Proceedings of AGARD Conference on Future Aerospace Technology, Paris, April 14-16, 1997

Technology Requirement to Implement Improved Situation Awareness: Machine Perception

James S. Albus
Chief, Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, Maryland 20899
United States of America

1. ABSTRACT

Situation awareness is the ability of an unmanned vehicle intelligent control system to model the world. A world model is an intelligent system's current internal estimate of the state of the world, plus its prior knowledge of the history of the world, plus knowledge about the rules of physics and mathematics, plus rules of behavior, task skills, and basic values. World modeling is the ability of the intelligent system to maintain and use a world model to predict and filter sensory experience, to understand the past, and to simulate the future. Perception is the functional transformation of data from sensors into situational awareness.

The technology required for machine perception exists. The ability to design machine perception systems with the sophistication and quality to be useful to field commanders is within reach.

2. PREFACE

Perception and situation awareness are processes that occur naturally in the human mind. In order to discuss these concepts in terms of technology requirements that can be implemented by machines, we make the following operational definitions:

Df: perception ::= the transformation of sensor signals into knowledge about situations and events in the world.

Df: knowledge ::= data structures and information that define the intelligent system's world model

Df: situation ::= a set of relationships that exist between entities in the world

Df: situation awareness ::= correspondence between knowledge in the system's world model and a situation in the world

When a system has situation awareness, it can take appropriate action.

Perception occurs as a result of the interactions between three functional elements: sensory processing, world modeling, and value judgment.

2.1 Sensory processing

Df: sensory processing ::= a process by which sensory data interacts with prior knowledge in order to recognize and track objects and situations, and generate and maintain useful internal representations of the world.

Sensory processing consists of five basic functions:

1) Windowing (or its inverse, masking) selects the regions of space and/or time to be considered. The shape, position, and duration of spatial and temporal windows are determined by the shape, position, and duration of regions in an image (or collection of sensors), or in the world, that are labeled as worthy of attention. Regions may be worthy of attention either because they are goal related, or because they exhibit properties that are unexpected or dangerous, or because they are otherwise noteworthy. The size and shape of each windows is determined by the size, shape, and level of confidence in the position and motion of the entity defining the window.

2) Grouping integrates or organizes spatially and temporally contiguous subentities with similar attributes into entities. The grouping process segments, or partitions, an image into topological regions with entity labels, or names. Any particular grouping is a hypothesis based on gestalt heuristics. A grouping hypothesis is confirmed or rejected based on the usefulness of the hypothesis in predicting sensory input.

3) Computation calculates observed entity attribute values generated by the grouping hypothesis

4) Filtering (e.g. by recursive estimation) computes a best estimate (over a window of space and time) of entity

attribute values based on correlation and differences between predicted and observed entity attribute values. Filtering also computes statistical properties such as confidence in estimated attribute values.

5) Recognition (classification, or detection) establishes a correlation or match between estimated attributes of entities observed in the world and attributes of entity classes stored in the system's knowledge database. Topological, generic, and specific entity classes may be stored in the knowledge database, and observed entities may be recognized as belonging to topological, generic, and/or specific object classes.

2.2 World Modeling

Df: world modeling ::= a process that constructs, maintains, and uses a world model to predict sensory inputs and simulate behavioral plans.

Df: world model ::= an estimate of the state of the world,
plus knowledge of the history of the world,
plus knowledge about the rules of physics and mathematics that govern how the world works,
plus rules of behavior for external agents in the world,
plus task skills that describe how internal agents should act so as to accomplish goals,
plus values that define
what is good and bad,
what is valuable and worthless,
what is important and unimportant,
and what is the level of confidence that can be attached to any part of the world model.

2.3 Value Judgment

Df: value judgment ::= a process that:
a) computes cost, risk, and benefit of actions and plans,
b) estimates the importance and value of objects, events, and situations,
c) assesses the reliability of information,
d) calculates the rewarding or punishing effects of perceived states and events.

Value judgment is a set of cost/benefit functions that determine how much priority an intelligent system should assign to tasks and goals, and what worth should be assigned to objects, agents, relationships, or regions of space. Value judgment calculates how rewarding or punishing specific actions and events are, or can be expected to be. Value judgment also computes statistics on how well observations correlate with expectations, and assigns uncertainty factors to all entities and attributes stored in the world model.

