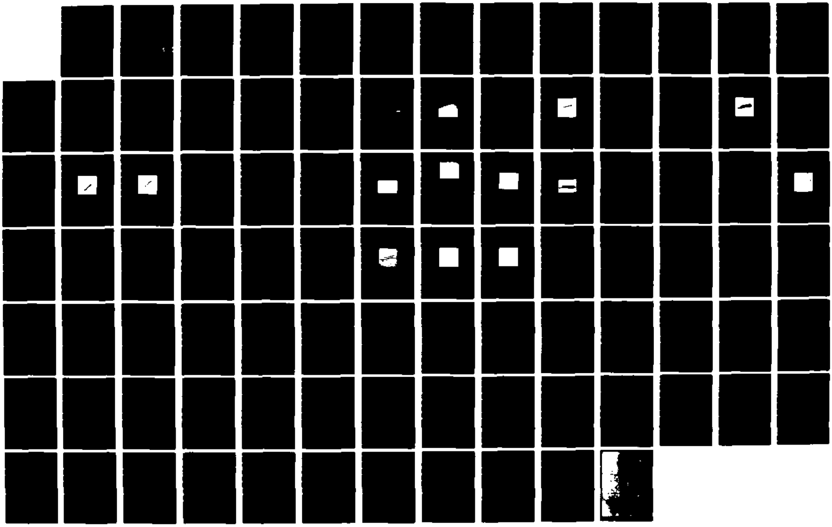
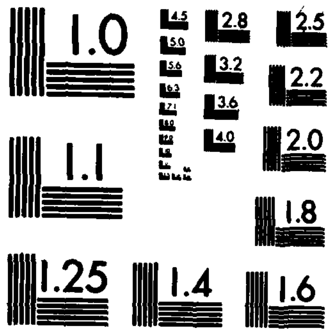
AD-A127 710 INTELLIGENT INTERPRETATION OF 3-D IMAGERY(U) ADVANCED 1/1
INFORMATION AND DECISION SYSTEMS MOUNTAIN VIEW CA
D KUAN ET AL. FEB 83 AI/DS-1027-1 AFOSR-TR-83-0331

UNCLASSIFIED F49620-82-C-0071 F/G 20/6 NL





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

AI&DS

AI&DS-1027-1 February, 1983

INTELLIGENT INTERPRETATION OF 3-D IMAGERY

Darwin T. Kuan and Robert J. Drazovich

Final Technical Report

1 April 1982 - 31 December 1982

Prepared for

Directorate of Mathematical & Information Sciences
Air Force Office of Scientific Research
Bolling Air Force Base, DC 20332

U.S. Government Contract No. F49620-82-C-0071

Approved for Public Release; Distribution Unlimited

ADVANCED INFORMATION & DECISION SYSTEMS
Mountain View, CA 94040

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0331	2. GOVT ACCESSION NO. AD-A127 710	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) INTELLIGENT INTERPRETATION OF 3-D IMAGERY		5. TYPE OF REPORT & PERIOD COVERED FINAL, 1 APR 82-31 DEC 82
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Darwin Kuan and Mr. Robert Drazovich		8. CONTRACT OR GRANT NUMBER(s) F49620-82-C-0071
9. PERFORMING ORGANIZATION NAME AND ADDRESS Advanced Information & Decision Systems 201 San Antonio Circle, Suite 286 Mountain View CA 94040		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS Mathematical & Information Sciences Directorate Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE Feb 83
		13. NUMBER OF PAGES 87
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION ST... MENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY ...TES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Model-based interpretation; computer vision; artificial intelligence; target classification; geometric modeling; model prediction; range imagery; knowledge data base; image feature extraction; edge detect...; symbolic reasoning.

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

In this research effort, techniques for 3-D object classification were developed for classifying vehicle targets in air-to-ground laser range imagery. An artificial intelligence approach was emphasized to intelligently interpret laser imagery in terms of 3-D symbolic models. The full classification system included 3-D image feature extraction, geometric modeling, model prediction, and feature-to-model matching. The approach used is to first extract low level image features with the guidance of model predictions. The object model was represented by a single viewpoint-independent 3-D model. The object (CONTINUED)

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENTS

The authors wish to thank several individuals who have contributed to the research efforts undertaken on this contract and to the preparation of this document. Dr. Thomas Binford of Stanford University made valuable contributions and provided guidance in the development of a model-based system. Dr. Ramakant Nevatia of the University of Southern California provided valuable suggestions in the development of 3-D image feature extraction techniques. Several other members of AI&DS's staff were involved. Dick Wishner provided project guidance and direction. Programming assistance was provided by Hans Muller, and Clif McCormick provided critical comment on this final report. Finally, we also would like to thank The Analytic Science Corporation of Reading, Massachusetts for providing synthetic range imagery.

CONTENTS

	Page
1. INTRODUCTION	1
1.1 RESEARCH OVERVIEW	1
1.2 SUMMARY AND CONCLUSIONS	4
1.3 ORGANIZATION OF THIS REPORT	8
2. THREE-DIMENSIONAL IMAGE FEATURE EXTRACTION	10
2.1 COORDINATE TRANSFORMATION	13
2.2 OBJECT-GROUND SEGMENTATION	18
2.3 PROJECTION IMAGES GENERATION	21
2.4 3-D PHYSICAL EDGE DETECTION	32
2.4.1 Occluding Edge Detection	33
2.4.2 Concave and Convex Edge Detection	34
2.4.3 Linear Feature Extraction	41
3. MODEL-BASED INTERPRETATION	47
3.1 OVERVIEW OF 3-D FEATURE-TO-MODEL MATCHING	49
3.2 3-D OBJECT MODEL REPRESENTATION	52
3.3 PREDICTION	58
3.3.1 Context Level Prediction	61
3.3.2 Object Level Prediction	61
3.3.3 Cylinder-level Prediction	63
3.3.4 Surface-level Prediction	65
3.3.5 Edge Level Prediction	65
3.4 INTERPRETATION	66
3.5 PROCESSING EXAMPLES	70
4. SUMMARY AND CONCLUSIONS	82
5. REFERENCES	86

ILLUSTRATIONS

	Page
1-1: Major Components of 3-D Object Classification System	7
2-1: The 3-D Image Feature Extraction Process	12
2-2: Coordinate Transformation Between a Sensor-Centered Coordinate System and the World Coordinate System	14
2-3(a): Range Imagery of a Decoy	15
2-3(b): Drawing of a Decoy	16
2-3(c): Z Coordinate of Figure 2-3(a)	17
2-4: Extracted Object Segment of Decoy	20
2-5: Ground Projection Points of Decoy	23
2-6: Refined Ground Projection of Decoy	24
2-7: Bounding Rectangles for Orientation Estimation	26
2-8: Side-View Projection of Decoy	28
2-9: Range Image of a Missile Launcher	29
2-10: Side-View Projection Points	30
2-11: Side-View Projection Image of Decoy	31
2-12: Exterior Boundary Obtained by Applying the Boundary Tracing Algorithm to the Decoy Image	35
2-13: Range Shadow Casting	36
2-14: Convex and Concave Angles	38
2-15: Region Diagram of Convex and Concave Angles	39
2-16: Physical Edge Angle Image of Decoy	42
2-17: Concave Edge Detection with Threshold Equal to 220°	43
2-18: Thinned Concave Edges	44
2-19: Recursive Line Fitting for Occluding and Concave Edges	45
3-1: Multi-level Matching and Processing Control Flow	50
3-2: Examples of Generalized Cylinders	54
3-3: Hierarchical Model of a Tank	56
3-4: 3-D Models of a Missile Launcher and a Decoy	59
3-5: Missile Launcher Decoy Model Information	60
3-6: Sample Set of Interpretation Rules	67
3-7: Extracted Feature Information from Decoy Image	72
3-8: Relevant Model Information	73
3-9: Likelihood Weights Associating Rules and Object Models	74
3-10: Line Segments of Occluding and Concave Edges	76
3-11: Extracted Cylinder Feature Information from Decoy Image	78
3-12: Extracted Feature Information from Missile Launcher Image	79
3-13: Likelihood Weights Associating Rules and Object Models	80
3-14: Line Segments of Occluding and Concave Edges	81

1. INTRODUCTION

This report documents the research efforts of Advanced Information & Decision Systems (AI&DS) to develop three-dimensional (3-D) object classification techniques for vehicle targets in air-to-ground laser range imagery. The authors ~~We~~ emphasize an artificial intelligence (AI) approach to intelligently interpret laser imagery in terms of 3-D symbolic models. The full classification system includes 3-D image feature extraction, geometric modeling, model prediction, and feature-to-model matching. ~~In~~ this report ~~we discuss~~ discusses new techniques for implementing these major system components, and provides overall conclusions and discussion about the feasibility of developing an automated 3-D object classification system.

1.1 RESEARCH OVERVIEW

A range sensor measures the distance from the sensor to the visible object surface along a given ray. Range images offer significant advantages over passive reflectance images because they preserve the 3-D geometry of the scene viewed from the sensor. The intrinsic properties of the scene such as depth, surface orientation, length, and size are of fundamental importance for scene segmentation, target recognition and scene interpretation. While these properties can only be obtained from 2-D images with extensive inference (due to the ambiguities introduced by the 2-D projection of the 3-D scene), they can be easily calculated from 3-D range images. Therefore, range data is becoming an increasingly important source of information for a variety of industrial and military applications. Military applications of laser or range imagery include automatic 3-D target classification, autonomous vehicles, and missile guidance. Industrial applications of range data include automatic inspection, and part handling and assembly by robots.

There are three types of range sensors: stereo vision, active illumination (light stripe) and laser range finder (time-of-flight). Stereo vision requires a correspondence process for matching and registering multiple images from different views. This is a difficult problem, and the accuracy of range estimation depends on the distance of the object and the baseline length. Active illumination uses the same triangulation principle to measure the depth information as in stereo vision. Instead of using two images corresponding to two views, this technique uses a stripe of light as illumination and records an image from another position. Active illumination can only be used in a controlled environment such as an industrial assembly line. Laser range finders directly measure the time delay or use modulation techniques to obtain the depth information. Laser range finders avoid the restrictions inherent in stereo and active illumination, and have potential applications in autonomous vehicles, missile guidance, space exploration, robotics and industrial inspection. Although laser reflectance imagery registered with the range imagery is also available, it may be degraded by severe speckle noise and is thus not very useful for target classification. In this project report we discuss 3-D object classification for laser range imagery.

Traditional methods for target classification in systems with laser range sensors involve standard statistical pattern recognition algorithms that match multiple pre-stored model images with the sensed images. These methods require a large number of pre-stored model images corresponding to various viewpoints and tend to perform adequately in normal situations where the observed laser image corresponds to one of the model images. In situations involving unusual target aspect angles, partially obscured or camouflaged targets, or targets of varying structure, most pattern recognition techniques based on matching global image features will fail.

Recent Image Understanding (IU) research programs have emphasized the symbolic interpretation of low level image features in terms of models. The ACRONYM system [Brooks-81] developed at Stanford is a powerful model-based vision system for 2-D image interpretation. The approach taken in ACRONYM is to match the extracted 2-D image features with the predictions from the 3-D model at multiple levels through the use of a geometric reasoning system. This system is shown to have good performance for aerial photo-interpretation applications. In our 3-D object classification task, the 3-D image features can be extracted directly from laser range imagery. Therefore, the interpretation process requires symbolic reasoning among 3-D models and 3-D image features. This capability does not exist in the ACRONYM system, and is our major research contribution in this project.

In this research, we adopt an artificial intelligence (AI) approach to interpreting 3-D range data in terms of 3-D symbolic models for object classification. The artificial intelligence approach brings together a wide variety of analysis techniques (e.g., algorithmic, heuristic, statistical) in order to perform interpretation in an "intelligent" manner. This approach is more robust and uses contextual information and common sense reasoning to aid in the analysis. The critical research issues involved are:

1. extraction of 3-D image features from laser imagery.

2. prediction of 3-D image features from 3-D object models.

3. interpretation of 3-D image features in terms of models for object recognition and classification.

These three functions interact heavily, and the control process is of fundamental importance. A summary of our research results and conclusions is presented in the next section.

1.2 SUMMARY AND CONCLUSIONS

AI&DS has developed a combined model-driven and data-driven 3-D object classification technique for analyzing laser range imagery. This research effort on air-to-ground vehicle target classification has adopted an AI approach that involves a hierarchical analysis process for extracting and matching multi-level features from laser imagery to object models. The low level image features are extracted from the image using little model knowledge. The high level symbolic features such as object components are extracted from low level features with the guidance of model predictions. The object model is represented by a single, viewpoint-independent 3-D model. The object classification task proceeds from coarse to fine by first comparing gross object features (e.g., object length, height, extreme points, etc.) and then finer component features (e.g., component volume, position, orientation, etc.) extracted from laser imagery with a model using a set of rules that produces a likelihood value to indicate the goodness of match. Since the 3-D information are available from the range image, the actual measurements (e.g., length, width, volume) are used for matching.

A bare-bones system was designed and developed. It is domain-independent in that it is applicable to a variety of tasks such as vehicle classification, ship classification, and industrial parts classification. This system has four major components as shown in Figure 1-1. The 3-D feature extraction techniques include object-ground segmentation, object orientation estimation, 3-D physical edge detection and linking, and projection image formation. We emphasize extracting physical features which are directly related to a 3-D object model from the range data. The object-ground segmentation algorithm extracts object segments from the laser image by use of a downward-continuation process. The object segment is then used to generate its ground projection and side-view projection images. The object orientation can be estimated from the orientation of the ground projection image since vehicle targets are usually elongated. The side-view projection images provide major object structure information and can be used to prune possible object models. Physical edges, such as occluding, convex and concave edges that are not distinguishable in an intensity image can be extracted from a range image directly. A 3-D physical edge detection algorithm is developed for this purpose. This physical edge detector is not only useful for edge detection, but also useful for extracting planar and curved surfaces.

The objects are represented by a viewpoint-independent volumetric model based on "generalized cylinders." After the initial feature extraction process has generated a candidate set of hypothesized target classifications, a prediction process can be used to further evaluate the hypotheses. The prediction process predicts the appearance of the model in the range image. Typical predictions are physical edge types (occluding, convex, or concave edges), cylinder contour, and invariant shape properties (parallel, collinear, connectivity). These knowledge-based predictions are very powerful for directing the feature extraction algorithms' search for particular features in a limited region. The results of feature extraction and prediction are gathered at multiple levels, and a reasoning process is applied to classify the target. The interpretation process uses the features of the individual components and the component structure of the object as the basis for matching.

