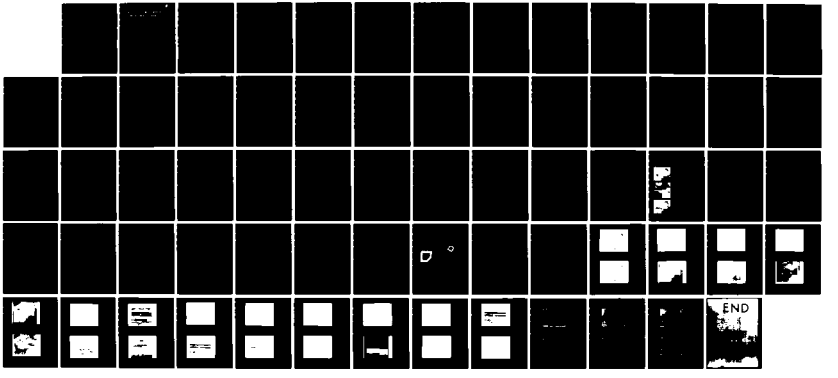
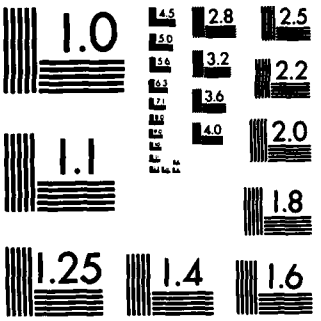
AD-A144 593 HIERARCHICAL MULTISENSOR IMAGE UNDERSTANDING(U) 1/1
HONEYWELL SYSTEMS AND RESEARCH CENTER MINNEAPOLIS MN
R K AGGARWAL JUL 84 AFOSR-TR-84-0639 F49620-83-C-0134
UNCLASSIFIED F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR-

(4) (2)

HIERARCHICAL MULTISENSOR IMAGE UNDERSTANDING

AD-A144 593

TECHNICAL REPORT
AFOSR F49620-83-C-0134
ANNUAL REPORT FOR PERIOD
OCTOBER 1983 - SEPTEMBER 1984

JULY 1984

DTIC FILE COPY

Approved for public release; distribution unlimited

Honeywell
SYSTEMS & RESEARCH CENTER
2600 RIDGWAY PARKWAY
MINNEAPOLIS, MINNESOTA 55413

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE

DTIC ELECTE AUG 20 1984 S D E

84 08 17 066

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 84-0889
6a. NAME OF PERFORMING ORGANIZATION: Honeywell, Inc. / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Honeywell Systems and Research, 2600 Ridgway Parkway, Minneapolis MN 55413	7b. ADDRESS (City, State and ZIP Code): Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR / 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-83-C-0134
8c. ADDRESS (City, State and ZIP Code): Bolling AFB DC 20332	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A7	

11. TITLE (Include Security Classification): HIERARCHICAL MULTISENSOR IMAGE UNDERSTANDING

12. PERSONAL AUTHOR(S): Raj K. Aggarwal

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Interim	FROM 1/7/83 TO 30/6/84	JUL 84	54

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Image processing; image understanding; artificial intelligence; scene analysis; and attributed graphs.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the research results on Honeywell's Hierarchical Multisensor Image Understanding program. Honeywell is developing a unified framework for the different hierarchical levels of image processing such as segmentation, detection, classification, and identification of outdoor scenes and across different sensor modalities such as millimeter wave, infra red, and visible. Current activities on the project are reviewed under the following headings: (1) AI-based generic image segmentation and object recognition; (2) evidence-confidence paradigms for image understanding; (3) hierarchical systems theory for control structures; and (4) invariant methods in image understanding. Also discussed are

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. Robert N. Buchal	(202) 767- 4939	NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

84 08 17 066

Hierarchical Multisensor

Image Understanding

Annual Report

1 October 1983 - 30 September 1984

Contract F49620-83-C-0134

Honeywell Systems and Research Center

Minneapolis, Minnesota 55413

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

ABSTRACT

This report describes the research results on Honeywell's Hierarchical Multisensor Image Understanding program. Honeywell is developing a unified framework for the different hierarchical levels of image processing such as segmentation, detection, classification, and identification of outdoor scenes and across different sensor modalities such as millimeter wave, infrared, and visible. Current activities on the project are reviewed under the following headings: (1) AI-based generic image segmentation and object recognition; (2) evidence-confidence paradigms for image understanding; (3) hierarchical systems theory for control structures; and (4) invariant methods in image understanding.

1. INTRODUCTION

This project is concerned with the study of a formal methodology for multisensor image understanding. It is being conducted under Contract F49620-83-C-0134 (AFOSR) monitored by Dr. Robert Buchal. Dr. King-Sun Fu and Mr. M. Eshera, both of Purdue University, are collaborating on some aspects of generic scene segmentation.

Conventionally, multisensor systems are treated as a set of domain specific subsystems (each optimized for one sensor domain such as infrared, visible or millimeter wave) that are integrated with each other only at the final output stage of processing. Honeywell is developing a unified framework for the flow of information and control between different hierarchical image processing levels (such as gradient, texture, context, etc.) and across different sensor modalities.

We are taking a multidisciplinary approach to the development of the unified framework. We are studying perceptual and physical invariants, developing and understanding of their mappings into different sensory domains at different representational levels, and developing machine intelligence techniques for image processing based on the invariants. Previously, we had developed image pixel level concommittant processing for simultaneous millimeter wave and infrared imagery and for simultaneous laser intensity and range imagery. This gave us some understanding of issues involved in multisensor image information integration. We also had previously developed knowledge based feed-forward control for scene segmentation in different infrared images with diverse characterstics.

We are now

(1) developing a functional model for the bidirectional(feedback and feed forward) control of information flow partially based on human visual and perceptual system.

(2) further analyzing and comparing image formation processes for multisensor vision.

(3) switching across different sensor modalities using physical scene invariants based on two-dimensional normalization filters.

(4) integrating the mode switching and level transition frameworks via appropriate production-rule structure with loose coupling of the hierarchical processing modules.

Thus far in the project, we

(1) have developed a successful context-independent scene segmentation approach which, unlike conventional approaches, does not depend on specific object models.

(2) have developed a dynamic spatio-temporal knowledge representation method that provides the knowledge base for multisensor vision control.

(3) have developed a hierarchical planner for control of information flow to automatically determine the optimum sequence of image processing operations and parameter values, and

(4) are developing a novel evidence accrual paradigm based on graphs with attributed lists as nodes and image processing operators as arcs.

This report reviews activities on the project during the period 1 October 1983 - 30 September 1984. This work is covered under the headings of segmentation and recognition; evidence accrual; control of information flow; and invariant methods. The work is summarized here since it is covered in greater detail in individual technical reports and conference papers. (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11).

2. AI-BASED GENERIC IMAGE SEGMENTATION AND OBJECT RECOGNITION

2.1 Context Independent Segmentation Inference Engine (CISIE)

Current generation image understanding systems cannot perform machine vision tasks in a wide variety of contexts (environments, conditions) without parameter modification. We have demonstrated the feasibility of autonomously processing digital imagery to discriminate regions which correspond to components of objects or areas of background terrain. Furthermore, the region discrimination was to be performed using only the information content of the original image. Thus, the system could be operated without restricting the context of input imagery. This is a

significant advance in the state of the art. Algorithms have been implemented in the image research laboratory and their performance measured on a database of tactical Forward Looking Infrared (FLIR) imager, typical of that used to test target acquisition systems.

The key milestones were:

1. Development of a set of combining rules for image primitives such as edges, bright blobs, and contours
2. Development of a consistent set of conflict resolution rules to resolve region conflicts between different "combined" images.
3. Laboratory demonstration of a rule-based region discrimination concept

Technical Approach - The first stage of an image processing system extracts image primitives such as edges, textures, and contours. Our approach is to apply rules which discriminate the structural regions in the original scenes based on the spatial coincidence of the various image primitives. These rules depend only on the primitives derived from the image, not on knowledge of expected scene content (e.g., tanks or road). Thus, the second stage of CISIE combines a set of image primitives to yield a single labeled image. The third and final stage applies a set of conflict resolution rules to a set of labeled images, each of which has been produced by processing the same image through different combining procedures. Conflict reolution yields a region discriminated image. The approach is illustrated in Figure 1.

Key Accomplishments - We have succeeded in demonstrating the feasibility of our technical approach for CISIE. Reasonable region discrimination was performed on FLIR imagery whose content and image quality varied widely, without the use of contextual information such as knowledge of scene objects or range to ground.

Milestone Specific Results

- Milestone 1-Eleven processes for combining image primitives were considered. Three combining processes were rejected because of expense of implementation. We experimented with the other eight. Four of these proved to require context-based information in order to perform reasonable region discrimination. The remaining four

combining processes were a homogeneity operator which finds regions of little intensity change, an inhomogeneity operator which finds coarse boundaries, Imaging Sensor Autoprocessor's (ISA) texture boundary locator which finds changes in texture, and ISA's prototype similarity transformation which finds areas of common texture. Figure 2B shows the results of these four operators on the original FLIR image pictured in Figure 2A. These procedures define the combination of primitives specified in milestone 1. Note that a procedural (algorithm) rather than declarative (rule-based) implementation was chosen because of the high computational overhead in pixel-level rule-based decisions.

Each of the four combinations of primitives results in an (possibly different) initial discrimination of the image into significant areas. This results in four labeled images.

- Milestone 2-A set of heuristic rules codifying the behavior of each of the operators, both individually and relative to each other, was used to formulate heuristics for conflict resolution between area discriminations. Each area in each of the labeled images was viewed as a characteristic (i.e., "on-off") function in order to achieve a novel, real-time implementation of the conflict resolution rules. Each pixel in each labeled image is labeled "on" if it is part of a region, and "off" if it is part of a boundary which separates regions. Thus each labeled image is binarized. The four binarized labeled images are put in adjacent bit planes in image memory, forming a new Four-bit image. Now the heuristic rules are coded into a look-up table mapping the four-bit image to a binary image. This performs conflict resolution in real time. The results of conflict resolution are shown in Figure 2C.

- Milestone 3-The CISIE approach was tested on eight digital FLIR images which ranged in complexity from single isolated targets in a relatively homogeneous field to highly cluttered imagery of power plants and general terrain. Reasonable context-free region discrimination was demonstrated. The process is illustrated in Figure 2 on a FLIR image of a power plant.

In addition to the research being conducted at Honeywell under this contract, world known Professor King-Sun Fu and his student Mr. Mohamed Eshera, both from Electrical Engineering Department at Purdue University, are also involved in investigating generic context independent image segmentation via Attributed Relational Graphs (ARG). Further details are contained in (12, 13).

2.2 AI-based Feedback Resegmentation

We have found that in general, an image cannot be totally segmented in a single pass of one segmentor. Better segmentation results when the image is first segmented at a lower resolution into a few regions. These regions are then in turn segmented into smaller regions. This process continues until the image has been segmented into homogeneous regions. This process can be represented as a tree where each node is a region. Branches form a node show it being made up of smaller regions. The head node of the tree is the entire image are leaf nodes are the smallest elementry regions that the image is made up of. The size of these smallest regions, and therefore the resolution of the segmentation depends on the size and type of the objects we re looking for in the image. This means that in an image understanding system even the image segmentation needs to be driven by higher level information such as range and the system goals.

At each node in the segmentation tree we have many different segmentation operators to choose from. The choice of an operator to use is based on knowledge about the sensor, the goals of the system at a given time, and knowledge about the behavior of the different operators on different input conditions. It is also possible to choose more than one operator at a given node. This creates parallel branches in the tree and multiple representations for a region. These multiple representations give more information about the region that can be used when classifying the region.

Two examples are described of how this segmentation tree is created. To keep the examples simple only two different operators were used.

The first example is a FLIR image of a power plant. Figure 3 shows the segmentation tree for this image. The image was first preprocessed to do noise cleaning. The type of preprocessing used depends on the type of sensor and on actual measurements from the original image. Operator 1

yielded a poor segmentation while operator 2 separated the image into good regions. These regions were then cut out of the original preprocessed image and segmented further using the two operators. The numbers in the tree nodes correspond to the picture number. See pictures 1-10.

Our second segmentation example shows a FLIR image of a multi-lane highway with vehicles on it. The segmentation tree for this image is shown in Figure 4. After preprocessing, the image was segmented at a low resolution to find the general regions of interest shown in picture 13. Picture 5 shows one of these regions cut out and picture 6 shows the segmentation of this region at a higher resolution. The other side of the segmentation tree shows the same type of processing on different regions of interest found by operator 2. See pictures 11-26.

2.3 Dynamic Scene Analysis Inference Engine (DSA)

Conventional image processing techniques deal mainly with the detection, extraction and recognition of objects, often by operating in a local area around the target. They fall short of utilizing spatial, temporal, relational and in general global information available in the scene.

To demonstrate the existence of this information in a dynamically changing scene and to appreciate the benefits to scene anlysis, we have synthesized a sequence of Frames 1 through 6, on our SYMBOLICS 3600 LISP machine.

For instance Frames 1 and 2 suggest that the three tanks will cross the bridge and the segmentor operators and parameters (edge operators, background estimators, thresholds, etc.) should be chosen according to the bridge contrast, texture, noise, etc. However, Frames 5 and 6 strongly indicate that the tank convoy is turning upstream and therefore the segmentor should be directed for tanks in terrain segmentation. This results in more robust target segmentation and better scene understanding, as it can "reasonably" be expected that the second and third tank will follow the fist tank.

DSA is designed to exploit the synergistic benefits of these aspects in scene understanding through a combination of reasoning and inferencing techniques. The reasoning process is modeled after the expert's

sequential steps in understanding a scene from low to high level entities, through the representation of operative knowledge by a frame/production rule structure. The inferencing is designed to enhance the robust users of the system by handling incomplete information, requesting missing information and identifying incorrect information. This could for instance correspond to lowering the thresholds to search for a "suspect" road that the initial segmentation missed. The system also includes query and explanation facilities for the man-in-the-loop mode.

3. EVIDENCE-CONFIDENCE PARADIGMS AND INFORMATION FUSION

In a distributed sensor environment symbolic information fusion is the integration of partial knowledge obtained from each sensor to arrive at a complete sensor-derived representation. It uses model-based representation of targets stored in the knowledge base and is guided by an inference engine in the accrual of evidence and matching. The flow of information during symbolic information fusion is depicted in Figure 5.0. The outcome of the symbolic information fusion is the full identification of objects in the scene.

The inference engine performs the reasoning process that involves the matching of two or more representations to identify their differences and similarities and also updates the confidence measures of the different possible targets and their components. These two functions are accomplished by a semantic net comparator and an evidence accrual module.

Identifying the best match between the sensor-derived representation and the model-based representation previously stored in the knowledge base provides the following:

- Object recognition - identifies the object type from the sensor derived representation.

- Directed rederivation of target representation - the inference engine is used to cue rederivation of the sensor derived representation so as to improve the level of confidence in the object recognition.

- Occlusion prediction - postulates a partial sensor derived representation to predict missing or occluded components. This enhances the recognition of occluded objects.

Matching tree data structures is generally carried out through classical search methods and hypothesize-test paradigms. Classical search methods include depth first, breadth first and A* optimal search; these methods are well documented in [14]. The hypothesize-test paradigm can be performed in either forward chaining or backward chaining modes. The forward chaining is also called data driven or bottom-up and the backward chaining is called the event driven, or top down.

As an example consider a particular target such as a tank. In forward chaining, components of the sensor derived representation are matched with the model-derived representation trunk, tread, and engine. If the match succeeds, then the componets are grouped and named body. Turret and barrel are then identified. Then the sensor-derived representation is a tank is true, or vice versa. Heuristics can be applied to facilitate a fast match. Honeywell is investigating both approaches.

The matching mechanism first constructs a network fragment, representing a sought-for-object, e.g., a tank, and then matches the network fragment against the network data base to see if such an object exists. For example, if a tank is sought, the fragment network depicted in Figure 6 will be generated and the components will be matched. The matches wil make inferences to create extended network structures, e.g., trunk, engine, and tread during the matching process.

