# Qualitative Reasoning and Modeling for Robust Target Tracking and Recognition from a Mobile Platform

**Bir Bhanu and Durga Panda**

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

## ABSTRACT

In the DARPA Strategic Computing Computer Vision Program, we focus on demonstrating robust techniques for target tracking and recognition from a moving robotic combat vehicle. Our work is specifically directed towards significant enhancements in the performance of existing target tracking techniques under high clutter and low contrast situations in a ground-to-ground scenario when the robotic combat vehicle is in motion and multiple targets may appear at varying ranges. The topics currently under investigation are: decomposition of complex vehicle motion into its constituent parts; qualitative 3-D scene modeling; target motion detection and tracking; landmark recognition; 3-D target model acquisition and refinement; and use of recognition and map information in an integrated motion detection and tracking system. The results from our research are useful in vision controlled navigation/guidance of a robotic combat vehicle for practical military missions such as targeting, reconnaissance and surveillance. This report summarizes the progress made during the period from March 1987 to January 1988. We also discuss the technology transfer aspects of our work.

## 1. INTRODUCTION

The goal of our research in the Strategic Computing Computer Vision Program is to demonstrate robust techniques for target tracking and recognition from an autonomously-moving robotic combat vehicle. In our experience in implementing vision controlled navigation/guidance for reconnaissance, surveillance, search and rescue, and targeting missions, we find that for spatio-temporal vision problems, purely quantitative approaches are unsuitable and insufficient because of the inexact nature of vision. As such, the technical basis of our work is qualitative reasoning and modeling for dynamic scene understanding.

To achieve our goal, we are engaged in developing efficient and reliable techniques for qualitative motion understanding, dynamic model matching, automatic 3-D model acquisition, spatial reasoning, geographic knowledge representation and its use in recognition and tracking. This work is specifically directed towards significant enhancements in the performance of existing target tracking techniques under high clutter and low contrast situations in a ground-to-ground scenario when the robotic combat vehicle is in motion and multiple targets may appear at varying ranges.

### 1.1 Qualitative Reasoning and Modeling

The choice of a suitable scheme for representing the perceived state of the scene, observed by a moving robotic combat vehicle, is a crucial question. It has an immediate impact upon the efficiency, versatility, and robustness of the reasoning processes that are attached to this representation. It is questionable whether an accurate numerical description of the 3-D environment is really necessary to facilitate efficient reasoning of spatio-temporal processes. The use of *qualitative* descriptions of physical properties has raised considerable interest in the area of Artificial Intelligence.[16,21] Its potential significance to the field of computer vision has been addressed only recently.[4,38,41] The main argument is that many of the error-prone, computationally expensive techniques which are commonly used can be replaced by emphasizing the qualitative effects and utilizing less precise representations without sacrificing the usefulness of the results.

Most previous work in motion understanding attempted to obtain the 3-D scene structure from motion in the form of a quantitative, numerical description of the spatial layout of the environment relative to the camera. The problems related to this approach are well-known and applications using real imagery have been rare. The systems of nonlinear equations that must be solved for this purpose are numerically unstable; small errors in the estimate of image displacement lead to unproportionally large errors in the estimated 3-D geometry.

Since numerical schemes are designed to converge towards a single solution which is optimal in some sense, there seems to be no practical mechanism that would reflect the *uncertainty* of the input data on the final result. Furthermore, the necessary assumption of rigidness cannot be guaranteed. When features are assumed to form a rigid configuration in space but are actually moving relative to each other, this may still result in a rigid interpretation. The problem with this approach is how the numerical model responds when moving features and stationary features are inadvertently grouped. In the best case, the deviation from a rigid configuration would be indicated by a high error value for the feature which is actually in motion. If this is not the case, the model may converge towards a completely different solution.

