Evaluation of Autonomous Ground Vehicle Skills

Phillip L. Koon

CMU-RI -TR- 06-13

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

March 2006

Abstract

Autonomous ground vehicles must achieve bold performance and solid reliability to mature from laboratory curiosities to fielded systems. Currently, there are no standard methods to measure, validate and compare the performance of autonomous unmanned ground vehicles, hence the impetus for this research. This paper documents the test methods implemented by Carnegie Mellon University's Red Team while preparing robots for the DARPA Grand Challenge. The Red Team's test methods were developed to enable quantitative evaluation of the effects of unit changes to the robots' hardware and software on the robots' over all ability to blindly track, track with perception assistance and modify based on perception a preplanned path. This paper describes tests for evaluating and comparing navigational skills of autonomous ground vehicles. Test data collected from the Red Team's H1ghlander and Sandstorm autonomous unmanned ground vehicles is presented. Suggestions for future test methodology research and test standardization are also discussed.

1 Introduction

This research was conducted in conjunction with Carnegie Mellon University's Red Team's preparations for the 2005 DARPA Grand Challenge. The 2005 DARPA Grand Challenge was a 132 mile (212 kilometer) race of autonomous ground vehicles along paved roads, dirt roads and trails through the Mojave Desert. The following describes the Red Team's formulation and execution of a test program to measure the quality of autonomous ground vehicle driving skills. The program pushed performance to the edge of driving ability without taking extraordinary risks. The paper documents how Red Team regressively used the test program to evaluate effects of unit changes to hardware and software on the overall driving skills of the autonomous ground vehicles.

2 Acknowledgement of Support

This research would not have succeeded without the support of Red Team's Systems Test group. In particular the hard work and dedication of Michael Clark, Test Conductor, for organizing a team to carry out and execute the tests as planned. The development of the test concepts described in this document was closely coordinated with Red Team's Software Leads Dr. Chris Urmson and Kevin Peterson. Members of the Systems Test Group include: Jason Ziglar, Josh Johnston, David Ray, Tim Reid, Chris Pinkston, Evan Tahler, Bhas Nalabothula, Josh Struble, Aaron Mosher and Jarrod Snyder.

The author would also like to thank Dr. Red Whittaker, Dr. Dimi Apostolopoulos and Juan Pablo Gonzalez for review and advise during the development of this paper.

3 Red Team Racing System

The Red Team Racing System creates path plans that its autonomous ground vehicles follow at high speed. The vehicles modify the preplanned paths as required based on local perception of the world. Figure 1 shows the major components of the Red Team Racing System. Figure 1 and the description of the Red Team Racing System are paraphrased or quoted from the Red Team[1] and Red Team Too[2] 2005 DARPA Grand Challenge Technical Papers.

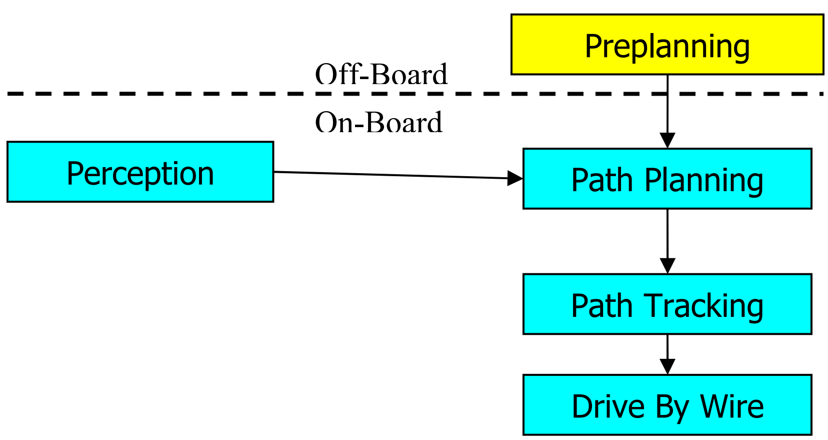

Figure 1 Major architectural components of Red Team Racing System.