A preliminary classification experiment was performed on a class of military vehicles which included such objects as tanks, missile launchers, and decoys. These targets have distinct structures and components. Past work on object recognition using 3-D range data only applied to simple objects without much self-occlusion (one component occludes another component). To classify the targets that we modeled is a

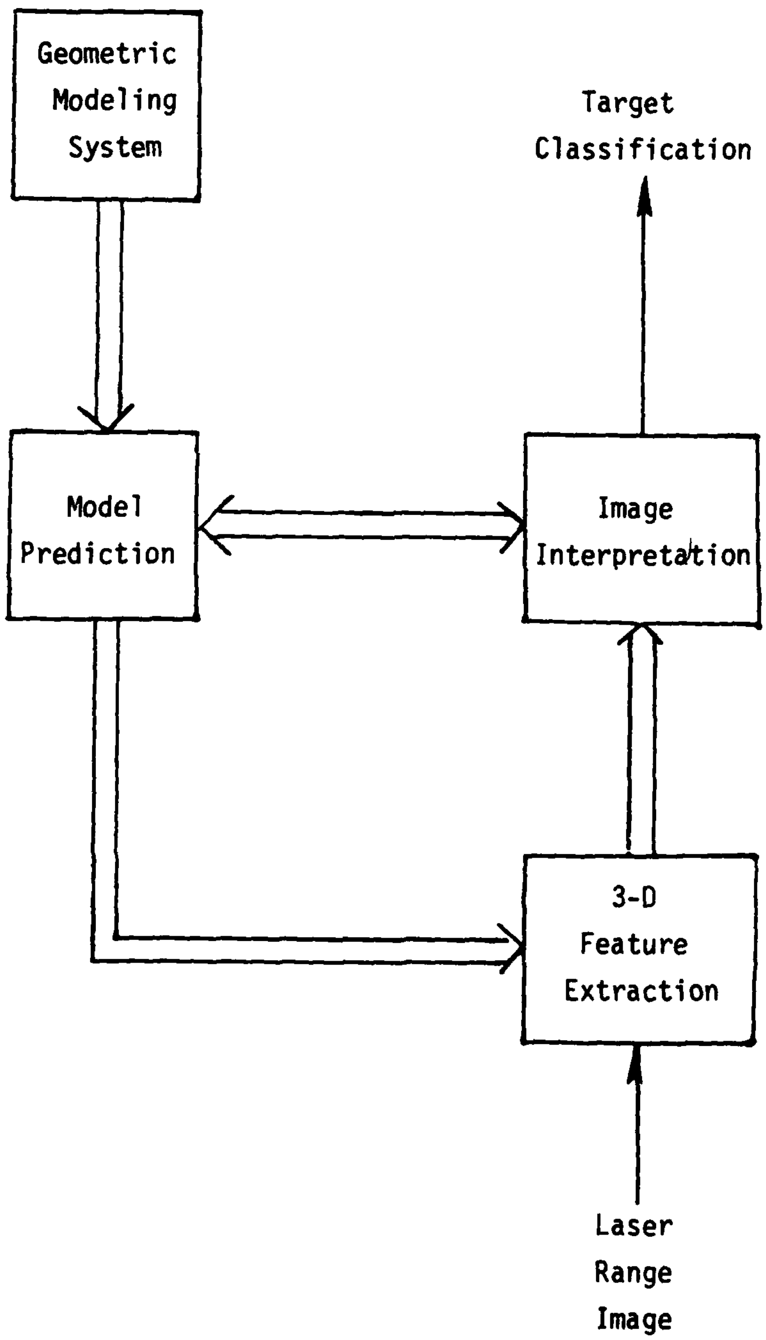


Figure 1-1 Major Components of 3-D Object Classification System

challenging task.

The experiment consisted of processing synthetic range imagery with additive range noise (provided by The Analytic Science Corporation). The experiment consisted of processing imagery through a set of low level feature extraction and high level feature interpretation algorithms and inference rules. All the processing was automated but done in a nonintegrated fashion. Although only a small amount of imagery was processed based on a few models, the successful results with a primitive inference system are encouraging.

1.3 ORGANIZATION OF THIS REPORT

This project report is organized as follows: Section 2 presents basic 3-D feature extraction techniques for laser imagery. These include object-ground segmentation, projection image generation, and 3-D physical edge detection techniques. Examples of these 3-D feature extraction techniques are presented.

Section 3 discusses techniques for classifying vehicle objects by matching the extracted image features with models. Objects are represented by volumetric models based on generalized cylinders. Multi-level predictions are generated from the model to guide the extraction of higher level features from edge segments.

Section 4 presents conclusions and discussion about the feasibility of using artificial intelligence techniques for 3-D laser target classification. Future directions pertaining to the critical research issues are also addressed.

2. THREE-DIMENSIONAL IMAGE FEATURE EXTRACTION

There are two major tasks in our laser target classification system, namely (1) extracting 3-D image features from laser imagery and (2) model-based interpretation of extracted image features for object recognition. In this section, we discuss the first task, where a largely data-driven approach is used to extract 3-D image features. Several techniques have been studied for extracting 3-D features from range data. Duda et al. [Duda et al.-79] used registered range and reflectance data to find planar surface regions in a sequential fashion. Oshima and Shirai [Oshima, Shirai-79] fitted local range data with planar surfaces, and merged local planes into planar and curved surface regions. Agin and Binford [Agin, Binford-76] extracted "cylinder" features from range data and segmented complex objects into simpler subparts in terms of generalized cylinders. All of these methods are basically concerned with various ways of fitting planar and curved surfaces to the range data, and require the 3-D coordinates of the surface points. Another approach is to obtain edge boundaries from range data. Nevatia and Binford [Nevatia, Binford-77] extracted jump boundaries from range imagery for object recognition. Sugihara [Sugihara-79] used a junction dictionary to guide the extraction of physical edges. These 3-D edge detection techniques operate on the range image and only provide range difference information as seen from the sensor without directly referring to the true physical properties of the object.

In this section we present several new techniques for 3-D feature extraction. We emphasize the extraction of 3-D physical features of the object from range data. The approach used is to first transform the range image (in a sensor-centered coordinate system) to the surface data (in a world coordinate system) from knowledge of the sensor position. We then separate the object from the background by an object-ground segmentation algorithm. Once the object segment is extracted from the image, the ground projection and side-view projection images of the object segment are generated. These projection images are useful for extracting gross object features and major object structure. The object orientation can be estimated from the orientation of the ground projection image since vehicle targets are usually elongated. The side-view projection image can be used to locate major object structure positions such as wheel and missile positions of a missile launcher. After extracting those global features, a 3-D edge detection algorithm is used to extract physical edge segments for fine feature-to-model matching. Our 3-D edge feature extraction algorithm directly calculates the physical angle of the object surface from surface data. Convex and concave edges can be distinguished according to the value of the physical edge angle. This physical edge angle image is not only useful for physical edge detection, but also provides relative surface orientation information for extracting planar and curved surfaces. Figure 2-1 presents the structure of our 3-D image feature extraction process. The extraction process moves from the laser image to symbolic feature information that is provided to other levels of the system.

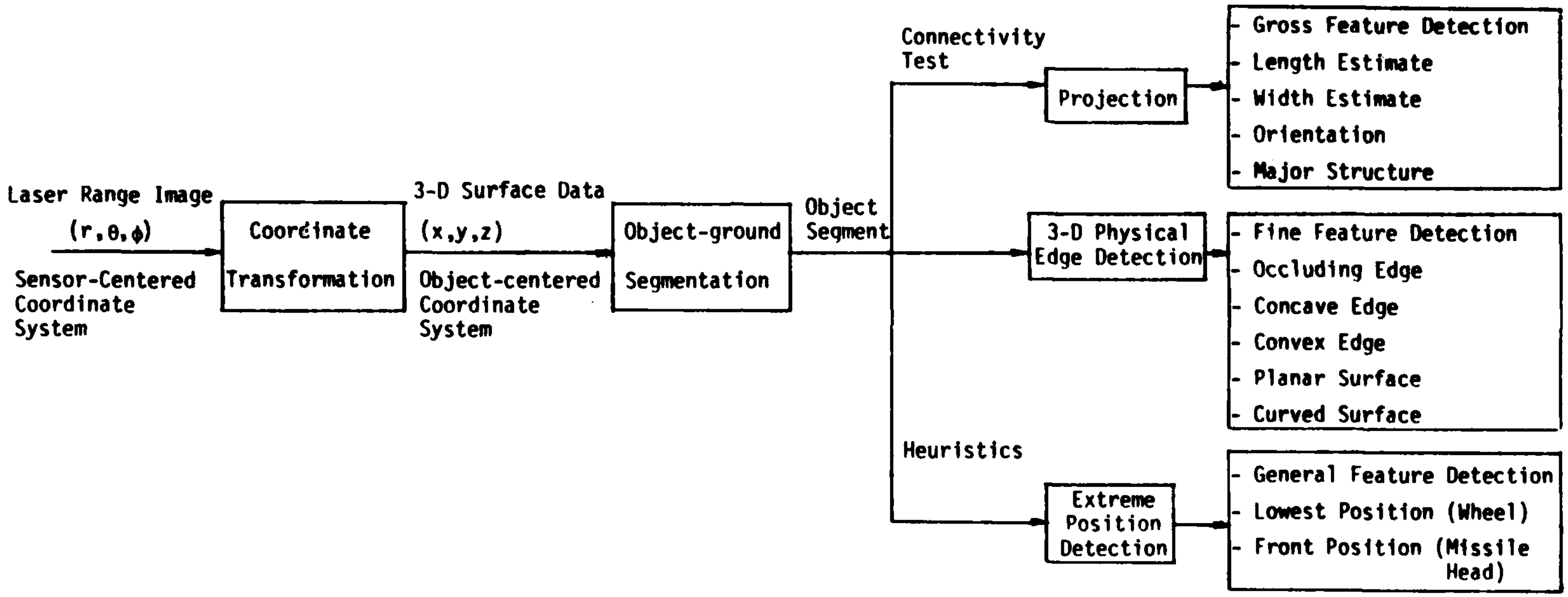


Figure 2-1 The 3-D Image Feature Extraction Process

2.1 COORDINATE TRANSFORMATION

There are two basic methods of range data acquisition: triangulation and time-of-flight. The range information acquired from both systems is described using a sensor-centered coordinate system, and we call it a "range image." The range image is a special ordering of 3-D range data viewed from the sensor, and the topology (neighborhood pixel relationships) defined in the range image is useful for feature extraction. However, in order to manipulate the range data more effectively, we need to transform the range image into a sensor-independent world coordinate system. This transformation can be carried out through a camera calibration procedure [Sobel-70] or from the geometry of the data acquisition system. For an air-to-ground laser sensor, the information required for this coordinate transformation are the depression angle and the angular scanning resolution along the azimuth and elevation directions. Figure 2-2 illustrates this transformation. The resulting 3-D Cartesian coordinates of the visible surface defined in a world coordinate are called "surface data."

The surface data obtained through the coordinate transformation provide direct 3-D scene information. Figure 2-3(a) is a synthetic range image of a decoy produced by The Analytic Science Corporation. In this image the range image intensity is proportional to the range value, and a bright region in the image corresponds to an area far away from the sensor. A drawing of a decoy is shown in Figure 2-3(b) for comparison. The range image in Figure 2-3(a) is defined in a sensor-centered coordinate system and does not give us direct information about

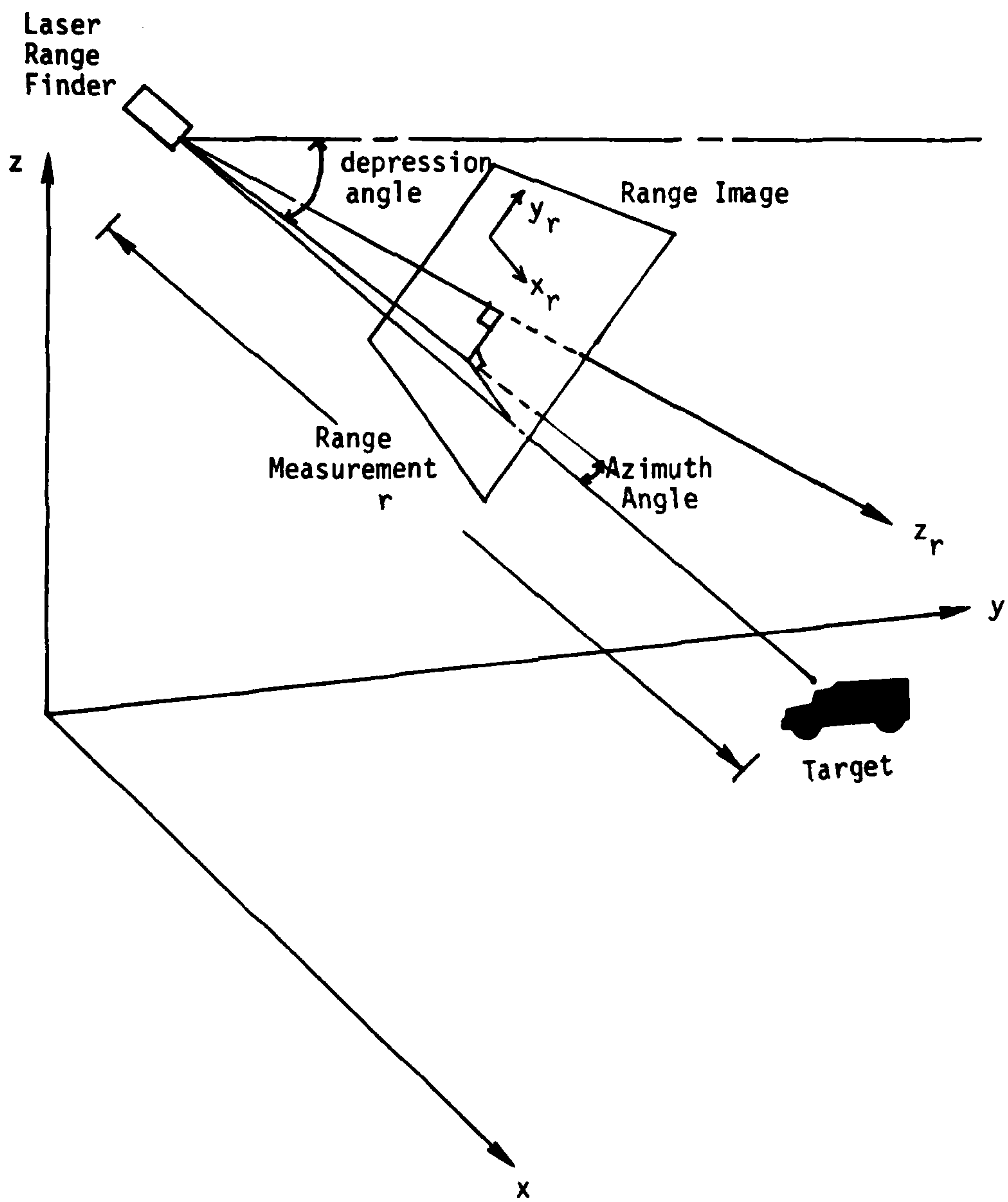


Figure 2-2 Coordinate Transformation Between a Sensor-Centered Coordinate System and the World Coordinate System

