RL-TR-91-304
Final Technical Report
November 1991

SYMBOLIC IMAGE UNDERSTANDING

University of Massachusetts

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. 6068

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-91-304 has been reviewed and is approved for publication.

APPROVED:

LEE A. UVANNI
Project Engineer

FOR THE COMMANDER:

GARRY W. BARRINGER
Technical Director
Intelligence & Reconnaissance Directorate

SYMBOLIC IMAGE UNDERSTANDING

Ed Riseman
Al Hanson

Contractor: University of Massachusetts
Contract Number: F30602-87-C-0140
Effective Date of Contract: 1 September 1987
Contract Expiration Date: 31 August 1990
Short Title of Work: Symbolic Image Understanding
Program Code Number: 8E20
Period of Work Covered: Sep 87 - Aug 90

Principal Investigator: Ed Riseman & Al Hanson
Phone: (413) 545-2450

RL Project Engineer: Lee Uvanni
Phone: (315) 330-4863

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by Lee Uvanni, RL (IRRE) Griffiss AFB NY 13441-5700 under Contract F30602-87-C-0140.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1991	Final Sep 87 - Aug 90

4. TITLE AND SUBTITLE
SYMBOLIC IMAGE UNDERSTANDING

5. FUNDING NUMBERS
C - F30602-87-C-0140
PE - 62301E
PR - F068
TA - 00
WU - 01

6. AUTHOR(S)
Ed Riseman and Al Hanson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Massachusetts
Dept of Computer & Information Science
Amherst MA 01003

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency
1400 Wilson Blvd
Arlington VA 22209

Rome Laboratory (IRRE)
Griffiss AFB NY 13441-5700

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-91-304

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Lee A. Uvanni/IRRE/(315) 330-4863

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
This report summarizes the work done at the University of Massachusetts in the area of Image Understanding of static scene domains. The goal of this research project was to demonstrate a practical knowledge-based approach to computer vision that utilized multiple levels of parallel processing. Some problems dealt with included extracting lines, grouping together regions and lines from the same surface, computing pose from geometry, and matching corresponding features in different views. Special attention was also given to the extraction of curved lines and curved surfaces. The static scene domains consisted of both ROAD and HOUSE scenes, each containing approximately fourteen (14) different objects per scene.

14. SUBJECT TERMS
Image Understanding, Knowledge-based object recognition, Model Matching, Tokens, Schemas, Knowledge Sources

15. NUMBER OF PAGES
40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

1 Overview

This report summarizes the progress in image understanding of static scenes at the University of Massachusetts for the period September 1, 1987 to August 31, 1990. This work was supported by contract F30602-87-C-0140 and partially by DARPA contract DACA76-89-C-0017. The goal of this research program is the development and practical demonstration of a knowledge-based approach to computer vision, which can take advantage of parallel processing on multiple levels.

One of the main aspects of our work is knowledge-based object recognition. Knowledge-based object recognition itself has many aspects including geometric model matching [Beveridge, Weiss, et al. 1989; Beveridge, Weiss, et al. 1990], computing pose from geometry [Kumar and Hanson 1989a; Kumar and Hanson 1989b; Kumar and Hanson 1990a; Kumar and Hanson 1990b], indexing to determine which models to match [Burns 1987], and general recognition strategies [Draper, Collins et al. 1989].

A prerequisite for model-based object recognition is having models of the important objects. Many of the same problems that are found in the object recognition task are also present in the model construction task, e.g. extracting lines [Dolan and Weiss 1989; Boldt, Weiss et al. 1989], grouping together regions and lines that are part of the same surface [Williams 1990], and matching corresponding features in different views [Williams and Hanson 1988]. Our work also includes these aspects with particular attention to extraction of curved lines [Dolan and Weiss 1989] and surfaces [Giblin and Weiss 1987; Grupen, Weiss et al. 1990b; Grupen, Weiss et al. 1990a].

2 Object Recognition

In this section, we will first give an overview of techniques relevant to object recognition and then give the details of each. A general system for object and scene interpretation, called the Schema System, has evolved as part of a long-term research effort at UMass (Draper 1989; Draper and Riseman 1990; Draper, Brolio et al. 1989; Hanson and Riseman 1978; Hanson and Riseman 1987; Riseman and Hanson 1984). The results of successful experiments in the outdoor scene domain has led to the conclusion that a declarative representation of knowledge is easier to work with than a procedural representation especially when developing automatic mechanisms for learning object recognition strategies (Hanson and Riseman 1989).

A particular class of object recognition problems is recognition from specific geometric models. An approach developed by Beveridge, Weiss, et. al [Beveridge, Weiss, et al. 1989; Beveridge, Weiss, et al. 1990] matches straight lines extracted from an image with model lines projected to the image plane using an assumed location of the camera. This two-dimensional matching scheme uses local search over the space of correspondences and solves for scale, rotation, and translation parameters of the model which produce the best fit to the data. The model-to-image line correspondences are then used as input to a 3D pose computation. The 3D pose refinement technique (Kumar and Hanson 1989a; Kumar and Hanson 1989b;

Kumar and Hanson 1990a; Kumar and Hanson 1990b) has been developed to work in the presence of outliers, i.e. incorrect matches. The robustness is achieved at some computational cost, since the median of the error function is minimized by combinatorial methods over the power set of all matched image and model lines, which is why the 2D matching is important for selecting the line correspondences. The method is capable of handling up to, but not including 50% outliers. A recent paper by Kumar and Hanson [Kumar and Hanson 1990a] recently established the superiority of the least-median squares algorithm over traditional least-mean squares algorithms as well as those based on statistical M-estimation techniques. The sensitivity of pose refinement and other related 3D inference methods to inaccurate estimates of the image center and focal length has been theoretically established and experimentally verified [Kumar and Hanson 1990b]. The results show that for a field of view of 24° , an error of 10 pixels in the image center does not affect the recovered location of the camera significantly. The error in the recovered orientation of the camera is the angular displacement of the estimate of the image center from the true value.

Model-directed object recognition becomes much more difficult when the viewpoint of the three-dimensional object is unknown. A popular approach is based upon the use of multiple two- dimensional views of three-dimensional structures, and is referred to under a variety of terms such as "aspect graphs", "generic views", and "characteristic views" (Burns 1987; Burns and Kitchen 1987a; Burns and Kitchen 1987b; Burns and Kitchen 1988; Ikeuchi 1987). If such systems are going to be effective, a clear understanding is required of the manner in which the features of 2D projections vary as a function of the 3D viewing position of the object. It is important to find metric features of an object whose variation is small over a large range of views in order to constrain the number that must be stored.

2.1 Learning 3D Object Recognition Strategies

The basis for recognizing objects in complex outdoor scenes varies widely in terms of the processes utilized, the reliability of the information extracted, the efficiency of the underlying mechanisms, and the manner in which the evidence is combined into an object hypothesis. All of this information is certainly object- and domain-dependent. Some objects can be distinguished on the basis of color, while others can only be identified by scene and object context. Three-dimensional information about shape or texture of some objects might be recovered through bottom-up vanishing point analysis, while the locations of other objects are more easily determined by model-based point matching.