The Preplanning function was performed prior to test sessions and prior to running races. Preplanning converts the DARPA or Test Designer provided route data definition file (RDDF) into a path definition file (PDF) that Red Team autonomous ground vehicles can follow. The PDF, a preplanned path, is loaded onto the autonomous ground vehicles. This file defines the waypoints, corridor and speeds the robots will attempt to pass, stay within and achieve during a race or test.

The Perception function interprets range data collected from lasers and radar creating a terrain cost map and binary obstacle map. These maps are delivered to the Path Planning function.

The Path Planning function fuses the preplanned path, terrain cost map and binary obstacle map into a world model. Items in the binary obstacle map are fused to the terrain cost map by adding them as high or infinite cost, while clear traverses are added as low or no cost. Path Planning uses an A-star algorithm which considers multiple possible traversable arcs forward of vehicle position within the preplanned path's route corridor. Each possible arc is evaluated in terms of least cost to goal. The "best" path at any given interval is then communicated to the Path Tracking function. Areas outside of the path definition file's route corridor are not considered in path planning[3].

The Path Tracking function evaluates a best path relative to current vehicle position and pose. The algorithm sets maximum speed and curvature and constrains the trajectory to ensure against skidding and tip over. Path tracking passes calculated commands of desired curvature and speed to the vehicle's drive-by-wire system.

The Drive By Wire function receives curvature and speed commands and converts them into control signals which position the steering, brake and throttle actuators appropriately. The Drive By Wire function monitors vehicle steering and speed and updates actuator commands appropriately to maintain the last commanded values.

4 Test Objective

The impetus for the Red Team's test methodology was a desire to measure the threshold of its robots' driving skills. The team used the tests in a regressive manner to evaluate the effects of unit changes in hardware and software on the robots' over all ability to drive. Three major skills constitute driving ability. The first skill is the ability of the robots to follow a preplanned path based on position sensing only. The second skill is the ability of the robots to track a preplanned path while assisted by perception sensors. The third skill is the ability of the robots to dynamically modify the preplanned path to avoid sensed obstacles.

5 Literature Review

The literature does not yet chronicle the subject of autonomous ground vehicle test related to measuring driving skills at speed. Review of technical reports of DARPA Grand Challenge 2005 finishers Stanford Racing Team[4], Team Gray[5] and Team Terramax[6] found they all placed great value on testing but did not mention specific tests to measure driving skill. Literature review found documentation of tests to measure an unmanned vehicle's ability to track a path at low speed[7&8]. An example of this is Shilcutts, Apostolopoulos' and Whittaker's tests to validate a meteorite search robot's ability to navigate a pattern path. Although this test did validate path tracking it required post test analysis to determine pass or fail. The test did not validate the integration of perception into the navigation process. Tests of driver assist technologies are also well documented[9,10&11]. These works describes tests of lane departure and collision avoidance warning systems. They all involve humans in the loop and are more qualitative than quantitative in nature. This lack of documented tests for autonomous vehicle driving skills at high speed led to a search of current industry standards for testing automobiles.

6 Test Formulation

Research of published literature on the subject of automotive dynamic testing revealed International Organization for Standardization standard ISO-3888-1, Passenger cars — Test track for a severe lane-change maneuver Part 1 – Double lane-change[12] (Reference Figure 2). This test was designed as a means to subjectively evaluate vehicle dynamic performance. The test is subjective because it only quantifies a small part of a vehicle's handling characteristics and is highly dependent on the input from the driver. This dependence on driver skill is what made the test attractive to the author for adaptation to autonomous ground vehicle driving skill testing.

