Achieving intelligent performance in autonomous on-road driving

Craig Schlenoff[1], John Evans[2], Tony Barbera[1], Jim Albus[1], Elena Messina[1], Stephen Balakirsky[1]
[1]NIST, 100 Bureau Drive, Stop 8230, Gaithersburg, MD USA 20899
[2]John M Evans LLC, 13 Mount Pleasant Road, Newtown, CT 06470

ABSTRACT

This paper describes NIST's efforts in evaluating what it will take to achieve autonomous human-level driving skills in terms of time and funding. NIST has approached this problem from several perspectives: considering the current state-of-the-art in autonomous navigation and extrapolating from there, decomposing the tasks identified by the Department of Transportation for on-road driving and comparing that with accomplishments to date, analyzing computing power requirements by comparison with the human brain, and conducting a Delphi Forecast using the expert researchers in the field of autonomous driving.

A detailed description of each of these approaches is provided along with the major finding from each approach and an overall picture of what it will take to achieve human level driving skills in autonomous vehicles.

Keywords: autonomous navigation, on-road driving, human-level performance, 4D/RCS, MARS

1. INTRODUCTION

The Intelligent Systems Division at the National Institute of Standards and Technology (NIST) has been supporting the Defense Advanced Research Projects Agency (DARPA) Mobile Autonomous Robot Software (MARS) program over the past five fiscal years.

Dr. Doug Gage, the MARS Program Manager, proposed that a significant benchmark for autonomous driving would be a system equivalent to a human chauffeur. This "robot chauffeur" would be able to navigate roads and traffic on highways and in cities, finding and driving to a requested destination. This is, more or less, the capability that Army recruits bring with them to boot camp. The Army then provides additional training for those selected to be Scouts, adding specific skills in off-road driving and understanding of tactical behaviors. The Army could provide the same incremental training for an autonomous system to produce a capable robot scout.

The questions important to planning at DARPA and the Army are, then, when will we achieve human equivalent driving capability and how much effort will it take?

NIST has addressed this question in four different ways:

- Extrapolating from the State of the Art as represented by the Army Demo III Experimental Unmanned Ground Vehicle project
- Estimating the amount of effort to build an autonomous driving system with the capabilities defined by the Department of Transportation Manual <u>Driver Education Task Analysis</u>
- Estimating necessary computer processing capability by comparison with the human brain; and
- Using a Delphi Forecast to poll the MARS researchers to obtain a consensus estimate of experts in the field of autonomous driving.

This paper argues from several different standpoints as to what might be the levels of effort required to achieve the "robot chauffer." There is a trade-off between time to achieve a goal and levels of funding; we estimate time frames assuming current levels of funding and then point out the chances to reduce those time frames.

Mobile Robots XVII, edited by Douglas W. Gage, Proceedings of SPIE Vol. 5609
(SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.580221

The first question to address is the target goal point: what vehicles are we trying to drive and where are we driving them? This paper assumes that appropriate vehicle platforms are being developed under other programs. For example, the XUV platform used in the Demo III program was specifically developed for autonomous scout missions. Future Combat Systems is developing three new platforms, a small sensor platform, the Unmanned Armed Reconnaissance Vehicle (UARV), and a robot "Mule" transport vehicle. In addition, the UGCV program has an articulated vehicle under development and the Tactical Mobile Robot (TMR) program developed the "Packbot" and "Throwbot" platforms that will be wrapped into FCS but which are not suitable for highway driving. Finally, many different vehicles have been converted for tele-operation by the Department of Defense and could be further modified for autonomous driving by the addition of an Autonomous Navigation System package of sensors, computers, and software.

The primary targets for advanced autonomous driving capability are the FCS and Demo III platforms. These are under development with substantial funding commitments and will be available in production versions before intelligent on-road driving is achieved. Production versions of wheeled vehicles are expected to be qualified for highway driving.

Appropriate vehicle platforms and the Autonomous Navigation System (ANS) baseline are assumed. The problem set that needs to be addressed, then, is the sensors, the computing platforms and the software beyond the required ANS capabilities of supervised teleoperation that are needed for intelligent on-road driving.

This paper focuses on the sensors, computers and software for autonomous on-road driving, the "robot chauffeur," with Future Combat Systems as the primary ultimate customer. Section 2 summarizes the state of the art in terms of the Demo III experience. Section 3 provides a task analysis based on the DOT manual. Section 4 considers the needs for improved sensors. Section 5 analyzes computing power requirements. Section 6 presents the results of the Delphi Forecast carried out at the April 2003, MARS Principal Investigators' meeting in San Diego. Finally, Section 7 itemizes the main conclusions drawn in earlier chapters.