Figure 2-3(a) Range Imagery of a Decoy

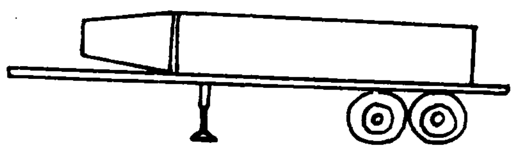


Figure 2-3(b) Drawing of a Decoy

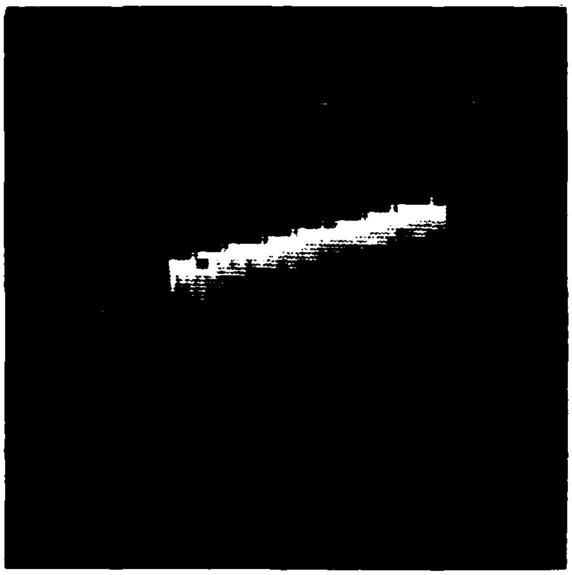


Figure 2-3(c) Z Coordinate of Figure 2-3(a)

The surface data obtained through the coordinate transformation provide direct 3-D scene information. Figure 2-3(a) is a synthetic range image of a decoy produced by The Analytic Science Corporation. In this image the range image intensity is proportional to the range value, and a bright region in the image corresponds to an area far away from the sensor. A drawing of a decoy is shown in Figure 2-3(b) for comparison. The range image in Figure 2-3(a) is defined in a sensor-centered coordinate system and does not give us direct information about the 3-D physical properties of the object. However, the z coordinate (defined in Figure 2-2) of the surface data obtained from the range image as shown in Figure 2-3(c) gives us a clear knowledge of the object elevation distribution along the z-axis.

In this report, we assume that surface data is available, and our 3-D feature extraction techniques operate directly on surface data independent of the techniques used for range image acquisition (e.g., stereo, light stripe, or laser range finder).

2.2 OBJECT-GROUND SEGMENTATION

To analyze range images, we need to first separate interesting objects from the background. In most applications (such as autonomous vehicles and robot vision), the background is the ground that supports the objects. Object-ground segmentation can be done by finding the jump boundaries in the range image and linking them into a complete bounding contour of the object. However, for noisy range data or objects that touch the ground surface, this process cannot be done reliably. A more robust segmentation technique that considers object and ground characteristics seems more appropriate.

Knowledge of the relationship between object and ground can aid the segmentation. The object is supported below by a ground plane that is locally flat. Furthermore, there are no data points below the ground plane. These support relations allow us to select a threshold z_o (along the z axis in the world coordinates as in Figure 2-2) to separate those object points at least z_o distance above the lowest ground point. The selection of threshold z_o depends on the heights of the objects that we are interested in.

The object segment extracted consists of data points on the upper part of the object. The next step is to perform a local downward-continuation process such that we can adaptively lower the threshold in order to extract more object points without including ground points. Since an object is a connected 3-D blob with finite extent, the connectivity property of the object segment and concave edge evidence (which will be discussed in Section 2-4) can be used for this purpose. If we lower the threshold too much, some ground points will be included in the object segment. For this case, the segmented object points will not be connected and the extent of the object blob will increase sharply due to the inclusion of randomly spatially distributed ground points. Figure 2-3(a) is a synthetic range image of a decoy above an uneven ground support. The object segment extracted from the range image through downward continuation is shown in Figure 2-4.

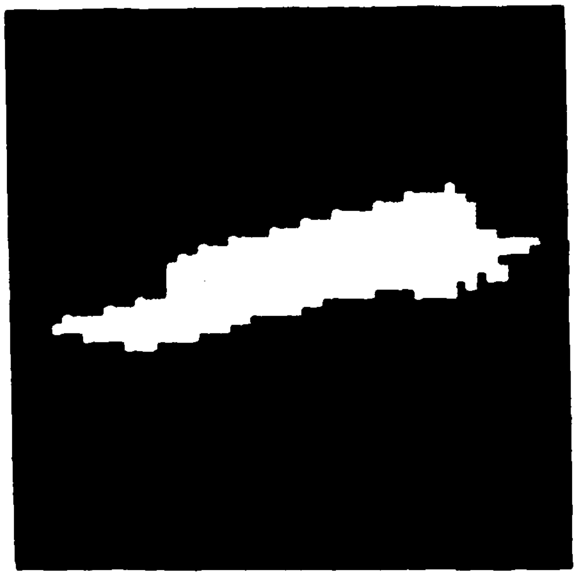


Figure 2-4 Extracted Object Segment of Decoy

Note that the downward-continuation process is performed locally such that the assumption about object and ground are valid. To further extract object points, we need to use an adaptive thresholding scheme. The extent of the object segment is fixed (e.g., its ground projection) and points around the object segment are examined. New points are included in the object segment if they do not increase the extent of the object and no concave edge has been reached. Concave edges are used because they occur at the junction of object and ground.

The object-ground segmentation algorithm generates a binary mask where the object points are set to one and the background points are set to zero as shown in Figure 2-4 of the object segment. Subsequent processing can be concentrated inside the object segment to reduce computation time and unnecessary analysis. This binary mask also provides a suitable form for extracting some important object features such as boundary curves.

2.3 PROJECTION IMAGES GENERATION

Once we extract the object segment from the range image, the next step is to extract and analyze the global object features such as orientation, length, width, height, and the boundary. The object-ground segmentation process generates a binary mask of the object segment, and this allows us to focus attention inside the object.

The range image is recorded as a perspective view of the scene viewed from the sensor. The ordering or the neighborhood pixel relationships of the range image is defined in a sensor-centered coordinate system. The coordinate transformation discussed in Section 2-1 transforms the range image into 3-D surface data in an object-centered coordinate system and thus removes this specific ordering and permits effective manipulation of 3-D data. Projection image generation is a good example of reordering these surface data points to form useful new images. Gross object features and major object structure can be extracted from projection images.

One of the most important pieces of information about the object in the scene is its orientation. This information is not directly available from the range image. However, with a specific ordering according to the x and y coordinates of the surface data, i.e., a ground projection, we can estimate the object orientation easily. Figure 2-5 is the projection of the object segment in Figure 2-4 onto the ground plane. The jump points in the upper left side of the ground projection image are caused by the split of the scanning laser beam on the object boundary points. These error points can be eliminated by checking the connectivity of the object segment. The connectivity algorithm removes isolated points in the object segment according to their 3-D coordinates based on the assumption that the object is 3-D connected and a single jump point must be due to noise. The ground projection of the refined object segment is shown in Figure 2-6. The ground projection image provides the top view of the object that is not available from the sensor position. However, this is only a partial top view. The occluded part

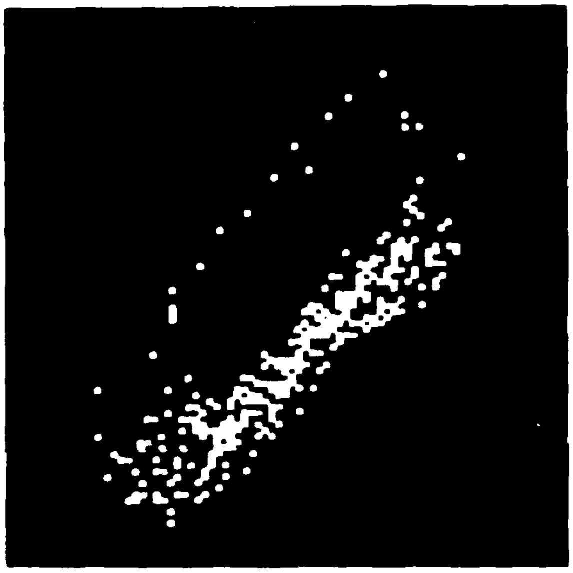


Figure 2-5 Ground Projection Points of Decoy

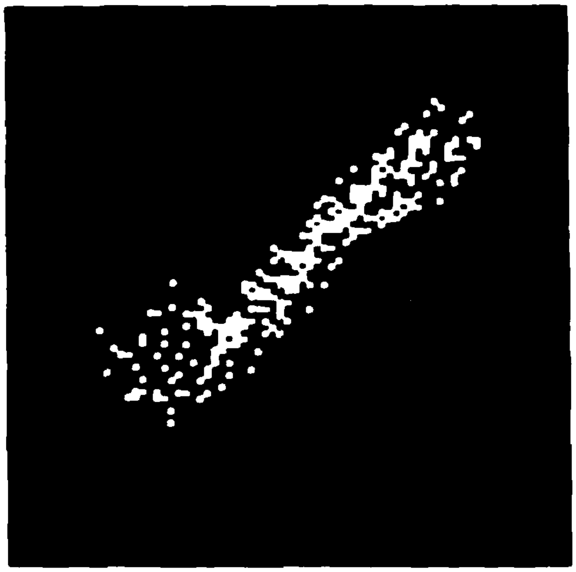


Figure 2-6 Refined Ground Projection of Decoy

of the object seen from the sensor will not appear in the projection image. Therefore, in Figure 2-6, we are more certain about the object length direction than the object width direction. The occlusion problem has to be resolved by a higher level model prediction system, and the occlusions in the projection image may put back-constraints on the model.

Because the imaged targets are vehicles whose shape is known (i.e., they are usually elongated), the orientation of these objects can be obtained by finding the orientation of the most elongated bounding rectangle on those ground projection points as illustrated in Figure 2-7. The length and width estimates of the object are equal to those of the bounding rectangle. In Figure 2-6, the object orientation estimate is 45 degrees with respect to the x-axis in the world coordinate system. The length estimate is 199, and the width estimate is 55.86 where each resolution unit is 2 inches. Due to the object-self-occlusion along the width direction, we are not certain about the width estimate. This fact is included in the rule-based interpretation system in that we allow a larger tolerance for object width in feature-to-model matching. Note that the actual measurements for scene geometry (e.g., lengths, widths, heights, etc.) are available from surface data and can be compared directly to similar parameters of an object model.

Other important characteristic views can also be obtained from projections. For example, we can project the surface data to the plane defined by the orientation of the object and the z-axis to obtain the side view of the object. Figure 2-8 is the side view projection of the

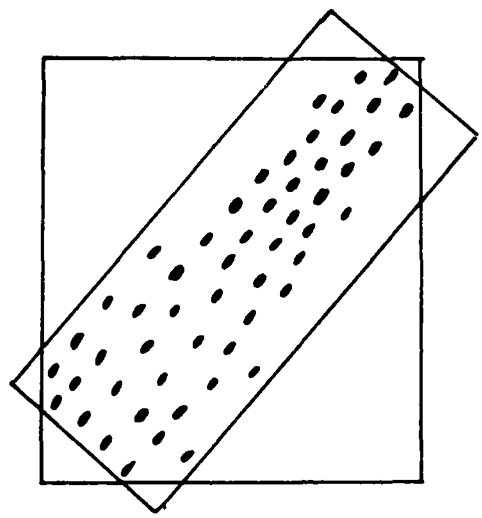


Figure 2-7 Bounding Rectangles for Orientation Estimation

object segment in Figure 2-3(a). Much structural information not available from the sensor position shows up in this side view projection.

Figure 2-9(a) is the synthetic range image of a missile launcher. Figure 2-9(b) is a drawing of a missile launcher. It is difficult for human analysts to determine whether this object is the same object as in Figure 2-3(a) or not. Figure 2-10 is the side view projection of the object segment in Figure 2-9(a). Unexpected structures are revealed in this side view picture and are very useful for target classification. For simple objects, the orientation and the characteristic side views may be sufficient for object recognition and manipulation.

This suggests a simple 3-D recognition scheme; that is, we only store two characteristic views (ground projection and side view projection) as the model images, and compare the projection images generated from 3-D surface data to the model images. However, this scheme does not work for complex objects and situations where severe occlusion occurs.