2.4 Knowledge

The world model contains knowledge about the world. Knowledge is represented in data structures in the form of state variables, attributes, entity frames, event frames, relationships, images, maps, rules, equations, and recipes. State variables define estimated conditions in the world. Attributes describe properties. Entities and events can be represented in frames that contain lists of state variables, attributes, and relationships. Images contain information about the position of entities in the world, and maps give an overhead view of the terrain overlaid with labels, icons, and text that provide information necessary for situation assessment and planning of action.

Representation is important both for the machine system and for the human operator. Maps provide information about the battlefield. Where are the friendly and enemy forces? Where is the high ground? Where are the roads, the rivers, the barriers to movement?

Images are windows into the world from the viewpoint of a sensor. It is important to know where the sensor is located, and where observed entities are relative to the sensor viewpoint. Images are also important for display of situations to the human user. The user should be able to choose a viewpoint that is optimal to the user's purpose. This typically requires specification of scale (or resolution) and range (or field of view). For high level observation and planning, maps with a range of many kilometers and resolution of 30 m may be optimal. For tactical observations and planning, maps with a range of about a kilometer with resolution of 3 m may be needed. For individual units to maneuver through the environment, images with both a wide and a narrow field of view are required. Resolution in the wide field of view can be around 0.5 ˚ per pixel, whereas in the narrow field of view, resolution of about 0.02 ˚ per pixel are needed to match the performance of the unaided human eye.

Symbolic information is also important? Spoken words and text are the primary means of communication between humans. As machine systems become more intelligent, speech and written text will grow in importance for human-machine communication as well.

Relationships between maps, images, and words are critical. To be useful, points and regions on a map must be characterized and labeled. Attributes (such as range, texture, color, shape, size, and motion) in visual images need to be measured, and objects need to be recognized (i.e., identified).

3. SITUATION ASSESSMENT

The types of situations that are important for future NATO missions scenarios involve urban, rural, wooded, and mountainous terrains and the sky above them. The type of entities that are important are friendly and enemy forces, buildings, roads, bridges, trees, vehicles, humans, and animals. The entity attributes and relationships between entities that matter are their position, movement, state, and size -- and in the case of intelligent entities -- their capabilities and intentions.

3.1 The Problem of Complexity

The world is infinitely rich with detail. The mission environment contains a practically infinite variety of real objects, such as the ground, rocks, grass, sand, mud, trees, bushes, buildings, posts, ravines, rivers, roads, enemy and friendly positions, vehicles, weapons, and personnel. The environment also contains elements of nature, such as wind, rain, snow, sunlight, and darkness. All of these objects and elements have states, and may cause, or be part of, events and situations. The environment contains a practically infinite regression of detail, and the world itself extends indefinitely far in every direction.

Yet, the computational resources available to any intelligent system are finite. No matter how fast and powerful computers become, the amount of computational resources that can be embedded in any practical system will be limited. Therefore, it is imperative that the intelligent system focus the available computing resources on what is important, and ignore what is irrelevant.

3.2 Focus of Attention

Fortunately, at any point in time and space, most of the detail in the environment is irrelevant to the immediate task of the intelligent system. Therefore, the key to building practical intelligent systems lies in understanding how to focus the available computing resources on what is important, and ignore what is irrelevant. The problem of distinguishing what is important from what is irrelevant must be addressed from two perspectives: top down and bottom up.

Top down, what is important is defined by behavioral goals. The intelligent system is driven by high level goals and priorities to focus attention on objects specified by the task, using resources identified by task knowledge as necessary for successfully accomplishing given goals. Top down goals and high level perceptions generate expectations of what objects and events might be encountered during the evolution of the task and which are important to achieving the goal.

Bottom up, what is important is the unexpected, unexplained, unusual, or dangerous. At each level of the sensory processing hierarchy, processing functions detect errors between what is expected and what is observed.

3.3 System Complexity

Intelligent systems are inherently complex. In order to be intelligent, systems must have a rich and detailed model of the world that is kept up to date by a sensory processing system that collects information from a large number and wide variety of sensors. An intelligent system must be able to reason about the past and plan for the future over a time horizon that extends many hours, or even days and weeks into the future.

Hierarchical layering is a common method for organizing complex systems that has been used in many different types of organizations throughout history for effectiveness and efficiency of command and control. In a hierarchical control system, higher level nodes have broader scope and longer time horizons, with less concern for detail. Lower level nodes have narrower scope and shorter time horizons, with more focus on detail. At no level should a node have to cope with both broad scope and high level of detail. This enables the design of systems of arbitrary complexity, without computational overload in any node and any level.