The criteria in evaluating the matching technique we are developing are:

1. Capability to recognize objects and reject clutter with partial information.

2. Capability to provide resegmentation direction.

3. Capability to determine occlusion.

In any matching technique, a similarity measure is used to determine the best match between the sensor-derived representation and one the model-derived representation. Commonly used similarity measures are Euclidian distance, mean square error, and Bayesian probabilities. These similarity measures work well in statistical pattern recognition, but their applicability in partial knowledge matching is limited. In artificial intelligence research, inexact reasoning has proved its usefullness in medical diagnosis, chemical analysis, and natural language understanding. Subjective Bayesian models have been used in expert systems. In particular, Shortliffe, and Buchanan [15] have devised a method for incremental accrual of classification confidence which is based on confirmation theory. The theory assumes that one can formulate approximations for a priori and conditional probabilities by using them to determine measures of "belief" and "disbelief". These belief measures are in turn used to define measures of confidence and rules for incrementally updating both the belief and confidence measures (see Figure 7).

The belief and disbelief measures as they are implemented in expert systems, are not spatial adaptive. That is, once the evidence for belief is accrued, its significance never changes regardless of the outcome of other spatially located sensors. Honeywell is extending the belief and confidence measures to incorporte the incremental evidence provided by the distributed sensors. Such a framework would have the potential for providing a unified inferencing framework that can work with partial representations and provide direction for rederiving the sensor-derived representation.

Sensor-Derived Representation (SRD) - At every step, the semantic net comparator produces intormation that validates or invalidates previous evidence obtained from the sensors. The sensor-derived representation, as a component in the distributed sensor target recognition system, maintains and updates the values of confidence measure for each of the possible targets, e.g., truck, tank, jeep, etc. To accomplish this task the SDR makes a copy of the model-based representations from the knowledge base for each of the candidate targets. Figure 6 depicts the initial state of the sensor-derived representation. The graphs initially contain zero confidence measures for all the targets and their components. As the interence engine obtains information from the 2-D views it builds or

accrues evidence for each of the target graphs. For example, Figure 8 displays the graph for a tank with a total confidence measure of 0.492. Table 1 is an example of how the SDR updates the confidence measures as the 2-D views are analyzed. The distributed sensor system arrives at a specific target indentification whenever the target's confidence measure exceeds a suitable threshold or it exhausts the available sequence of 2-D views available, in which case it chooses the target with the highest confidence measure and provides suitable warnings.

The sensor-derived representation evolves as more information about the target(s) is obtained by the different images or derived by the inference engine. Honeywell's unique approach to symbolic information fusion is based on coordinating, updating, and validating the attributes of targets to achieve a parsimonious representation. Redundancies and conflicts are resolved by the inference engine at the fusion level.

4. CONTROL OF INFORMATION FLOW

In a recent work [4], we explored ways in which fundamental concepts in artificial intelligence (AI) can be applied to image understanding (IU) systems. These three basic areas were the use of:

1. Knowledge Representation Levels

2. Control Structures

3. Constraints (both natural and domain-specific).

The use of difference representation levels is an area that is now well known in the IU field. Several notable researchers have developed this concept and its application to IU systems in some detail. [16,17] This is an area, however, where it may still be somewhat too early to attempt to classify and clarify the numerous types of constraints available, their interactions, and the ways in which they can best be incorporated into intelligent IU systems.

In comparison with the above two areas, the use of control structures in image understanding systems stands out as an area which has not yet received strong conceptual development, but yet is ripe for just such an approach. For purposes of this report, we will distinguish control structures from knowledge representation by stating that representation levels will be used to store static knowledge at differenct levels of refinement throughtout the system. We can imagine looking at "snapshots" of the contents of the different representation levels at different times during the processing in the system. At any given moment, the contents of a representation level essentially portray a static form of information regarding the objects in a scene and their relationship with one another. In contrast, the control structure for an IU system will contain inherently dynamic or process knowledge; it will be the knowledge about how to use or operate the information in the different representation levels in order to generate further or more refined knowledge. These definitions, of course, have been adapted from the classical AI concepts for IU purposes, and thus may be somewhat different from other points of view as a result.

The Nature of Hierarchical Systems -

The concept of a hierarchical system may be well-defined. At the outset, hierarchical systems may be categorized as those which have the following features: (18)

1. There is a vertical arrangement of the subsystems.

2. The higher subsystems have the right to intervene in the actions of the lower subsystems.

3 The effectiveness of the higher level subsystems depends on the actual performance of the lower levels.

There are three basic views of hierarchical systems . Each of these may be applied to any given system, although often one view may contain more usable information than another.

A system may be viewed as successive levels of description or abstraction. These levels, or strata, all describe the same system, but each description carries different information. Another hierarchical view of a system would be based on levels of decision complexity. This view, in which the levels will be referred to as layers, will prove quite useful for AI applications, and so will be discussed in some depth. A final view of systems is based on the organization of decision making units. In this view, where the levels are referred to as echelons, the distinction is made on the horizontal relations of units, where there should be more than one unit on the lowest level. The layers, mentioned above, are distinguished on the basis of vertical decomposition into subsystems. These two concepts are quite similar, and will often blend together in the discussion which will follow.

With respect to hierarchical arrangements of decision units which comprise a system, the following categories of decision-making systems can be recognized: single-level, single-goal; single-level, multigoal system; and multilevel, multigoal systems. The AI/IU programs to which this theory will be applied will fall into the latter category.

Communication between supremal and infimal units must be bidirectional. The supremal unit can signal downward to the infimal unit, where the signal will represent intervention. This intervention should specify decision problems for infimal units. Infimal units should also be able to signal upward to the supremal units. Generally, their signals will represent the status of activities undertaken.

In general, the supremal units have two broad responsibilities in dealing with the infimal units. The first is to instruct infimal units in how to proceed by selecting for them the rules and procedures to be followed. This is referred to as "selection of a coordination mode." The second major responsibility is to influence the infimal units to change their actions (if necessary), or to adjust the roles of infimal units in order to improve performance. The selection of actual intervention or use of a control variable will be referred to simply as "coordination".

It is also necessary that units on the same level be allowed some form of communication. The relationships between units of the same level will be characterized the action of a units and by the response of the rest of the system as it influences that units. This response is referred to as an interface input. Supremal units determine how infimal units will account for the interface input. There are five general ways in which this can be done. Summarized below, they are:

1. Interaction Predication Coordination, in which the suremal unit specifies the interface output. The infimal units solve their local decision problems based on this information alone from the other units, and must assume that it correctly represents the known state of the system.

2. Interaction Estimation Coordinations, in which the supremal unit specifies a range of possible values which the interface inputs may have. The infimal units treats the inputs as distriburances which may assume any value within the given range.

3. Interaction Decoupling Coordination, allows the infimal units to treat the interface input as an additional decision variable. They solve their decision problems as thought the value of the interface input could be chosen independently.

4. Load-Type Coordination, is the first case in which the infimal units actually recognize the existance of other units on their level. The supremal unit provides the infimal units with a model of the relationship between its action and the response of the system.

5. Coalition-Type Coordination, expresses the situation in which the infimal units not ony recognize the existance of other units on their level, but are allowed an interaction with other units. The form of the interaction is controlled by supremal units.

Of these approaches, the last is the most sophisticated and comes the closest to the organization of human hierarchical systems. However, it is quite complex. The first three appraoches are those that would be easiest to implement in an AI/IU application. It is worth noting, however, for possible future reference, that the usefullness of communication between infimal units depends on how the infimal problems are defined. Some problems benefit from communication, whereas others (generally the simplest sort) do not. Also, it can be shown that the effect of excessive communication between units may have the same effect as the lack of communication in leading to an overall deterioration of performance.

Complex Hierarchical Systems

It can be useful at this point to review the terminology introduced previously to describe the levels in a hierarchical system. The concept of strata was introduced to indicate the choice of abstraction layers that could be used. The concept of layers refers to the vertical decomposition of a decision problem into subproblems, and the concept echelons is used when there are more than one decision units on one layer. In the previous discussion, the concept of communication between units on the same layer implicity referred to a multiechelon decision-making system.

It is possible that a complex problem or situation would require a complex multilayer hierarchy to adequately represent the system. In this case, each of the decision units in a multiechelon hierarchy may use a multilayer approach to solve their own, local subproblems. This concept is illustrated in Figure 10.

It would also be possible to decompose a larger system into several vertically arranged decision units. Each decision unit could be viewed as both a multilayer hierarchy and as a multiechelon hierarchy, as is illustrated in Figure 11. In this example, the learning and adaption layer corresponds to an echelon with two units, and the selection layer to an echelon of four units.

Regardless of how complex a hierarchy becomes, several features will remain constant:

1. A supremal unit will always be concerned with a larger portion or broader aspect of the overall system behavior.

2. The decision period of supremal unit will be longer than that of its infimal units.

3. The supremal units will be concerned with the slower aspects of the overall system behavior.

4. Decisions and problems on higher levels are less structured, will contain more uncertainties, and will be more difficult to formalize quantitatively.

Using MLMGHST as a control structure paradigm

After the preceeding discussion of the nature of Multi Level, Multi Goal Hierarchical System Theory (MLMGHST), we might ask ourselves: Why should this apply to control structure issues in AI/IU systems?

The purpose of an IU control structure is to guide the choice of appropriate algorithms in order to achieve certain goals. The control structure of an IU system is both goal-oriented and process oriented. It is easy to see that in a complex system, the hierarchical ordering of process knowledge into strategic, tactical, and operational (algorithmic) processes could prove useful. If this is so, then the design schema presented before could prove a useful basis for system design. Further, the five types of interlevel communication protocols presented earlier could prove useful, not only in designing an initial system, but also in

planning for future system development and upgrading, by successively modifying the communication protocols into more complex structures.

That MLMGHST is primarily useful for structuring control in a system, rather than designing the representation levels or the use of constraints, can be shown by examining each of these systems components separately. Although the commonly suggested structrure of IU knowledge representation levels is hierarchical, each level may be viewed as a means of structuring the relatively static knowledge in each image (Figure 9). Algorithms are used to pass information from one level to the next. Although there is some possibility for using the ideas of self-organizing, learning and adaptation, and selection levels in designing more complex representation levels, it appears that the most immediate and profitable use of MLMGHST is to apply it to control structure which contains the process knowledge for the system. In this way, we would have to view the multilayer control sturcture as a construct superimposed on the representation levels and algorithms. The algorithms or operators would be in the same plane as the representation levels, but we would have to view the strategic and tactical control levels as projecting out of the plane, forming a 3-D construct.

The use of MLMGHST is similarly more suited for control than for structuring the use of constraints. There are two types of constraints which are currently in use in AI/IU system; natural and domain-specific. The natural constraints which are currently in use are low level (e.g., "surfaces tend to be continuous"), and hence are incorporated into the system at the algorithmic level. Higher level natural contraints are not well evolved.

On the other hand, domain-specific constraints embody knowledge about subjects or attributes that are likely to pertain to the image being used. This knowledge often expresses relationships about the observable features or regions. This is often stored in knowledge bases which are separate from but associated with the representation levels. Often, the use of domain-specific constraints is crucial to the success of an IU system, but since the constraints are unique for each system, examples of their use is defered until later. The MOLGEN system [19] is an example of how domain-specific constraints may be used as an integral part of the

classical expert system with hierarchically organized process knowledge or control.

Given that MLMGHST is best suited for application to organizing the control structure of a system, it is possible to further specify those types of systems for which it can best be used. First, the system should be sufficiently complex so that the organization of processes into strategic, tactical and algorithmic groupings seem natural. This could mean that a large number of algorithms and/or algorithm pathways connecting representation levels should be available. A second aspect of a sytem that would benefit from this type of control is that the system should be designed to handle multiple (but not necessarily simultaneous) goals. For example, an IU system for robot bin-picking would benefit from this type of control structure if there is more than one type of part for which it will search. An IU system for an autonomous vehicle would have the multiple goals of needing to characterize both objects and terrain. Different types of representation could be required for the different types of objects and the terrain, necessitating a strategic/tactical approach to determining which of the several representation schemas should be employed. As a third example, multisensor IU systems will need to determine which, among several sensors or combination of sensors, would be most efective under different conditions, or when searching for different types of objects. Each of these example areas present a compelling need for designing robust control structures which are more sophisticated than those currently in use.

5. INVARIANT METHODS IN IMAGE UNDERSTANDING

5.1 Object recognition and scene parametric analysis

An analysis of the effect that scene parameters have on object recognizability is fundamental to the understanding of distributed sensor phenomenology. Accordingly, we started out with a parametric analysis which includes the following:

- Evaluation of the number and location of 3-D sample points which are necessary for the discrete representation of objects and

- Investigation of the effect of sensor aspect angle, depression angle, range to an object and sensor angular separation on the recognizibility of the objects.

The importance of the sampling point selection of the discrete representation of any object becomes apparent once one is reminded that in all the past attemps at 3-D object recognition, assumptions have been made about the availability of 3-D points which completely represent an object. These points usually are obtained arbitrarily or imposed on the 3-D objects by projecting a rectangular mesh over their surfaces. No attempt has been made to justify these approaches or study the effect of point selection on the recognition of the object.

The effect of the angle of separation between the sensors, aspect and depression angles as well as object range on the recognizability of the object has also not been addressed in the past. It is intuitively apparent that as the angle of separation increases, more of the object is viewed, hence more information is obtained for recognition. While such intuitive feelings are helpful, a quantitative parametric analysis is essential for a thorough phenomenological understanding.

In order to permit a parametric analysis one needs a suitable figure of merit of target characterization, i.e., a barometer of target recognizability. This target attribute has to be capable of adequately characterizing the target object. One can then deduce a parametric analysis by studying how this attribute degrades from its ideal value as the parameters are varied.

One such attribute which we have used is a 3-D moment invariant [20]. In this method, a set of functions that may be used to represent 3-D objects independent of size and coordinate system is derived. Knowing the proper number of discrete points and their position on the object is a prerequisite for this calculation, and hence the importance of a theoretical analysis of the 3-D sampling phenomenon. The next logical step is to determine the effect of undersampling on the calculation of 3-D moment invariants. This undersampling can be due to the fact that based on the sensors' geometrical location and orientation only part of object may be "seen" and so the sample points may not "cover" the whole 3-D

target surface. Likewise, the other scene parameters may also be varied and their effect on target recognizability analyzed y observing the 3-D moment invariants. We now provide a brief review of the moment invariants.

3-D Moment Invariants - In order to recognize any 3-D object indepent of size, position, and orientation, one must obtain measurements which convey the invariant attributes of the object. The use of three-dimensional moment invariants provides an excellent representation of 3-D objects [20].

The three-dimensional central moments of order p + q + r of a density g (z_1, x_2, x_3), are defined as:

$$\mu_{pqr} = \iiint_{-\infty}^{\infty} x_1^p x_2^q x_3^r g(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

where, for the sake of simplicity, the centroid is assumed to be at the origin.

It has been shown [20] that for quadratic surfaces, which form a special but important subset of general ternary quantics, a set consisting of two moment variables can be derived with the following results:

$$\frac{J_{1\mu}^2}{J_{2\mu}} \quad \text{and} \quad \frac{\Delta_{2\mu}}{J_{1\mu}^3} \quad \text{where } J_{1\mu} = \mu_{200} + \mu_{020} + \mu_{002}$$

$$J_{2\mu} = \mu_{020}\,\mu_{002} - \mu_{011}^2 + \mu_{200}\,\mu_{002} - \mu_{101}^2 + \mu_{200}\,\mu_{020} - \mu_{110}^2$$

$$\Delta_{2\mu} = \det \begin{bmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{bmatrix}$$

Analysis Procedure - For the parametric analysis, one needs to study the variation of moment invariants from their ideal value for each object class, as the sensor parameters are varied. the following steps will be taken for the analysis:

1. The object is first encoded into a computer by using a scale model of the object and any one of the various 3-D coordinate machine devices that are commercially available [21].

2. In order to study the effects of 3-D sampling, in analogy with 1-D and 2-D cases, we first select a set of suitable cut-off frequencies (since the Fourier transform of any finite object is infinite [22]) using appropriate criteria, e.g., maximum volume criteria which leads to the selection of the first zero crossings. This then serves as the baseline for studying the effects of undersampling, which can be simulated on the computer.

3. In order to study the effects of the scene parameters, the desired view of the object is obtained by means of computer software such as commercially available MOVIE-BYU [23] which can give the desired view of the object when range, depression and aspect angles are specified.

