimensional space position, and the orientation of a target mirror system. If the target is attached to a moving object, like an UGV for instance, the LTI will follow the moving target and measure its position and orientation in real time at a high sampling frequency (up to 1000 times per second). The LTI has been used for many industrial applications requiring high accuracy. Following is a list of some applications which have come to our attention:

a. The calibration and measurement of the performance of industrial robots.
b. The calibration of very large milling machines (up to 30 m long table) used for the manufacture of airplane parts (Boeing Commercial Aircraft Co.).
c. The manufacture and assembly of multistage rockets (Martin-Marietta, New Orleans).
d. Surface contour mapping of large composite components (Martin-Marietta, Oak Ridge)

e. The calibration of jigs and fixtures for helicopter refurbishment (U.S. Army, Corpus Christi).
f. Tracking the motion of SDI smart pebbles rockets in a laboratory (Edwards Air Force Base).
g. Development of topographical maps of small landscape models (NIST).
h. Positioning sensor for the AEGIS phased-array antennae calibrator (Westinghouse).

The Robot Systems Division of NIST purchased a three axis LTI several years ago. It has been used for several years for the performance testing and calibration of robots. Its reliability and performance have been very good.

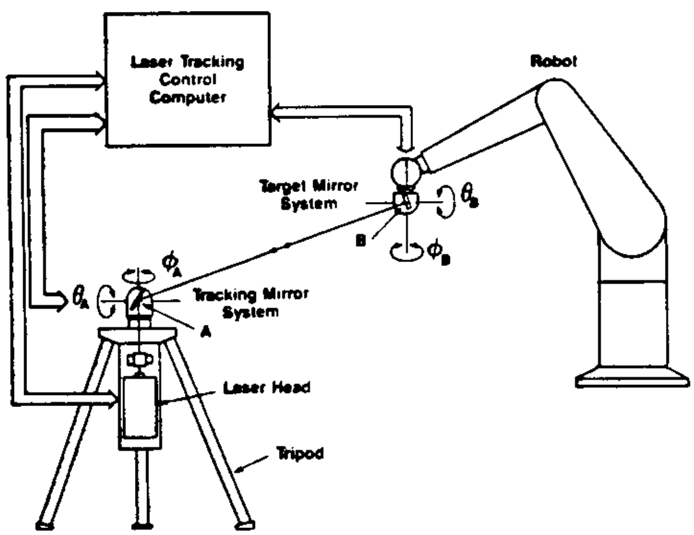

Figure I.1. Schematic of a Laser Tracking Interferometer System

I.1.1. Brief Description

Figure I.1 shows the schematic of a six axis Laser Tracker Interferometer (LTI), which is measuring the position and orientation of a robot arm end-plate. The tracking unit, shown in the figure mounted on a tripod, consists of a laser interferometer and a dual-axis-servoed mirror. A second dual-axis-servoed mirror is mounted on the robot end-plate and becomes the target of tracking and measurement. The idea is to direct the laser beam to the target continuously by controlling the angles of the tracking mirror. In the meantime, the target mirror is also controlled to stay perpendicular to the beam and return the beam precisely to the laser system. Two bilateral-effect photo diodes are installed at the back of the target mirror to supply misalignment information of the tracking process to a control computer. The computer then computes and issues the appropriate servo-commands to the four servo-axes to null the misalignment. The change in length measurement obtained from the laser system, when combined with the azimuth and elevation angles of the tracking mirror, yields the position of the target in spherical coordinates. The pitch and yaw angles of the target with respect to the tracking system can also be obtained by measuring the target-mirror angles. With a polarized laser beam it is possible to measure the roll angle of the target-mirror.

I.1.2. Macro Metrology Application

Since the LTI requires maintaining line-of-sight contact with the target, it will have to be mounted on a tower, tall building, or hill overlooking the whole test track. If contact is lost between the laser beam and the target, even temporarily, an error will be introduced in the test measurement. Most laser interferometers can detect when this occurs and warn the operator. The maximum distance capability of off-the-shelf laser interferometers is up to 60 meters; however, special interferometer systems (such as Leica's ME 5000) can have a range from 8 to 10 kilometers. Above a certain level of dense fog or smoke, between the tracking unit and the target, the LTI will become inoperable, either due to distortion of the laser beam wave form or low intensity of the reflected beam.