A major problem in knowledge-directed vision is the *construction* of object-directed control strategies. Existing knowledge-directed systems rely on user-supplied heuristics to guide the recognition process. Specifying these heuristics has proven a difficult and time-consuming process. Worse still, there is no guarantee that the resulting strategies are either effective or optimal. The goal of each recognition strategy is to identify any and all instances of the object in an image, and give the 3D position (relative to the camera) of each instance. The goal of the learning process is to build a strategy that minimizes the expected cost of recognition, subject to accuracy constraints imposed by the user.

The problem of automatically learning knowledge-directed control strategies for object recognition is being addressed in (Draper and Riseman 1990). The system is given a description of the object and a set of user-interpreted training images. The task is to build the most efficient object recognition strategy possible within performance constraints set by the user. Three-dimensional 3D object recognition is approached within a generate-and-verify paradigm. The task of learning to generate the minimal necessary set of hypotheses is phrased as a search problem. The task of learning to verify a hypothesis is cast as a classification problem, followed by graph optimization.

In order to recognize an object in an image, a system must compare data extracted from the image to a description of the object class. This description, in turn, can suggest what features to look for in the image. For example, a description in memory of "house" might constrain the shape of a house, but not its color, since houses can be painted almost any color. Therefore, when looking for houses, shape primitives should be extracted, but not color features. The best strategy for recognizing a house (or any object) is determined by its properties.

In the Schema System, object recognition is modeled as a process of applying visual knowledge sources(KSs) to hypotheses. Knowledge sources are processing routines for image understanding, such as 2D$\rightarrow$3D point matching, vanishing point analysis and straight line extraction. Hypotheses are intermediate-level statements about the image and/or 3D world, and can occur at many levels of abstraction. Examples of hypotheses include straight line segments, 3D orientation vectors and volumes. At each step in the recognition process, a knowledge source is applied to one or more hypotheses. The result is either a new hypothesis or a discrete evidence value reflecting the quality of the original hypotheses.

An object's description determines the most efficient and accurate method for recognizing the object. The problem, therefore, is learning which KSs to apply, when to apply them, and how to integrate their results. Recognition strategies are represented by recognition graphs, which are similar in many ways to decision trees. Unlike decision trees, however, recognition graphs direct hypothesis creation as well as hypothesis classification or verification. Object-specific strategies are learned in a two step process. The first step involves learning which hypotheses should be generated. The second step learns how to verify them efficiently.

This work extends the knowledge-based approach by replacing the ad-hoc control heuristics of other systems with Bayesian control and classification decisions. The user, instead of supplying heuristics in the form of if-then rules or confidence functions, specifies accuracy requirements. We solve the problem of selecting when to execute which knowledge source by selecting the KS that minimizes the expected cost of recognition, subject to accuracy constraints imposed by the user. We select these KSs at compile-time in a two-step process. The first step minimizes the number of incorrect or unnecessary hypotheses generated by selecting those with the highest likelihood of beng correct. The second step learns how to verify the remaining hypotheses by building a classifier and then ordering the verification KSs so as to minimize the expected cost of verification. The resulting recognition strategy in embedded in a recognition graph.

2.2 Model matching using local search

A specific but important problem in computer vision is the identification of objects in the world by matching a geometric object model to data extracted from an image. The identification task has two parts: a) determining the correct correspondence between object features and image features and, b) determining the position of the object with respect to the camera. These sub-tasks are interdependent, since an object's position cannot be determined without assuming a correspondence to image features, while identifying the correct correspondence depends on the object's 2D appearance and hence its relative position and orientation in space.

Assuming a roughly known viewpoint, the projection of a 3D object may be translated, rotated and scaled in the plane until it is aligned with its corresponding image features. When both the object projection and image features are made up of straight line segments, the type of matching problem that results is illustrated by Figure 1. We formulate this problem in terms of combinatorial optimization, and utilize a novel local search algorithm to solve it.

Local search is a widely recognized and effective means of solving many difficult combinatorial optimization problems [Papadimitriou82; Lin73]. Put most simply, local search is an iterative generate-and-test procedure which moves from an initial solution, via transformations, to one that is locally optimal. Figure 1 illustrates this approach. One common means of finding global optima is to take the best result from a set of independent trials. Consider an algorithm which finds the globally optimal match with probability $1/3$. The global optimum will be found one or more times in t trials with probability $1-(2/3)^t$. In this case, 6 trials produces the optimal match with probability above 0.9. For a given domain, empirically estimating this probability provides a principled basis for selecting the number of trials required to achieve a desired level of confidence.

Local search matching departs substantially from the previous work in model matching, much of which falls loosely into three categories: generalized Hough and geometric hashing, key-feature and perceptual organization, and tree search.

How a matching algorithm performs when data is imperfect and cluttered is paramount. Most algorithms perform well on clean data. Figure 1 illustrates the types of 'errors' we expect to find in the output of common bottom-up line extraction algorithms such as [Burns, Hanson et al. 1986]. Data lines are often fragmented; they may extend beyond or fall short of the point predicted by the model, and often they are skewed. Hence, the correct correspondence often involves mapping many data line segments to a single model line segment. Moreover, point-wise correspondences are difficult to obtain in a reliable manner. Therefore, as observed by Lowe [Lowe85] and Ayache [Ayache84] before us, a point-to-line measure is preferable. Our exact measure, the integrated squared perpendicular distance between model lines and data segments is new, and the closed form solution to the problem of minimizing this measure subject to 2D rotation, translation and scaling is presented below.

The local search matching algorithm is implemented on a TI Explorer II Lisp machine. In addition, a special C version is running on the Sequent multi-processor and is used to