4. The coordinates of the points that are viewed by the sensor are evaluated by changing one parameter at a time while keeping other constant.

5. The 3-D moment invariants of the target as it is viewed by the sensors are then calculated.

6. The moment invariant sets are then obtained for other objects under identical goemetrical conditions.

7. A distance measure is defined in terms of sets of moment invariants of the objects as functions of the various parameters. This provides a quantification of the objects recognizability.

To obtain a phenomenological understanding of some of the parameters that have effects on object recognizability we will further investigate the following:

1. The effect of the number of sampling points and their location on the object on the recognizability of the object,

2. The effects of the sensors' angle of separation on the object recognition ability,

3. The effects of the depression and azimuth angles of the sensors on the object recognition performance, and

4. The effect of the range of the object from each sensor on the recognition performance.

5.2 Invariant Analogical Image Representation

We present a summary of an image representation that uniquely encodes the information in a gray-scale image, decouples the effects of illumination, reflectance, and angle of incidence, and is invariant, within a linear shift, to perspective, position, orientation, and size of arbitrary planar forms. A detailed description of our research results can be found in a separate Technical Report (11).

The challenge of the visual information problem stems from the fact that the interpretation of a 3-D scene from a single 2-D image is confounded by several dimensions of variability. Such dimensions include uncertain perspective, position, orientation, and size, (pure-geometric variability) along with sensor mode, object occlusion, and non-uniform illumination. Vision system must not only be able to sense the identity of objects despite this variability, but they must be able to explicitly characterize such variablility. This is so because the variability in the image formation process (particularly that due to geometric distortion and varying angle of incident illumination) inherently carries much of the valuable information about the imaged scene. Consider human vision for the moment. In spite of the complications introduced by geometric distortion, it is precisely the "unraveling" of such distortions that enables a human to readily perceive the three-dimensionality of any static 2-D image--be it a single face of a Necker cube, or a 15th Century painting of Da Vinci. Indeed, humans seem capable of unraveling the physical and geometric distortions in an image almost as precisely as the physics and geometry of the world created in the first place.

Contrasted with the apparent ease and elegance of human visual interpretation of scene geometry, current vision algorithms are clearly lacking. It is becoming increasingly clear that much of the blame lies with conventional image representations. Good image representations must satisfy a number of requirements which seem to be mutually incompatible.

First, they should be simultaneously compact and complete in their representation of gray-scale image information. Compactness is synonymous with ease of computation and efficient use of memory. Completeness, on the other hand, implies that, if desired, one could fully reconstruct the original image from which the representation is derived. Secondly, image

representations must provide good intra-object clustering and inter-object separability independent of image distortion while at the same time preserving information about pattern distortion. No conventional image representation satisfies all of the above conditions. Though many conventional representations claim compactness, most do not make a credible attempt to decouple information about object identity from information about viewing geometry and illumination, nor do current representations fully exploit the abundance of gray-scale information in an image. In contrast, invariant "analogical" image representations , such as that advocated in this project, can satisfy the above requirements.

Decoupling Multiplicative Processes in Image Formation.

For the case of an ideally diffusing surface, an image can be modeled as a product of three independent signal componets:

$$f(x,y) = i(x,y) \cdot r(x,y) \cdot \cos t \qquad (1)$$

Where $i(x,y)$ is the illumination, $r(x,y)$ is the reflectance, and t is the angle of incidence of the illuminating light (24). Let's assume that any additive noise that is present has small magnitude relative to the above three componets of the signal. Then there is a well-known and effective method known as homomorphic filtering that allows one to individually filter out such multiplicative signal components when certain reasonable conditions are met. This method is briefly reviewed below.

Suppose one takes the logarithm of our function $f(x,y)$ as defined above. There results the so-called "density image".

$$\ln f(x,y) = \ln i(x,y) + \ln r\,(x,y) + \ln \cos t \qquad (2)$$

Hence the product becomes a sum of three density componets $\ln i$, $\ln r$, and $\ln \cos t$. If these three additive density componets have Fourier spectra which overlap very little in their regions of significant energy, then linear filtering can be used to extract any one of them. By taking the exponent of the extracted componet, one obtains the corresponding multiplicative signal component which appeared in the original expression

for the image f(x,y). Homomorphic filtering has been applied to enhance imagery by selectively filtering out the slowly varying illumination component with very impressive results (25).

The success of homomorphic filtering clearly stems from the fact that the Fourier transform of a density image often has the effect of "representationally" decoupling multiplicative image componets. If image pattern recognition (rather than image filtering) could be based on such a Fourier representation, then perhaps the effects of surface reflectance, illumination, and angle of incidence could be decoupled. This in turn could lead to methods for form recognition that are insensitive to varying illumination conditions. Unfortunately, as is well know, neither the Fourier transform, nor its (phaseless) power spectrum, have proved to be especially useful for image pattern recognition. However, there are alternative representations - namely, simultaneous spatial/spatial-frequency representations - which, like the Fourier spectrum, can provide the decoupling of multiplicative image density componets and, at the same time, overcome the classical shortcomings of the Fourier spectrum as an image representation.

Simultaneous Representations of Space and Spatial Frequency.

Vision researchers have traditionally emphasized the importance of either the spatial or the spatial-frequency domain, but not both. This should be contrasted with conventional representations of acoustic signals where simultaneous time/frequency representations (e.g. the spectrogram) have long been used. Nonetheless, vision researchers have very recently begun to express an active interest in simultaneous spatial-spatial-frequency image representations (26, 27, 28, 29, 30). Such representations can provide for improved separability of information characterizing visually relevant patterns and they are also compatible with the representation of gray-scale characteristics ranging from textures to object contours (27). Furthermore, simultaneous spatial/spatial-frequency image representatons can be used to decouple the effects of illumination, object reflectance, and angle of incidence.

The Wigner Distribution (WD).

Two simultaneous spatial-spatial-frequency representations have recently received much attention. They are the Gabor function representation (31, 32, 33) and the Wigner distribution (26, 27, 28, 29, 30, 34). (Note that we are here concerned only with 4-D directionally selective representations; this excludes, for example, Marr's 3-D DOG representation (24) and others like it which do not have a fourth dimension covering spatial-frequency angle. As discussed elsewhere (30) the Wigner distribution provides higher simultaneous resolution than is possible using the Gabor functions. In fact as discussed in (35), every simultaneous representation ever proposed can be expressed in terms of averages of the Wigner distribution over its independent spatial and spatial-frequency variables. Like the Fourier transform, the Wigner distribution (WD) is not, in general, a computable function since this would require the evaluation of infinite integrals. However, just as the Fourier transform has proved to be an elegant and convenient transformation with which to handle many problems in the spatial-frequency domain, the WD is an invaluable tool for problems <u>simultaneously</u> involving the spatial and spatial-frequency domains. As with the Fourier transform, the WD can be used in practice by employing an approximation obtained via finite integration windows. A particularly attractive approximation to the WD, denoted as the "<u>c</u>omposite <u>p</u>seudo <u>W</u>igner <u>d</u>istribution" (CPWD) has been introduced (30).

The WD and Multiplicative Signal Component Separability

The essence of the homomorphic filtering was to transform the image density function (i.e. the logarithm of the sensed image function) into the spatial-frequency domain in order to decouple the image components due to illumination, reflectance, and angle of incidence. To simplify our calculations, we assume a two component model including only illumination and reflectance:

$$f(x,y) = i(x,y) \cdot r(x,y) \qquad (3)$$

Taking the logarithm:

$$\ln f(x,y) = \ln i(x,y) + \ln r(x,y) \tag{4}$$

and computing the WD of both sides yields:

$$W_{\ln f}(x,y,u,v) = W_{\ln i}(x,y,u,v) + W_{\ln r}(x,y,u,v) + 2\, W_{\ln i,\, \ln r}(x,y,u,v). \tag{5}$$

where (x,y) and (u,v) specify the spatial and spatial-frequency domains, respectively. The third component, representing the cross-Wigner distribution (6) occurs because computing the WD involves a nonlinear correlation operation. Nevertheless, if the auto-WD's of the functions in i and r do not overlap substantially in space and spatial frequency, then approximations to the cross-WD contribution will in practice contain negligible energy. (This non-obvious fact follows from considering the definition of the CPWD (30)). Therefore, as with the Fourier transform, if regions of significant spectral energies of $W_{\ln i}$ and $W_{\ln r}$ are disjoint, then separability of the multiplicative signal components will have been achieved. These observations generalize to the case where angle of incidence effects are included; then the right hand side of equation (5) would include the auto-WD of the density function of this component along with its cross-WD's with the illumination and reflectance density functions. The Wigner distribution, like the Fourier transform, can therefore decouple the effects of illumination, reflectance, and angle of incidence when the spectra of these components are mutually disjoint. We describe in the remainder of this report how the Wigner distribution can be employed to define a unique image representation that, in addition to providing decoupling of multiplicative image components, also provides invariance to geometric distortions introduced by the imaging process.

Invariant Form Recognition In The Frontoparallel Plane.

As discussed earlier, good image representations should not only decouple the effects of illumination, reflectance and angle of incidence, but they should also allow objects to be recognized irrespective of the a priori unknown geometric distortions introduced by the image formation process. We begin this section by discussing only methods for obtaining invariance

to linear transformations of non-occluded, planar gray-scale patterns in front parallel view. The more general problem including perspective distortion is reviewed in (11). One particular approach for obtaining invariance to linear transformations makes use of the complex-logarithmic (CL) conformal mapping. Some researchers (36,37) have advocated use of the representation derived by CL conformally mapping the gray-scale image function itself. Others (38,39,40) have suggested that the CL conformally mapped Fourier power (or magnitude) spectrum should be used. In fact, neither representation seems entirely adequate. The former representation (CL mapped image function) is indeed invariant, within a linear shift (WALS-invariant), to rotation and scaling of the image about a single image point. However, such a representation is not invariant to translation of an image and the effects of illumination and reflectance are not in any way decoupled from one another. The second representation mentioned above, (Cl conformally mapped power spectrum of an image), is strictly invariant to translation and WALS-invariant to rotation and scaling of an image, but it does not uniquely represent an image since Fourier phase information is discarded. As alluded to earlier in our discussion of multiplicative signal separability, the loss of phase information makes Fourier spectra especially ill-suited for imagery containing clutter or multiple objects to be recognized. Since the WALS-invariance properties of the above-mentioned representations arise exclusively from using the CL conformal mapping, one is naturally led to consider whether this mapping could be used to develop considerably more robust image representations (perhaps simultaneous spatial/spatial-frequency representations) which are similarly WALS-invariant to rotation and scale changes. We show in (11) that this is possible.

To review our progress thus far, we now have an 8-dimensional representation which is WALS-invariant to all common geometric distortions of rigid planar forms and which, given reasonable assumptions, is also invariant to non-uniform illumination. We briefly describe next how such a representation can be used to actually perform image pattern recognition.

The memory prototype pattern characterizing some arbitrary planar form is just the 4-D CL conformally mapped WD of a frontoparallel view of that form, where spatial domain mapping has been performed about some arbitrary

point (see Figure 12). If that same planar form is ever "seen" again, its presence is detected by mathematically correlating the 4-D prototype with the previously defined 8-D image representation $W_{\ln f}$ for all linear shifts in a 6-D hyperspace. If the resulting 6-D correlation function exceeds threshold somewhere, then recognition is achieved. The location of the suprathreshold peak in the 6-D correlation space furthermore specifies the object position (in distance-normalized coordinates) within the object plane, the distance-normalized size, the orientation within object plane and finally, the slant and tilt of the object plane within which the planar form lies -- all relative to the fixation point, size, and orientation of the pattern (in frontoparallel view) from which the matching template was formed.

We have described an approach to invariant image pattern recognition which utilizes an 8-dimensional analogical image representation. This representation decouples common dimensions of variability in the image formation process to reveal particular 4-D canonical patterns that characterize arbitrary planar gray-scale forms invariant to imaging geometry and scene illumination. Canonical patterns that are embedded within the 8-D representation can be detected by mathematically correlating the 8-D representation with corresponding 4-D canonical patterns stored in visual memory. Furthermore, once a given canonical pattern has been identified, a number of geometric attributes of the corresponding planar objects in a scene are specified by the location of the suprathreshold peak in the corresponding correlation function.

Work is under way to investigate the performance and computational feasibility of the methods described in this paper. In particular, we will be using a discretely sampled approximation to the proposed 8-D representation. Though the amount of computation required is substantial, combinatorial explosion is not a problem since each dimension of the 8-D representation need only be encoded at a small sampling of discrete points. This follows from the fact that five of the eight dimensions correspond to finite angular axes, and the other three dimensions are logarithmic distance axes. Furthermore, referring to the computation tree for the 8-D representation (Figure 13), it should be apparent that computation of each 4-D function found at any leaf node of the tree can be carried out independent of all other such nodes. A leaf node

representation provides WALS-invariance to rotation and scaling about a single point within a single object plane. Therefore, a reasonable strategy would be to seek recognition of planar forms by sequentially shifting "attention" from one leaf node to the next. This would eliminate the need to compute the entire 8-D representation in parallel. It could also lead to efficient strategies for handling image frame sequences by exploiting the context of earlier frames to guide an attentional mechanism.

6. IMPACT ON STATE-OF-THE-ART

As seen in the previous sections, the Hierarchical Multisensor Image Understanding program has resulted in the development of various concepts and techniques which have the effect of extending the knowledge and capabilities of image understanding systems. The most noteworthy of these new concepts and techniques are:

- o Context Independent Scene Segmentation. This is a unique concept for segmenting a scene in such a way that, unlike conventional approaches, the procedure does not depend on specific objects on specific object or scene models. This context-free scene segmentation is of high significance because it removes the limitations of traditional scene segmentation.

- o Dynamic Spatio-Temporal Knowledge Representation and Inference. This aspect of the research investigates approaches for representing and processing knowledge on spatial constraints among different scene objects and their temporal changes. Specific results are contained in the Archival Scene Model (ASM), a highly compact data structure. ASM represents all the important scene information such as relationship between objects in terms of position and velocity, image-motion compensation, relative motion of objects with respect to the image window, and background description.

- o Hierarchical Planning for Control of Information Flow. Automated planning is the specification of a path through internally represented world states beginning at the current state and ending

at a goal world state. Honeywell's automated planning consists of obtaining or deducing basic informational steps, allocation of information processing resources, and the verification of the internal world states with the observed real world. Honeywell's concept of an automated planner includes provisions to deal with potential false information (inferred and/or acquired), parallel goal expansion which considers resource allocation as an additional dimension of planning, conflict resolution knowledge base, and replanning mechanisms.

However, besides the development of various new concepts and techniques, the program effort also opened many new doors in the image understanding area by demonstrating feasibility and potential of various additional concepts. These new areas of research we recommend further work fall into:

1. Spatio-temporal evidence accrual concepts and belief systems. The objective is to develop a goal-directed system to search spatially (across a scene) and temporally (through a sequence of scenes) for information that will minimize a measure of entropy in outdoor scenes (or maximize a measure of scene explanation).

2. Modeling and representation of scene invariant multisensor information via invariant matrices and normalization filters. This representation captures the essential information about the structural and perceptual invariants in the scene, regardless of relative size and orientation, we want to extend our work in understanding from scene synthesis to actually infer causal models that explain the dynamic and purpose of scene objects via AI-based inference engines.

The results of our efforts will provide a stepping stone for the future development of mature multisensor vision system control structures, a framework for the development of specific multisensor systems and an understanding of the commonalities within and between sensor domains. Furthermore, a functional model for synergistic use of complementary and redundant information in all levels of processes of a multisensor vision system across sensory domains will provide investigators with a tool to experiment and evaluate new concepts and paradigms in computer vision.

REFERENCES

1. D. Panda and R. Aggarwal, Image understanding Technology, January 1984.

2. T. Levitt and K. Schaper, Contect Free Region Discrimination (CRFD), Technical Report, Artificial Intelligence/Image Understanding Section, Honeywell Systems and Research Center, Minneapolis, MN, December 83.

3. T. Levitt, on Expert Systems for Sensor Data Understanding, Technical Report, Artificial Intelligence/Image Undersanding Section Honeywell.

4. A. Maren, Application of Artificial Intelligence to Image Understanding, Technical Report, Artificial Intelligence/Image Understanding Section, Honeywell Systems, November 83.