Note that this paper is a condensed version of a longer, more detailed paper which can be found at [2].

2. CURRENT STATE OF THE ART

In order to determine how much is it going to take to reach intelligent performance in on-road and off-road driving, we must first understand what is achievable now. We can use our current capabilities as a benchmark, and extrapolate out to determine what it would take to achieve an intelligent level of on-road driving performance.

The Demo III Experimental Unmanned Vehicle (XUV) effort seeks to develop and demonstrate new and evolving autonomous vehicle technology, emphasizing perception, navigation, intelligent system architecture, and planning. [10] Many believe that this effort represents the state of the art in autonomous driving. As such, we will use this effort to serve as a benchmark to represent what we can do now, and then project to the capabilities needed to enable intelligent levels of performance.

The autonomous navigation system (ANS) within the Demo III effort was recently declared to have reached Technology Readiness Level 6 (TRL-6), indicating that the ANS has been demonstrated and tested in a relevant environment. [3] Though focusing primarily on off-road driving, the authors believe that the technology used in Demo III will lend itself well as a starting point for on-road driving.

Within Demo III, the 4D Real-Time Control System (4D/RCS, the 4D referring to planning in three spatial dimensions plus time, as used in the German autonomous driving program) was used as the underlying architecture within the autonomous mobility system. This architecture provides a reference model for the identification and organization of software components for autonomous driving of military unmanned vehicles. 4D/RCS defines ways of interacting to ensure that missions, especially those involving unknown or hostile environments, can be analyzed, decomposed, distributed, planned, and executed intelligently, effectively, efficiently and in coordination. To achieve this, the 4D/RCS reference model provides well-defined and highly coordinated functional modules for sensory processing, world modeling, knowledge management, cost/benefit analysis, and behavior generation, and defines the interfaces and messaging between those functional modules. The 4D/RCS architecture is based on scientific principles and is consistent with military hierarchical command doctrine. [1]

Sensory processing algorithms use sensor data to compute vehicle position, range, obstacle lists, obstacle positions, and terrain information. The suite of sensors used in the mobility system include a General Dynamics/Schwartz Electro-Optics Scanning Laser Rangefinder (LADAR)[1], a pair of color cameras for stereo vision, a stereo pair of Forward-Looking Infra-Red (FLIR) cameras, a stereo pair of monochrome cameras, a pan-tilt platform, a global positioning system (GPS) sensor, a force bumper that alerts the system to obstacles in the vehicle's immediate path, and an inertial navigation system (INS) sensor. All sensors are mounted on the vehicle, which is equipped with electric actuators on the steering, brake, transmission, transfer case, and parking brake. Feedback from the sensors provides the controller with engine rotations per minute, speed, temperature, fuel level, etc. A Kalman filter computes vehicle position and orientation using data from the internal dead reckoning system and the differential GPS unit.

Within Demo III, positive and negative obstacles can be detected, but little object classification is performed. Using the LADAR, terrain is only classified as either vegetation or ground. By adding color images from cameras, terrain can be further classified as green vegetation, dry vegetation, soil/rock, ruts, tall grass, and outliers, but only at very course resolution.

The primary form of knowledge representation in the world model is multiple occupancy grid maps with different size cells as a function of the planning horizon at different levels of control. Underlying data structures are used to associate terrain features with cells in the map. Because of limitations in the object classification, only a small set of data structures are available based on sensor data, while a larger set of data structures is available based upon *a priori* information.

Planners in the Demo III vehicle use value-driven graph search techniques based upon cost-based computations at all levels within the 4D/RCS hierarchy. Multiple planners work concurrently at differing time horizons. Though higher-level planners have been developed to support tactical behaviors and have been tested in simulation, they have not been implemented in any substantial way on the Demo III vehicle. Planners have primarily performed waypoint following, obstacle avoidance, and ensuring stability of the vehicle based on the sensed support surface characteristics.

The discussion above highlights perception as the "tallest pole in the tent." Demo III has had some significant success, but it is badly nearsighted and can see only with very coarse resolution. Cost-based planning has been quite successful to the extent that the sensor generated obstacle maps contain adequate data. Large obstacles, both positive and negative, are routinely avoided and the vehicles can successfully follow roads and waypoints across modest off-road terrain.