For example, at the top of the military command and control hierarchy, strategic objectives and priorities influence the selection of goals and the prioritization of tasks throughout the entire hierarchy. However, the details of execution are left to subordinates.

At intermediate levels, tasks with goals and priorities are received from the level above, and sub tasks with sub goals and attention priorities are output to the level below. In the intelligent vehicle environment, intermediate level tasks might be of the form <go to position at map coordinates x,y>, <advance in formation along line z>, <engage enemy units at time t>, etc. The details of execution are left to subordinates.

At each level in the task decomposition hierarchy, higher level more global tasks are decomposed and focused into concurrent strings of more narrow and finer resolution tasks. The effect of each hierarchical level is thus to geometrically refine the detail of the task and limit the view of the world, so as to keep computational loads within limits that can be handled by individual agents, such as intelligent computational nodes, or human beings.

4. TECHNOLOGY READINESS

For the most part, the individual technologies necessary for intelligent semi-autonomous vehicle systems are available. To the extent that the individual technologies have short comings, it is because they are not integrated into a system architecture that enables them to draw information from, other technologies. For example, image processing systems have problems in analyzing and understanding scenes largely because they are not integrated with inertial sensors, or provided with knowledge of lighting conditions, or with range data from radar, stereo cameras, or laser range imaging devices; nor do they use information from topographical maps or navigational instruments. In most cases, use of all of this additional information would remove the ambiguity from image analysis, and vastly improve the performance of image processing systems.

A great deal is known about sensing, filtering, recursive estimation, image analysis, uncertainty, knowledge representation, system identification, game theory, optimization, reasoning, planning, prediction, simulation, and control. The most important technology that is missing, or needs improvement, is a system architecture that provides for the integration of data from sensors with other sources of knowledge into a dynamic world model with a value judgment system, and supports intelligent behavior generation at a multiplicity of levels of range and resolution in space and time. Specifically, what is required is a system theory and reference model architecture that integrates and coordinates:

1) planning for the future with many different time horizons, each with a different level of range and resolution;
2) reasoning about discrete events, objects, and symbols with many different levels of abstraction;
3) representation of information about the world in maps, images, lists, and frames with many different levels of range, resolution, and abstraction;
4) closing of reactive dynamic control loops with continuous feedback at many different sampling rates and bandwidths; and
5) focusing of attention and computational resources on what is important, ignoring what is unimportant, at all levels.

4.1 A System Architecture

4-D/RCS provides a theoretical framework and reference model architecture that meets the requirements stated above. The NIST (National Institute of Standards and Technology) RCS (Real-time Control System)[1,2] with the German (Universitat der Bundeswehr Munchen) VaMoRs 4-D approach to dynamic machine vision [3,4]. It incorporates a Laser Range Imager (LADAR) camera build by Dornier[1], a computer-aided mission planning system designed by Hughes/STX, and terrain visualization technology developed by the U.S. Army Research Laboratory. 4-D/RCS is designed to provide the U.S. DoD Demo III unmanned ground vehicle project with an open system operational architecture that will facilitate the integration of a wide variety of subsystems, such as a foveal/peripheral CCD color camera pair, a LADAR, an inertial guidance package, a GPS satellite navigation system, stereo image processing algorithms developed at the NASA Jet Propulsion Lab, and a HMMWV telerobotic driving system with interfaces for a wide variety of mission packages.

The intended scope of the 4-D/RCS architecture is the design, development, integration, and testing of intelligent supervised autonomy controllers for experimental military vehicle systems. 4-D/RCS provides mechanisms by which intelligent vehicle controllers can analyze the past, perceive the present, and plan for the future. It enables systems to assess cost, risk, and benefit of past events and future plans, and to make intelligent choices between alternative courses of action. A single node in the 4-D/RCS architecture is illustrated in Figure 1.

An example of a 4-D/RCS reference model architecture for a semi-autonomous unmanned ground vehicle is shown in Figure 2. Task Decomposition (TD) modules in nodes at the upper levels in the hierarchy make long range strategic plans consisting of major milestones, while lower level Task Decomposition modules successively decompose the long range plans into short range tactical plans with detailed activity goals. Sensory Processing (SP) modules at lower levels process data over local neighborhoods and short time intervals, while at higher levels, they integrate data over long time intervals and large spatial regions. At low levels, the knowledge database in the World Model (WM) is short term and fine grained, while at higher levels it is broad in scope and generalized. At every level, feedback loops are closed to provide reactive behavior, with high-bandwidth fast-response loops at lower levels, and slower more deliberative reactions at higher levels.