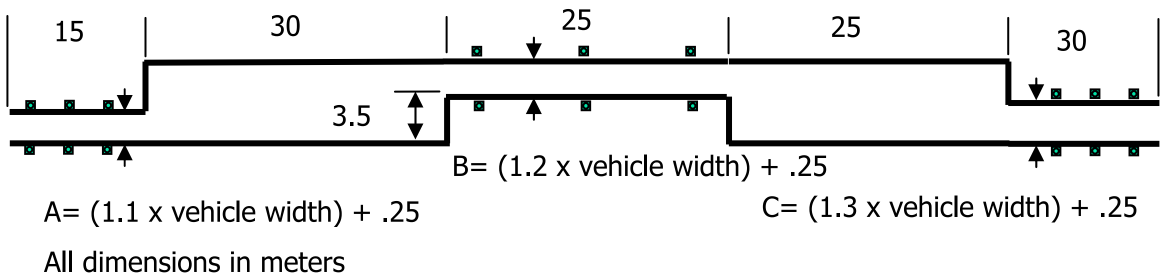

Figure 2 ISO-3888-1 Test track for a severe lane change maneuver.

Known error in the Red Team's autonomous ground vehicles' pose sensor of 1 meter RMS and estimated path tracker of 0.5 meter RMS drove the team to modify the original ISO-3888-1 course adding 1.5 meter to lane width in all sections. This additional lane width enabled quantitative measurement of performance considering total system error. The course shown in Figure 2 can be used with any autonomous ground vehicle by adjusting the lane width as shown. The length of the course is fixed for all size vehicles at 125 meters. Table 1 describes the basic steps used in the autonomous ground vehicle path tracking skills assessment. Ideally reliability would be validated by running the test over several days without system configuration changes.

Table 1Test steps for autonomous ground vehicle path tracking skill assessment.

1.	Create a route file which traverses through the ISO-3888-1 test track.
2.	Create a path definition file for the route created in step 1 setting the corridor width slightly wider than the test track's lane width and the speed to a constant (e.g., 5 meters/sec). Path definition file must include an area before the test course begins for the robot to achieve the required constant velocity.
3.	Load the path definition file into the autonomous ground vehicle
4.	Command the autonomous ground vehicle to drive the route described in the path definition file.
5.	Record the time the autonomous ground vehicle is on the test track entry to exit.
6.	Record the number of times the autonomous ground vehicle touches or exits the test track's boundaries.
7.	Repeat steps 2 through 6 increasing the speed by an incremental value (e.g., 2 meters/second) until the autonomous ground vehicle can no longer successfully traverse the course or the operation is deemed to be unsafe. Multiple runs at each speed increment are required to demonstrate consistency.

7 Blind Path Tracking Test

The initial skill test conducted on the modified ISO-3888-1 test course was blind path tracking. Red Team created the route file for this test with a human driving the robot through the test course and recording the position output of the robot's pose sensor. The path definition file was created setting the corridor to 4 meters on either side of the path center line. Figure 3 is a graphical representation of the path definition file used in the blind and perception assisted path tracking tests. When conducting the blind path tracking test the autonomous ground vehicle is configured such that only the pose sensor is considered in path planning. The absence of all perception sensor input will limit path tracking error to that induced from the pose sensor, path tracking algorithm and drive by wire actuation control errors.

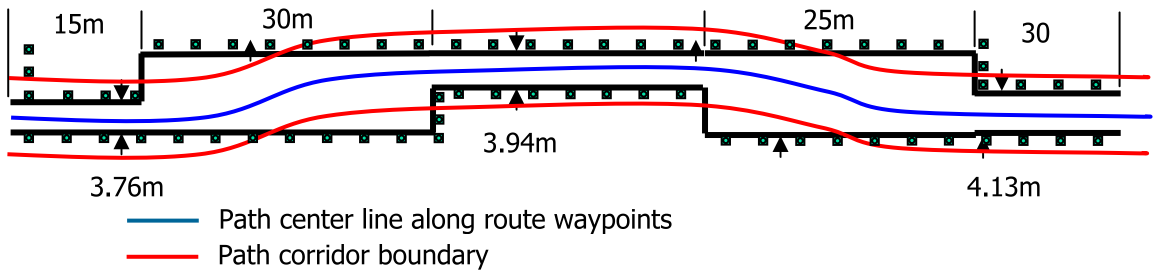

Figure 3 Blind and perception assisted route through modified ISO-3888-1 test track.