The LTI could be used to track the position and orientation of the UGV in real time and it could also be used to develop topographic maps of the test site. The LTI target system will be portable and can be mounted on the UGV or a topographic mapping vehicle. The topographic mapping vehicle could be a motorized all terrain vehicle or other appropriate mobile platform. The most simple map could be that of the outline of special test track features like roads, lakes, obstacles, etc. A grid map of the whole test site could also be created. The test site would be divided into a certain number of rectangles, which lie on a horizontal reference plane, defined by the LTI controller as specified by the operator. The topographic mapping vehicle would be guided to visit each rectangle and its vertical axis distance and orientation from the horizontal reference plane would be recorded. With various interpolation techniques the data will be combined to create three dimensional stereo maps of the test site.

I.1.3. Performance Characteristics

Some of the performance characteristics relevant to this application are:

a. Maximum Tracking Speed: This is the maximum speed the UGV can travel without the tracking mirror losing track of the target mounted on the vehicle and which must be below the speed limit of the interferometer. Commercially available LTIs can track a target moving at a speed of 1 to 2 m/s at a distance of 2 m. Since the distance between the tracking mirror and the UGV is much greater there should not be any significant problem in tracking the vehicle.
b. Sampling Frequency: This is the number of observations the LTI can make of the target per second of time. The UGV speed divided by the sampling frequency gives the resolution of the observations. Commercially available LTIs can sample at frequencies of up to 1000 times-per-second.
c. Resolution: This is the smallest distance the UGV can move that can be observed by the LTI. Commercially available LTIs have a resolution of approximately 3.5 μm, which corresponds to approximately 3 mm/km.
d. Accuracy Error: This is the offset or bias error of the instrument. The main sources of this error are the drive mechanisms and the encoders. Commercially available LTIs can have an accuracy error of approximately 10 μm (under laboratory conditions), which corresponds to approximately 10 mm/km.
e. Repeatability Error: This is the random component of the error. The main source of this error is the turbulence in the air between the tracking mirror and the target. It should depend on the weather conditions and is very difficult to predict without experimental data. If the target is not moving, averaging of 10 to 50 observations should be used. This could be possible during the collection of data for the development of the topographic map. Under laboratory conditions an error of approximately 15 mm/km could be expected.

The NIST Robot Systems Division has several years experience with the specification and testing of LTIs and could purchase and test this instrument.

I.1.4. Estimated Cost and Delivery Time

Commercially available LTIs cost $100 K to $150 K. An LTI which could be used for the UGV testing would require a laser interferometer with a range of several kilometers. This would require significant redesign of the LTI mechanical and electrical components and thus raise the price and delivery time. A price of $500 K to $750 K and a delivery time of 1 to 2 years could be expected.

I.2. LIDAR Tracking System

The LIDAR Tracking System (LTS) is very similar to a Laser Tracker Interferometer (LTI) except that in the case of an LTS the distance between the tracking mirror and the target is measured by a LIDAR system which uses an absolute ranging repetitive time-of-flight laser technique. It is a real-time measuring system with a sampling rate of up to 100 samples per second. Each sample represents an averaging of about 1,000 trips of time-of-flight measurement for improvement of accuracy. For better accuracy, the sampling frequency can be reduced for larger sample averaging.

Following is a list of advantages and disadvantages with the use of one laser tracking system as compared to the other:

I.2.1. Advantages of using an LTS instead of an LTI

a. The LTS measures the absolute range between the tracking mirror and the target. Thus at the beginning of the test the LTS target does not have to be placed at a location of known distance from the tracking mirror as the LTI requires.
b. The line-of-sight of the LTS can be momentarily interrupted.
c. The LTS can work under more adverse weather conditions.

I.2.2. Disadvantages of using an LTS instead of an LTI

a. The LTS has a larger accuracy error than a LTI, perhaps of the order of 100 mm/km.
b. The LTS has a lower sampling rate.

Appendix II. Robot Performance Testing

Introduction

Several standards have been established for the testing and evaluation of the performance of robots under different operating conditions. On the national level the effort to establish this type of standards is lead by the Robotic Industries Association (RIA) in cooperation with the American National Standards Institute (ANSI), and with the help of several committees and technical advisors groups. On the international level the International Standardization Organization (ISO) is responsible for this type of activity in cooperation with the national robotic societies of all the industrialized nations.

The objective of their effort is to come up with universally accepted practices for the testing of robots and universally accepted figures of merit for expressing the results of the tests. These tests have been thoroughly studied and debated, for their usefulness, by experts from the robot manufacturing industry, the robot users and members of government labs and academia. One benefit of the existence of these standards is that tests can be performed anywhere there are suitable facilities and there is no need to transport heavy machines, like robots, to central testing laboratories, thus reducing testing cost. The designers of robots benefit because they can compare the performance of their robots against that of their competitors and identify the deficiencies in the design, thus improving robot performance. But the main beneficiary, of standardized robot testing is the robot buyer, who can easily compare those meeting his price limit requirement and select the best performing robot for his application.