Considering the time and resources that have been spent on Demo III, it is roughly estimated that another decade and total funding of the order of several hundred million dollars will be needed to achieve capability close to intelligent performance in driving. This is roughly a continuation of current levels of funding for approximately another fifteen calendar years. There is a trade-off between levels of funding and time to realize needed capabilities. In this case we estimate that doubling of effort (i.e. doubling of funding) would cut the time to realize intelligent driving to no more than a decade. That is, intelligent driving could be achieved within one decade and possibly as soon as 2010 if adequate funding were provided.

---

[1] Certain commercial software and tools are identified in this paper in order to explain our research. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the software tools identified are necessarily the best available for the purpose.

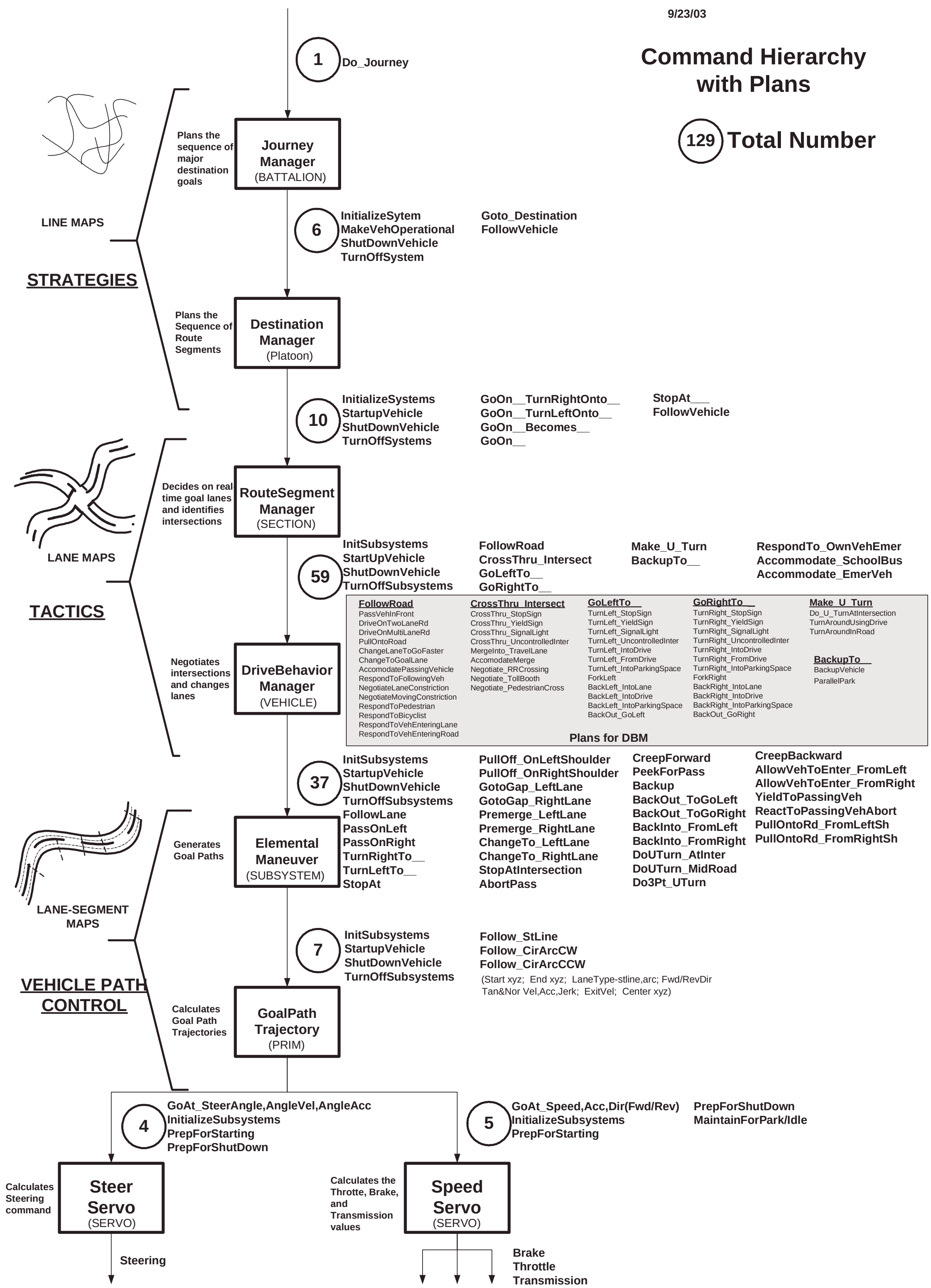

Figure 1: Command Hierarchy With Plans

3. TASK DECOMPOSITION

As part of the DARPA MARS program, an effort has focused on analyzing what it would take to achieve intelligent performance for on-road driving. The goal of this effort is to provide a task analysis description of the on-road driving task at a level of detail to be able to support work in the design and development of autonomous driving systems. This effort, therefore, requires the collection, ordering, and representation of the knowledge set that encompasses all of the on-road driving activities. This knowledge set has been assembled from a number of different sources. The single largest source document has been the Department of Transportation (DOT) manual entitled Driver Education Task Analysis, Volume 1, Task Descriptions [8], authored by James McKnight and Bert B. Adams. Significant additional sources have been the DOT Manual of Uniform Traffic Control Devices (MUTCD) document [11], numerous state traffic law documents, and considerable discussion by the authors in attempting to mine their own task knowledge.

The above documents provided a large set of the on-road knowledge as it applies to human drivers. These documents, however, have the shortcoming of not detailing the assumed driving knowledge such as the understanding of what attributes of roads and intersections are to be perceived, how vehicles are to be characterized, how objects (both animate and inanimate) are to be sensed in order to allow an autonomous computer control system to recognize and reason about them relative to the driving task context. As a result, a major effort of this work has been to attempt to define the database structures that might be used to represent all of the knowledge required about roads and entities.