5. A. Maren, Multilevel, Multigoal Hierarchiacal Systems Theory: A Control Structure Paradigm for AI/Image Understanding Systems, Technical Report, Artificial Intelligence/Image Understanding Section, Honeywell System, and Research Center, Minneapolis, MN, April 84. To be presented at the SPIE Cambridge '84 Conference on Intelligent Robots and Computer Vision, Cambridge, MA, NOVEMBER 84.

6. J. Budenske, A. Hierarchical Generator for non-linear Research Allocation and Planning in Image Understanding Systems, Technical Report, Artificial Intelligence/Image Understanding, Honeywell Systems and Research Center, Minneapolis, MN, April 84. Presented at the Eight Annual Honeywell International Computer Science Conference, Minneapolis, MN, May 84.

7. R. Whillock, Using Production Rules for Control of Image Understanding Systems, Technical Report, Artificial Intelligence/Image Understanding, Honeywell Systems and Research Center, Minneapolis, MN, April 84. Presented at the Eight Annual Honeywell International Computer Science Conference, Minneapolis, MN, May 84.

8. N. Marquina, a Rule-based Evidence Accruel Paradigm for Image Understanding. To be presented at SPIE Cambridge '84 Conference on Intelligence Robots and Computer Vision, Cambridge, MA, Novenber 84.

9. P. Schenker, et al, three Generations of Image Understanding Architecture: Studies in Automatic Target Recognition System Design. To be presented at SPIE Cambridge '84 Conference on Intelligence Robots and Computer Vision, Cambridge, MA, Novenber 84.

10. F. Sadjadi, Recognition of Complex Three Dimensional Objects using 3-D Moment Invariants. To be presented at To be presented at SPIE Cambridge '84 Conference on Intelligence Robots and Computer Vision, Cambridge, MA, Novenber 84.

11. L. Jacobson and H. Wechsler, Invariant Analogical Image Representation and Pattern Recognition, Technical Report, Artificial Intelligence/Image Understanding Section, Honeywell Systems and Research Center, Minneapolis, MN, March 84. To be presented at SPIE Cambridge '84 Conference on Intelligence Robots and Computer Vision, Cambridge, MA, Novenber 84.

12. M. Eshera and K.S. Fu, A Similarity Measure between Attributed Relational Graphs for Image Analysis, IEEE Transactions in Pattern Analysis and Machine Intelligence, 1984.

13. M. Eshera and K.S. Fu, A Graph Distance Measure for Image Analysis, to appear in IEEE Transactions on Systems, Man, and Cybernetics, 1984.

14. A Barr and E. Feigenbaum, Handbook of Artificial Intelligence, Vol 1, William Kauffman, Inc., 1981.

15. E. Shortliffe and G. Buchanan, A Model of Inexact Reasoning in Medicine, Mathematical Bioscience, Vol. 23, 1975.

16. H. Barrow, and J. Tenenbaum, Computational Vision, Proc. of IEEE, 69 (5), May 1981, PP> 572-595.

17. R. Brooks, Model-based Three-dimensional Interpretation of Two-dimensional Images, IEEE Trans. on PAMI, PAMI-5, March 1983, pp. 140-150.

18. R. Brooks, "Symbolic Reasoning among 3-D models and 2-D images," Artificial Intelligence 17 (1981) 285-348.

19. M. Stefik, Planning with Constraints (MOLGEN: Part 1); and Planning and Meta-Planning (MOLGEN: Part 2); Artificial Intelligence, 14 (2) (1980), pp. 111-169.

20. F. Sadjadi and E. Hall, Three Dimensional Moment Invariants, IEEE Transactions in Pattern Analysis and Machine Intelligence, Vol. 2, 1980.

21. M. Ohlson, System Design Considerations for Graphics Input Devices, Computer, Vol. 11, 1978.

22. F. Sadjadi and E. Hall, Three Dimensional Sampling and Object Recognition, Proceedings of IEEE Region 3 Conference, Destin, Florida, 1982.

23. Movie-BYU Program Documentation, Brigham Young University.

24. D. Mara, Vision, San Francisco: W. H. Freeman and Company (1978).

25. A. V. Oppenheim, et al., Nonlinear Filtering of Multiplied and Convolved Signals, Proc. IEEE 56, 1264-1291 (1968).

26. R. Bamler, and H. Glunder, Coherent-optical generation of the Wigner distribution function of real-valued 2D signals, in Proc. 10th Int. Optical Computing Conf., Cambridge, Massachusetts, 117-121 (1983).

27. L. Jacobson, and H. Wechsler, The Wigner distribution as a tool for deriving an invariant representation of 2-D images, in Proc. IEEE Conf. Pattern Recognition and Image Processing, Las Vegas, Nevada, 218-220 (1982).

28. L. Jacobson, and H. Wechsler, The Wigner distribution and its usefulness for 2-D image procesing, in Proc. 6th. Int. Conf. Pattern Recognition, Munich, West Germany (1982).

29. L. Jacobson, and H. Wechsler, A paradigm for invariant object recognition of brightness, optical flow and binocular disparity images, Pattern Recognition Letters 1, 61-68 (1982).

30. L. Jacobson, and H. Wechsler, The composite pseudo-Wigner distribution, in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Boston, Massachusetts (1983).

31. D. Gabor, Theory of communication, J. IEE (London) 93, 429-457 (1946).

32. J. J. Kulikowski, et al., Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex, Biol. Cybernetics 43, 187-198 (1982).

33. S. Marcelja, Mathematical description of the responses of simple cortical cells, J. Opt. Soc. Am. 70, 1297-1300 (1980).

34. W. Wigner, On the quantum correction for thermodynamic equilibrium, Phys. Rev. 40, 749-759 (1932).

35. T. A. C. M. Claasen, and W. F. G. Mecklenbrauker, The Wigner distribution - a tool for time-frequency analysis. Parts I-II, Philips J. Res. 35, 217-250, 276-300, 372-389 (1980).

36. P. S. Schenker, et al, New sensor geometries for image processing: Computer vision in the polar exponential grid, in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 1144-1148, Atlanta, Georgia (1981).

37. E. L. Schwartz, Cortical anatomy, size invariance, and spatial frequency analysis, Perception 10, 455-468 (1981).

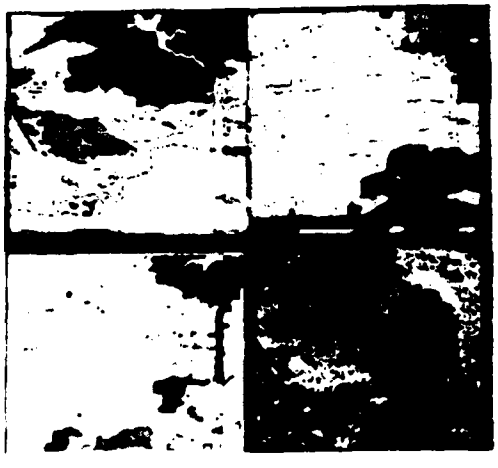
38. J. K. Brousil, and D. R. Smith, A threshold logic network for shape invariance, IEEE Trans. on Elec. Computers EC-16, 818-828 (1967).

39. D. Casasent, and D. Psaltis, Position, rotation and scale invariant optical correlation, Appl. Optics 15, 1795-1799 (1975).

40. P. Cavanagh, Size and position invariance in the visual field, Perception 7, 167-177 (1978).

Figure 1. Approach to a Region Discriminated Image

A. Original FLIR Image of Part of a Power Plant



B. Results of Four Different Operators for Combining Primitive Region Discrimination Cues into a Labeled Image

C. Discriminated Regions Superimposed on the Original Image. This is the Result of Conflict Resolution on the Four Labeled Images of B

Figure 2. Context Free Region Discrimination

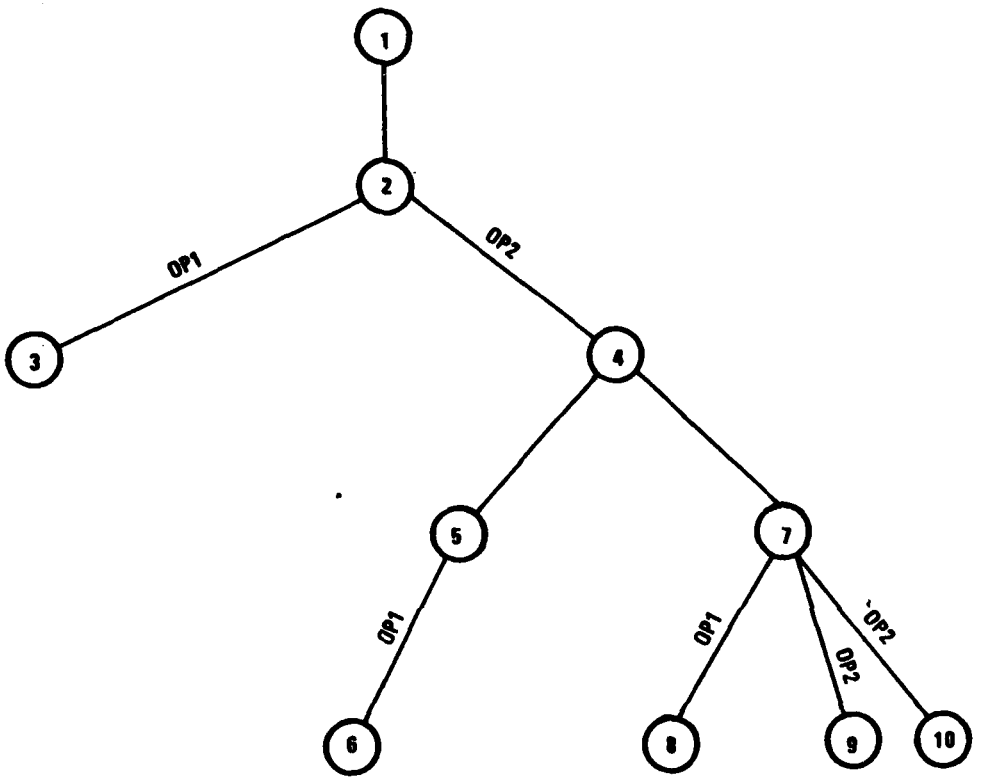


Figure 3. Segmentation Tree for Power Plant Scene.

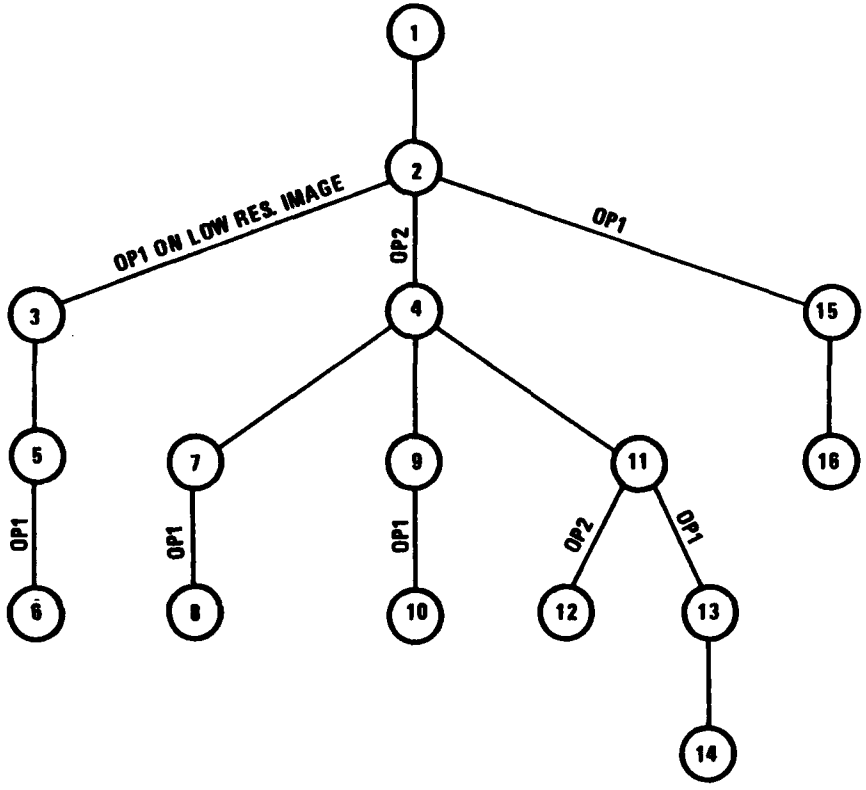


Figure 4. Segmentation Tree for Highway Scene.