Following the qualitative reasoning and modeling approach, the central building block of our DRIVE system[4] for target motion detection and tracking is a *Qualitative Scene Model* (QSM), which can be considered as the "mind" of the motion understanding system. This model is a 3-D, camera-centered representation of the scene which describes the observed environment by using a set of simple qualitative relationships. The set of entities in the QSM is conceptually split into two parts, the *stationary world* and a set of independently moving objects. Construction of the QSM over time is accomplished by a reasoning process which draws conclusions from significant configurations and changes in the image. As the vehicle travels through the environment, the model is continuously updated and revised by adding or deleting hypotheses.

Additionally, the state of the QSM is a not a single interpretation but a *set* of interpretations which are all pursued simultaneously. This provides a very flexible mechanism for handling the inherent ambiguities encountered in image understanding. Each interpretation is a collection of hypotheses, called *partial* interpretations, which cover overlapping subsets of the entities in the model. The structure and dynamic behavior of the Qualitative Scene Model are described in more detail in the paper by Bhanu and Burger.[6]

Qualitative reasoning and modeling is also emphasized in our work on landmark and target recognition from a mobile platform.[7,29] Using qualitative information, we do not have to rely on obtaining precise geometric representations of a target. To handle continuous changes in the target's appearance caused by range and perspective, we use a dynamic model matching technique,[29] which combines a camera model, 3-D target models, and predicted range and perspective to generate multiple 2-D image models for matching. TRIPLE's[7] machine learning approach allows for automated 3-D model acquisition and refinement. It uses qualitative and quantitative shape descriptions.

The research results described in this report are partitioned into the following topic areas: (a) target motion detection and tracking and (b) landmark and target recognition. We also discuss the technology transfer aspects of our application in the discussion.

## 2. TARGET MOTION DETECTION AND TRACKING

Motion becomes a natural component of visual information processing as soon as moving objects are encountered in some form; while following a convoy, approaching other vehicles, or detecting threats. The presence of moving objects and their behavior must be known to provide appropriate counteraction. In addition, image motion provides important clues about the spatial layout of the environment and about the actual movements of the vehicle. As part of the vehicle control loop, visual motion feedback is essential for path stabilization, steering, and braking. Results from psychophysics[24,34] show that humans rely heavily on visual motion for motor control.

While the vehicle is moving itself, the entire camera image is changing continuously, even if the observed part of the environment is completely stationary. The interpretation of complex dynamic scenes is therefore the continuous task for the vision system of an autonomous robotic combat vehicle. Previous work in motion analysis has mainly concentrated on numerical approaches for the reconstruction of motion and scene structure from image sequences. Recently Nagel[28] has given a comprehensive review. While a completely stationary environment has been assumed in most previous work on the reconstruction of *camera motion*, the possible presence of moving objects must be accounted for in this scenario. Similarly, one cannot rely on a fixed camera setup to *detect* those moving objects. Clearly, some kind of common reference is required against which the movement of the vehicle as well as the movement of objects in the scene can be related.

Extensive work has been done in determining the relative motion and rigid 3-D structure from a set of image points and their displacements, basically following two approaches.

In the first approach, 3-D structure and motion are computed in one integral step by solving a system of linear or nonlinear equations[27,39] from a minimum number of points on a rigid object. The method is reportedly sensitive to noise.[15,42] Recent work[10,11,17,37,40] has addressed the problem of recovering and refining 3-D structure from motion over extended periods of time, demonstrating that fairly robust results can be obtained. However, these approaches require large amounts of computation, convergence is slow and require many distinct views of the object (the environment), which are generally not available to a moving vehicle. In addition, it seems that the noise problem cannot be overcome by simply increasing the time of observation.

The second approach[10,18,24,25,33,35] makes use of the unique expansion pattern which is experienced by a moving observer. Arbitrary observer motion can be decomposed into translational and rotational components from the 2-D image without computing the structure of the scene. In the case of pure camera translation in a stationary environment, every point in the image seems to expand from one particular image location termed the *Focus of Expansion* (FOE). The closer a point is in 3-D, the more rapidly its image expands away from the FOE. Thus, for a stationary scene, the 3-D structure can be obtained directly from the expansion pattern. Certain forms of 3-D motion become apparent by local deviations from the expanding displacement field and therefore can be detected immediately. The views of the scene need not be very distinct in this approach and there seems to be evidence from psychophysics that the human visual system employs similar techniques.[32,34]

The primary goal for *Dynamic Scene Understanding* in this particular context is to construct and maintain consistent and plausible interpretations of the time-varying images obtained from the camera on the moving vehicle by determining:

- How is the vehicle itself moving ?
- What is the approximate 3-D structure of the scene ?
- What is moving in the scene and how does it move ?