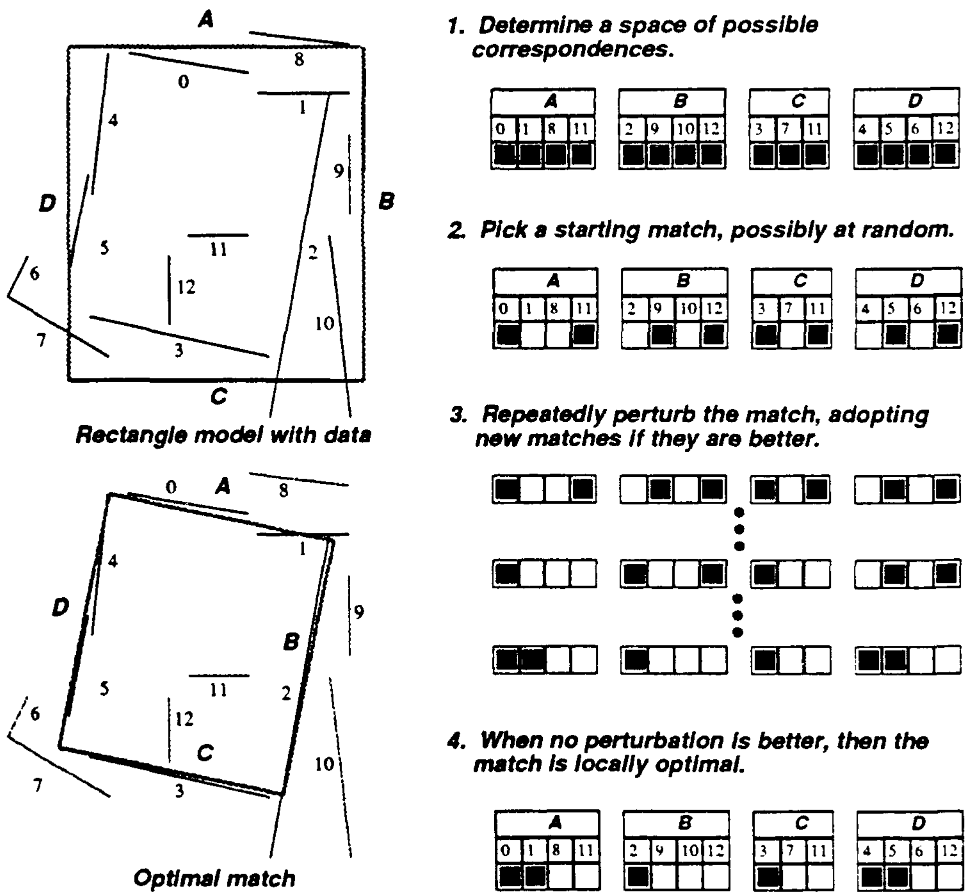

Figure 1: A typical matching problem and a sketch of our basic approach. In the upper left, a rectangle model is shown in proximate registration with an instance of the model. The four straight line segments forming the model are identified by letters, the data line segments are numbered. A space of possible matches is determined and indicated by the tables shown on the right. A filled in entry in the table indicates a match between a model segment and data segment. The best match is shown in the lower left with the model rotated, translated and scaled to fit the corresponding data. The associated correspondence is indicated in the table on the lower right.

perform landmark recognition for our mobile robot. The goal of these experiments is to periodically identify known landmarks and thereby confirm and correct a mobile robot's estimated position and orientation [Fennema, Hanson et al. 1989]. Errors in position and orientation are assumed to be modest so that distortion due to perspective projection is constrained.

To obtain an updated position estimate, the robot navigation system first selects prominent straight lines in the scene. These features are then projected into the 2D image plane and local search matching finds an optimal correspondence between projected model lines and extracted data lines. Data lines are extracted using the Burns algorithm [Burns, Hanson et al. 1986]. Since each 2D model line segment is a projection of a 3D segment in the scene, 2D matching establishes a correspondence between 3D line segments and 2D line segments. This correspondence is used by a pose refinement algorithm [Kumar 1989] to determine the 3D position and orientation of the robot.

Figure 2 shows a 512x512 image of a hallway taken from our mobile robot. The robot navigation system uses a partial model of this environment, along with its estimated position, to select straight lines expected to be visible and of moderate to high contrast. These are then projected into the 2D image plane as shown in Figure 3a. Data line segments found to be 'near' a model line and of similar orientation are considered potential matches. In this example data line segments within 50 pixels and 0.3 radians (17 degrees) of a projected model line segment are considered to be possible matches.

The projected landmarks shown in Figure 3a were obtained by introducing an artificial error into the robot's position estimate in order to test our ability to match and recover the correct 3D pose. We tested three position estimates: the true estimate, 6 inches to the left, and 6 inches to the right. These shifts introduce some perspective distortion into the projected 2D landmarks. Hence each case represents a different 2D matching problem. Figure 3a shows the projections associated with shifting the estimate to the left. In 100 trials, the optimal match was found 93/100 times for the correct estimate, 48/100 for the error to the left, and 100/100 times for the error to the right. In each case the search space contained roughly 280 possible individual model-data pairs. The final results of pose refinement using the optimal 2D matches are summarized in Table 1. Note the final pose derived from matching is always within an inch and a half of the true position. This demonstrates that we can recover pose when small amounts of perspective distortions are present in the 2D landmarks. Clearly, as position errors grow larger there comes a point where this method breaks down.

2.3 View Variation of Line Segment Features

The recognition of 3D objects becomes much more difficult as the position of the viewer relative to the object becomes less constrained or more uncertain. Part of this difficulty comes from the fact that many important features of an object's image vary with view. For an image feature to be useful in discrimination, its distribution of values with respect to each object should be narrow, and the distributions with respect to different objects should