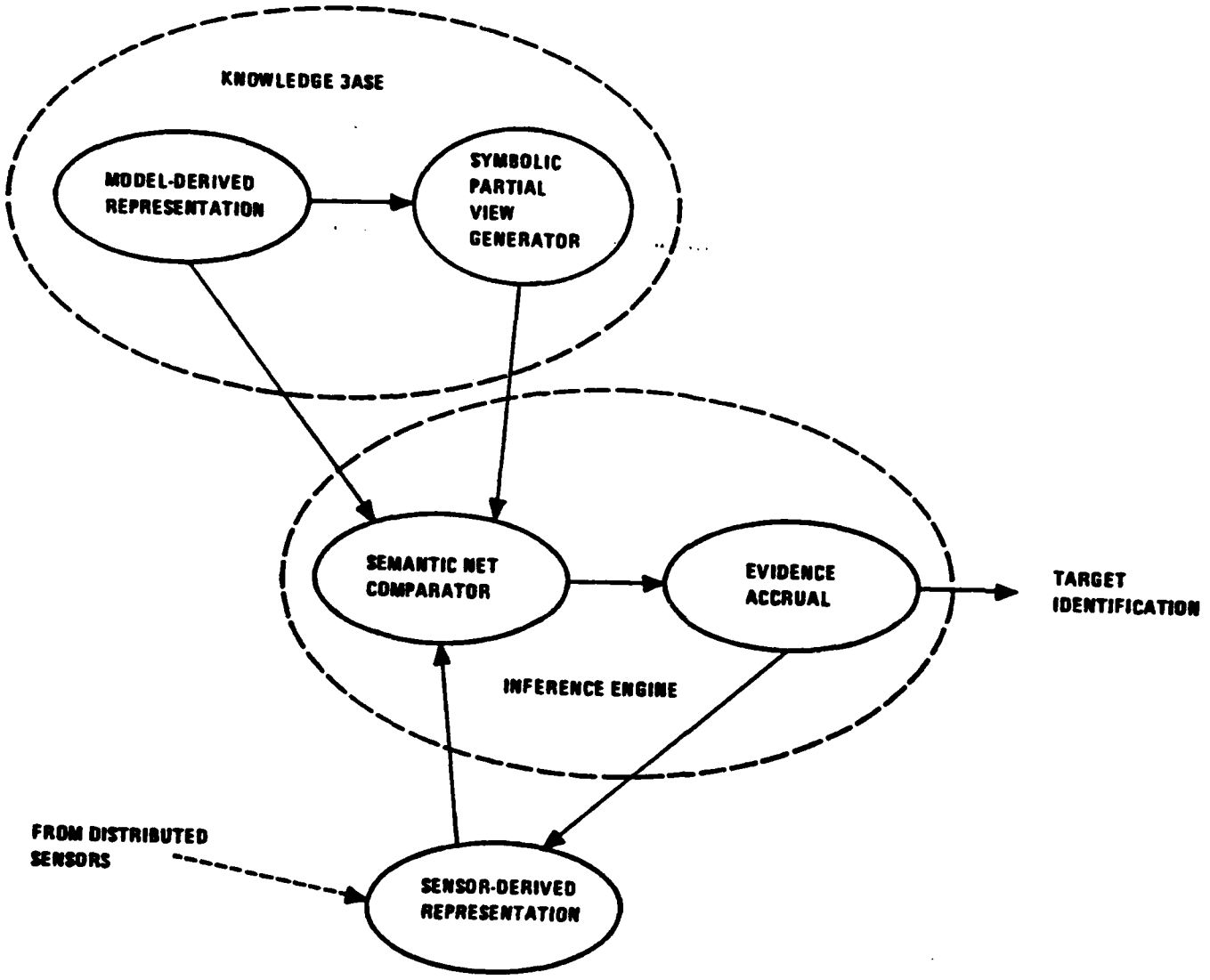


FIGURE 5 SYMBOLIC INFORMATION FUSION

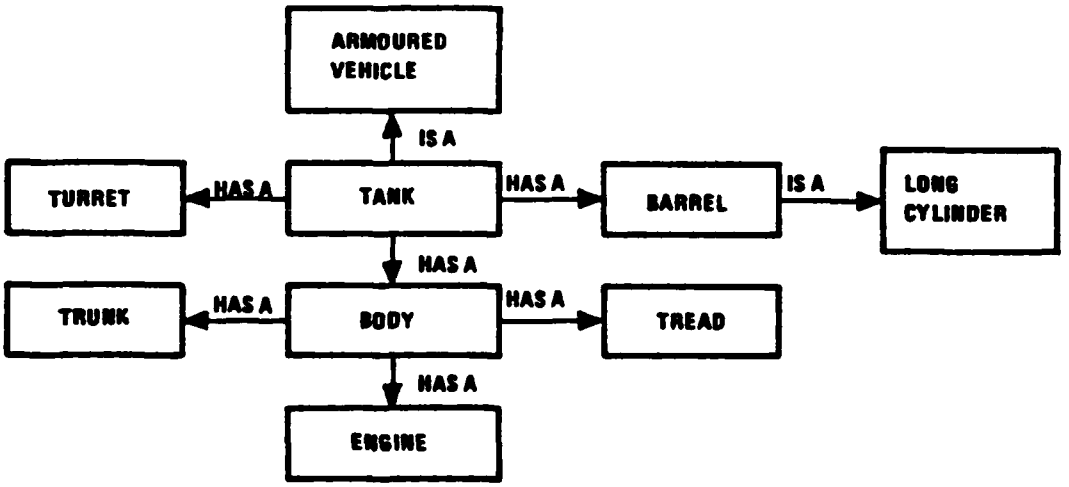


FIGURE 6 REPRESENTATION OF A TANK AS A SEMANTIC NETWORK

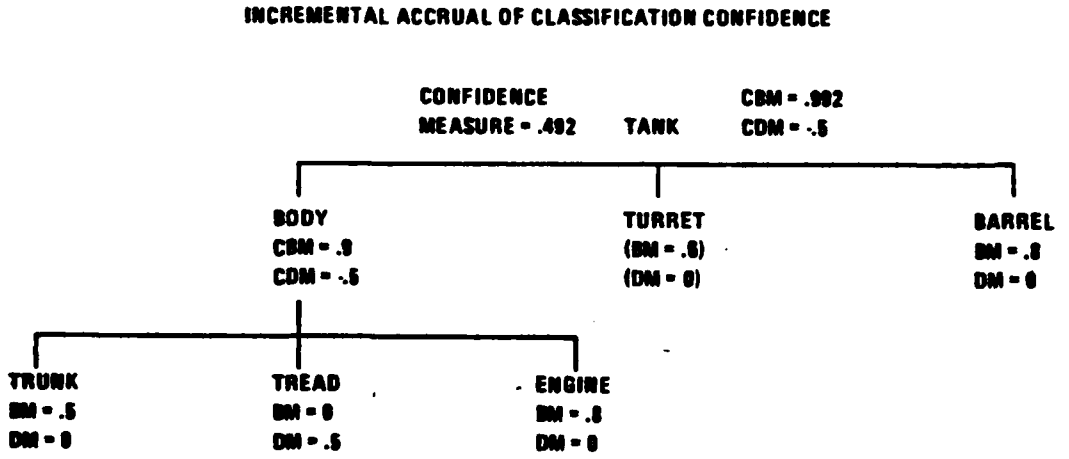


FIGURE 7 INFERENCE DRIVEN RECURSIVE REFINEMENT

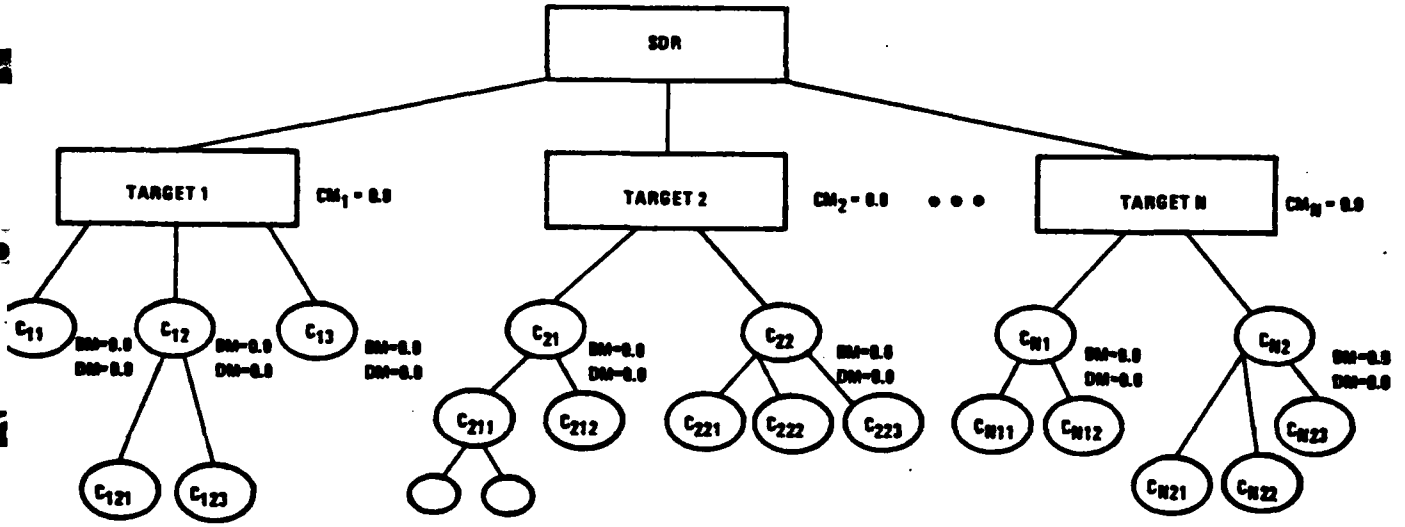


FIGURE 8 INITIAL STATE OF SENSOR-DERIVED REPRESENTATION

Self-organization
(Strategic)

Learning and Adaption
(Tactical)

Selection
(Operational)

Process

Figure 9. Layers of Decision Complexity[18]

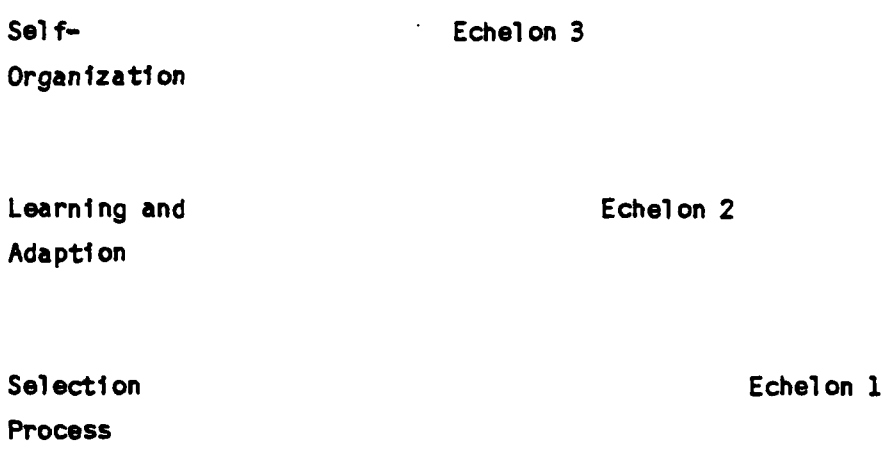


Figure 10. Multilayer hierarchies within decision units of a multiechelon systems. (18)

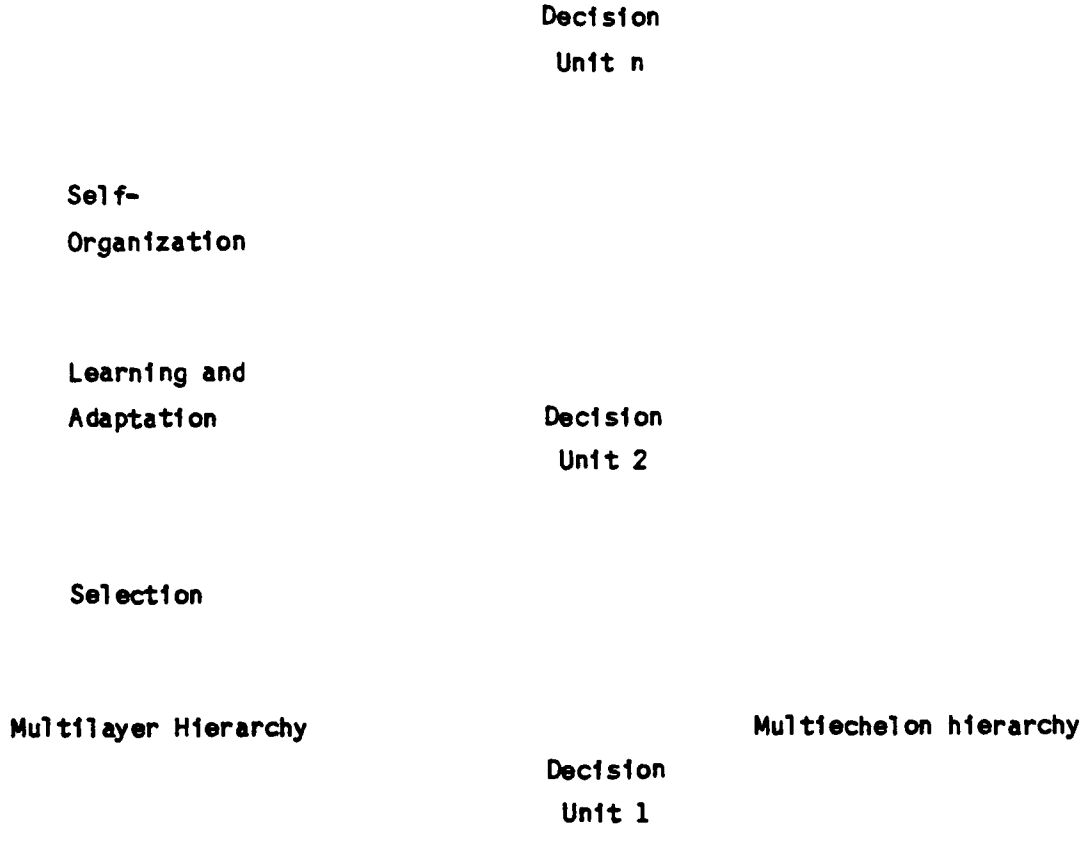


Figure 11. Decision units of multilayer hierarchy presented as multilayer or multiechelon hierarchies. (18)

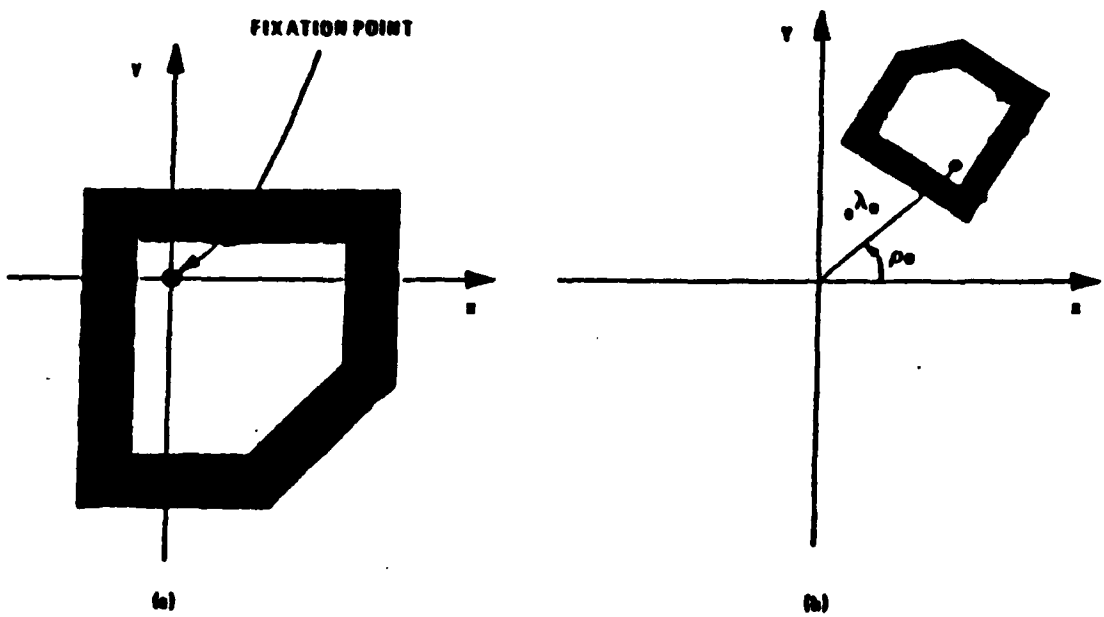
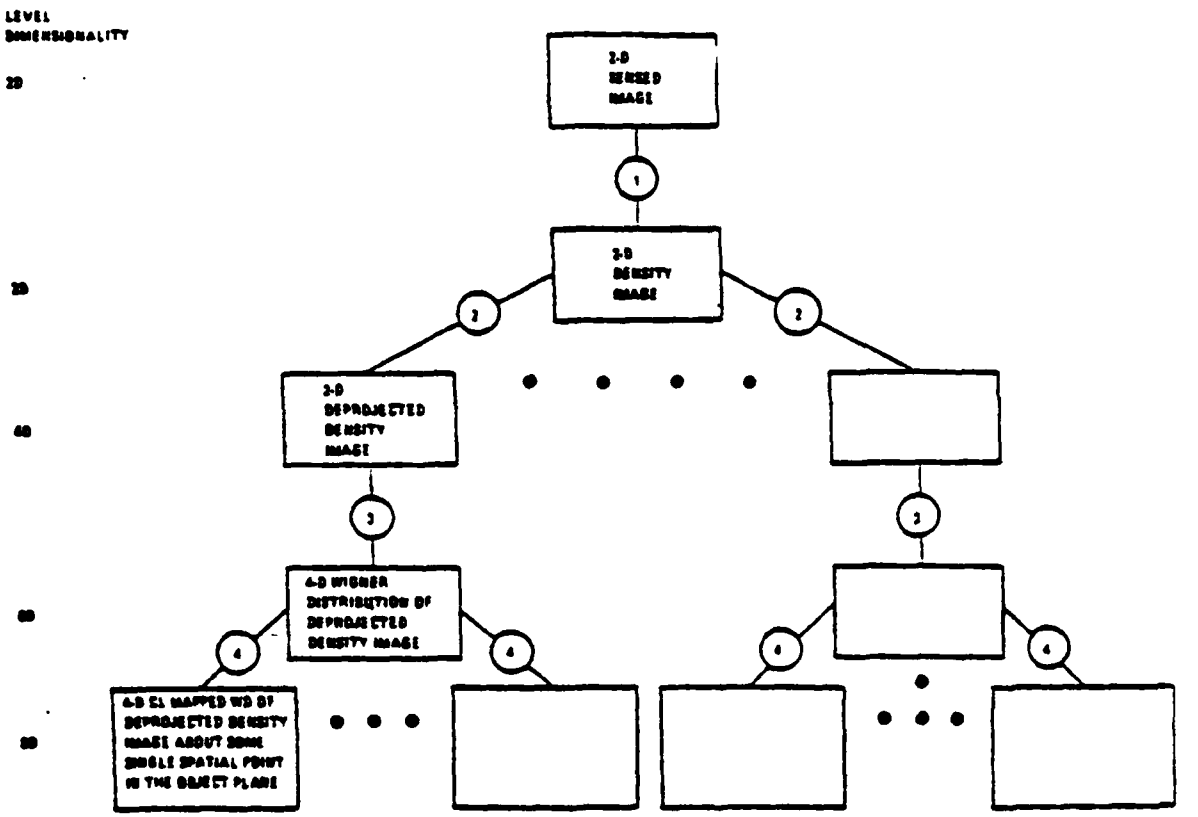


Figure 12. (a) The "fixation point" about which CL mapping is performed when deriving the template; (b) The identified form whose distance-normalized position in its object plane (σ_o, τ_o, ζ) is specified in polar coordinates by ($e^{\lambda_o}\cos\rho_o$, $e^{\lambda_o}\sin\rho_o$), and whose distance-normalized size and orientation, relative to the template form, is denoted by K_o and Θ_o, respectively.