Obviously, these three goals are in very close interaction: any form of motion, whether vehicle motion or actual target motion, must be measured against some stationary reference in the environment.

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 [16,21] 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.[26,36] 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.

Details of the qualitative reasoning concept emphasizing the motion aspects of the DRIVE system are presented in papers by Bhanu and Burger.[4,5,6,12,13,14]

## 2.1 Estimation of Vehicle Motion

The problem of determining the motion parameters of a moving camera relative to its environment from a sequence of images is crucial for the application of computer vision to practical military missions. In addition to translating in an unknown direction, the vehicle also rotates about an arbitrary axis (roll, pitch, and yaw), though not drastically. However, due to the design of the vehicle, the direction of travel is quite restricted, e.g., vehicle orientation does not change rapidly and target stays within the field of view. The observed displacement field is the addition of the vector fields caused by vehicle translation and rotation, such that the vehicle motion cannot be obtained from the displacement field directly. However, the displacement field caused by the vehicle's motion can be decomposed into its rotational and translational components exclusively in the 2-D image, without any 3-D information.

In our work the computation of camera motion is performed from sets of displacement vectors obtained from consecutive pairs of images.[19] First, the decomposition of 3-D camera motion into rotation and translation components and their individual effects upon the image are analyzed in detail. Two basic approaches for computing camera motion are evaluated. In the *FOE-from-Rotation* approach, the direction of camera translation (marked by the *Focus of Expansion* - FOE) is derived for a given estimate of the camera's rotation. Alternatively, in the *Rotation-from-FOE* approach, the rotational components are determined from a given estimate of the location of the FOE. It is shown that the latter approach is highly robust against disturbances of the displacement field, since it works without extending the displacement vectors. Instead of searching for one particular FOE, the final algorithm computes a connected *region* of possible FOE locations, which accounts for noise and distortions in the image. Finally, the absolute velocity of the vehicle towards the FOE is estimated from the expansion pattern by knowing the height of the camera above the (approximately flat) ground. We show the results on real image sequences in the paper by Bhanu and Burger.[6]

## 2.2 Estimation of Stationary 3-D Structure

The environment is modeled as a 3-D, time-varying configuration of rigid objects whose structures, relative positions, and motions are estimated from visual information. The stationary part of the world is represented by a subset of the rigid objects, which form a rigid configuration in 3-D space. This definition, however, is not sufficient to identify the stationary world a priori, because more than one rigid subset of world objects may be observed. To operate in a real environment, some description about the 3-D layout of the scene must be available. 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 of vision use this model as a central reference. The motion of the vehicle, for instance, is 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 effective for this problem.

## 2.3 Detection of Moving Targets

For intelligent action in the presence of potential threats and targets, navigation in a traffic environment, etc., information on actual motion in the scene is indispensable. Moving objects must be detected and isolated from the stationary environment, their current motions must be estimated to track them, and expectations about their future behavior must be created. Since the camera itself is moving, the stationary part of the scene cannot be assumed to be registered in subsequent images, as in the case of a stationary sensor. Simple frame-differencing techniques to detect and isolate moving objects do not work in this case because image changes, due to sensor motion, would generate too many false alarms. More sophisticated image-based techniques, which apply 2-D transformations (warping) to the image to compensate for background motion, work well only when objects are moving in front of a relatively flat background, such as in some air-to-ground applications. To detect actual object motion in the complex scenario of a robotic combat vehicle, the 3-D structure of the observed environment together, with the vehicle's motion, must be taken into account.

In our DRIVE approach, 3-D motion is detected in two ways:

- *First*, some forms of motion are concluded directly from the 2-D displacement vectors without any knowledge about the underlying 3-D structure.
- *Second*, motion is detected by discovering inconsistencies between the current state of the internal 3-D scene model and the changes actually observed in the image.