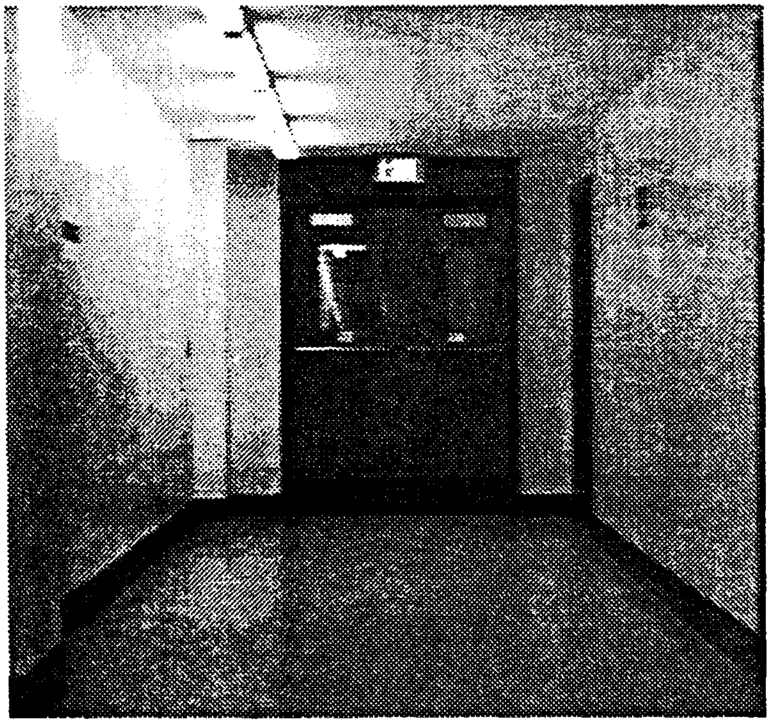

Figure 2: 512x512 Hallway Image

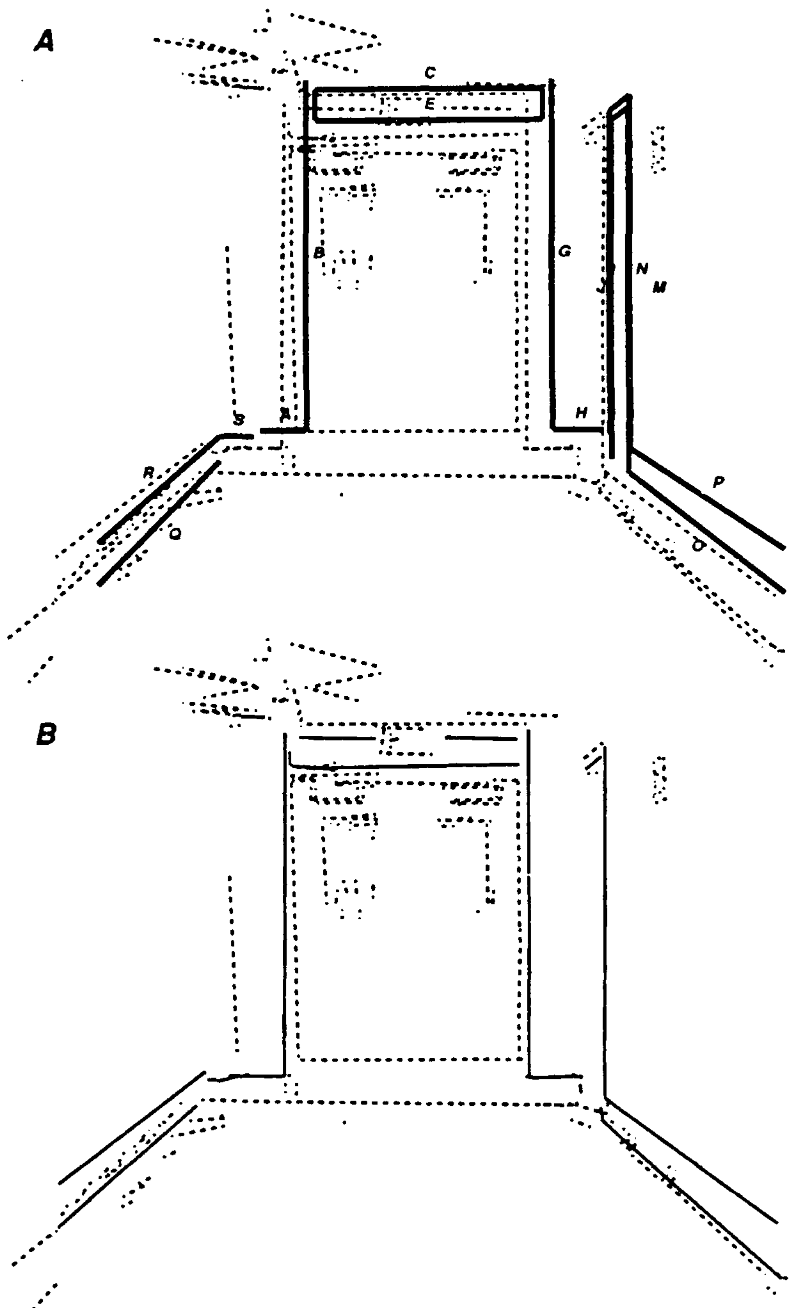

Figure 3: Matching 2D projections of 3D lines to image data. A) The projected line segments in black superimposed over the data extracted from the image shown in Figure 2. B) Data line segments found to match the projected line segments are shown in black, those found not to match are dashed.

	Correct Estimate			6 Inches Left			6 Inches Right		
	X	Y	Z	X	Y	Z	X	Y	Z
True	40.00	4.00	3.57	40.00	4.00	3.57	40.00	4.00	3.57
Est.	40.00	4.00	3.57	40.00	3.50	3.57	40.00	4.50	3.57
Derived	39.96	4.05	3.58	39.87	3.98	3.57	39.90	4.12	3.59
Error	0.04	0.05	0.01	0.13	0.02	0.00	0.10	0.12	0.02

Table 1: Results of 2D matching followed by 3D pose refinement. The units shown are feet, the robot is 40.0 feet from the door shown in Figure 3 and 4.0 feet from either wall. These results show successful recovery of 3D position for a 6 inch error in the initial position estimate.

be well separated. Burns (Burns, Weiss et al. 1990) presents a study of the variation with respect to viewpoint of features for projected point sets and line segments. In this paper it is first established that general-case view-invariants do not exist for any number of points, given true perspective, weak perspective or orthographic projection models.

The use of weak perspective makes it possible to carry out the analysis of feature variation by analytical methods. This approximation to perspective projection is applied extensively in object recognition research to simplify the analysis and computation for 3D object recognition; it produces reasonable results when the camera distance is great enough, relative to the object's extent in depth. Though there are no general-case weak-perspective invariants, there are special-case invariants of practical importance, such as the cross ratio. The special-case weak-perspective invariants cited in the literature are derived from linear dependence relations and the invariance of this type of relation to linear transformation. The variation with respect to view is then studied for an important set of 2D line segment features: the relative orientation, size, and position of one line segment with respect to another. The analysis includes an important evaluation criterion for feature utility in terms of *view-variation*: the relationship between the fraction of views (over a view sphere) and the range of values assumed by a feature over these views. Ideally, a feature should be *view-invariant*, that is, unaffected by change in view. Even if one were to use only those features whose values are bounded over the entire sphere, that would severely restrict the possibilities. Since features of a projection usually "blow-up" to extreme values at some (usually small) set of views, a feature is considered here to have *low view-variation* if the variation is small in extent over a large fraction of the views. For example, the relative orientation of the projection of two lines that are parallel in 3D is always 0° under weak-perspective. The most variation in relative orientation is for lines that are perpendicular in 3D. Even in this case, the relative orientation will be between 60° and 120° for 50% of the view sphere.

The information in the view-variation analysis allows determination of semi-invariant features of an object over areas of the 3D viewing sphere, i.e. features which have a small variation over a large fraction of views. The relationships between the range of feature variation and the fraction of views are presented in (Burns, Weiss, et al. 1990) as a series of graphs for the features described above, and for varying instances of 3D line segments pairs. The mathematical analysis embodied in this paper is generally relevant to techniques for matching 3D models to 2D images.

The prediction-based methods developed in [Brooks81, Lowe85, Burns87, Korn87] constitute a promising approach to the recognition of 3D objects in 2D images. In this approach, recognition is achieved by (1) predicting characteristics of the object projections from all views, (2) matching these predictions against the input 2D image and (3) verifying all promising matches by dete :nining the 3D pose of the object given the data matched. In its most general form a *prediction* expresses the expected values for a set of features of the object's projection. A *feature* can be any measurement or function of the projection, and the expectations can be any valid statement about the feature distribution.

3 Automated Model Generation and Extension

The problem of acquiring models or modifying incorrect models is an important aspect of object recognition and navigation. The major functional requirements of modeling for these tasks are: accurate prediction of visual features, accurate surface orientation and curvature, and accurate feature dimensions. The construction of positionally accurate environmental models is a time consuming, tedious task. Ultimately, the only feasible approach for intelligent systems which are required to interact with large scale changing environments is to provide them with methods for automatically acquiring their internal models during goal-oriented activities or unrestricted exploration. We have been active both in developing a geometric modeling system and in developing techniques for automatic modeling. There are many types of surface that one can fit to 3-D data. We have chosen a representation based on a vertices, edges, and faces. This type of model is supported by Geometer (Connolly 1989; Connolly, Kapur et al. 1989) which provides an environment that includes both planar and algebraic faces. The techniques for automatic modeling use contours (Giblin and Weiss 1987; Collins and Weiss 1990) and intensity information (Oliensis 1990b; Oliensis 1990c).

3.1 GeoMeter

GeoMeter was developed jointly by the University of Massachusetts at Amherst and the GE Research and Development Center as a Common Lisp solid modeling environment specifically oriented toward image understanding. An important goal was to have a system available with source code on a variety of workstations, which would provide the kernel of a modeling system for many applications including visual navigation. The system currently runs on Symbolics, TI Explorer, VaxWorkstation, and Suns. Workstation environments are very useful for interactive debugging of complex models, and the use of X Windows has made the system easily portable.

The advantages of using *GeoMeter* are 3-fold:

1. *GeoMeter* provides the ability to perform the required predictions within a Lisp environment, along with other modules in the navigation system. While there are many other modeling systems, most of them cannot be linked directly with application software.

2. *GeoMeter* contains a simple camera model expressly designed to simulate a mobile, physical camera within a global coordinate system.

3. *GeoMeter* uses a multilevel representation scheme allowing *parts* of objects to be isolated as *landmarks* for navigation.

For the task of modeling objects for recognition and navigation, a boundary representation of surfaces of objects is most useful. With this type of representation, the edges and visual characteristics of the object surface are made explicit. Edges from an existing model of the environment are projected using robot and camera parameters to obtain an ideal "robot's eye" view of the world. For example, the projection of model edges can be used either for prediction to restrict the search for line tokens in the image or for verification to be compared with image line tokens already extracted.