8 Perception Assisted Path Tracking Test

The perception assisted tracking test is conducted using the same path definition file used in the blind path tracking test and shown in Figure 3. Note that at the entrance to each lane segment a perpendicular boundary wall has been added to the original ISO-3888-1 test track. These boundary walls ensure the autonomous ground vehicle has no possibility of planning a trajectory that does not go through the desired lane. The autonomous ground vehicle is configured to use all of its perception sensors and the pose sensor date when conducting path planning. The test measures driving skill when given a nominal path through an area constrained by boundary obstacles and assisted by perception.

9 Perception Planning Test

The perception planning test is conducted on the same modified ISO-3888-1 test track but uses a route file that follows the center line of test tracks outside boundaries as depicted in Figure 4. This test measures an autonomous ground vehicle's ability to modify the preplanned path based on perception.

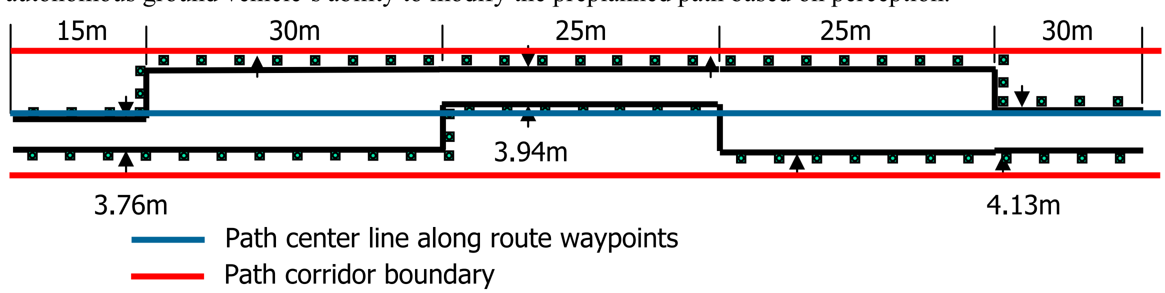

Figure 4 Perception planning route through modified ISO-3888-1 test track.

10 Test Tracks Utilized

Red Team conducted field tests at the LTV Steel facility along the Monongahela River in Pittsburgh, Pennsylvania and at the Nevada Automotive Test Center (NATC) in Silver Springs, Nevada. The test tracks at both sites operated on roads that had been created or used for conduct the DARPA Grand Challenge 2005 mandatory site visit demonstrations of Red Team's two autonomous ground vehicles. The site visit course was 200 meters in length and included two 31 degree turns (Turns were required to be ≥ 30 degrees). Figure 5 is a graphical representation of the modified ISO-3888-1 test track implemented by the Red Team. The implementation of the modified ISO-3888-1 test track featured the double lane change maneuver in the long leg of the 200 meter long site visit course. The road boundary of the track was defined with traffic cones. While testing at the LTV Steel site Red Team used small cardboard boxes wrapped in plastic bags as lane boundaries. While testing at the NATC Red Team migrated from boxes to traffic cones for lane boundaries. This migration consisted of defining the lane center line with cones and the perpendicular wall with boxes at first then cones later. Figure 6 includes images of Red Team's autonomous ground vehicles operating on the test tracks at LTV Steel and NATC.

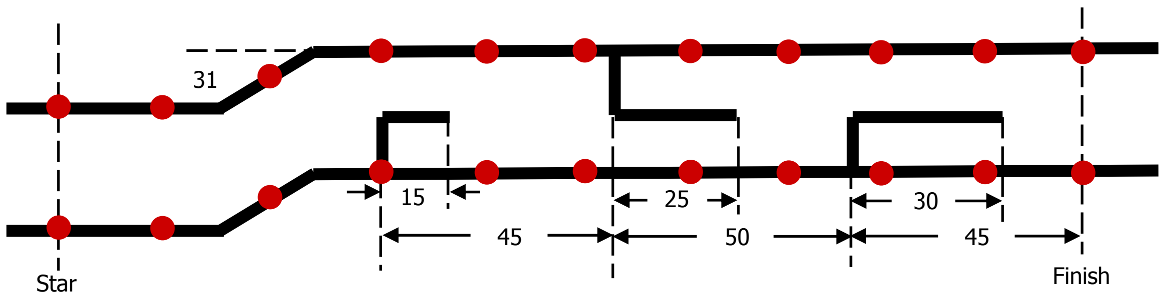

Figure 5 Modified ISO-3888-1 test track as implemented by Red Team.