3.1 Approach

The overall approach is to analyze the driving tasks through a discussion of a large number of scenarios of particular on-road driving subtasks and to derive from these descriptions a task decomposition tree representation of all the task activities at various levels of abstraction and detail. From this task tree we can organize the activities into a more rigorous layering by the artifice of identifying an organizational structure of agent control modules that are responsible for executing the different levels of the task decisions, as shown in Figure 1.

This use of separate executing agents organized into an execution hierarchy provides a mechanism to formalize the task decision tree by assigning certain decisions to particular agent control modules. This creates well-defined sets of subtask commands from each supervisor agent control module to its subordinate agent control module, thus forcing us to group and label various sets of related activities of the driving task with a context identifier such as “PassVehInFront”, “TurnLeftAt_StopSign”, “PullOffOnto_LeftShoulder” etc. Each of these identifiers is really a subtask goal command at different levels in the execution hierarchy. The task decision rules appropriate to each of these subtask goal commands that identify the partial task decomposition of the driving task that occurs within the one agent control module’s level of responsibility can be encoded within Finite State Machines (FSMs), as shown on the right side of Figure 2.

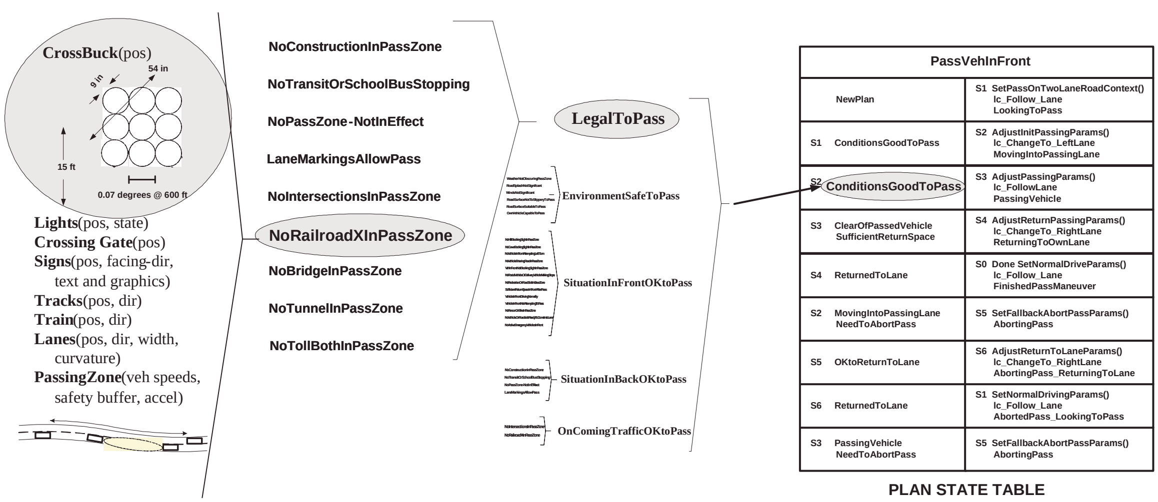

Figure 2: Task Decomposition Procedure

The FSMs are used to encode the task decomposition knowledge. Each line of each state table uses some symbolic value to describe the present situation that must be matched in order to execute the corresponding output action of that rule. The processing required to evaluate that this particular situation is true can be thought of as a knowledge tree lying on its side, funneling left to right, from the detailed sensory processing branching until all of the values have been reduced to the one appropriate situation identification encoded in a symbolic value such as "ConditionsAreGoodToPass." This lateral tree represents the layers of refinement processing made on the present set of world model data to come to the conclusion that a particular situation now exists such as "ConditionsAreGoodToPass".