Currently, for computational simplicity, straight lines are used for edges and planes for faces, so that curved surfaces are approximated by polyhedra. However, because of the generality of the mathematical framework, it is possible to represent semi-algebraic curves and surfaces, and there are plans for the future to provide numerical procedures for manipulating these objects.

3.2 Modeling curved surfaces from profiles and other features

Many researchers have worked on acquiring models automatically [Agin73; Baker77; Chien89; Connolly and Stenstrom 1989; Shirai71]. Recently, work on surface reconstruction for curved surfaces has produced notable results [Baker89; Pentland; Vaillant90]. Nevertheless, recovering the structure of curved surfaces is still a difficult, open problem. The results for simultaneously recovering motion parameters and depth of points have been poor because of the difficulty in accurately resolving motion into rotation and translation parameters. This problem is avoided by using known camera motion.

Giblin and Weiss [Giblin and Weiss 1987] have developed a method to compute depth and curvatures for occluding contours as well as creases and surface markings. This approach has been extended to general motion [Vaillant90] and to use differential measurements to reduce errors in recovered curvature [Blake and Cipolla 1990].

There are many ways to model the geometry of a surface. Since one of the goals of the system is to integrate data from multiple views of the surface, it is important that the model be easily modified locally without producing changes to parts of the surface that are not being measured. In addition, since the data is sampled non-uniformly on the surface and contains errors, it important that the model be able to reflect the uncertainty of the observations. The approach that we have selected for modeling the data is triangular planar patches with uncertainties in the surface normal and the distance from the plane to the origin. The triangles are produced by tracking two-dimensional line segments over three or more frames.

An important issue is how to maintain consistency as new data is added. Occluding contours provide estimates on the maximal extent of a surface and therefore one endpoint of a confidence interval for a surface patch. An extremal boundary defines a *cone* from the camera focal point, and the surface must lie within this cone [Connolly and Stenstrom 1986]. At any time, the set of cones from multiple views can be used to detect inconsistency of the model with the visual data. It would also be possible to integrate tactile sense data, which would define a normal to the surface, and the extent of the object would also be constrained in that direction. This is particularly useful for concave surface elements where visual data may not be available. In addition, surface curvatures provide information about the reliability of a planar approximation to the surface, i.e. how far on the surface one can extrapolate the values from a planar patch.

3.3 Recovering Shape from Shading

Shape from shading has traditionally been considered an ill-posed problem. However, in recent work, Oliensis (Oliensis 1990a; Oliensis 1990c) has demonstrated that the solutions to shape from shading are often well-determined, with little or no ambiguity. For the case of illumination that is symmetric around the viewing direction (i.e. the light source is behind the camera), it was shown in (Oliensis 1989) that there is in general a unique solution to shape from shading. This proof is valid for general Lambertian objects (without holes), and is the first proof that the problem of shape from shading can be well-posed in general. These arguments were extended to the case of general illumination direction in (Oliensis 1990c), where it was demonstrated that, in this case also, the solutions to shape from shading are strongly constrained over much of the image. These results follow from a combination of local uniqueness theorems and global arguments concerning the properties of the flow of characteristic strips, both derived from the mathematical theory of dynamical systems theory. The essential constraints restricting the solution space are shown to be provided by the singular points in the image. Also, characteristic strips are given a simple interpretation as space curves, and demonstrated to be independent of the viewing direction.

It has long been an open question whether the image of the occluding boundary provides additional constraints on the solution to shape from shading. In (Oliensis 1990c), it is proven analytically that the answer to this question is negative. Specifically, for a local image patch containing a portion of the boundary, the problem of shape reconstruction is shown to be ill-posed. Shape reconstruction is actually more ambiguous in the neighborhood of an occluding boundary segment than it is in the neighborhood of an interior image curve. The proof, which applies to a Lambertian surface illuminated from a general light source direction, is based on recasting the basic characteristic strip equations of Horn in a form that is completely non-singular on the occluding boundary.

Also, an example is presented in (Oliensis 1990c) in which a small image region bordering the image of the occluding boundary yields an ambiguous shape reconstruction, even though the image contains both singular points and the whole of the occluding boundary. This example demonstrates that shape from shading can be well-posed and ill-posed simultaneously: although the shape corresponding to most of the image is actually uniquely determined, the shape corresponding to the specified small image region is ill-determined. It is argued that, in general, these ill-posed regions are probably small fractions of the image, but that they can occur frequently, in images both with and without visible occluding boundaries, and in practice may lead to instabilities and errors in shape reconstruction algorithms.

Finally, Oliensis has developed (Oliensis 1990b; Oliensis 1990c) a new local algorithm for reconstructing shape from shading using a general quadratic surface model. The new constraints for shape from shading should be investigated for their potential for robust surface reconstruction in combination with the information obtained from contours by the Giblin-Weiss algorithm.

3.4 Determining orientation of planar surfaces from stereo line correspondences and vanishing points

One particular aspect of modeling is determination of orientation of straight lines and planar surfaces. Collins (Collins and Weiss 1990) has developed a system to compute the orientation of 3D lines directly from stereo line correspondences without first computing point depths, in contrast to methods that compute depth in order to obtain the orientation of line segments. After computing line directions, it is possible to efficiently discover coplanar lines and thereby recover the orientation and position of planar surfaces in the scene.

The problem of obtaining reliable and accurate estimates of line and surface orientations obtined either from stereo line correspondences or vanishing point analysis is addressed by confidence regions. Since orientations are represented as unit vectors, statistical techniques for estimating the axes and uncertainty of point distributions on the unit sphere must be used. Bingham's distribution is versatile in that it can describe both equatorial and bipolar distributions depending on the parameter values. Statistical parameter estimation based on Bingham's distribution, as well as a general nonparametric estimation method, are used to solve for the polar axis of a great circle of points and to represent the statistical uncertainty in the resulting orientation estimate.

One of the methods we have for deriving orientation information from static images is the estimation of a unit vector perpendicular to a number of derived unit vectors. For instance, under perspective projection a ray pointing towards the intersection of a group of converging image line segments is perpendicular to their projection plane normals. This has applications in finding vanishing points and in locating the focus of expansion of a pure translational flow field. Furthermore, the normal to a planar surface is perpendicular to the direction of all lines lying on that surface. The problem of estimating a vector mutually perpendicular to several unit vectors can be characterized as estimating the polar axis of a great circle on the unit sphere. Bingham's distribution, which represents the intersection of a 3D Gaussian distribution with the surface of the unit sphere, is introduced to describe such an equatorial distribution of unit vectors. However, statistical parameter estimation based on Bingham's distribution is somewhat expensive computationally. Collins and Weiss (Collins and Weiss 1990) develop a more convenient alternative based on linear-least-squares plane fitting. In addition, they consider the problem of estimating the orientation and uncertainty of the cross product of two uncertain unit vectors. The tentative solution is to form a Gaussian approximation to the "intersection" of two equatorial Bingham distributions.

The above methods are illustrated using two examples (Collins and Weiss 1990). The first involves reconstruction of planar surfaces using stereo line correspondences . If relative pose of the stereo cameras can be determined, then the orientation of lines in the world can be computed as the cross product of the projection plane normals of its two corresponding images, one in each image plane. The projection plane is the plane containing the image line and the camera focal point. Given a set of lines hypothesized to lie on a single planar surface, the plane orientation and uncertainty can be computed as the pole of a great circle formed from the uncertain line orientation estimates.