### 2.4 Interpretation of Terrain

An autonomous robotic combat 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. Our approach for terrain interpretation employs a robust texture-based algorithm and a hierarchical region labeling scheme 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-level 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.

Details of the HSGM technique with results and examples from real imagery are given in papers by Bhanu and Symosek.[8,9]

### 2.5 Map Integrated Motion Detection and Tracking

A priori information for scene content, in the form of digital map data, is an invaluable resource for tracking algorithms. Contextual information, derivable from digital maps, is especially critical to high-level reasoning paradigms which carry out the mission tasks such as estimation of vehicle location, condition monitoring, target acquisition, target classification, target tracking, target engagement, and sensing of vehicle orientation.

Using techniques developed at Honeywell, digital map databases can be transformed into digital visibility maps, from which intervisibility predictions can be computed.[20] We are in the process of implementing a map reasoning system that will be able to identify the world position of moving and tracked targets. The system will incorporate map/terrain and cartographic data bases and will be integrated with the DRIVE system. DRIVE will select moving targets in the image, give their 2-D image location, velocity vector, and approximate range. Given this information from DRIVE, the vehicle geodetic location, and a camera model, the system will establish an image-to-map registration and search in the map data base for possible roads/terrain on which the targets may be moving. The system will also provide information about neighboring landmarks to the target and possible occlusion information.

## 3. Landmark and Target Recognition

A few of the desirable features to be incorporated in an advanced target recognition system[3] are: (a) The models used by the system to represent targets, contexts, and other system knowledge should be dynamic data structures; (b) Most data should be of a symbolic, qualitative nature, thus avoiding the problems encountered in dealing with quantitative information. Using qualitative information, we do not have to rely on obtaining precise geometric representations of target; (c) The system has to be able to handle problems such as imprecise segmentation, occlusion, noise, etc. and ; (d) The system should exhibit improved performance over time. This improvement may come in the form of faster recognition times, improved recognition accuracy, and higher confidence in system results.

Our work on landmark and target recognition is directed towards emphasizing the above features in a dynamic scenario. Target recognition from a mobile platform requires the ability to recognize targets from varying range and perspectives under changing environmental conditions.

### 3.1 Landmark Recognition

Landmark recognition is used to update the land navigation system which accumulates a significant amount of error after the vehicle traverses long distances, which is typically the case in surveillance and targeting missions. The vision system of the autonomous vehicle is required to recognize the landmarks as the vehicle approaches from the road or terrain.

We have developed an expectation-driven, knowledge-based landmark recognition system, called PREACTE[31] that uses map, and landmark knowledge, spatial reasoning and a novel dynamic model matching technique.[29] PREACTE's mission is to predict and recognize landmarks as the vehicle approaches them from different perspective angles and at varying ranges. Once the landmarks have been recognized, they are associated with specific map coordinates, which are then compared to the land navigation system's readings, and corrections are made. Landmarks of interest include buildings, gates, poles, and other man made objects.

Dynamic model matching generates and matches target landmark and map site descriptions dynamically. These descriptions are a collection of spatial, feature, geometric and semantic models. From an approximate range and view angle, and using a priori map information, 3-D landmark models, and the camera model, PREACTE generates predictions about individual landmark locations in the 2-D image. The parameters of all models are a function of range and view angle. As the vehicle approaches the expected landmark, the image content changes, which in turn requires updating the search and match strategies. Landmark recognition, in this framework, has been divided into three stages: detection, recognition, and verification. At far ranges, only "detection" of distinguishing landmark features is possible, whereas at close ranges, recognition and verification are more feasible, since more details of the object are seen. The salient features of the technique are: (a) landmark models are dynamic; (b) different landmarks require different representations and modeling techniques; (c) a single landmark requires hybrid models; and (d) at different ranges, different matching and recognition plans are performed.

