C-1
[illegible]

Knowledge-Based Robust Target Recognition & Tracking

Bir Bhanu and Durga Panda

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

ABSTRACT

In the Honeywell Strategic Computing Computer Vision Program, we are working on demonstrating knowledge-based robust target recognition and tracking technology. 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 information, 2) Robust target motion detection and tracking using qualitative reasoning, 3) Interpretation of terrain using symbolic grouping. An integrated system concept for these topics is presented, along with results on real imagery. Practical applications of our work involve vision controlled navigation/guidance of the autonomous land vehicle, reconnaissance, surveillance, photo-interpretation, and other military applications such as search and rescue and targeting missions.

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.[16] It is written in LogLisp, a logic programming system.[12] 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 report 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 PREACTE (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.[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.[11] Our map representation relies heavily on declarative and explicit knowledge instead of procedural methods on relational databases.[20] Davis[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.[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.

Details of the landmark recognition system PREACTE together with results on ALV imagery are given in.[24]

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,[26] 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.[22] 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 [14] 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.[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 *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.

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 presented in.[8]

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[10] 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.

Details of the HSGM technique with initial results and examples from real ALV imagery are given in.[10]

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. Precision Guided Weapons (PGWs) are one such application. Conventional technology such as Automatic Target Recognition (ATR) has come a long way but it needs help.[7] It is clear that for the vision technology to succeed in practical brilliant weapons application, it must be optimaly suited for such multisensor combinations as millimeter wave/infrared,[2,4,17] and CO_2 laser.[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.[3]

6. CONCLUSIONS

In this report we have presented a summary of our work completed during the last seven months. 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

1. *Proc. IEEE , Workshop on Motion: Representation and Analysis, Kiwah Island Resort, Charleston, South Carolina.* May 7-9, 1986.
2. *Dual-Mode Seeker Study, Final Report, Airforce/Army Contract No. F08635-85-C-0156, Honeywell Systems & Research Center.* May 1986.
3. *Smart Weapons Program, DARPA/U.S. Army AMCCOM Armament Research & Development & Engineering Center, Contract no. DAAA21-86-C-0305, Honeywell Systems & Research Center.* 1986.
4. W. Au, S. Mader, and R. Whillock, ''Scene Analysis,'' Third Triannual Technical Report, Night Vision & Electro-Optics Lab, Contract No. DAAL01-85-C-0429, Honeywell Systems & Research Center (July 1986).
5. B. Bhanu and O.D. Faugeras, ''Segmentation of Images Having Unimodal Distributions,'' *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-4**(4) pp. 408-419 (July 1982).
6. B. Bhanu and T. Henderson, ''CAGD Based 3-D Vision,'' Proc. IEEE International Conference on Robotics & Automation (March 1985).
7. B. Bhanu, ''Automatic Target Recognition: State of the Art Survey,'' *IEEE Trans. on Aerospace & Electronic Systems* **AES-22**(4) pp. 364-379 (July 1986).
8. B. Bhanu and W. Burger, ''DRIVE: Dynamic Reasoning from Integrated Visual Evidence,'' *Proc. DARPA Image Understanding Workshop,* pp. 581-588 (Feb. 1987).
9. B. Bhanu and B. Parvin, ''Segmentation of Natural Scenes,'' *Pattern Recognition, in Press,* (1987).
10. B. Bhanu and P. Symosek, ''Interpretation of Terrain Using Hierarchical Symbolic Grouping for Multi-Spectral Images,'' *Proc. DARPA Image Understanding Workshop,* pp. 466-474 (Feb. 1987).
11. T.O. Binford, ''Survey of Model-Based Image Analysis,'' *The International Journal of Robotics Research* **1** pp. 18-64 (Spring 1982).
12. J. Carciofini, T. Colburn, and R. Lukat, ''LogLisp Programming System User's Guide,'' Internal Report, Honeywell Systems & Research Center (July 1986).

13. E. Davis, *Representing and Acquiring Geographic Knowledge,* Morgan Kaufmann Publishers, Inc. (1986).

14. B. Kuipers, ''Qualitative Simulation,'' *Artificial Intelligence* **29** pp. 298-338 (1986).

15. K. Landman, ''Automatic Laser Target Classification,'' Phase I Interim Report, AFWAL/Avionics Lab Contract No. F33615-84-C-1519, Honeywell Systems & Research Center (April 1985).

16. A. Larson, ''Real-Time Blackboard Architecture System,'' Internal Report, Honeywell Systems & Research Center (August 1986).

17. B. Lee, W. Higgins, and J. Gillberg, ''Multisensor Algorithm Development: First - Fourth Quarterly Reports,'' Night Vision and Electro-Optics Lab, Contract no. DAAL 01-85-C-0443, Honeywell Systems & Research Center (1986).

18. J. Lowrie, ''The Autonomous Land Vehicle Second Quarterly Report,'' Martin Marietta, Denver, Colorado (September 1986).

19. T. Matsuyama and V. Hwang, ''SIGMA: A Framework for Image Understanding - Integration of Bottom-up and Top-Down Analysis,'' Proc. Int. Joint Conference on Artificial Intelligence (August 1985).

20. D.M. McKeown, Jr., W.A. Harvey, Jr., and J. McDermott, ''Rule-Based Interpretation of Aerial Imagery,'' *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-7** pp. 570-585 (September 1985).

21. G. Medioni and Y Yasumoto, ''Robust Estimation of 3-D Motion Parameters from a Sequence of Image Frames Using Regularization,'' Proc. DARPA Image Understanding Workshop (Dec. 1985).

22. H.-H. Nagel, ''Image Sequences - Ten (octal) Years - From Phenomenology Towards a Theoretical Foundation,'' Proc. International Conference on Pattern Recognition (October 1986).

23. P. Nagin, A. Hanson, and E. Riseman, ''Studies in Global and Local Histogram Guided Relaxation Algorithms,'' *IEEE Trans. on Pattern Analysis and Machine Intelligence,* pp. 263-277 (May 1982).

24. H. Nasr, B. Bhanu, and S. Schaffer, ''Guiding the Autonomous Land Vehicle Using Knowledge-Based Landmark Recognition,'' *Proc. DARPA Image Understanding Workshop,* pp. 432-439 (Feb. 1987).

25. R. Ohlander, K. Price, and D.R. Reddy, ''Picture Segmentation Using a Recursive Region Splitting Method,'' *Computer Graphics and Image Processing* **8** pp. 313-333 (1978).

26. K. Prazdny, ''On the Information in Optical Flows,'' *Computer Vision, Graphics and Image Processing* **22** pp. 239-259 (1983).

27. E.M. Riseman and A.R Hanson, ''Summary of Progress in Image Understanding at the University of Massachusetts,'' Proc. DARPA Image Understanding Workshop (Dec. 1985).