The projection points are unstructured. A binary image or even a range image can be formed from these data. Figure 2-11 is the binary image formed from projection points in Figure 2-8. To get this picture, we first define a sampling distance D_s on the projection plane. A resolution cell is a D_s by D_s square. If a certain number of points fall inside a resolution cell and its 8-connected neighboring cells, the center cell is set to one, otherwise set to zero. This procedure creates a silhouette projection image of the object from a new viewing

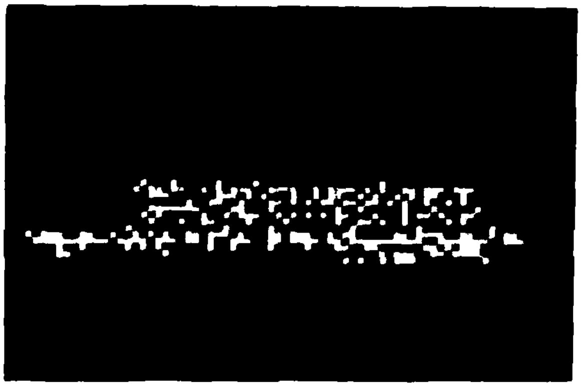
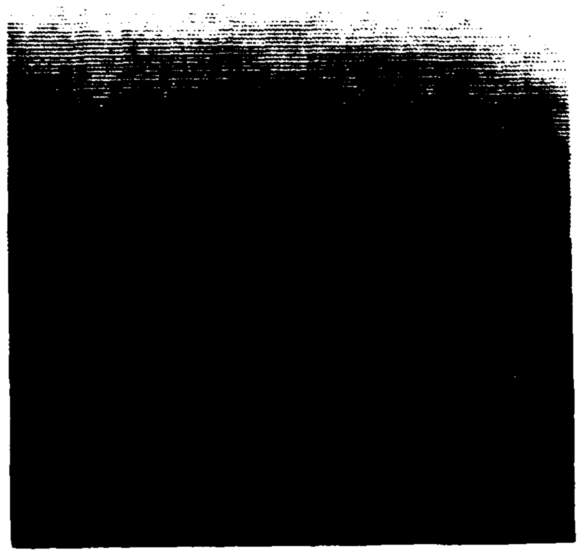
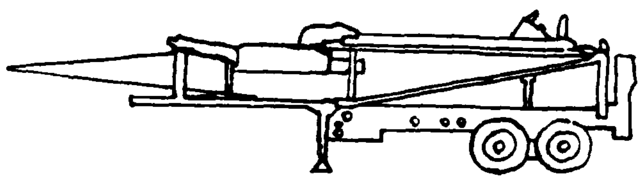


Figure 2-8 Side-View Projection of Decoy



(a) Range Image



(b) Missile Launcher Drawing

Figure 2-9 Range Image of a Missile Launcher

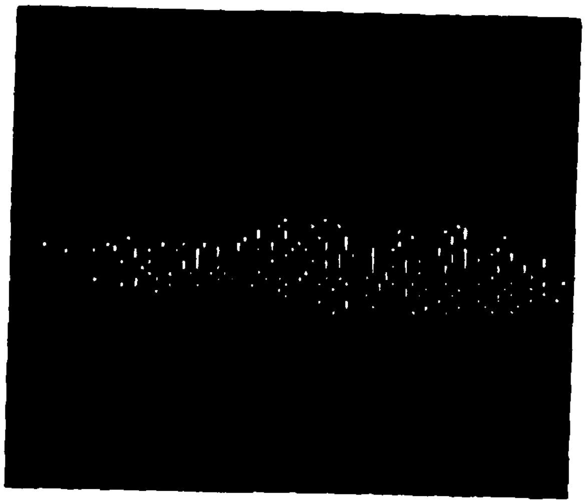


Figure 2-10 Side View Projection Points

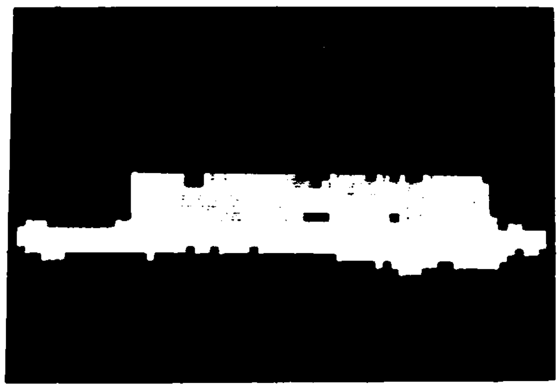


Figure 2-11 Side-View Projection Image of Decoy

position. The range image seen from this new position can also be generated by a hidden point elimination algorithm similar to the z-buffer algorithm [Newman, Sproull-73] used for hidden surface removal. Essentially, for each resolution cell and its 8-connected neighboring cells, we keep a record of the closest data point from the new viewing position and the resulting image is the side-view range image.

Projection images provide important information about the object that is not directly available from the range image. Standard image analysis algorithms for binary images can be applied to these projection images to extract gross object features at the object level. The same techniques can also be used at the component level for component orientation estimation, length and volume estimation, and major structure identification.

2.4 3-D PHYSICAL EDGE DETECTION

The feature extraction techniques discussed in the last two sections extract gross features at the object level, and are useful for object orientation estimation and major structure identification. However, they do not provide internal object component structure and fine details which are of critical importance for complex 3-D object recognition. In this section, we introduce a physical edge detection technique for extracting fine edge features. These low level edge features are the primitives in goal-directed shape feature extraction.

2.4.1 Occluding Edge Detection

Physical edges such as occluding, convex, and concave edges that are not distinguishable in an intensity image can be extracted from a range image directly. Past work on 3-D edge detection concentrated on jump boundary extraction. These jump boundaries correspond to large range discontinuities caused by object occlusion and can be easily extracted from the range image. However, jump boundaries are also regions subject to large measurement errors due to the splitting of the scanning laser beam across occluding boundaries. Although this will only introduce one or two pixel error in the range image, it causes large errors in 3-D coordinates and restricts the use of 3-D position and orientation of the occluding boundaries. A more reliable way to extract/calculate the occluding exterior boundary of an object is to trace the boundary of the extracted object segment. Since the connectivity analysis has been performed on the segmented object, large error points have been removed and we have more confidence in the 3-D edge information. Figure 2-12 is the exterior boundary obtained from the refined object segment (after the connectivity test) in Figure 2-4 by a boundary tracing algorithm [Rosenfeld, Kak-76]. This boundary contour only describes the exterior occluding edges between the object and background. Another type of occluding edge that occurs inside the object segment is called an interior occluding edge. Interior occluding edges are due to object self-occlusion at the component level (e.g., one component occludes another). There are two advantages to explicitly distinguishing these two types of occluding edges. First, the interior occluding edges usually have smaller range jumps and we need to use a smaller threshold for edge detection. Second, the exterior occluding segments are global features at the object level while the interior occluding segments are features at the component level. To distinguish features at multiple levels can facilitate and hasten the feature to model matching process.

In the world coordinate system defined in Figure 2-2, if the azimuth scanning direction of the sensor lies roughly along the x axis direction, then the x coordinate values of the range image have more or less uniform sampling distance. The y and z coordinates vary according to the spatial detail of the object. The large range discontinuity is the result of occlusion or range shadow casting as illustrated in Figure 2-13, where the z difference is the reason for occlusion. This suggests that the interior occluding boundaries can be extracted from large differences of z or y coordinates instead of the range difference. Thus occluding edges can be extracted from surface data instead of the range image and are more closely related to the physical properties of the object. The importance of this direct physical relationship becomes more clear for concave and convex edge detection.

2.4.2 Concave and Convex Edge Detection

There are certain invariant properties of shapes that are in general independent of the sensor position. For example, three collinear points in 3-D space will be collinear in the 2-D projection images viewed from different positions. This invariant property of the collinear relationship is a singular case of a more general invariant

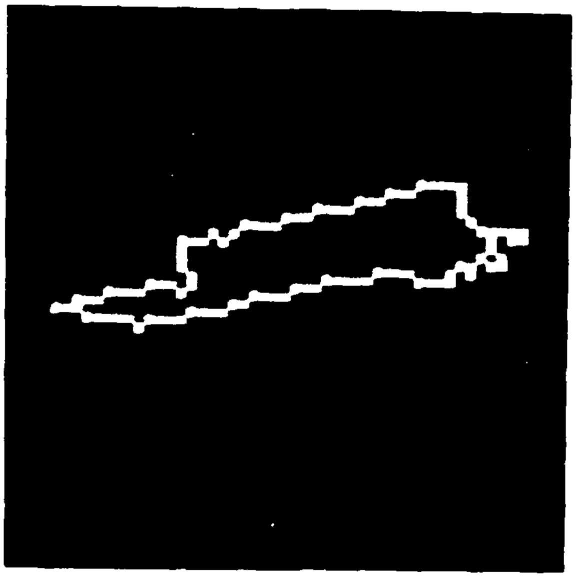


Figure 2-12 Exterior Boundary Obtained by Applying the Boundary Tracing Algorithm to the Decoy Image

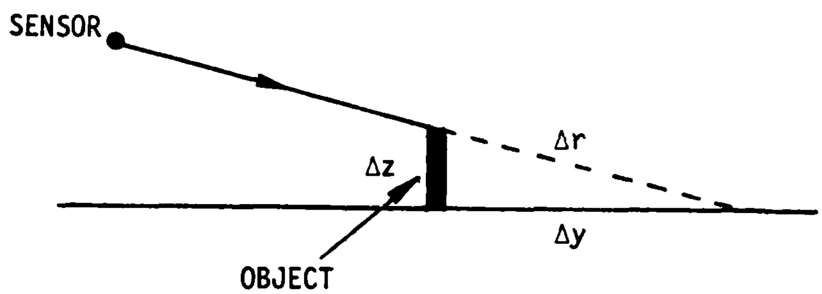


Figure 2-13 Range Shadow Casting

property of concave and convex angles. Assume that we have three points in the y-z plane as illustrated in Figure 2-14. The two vectors a(k-1) and b(k+1) are defined as

$$a(k-1) = r(k-1) - r(k) \text{ and } b(k+1) = r(k+1) - r(k)$$

where r(k) is the position vector of the center point. The order of these vectors has been defined, and a(k-1) is the first vector, b(k+1) is the second vector. The direction of the cross product a(k-1)*b(k+1) will be pointing in the positive x direction if the counter clockwise angle from a(k-1) to b(k+1) is between 0 and 180 degrees; i.e., a convex angle. On the other hand, for a concave angle where the counter clockwise angle from a(k-1) to b(k+1) is between 180 degrees and 360 degrees, the direction of a(k-1)*b(k+1) will be pointing along the negative x axis. This property is invariant as long as the viewer is in the positive x half-space. Thus convex angles and concave angles have different polarities and can be determined once we define our relative viewing position (the particular half-space). Although the apparent angle between these two vectors projected on an image plane will vary according to the plane orientation, it will not exceed 180 degrees if it is a convex angle in the 3-D space. That is, a convex angle in 3-D will always appear to be nonconcave in 2-D images. Similarly, a concave angle in 3-D will always be a nonconvex angle in 2-D images. The singular case (180 degrees) occurs when the viewing position is on the y-z plane. Figure 2-15 shows the region of the apparent angle on the image plane for determining 3-D convex and concave angles.

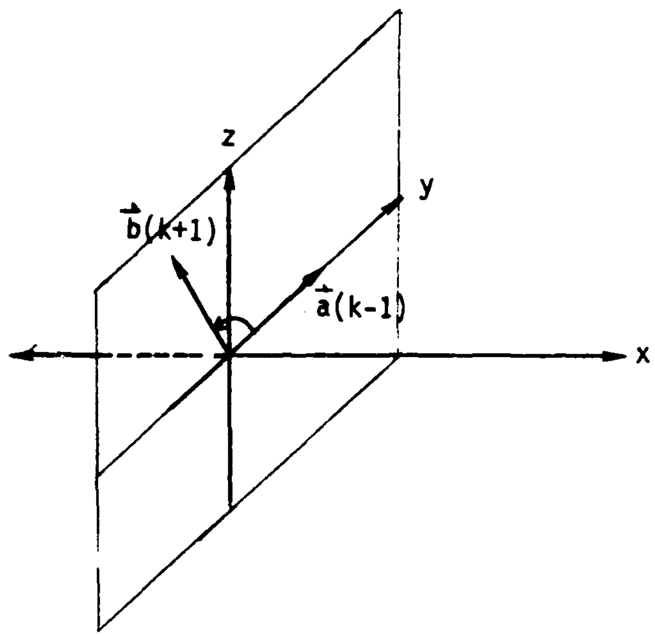


Figure 2-14 Convex and Concave Angles Calculation

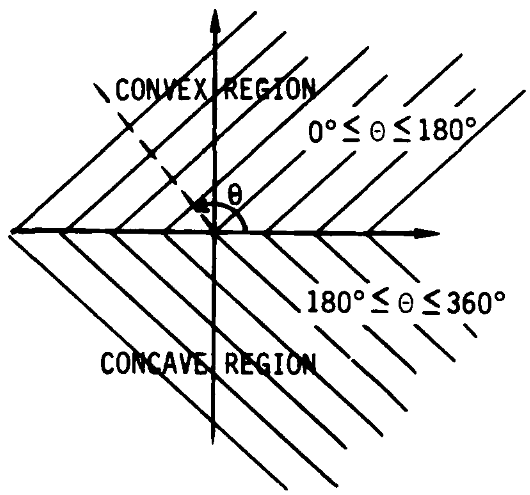


Figure 2-15 Region Diagram of Convex and Concave Angles

From the above discussions, it is clear that the extraction of convex and concave edges from range images can be done by calculating the polarity of the data. The polarity information is sufficient to distinguish between convex and concave edges, but it is not a physical quantity that directly related to the physical properties of the object. Here, we calculate the physical angle of three surface data points along a specified direction. For example, let the surface data of a range image at pixel (i,j) be r(i,j). The distance from the sensor to the object point is $||r(i,j)||$. If the column direction is chosen, the vectors a(k-1) and b(k+1) are defined as

$$a(k-1) = r(i-s,j) - r(i,j) \qquad b(k+1) = r(i,j+s) - r(i,j);$$

where s is the step size. The step size is chosen according to the image resolution and the amount of noise in the range image. The physical angle between a(k-1) and b(k+1) can be determined from their inner product and cross product. Those points with concave angle (180 to 360 degrees) are candidates for concave edges. To avoid the singular case (edges along the column direction), we need to calculate the edge angle along the row direction. From the invariant property of convex and concave angles, the two physical edge angle images along column and row directions are sufficient to detect convex and concave edges. The threshold for physical edge detection corresponds to a physical quantity (i.e., edge angle) of the object. Figure 2-16 is the physical edge angle image of Figure 2-3(a) along the column direction with step size equal to 2. The edge angle is between 0 and 360 degrees, and the intensity of the image is proportional to the edge angle. Bright edges are concave edges which usually occur at the junction of two object components and are of fundamental importance for segmenting complex objects into simpler components. Figure 2-17 is the detected concave edges of Figure 2-16 with threshold equal to 220 degrees.

The physical edge angle image also provides local surface orientation information. Instead of thresholding for convex or concave edge detection, we can merge connected points with physical angle close to 180 degrees to extract a planar surface. A curved surface such as the cylinder on top of the vehicle platform in Figure 2-3(a) will show as connected points with their edge angles clustered in a convex angle region. Thus our 3-D edge feature extraction algorithm is not only useful for physical edge detection, but also suitable for surface reconstruction.