It is possible that by standardizing at least a portion of the UGV testing, data gathering, and analysis procedure, the Department of Defense can realize similar benefits. Furthermore, this could become a dual use technology if these tests are adopted by the civilian UGV research programs sponsored by the state and federal Departments of Transportation. NIST, with its experience in the development of this type of standards, and its contacts with the standards setting organizations and committees could provide valuable help with that type of effort.

This appendix presents a very brief review of some of the philosophies and figures of merit used in the establishment of standard robot performance tests.

II.1. Test Classes

Because of the complexity of a robot controller it is sometimes better if different settings (classes) are selected for various types of testing. For example, in the case of the dynamic testing of robots three different classes of tests have been established:

a. Path Prioritized,
b. Speed Prioritized, and
c. Optional.

During path prioritized testing the objective is to maximize path following accuracy. When that objective is adhered to, significant variations in the speed could occur. This test class, for example, is more important for a scout UGV. During speed prioritized testing the objective is to optimize the speed characteristics of the robot motion and to follow the commanded speed as closely as possible. This test class, for example, is more important for a sealing robot, or a UGV transporting explosives.

II.2. Test Paths and Payloads

All standard robot tests specify the paths and payloads to be used for testing, but they also accept an optional category of non-standard ones. There are two possible choices in the selection of these standards:

a. Standard with Specific Dimensions and Payloads,
b. Standard with Non-Specific Dimensions and Payloads.

In the case of test paths, the path shape is specified, for example a circle, but in choice (a) the radius of the circle is specified (usually three to four choices are given), while in choice (b) the value of the radius is left open to be specified by the size of the workspace of the robot. Similarly for payloads, while in choice (a) the weights are given, in (b) they are certain per-cent values of the robot payload. If the objective is to select a specific robot for purchasing, among several of similar capabilities, the test results based on (a) are preferable.

II.3. Programming (or Command) Modes

The standard tests are designed around the following robot programming (or command) modes:

a. Manual Teach Programming.
b. Manual-Off-Line Programming.
c. Off-Line Programming.

In the case of UGV testing, mode (a) would correspond to a full master-slave control. Case (b) would correspond to the type of control used presently, which is a mixture of autonomous and master-slave control. Case (c) would correspond to completely autonomous control.

II.4. Standard Figures of Merit

The Figures of Merit (FOM) are defined very carefully through mathematical formulas to minimize the possibility for misunderstanding. The desired minimum number of tests is in most cases 50. Recognizing that some of the tests are time consuming and expensive to perform, a minimum of 10 is sometimes accepted. Following is a list of FOM which might be relevant to the present UGV performance testing program:

a. Maximum path accuracy.
b. Average path accuracy.
c. Average path repeatability.
d. Cornering round-off.

e. Cornering overshoot.
f. Path speed accuracy.
g. Path speed repeatability.
h. Path speed fluctuation.
i. Acceleration time.

II.5. Categories of Test Standards

There are two broad categories of standard robot performance tests. The first covers basic performance testing and the second application specific testing. The first category consists of two standards; one covers basic motions static performance and the second covers basic motions dynamic performance. The second category consists of numerous application specific performance test standards; like spot welding, arc welding, sealing, etc.

More relevant to the present UGV performance testing program is the robot dynamic performance characteristics evaluation standard. The future UGV tests, away from the Denver Martin-Marietta site, would come closer to application specific performance tests.

Appendix III. Topographic Engineering Center Support

DEPARTMENT OF THE ARMY
UNITED STATES ARMY TOPOGRAPHIC ENGINEERING CENTER
FORT BELVOIR, VIRGINIA 22060-5546

CETEC-RI-T

28 January 1993

MEMORANDUM TO NIST (Ken Goodwin)

SUBJECT: Potential TEC Support for UGV Performance Evaluation

1. INTRODUCTION:

This memorandum outlines potential support by the U.S. Army Topographic Engineering Center (TEC) to the National Institute for Standards and Technology (NIST) to the DoD Unmanned Ground Vehicle (UGV) Performance Evaluation Program. This response follows discussions held with NIST at TEC on 19 January 1993. Two initial areas of activity are suggested:

- Geolocation of moving and/or static vehicles and targets using Global Positioning System (GPS) technology;
- Generation of a custom high-resolution Digital Elevation Model (DEM) and Digital Orthophoto for the RSTA field-of-view at the UGV Denver Testsite from Controlled Stereo Imagery (CSI).