The identification of these layers of knowledge processing to evaluate to the situation value is done in reverse. We know that we cannot change into the oncoming traffic lane (the "ChangeToLeftLane" action) during the passing operation until "ConditionsAreGoodToPass". Now we have to determine what are all of the things that have to be taken into consideration in order for this to be true. To determine this, many different example scenarios are reviewed to determine all of the pieces of knowledge required for all of these variations. The results are grouped by category into (in this example) five major evaluation areas. Thus, to be able to say that the "ConditionsAreGoodToPass", we first had to evaluate that each of the five sub groups were true, namely, the five situations of "LegalToPass", "EnvironmentSafeToPass", "SituationInFrontOKtoPass", "SituationInBackOKtoPass", and "OncomingTrafficOKtoPass", all had to be true.

In the example described above and shown in Figure 1, we have clustered all of the rules of the road that pertain to the passing operation at this level of task detail into the "LegalToPass" sub group evaluation. We have itemized nine world states to be evaluated and we have named them with the identifiers such as "NoConstructionInPassZone", "NoTransitOrSchoolBusStopping", "NoPassZone-NotInEffect", "LaneMarkingsAllowPass", "NoIntersectionsInPassZone", "NoRailroadXInPassZone", etc.

These world states can now be further broken into the primitive world model entities we need to be able to measure (such as vehicles, their speed, direction, location, lane markings, signs, railroad tracks, etc.) in order to determine that these world states exist. These primitive world model entities then set the requirements for the sensory processing system we need to build to support these control tasks. Everything has been determined in the context of individual tasks we want the system to be able to do.

Based upon preliminary work performed using the above analysis technique, we can estimate:

- the number of state tables that are necessary to capture all the behaviors that we wish the vehicle to execute,
- the number of situations that are needed to trip the actions in the state table,
- the number of world model states that must be true for a situation to be evaluated as true,
- the number of world model entities that must exist to be able to evaluate the world model states, and
- the number of attributes that must exist for the world model entities.

Table 1 summarizes our estimation of the number of state tables, situations, world model states, world model entities, and world model attributes we believe are necessary to enable autonomous on-road driving, as described above.

Knowledge	**Total Number**
State Tables (behaviors)	**129**
Situations	**1000**
World Model States	**10000**
World Model Entities	**1000**
World Model Attributes	**7000**

Table 1: Knowledge Summary

3.2 Comparison to capabilities in Demo III

Now that we have estimated what knowledge is necessary to enable autonomous on-road driving, we will explore how much of this knowledge has been encoded in the Demo III effort described in Section 2 to determine where we are now and how far we have left to go.

Although Demo III is focusing on off-road driving as opposed to on-road driving, it is the authors' belief that many of the same underlying functionalities at the lower levels are fundamentally the same. In both case, the vehicle is recognizing objects, planning trajectory paths, and performing lane/path following. As such, the authors feel that the Demo III effort serves as a reasonable benchmark to set time and funding estimates for implementing autonomous on-road driving.

It should be noted that although Demo III uses a cost-based planning approach as opposed to the finite state machine approach described in this section, it is still possible to draw meaningful correlation between the approaches by comparing the functionality that are able to be accomplished in each approach.

As mentioned in Section 2, much of the work exhibited in Demo III focused on waypoint following and trajectory generation. By comparing this functionality to the number of state tables, situations, world model states, world model entities, and world model attributes, we can estimate that Demo III was able to accomplish about 8% (10/129) of the tasks that are needed to achieve acceptable behavior while driving on-road.

Now, if we look at the amount of time and money that have been put into Demo III to realize that 8%, we can estimate that there has been approximately 10 calendar years of effort at a funding level of approximately 30 million dollars, fairly evenly split between the efforts of General Dynamic Research Systems (GDRS) and NIST. This is only the money that has been applied to the vehicle navigation system, not what has been applied to building the hardware for the vehicle. Assuming that all commands are at equal level of complexity, namely, that the effort needed to realize the command is equivalent for all commands, then if 30 million dollars gets you 8% of the way there, that it would take between 350 and 400 million dollars to get you 100% of the way to achieving acceptable behavior while driving on-road.