2.4.3 Linear Feature Extraction

The detected edge points in Figure 2-17 are clustered and require edge thining. The edge thinning algorithm we used is similar to the Nevatia-Babu edge detector [Nevatia, Babu-80] and it proceeds as follows: If the edge angle at the pixel is larger than the edge angles of its two neighbors in a direction normal to the direction of the edge, then the edge point is considered to be present at the pixel. The thinned concave edges of Figure 2-17 are shown in Figure 2-18. The next step is to link the edge points into an edge segment. Finally, each edge segment is approximated by piecewise linear segments. This is accomplished by using the well-known recursive line fitting algorithm [Duda, Hart-73]. This algorithm proceeds recursively in approximating a

Figure 2-16 Physical Edge Angle Image of Decoy

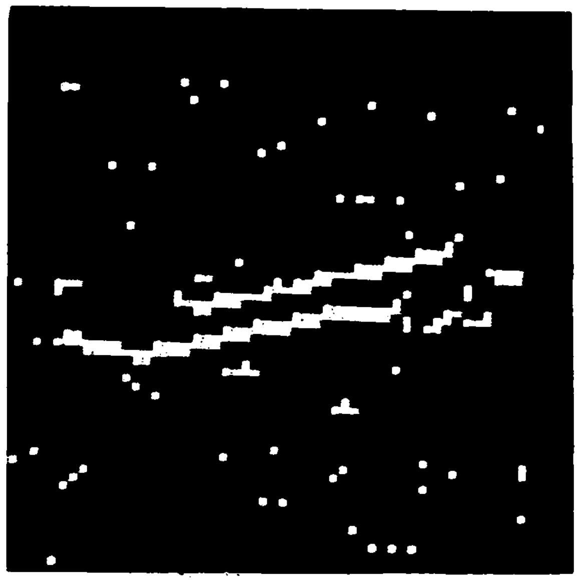


Figure 2-17 Concave Edge Detection with Threshold Equal to 220°

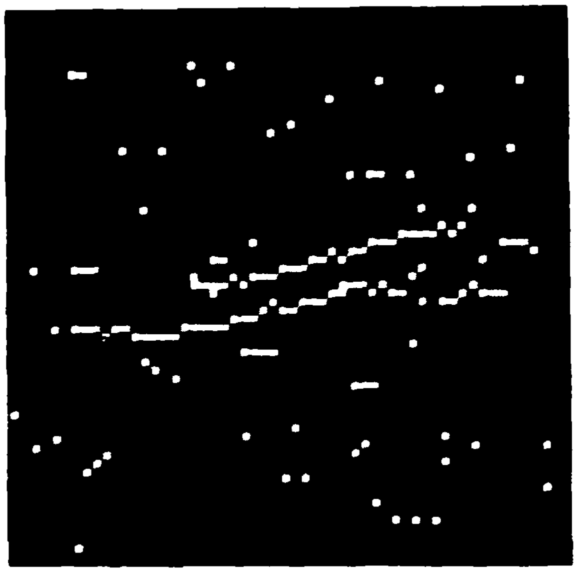


Figure 2-18 Thinned Concave Edges

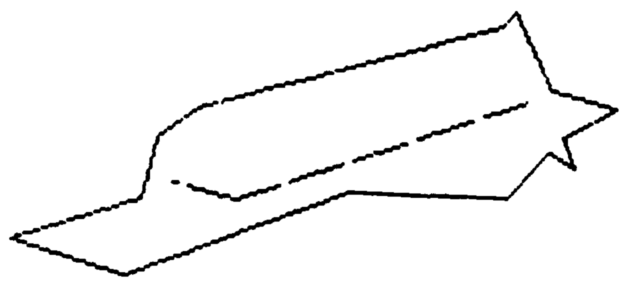


Figure 2-19 Recursive Line Fitting for Occluding and Concave Edges

segment by joining the end points and then dividing the segment into two segments at the point where the original segment has maximum deviation from the data. This process is repeated until all the segments fit within the threshold. Figure 2-19 shows the results of applying the recursive line fitting algorithm to the exterior occluding boundary (in solid-line) of the object segment and the concave edge segments (in dashed-line). Note that the cylinder on top of the vehicle platform is bounded by occluding and concave edge segments and can be extracted by a cylinder extraction algorithm. These physical edge segments are inputs to a high level cylinder extraction algorithm for segmenting complex objects into simpler components.

3. MODEL-BASED INTERPRETATION

The task of 3-D object classification is one of comparing image features extracted from the laser imagery with object models. Much of the work in computer vision is based on trying to extract image features without any *a priori* knowledge of the object model. This approach faces the difficult problem of developing a meaningful interpretation based on image features that are ambiguous and incomplete due to inadequate feature extraction processes. Furthermore, the inherent ambiguities of perspective projections and occlusion prevent this data-driven approach from always making a unique interpretation.

A more powerful approach is to make use of "higher level" information to aid in the image feature interpretation process. This higher level knowledge can be any information that reduces the ambiguities of the feature extraction process. It can take many forms including contextual information about what is in the scene, or pre-developed computer models stored in the system to help determine the position or location of image objects. This approach is known as the model-driven approach, and in this research, a largely model-driven interpretation is undertaken.

The model-driven approach involves:

(i) developing computer models of a set of objects that are likely to appear in the image

(ii) developing techniques for predicting the appearance of the objects (or portions of those objects) in a given view or situation

(iii) developing techniques for associating these predicted features with features extracted from the image.

Sections 3.2, 3.3, and 3.4 deal with these topics. Section 3.1 discusses the general approach and interactions in more detail.

The model representation is based on the concept of generalized cylinders which represent objects as volume primitives. These primitives can be hierarchically organized to provide alternative levels of detail in prediction and interpretation.

The image feature-to-model matching process predicts invariant image features and the appearance of the modeled object in the range image with a rule-based prediction system. Typical predictions include information about physical edge types (occluding, convex, or concave edges), shapes (length, cylinder contour), linear segment relations (parallel, collinear, connectivity, angle), and possible shape occlusions. The model predictions guide the feature extraction process by providing guidelines for extracting the most useful image features to match the model, and are critical to interpreting partially occluded shapes.

The interpretation process compares the extracted image features with the object model according to a set of rules and produces a "goodness" measure of how well the two features match. The feature-to-model matching can occur on multiple levels (for example, from the edge segment level, component level, and object level up to the contextual level). The spatial relationships between locally matched features are checked for global consistency. Because the prediction and interpretation rules are domain independent (they operate on any model in a generalized cylinder form), our system is applicable to various 3-D classification problems such as vehicle classification, ship classification and robot vision.

3.1 OVERVIEW OF 3-D FEATURE-TO-MODEL MATCHING

The feature extraction techniques discussed in Section 2 extracts 3-D physical features from the laser range image. Once such feature descriptions are available, the recognition problem of intelligently interpreting these descriptions in terms of the object instance still remains. Instead of directly matching the low level laser image features with possible object models, the feature-to-model matching can occur on multiple levels (e.g., the object level, the component level, the edge segment level). Figure 3-1 illustrates this multi-level matching process.

The proposed multi-level matching can be either data-driven or model-driven, or a combination of the two. The data-driven approach proceeds by first extracting image features such as edges from laser

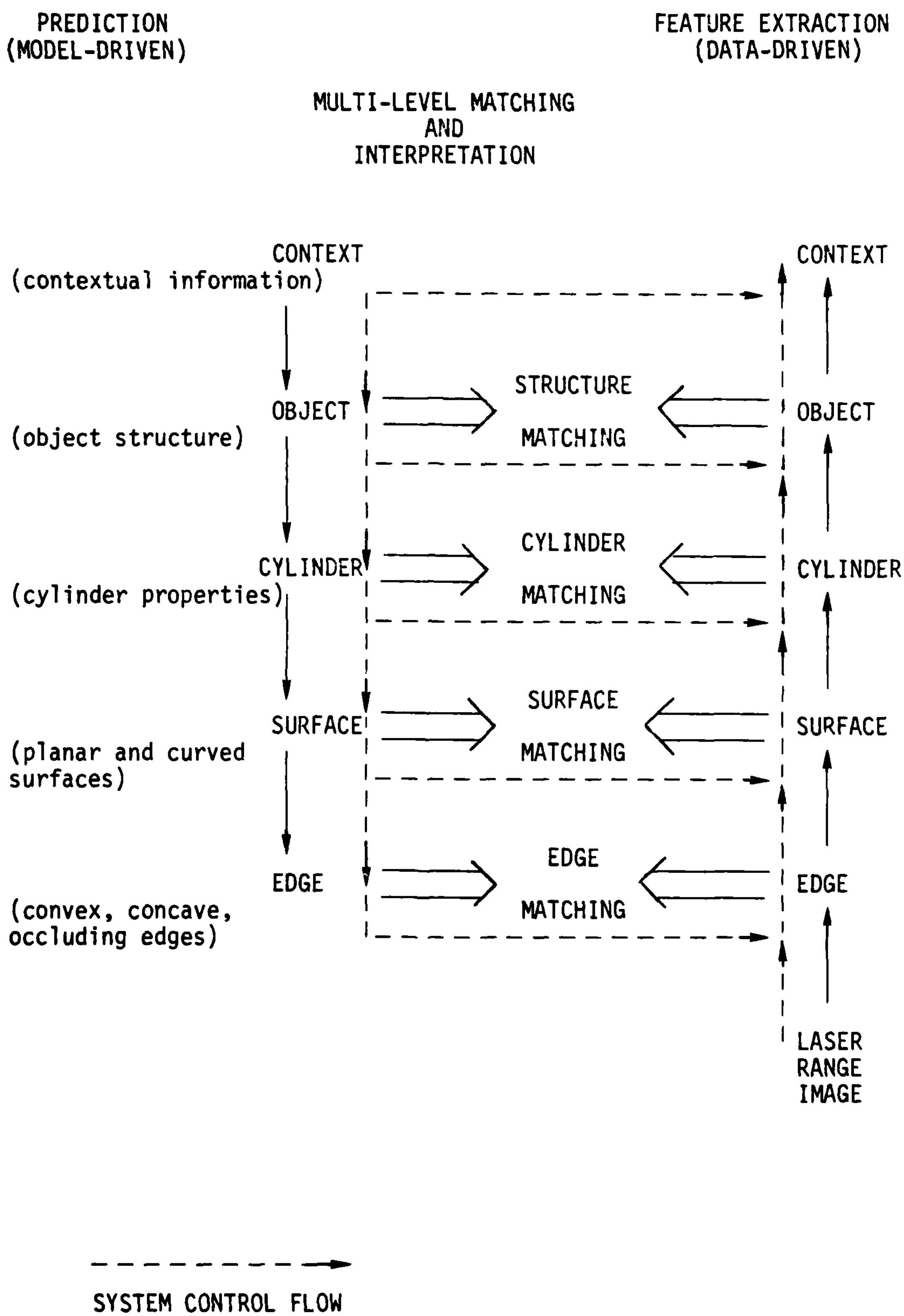


Figure 3-1 Multi-level Matching and Processing Control Flow

imagery, then grouping them into higher level features (e.g., cylinder contours), and finally matching these multi-level features to the object model. In general, the data-driven process requires extensive inferencing capability to resolve ambiguities caused by occlusion and error-prone image descriptions. A typical problem is how to associate edge segments into the contour of a cylinder based on incompletely detected edge segments. Partial occlusion makes this problem more difficult. The model-driven approach on the other hand predicts the appearance of the object in the image from the model, then a goal-directed feature extraction process tries to fit the low level image features to the prediction. In general, a model-driven approach can be more reliable and robust, and it uses knowledge of the model and contextual knowledge of the situation. However, the model-driven approach is inefficient for situations where the class of objects is large (especially if little contextual or collateral information is available).

As indicated above, the processing control can be either data-driven or model-driven. To avoid the inefficiency of the model-driven approach and the ambiguities associated with the data-driven approach, we use a combination of the two. A typical processing example proceeds as follows: An input laser range image is presented to the system. Initial feature extraction algorithms are used to extract an initial set of both low-level and global image features (e.g., physical edges, planar and curved surfaces, projections). This is accomplished without using model knowledge. The global image features (e.g., overall length, orientation, etc.) are compared with the coarsest level of the object model to eliminate unlikely object classes. Then the system searches for a single best-fit cylinder from the detected edge segments using some heuristic rules. If this is successful, the properties of the extracted cylinder (e.g., length, width, volume) and its relative position and orientation with respect to the object coordinate system are used to prune the set of likely models. Component occlusion rarely causes any problem in extracting the first cylinder because at least one cylinder is not occluded. In most situations, the properties of the first cylinder are sufficient for object classification.

At this point (regardless of whether the first-cylinder extraction is successful), the system switches to a model-driven mode to guide the search for high level symbolic features using model predictions and low level image feature data. Predictions such as physical edge types, collinear and parallel relations, cylinder contours, and spatial relations between cylinders are generated from object models. Higher level symbolic features are formed from low level features according to these predictions, and compared with the object model. Incompatible features can generate negative likelihood factors to reduce our confidence in the hypothesized object. Locally matched features are checked for global consistency. Figure 3-1 shows this multi-level image feature-to-model matching and the control flow of the system.

3.2 3-D OBJECT MODEL REPRESENTATION

There are basically two types of representation for 3-D shape recognition: visible surface representation and volumetric representation. The coordinate system used for visible surface representation is viewer-centered, and locations are indicated relative to the viewer. The visible surface representation uses the same types of features as are available in an image and thus provides direct matching capability between model and image features. However, a viewer-centered representation depends on the orientation of the object, and thus requires an enormous number of descriptions for different possible viewpoint positions. The volumetric representation, on the other hand, uses an object-centered coordinate system and is viewpoint-independent. The primitive elements of a volumetric representation are based on more global geometric features such as volumes and cylinders rather than on cumbersome surface details. Since the emphasis of our 3-D object classification system is to classify objects from various viewpoints, a volumetric representation is more suitable.