---

[1] Reference to specific brands, equipment, or trade names in this document are made to facilitate understanding and do no imply endorsement by the National Institute of Standards and Technology.

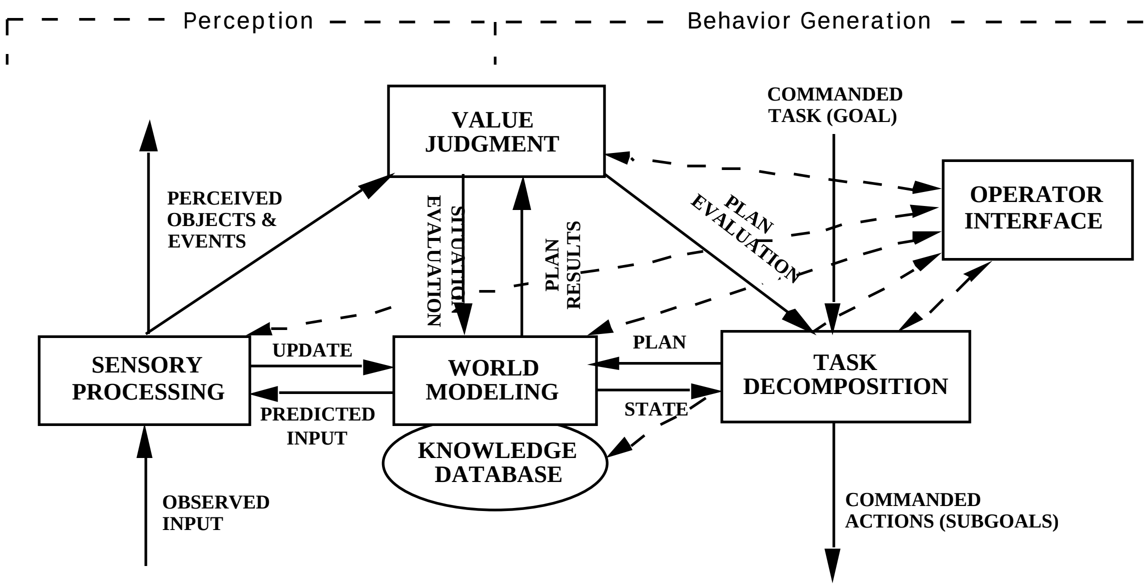

Figure 1. A node in the 4-D/RCS reference model architecture. The functional elements of an intelligent system are task decomposition (planning and control), sensory processing (filtering, detection, recognition, and interpretation), world modeling (store and retrieve knowledge and predict future states), and value judgment (compute cost, benefit, importance, and uncertainty). These are supported by a knowledge database, and a communication system that interconnects the functional modules and the knowledge database. This collection of modules and their interconnections make up a generic node in the 4-D/RCS reference model architecture. Each module in the node may have an operator interface.

At each level, state variables, entities, events, and maps are maintained to the resolution in space and time that is appropriate to that level. At each successively lower level in the hierarchy, as detail is geometrically increased, the range of computation is geometrically decreased. Also as temporal resolution is increased, the span of interest decreases. This produces a ratio that remains relative constant throughout the hierarchy. As a result, at each level, task decomposition functions make plans of roughly the same number of steps. Sensory processing functions compute entities that contain roughly the same number of sub entities. At higher levels, plans, perceived entities, and world modeling simulations are more complex, but there is more time available between replanning intervals for planning processes to search for an acceptable or optimal plan. Thus, hierarchical layering keeps the amount of computing resources needed in each node within manageable limits.

Hierarchical layering in the 4-D/RCS provides mechanisms for focusing the computational resources of the lower levels on particular regions of time and space. Higher level nodes with broad perspective and long planning horizon determine what is important, while the lower levels detect anomalies and attend to details of correcting errors and following plans. In each node at each level, computing resources are focused on issues relevant to the decisions that must be made within the scope of control and time horizon of that node.

The 4-D/RCS hierarchy also supports focusing of attention through masking, windowing, and filtering based on object and feature hypotheses and task goals, as well as by pointing high resolution regions of sensors at objects-of-attention. At each level, masks and windows are used to focus computational resources on objects and events that are important to the mission goal.

At each level along the time line from the present (t = 0), short term memory is much more detailed than long term memory, and plans for the immediate future are much more detailed than plans for the long term.

