Knowledge-Based Analysis of Scene Dynamics
For An Autonomous Land Vehicle

Bir Bhanu, Hatem Nasr, Peter Symosek, Wilhelm Burger, Marc Gluch
and Durga Panda

Honeywell Systems & Research Center
3660 Technology Drive, Minneapolis, MN 55418

ABSTRACT

In the Honeywell Strategic Computing Computer Vision Program, we are
developing knowledge-based technology applied to an Autonomous Land Vehicle.
The focus of our work has been to use artificial intelligence techniques in computer
vision, spatial and temporal reasoning and incorporation of a priori, contextual and
multisensory information for dynamic scene understanding. The topics currently under
investigation are: 1) Landmark and target recognition using multi-source a priori informa-
tion, 2) Robust target motion detection and tracking using qualitative reasoning, 3)
Interpretation of terrain using symbolic grouping.

We have developed a new approach, called PREACTE (Perception-REasoning-
ACTion and Expectation), for knowledge-based landmark recognition for the purpose
of guiding ALV. It is based on the perception-reasoning-action and expectation para-
digm of an intelligent agent and uses dynamic model matching. For motion under-
standing we have developed a new approach called DRIVE (Dynamic Reasoning from
Integrated Visual Evidence) based on a Qualitative Scene Model. The approach
addresses the key problems of the estimation of vehicle motion from visual cues, the
detection and tracking of moving objects and the construction and maintenance of a
global dynamic reference model. Our approach to terrain interpretation employs a
hierarchical region labeling algorithm for multi-spectral scanner data. The technique
is called HSGM (Hierarchical Symbolic Grouping from Multi-spectral data). For this
approach, features used for image segmentation vary form macro-scale features at the
first level of the hierarchy to micro-level features at the lowest level.

We present results from the above algorithms and an integrated system where
these subsystems operate as interactive knowledge sources to a central blackboard.
These results are useful in vision controlled navigation/guidance of an autonomous
land vehicle, search and rescue, munitions deployment and other military applications.
1. INTRODUCTION

The goal of our research in Strategic Computing Computer Vision Program is to demonstrate that knowledge-based approaches as applied to the real-world computer vision problems, such as recognition and tracking of targets and interpretation of terrain, can provide significantly enhanced and robust performance. The results from our research are useful in vision controlled navigation/guidance of Autonomous Land Vehicle (ALV), reconnaissance, surveillance and other practical military applications.

To achieve our goal, we are engaged in developing new techniques for qualitative motion understanding, scene and object modeling, matching, spatial reasoning for recognition, reasoning under uncertainty, symbolic grouping for the interpretation of multi-spectral terrain data, geographic knowledge representation, etc.

Since the knowledge-based techniques, which are being developed here will be a part of a larger system where real-time considerations are very crucial, we are using Real Time Blackboard Architecture (RTBA) software developed at Honeywell. It is written in LogLisp, a logic programming system. LogLisp is composed of Common Lisp and extensions to Common Lisp for logic procedure language. Since the blackboard and LogLisp are written in Common Lisp, RTBA can be used on any hardware/operating system configuration that supports Common Lisp. Currently implemented systems are symbolics 3670 and DEC Vax 11/780 running BSD 4.3 Unix. RTBA is similar to Local Map Builder (LMB) developed at Carnegie Mellon University. However, it is directed towards defense applications where the interest is in building embedded expert systems to achieve real-time performance, to integrate expert system and host system, to establish reliability and to maintain correct inference across changing input data.

The research results described in this paper are partitioned into three topic areas: (a) Landmark and Target Recognition, (b) Target Motion Detection and Tracking, and (c) Interpretation of Terrain. We also discuss the applications of this work to Brilliant weapons.

2. LANDMARK AND TARGET RECOGNITION

An autonomous land vehicle has to traverse long distances to accomplish missions such as surveillance, search and rescue and munitions deployment. This results in the accumulation of significant amount of positional error in the land navigation system. Landmark recognition can be used to reduce this error by recognizing the observed objects in the scene and associating them with the specific landmarks in the geographic map knowledge base. In the current ALV test sites at Martin Marietta, Denver, the landmarks of interest include telephone poles, storage tanks, buildings, houses, gates, etc.