The volume primitives we use are generalized cylinders [Agin, Binford-76]. A generalized cylinder is defined by a space curve, called the axis, and planar cross-section functions defined on the axis. Figure 3-2 shows examples of typical generalized cylinders. A more complex object (e.g., a tank or missile launcher) is represented in terms of a set of individual cylinders and their spatial relationships to each other. Generalized cylinders are natural representations of elongated shapes which are common in the vehicle targets we are trying to classify. The shape of an object is represented in terms of its distinct components. Both the properties of individual components and their spatial relations are specified for recognition purposes. This model allows segmentation of a complex object into simpler components. The 3-D object model can be represented at several levels of coarse to fine

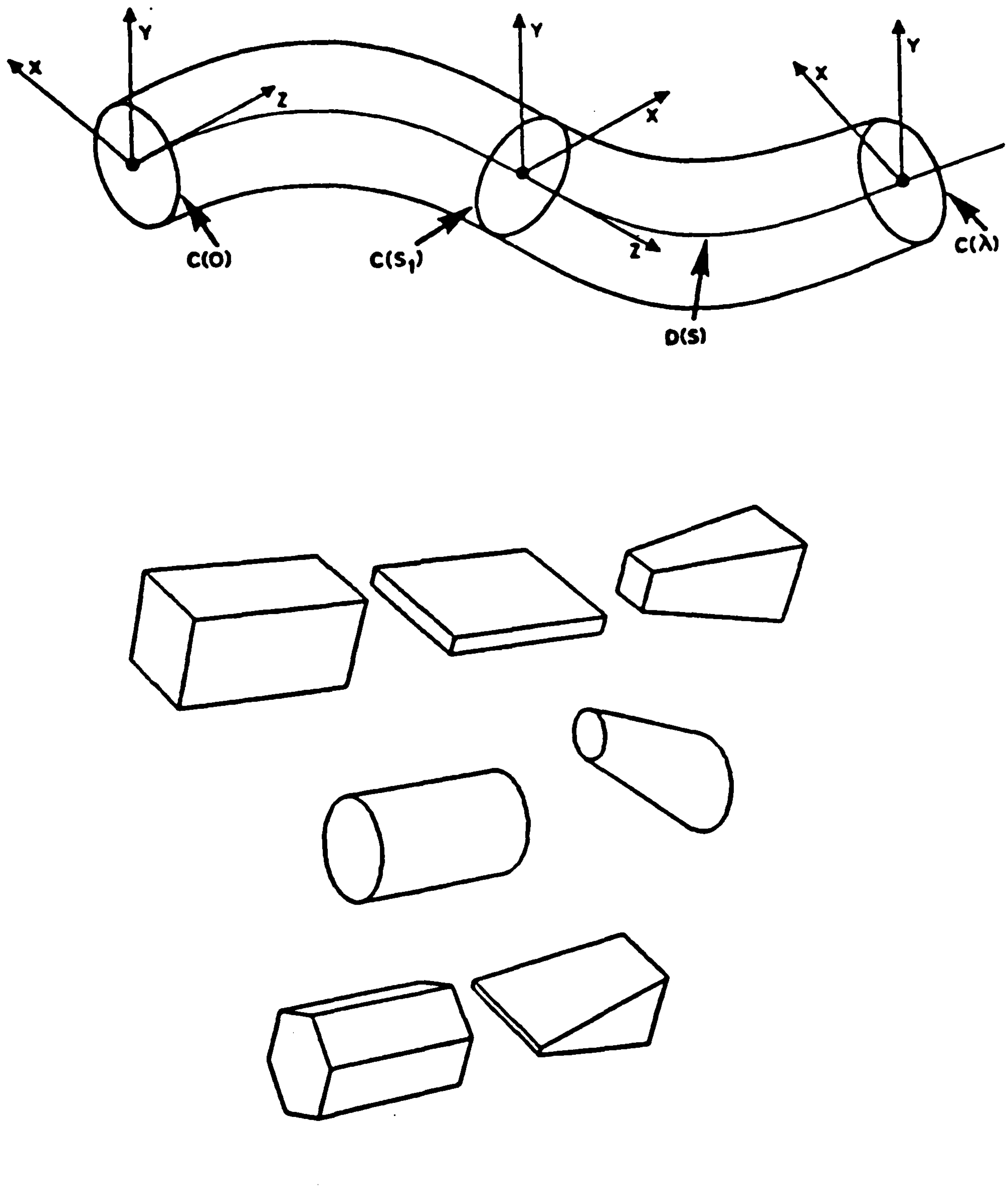


Figure 3-2 Examples of Generalized Cylinders

detail in a hierarchy. This allows for successive levels of increasingly refined analysis. The first level corresponds to the coarsest information of the object. For example in Figure 3-3, the tank is modeled at the top level as a single entity (a box). This top level representation gives global coarse information such as the orientation, volume, height, length and width of the object. At the next lower level, the tank is made up of several major components (gun barrel, gun turret and platform). The locations of components and their spatial relations are defined in the object coordinate system at the next higher level (the top level in this case). Each component has its own local coordinate system and is in turn made up of several smaller components. This hierarchical representation with coarse to fine details enables successful refinement of analysis and also provides a prediction generation mechanism at multiple levels.

The components in the same detail level may vary in importance for recognizing the object. For example, the gun barrel of a tank is unique in vehicle models and provides sufficient evidence to distinguish a tank from a truck or other vehicles. Therefore, to recognize a tank, we may first look for the gun barrel in the image. This kind of knowledge is explicitly implemented in our object models by using a model priority index. Another kind of component priority index is determined by the geometric properties of each cylinder. For example, elongated components and large components show distinct cylinder properties and are easy to distinguish from other components. These distinguished pieces can be used similarly for model selection. The model priority index is viewpoint-independent and provides a mechanism for efficient model

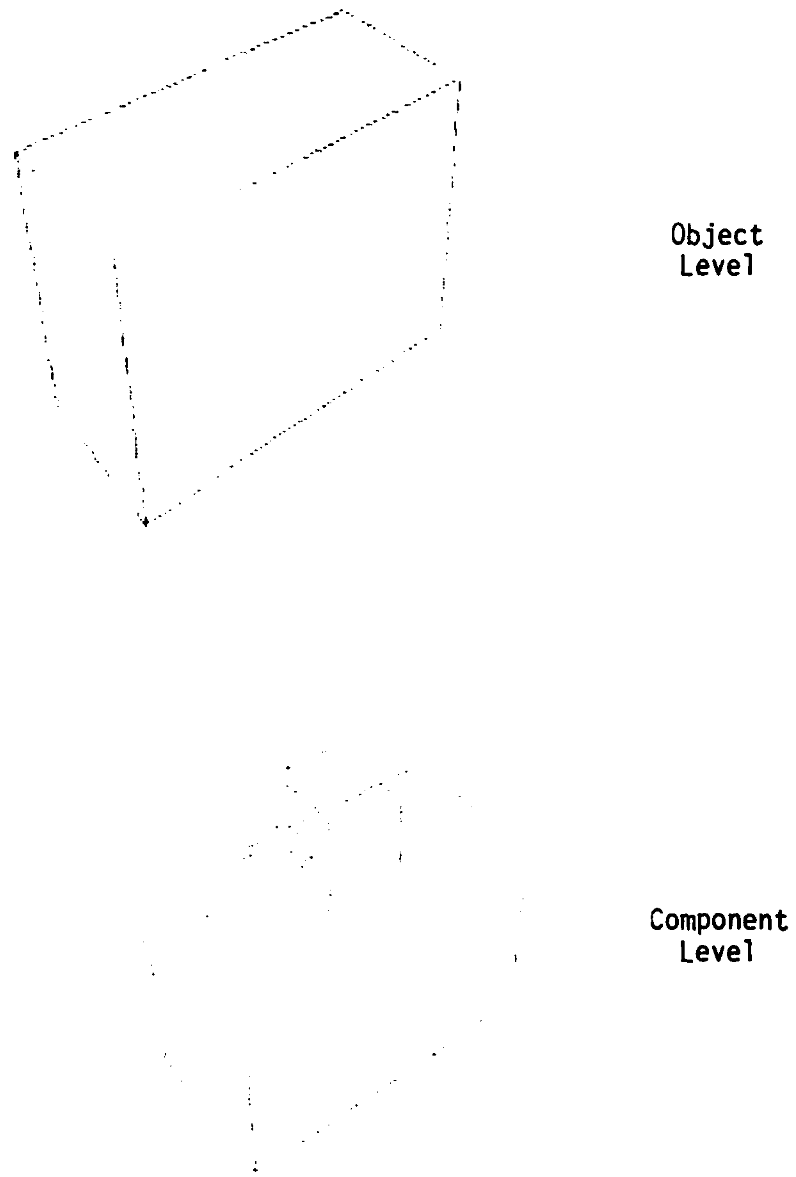


Figure 3-3 Hierarchical Model of a Tank

access and selection. However, this index does not give us any information about the visibility of the component or the ease of cylinder extraction. Therefore, it provides only part of the information needed to select the appropriate feature to extract.

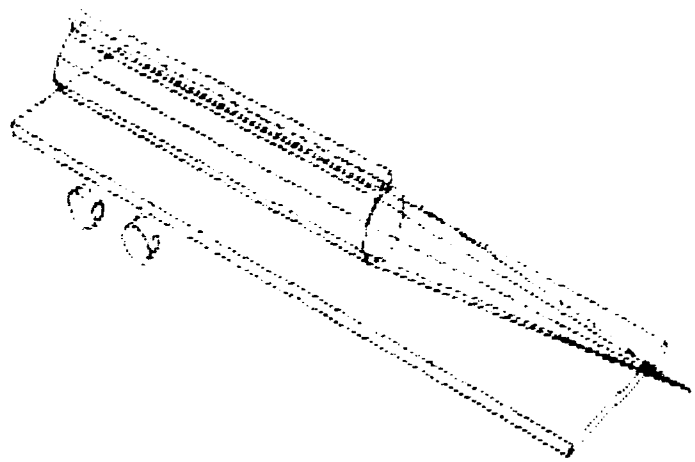
Shapes that are occluded or have low contrast in terms of the sensor are generally more difficult to extract from images. Hence an obscuration priority index has also been introduced to represent the occlusion and visibility relations among components. This index depends on the viewpoint and indicates an order for cylinder extraction based on relative computational simplicity. In our model-driven system, the obscuration priority index allows us to extract a totally visible cylinder first for reliable analysis. This will be discussed in more detail in Section 3.3.2.

As was discussed in Section 2, physical edges such as occluding, convex and concave edges can be distinguished in the range image. The prediction of the physical edge types of a cylinder contour strongly constrain the possible interpretations of each edge segment. Hence our models should facilitate prediction of the physical edge type. To do this, the attachment relations between cylinders in the object are explicitly specified. The convex edges correspond to the internal edges of cylinders, and thus are not useful for segmenting a complex object into simpler components (cylinders). The occluding edges occur at the occluding boundaries of a cylinder that are not attached to any cylinder in the model. The concave edges correspond to the occluding boundaries of a cylinder that is attached to another cylinder. These relationships are useful for physical edge type prediction and will be discussed in more detail in Section 3.3.

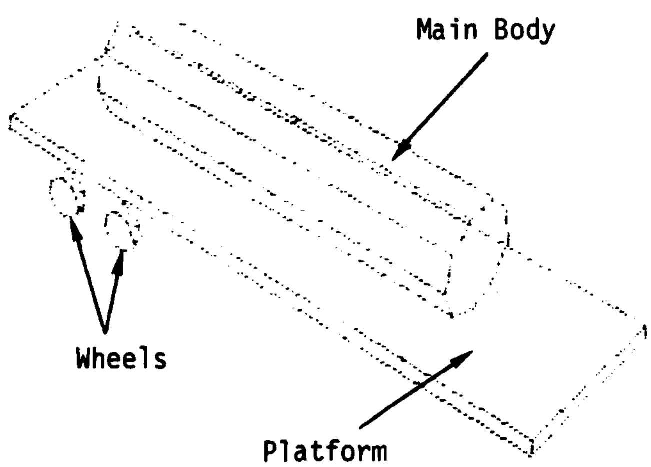
Figure 3-4 shows models of a missile launcher and a decoy. Due to the relatively low resolution of the air-to-ground laser imagery used for these examples, two levels of detail are adequate for target classification (additional levels can be used in higher resolution imagery). Figure 3-5 gives an example of the information stored in the decoy model.

3.3 PREDICTION

Prediction is the process of making estimates about the image appearance of image objects using volumetric models of the object and given some information about the objects' relative position, orientation, and shape in the image. Predictions first give guidance to low level image feature extraction processes for goal-directed shape extraction; then they provide mechanisms for feature-to-model matching and interpretation. The best features for prediction are those invariant features that will always be observable in the image independent of the object's orientation and sensor position. Examples of these invariant features in range imagery are physical edge type (i.e., occluding, convex, concave), surface type (planar, curved), collinear and parallel relations and connectivity. Some invariant features are also independent of the class of objects modeled (e.g., parallel, collinear relations) and thus provide data-driven capability and improve the efficiency of an object classification system.