Figure 6 Sandstorm (Left) and H1ghlander (Right) on the LTV and NATC test tracks.

11 Test Execution

Red Team conducted formal tests on the modified ISO-3888-1 test track on eight separate occasions. Test personnel included test conductor, robot operator, chase car driver and timers. The basic plan was to conduct a complete set of blind path tracking, perception assisted tracking and perception planning during each test session. Table 2 is an example of data collected during a typical test session on the modified IS)-3888-1 test tracks.

Table 2 Typical data collected during tests.

Test #	TOD	Type	200 m (sec)	Speed m/s	mph	Notes
1	2:35	BT-5	40.98	4.88	10.92	Cone 9L brushed
2	2:40	BT-5	41.20	4.85	10.86	Cone 4L brushed
3	2:44	BT-5	41.13	4.86	10.88	Good
4	2:49	BT-7	29.20	6.85	15.32	Cone 4L hit hard
5	2:52	BT-7	29.45	6.79	15.19	Cone 4L hit hard
6	3:09	BT-7	31.70	6.31	14.11	Cone 4L hit more hard, robot moving left
7	3:12	BT-7	31.67	6.32	14.13	Cone 4L hit more hard, robot moving left
8	3:15	BT-9	ABORT			E-Stop because robot leaving course at 4L
9	3:40	PT-5	45.58	4.39	9.82	Good
10	3:58	PT-5	45.26	4.42	9.89	Good
11	4:04	PT-5	43.64	4.58	10.25	Good
12	4:09	PT-7	36.67	5.45	12.20	Good
13	4:11	PT-7	37.29	5.36	12.00	Good
14	4:15	PT-7	38.70	5.17	11.56	Cone 7R hit
15	4:17	PT-9	29.70	6.73	15.06	Cone 7R hit
16	4:21	PT-9	31.27	6.40	14.31	Cone 7R hit
17	4:25	PP-5	45.89	4.36	9.75	Ctr 3rd wall hit
18	4:28	PP-7	39.72	5.04	11.26	Corner 1st wall brushed, 3rd wall corner crushed
19	4:30	PP-9	37.60	5.32	11.90	Corner 1st wall brushed, 3rd wall corner crushed
20	4:32	PP-9	NA			2nd outer and ctr wall hit, 3rd corner crushed
21	4:37	PP-9	33.61	5.95	13.31	3rd wall corner, 2 wall ctr punch & corner, 1st wall corner box crushed

Red Team started sessions on the modified ISO-3888-1 track with the blind path tracking test at an initial speed of 5 meters per second. The team would execute a minimum of three runs then increase the speed by 2 meters per second. At each speed increment three runs were executed. Speed was increased until the vehicle left the course and had to be stopped via the emergency stop link or the test team deemed operations at higher speeds were unnecessary. The blind path tracking test was never conducted at speeds above 13 meters per second. As confidence in the robot's blind tracking ability increased the number of blind tracking tests were decreased and eventually were not included in the test routine. The test was held in reserve for regression testing after hardware or software changes were made to a robot that would affect the basic path tracking. Examples of system changes affecting path tracking include steering position sensor, steering control algorithm, pose sensor, path tracking algorithm, etc.

Execution of perception tracking tests started with recording of the active perception sensor configuration. Initial speeds for perception tracking were set at 5 meters per second. Multiple runs at the slower speeds were eventually found to be unnecessary. As in the blind path tracking incremental speed changes were of 2 meters per second and maximum speed was 13 meters per second. As confidence in the perception sensing systems ability to correctly sense and localize obstacles in the world view increased emphasis on conducting perception tracking was diminished.

Execution of the perception planning test was identical to the perception tracking test with the exception of using a different path definition file (Reference Figure 4).