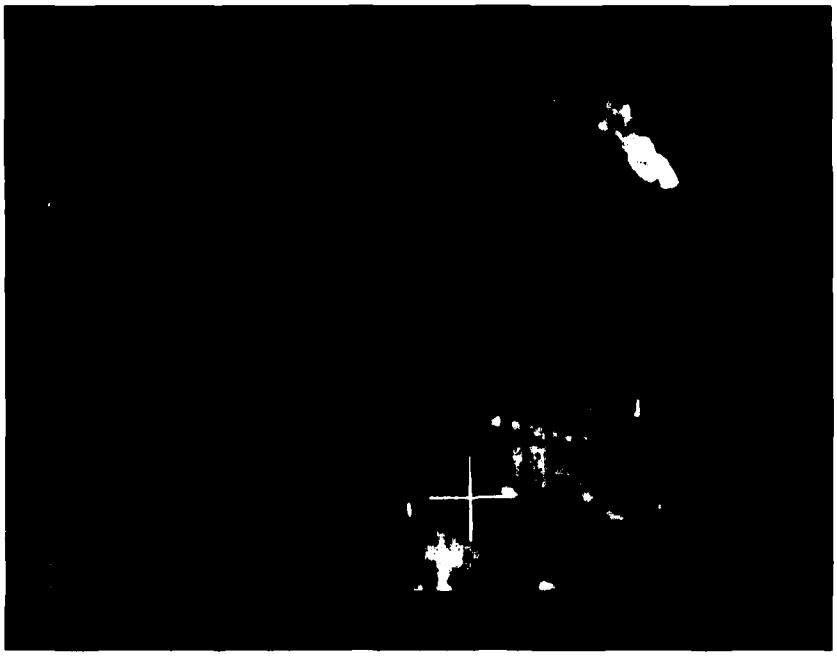
STEP	TYPE OF TRANSFORMATION	NAME OF TRANSFORMATION	EFFECT OF TRANSFORMATION
1	FUNCTIONAL	LOGARITHM	CHANGES PRODUCT OF SIGNAL COMPONENTS INTO A SUM
2	GEOMETRIC	DEPROJECTION	PROVIDE INVARIANCE TO PLANAR PERSPECTIVE
3	FUNCTIONAL	WIGNER DISTRIBUTION	PROVIDE SEPARATION OF MULTIPLICATIVE SIGNAL COMPONENTS PROVIDE A LOCAL 2D TEXTURE REPRESENTATION
4	GEOMETRIC	CI CONFORMAL MAPPING	PROVIDE ROTATION, SCALE, AND POSITION INVARIANCE WITHIN EACH OBJECT PLANE

Figure 13. The computation tree for the 8-dimensional representation $W_{Inf}(\zeta',\theta',w',\xi';\zeta,\theta,\sigma,\tau)$.

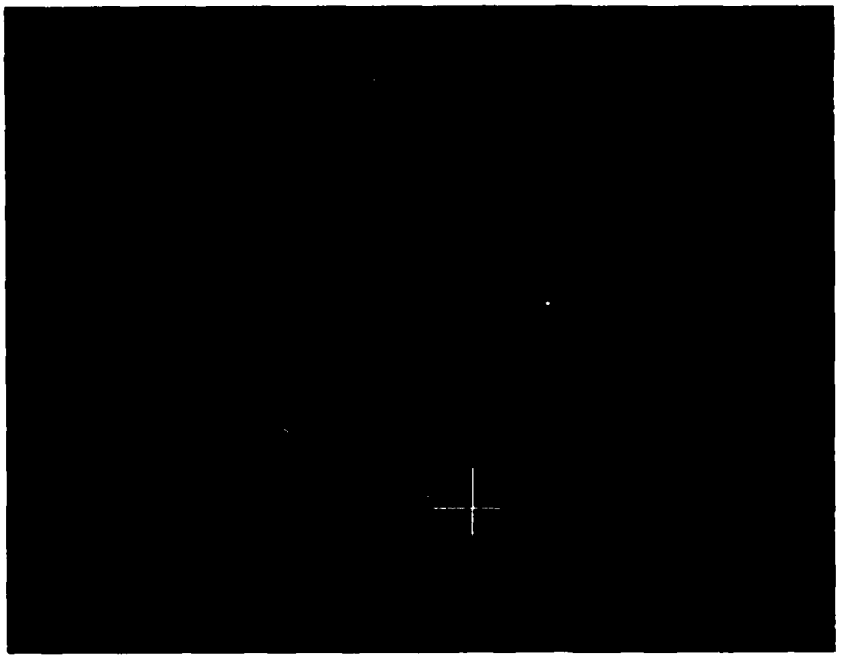
	TARGET 1		TARGET 2	
STEP 1	$CM_1 = 0.2$		$CM_2 = 0.3$	
	$BM_{11} = 0.3$	$DM_{11} = -0.1$	$BM_{21} = 0.8$	$DM_{21} = -0.1$
	$BM_{12} = 0.6$	$DM_{12} = -0.0$	$BM_{22} = 0.2$	$DM_{22} = -0.0$
STEP 2	$CM_1 = 0.6$		$CM_2 = 0.25$	
	$BM_{11} = 0.5$	$DM_{11} = -0.0$	$BM_{21} = 0.6$	$DM_{21} = -0.3$
	$BM_{12} = 0.7$	$DM_{12} = -0.0$	$BM_{22} = 0.2$	$DM_{22} = -0.1$
STEP 3				

TABLE 1 EXAMPLE OF EVIDENCE ACCRUAL

Picture 1. Original FLIR image of power plant scene.

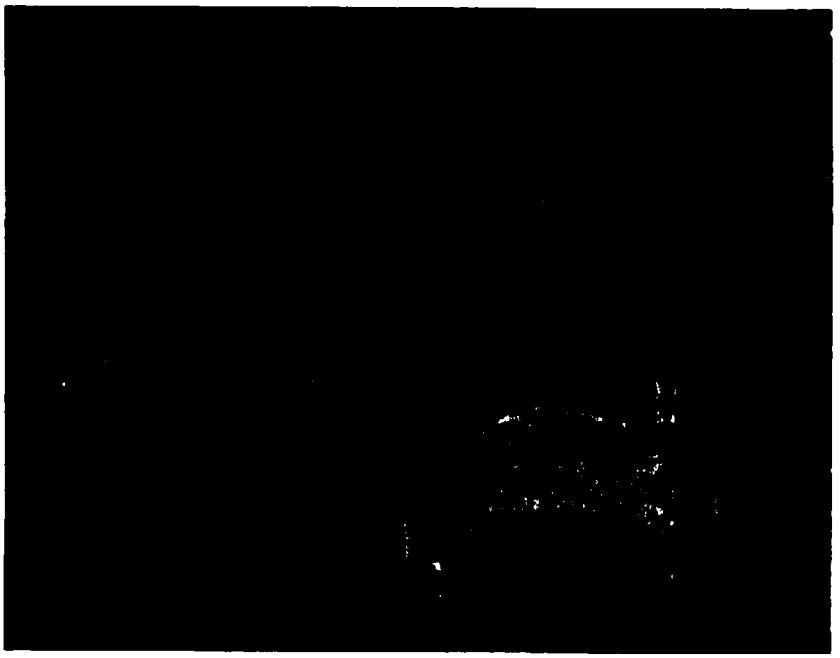


Picture 2. Power plant scene after noise cleaning.

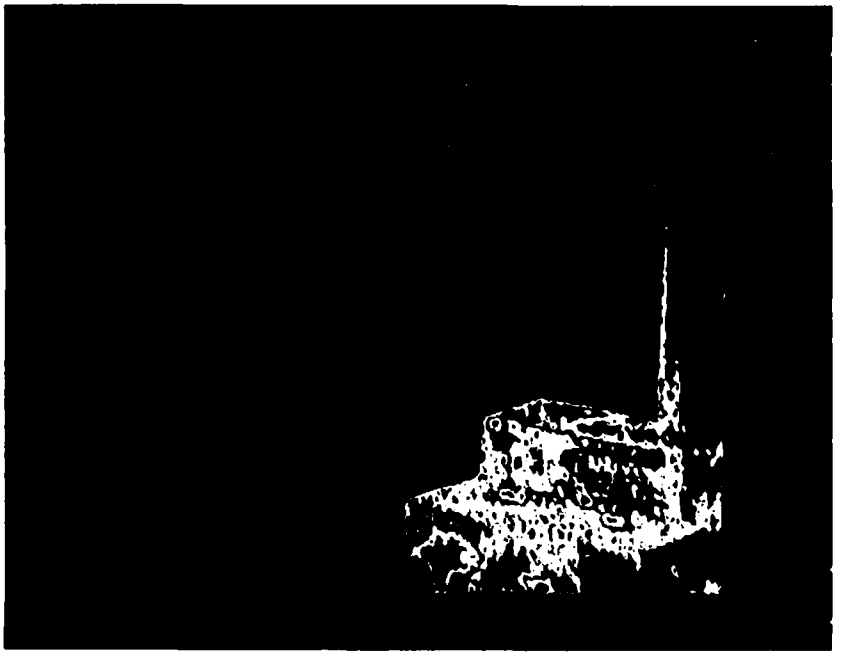


Picture 3. Operator 1 on full image of power plant scene.

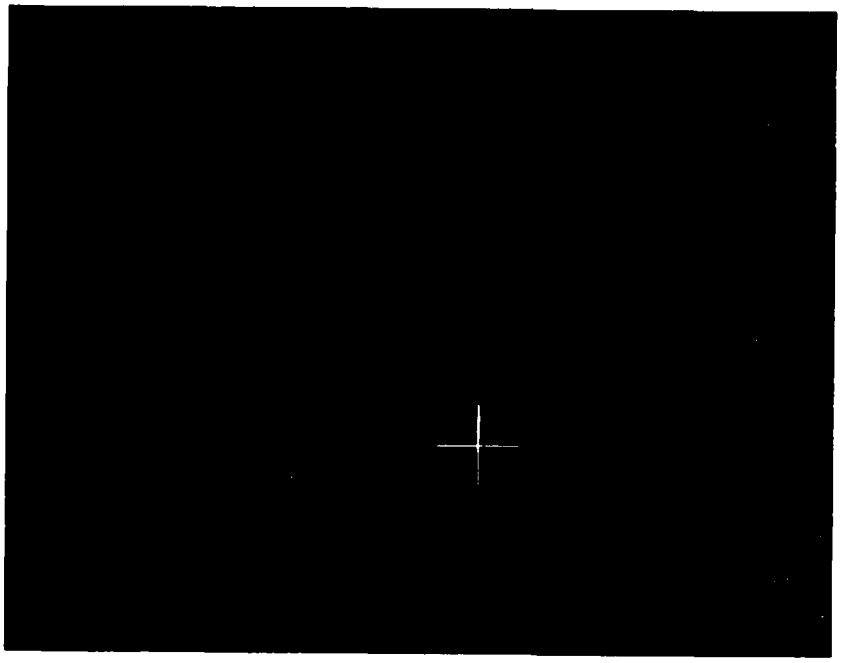
Picture 4. Operator 2 on full image of power plant scene.



Picture 5. Region from Picture 4.

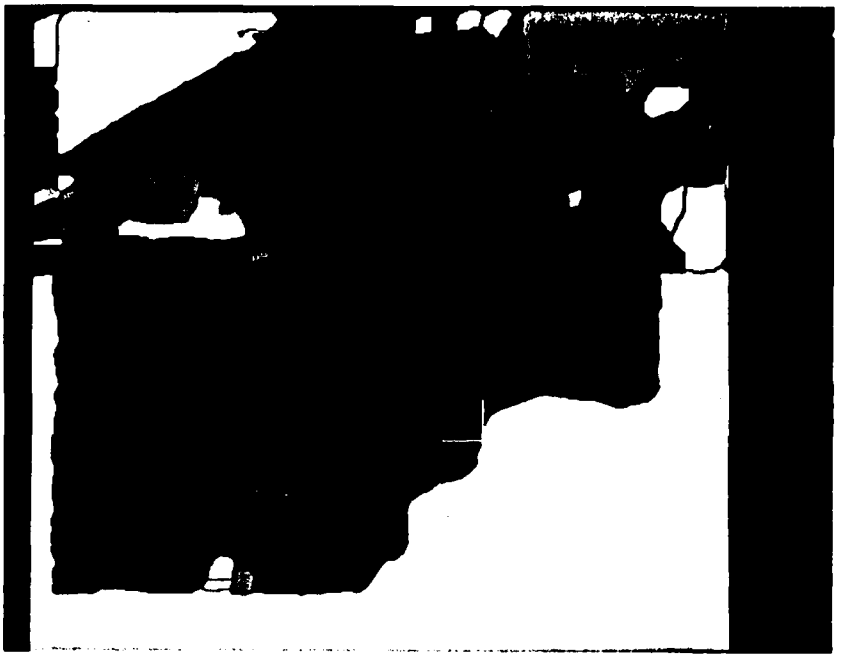


Picture 6. Operator 1 on same region as 5.

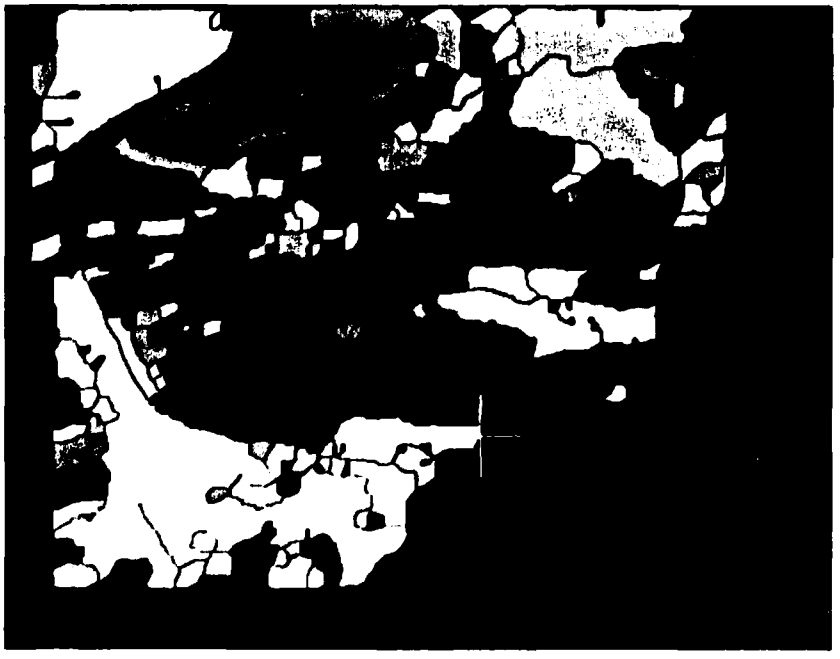


Picture 7. Another region from Picture 4.

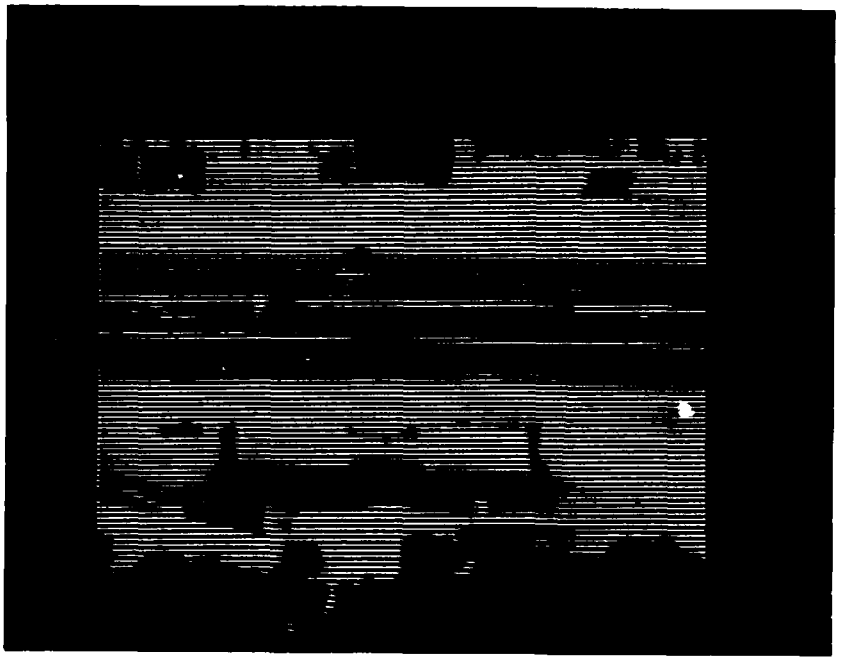
Picture 8. Operator 1 on same region as 7.