Component Level Model of Missile Launcher



Component Level Model of Missile Launcher Decoy

Figure 3-4 3-D Models of a Missile Launcher and a Decoy

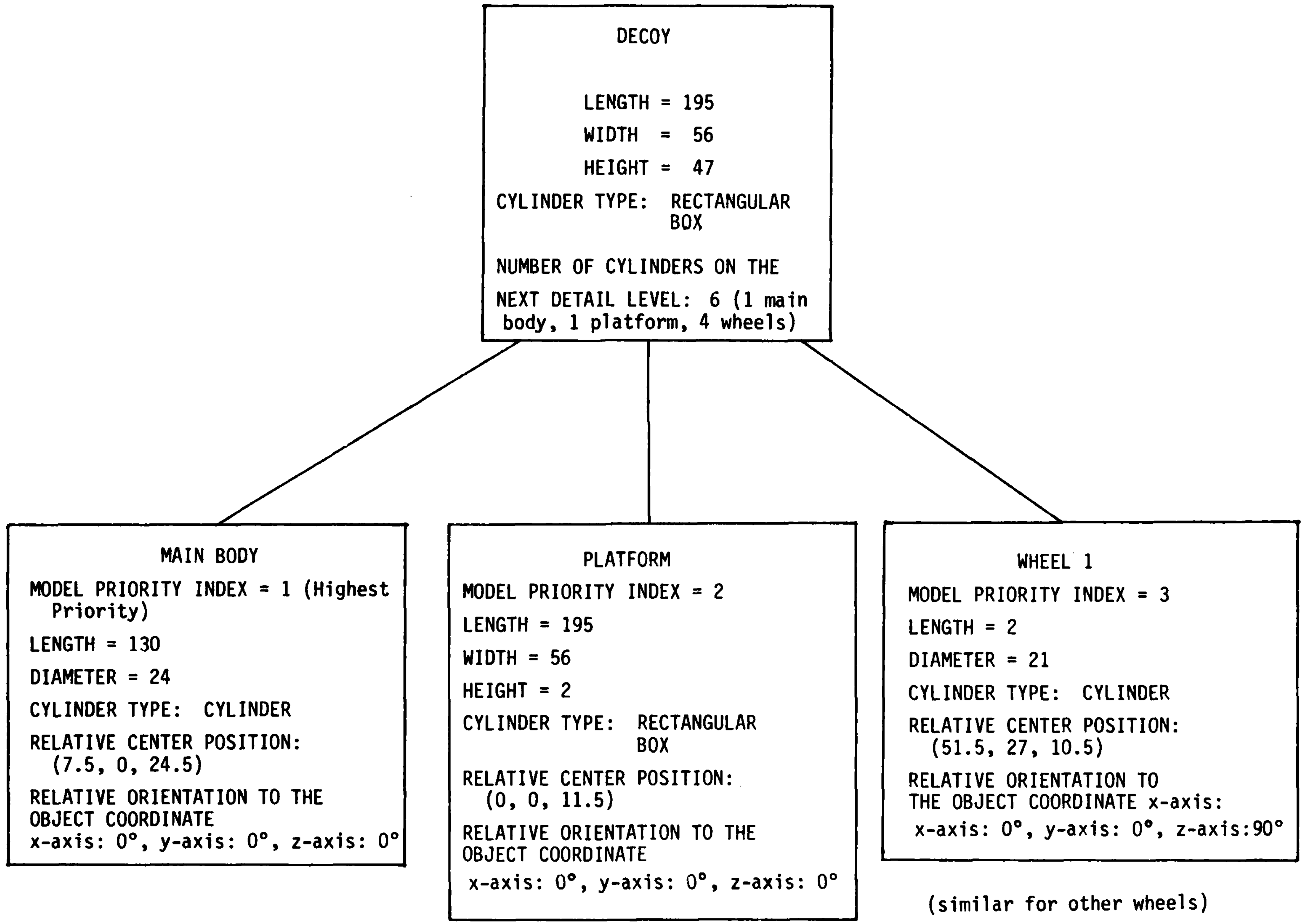


Figure 3-5 Missile Launcher Decoy Model Information

The prediction process proceeds in a top-down fashion. Various rule-based prediction algorithms developed for different levels are discussed in the following subsections.

3.3.1 Context Level Prediction

The context level prediction suggests relationships among objects in the scene. For example, tanks usually form a group and appear together. The classification of one object in the scene as a tank may easily generate a contextual prediction about the object next to it (e.g., another tank). Context level prediction provides strong conditions for searching and verification. This level of prediction is not implemented in our system, because the provided laser imagery used in this research contains only single vehicles and has little textual information.

3.3.2 Object Level Prediction

The object level predictions provide global object features and spatial relations among object components viewed from the sensor. Examples are the symbolic description of the object contour in the image, the spatial relationships between object components and the occluding relations among object components.

The occluding object contour can be predicted once we know the sensor position and the object orientation. In the range image, the object contour can be extracted from range jump boundaries or the object segment obtained from object-ground segmentation. Object classification can be performed by matching the extracted object boundary with the model prediction. However, this method only works for cases without object occlusion and when the objects considered have distinct boundary contours. For example, a missile launcher and a decoy are vehicle objects modeled in this research that have similar occluding contours when viewed from certain directions. These objects can only be classified according to their internal components' structure.

The model priority index discussed in Section 3.2 is a viewpoint-independent description of the importance of each cylinder for object recognition. Some components may not be visible in the range image due to self-occlusion. Since distinguished components (e.g., the gun barrel of a tank) may not be visible and occluded shapes are generally difficult to extract, we need to have a viewpoint-dependent obscuration priority index to indicate the ease of cylinder extraction in the range image. The obscuration priority index is similar to the priority algorithm [Newman, Sproull-73] used for hidden surface elimination. The idea is to arrange all cylinders in the scene in priority order based on their depth and obscuration relations. Cylinders nearer to the viewpoint and not obscured by other cylinders will have higher priority indices. Cylinders that are totally obscured are explicitly indicated, and no effort is spent trying to find them in the range image. This obscuration priority index is purely geometrical and determined from the object orientation and viewpoint. The combination of model priority index and obscuration priority index gives a new priority order that not only indicates the importance of a particular cylinder for object recognition, but also compares the ease of cylinder extraction in the range image. We currently use the obscuration priority index as the base index for cylinder extraction and matching. If two cylinders have the same obscuration priorities (they don't obscure each other), then the model priority indices are compared. In general, the priority index is valid for an interval of viewing angle and thus can be used based on rough object orientation estimates.

The structural relationships among object components are the basis for object recognition. The object level prediction indicates relationships between object components. The nodes contain both 3-D properties of each cylinder (e.g., volume, 3-D relative position and orientation in the object coordinate system) and their 2-D image properties (e.g., cylinder length, width and visibility). The arcs specify the spatial relations between cylinders (e.g., connectivity, obscuration, support and relative angle). The object prediction information is used for global structure matching after we extract and analyze the individual cylinders of an object.

3.3.3 Cylinder-level Prediction

Cylinder-level predictions provide goal-directed guidance for cylinder extraction from low level image features. This is the most important prediction level in our system because cylinders are the basic symbolic primitives we use to perform image feature-to-model matching.

A generalized cylinder is hierarchically characterized by first defining its axis and then the cross-sections along the axis. For the man-made vehicle objects we are trying to classify, the cylinder axis is a generally straight line segment, and the cross-sections are generally uniform and at most linearly varying. At the cylinder level, we predict the shape of a cylinder in the range image. Due to the internal structure of the objects we modeled and the general air-to-ground imaging geometry, only a few cylinder contours will be totally visible (cylinders with a high obscuration priority index), and most cylinders are partially obscured. The prediction of occluded cylinder contours in the image can be generated by polygon clipping algorithms [Weiler, Antherton-77]. In this case, the cylinder extraction process guided by predictions corresponds to finding a complete, closed polygon (or several polygons) from the detected edge segments. This polygon matching approach is cumbersome and difficult because of complex shape matching and possible missing edge segments. Here, cylinder level prediction is accomplished by using a hierarchy in defining generalized cylinders. The properties of the two cylinder boundaries along the cylinder axis are first predicted. These predictions include parallel relations (relative angles in general), physical edge types (concave, occluding), length, distance between two segments, missing parts due to occlusion, and collinear relations between disjoint segments. These predictions are sufficient to guide the coarse extraction of cylinders. After extracting the two cylinder boundaries along the axis, other boundaries on the cylinder contour can be predicted in limited regions relative to the two major boundaries, and heuristic rules can be used to construct the cylinder contour from incomplete edge segments.

3.3.4 Surface-level Prediction

Surface-level predictions provide the cylinder surface appearances and their spatial relations in the image. Due to the low resolution nature of air-to-ground laser imagery, surface properties are not easy to extract and to use (sometimes a single surface patch only has a few points). However, for industrial applications where high resolution range images are available, surface level predictions impose strong constraints on cylinder extraction. The physical edge angle images discussed in Section 2 can be used for planar and curved surface extraction. These surface primitives can then be grouped together to form a cylinder according to surface level predictions from the model.

3.3.5 Edge Level Prediction

Edge-level predictions assign physical edge types to each edge segment and thus strongly constrain the possible interpretation of each edge segment for cylinder contour extraction.

Prediction of the physical edge type of a cylinder contour is made possible by explicitly specifying the attachment relations between cylinders in the model. For example, if cylinder A is supported by cylinder B, the two touching faces of the cylinders are explicitly labeled in the object model. The occluding edge type is predicted for those cylinder contour segments that do not belong to a labeled face. The concave edge type is predicted for those segments that belong to a labeled face, and are inside the other surfaces with the same label.

The convex edges correspond to the internal edges of cylinders, and thus are not useful for cylinder extraction since we don't use surface level predictions. The physical edge type limits the search space for cylinder grouping, but more importantly, it can be used to verify the correctness of the extracted cylinder. Therefore, we have more confidence in the cylinder extraction because not only the geometric properties of edge segments are used, but also the physical properties of edge segments are examined.

3.4 INTERPRETATION

Interpretation proceeds by comparing the image features on multiple levels to the object models according to a set of if-then rules (a production system). Each rule for comparison produces a "goodness" measure of the system's confidence in how well the two features match. If a single object model has a much larger likelihood than others, the target in the range image is classified as an instance of that object. Besides the classification of the object, object position and orientation information are also available and can be used for higher level scene interpretation.

1. OBJECT LENGTH

<u>If</u> FEATURE_LENGTH > MODEL_LENGTH + DELTA <u>or</u>

FEATURE_LENGTH < MODEL_LENGTH - DELTA

<u>then</u> -0.5

<u>else</u> $0.5 \left(1 - \frac{|FEATURE\_LENGTH - MODEL\_LENGTH|}{MODEL\_LENGTH}\right)$

2. OBJECT WIDTH

<u>If</u> FEATURE_WIDTH > MODEL_WIDTH + DELTA <u>or</u>

FEATURE_WIDTH < MODEL_WIDTH - DELTA

<u>then</u> -0.5

<u>else</u> $0.5 \left(1 - \frac{|FEATURE\_WIDTH - MODEL\_WIDTH|}{MODEL\_WIDTH}\right)$

3. OBJECT HEIGHT

<u>If</u> FEATURE_HEIGHT > MODEL_HEIGHT + DELTA <u>or</u>

FEATURE_HEIGHT < MODEL_HEIGHT - DELTA

<u>then</u> -0.5

<u>else</u> $0.5 \left(1 - \frac{|FEATURE\_HEIGHT - MODEL\_HEIGHT|}{MODEL\_HEIGHT}\right)$

4. MINIMUM HEIGHT POINT LOCATION

<u>If</u> NUM.MODEL_MINHEIGHT_REAR > 0 <u>and</u>

NUM.FEATURE_MINHEIGHT_REAR > 0

<u>then</u> 0.5

<u>else</u> -0.5

Figure 3-6 Sample Set of Interpretation Rules

5. FRONT EXTREME POINT HEIGHT

<u>If</u> FEATURE_FRONTEXTREME_HEIGHT > MODEL_FRONTEXTREME_HEIGHT + DELTA

<u>or</u>

FEATURE_FRONTEXTREME_HEIGHT < MODEL_FRONTEXTREME_HEIGHT - DELTA

<u>then</u> -0.5

<u>else</u> $0.5 \left(1 - \frac{|\text{FEATURE\_FRONTEXTREME\_HEIGHT} - \text{MODEL\_FRONTEXTREME\_HEIGHT}}{\text{MODEL\_FRONTEXTREME\_HEIGHT}}\right)$

6. REAR EXTREME POINT HEIGHT

<u>If</u> FEATURE_REAREXTREME_HEIGHT > MODEL_REAREXTREME_HEIGHT + DELTA

<u>or</u>

FEATURE_REAREXTREME_HEIGHT < MODEL_REAREXTREME_HEIGHT - DELTA

<u>then</u> -0.5

<u>else</u> $0.5 \left(1 - \frac{|\text{FEATURE\_REAREXTREME\_HEIGHT} - \text{MODEL\_REAREXTREME\_HEIGHT}|}{\text{MODEL\_REAREXTREME\_HEIGHT}}\right)$

Figure 3-6 Sample Set of Interpretation Rules (Continued)

The set of rules for interpreting image features in terms of models can be classified into two classes according to the level of detail they compare. The first class of rules looks for general features and global characteristics; i.e., the object level features such as the object length, width, and height. Since the 3-D surface data can be obtained from the range image through a coordinate transformation, the actual length measurements are available and can be compared directly with the model parameters. A typical rule for object length matching is shown in the first rule of Figure 3-6. The rule assigns a negative likelihood to models that exceed the tolerance interval DELTA and prunes these objects from further consideration. For those object models within the tolerance interval, the rule returns a likelihood value as the goodness measure. Another set of general features is the extreme positions of the image object. For example, the lowest components of a truck are its wheels. Therefore, from a thin cross-section slice of the target, a segment with low z value (elevation) will probably show the wheel positions relative to the overall object length. The fourth rule in Figure 3-6 is one example. Other extreme positions such as the heights of the front and rear points of the side-view projection image can also be used for comparison. These extreme position rules usually pinpoint the end positions of cylinders and check for general structural features, and can be valuable even for partially occluded objects. Note that these rules are domain-independent and are derived from the object models. Since we are not sure which end of the image feature is the vehicle front, we need to hypothesize two possibilities (front and rear), and the results of an extreme positions match usually prune some more unlikely models and resolve the front-rear ambiguities.

The second class of rules compares finer object details at the cylinder level. The system first tries to extract a single cylinder from range image by some heuristic rules and predictions of invariants (e.g., antiparallels, angles). Once a cylinder is extracted, its 3-D properties (length, width, length, and volume) and relative position and orientation in the object coordinate system are compared with the model. These cylinder level features not only provide finer detail for feature to model matching, but also put strong constraints on the internal structure of the object. These constraints are often sufficient to make a unique interpretation of the image.

3.5 PROCESSING EXAMPLES

To assess the feasibility and capability of rule-based interpretation for classifying vehicle targets from extracted 3-D features, a sample set of rules was developed and tested on extracted image feature information obtained using the techniques described in Section 2. Figure 3-6 represents this set of rules. They are not meant to be comprehensive, just representative. The rules are applied in sequence according to the rule number. This sequential process is used to prune possible models at the early interpretation stage with the most important features and resolve the object front-rear ambiguity in the beginning of analysis.

A laser range image of a decoy was processed using the sequence of analysis steps shown in Figure 2-1. A summary of the extracted feature information is shown in Figure 3-7. The rules in Figure 3-6 were then used to compare this extracted feature information about the decoy with two vehicle models. These were a decoy and a missile launcher. They are chosen because their similarities in size and structure make the selection and correct classification a non-trivial task. The relevant model information for these two objects is obtained from their volume models and is shown in Figure 3-8.