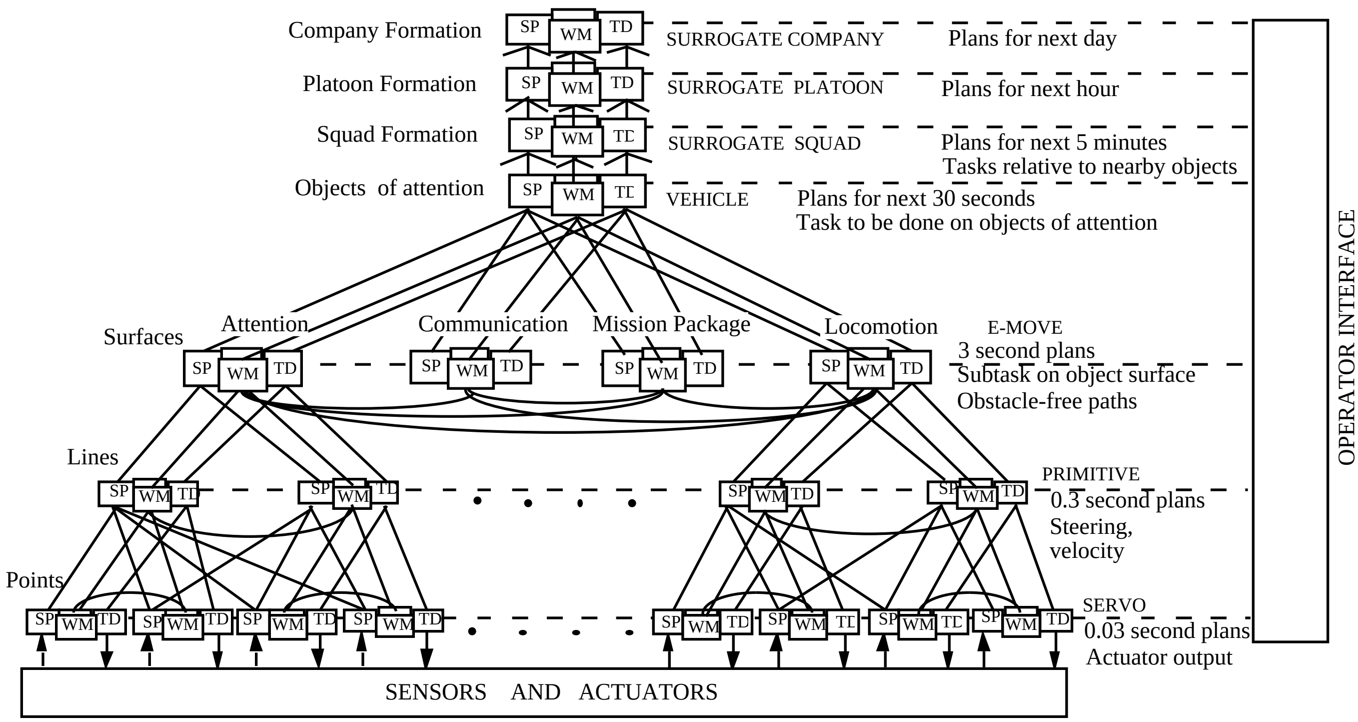

Figure 2. A 4-D/RCS reference model architecture for an individual vehicle. Processing nodes are organized such that the task decomposition (TD) modules form a command tree. Information in the knowledge database (KD) is shared between world modeling (WM) modules in nodes within the same subtree. KD modules are not shown in this figure. On the right, are examples of the functional characteristics of the task decomposition (TD) modules at each level. On the left, are examples of the type of entities recognized by the sensory processing (SP) modules and stored by the WM in the KD knowledge database at each level. Sensory data paths flowing up the hierarchy typically form a graph, not a tree. Value judgment (VJ) modules are hidden behind WM modules. A control loop may be closed at every node. An operator interface may provide input to, and output from, modules in every node.

At each level in the sensory processing hierarchy, lower level entities are grouped into higher level entities. The effect is to encapsulate the detail of entities and events observed in the world in higher level entities and events with broader scope but reduced resolution. This tends to keep the computational load of processing sensory data within manageable limits at all levels of the hierarchy.

4.2 Levels of Abstraction

The 4-D/RCS model addresses the problem of intelligent control at three levels of abstraction: 1) a conceptual framework, 2) a reference model architecture, and 3) an engineering guideline.