We have developed a new approach, called PRACTE (Perception-REasoning-ACTion and Expectation), for using knowledge-based landmark recognition for the purpose of guiding ALV. It is based on the perception-reasoning-action and expectation paradigm of an intelligent agent. This paradigm is different from the test-
hypothesize-act cycle of Matsuyama and Hwang.\textsuperscript{19} PREACTE uses expectations and allows the prediction of appearance and disappearance of objects in the field-of-view and therefore, it reduces the computational complexity and uncertainty in labeling objects. The approach makes use of extensive map and domain dependent knowledge in a model-based scheme.\textsuperscript{11} Our map representation relies heavily on declarative and explicit knowledge instead of procedural methods on relational databases.\textsuperscript{20} Davis\textsuperscript{13} at Yale University has worked on the problem of acquiring geographic knowledge by a mobile robot. In contrast to his work, in our research explicit knowledge about the map and landmarks is assumed to be given and it is represented in a relational network. It is used to generate an Expected Site Model (ESM) for search delimitation, given the ALV location and its velocity. Landmarks at a particular map site have their 3-D models stored in heterogeneous representations such as generalized cylinder or wire-frame.\textsuperscript{6} The landmark recognition vision system for PREACTE generates a 2-D and partial 3-D scene model from the observed scene. The ESM hypothesis is verified by matching it to the image model. The matching problem is solved by using object grouping and spatial reasoning. Unary, binary and ternary relations are used for spatial reasoning. Both positive and negative evidences are used in spatial reasoning and updating the positional uncertainty of ALV in the map. Evidence accumulation is accomplished by an extension of an efficient heuristic Bayesian formula. A common framework for abstracting image information and modeling objects in heterogeneous representations provides a flexible and modular computational environment for reasoning about image content. The system also provides feedback control to the low-level processes to permit adaptation of the feature detection algorithms parameters to changing illumination and environmental conditions. Currently the landmark recognition scheme assumes that the landmarks are along the sides of the road. In the future we will extend it to a more general and complex situation where the ALV may be traveling through terrain and it has to determine precisely where it is on the map by using landmark recognition.

Figures 1 to 4 show the concept and results of landmark recognition graphically. Table 1 shows the matching results.

Details of the landmark recognition system PREACTE together with results on ALV imagery are given by Nasr, Bhanu and Schaffer.\textsuperscript{24}
Figure 1: Schematic diagram of landmark recognition using expectations, and map and objects knowledge bases.
Figure 2: Original image collected at a predicted site containing a storage tank (right), telephone pole (left) and a building (upper right).

Figure 3: Segmentation results after region splitting and merging.
**Figure 4:** Final results are shown by overlaying rectangles over the classified regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Size</th>
<th>Color</th>
<th>MBR</th>
<th>Texture</th>
<th>Elongation</th>
<th>Shape</th>
<th>Location</th>
<th>Match-Evidence</th>
<th>Landmark Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Small</td>
<td>Black (25.3)</td>
<td>(99, 103, 61, 111)</td>
<td>Smooth</td>
<td>High (50:4)</td>
<td>Long and Linear</td>
<td>(101.2, 82.9)</td>
<td>0.92</td>
<td>T-POLE-110</td>
</tr>
<tr>
<td>79</td>
<td>Small</td>
<td>Black (27.3)</td>
<td>(104, 105, 112, 132)</td>
<td>Irregular</td>
<td>High (20:1)</td>
<td>Long and Linear</td>
<td>(104.6, 121.9)</td>
<td>0.83</td>
<td>T-POLE-110</td>
</tr>
<tr>
<td>95</td>
<td>Medium</td>
<td>White (224.5)</td>
<td>(445, 510, 140, 155)</td>
<td>Smooth</td>
<td>Low (05:15)</td>
<td>Not convex</td>
<td>(482, 148.3)</td>
<td>0.62</td>
<td>G-TANK-110</td>
</tr>
<tr>
<td>70</td>
<td>Medium</td>
<td>Gray (199.7)</td>
<td>(469, 510, 99, 127)</td>
<td>Irregular</td>
<td>Low (41:28)</td>
<td>Linear</td>
<td>(490.2, 115.3)</td>
<td>0.71</td>
<td>BLDG-110</td>
</tr>
</tbody>
</table>