Current funding for Army autonomous mobility programs at ARL and TACOM total approximately $50M per calendar year. This funding covers many projects and only a part of it is targeting the problem this paper addresses. Funding specifically for autonomous navigation is of the order of $15-20M per calendar year. The conclusion is that, if current funding is continued, it will take more than twenty calendar years to reach intelligent driving capability.

4. SENSORS

The task decomposition described in Section 3 assumes the availability of sensors and sensory processing systems that work at a specified level such that the vehicle control system can recognize objects, and characteristics of objects, and then make appropriate decisions based upon what it sees. The task decomposition effort has progressed to the point that the requirements on the sensors and sensory processing software can be specified, as described below.

4.1. Next generation LADAR

The LADAR used in Demo III is clearly inadequate in resolution and does not have the range required for full speed highway driving. A next generation of laser range sensors has appeared on the market in the past two years, with approximately ten times the speed (600,000 points per second) and much better range (beyond 100 m). This is a very high resolution scan which takes many seconds, but the same technology could produce a 256 x 256 range image at 10 frames per second or better.

Based upon experience from Demo III and a survey of available technology, a Broad Agency Announcement (BAA) was released in June 2002. Phase 1 of the BAA focused on the design of a LADAR for on road driving with the specs shown in Table 2.

Sensor Type	Range Resolution	FOV Vert and Horiz	Resolution – Vert and Horiz	Ground Range	Vertical Surface Range	Scan Rate	Stabilization*
Wide FOV LADAR	5-10 cm or better	About 40 x 90 degs	0.25 - 0.3 degs or better	40-50 m or better	125-200 m or better	10 frames/sec or better	0.3 deg
Narrow FOV LADAR	5-10 cm or better	About 5 x 5 degs	0.05-0.06 degrees or better	40-50 m or better	125-200 m or better	10 frames/sec or better	0.03 degs
Wrap around LADAR	10-15 cm	About 0.5 x 360 degs	0.5 x 0.5 degs	N/a	50 m	About 10 frames/sec	N/a

Table 2: Next Generation LADAR Specifications

In addition to these specifications, the LADAR must also:

- Operate in full sunlight
- Be eye-safe
- Be capable of penetrating dust, fog, smoke, grass and light foliage
- Be small sized, low cost, and ruggedly designed

Based on the BAA, four Phase 1 awards were made and the results of these awards have been reviewed. Phase 2 awards, focusing on the development of the LADAR, are pending the availability of funds. Based upon the four award results, it is estimated that a prototype of a LADAR with the above specifications will take anywhere from 16-30 months to manufacture and cost between one and three million dollars.

4.2. Next generation vision system

Similar to the LADAR specifications above, the Table 3 are the specifications for camera systems that we believe can be implemented with currently available commercial technology within the next 24 months at a cost of less than one million dollars.

Sensor Type	FOV Vert and Horiz	Resolution – Vert and Horiz	Scan Rate	Stabilization
Wide FOV camera	About 21 x 28 degs	0.1 degs or better	10 frames/sec or better	0.1 deg
Narrow FOV camera	About 2 x 2 degs	0.01 degrees or better	10 frames/sec or better	0.01 degs
Wrap around camera	About 90 x 360 degs	1.0 degrees or better	About 10 frames/sec	N/a

Table 3: Color Camera Specifications

The importance of high resolution foveal vision should be emphasized as a good solution to the resolution/processing load trade. For example, the MARS work on reading road signs shows that you need high resolution to be able to read road signs, and that means the signs get quite close before they are legible if you have a single fixed resolution camera.

High resolution in only a (steerable) part of the field of view would allow signs to be read at a much greater distance. As another example, consider Dickmann's camera configuration with a high resolution central field of view and multiple cameras providing peripheral fields of view [5]. The view of the central fields of view are shown in the figure below. Note how difficult it is to really see any detail in the low resolution image but how the high resolution image provides detail but lacks any context. The two scenes together make the highway scene understandable.

4.3. Comparison with requirements

In this section, we compare the required sensor resolution that we derived from the task decomposition effort (not described in this paper) to LADAR and vision specifications expected to be available in the next 16-30 months. Table 4 shows the results.

Speed (m/s)	Needed Resolution Based on Task Decomposition	Expected LADAR Resolution – Narrow FOV (degrees)	Expected Camera Resolution – Narrow FOV (degrees)
13	0.1042	0.05	0.02
18	0.0711	0.05	0.02
27	0.0406	0.05	0.02

Table 4: Comparison of Needed and Expected Sensor Resolution

As shown in Table 4, it appears that the needed resolution from both the vision and the LADAR sensor should be available within the next 16-30 months assuming that funding becomes available to pay for the required development effort.