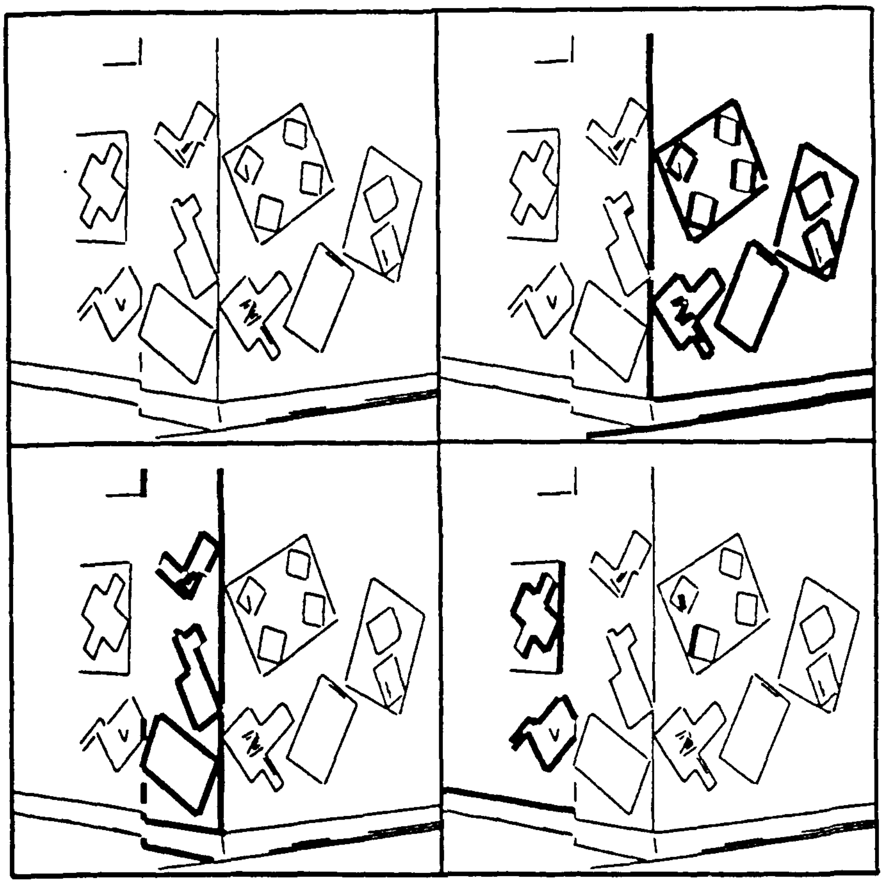

Figure 4: Matched input lines from the left image of a stereo pair are shown in the upper left; also shown are 3 sets of lines consistent with hypothesized surface planes.

The second example involves the analysis of vanishing points (Collins and Weiss 1989). The images of parallel 3D lines converge to a vanishing point in the projective image plane. A ray constructed from the camera focal point towards the vanishing point has the same 3D orientation as the original world lines. The line orientation and its approximate confidence region on the unit sphere is estimated as the polar axis of a great circle of projection plane normals. Furthermore, surface plane orientations are hypothesized as the cross product of these uncertain line directions.

Having first computed 3D line directions, it is possible to discover coplanar lines and thereby recover the orientation and distance of the planar surfaces that contain them. This is done in two stages. First the lines are broken into groups consistent with a family of parallel planes, then distances are finally computed to partition the lines into sets consistent with individual plane equations.

Figure 4 shows an example of the partition created for a stereo hallway image. The algorithm forms hypotheses of all three visible wall planes, and correctly identifies that one plane orientation is shared by two parallel planes at different depths.

3.5 Automated Model Extension

Two preliminary experiments have been performed using the 3D pose refinement algorithm to extend a partial model from a set of known points to include unknown points: these experiments are described in more detail in (Kumar and Hanson 1990b). The known model points are used to locate new points in the world coordinate system from pose refinement and triangulation over the induced stereo baseline obtained from a pair of 3D poses (e.g. location and orientation of the camera for each image). In both experiments, an image sequence was obtained for which the three- dimensional location of a set of points in the environment was known (the model). Image features are tracked over a sequence of frames using a token-based line tracker (Williams and Hanson 1988a; Williams and Hanson 1988b; Williams and Hanson 1988c), which provides the token correspondences. The 3D pose estimation algorithm described earlier is applied to each frame to map each feature into a stable world coordinate frame. The 3D pseudo-intersection of the rays passing through the camera center and the image feature point in each image frame is found using an optimization technique which minimizes the sum-of-squares of the perpendicular distances from the 3D pseudo- intersection point to the rays. In effect, this induces a stereo baseline between frames from which the 3D coordinates of the unknown features can be obtained by triangulation. Note that the computation of the location of new points in the world coordinate system is not sensitive to accurate estimation of the image center.

4 Perceptual Organization

Image understanding and model acquisition require the extraction of appropriate structures from images. This includes detecting lines and regions and forming geometric structures for matching or grouping them together into surfaces.

Dolan (Dolan and Weiss 1989) is extending the perceptual grouping mechanisms developed by Boldt (Boldt, Weiss et al. 1989) for straight lines to the case of general curves. Like the straight line system, the curve grouping algorithm relies on the Gestalt principles of proximity and good continuation and employs an iterative token-based approach to search for and describe significant curve structures (including straight lines, conic arcs, inflections, corners, and cusps).

Williams (Williams 1990) has developed a system for perceptually organizing surface boundaries based on figural clues alone, although results have only been demonstrated in the 'Colorforms' domain and other simple scenes. The system has, however, successfully extracted Kanizsa's occluding triangle and has correctly analyzed relatively complex scenes containing multiple occluding surfaces. Detailed results are presented in Williams (Williams 1990). The current system is designed to complete gaps in the straight sections of occluded contours but isn't yet able to cope with more complex occlusions, such as missing corners or missing sides.

4.1 Perceptual Organization of Curves

The system iterates a cycle of linking, grouping, and replacement over a range of perceptual scales, but within each iteration processing occurs independently at each token. Each token is linked to other tokens that are likely to be its neighbors along some contour. Sequences of linked tokens are analyzed and classified based on the geometric structure they exhibit. Appropriate replacement tokens are then generated to explicitly describe and replace each sequence. Beginning with initial edge tokens (unit tangents centered at edge locations), curved structure is discovered in a bottom-up, local-to-global fashion and a multiscale description results. The computational complexity inherent in any grouping process is managed here by searching locally within a perceptual window (which defines the local scale) and by explicitly replacing a sequence of tokens by a single token at the next scale.

Since the work previously reported in (Dolan and Weiss 1989), a parallel version of the grouping algorithm has been implemented in anticipation of parallel hardware. Here, the grouping process is simultaneously applied to the perceptual window (i.e. context) around each token for potential grouping and replacement, and parallel replacement of the aggregate tokens is assumed to take place simultaneously. A consequence of a highly distributed and parallel grouping process is that redundant descriptions arise because the contexts of nearby tokens overlap, and overlapping aggregate tokens are produced. Dolan is currently developing methods to identify and eliminate such redundancies by representing multiple types of relationships in the link graph; this will allow redundancy, as well textural structures to be dealt with in this parallel framework.