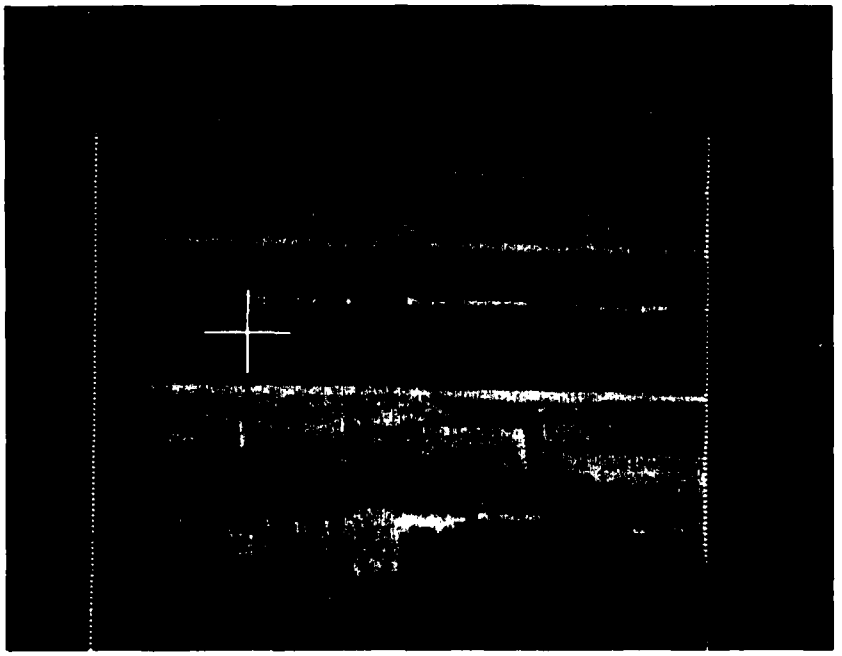
Picture 9. Operator 2 on region in Picture 7.



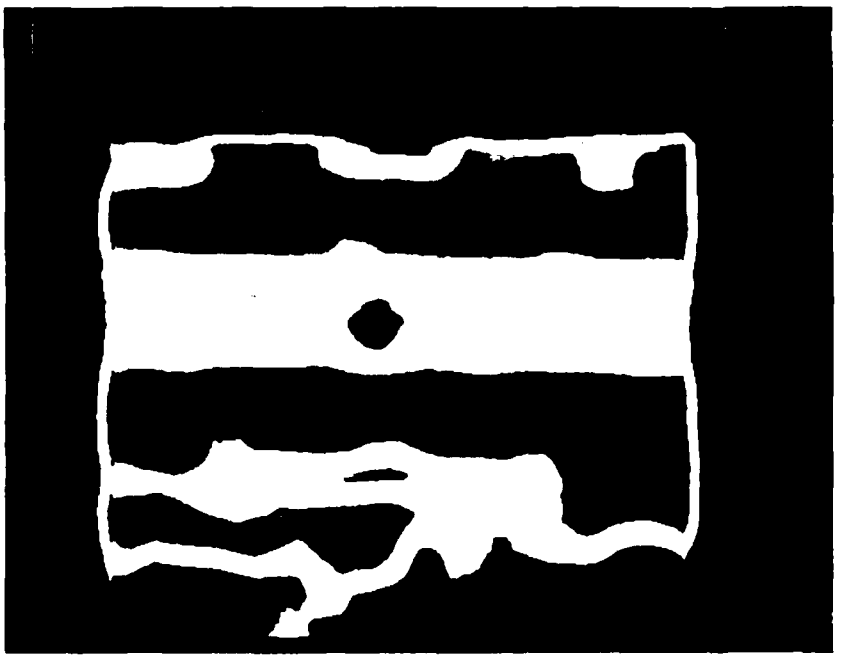
Picture 10. Operator 2 on region in Picture 7 with higher edge threshold.



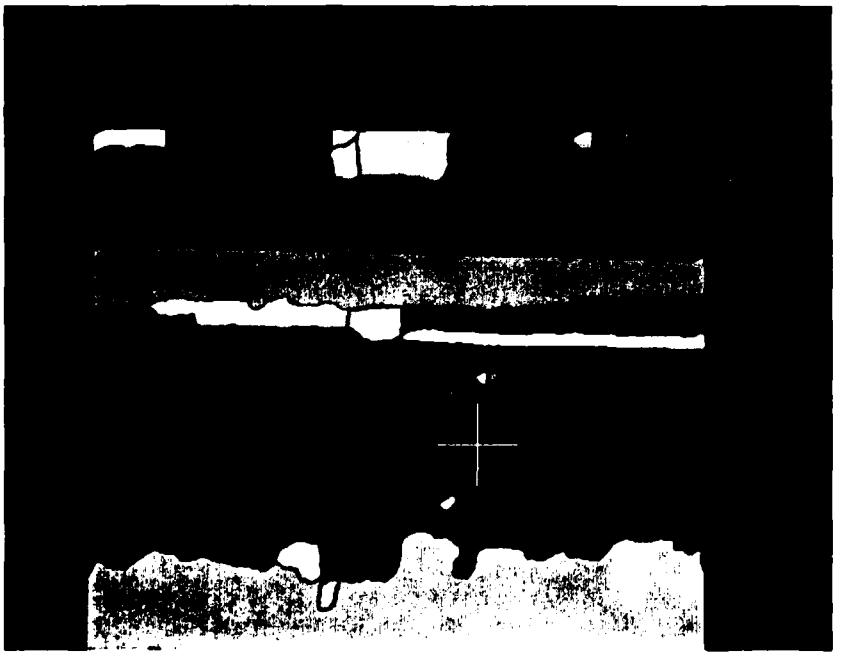
Picture 11. Original FLIR image of highway scene.



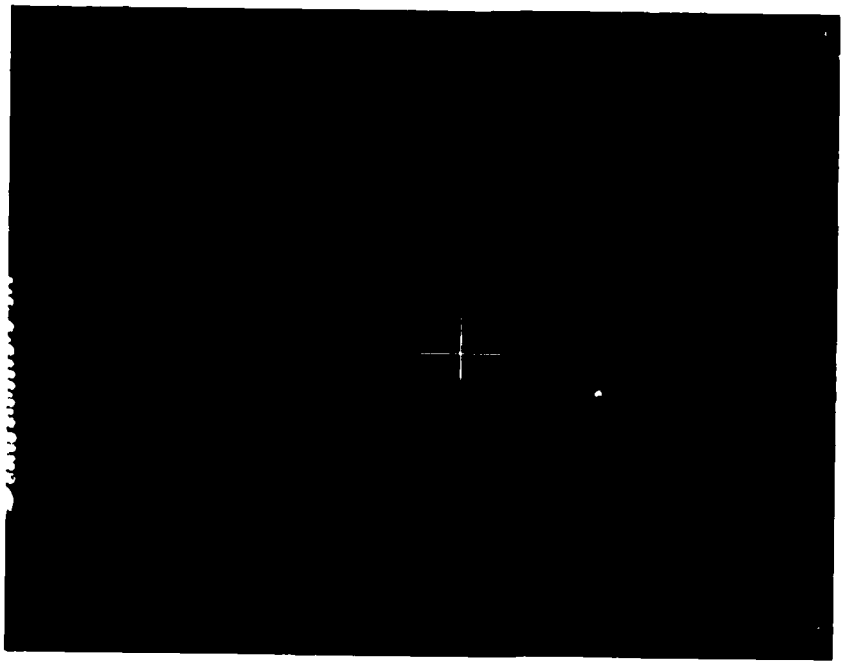
Picture 12. Highway scene after noise cleaning.



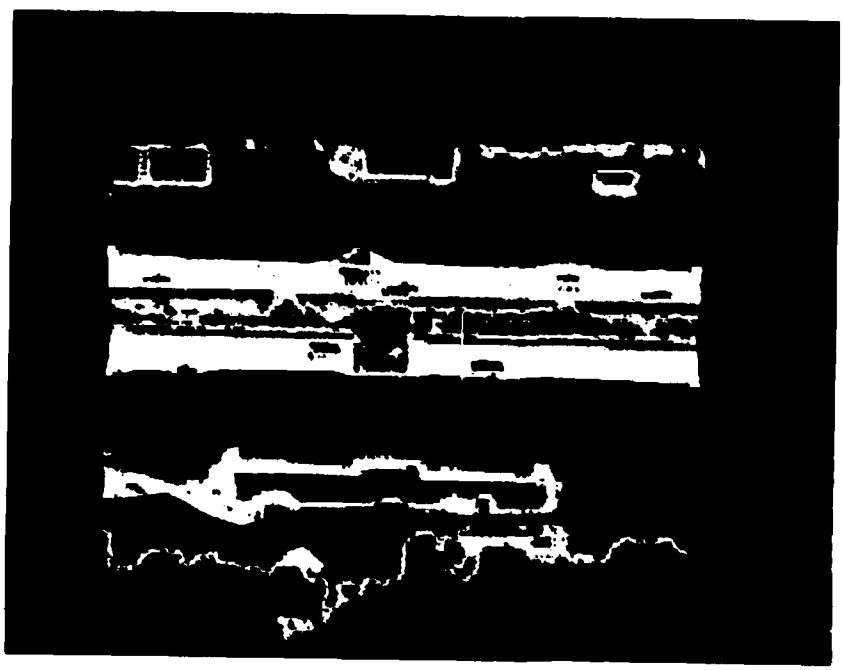
Picture 13. Operator 1 on averaged image of highway scene.



Picture 14. Operator 2 on full image of highway scene.

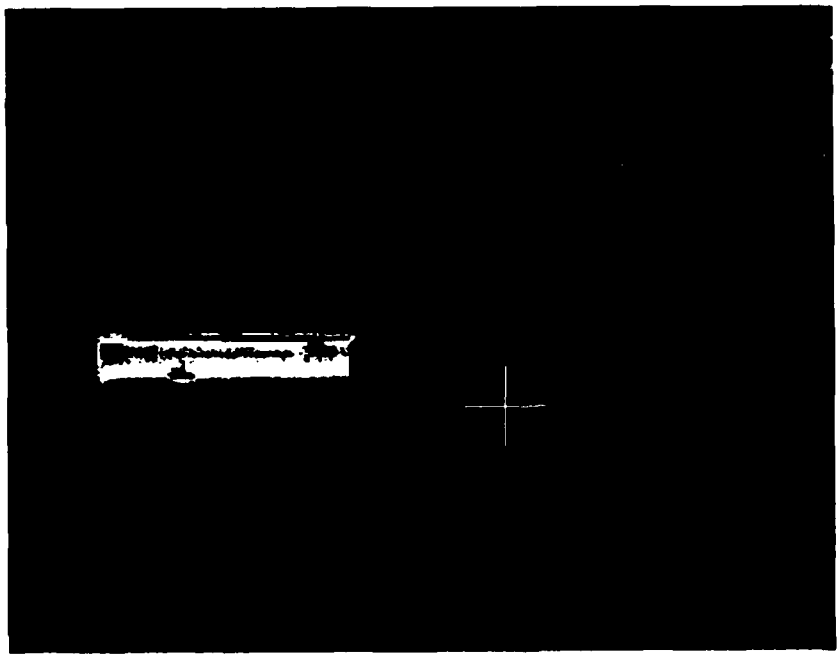


Picture 15. A region from Picture 13.

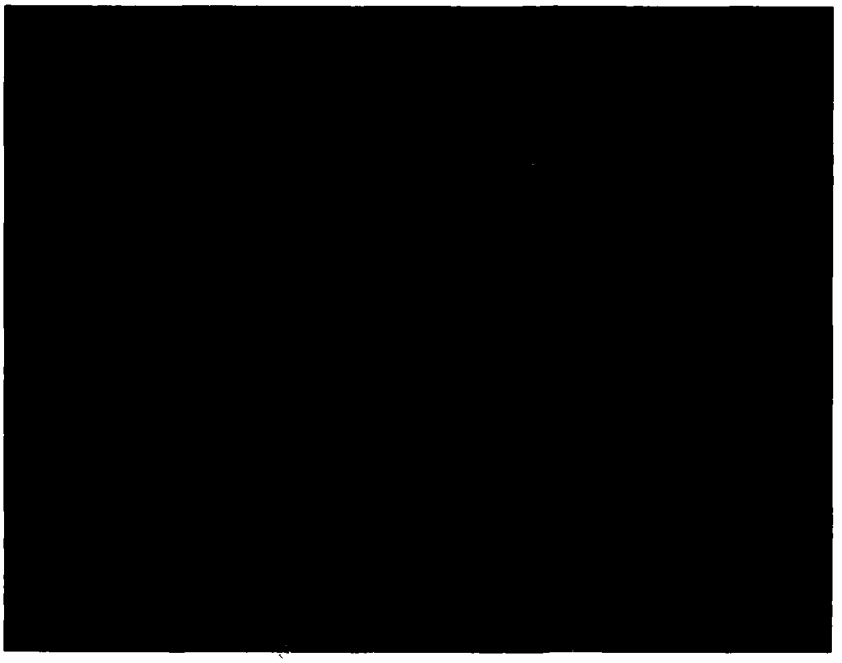


Picture 16. Operator 1 on same region as 15.

Picture 17. A region from Picture 14.

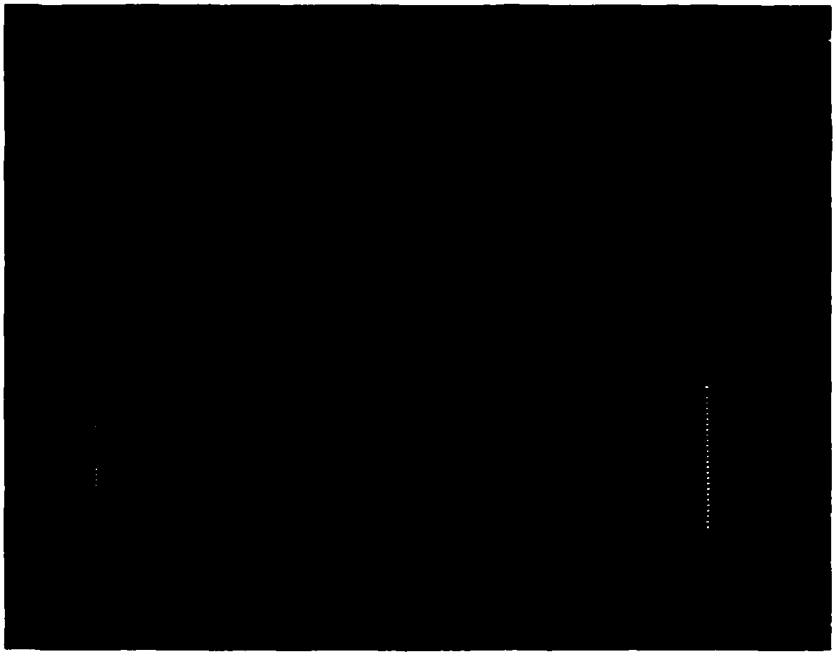


Picture 18. Operator 1 on same region as 17.