**Table 1:** Match evidences and the corresponding features of classified regions
3. TARGET MOTION DETECTION AND TRACKING

Successful operation of an autonomous land vehicle requires continuous interpretation of complex dynamic scenes utilizing multiple sources of visual and a priori information. Concepts from static scene analysis such as segmentation, feature extraction, spatial grouping and object recognition must be complemented by processes which deal with the temporal aspects of visual perception. Since the ALV moves through a 3-D environment, the resulting camera image is subject to continuous change and objects in the scene cannot be labeled as stationary or moving by simple 2-D techniques. Extensive work has been done in low-level (pixel based) motion analysis, e.g. the computation of optical flow and displacement fields, as well as reconstructive bottom-up approaches to determine 3-D structure and motion. Although some problems remain unresolved and researchers have often made unrealistic assumptions such as orthographic projection, stationary viewer or static environment, the field is well developed and useful techniques are available. However, the integration of motion information into the higher levels of vision is still in its infancy. The integration requires techniques for knowledge representation and reasoning about events in space and time. This capability is crucial to the navigation of an ALV in unstructured environments.

Dynamic scene understanding can be structured into basically three levels of processes. While the low and high levels have their equivalents in many other vision approaches, the important role of intermediate-level processes has not been clearly identified and defined. Low-level processes are purely two-dimensional, bottom-up and image-centered, such as feature extraction, feature matching or optical flow computation. High-level processes are three-dimensional and world-centered, attempting the semantic interpretation of the dynamic proceedings in the environment, using information about the structure and motion of 3-D aggregates. Note that long-term observation and understanding of the behavior of objects form the basis for intelligent actions, such as navigation, route planning and threat handling. In order to bridge the representational gap between low and high levels, we introduce processes operating at an intermediate level, characterized by the transition from image-centered features to world-centered objects. Unlike at the low and high levels, bottom-up and top-down strategies combine at the intermediate level and 3-D reasoning is supported by aggregation of physical and perceptual knowledge and expectations. In psychological terms this stage could be related to unconscious but active visual perception; those tasks that are performed continuously and automatically by the human visual system.

We have developed a new DRIVE (Dynamic Reasoning from Integrated Visual Evidence) approach based on a Qualitative Scene Model to solve the motion understanding problem. The approach addresses the key problems of the estimation of vehicle motion from visual cues, the detection and tracking of moving objects and the construction and maintenance of a global dynamic reference model. Object recognition, world knowledge and accumulation of evidence over time are used to disambiguate the situation and continuously refine the global reference model. The approach departs from previous work by emphasizing a qualitative line of reasoning and modeling, where multiple interpretations of the scene are pursued simultaneously in a hypothesis
and test paradigm. Different sources of visual information such as two-dimensional displacement field, spatial reasoning and semantics are integrated in a rule-based framework to construct and maintain a vehicle centered three-dimensional model of the scene. This approach offers significant advantages over "hard" numerical techniques which have been proposed in the motion understanding literature.\textsuperscript{1,21,27} These advantages include the tracking of objects in the presence of partial or total occlusion and use of this information for route planning and threat handling.

In the DRIVE approach a vehicle-centered model of the scene is constructed and maintained over time, representing the current set of feasible interpretations of the scene. In contrast to most previous approaches, no attempt is made to obtain an accurate geometric description of the scene. Instead a \textit{Qualitative Scene Model} is proposed which holds only a coarse qualitative representation of the three-dimensional environment. As part of this model, the "stationary world" is represented by a set of image locations, forming a rigid 3-D configuration which is believed to be stationary. All the motion-related processes at the intermediate level use this model as a central reference. The motion of the vehicle, for instance, must be related to the stationary parts of the environment, even if large parts of the image are in motion. This kind of reasoning and modeling appears to be sufficient and efficient for the problem at hand.

Figure 5 gives an example of the reasoning process in DRIVE.