4.2 Perceptual Organization of Occluding Contours

Contours corresponding to surface boundaries are readily perceived or completed by human observers even when local evidence in the form of measurable image brightness gradients is completely absent. A classic example of the former is the Kanizsa triangle, in which the illusory contours of the 'occluding' triangle are visually compelling, even though there is scant evidence for their existence. An example of completion occurs when one surface is partially occluded by a second (opaque) surface.

In Williams' system, the mechanisms of occlusion of one surface by another are captured in a set of integer linear constraints. These constraints ensure that the outputs of a contour grouping process is physically valid and consistent with the image evidence. Among the many feasible solutions, the most compelling is the solution which best explains the presence and form of the image structure. The problem of computing a complete and consistent surface boundary representation is thus reduced to solving an integer linear program.

Image contours corresponding to discontinuities in depth are called *occluding contours*. Robust computation of the $2\frac{1}{2}D$ *sketch* [Marr82] under the most general conditions is probably impossible without additional constraints derived from occluding contours. Accordingly, understanding the grouping processes which create occluding contours is an important problem in computer vision. Happily, the semantics of the $2\frac{1}{2}$D sketch can provide objective grouping criteria. Whether the human visual system is willing or unwilling to complete a

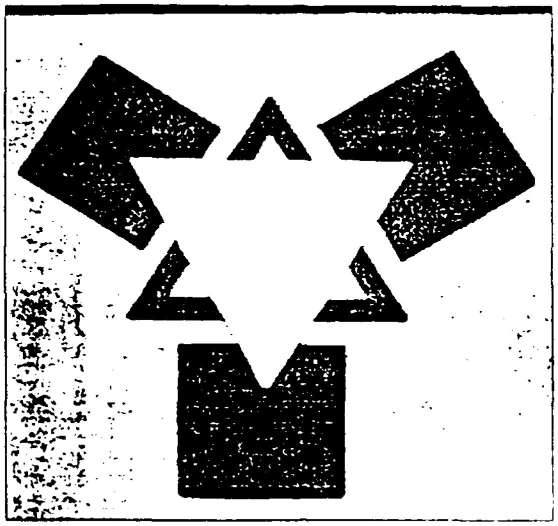

Figure 5: A Colorforms Kanizsa Triangle.

gap in an image contour is determined by a non-local process with knowledge of the mechanics of surfaces and occlusion.

As a test domain, we chose simple scenes built from flat, polygonal vinyl cutout surfaces of uniform reflectance called Colorforms. The current system is designed to complete gaps in the straight sections of occluding contours, but isn't able to cope with more complex omissions such as missing corners or missing sides. Nevertheless, even with this restriction, it is able to solve relatively complex perceptual problems such as the construction of an illusory triangle in a Colorforms equivalent of the Kanizsa Triangle[Kanizsa76] (Figure 5). The system operates in two stages: 1) A problem posing stage; and 2) A problem solving stage[1].

In the problem posing stage image evidence is collected and incorporated in a graph, called the *contour graph*. The contour graph is an explicit representation of *primitive* image structure [Witkin83] and corresponds approximately to Marr's *full primal sketch* [Marr82]. It is composed of two types of vertices and three types of edges. Every vertex is located at a point in the image and every edge is a contour joining two vertices. The initial edges of the contour graph are called *image lines*, and are the output[2] of Boldt's zero crossing grouping algorithm [Boldt, Weiss et al. 1989] (Figure 6).

Image lines are contours with a measurable image brightness gradient. Each image line joins its two *endpoints*, which are the initial vertex type. Proximal endpoints of image lines satisfying certain other simple criteria are joined with a second edge type called a *corner*. Next, all pairs of roughly collinear image lines (as determined by the mean square error of a line fit to the four endpoints) are identified. The near endpoints of each such pair are

[1]The notion that visual perception is cognitive problem solving is due to Rock [Rock83; Rock87].

[2]In the case of the Colorforms Kanizsa Triangle, all evidence of the center triangle was first removed by filtering the initial zero crossing segments on gradient magnitude.

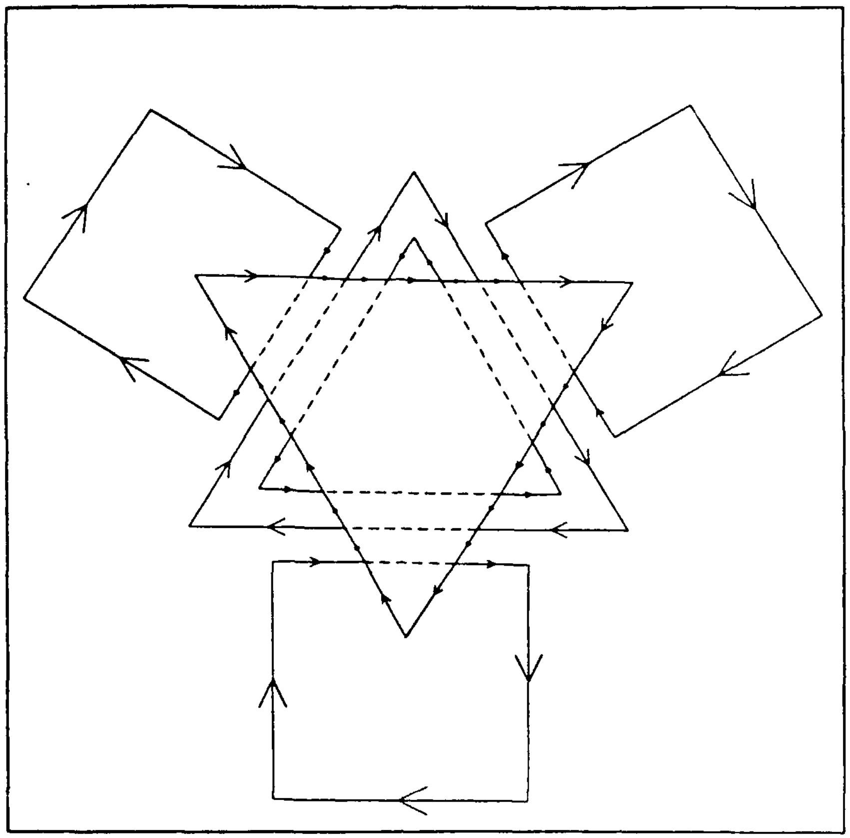

Figure 6: The optimal feasible solution.

joined by a third edge type, the *virtual line* [Kanizsa76; Marr82]. Finally, wherever a virtual line intersects another virtual line, or a virtual line intersects an image line, the two lines are split into four sub-segments and joined by a new type of vertex, called a *crossing*. This insures that the graph remains planar.

By writing a fixed number of linear constraints for each vertex and edge in the contour graph, an *integer linear program* is generated. During the problem solving stage, branch and bound search and the Simplex algorithm are used to find its optimal feasible solution[Burnett91]. The optimal feasible solution defines the *boundary graph*, which is a labeled sub-graph[3]. The edges of the boundary graph are labeled with a sign of occlusion and a depth index (hidden lines are displayed dashed). The boundary graph corresponding to the optimal feasible solution of the integer linear program is depicted in Figure 7. An alternate organization, which is a feasible but non-optimal solution, appears in Figure 8.

[3]This is consistent with Witkin and Tenenbaum's[Witkin83] view that "naively perceived structure survives more or less intact when a semantic context is established... the difference between naive and informed perception amounting to little more than labeling the perceptual primitives."