Details of the landmark recognition system, PREACTE, together with results on Autonomous Land Vehicle imagery, are given in the papers by Nasr and Bhanu.[29,30,31]

### 3.2 Target Model Acquisition and Refinement

Target recognition systems currently lack the ability to adapt to changing environmental conditions and to modify their behavior based on the context of the situation in which they are operating. In order to be effective in dynamic outdoor scenarios, a robust recognition system should be able to automatically acquire necessary contextual information from the environment. Most target recognition systems lack this capability. Their performance begins to quickly degrade when subjected to the problems of variable lighting conditions, image noise and object occlusion.

Due to recent advances in machine learning technology, some of the problems encountered in the target recognition domain seem to be resolvable. Learning allows an intelligent recognition system to use situation context in order to understand images. This context, as defined in a machine learning scenario, consists of a collected body of background knowledge as well as environmental observations which may impact the processing of the scene.

Machine learning will facilitate two main breakthroughs in the target recognition domain: automatic knowledge base acquisition and continuous knowledge base refinement. The use of learning in the knowledge base construction will save the user from spending the enormous amount of time necessary to derive target models and databases. Knowledge base refinement can then be incorporated to make any necessary changes to improve the performance of the recognition system. These two modifications alone will serve to significantly advance the present abilities of most target recognition applications.

To validate the concept of a target recognition system with integrated machine learning capabilities, the paper by Bhanu and Ming[7] presents an overview of a new approach to target recognition. The system currently under implementation is called TRIPLE: *Target Recognition Incorporating Positive Learning Expertise*. The system uses a multistrategy technique; two powerful learning methodologies are combined with a knowledge-based matching technique to provide robust target recognition. Using dynamic models, TRIPLE can recognize targets present in the database. If necessary, the models can be refined if errors are found in the representation. Additionally, TRIPLE can automatically store a new target model and recall it when that target is encountered again. Finally, since TRIPLE uses qualitative data structures to represent targets, it can overcome problems such as image noise and occlusion.

The two main learning components of the TRIPLE system are Explanation-Based Learning (EBL) and Structured Conceptual Clustering (SCC). Explanation-based learning provides the ability to build a generalized description of a target class using only one example of that class. Structured conceptual clustering allows the recognition system to classify a target based on simple, conceptual descriptions rather than using a predetermined, numerical measure of similarity. While neither method, used separately, would provide substantial benefits to a target recognition system, they can be combined to offer a consolidated technique which employs the best features of each method and is very robust.

## 4. TECHNOLOGY TRANSFER

In this report, we have presented a summary of our work completed during the last twelve months. Our work is directed towards providing key functionalities of target motion detection and tracking, which are needed in autonomous robotic combat vehicle missions of targeting, reconnaissance, and surveillance. In our experience with accomplishing these practical military missions, we find that for spatio-temporal vision problems, purely quantitative approaches are unsuitable and insufficient because of the inexact nature of vision. As such, the technical basis of our work is qualitative reasoning and modeling for dynamic scene understanding. Our PREACTE module for man made landmark recognition and DRIVE module for motion detection and tracking are ready to be transferred and integrated with Carnegie Mellon University's software. We are also working on integrating the PREACTE and DRIVE modules for an end-to-end simulation demonstrating Honeywell's knowledge-based scene dynamics approach for technology transfer to a robotic combat vehicle.

In addition to the robotic combat vehicle applications as discussed above, our interest is also to transfer this technology to other practical military applications. Precision Guided Weapons (PGWs), such as Honeywell's next generation SADARM, are one such application. Conventional technology, such as Automatic Target Recognition (ATR), has come a long way but it needs help.[3] It is clear that for vision technology to succeed in practical smart weapons applications, it must be optimally suited for such multisensor combinations as millimeter wave/infrared,[1,2,23] and $CO_2$ laser (range, reflectance, Doppler and vibration measurements).[22] We are transferring the knowledge-based technology under development here to smart weapons relevant multisensor applications in order to provide significant improvement in performance in diverse scenarios. We are using multisensory and a priori information (map) in a knowledge-based framework to achieve the required performance which is beyond what conventional ATR technology can provide.