Clearly there is a great deal of work to be done in model based perception. A new generation of sensors is the starting point for attacking this problem. While prototypes of next generation sensors have been estimated at $3-4 Million over 2 to 3 calendar years, the total engineering effort in achieving refined, field tested and hardened deployable versions will take up to a decade and will cost $20-30 Million. The software engineering effort is at least twice that great and probably more. Achieving the required level of perception is a decade long effort costing in excess of $100 Million.

5. COMPUTER PROCESSING CAPABILITY ANALYSIS

5.1. Computing power of the human brain

Research to date has indicated the need for massive computing power to provide the necessary perception and world modeling capabilities for autonomous driving, well beyond the levels employed to date. In attempting to ballpark resources and time scales to reach minimum levels of human equivalent performance in autonomous driving, it is necessary to quantify what levels of computing are needed.

Several authors have addressed this issue, with greater or lesser credibility and generating greater or lesser levels of hostility from those who disagree with them [7][2][4,6,9]. The citations point to a range of estimates of the processing power of the human brain in the range of 10^{12}-10^{14} instructions per second.

Perception is the most compute-intensive task in routine driving. Visual processing accounts for some 10-20% of the visual cortex, auditory processing another 10% and motor control about 10%. Add to that some level of planning and symbolic reasoning needed for following traffic laws and analyzing various road situations and a level of 50% or so of the total computational capability of the brain might be employed, on an intermittent basis, in driving.

If we expect robot vehicles to be always focused on the task at hand and not subject to distraction, then we will need to be at least within an order of magnitude of the computing power of the brain to achieve human levels of performance.

We thus argue that 10^{11} instructions per second would be a good estimate of the lower end of the computing power needed (an order of magnitude below the lowest level argued above), and 10^{14} would be a highest end estimate (the highest level above).

5.2. Projecting when enough computing power will be available

Gordon Moore, one of the inventors of the integrated circuit and founder and Chairman of Intel, noted in about 1970 that the number of transistors on a chip was doubling every eighteen months[2]. This was an observation of manufacturing efficiency using ever better lithography process technology. Since the cost of a chip is more or less constant, the implication is that you get twice as much computing power per dollar every 18 months.

Moore's Law has held true for more than three decades. In fact, the doubling period has been decreasing and was approximately twelve months between 1995 and 2002 before lagging this year.

Computing power per dollar has been nearly doubling every year since 1995. This is faster than historical trends and may not continue unabated, and various doubling periods should be considered in forecasting. Using a baseline of 10^9 instructions per second per $1000 in the year 2000, we can extrapolate when different levels of processing power will be available for different assumptions of doubling periods, as shown in Table 5:

	12 Month Doubling	15 Month Doubling	18 Month Doubling
10^{11} instructions/sec	2007	2009	2011
10^{12} instructions/sec	2010	2013	2015
10^{13} instructions/sec	2014	2017	2020

Table 5: Moore's Law Predictions of Available Computing Power per $1000

It would seem that adequate computing power will be available in single processors for only $1000 between 2007 and 2015 if the estimate of 10^{11}-10^{12} instructions per second is correct.

The military is not constrained to using $1000 computers. Cluster computers with processing power of 10^{11} ips could be assembled for less than $20,000 with today's P4 or G5 or Itanium processors and 10^{12} ips could be similarly attained in three or four calendar years.

Given a development period of three to four calendar years for software to run on new computing architectures, a forecast is made of 2010 or 2011 for reaching a minimal level of human-equivalent performance in autonomous driving.

6. DELPHI FORECAST

As another approach to technology forecasting, NIST received approval from Dr. Gage to carry out a Delphi forecast on autonomous driving at the spring MARS PI meeting, held in San Diego April 6-10, 2003.

A Delphi forecast, named for the Oracle at Delphi who was said to be able to forecast the future, is a poll of experts as to when a certain future event might take place. The concept is that a mean prediction of experts is as good an indicator as is possible to achieve.

While the results are not considered definitive, researchers generally felt that it would take at least ten calendar years and probably closer to twenty calendar years to achieve the capabilities of autonomous driving desired for Future Combat Systems, and that funding of the order of $500M would be needed.

---

[2] Moore's original estimate was a twelve-month doubling period; apparently he revised that to twenty-four months some ten years later. An eighteen-month doubling period has been widely used as "Moore's Law" since the 1970's. Actual doubling periods have ranged between twelve and twenty-four months.

It was further clear that setting general human levels of autonomy is not the correct approach, that specific military needs and modes of driving need to be addressed and solved, and that this involves continued research in sensors, perception, knowledge management and planning.