1. A Conceptual Framework

At the highest level of abstraction, 4-D/RCS is intended to provide a conceptual framework for addressing the general problem of intelligent vehicle systems operating in man-made and natural environments to accomplish mission goals supervised by human commanders.

The 4-D/RCS conceptual framework spans the entire range of operations that affect intelligent vehicles, from those that take place over time periods of milliseconds and distances of centimeters to those that take place over time periods of months and distances of many kilometers. The 4-D/RCS model is intended to allow for the representation of activities that range from detailed dynamic analysis of a single actuator in a single vehicle subsystem to the combined activity of planning and control for hundreds of vehicles and human beings in full dimensional operations covering an entire theater of battle. The 4-D/RCS architecture is also designed to integrate easily into the information intensive structure of Force XXI Operations and advanced concepts for the strategic Army and Marine Corps of the early 21st century.

In order to span this wide range of activities within a single conceptual framework, 4-D/RCS adopts a multilevel hierarchical architecture, with different range and resolution in time and space at each level.

2. A Reference Model Architecture

At a lower level of abstraction, 4-D/RCS is intended to provide a reference model architecture for supporting the design and development of intelligent vehicle systems, and to provide a theoretical basis for the development of future standards. In order to accomplish this, the 4-D/RCS architecture follows, as closely as possible, the existing command and control structure of the military hierarchy in assigning duties and responsibilities and in requiring knowledge, skills, and abilities.

4-D/RCS defines functional modules at each level such that each module embodies a set of responsibilities and priorities that are typical of operational units in a military organization. This enables the 4-D/RCS architecture to map directly onto the military command and control organization to which the intelligent vehicles are assigned. The result is a system architecture that is understandable and intuitive for human users and integrates easily into battle space visualization and simulation systems.

3. Engineering Guidelines

At a still lower level of abstraction, 4-D/RCS is intended to provide engineering guidelines for building and testing, and eventually using, specific instances of intelligent vehicle systems. In order to build a practical system in the near term, 4-D/RCS engineering guidelines will be developed bottom-up, starting with a single vehicle and its subsystems. The 4-D/RCS engineering guidelines define how intelligent vehicles should be configured in order to work together in groups with other intelligent vehicles, both manned and unmanned, in units of various sizes.

The type of problems to be addressed by the 4-D/RCS engineering guidelines include:
1) navigation and driving both on and off roads,
2) responding to human supervisor commands and requests,
3) accomplishing mission goals and priorities amid the uncertainties of the battlefield,
4) cooperating with friendly agents,
5) acting appropriately with respect to unfriendly agents, and
6) reacting quickly, effectively, and resourcefully to obstacles and unexpected events.

Intelligent vehicle systems will consist of a variety of sensors, actuators, navigation and driving systems, communications systems, mission package interfaces, and weapons systems controlled by an intelligent controller.

The intelligent vehicle must be able to communicate easily and naturally with human operators, and be integrated into the military command and control structure in a manner that is natural and intuitive to military personnel. Specifically, the system should be able to take commands, to ask or offer advice, to report what is important, and to respond to queries.

5. SUMMARY AND CONCLUSIONS

The technology requirements to implement improved situation awareness have been specified. The principal impediment to machine perception is seen to lie in a lack of system architecture that integrates the currently available technologies in signal processing, scene analysis, image understanding, knowledge representation, value judgment, and behavior generation. A theoretical framework and reference model architecture for semiautonomous intelligent unmanned ground vehicles has been discussed. It is suggested that implementation of this architecture will result in a intelligent vehicles systems with a level of performance that would be useful to field commanders in tactical situations.

6. REFERENCES

[1] Albus, J.S., Meystel, A.M. "A Reference Model Architecture for Design and Implementation of Intelligent Control in Large and Complex Systems," International Journal of Intelligent Control and Systems, Vol. 1, No. 1, pgs. 15-30, March 1996.

[2] Albus, J.S., "Outline for a Theory of Intelligence," IEEE Transactions on Systems, Man and Cybernetics, Vol. 21, No. 3, May/June 1991.

[3] E. D. Dickmanns, A General Dynamic Vision Architecture for UGV and UAV, *Journal of Applied Intelligence* 2.251-270 (1992), 1992 Kluwer Academic Publishers, Boston, Manufactured in The Netherlands

[4] E. D. Dickmanns, et. al., The Seeing Passenger Car 'VaMoRs-P', *International Symposium on Intelligent Vehicles'94*, Paris, October 24-26, 1994