Details of the qualitative reasoning concept emphasizing the motion aspects of intermediate-level processes and interfaces to the adjacent levels in the DRIVE system are given by Bhanu and Burger.\textsuperscript{8}
Figure 5: A typical situation: the ALV is approaching an intersection. Two feature points are tracked in the image, point A on the (moving) truck, and point B on the building. The stationary part of the image seems to expand, while the truck stays at a constant image location. The shaded area in the center represents the region of possible FOE locations. The reasoning is like this: Geometric Information says B "expands" faster than A, so B must be closer than A. Spatial information says that object with A occludes the object of B, so A must be closer than B. Therefore, there is a contradiction. This implies that A or B must be moving. Semantic information says that A is on the truck so A must be moving from right to left.
4. INTERPRETATION OF TERRAIN

An autonomous land vehicle must be able to navigate not only on the roads, but also through terrain in order to execute its missions of surveillance, search and rescue and munitions deployment. To do this the vehicle must categorize the terrain regions it encounters as to the trafficability of the regions, the land cover of the regions and region-to-map correspondence.

Predominantly, the segmentation algorithms used for outdoor scene segmentation are region analysis algorithms $^9,23,25$ which attempt to identify regions of the scene on the basis of the homogeneity of the region’s features. Recursive segmentation based on the analysis of distribution of features is one of the most popular and commonly used techniques for image segmentation. Many of these techniques make use of an elaborate peak location and selection procedure which provides threshold values for the purpose of image segmentation. The computation of peak maxima and minima is complicated since minor changes must be distinguished from major ones. One of the shortcomings of these techniques is that small regions in a large image may not show a distinct peak in the histogram, even if they are distinct from their immediate neighborhood. Therefore, in the application of these techniques, normally the image is partitioned artificially into a set of subimages and each subimage is segmented and split further independently. As a result, a remerging measure may be required to merge regions that are arbitrarily broken at the subimage boundaries. Very often this merging step leads to some regions which remain unmerged or overmerged.

Bhanu and Parvin$^9$ have presented a simpler and computationally efficient technique which does not have the above disadvantages and provides good results. It is based on the generalization of a two-class gradient relaxation algorithm for the segmentation of natural scenes.$^5$

However, since the outdoor scenes in the ALV scenario have immense variability, purely region based segmentation algorithms do not perform adequately, because they fail to integrate constraints derived from the three-dimensional attributes of the scene and other auxiliary data into the segmentation process such as the information from a standoff-sensor. Scene variability leads to poorly defined region boundaries and spurious noise regions in the segmented image and this degrades the performance of high-level region labeling schemes which operate on these low-level results. Also the unstructured nature of the outdoor scenes makes their segmentation very difficult with a single set of rules. Currently very simple segmentation methods are used for road-following which are severely limited in robustness and flexibility.$^{18}$

The use of three-dimensional qualities for segmentation is critical for the ALV scenario because the range varies significantly with respect to scan line of the image and feature measures which are valid for a specific range may produce unsatisfactory results at other ranges. Also it is very important that we make use of the spectral properties of the objects in the world and a priori and contextual information to achieve robust interpretation of terrain in a flexible manner.

Our approach for terrain interpretation employs a hierarchical region labeling algorithm for ERIM 12 channel Multi-Spectral Scanner data. The technique called,
HSGM (Hierarchical Symbolic Grouping for Multi-spectral data), is specifically designed for multi-spectral imagery, but is appropriate for other categories of imagery as well. For this approach, features used for segmentation vary from macro-scale features at the first level of the hierarchy to micro-scale features at the lowest level. Examples of labels at the macro label are sky, forest, field, mountain, road, etc. For each succeeding level of the hierarchy, the identified regions from the previous stage are further subdivided, if appropriate, and each region's labeling is made more precise. The process continues until the last stage is reached and no further subdivision of regions from the preceding stage appears to be necessary. Examples of region labels for this level of the hierarchy are gravel road, snowberry shrub, gambel oak tree, rocky ledge, etc.

The HSGM approach is distinct from the classical tree classifier approaches used in the remote sensing literature. The approach operates as follows: For the first stage of the hierarchy, each of the 12 channels of the Multi-Spectral Scanner data is segmented with a textured region detection scheme. The individual segmentations for the 12 channels are combined by using a edge linking relaxation operator to define a "plan" region boundary image. Representative features for each region of the "plan" image are calculated by averaging the feature values for each pixel of the region across the entire region area. These features are classified with a rule-based classification scheme which uses contextual as well as spectral cues for region labeling. Then, at the succeeding stages of the hierarchy, the regions are subdivided by a variety of region and edge-based segmentation algorithms which are optimized for the specific category of region under consideration. These segmentation algorithms also employ spectral, contextual and auxiliary information cues for the specification of region boundaries. Examples of the applied constraints for these stages are a priori terrain elevation data, land cover map information, geological data, time of day and seasonal information. HSGM approach possesses several attractive features, the most important of which is its robustness in the presence of high scene variability. Because the finalized region classifications are derived with global support, both from neighboring regions and from other spectral images, the likelihood that a region will be misclassified because of arbitrary noise is greatly reduced. This approach is also computationally less expensive than many rule-based region labeling schemes because the application of auxiliary constraints decreases the branching factor of the search process significantly.