7. CONCLUSIONS

Useful and practical autonomous driving is in its infancy. As such, there will certainly be unforeseen challenges and periods of both pessimism and over-optimism. Nonetheless, a review of the accomplishments to date, and a survey of current views of experts in the research community is useful, and has provided a basis for a best-estimate at this time of the nature and size of the challenge. While not unanimous, the most prevalent views lead to these overall conclusions:

- Militarily useful autonomous driving capabilities can be developed in approximately ten to twenty calendar years on continued research. The time scale will depend upon the level of funding available.
- The cost will be in the range of three to five hundred million dollars, which is consistent with current funding levels of Army autonomous mobility programs extended over twenty calendar years.
- The biggest single problem is perception. The attack on the problem should start with development of a new generation of sensors designed specifically for autonomous driving.

The conclusions of the different approaches to estimating time and cost for achieving intelligent on-road driving, which support the overall conclusions above, are summarized below.

First: Based on extrapolation from the Demo III experience, it will take another fifteen calendar years of work at the current level of effort to achieve intelligent on-road driving capability.

Second: Based on the Task Decomposition of driving tasks using the DoT manual, it is estimated that approximately $300-400 Million in funding will be needed to achieve intelligent on-road driving skills. Over a twenty calendar year period, this is $15-20 M per year, roughly the level of funding now provided under the ARL and TACOM programs. Increased funding would reduce the time scale.

Third: A new generation of sensors designed specifically for autonomous driving is needed to provide the necessary visual acuity. This is critical because perception emerges as the largest problem in autonomous driving.

Fourth: Engineering attention needs to be paid to providing adequate processors with adequate inter-processor communication to researchers along with software development and debugging tools. Adequate computing power using cluster computers is now or will soon be available, making it possible to address these engineering issues in the near future. Computing power should not be a gating element.

Fifth: Based on the Delphi Forecast of MARS researchers, it will take 15-20 calendar years and of the order of $500M to achieve intelligent driving skills.

Sixth: Several MARS researchers emphasized that setting intelligent driving skills as the goal was not the correct approach, that militarily useful capabilities would be achieved short of that goal

Seventh: Continued research in sensors, perception, knowledge management and planning, at a level at least equal to current funding is essential, even if the scope is reduced to targeting specific military driving modes to be solved in the near term.

ACKNOWLEDGEMENT

This work was supported by the Defense Advanced Research Projects Agency (DARPA) Mobile Autonomous Robot Software (MARS) program (PM. D. Gage).

References

1. Albus, J. and et al, "4D/RCS Version 2.0: A Reference Model Architecture for Unmanned Vehicle Systems," NISTIR 6910, National Institute of Standards and Technology, Gaithersburg, MD, 2002.
2. Albus, J., Evans, J., Schlenoff, C., Barbera, T., Messina, E., and Balakirsky, S., "Achieving Intelligent Performance in Autonomous Driving," National Institute of Standards and Technology, Gaithersburg, MD, Oct. 2003.
3. Camden, R., Bodt, B., Schipani, S., Bornstein, J., Phelps, R., Runyon, T., French, F., and Shoemaker, C., "Autonomous Mobility Technology Assessment Interim Report," *Army Research Laboratory (ARL-MR-565)*, 2003.
4. Churchland, P. S. and Sejnowski, T. J., *The Computational Brain*, MIT Press, Cambridge, 1999.
5. Dickmanns, E. D., "An Expectation-Based Multi-Focal Saccadic (EMS) Vision System for Vehicle Guidance," *Proceedings of the 9th International Symposium on Robotics Research (ISRR'99)*, Salt Lake City, 1999.
6. Grossberg, S., *Boston University, private communication*, 2003.
7. Kurzweil, R., *The Age of Spiritual Machines*, Penguin Books, New York, 1999.
8. McKnight, J. and Adams, B., *Driver Education Task Analysis. Volume 1. Task Descriptions*, Human Resource Research Organization, Department of Transportation, National Highway Safety Bureau 1970.
9. Moravec, H., *Robot, Mere Machine to Transcendent Mind*, Oxford University Press, New York, 1999.
10. Shoemaker, C. and Bornstein, J. A., "Overview of the Demo III UGV Program," *Proceedings of the SPIE Robotic and Semi-Robotic Ground Vehicle Technology Conference*, Vol. 3366, 1998, pp. 202-211.
11. U.S. Department of Transportation, F. H. A., *Manual on Uniform Traffic Control Devices (MUTCD 2000) Millenium Edition* 2000.