2. BACKGROUND:

a. Since 1983, TEC and its predecessor, the Army Engineer Topographic Laboratories (ETL), has supported aspects of the DARPA Autonomous Navigation Program including in-house efforts for terrain database generation, technical consulting and contract management.

b. Currently, TEC's Autonomous Technology Division has responsibility for digital terrain data under the DARPA Autonomous Navigation Program and the DoD Unmanned Ground Vehicle (UGV) Program. Various digital terrain products are being generated and distributed to the UGV research and development community to support UGV mission planning and operations (see attachment).

c. TEC's Surveying Division conducts a research and development program in precision survey technology. On-going efforts emphasize exploitation of Global Positioning System (GPS) technology for positioning and navigation of both static and moving platforms.

3. REAL-TIME GEOLOCATION

a. Purpose. To evaluate feasibility and demonstrate capability of using GPS in a differential mode to provide time-tagged object positions for performance evaluations of unmanned ground vehicles.

b. Functional Objective: For UGV Demo A in Spring 1993 at the Denver Testsite, position one (1) vehicle to an accuracy of at least one (1) meter with a potential for 0.1 meter in three-dimensions based on positional time-tags to within one (1) microsecond and post-processed positional computations.

c. Products. Time-tagged positions of one (1) vehicle for the duration of the demonstration; a report describing data collection and analysis; and preliminary design for full integration of GPS performance evaluation system into UGV Demos B, C & II.

d. Costs and Resources:

	Man-Weeks	$K
System Design	6	11.0
Hardware Procurement	-	60.0
Installation	1	2.0
Data Collection	3	6.0
Analysis/Report	4	8.0
Design Report	5	10.0
TOTAL	16	97.0

4. HIGH-RESOLUTION DIGITAL TERRAIN DATA

a. Purpose. To support UGV performance evaluation of off-road navigation and RSTA with high-resolution digital terrain data for a limited area at the UGV Denver Testsite.

b. Functional Objective: Provide custom digital terrain data for the RSTA field-of-view to support UGV Demo A and B performance evaluation.

c. Products. A custom one-meter Digital Elevation Model (DEM) and Digital Orthophoto for the RSTA field-of-view at the UGV Denver Testsite from Controlled Stereo Imagery (CSI)

d. Costs and Resources:

	Man-Weeks	$K
Custom DEM/Orthophoto	4	8.0
Consultation	2	4.0
Report	3	6.0
TOTAL	9	18.0

5. We have not attempted to formulate a detailed proposal for future UGV demos without a better understanding of the performance evaluation requirements of each demonstration. A detailed proposal will require coordination to determine accuracy requirements, data timeliness, and the number of objects to be evaluated. For very preliminary planning purposes, estimate the costs for precise positioning of one (1) vehicle to an accuracy of 0.1 meter throughout the duration of a demo at approximately $100,000; real-time positioning of each additional vehicle would be about $50,000. Costs associated with detailed terrain material characterization would depend on the area to be mapped as well as the level of generalization.

5. TEC TECHNICAL POINT-OF-CONTACT:

George E. Lukes — Phone: 703/355-2732
Autonomous Technologies Division — FAX: 703/355-3176
Research Institute — Internet: lukes@tec.army.mil

GEORGE E. LUKES
Chief, Autonomous Technologies
Division and DARPA Agent

Unmanned Ground Vehicle Program

Concept of Operations for Digital Terrain Data

For the full area of operations:
- DTED Level 2 (≈ 30 meter posts) or equivalent
- Interim Terrain Data (ITD) or equivalent
- Controlled Stereo Imagery (CSI) at one meter GSD

For UGV mission corridors:
- Custom DEM (5 meter posts) derived from CSI
- Waypoint and landmark coordinates via point-positioning from CSI

On-board the unmanned ground vehicle:
- Real-time range sensor(s) ($\Delta r \approx 30$ cm)

UGV Terrain Data Status

Initial terrain data package (Killeen, TX)
- DTED Level 2 -- distributed
- Interim Terrain Data (ITD) -- distributed

Demo A/B Testsite (Martin Marietta Facility, Denver, CO)
- USGS 30 meter DEM -- distributed
- Controlled Stereo Imagery -- distributed
- Custom (5 meter) DEM -- distributed
- Digital Orthophoto (1 meter) -- distributed
- Updated Feature Data -- distributed

Demo C/II Testsite (Blackwell Mountains, Fort Hood, TX)
- ITD -- distributed
- DTED Level 1 -- distributed
- DTED Level 2 -- TBD
- Controlled Stereo Imagery -- TBD
- Custom (5 meter) DEM -- TBD
- Digital Orthophoto -- TBD