Figure 6 displays 12 channels of the multispectral data and Figure 7 shows the terrain interpretation results at the macro level.

Details of the HSGM technique with initial results and examples from real ALV imagery are given by Bhanu and Symosek.
Figure 6: Display of 12 channels of multispectral scanner image.
Figure 7: Results of terrain interpretation at the macro level.
5. BRILLIANT WEAPONS APPLICATIONS

In addition to the ALV applications as discussed in the above, our interest is also to transfer this technology to other practical military applications. Brilliant Weapons (BW) are one such application. Conventional technology such as Automatic Target Recognition (ATR) has come a long way but it needs help.\textsuperscript{7} It is clear that for the vision technology to succeed in practical brilliant weapons application, it must be optimally suited for such multisensor combinations as millimeter wave/infrared,\textsuperscript{2,4,17} and CO$_2$ laser.\textsuperscript{15} One of our objectives is to transfer the knowledge-based technology under development here to brilliant weapons relevant multisensor applications to provide significant improvement in performance in diverse scenarios, especially in inclement weather and battlefield scenarios. The vision system performance must demonstrate robustness in hundreds of hours of classified flight test sensor data in presence of target camouflage, concealment and deception (CCD). We are using multisensory and a priori information in a knowledge-based framework within RTBA to achieve the required performance which is beyond what conventional ATR technology can provide.\textsuperscript{3}
6. CONCLUSIONS

In this paper we have presented a summary of our ongoing work for knowledge-based analysis of scene dynamics for an autonomous land vehicle. In the future we plan to integrate our PREACTE module for landmark and target recognition with DRIVE module for qualitative motion understanding and HSGM module for terrain interpretation for an end-to-end simulation demonstrating knowledge-based scene dynamics approach for target motion detection, recognition and tracking.

References


AUVS-87
PRELIMINARY CONFERENCE PROGRAM

Sunday, July 19
12:00 noon - 8:00 p.m.
  Registration
5:30 p.m. - 8:30 p.m.
  Exhibits Open and Reception

Monday, July 20
7:30 a.m. - 7:00 p.m.
  Registration
8:30 a.m.
  WELCOME
8:45 a.m.
  Keynote Address: Donald C. Latham, Assistant Secretary of Defense for Command, Control, Communications and Intelligence
9:15 a.m.
  Overview: Past, Present and Future of Combat Robotics
10:00 a.m.
  SESSION I: LAND APPLICATIONS
  Session Chair: Dr. Raymond E. Bowles, U.S. Army Laboratory Command
  Session Co-Chair: Charles M. Shoemaker, U.S. Army Laboratory Command
  Assessments of Maneuverability with the Teleoperated Vehicle (TOV)
  Edward H. Spain
  NAVAL OCEAN SYSTEMS CENTER
  Robotic Light Armored Vehicle (RLAV) Testbed
  D. Collier, G. Vossler, D. Miley
  GENERAL MOTORS
  Obstacle Avoidance for the Autonomous Land Vehicle Perception
  Terry Dunley, Stephen Hennessy, David Margenthaler
  MARTIN MARITETTA DENVER AEROSPACE
  Knowledge Based Analysis of Scene Dynamics for an Autonomous Land Vehicle
  Bir Bhanu, Hatem Nasr, Peter Symosek, Wilhelm Burger, Marc Gluch, Durge Panda
  HONEYWELL SYSTEMS & RESEARCH CENTER
  High Strength to Weight. Precision Manipulation
  Thomas Bartholet
  ODETEC INC.