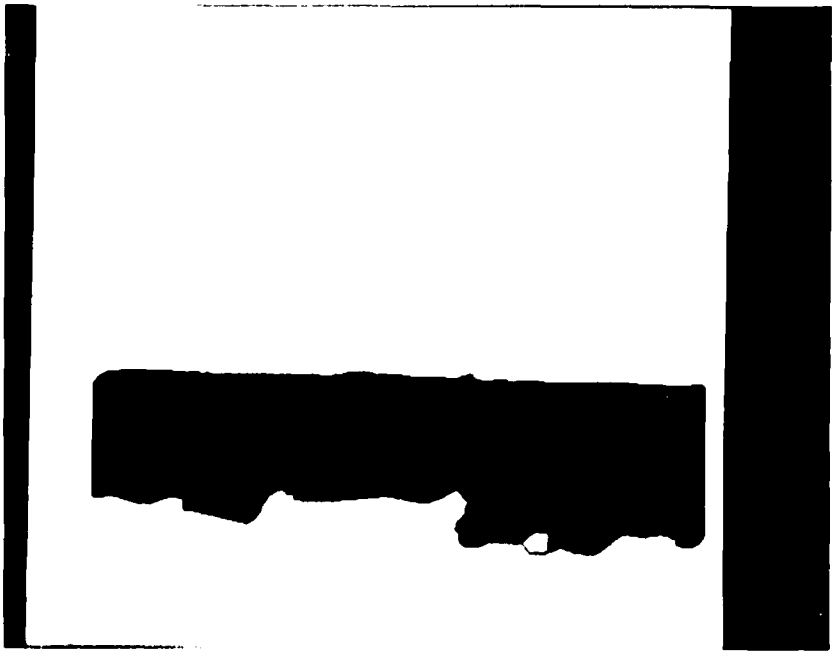


Picture 19. Another region from Picture 14.

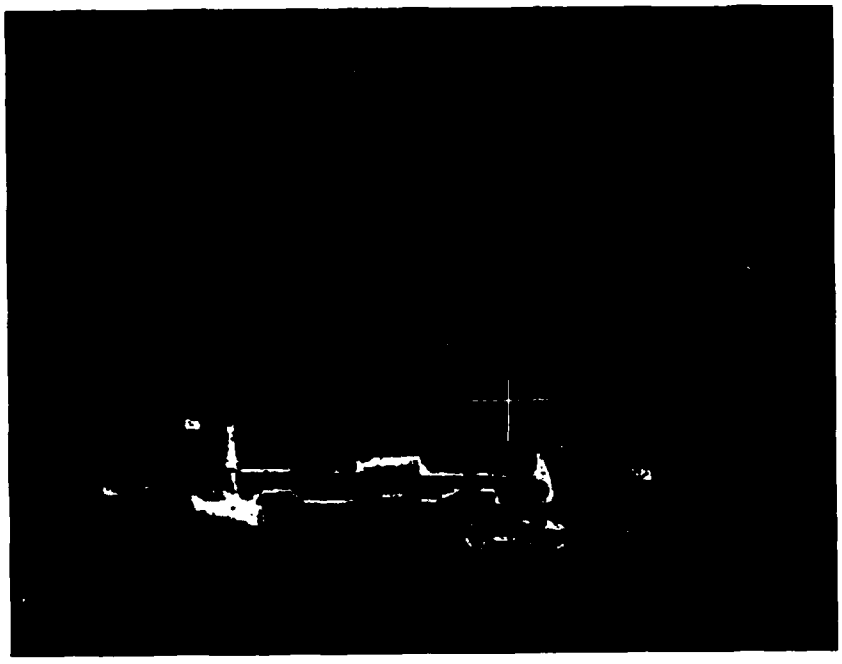
Picture 20. Operator 1 on same region as 19.



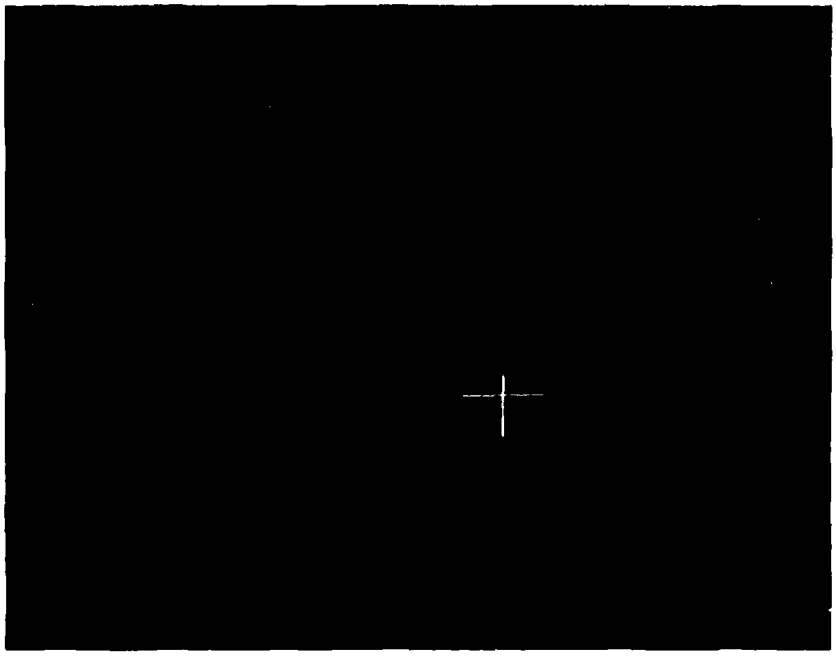
Picture 21. Another region from Picture 14.



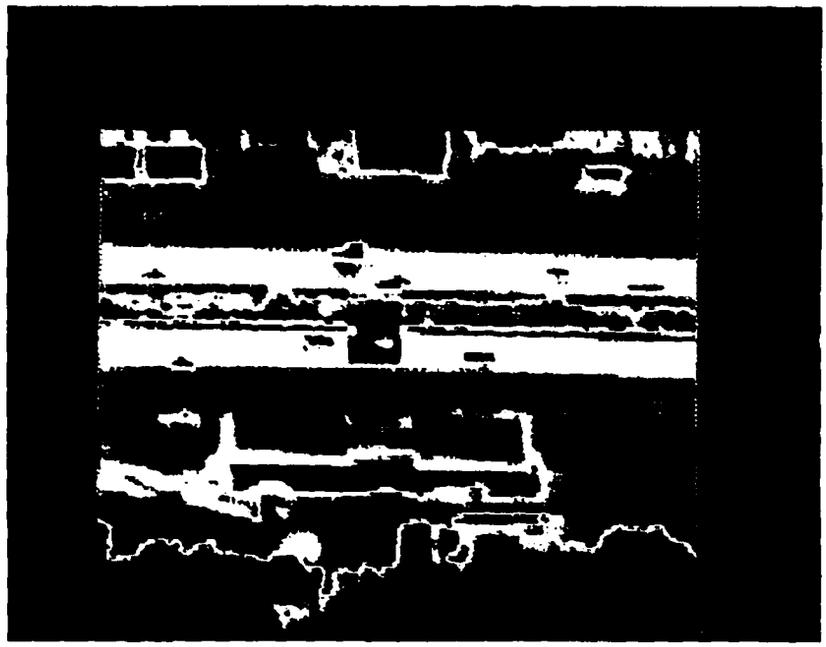
Picture 22. Operator 2 on same region as 21.



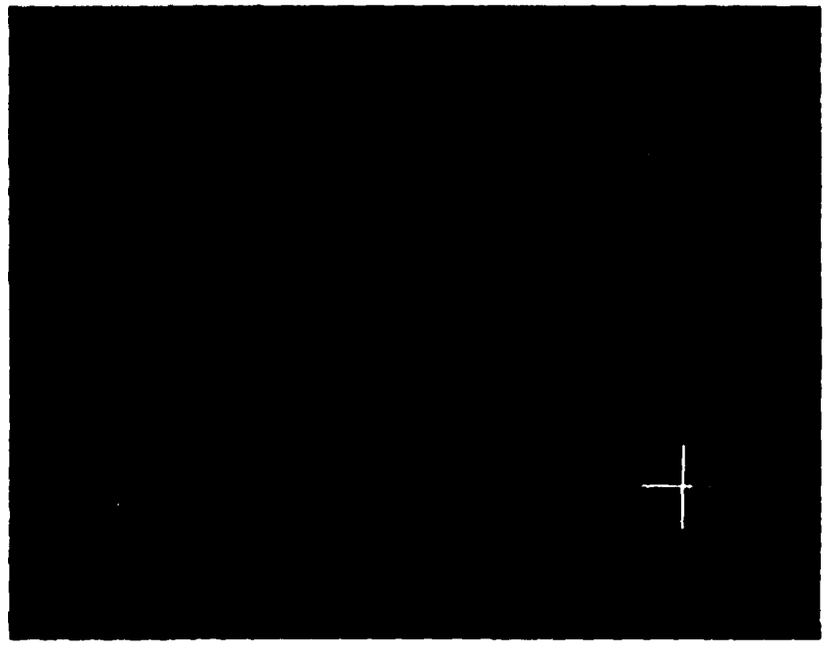
Picture 23. Operator 1 on region from Picture #21.



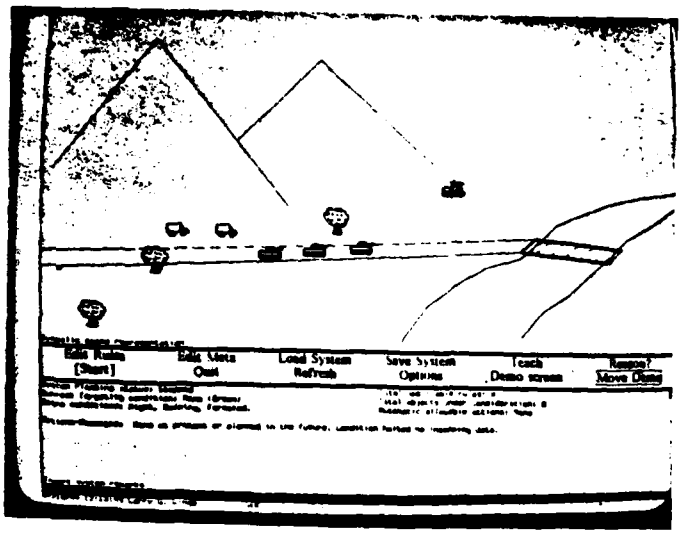
Picture 24. Region from same segmentation as 23.



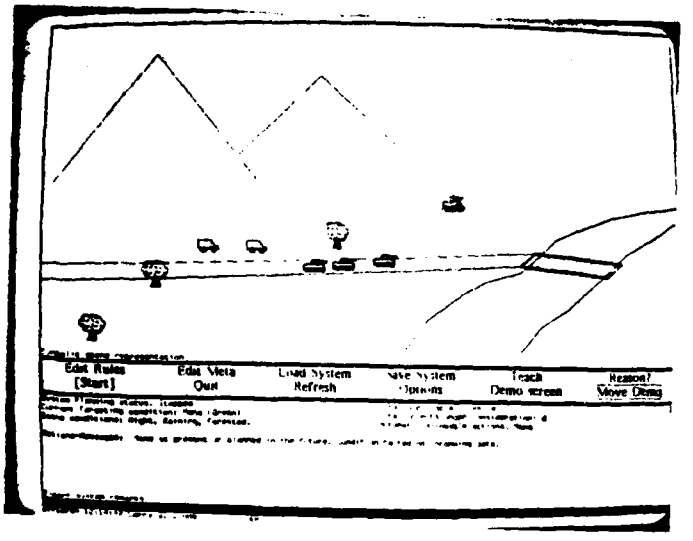
Picture 25. Operator 1 on full image of highway scene.



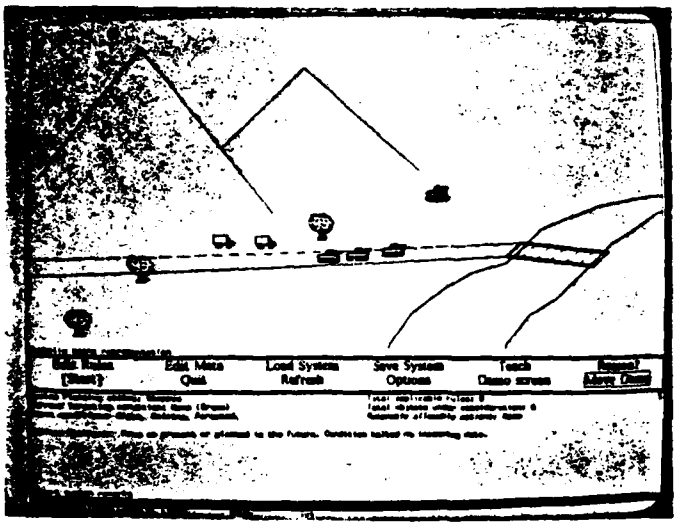
Picture 26. A region from Picture 25.



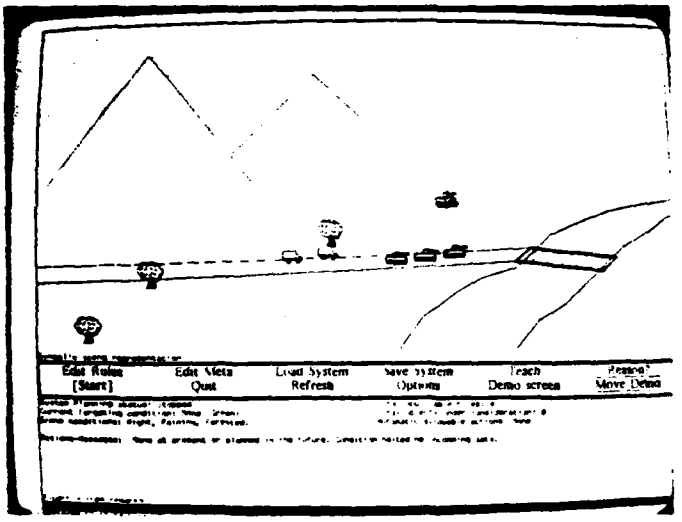
Picture 27. Outdoor Scene with Trucks and Tanks: Frame 1.



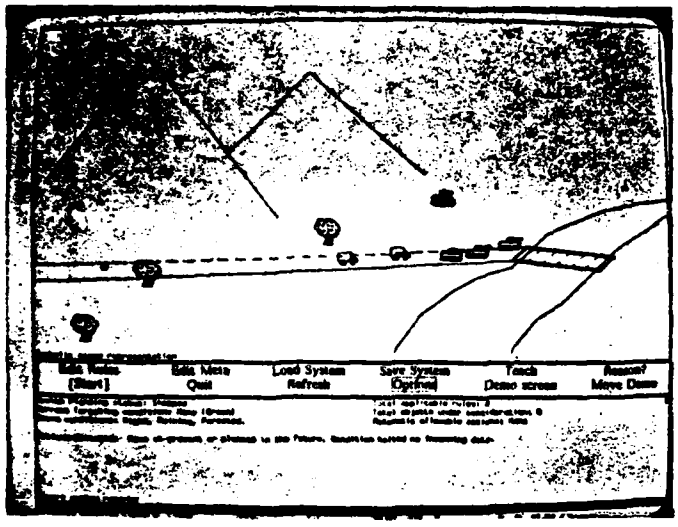
Picture 28. Outdoor Scene with Trucks and Tanks: Frame 2.



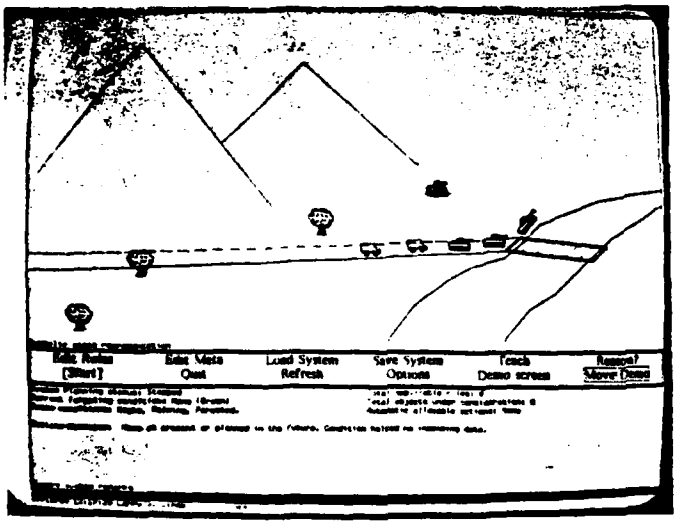
Picture 29. Outdoor Scene with Trucks and Tanks: Frame 3.



Picture 30. Outdoor Scene with Trucks and Tanks: Frame 4.



Picture 31. Outdoor Scene with Trucks and Tanks: Frame 5.



Picture 32. Outdoor Scene with Trucks and Tanks: Frame 6.

END

FILMED

DTIC