## REFERENCES


1. *Dual-Mode Seeker Study, Final Report, Airforce/Army Contract No. F08635-85-C-0156, Honeywell Systems & Research Center.* May 1986.
2. 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).
3. B. Bhanu, "Automatic Target Recognition: State of the Art Survey," *IEEE Transactions on Aerospace & Electronic Systems* **AES-22**(4) pp. 364-379 (July 1986).
4. B. Bhanu and W. Burger, "DRIVE: Dynamic Reasoning from Integrated Visual Evidence," *Proc. DARPA Image Understanding Workshop*, pp. 581-588 Morgan Kaufmann Publishers, (Feb. 1987).
5. B. Bhanu and W. Burger, "Qualitative Reasoning in Dynamic Scene Understanding," *Submitted to Computer Vision, Graphics and Image Processing*, (1987).
6. B. Bhanu and W. Burger, "Qualitative Motion Detection and Tracking of Targets from a Mobile Platform," *Proc. DARPA Image Understanding Workshop*, (April 1988).
7. B. Bhanu and J.C. Ming, "TRIPLE: A Multi-Strategy Machine Learning Approach to Target Recognition," Proc. DARPA Image Understanding Workshop (April 1988).
8. 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).
9. B. Bhanu and P. Symosek, "Interpretation of Terrain Using Multispectral Images," *Submitted to Pattern Recognition*, (1987).
10. S. Bharwani, E. Riseman, and A. Hanson, "Refinement Of Environmental Depth Maps Over Multiple Frames," *Proc. IEEE Workshop on Motion, Kiawah Island Resort*, pp. 73-80 (May 1986).
11. T.J. Broida and R. Chellapa, "Kinematics and Structure of a Rigid Object from a Sequence of Noisy Images," *Proc. IEEE Workshop on Motion, Kiawah Island Resort*, pp. 95-100 (May 1986).
12. W. Burger and B. Bhanu, "Estimating 3-D Egomotion from Perspective Image Sequences," *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1987).
13. W. Burger and B. Bhanu, "Qualitative Motion Understanding," *Proc. Tenth International Joint Conference on Artificial Intelligence, IJCAI-87, Milan, Italy*, Morgan Kaufmann Publishers, (August 1987).
14. W. Burger and B. Bhanu, "Qualitative Understanding of Scene Dynamics for Autonomous Mobile Robots," *Submitted to International Journal of Robotics Research*, (1987).
15. J.Q. Fang and T.S. Huang, "Some Experiments on Estimating the 3-D Motion Parameters of a Rigid Body from Two Consecutive Image Frames," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(5) pp. 545-554 (September 1984).
16. K.D. Forbus, "Qualitative Process Theory," *Artificial Intelligence* **24**(1-3) pp. 85-168 (December 1984).
17. E.C. Hildreth and N.M. Grzywacz, "The Incremental Recovery of Structure from Motion: Position vs. Velocity Based Formulations," *Proc. IEEE Workshop on Motion, Kiawah Island Resort*, pp. 137-143 (May 1986).
18. R. Jain, S. L. Bartlett, and N. O'Brien, "Motion Stereo Using Ego-Motion Logarithmic Mapping," Proc. Int. Conf. Computer Vision and Pattern Recognition, pp. 188-193 (1986).
19. J. Kim and B. Bhanu, "Motion Disparity Analysis Using Adaptive Windows," Technical Report 87SRC38, Honeywell Systems & Research Center (June 1987).
20. A. Kramer, H. Funk, and J.A. Hawthorne, "Topological Representation," Proc. 1986 IR&D Initiatives Program, Honeywell Systems & Research Center (June 1987).
21. B. Kuipers, "Qualitative Simulation," *Artificial Intelligence* **29** pp. 298-338 (1986).
22. 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).
23. 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).
24. D.N. Lee, "The Optic Flow Field: The Foundation of Vision," *Phil. Trans. R. Soc. Lond. B* **290** pp. 169-179 (1980).
25. H.C. Longuet-Higgins and K. Prazdny, "The Interpretation of a Moving Retinal Image," *Proc. R. Soc. Lond. B* **208** pp. 385-397 (1980).