12:00 noon
  Reception in the Exhibit Hall
12:45 p.m.
  Luncheon with Speaker
2:00 p.m.
  SESSION II: SEA APPLICATIONS
  Session Chair: Charles E. Stuart, OSD/Defense Advanced Research Projects Agency
  Session Co-Chair: Mack D. O'Brien, Charles Stark Draper Laboratories
  A Control System Architecture for Multiple, Autonomous Underwater Vehicles (MAUVE)
  Dr. James S. Albus
  NATIONAL BUREAU OF STANDARDS
  Development of a Night Surveillance Autonomous Underwater Vehicle
  H. A. Perry
  PERRY OFFSHORE
  The Artificial Gill: Oxygen Extraction for Long Range Autonomous Underwater Vehicles
  Steven J. Carnevale
  AQUANAUTICS
  A Multiple Knowledge Base Approach to Autonomous Underwater Vehicle Control
  Dr. David Wendstrand, R. L. Stewart, W. J. Grabowski, D. J. Busher
  APPLIED PHYSICS LAB/JOHNS HOPKINS UNIVERSITY
  Advanced Propulsion and Control for Autonomous Underwater Vehicles Using the Tandem Propeller System
  G. M. Stenovec, J. E. Snow
  AMETEK, STRAZA DIVISION
4:00 p.m.
  SESSION III: TECHNOLOGY & SUB-SYSTEMS
  Session Chair: Dr. James S. Albus, Robotic Div., National Bureau of Standards
  Session Co-Chair: Dr. Alan Meyrowitz, Office of Naval Research
  Value-Driven Logic for the Coordination of Cooperating Autonomous Underwater Vehicles
  G. E. Pugh, J. C. Krupp
  DSA
  AI Enhanced Path Planning for Robotic Vehicles
  W. J. Pollard, G. G. Gildner
  KMS FUSION, Inc.

(Continued on next column)
Tuesday, July 21

7:30 a.m. - 6:00 p.m.
Registration

8:30 a.m.
SESSION IV: AIR & SPACE APPLICATIONS
Session Chair: Ronald D. Murphy, Defense Advanced Research Projects Agency
Session Co-Chair: Capt. Penn Mullowney, U.S. Naval Air Systems Command
Command, Control and Communications (C3) for Unmanned Vehicles in Strategic Air Defense
Lt. Col. James D. Taylor, USAF ELECTRONIC SYSTEMS DIVISION (ESDR-1)
Peacetime Operation of the Unmanned Air Reconnaissance System (UARS)
Robert B. Piper ANSER
Optimizing the Medium-Range Remotely Piloted Vehicle Reconnaissance System
Charles M. Undermann NORTHROP VENTURA DIVISION
Operational Testing of the Aquila System
R. F. Nettles LOCKHEED MISSILES & SPACE CO.
Aeronautical Systems Division’s Avionics Integrity Program
Gary Ludwig WRIGHT-PATTERSON AFB
RPV and UAV Strike Weapons: An Approach to Augment Manned Aircraft Effectiveness
Milan M. Skrtic LTV MISSILES & ELECTRONICS GROUP
Engine Management Systems for Airborne UVS
C. W. Humphries GEC AVIONICS LTD.

12:00 noon
Reception in Exhibit Hall

12:45 p.m.
Luncheon with Speaker

2:00 p.m.
SESSION V: AIR & SPACE APPLICATIONS (continued)

The Phoenix Trials Programme
R. W. Dennis GEC AVIONICS LTD
The Aerobat, A Robotic VTOL Aerial Vehicle
Paul S. Moller MOLLER INTERNATIONAL
SKYEYE R4E-50—A Fourth Generation System
Gerald R. Seemann DEVELOPMENTAL SCIENCES
Optimizing the Use of Telerobotic Vehicles for On-Orbit Maintenance of SDI Systems
Pat R. Odem ADVANCED TECHNOLOGY INC.
A Coordinate, Dual Arm Control System for Dexterous Manipulation of a Common Object
Samuel J. Harris ADVANCED INTELLIGENT MACHINES DIVISION
A Survey of High-Altitude, Long Endurance (HALE) Technologies
David Hall SAIC

5:00 p.m.
Panel Discussions

6:00 p.m.
Exhibition and Reception

8:00 p.m.
Annual Banquet

— ADJOURNMENT —

Preliminary program is subject to revision.