12 Data Evaluation

The initial blind tracking test of Sandstorm, Figure 7, conducted on 17-June-2005 revealed the robot did not have a robust path tracking ability. The observed quality of driving was poor with significant overshoot when cornering. The test was aborted after Sandstorm left the roadway at 9 m/sec. The path tracking control algorithm was modified adding an integral term and when the test was repeated on 17-August-2005 performance was notably better. Observed quality of driving found it to corner smoothly without significant overshoot. The team was satisfied with Sandstorm's blind tracking performance at this point and did not conduct the test on Sandstorm again.

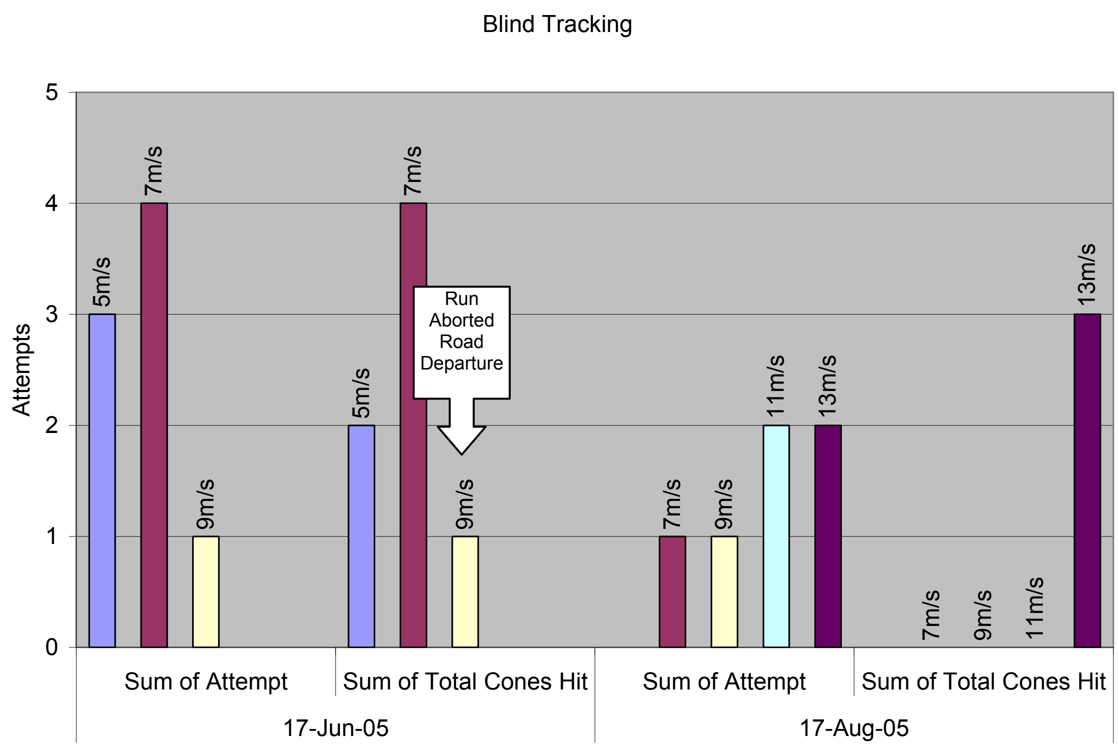

Figure 7 Sandstorm Blind Tracking

H1ghlander's blind tracking data, Figure 8, is a little deceptive. Observed quality of driving on 1-July-2005 was good with smooth cornering and minimal overshoot. Minor adjustments were made to the path tracking control algorithm (H1ghlander did not use an integral term) and performance marginally improved on 21-August-2005. The team was satisfied with H1ghlander's blind tracking performance at this point and did not repeat the test.