26. 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, pp. 117-128 (Dec. 1985).

27. A. Mitiche, S. Seida, and J.K. Aggarwal, "Determining Position and Displacement in Space from Images," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, pp. 504-509 (June 1985).

28. H.-H. Nagel, "Image Sequences - Ten (octal) Years - From Phenomenology towards a Theoretical Foundation," Proc. Intern. Conf. on Pattern Recognition, Paris, pp. 1174-1185 (1986).

29. H. Nasr and B. Bhanu, "Dynamic Model Matching for Target Recognition from a Mobile Platform," *Proc. DARPA Image Understanding Workshop*, (April 1988).

30. H. Nasr and B. Bhanu, "Landmark Recognition for Autonomous Mobile Robots," *Proc. IEEE International Conference on Robotics and Automation*, (April 1988).

31. 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).

32. J.A. Perrone, "Anisotropic Responses to Motion Toward and Away from the Eye," *Perception & Psychophysics* **39**(1) pp. 1-8 (1986).

33. K. Prazdny, "Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinear Moving Observer," *Computer Graphics and Image Processing* **17** pp. 238-248 (1981).

34. D. Regan, K. Beverly, and M. Cynader, "The Visual Perception of Motion in Depth," *Scientific American*, pp. 136-151 (July 1979).

35. J.H. Rieger, "Information in Optical Flows Induced by Curved Paths of Observation," *J. Opt. Soc. Am.* **73**(3) pp. 339-344 (March 1983).

36. E.M. Riseman and A.R Hanson, "Summary of Progress in Image Understanding at the University of Massachusetts," Proc. DARPA Image Understanding Workshop, pp. 48-60 (Dec. 1985).

37. H. Shariat and K.E. Price, "How to Use More Than Two Frames to Estimate Motion," *Proc. IEEE Workshop on Motion, Kiawah Island Resort*, pp. 119-124 (May 1986).

38. W.B. Thompson and J.K. Kearney, "Inexact Vision," *Proc. IEEE Workshop on Motion, Kiawah Island Resort*, pp. 15-21 (1986).

39. R.Y. Tsai and T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(1) pp. 13-27 (January 1984).

40. S. Ullman, "Maximizing Rigidity: The Incremental Recovery of 3-D Structure from Rigid and Rubbery Motion," MIT A.I.Memo No. 721 (June 1983).

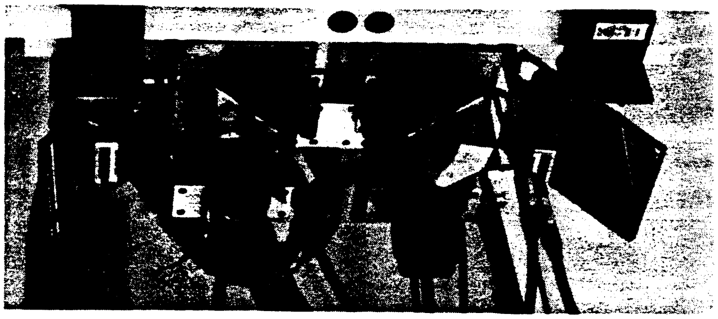
41. A. Verri and T. Poggio, "Qualitative Information in the Optical Flow," Proc. DARPA Image Understanding Workshop, Los Angeles, pp. 825-834 (February 1987).