The results of applying the rules in Figure 3-6 to these vehicle models are shown in Figure 3-9. The first three rules check the object length measurements and the system prefers the decoy slightly. No classification can be made at this stage. Rules 4 to 6 check the general features of each model and try to resolve the object front rear ambiguity. Rule 4 compares the relative positions of object points with the minimum z coordinate and gives likelihood values to the front and rear hypotheses of the same object. Both the missile launcher front and decoy front are favored because the bottom points are on the rear part of the segmented object and the model wheels are on the rear part of the vehicle. Rule 5 and Rule 6 check the height (z coordinate) of the extreme points (front and rear) to figure out major structural differences. Note that we used the relative z coordinate difference between the extreme points and the maximum z coordinate value of the segmented object, because we are not certain about the ground level. The rule matching results apparently favor the decoy model with the model front corresponding to the feature front. Reasonable classification is achieved at this level.

OBJECT ORIENTATION = 45 degrees

OBJECT LENGTH = 199.24

OBJECT WIDTH = 55.86

OBJECT HEIGHT = 38

MINIMUM OBJECT POINT HEIGHT = 10

MAXIMUM OBJECT POINT HEIGHT = 48

MINIMUM HEIGHT POINT RELATIVE POSITION

IN SIDE-VIEW PROJECTION

= 2 at rear (from 60 ~ 100%) of vehicle

positions (along the vehicle length direction): 152.83, 156.66

EXTREME POINTS (FRONT-REAR) HEIGHT

IN SIDE-VIEW PROJECTION

= front extreme has height 24

rear extreme has height 21

(Resolution: 2 inches per unit)

Figure 3-7 Extracted Feature Information from Decoy Image

MODEL INFORMATION

RULE	MISSILE LAUNCHER	DECOY
1. OBJECT LENGTH	225	195
2. OBJECT WIDTH	48	56
3. OBJECT HEIGHT	60	47
4. MINIMUM HEIGHT POINT LOCATION	MIDDLE (40-60%), REAR (60-100%)	REAR (60-100%)
5. FRONT EXTREME POINT HEIGHT	12 (measured from the top of the missile launcher)	25 (measured from the top of the decoy)
6. REAR EXTREME POINT HEIGHT	25 (measured from the top of the missile launcher)	25 (measured from the top of the decoy)

RESOLUTION: 2.0 inches per unit

Figure 3-8 Relevant Model Information

RULE	MISSILE LAUNCHER		DECOY	
1. OBJECT LENGTH	0.443		0.489	
2. OBJECT WIDTH	0.418		0.499	
3. OBJECT HEIGHT	0.400		0.489	
SUBTOTAL	1.261		1.477	
	MISSILE LAUNCHER FRONT	MISSILE LAUNCHER REAR	DECOY FRONT	DECOY REAR
4. MINIMUM HEIGHT POINT LOCATION	0.5	-0.5	0.5	-0.5
5. OBJECT FRONT EXTREME POINT HEIGHT	-0.5	-0.5	0.48	0.48
6. OBJECT REAR EXTREME POINT HEIGHT	-0.5	-0.5	0.46	0.46
TOTAL	0.711	-0.239	2.917	1.917

Figure 3-9 Likelihood Weights Associating Rules and Object Models

If the gross object features and major structures do not provide sufficient evidence for classificaiton, the system tries to extract finer details at the component level. Figure 3-10 is a reproduction of Figure 2-19, and the occluding exterior boundaries are in solid line, and the concave edge segments are in dashed line. The cylinder extraction algorithm first finds the major cylinder boundaries along the cylinder axis by using the prediction that these two segments will appear parallel in the image and one of them is a concave edge. This prediction strongly constrains the possible edge segments for the major cylinder boundaries. The cylinder extraction algorithm successfully finds that two edge segments (with labels A and B in Figure 3-10) satisfy the cylinder prediction and that they have a significant amount of overlap between them. Using these two edge segments as two sides, a closed polygon is formed and used as an approximation of the cylinder contour. The length, width, and height of this cylinder can be extracted by the same techniques used for gross object feature extraction. Its length measurements, relative orientation and position to the object coordinate system are shown in Figure 3-11. These features strongly restrict the possible object models for matching.

Another example is demonstrated using the missile launcher image in Figure 2-9(a). The extracted image feature information is shown in Figure 3-12 and the interpretation results are shown in Figure 3-13. The system correctly classified the image as the front view of the missile launcher. For cylinder level fine detail analysis, the exterior

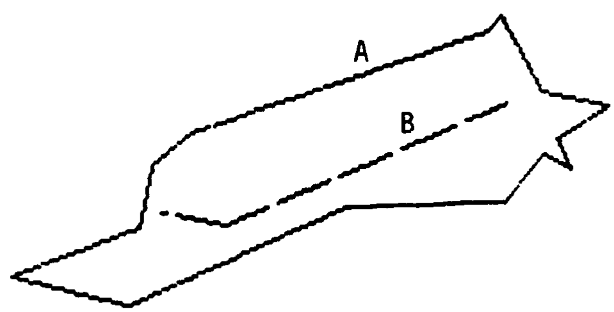


Figure 3-10 Line Segments of Occluding and Concave Edges

occluding boundaries and concave edge segments were extracted from the image and are shown in Figure 3-14. The system successfully extracted the cylinder bounded by edge segments A and B, but fail to extract the major missile component due to missing edges. Future research efforts are required to make the cylinder extraction algorithm more robust and to allow the physical edge detector to operate in a prediction mode, i.e., first predict missing edges then drive the edge detector to find them.

CYLINDER FEATURE INFORMATION

ORIENTATION = 54 degrees

RELATIVE ORIENTATION TO THE OBJECT = 9 degrees

LENGTH = 136

WIDTH = 26.48

HEIGHT = 30 (minimum height point = 18, maximum height point = 48)

RELATIVE CENTER POSITION (TO THE OBJECT CENTER POSITION (0,0,0))
= (14, 4, 4)

Figure 3-11 Extracted Cylinder Feature Information from Decoy Image

OBJECT ORIENTATION = 54 degrees

OBJECT LENGTH = 232.12

OBJECT WIDTH = 45.66

OBJECT HEIGHT = 49

MINIMUM OBJECT POINT HEIGHT = 12

MAXIMUM OBJECT POINT HEIGHT = 61

MINIMUM HEIGHT POINT RELATIVE POSITION

IN SIDE-VIEW PROJECTION

= 3 at rear (from 60 ~ 100%) of vehicle

positions (along the vehicle length direction): 200.66, 168.46, 170.96

1 at middle (from 40 ~ 60%) of vehicle

position (along the vehicle length direction): 110.38

EXTREME POINTS (FRONT-REAR) HEIGHT

IN SIDE-VIEW PROJECTION

= front extreme has height 49

rear extreme has height 26

(Resolution: 2 inches per unit)

Figure 3-12 Extracted Feature Information from Missile Launcher Image

RULE	MISSILE LAUNCHER		DECOY	
1. OBJECT LENGTH	0.484		-0.5	
2. OBJECT WIDTH	0.476		0.408	
3. OBJECT HEIGHT	0.492		0.372	
SUBTOTAL	1.452		0.280	
	MISSILE LAUNCHER FRONT	MISSILE LAUNCHER REAR	DECOY FRONT	DECOY REAR
4. MINIMUM HEIGHT POINT LOCATION	0.5	-0.5	0.5	-0.5
5. OBJECT FRONT EXTREME POINT HEIGHT	0.5	-0.5	-0.5	-0.5
6. OBJECT REAR EXTREME POINT HEIGHT	0.46	-0.5	-0.5	-0.5
TOTAL	2.912	-0.048	-0.220	-1.220

Figure 3-13 Likelihood Weights Associating Rules and Object Models

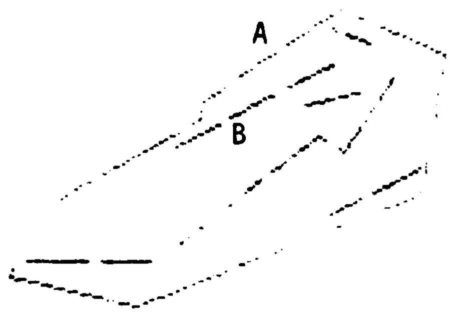


Figure 3-14 Line Segments of Occluding and Concave Edges

4. SUMMARY AND CONCLUSIONS

This document addresses AI&DS's research effort to develop 3-D object classification techniques for laser range imagery. Our conclusion is that even though there are several remaining research issues, laser range imagery does provide more useful information than intensity imagery and it is feasible to build an automated 3-D object classification system. The techniques which we have examined from the fields of artificial intelligence and image understanding have shown promise of providing assistance in achieving this goal. Techniques for extracting 3-D image features were developed and estimates of gross object features such as object length, width, height, object orientation, and major structure were obtained. Fine object features at component level such as component length, width, height, and its relative orientation and position to the object were obtained through a cylinder extraction process based on the detected physical edges. An initial knowledge data base of information about three military vehicles (i.e., tank, missile launcher, and decoy) was developed. Prediction rules and interpretation rules were formulated to associate the extracted image features with the data base. Using a few synthetic range images, the rules were tested on automatically extracted information and proper object classification was achieved. These results are encouraging and are optimistic indications of the potential power of the technology.

Efforts to develop and test object-ground segmentation techniques and techniques indicate that gross object features can be obtained. These features will also provide information to drastically prune the list of possible alternatives by checking the general structural features of the models. However, some analysis of fine detail in the image is required to classify objects with similar general structures and to have more robust and reliable performance. The cylinder extraction technique used the model prediction rules as guidelines to extract fine object detail at the component level (cylinders) from low level edge features. As illustrated in Section 3.5, this fine feature usually provides sufficient evidence to make a final classification.

The above conclusion should be considered with the knowledge that only the laser range imagery was analyzed and no external object occlusion was assumed (component occlusion by other vehicle parts was considered). In the air-to-ground laser sensor, the laser reflectance imagery registered with the range imagery is also available. Even though the reflectance imagery is noisy due to the coherent nature of the laser sensor, adequate analysis based on the object-surface reflectivity model can provide useful information to aid object classification. This aspect was not discussed in our work, and it is a good direction for future research. Another important issue is object occlusion. Our gross feature extraction techniques may not provide meaningful information for occluded objects. A flexible control process is desirable to bypass this useless processing and to analyze available features (e.g., cylinder features, fine details).

AI&DS believes that with more research efforts, the construction of an automated 3-D object classification system is feasible and in many applications has the potential for exceeding human performances or the performance of other imaging sensors in classifying objects. An automated system can take advantage of the rules and knowledge used by human analysts while at the same time using 3-D information effectively and using the more precise measurement capabilities available with a machine. The system also has the memory capabilities to consider and manipulate multiple hypotheses or explanations and quickly evaluate or update them. The critical issue is the development of a more flexible control process for bottom-up and top-down processing.

As has been mentioned above, several key research issues remain to be explored before the development of an automated 3-D object classification system for laser range imagery. This research has partially addressed some of these tasks but further work is still required. These include:

- Extraction and interpretation of finer image features
  - extraction of weak edge features using prediction
  - refinement of cylinder extraction algorithm
  - extension of cylinder extraction techniques to deal with partially occluded cylinders
- Use of laser reflectance imagery and object surface reflectivity modeling

- Development of robust interpretation and prediction rules, heuristics and algorithms
- Development of flexible processing mechanisms and more sophistication control techniques
  - effective use of combined top-down and bottom-up processing
  - opportunistic control to analyze available features (occlusion case)

The development of 3-D object classification techniques has been an interesting and challenging effort. We gained much insight in 3-D data analysis. We believe that a prototype classification system should be developed and tested. Because our approach is domain-independent and somewhat independent of the types of range sensor used, it can have wide applications to vehicle, ship and aircraft classification, robot vision, and in the use of autonomous vehicles.

5. REFERENCES

[Agin, Binford-76] Agin, G.J., and Binford, T.O., "Computer Description of Curved Objects," IEEE Trans. Computer, C-25, pp. 439-449, 1976.

[Brooks-81] Brooks, R.A., "Symbolic Reasoning Among 3-D Models and 2-D Images," Artificial Intelligence, Vol-17, pp. 285-348, 1981.

[Duda, Hart-73] Duda, R.O, and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, Inc., New York, 1973.

[Duda, et al.-79] Duda, R.O., Nitzan, D., and Barrett, P., "Use of Range and Reflectance Data to Find Planar Surface Regions," IEEE Trans. PAMI, Vol. 1, pp. 257-271, July 1979.

[Nevatia, Babu-80] Nevatia, R., and Babu, K.R., "Linear Feature Extraction and Description," Computer Graphics and Image Processing, Vol-13, pp. 257-269, 1980.

[Nevatia, Binford-77] Nevatia, R., and Binford, T.O., "Description and Recognition of Curved Objects," Artificial Intelligence, Vol. 8, pp. 77-98, February 1977.

[Newman, Sproull-73] Newman, W.M., and Sproull, R.F., Principles of Interactive Computer Graphics, McGraw-Hill, New York, 1973.

[Oshima, Shirai-79] Oshima, M., and Shirai, Y., "A Scene Description Method Using Three-Dimensional Information," Pattern Recognition, Vol. 11, pp. 9-17, 1979.

[Rosenfeld, Kak-76] Rosenfeld, A., and Kak, A.C., Digital Image Processing, Academic Press, New York, 1976.

[Sobel-70] Sobel, I., "Camera Models and Machine Perception," Stanford AI project, Memo AIM-121, Computer Science Department, Stanford University, May 1970.

[Sugihara-79] Sugihara, K., "Range Data Analysis Guided by a Junction Dictionary," Artificial Intelligence, Vol. 12, pp. 41-69, 1979.

[Weiler, Atherton-77] Weiler, K., and Atherton, P., "Hidden Surface Removal Using Polygon Area Sorting," Computer Graphics, Vol-11, pp. 214, Summer 1977.

PUBLICATIONS

Papers submitted to technical conferences as a result of this research are:

- "Three-Dimensional Feature Extraction," submitted to Computer Vision and Pattern Recognition Conference, to be held June 21-23, 1983
- "Model-Based Interpretation of Range Imagery," submitted to the 1983 National Conference on Artificial Intelligence, to be held August 22-26, 1983

SUMMARY OF KEY PERSONNEL

Darwin T. Kuan, Ph.D., Electrical Engineering, University of Southern California, 1982. Thesis title: Nonstationary Recursive Restoration of Images with Signal-Dependent Noise with Application to Speckle Reduction.

Robert J. Drazovich, M.S., Operations Research, Stanford University, 1975.

END

FILMED

6-83

DTIC