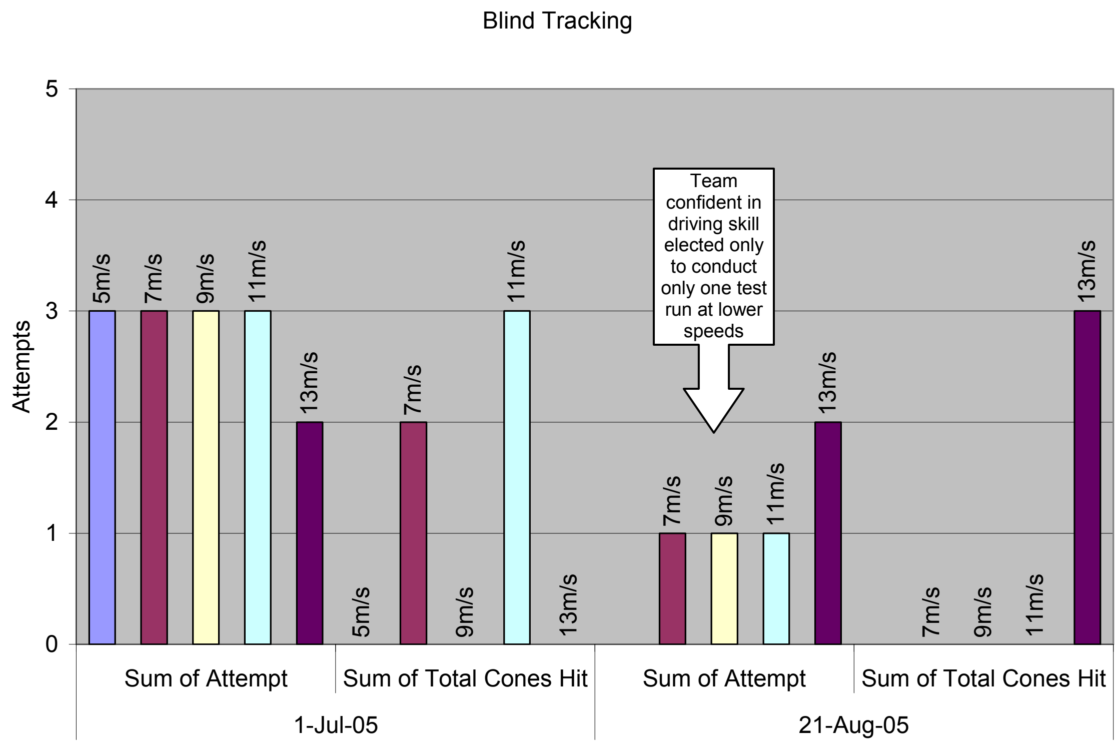

Figure 8 H1ghlander Blind Tracking

Analysis of the data collected during the perception tracking and planning tests is inconclusive of performance gains due to changing hardware and software configurations. As an example, performance of the vehicles is observed to decline in September after the short range LIDARS were removed and replaced during addition of sensor washing hardware. This decline was directly attributable to the lack of adequate calibration of the sensors for object localization.

13 Conclusions

The Blind Tracking, Perception Tracking and Perception Planning tests are an effective tool for measuring autonomous ground vehicle driving skill. The tests are relatively easy to set up and inexpensive to conduct. Red Team has found the tests an effective means to evaluate hardware and software configuration changes.

The blind tracking test is an excellent tool for measuring an autonomous ground vehicle's ability to blindly follow waypoints. The test is applicable to all automotive and truck class autonomous ground vehicles.

The perception tracking test is a good tool to measure the effects of perception on path tracking and also as a subjective qualitative of driving quality. An example qualitative analysis, does the path planner attempt to maximize distance from perceived obstacles thus centering in the lane or does it attempt to smooth the trajectory maintaining the current trajectory as long as it is clear and within corridor constraints.

The perception planning test is a good tool to measure the effectiveness of an autonomous ground vehicle's ability to dynamically adapt to items impeding the preplanned path. The perception planning test is limited in that it measures the vehicle's ability to rapidly adapt to obstacles. The test does not measure the smoothness of a vehicle's reaction. For example, when humans are driving and see obstacles they will generally react as early and smoothly as possible to change their trajectory to avoid obstacles. The perception planning test could be adapted for this purpose by elongating the length of the track segments between lane change barriers.