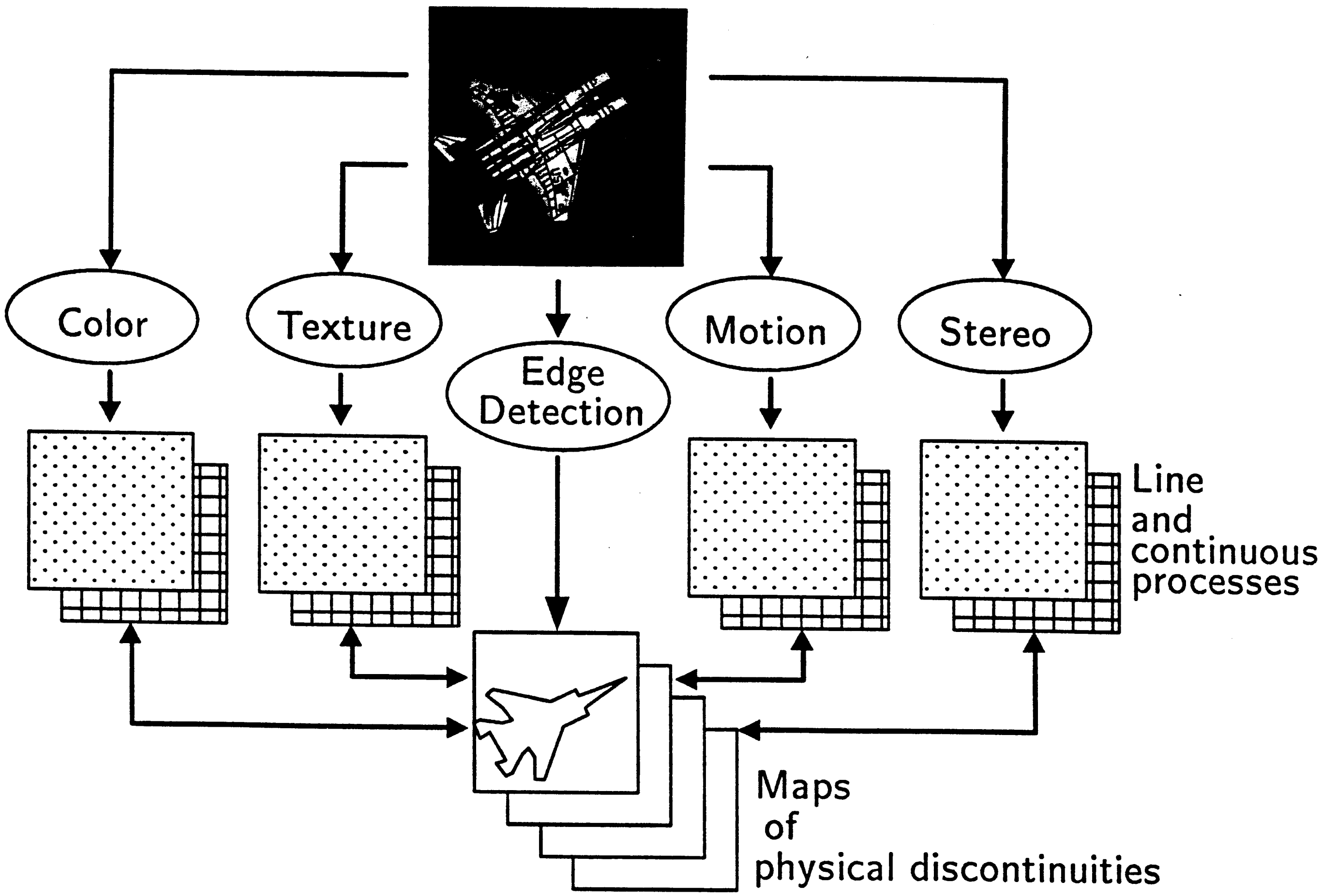
42. Y. Yasumoto and G. Medioni, "Experiments in Estimation of 3-D Motion Parameters from a Sequence of Image Frames," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 89-94 (1985).

Volume I

PROCEEDINGS:

# Image Understanding Workshop






**Sponsored by:**

Defense Advanced Research Projects Agency
Information Science and Technology Office

DARPA

SC

**April 1988**

Bri Blam

# Image Understanding Workshop

Proceedings of a Workshop
Held at
Cambridge, Massachusetts

**April 6-8, 1988**

Volume I

Sponsored by:

**Defense Advanced Research Projects Agency**
**Information Science and Technology Office**

**This document contains copies of reports prepared for the DARPA Image Understanding Workshop. Included are results from both the basic and strategic computing programs within DARPA/ISTO sponsored projects.**