14 Future Work

The Red Team has advanced the technology readiness level [13,14,15] of its robotic racing system from about TRL 3 (analytical or experimental characteristic proof of concept) to about TRL 5 (Technology component demonstration in a relevant environment) while preparing for the DARPA Grand Challenge. This change is largely attributable to the rigorous test program implemented. In order to achieve TRL 9, actual technology system qualified, a much wider set of systems tests must be developed. Although the tests described above effectively measure an autonomous ground vehicle's abilities to track and plan a path they are not effective at measuring perception skills. Standard tests need to be developed that measure an autonomous ground vehicles ability to sense and accurately localize obstacles of varying size. These tests should account for differing perception sensing modes. Standard tests that measure an autonomous vehicle's ability to safely and reliably interact with other vehicles and humans are needed. This is only a fraction of the tests needed to move autonomous ground vehicles from technological curiosities to common tools used by people everywhere.

15 References

[1] Red Team Racing. Red Team DARPA Grand Challenge 2005 Technical Paper. http://www.darpa.mil/grandchallenge/TechPapers/RedTeam.pdf

[2] Red Team Racing. Red Team Too DARPA Grand Challenge 2005 Technical Paper. http://www.darpa.mil/grandchallenge/TechPapers/RedTeamToo.pdf

[3] Vanessa Hodge, Kevin Peterson, William "Red" Whittaker, "Scaled, Curvilinear Grids for Maneuvering Extreme Routes at High Speeds,"
Proceedings of the International Conference on Field and Service Robotics, 2005.

[4] Stanford Racing Team. Stanford Racing Team's Entry In The 2005 DARPA Grand Challenge. http://www.darpa.mil/grandchallenge/TechPapers/Stanford.pdf

[5] Paul G. Trepagnier, Powell M. Kinney, Jorge E. Nagel , Matthew T. Dooner and Jason S. Pearce. Team Gray Technical Paper DARPA Grand Challenge 2005. http://www.darpa.mil/grandchallenge/TechPapers/GreyTeam.pdf

[6] Oshkosh Truck Corp, Oshkosh, WI Rockwell Collins, Cedar Rapids, IA Rockwell Scientific, Thousand Oaks, CA University of Parma, Parma, Italy. TEAM TERRAMAX DARPA GRAND CHALLENGE 2005. http://www.darpa.mil/grandchallenge/TechPapers/TeamTerraMax.pdf

[7] K. Shillcutt, D. Apostolopoulos, and W.L. Whittaker. Patterned Search Planning and Testing for the Robotic Antarctic Meteorite Search 1999 International Topical Meeting on Robotics and Remote Systems for the Nuclear Industry, American Nuclear Society, April, 1999.

[8] S.A. Roth and P. Batavia. Evaluating Path Tracker Performance for Outdoor Mobile Robots, Automation Technology for Off-Road Equipment, July, 2002.

[9] L. Tijerina, J.J. Jackson, D. Pomerleau, R.A. Romano, and A.D. Pertersen. Driving Simulator Tests of Lane Departure Collision Avoidance Systems, ITS America Sixth Annual Meeting, April, 1996.

[10] P. Batavia. Driver Adaptive Lane Departure Warning Systems, doctoral dissertation, tech. report CMU-RI-TR-99-25, Robotics Institute, Carnegie Mellon University, September, 1999.

[11] IVsource.net and Richard Bishop Consulting (RBC). Advanced Collision Avoidance Field Test Launched in Michigan 29 March 2003, http://www.ivsource.net/archivep/2003/mar/030329_ACASlaunch.html#formoreinfo

[12] International Standards Organization. Passenger cars — Test track for a severe lane-change manoeuvre — Part 1: Double lane-change First Edition 1999-10-01

[13] J. C. Mankins, Technology Readiness Levels: A White Paper. Office of Space Access and Technology, NASA, http://www.hq.nasa.gov/office/codeq/trl/trl.pdf

[14] AMS Guidance on Technology Readiness Levels (TRLs), http://www.ams.mod.uk/ams/content/docs/trlguide.doc

[15] Performance Measures for Intelligent Systems: Measures of Technology Readiness PERMIS'03 White Paper. http://www.isd.mel.nist.gov/research_areas/research_engineering/Performance_Metrics/PerMIS_2003/Proceedings/WhitePaper.pdf