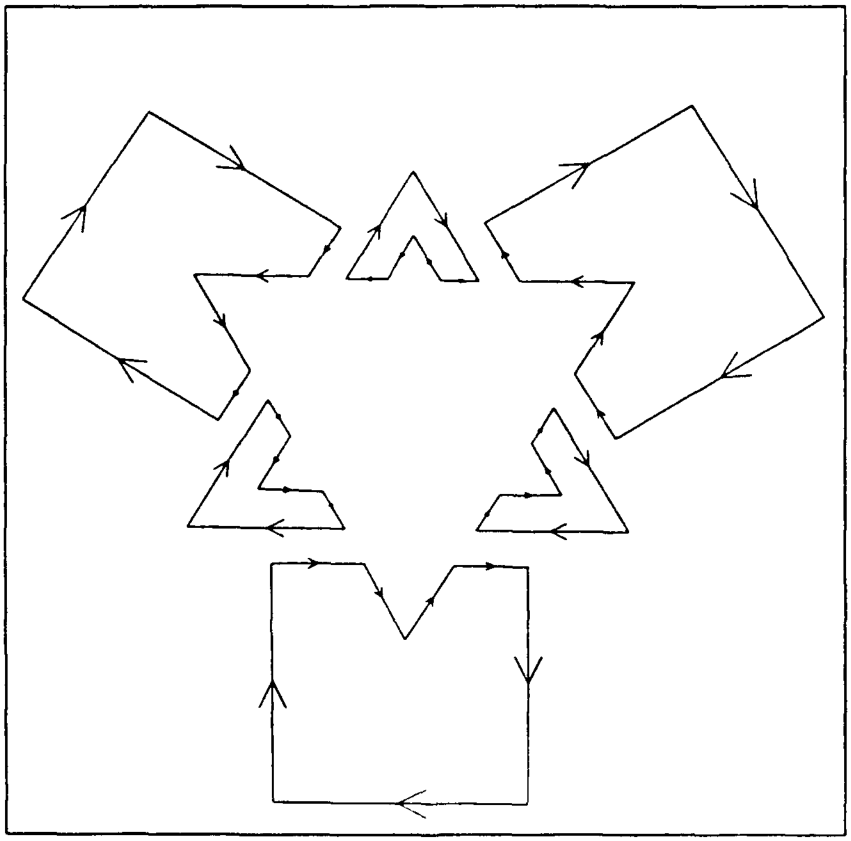

Figure 7: A feasible solution which is non-optimal.

References

[Agin73] G.J. Agin and T.O. Binford. Computer description of curved objects. In *Proceedings of the 1973 International Joint Conference on Artificial Intelligence*, pages 629–640, 1973.

[Allen84] P.K. Allen. Surface descriptions from vision and touch. In *Proceedings of the 1984 Conference on Robotics*, pages 394–397, Atlanta, GA, March 1984. IEEE.

[Allen86] P.K. Allen. Sensing and describing 3-d structure. In *Proceedings of the 1986 Conference on Robotics and Automation*, volume 1, pages 126–131, San Francisco, CA, April 1986. IEEE.

[Ayache84] N. Ayache and O. D. Faugeras. A new approach for the recognition and positioning of 2-d objects. In *Seventh Internation Conference on Pattern Recognition*, pages 1274 - 1278, 1984.

[Bajcsy85] R. Bajcsy. Active perception vs. passive perception. In *Proceedings of Workshop on Computer Vision*, pages 55–59, Bellaire, MI, October 1985. IEEE.

[Baker77] H. Baker. Three-dimensional modeling. In *Proceedings of the 1977 International Joint Conference on Artificial Intelligence*, pages 649–655, 1977.

[Baker89] H. Baker. Surface reconstruction from images. In *Proceedings of the Second International Conference on Computer Vision*, pages 334–343, Tampa, FL, December 1988.

[Beveridge, Weiss, et al. 1989] Beveridge, J. Ross, Weiss, R., and Riseman, E., "Optimization of 2-Dimensional Model Matching", *Proc. DARPA Image Understanding Workshop*, Palo Alto, CA, May 1989, pp. 815-830. Also COINS TR 89-57.

[Beveridge, Weiss, et al. 1990] Beveridge, J.R., Weiss, R., and Riseman, E., "Combinatorial Optimization Applied to 2D Model Matching Subject to Fixed and Variable Scale Transformations", *IEEE International Conference on Pattern Recognition*, Atlantic City, NJ, June 1990, Vol. 1, pp. 18-23.

[Blake and Cipolla 1990] A. Blake and R. Cipolla. Robust estimation of surface curvature from deformation of apparent contours. In *Proceedings of the European Conference on Computer Vision*, Antibes, FRANCE, 1990.

[Boldt, Weiss et al. 1989] Boldt, M., Weiss, R., and Riseman, E., "Token-Based Extraction of Straight Lines", *IEEE Transactions on Systems, Man and Cybernetics*, Vol 19, No. 6, November/December 1989, pp. 1581-1594.

[Bolles82] R. C. Bolles and R. A. Cain. Recognizing and locating partially visible objects: the local-feature-focus method. *International Journal of Robotics Research*, 1(3):57 - 82, 1982.

[Brolio, Draper et al. 1989b] Brolio, J., Draper, B., Beveridge, J.R., and Hanson, A., "The ISR: A Database for Symbolic Processing in Computer Vision", special issue of **Computer**, titled *Image Database Management*, December 1989, pp. 22-30.

[Brolio, Draper et al. 1989a] Brolio, J., Draper, B., Beveridge, J.R., and Hanson, A., "The ISR: A Database for Symbolic Processing in Computer Vision", COINS TR 89-111, November 1989.

[Brooks81] Brooks, R.A., "Symbolic Reasoning among 3D Models and 2D Images," *Artificial Intelligence*, vol 17, pp. 285-348, 1981.

[Bruce and Giblin 1984] J.W. Bruce and P.J. Giblin. *Curves and Singularities*. Cambridge University Press, Cambridge, 1984.

[Burnett91] Burnett, J., and Buy, U., Solving Integer Programming Systems Using the IMINOS Prototype, Technical Report, Department of Computer and Information Science, University of Massachusetts, Amherst, Mass., in preparation.

[Burns, Hanson et al. 1986] J. B. Burns, A. R. Hanson, and E. M. Riseman. Extracting straight lines. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(4):425 - 456, July 1986.

[Burns 1987] Burns, J.B., and L.J. Kitchen, "Recognition in 2D images of 3D objects from large model bases using prediction hierarchies," *Proc Int Joint Conf on Artificial Intelligence*, Milan, pp. 763-766, 1987.

[Burns and Kitchen 1987] "An Approach to Recognition in 2D Images of 3D Objects from Large Model Bases," Dept of Computer and Information Science, University of Massachusetts, Amherst, TR87-85.

[Burns and Kitchen 1988] Burns, J. Brian, and Kitchen, Leslie J., "Rapid Object Recognition From a Large Model-Base Using Prediction Hierarchies", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988, pp. 711-719.

[Burns, Weiss et al. 1990] Burns, J.B., Weiss, R., and Riseman, E., "View Variation of Point-Set and Line-Segment Features", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 650-659. Also COINS TR 90-84.

[Chien89] C.H. Chien and J.K. Aggarwal. Model construction and shape recognition from occluding contours. In *Pattern Analysis and Machine Intelligence*, volume 11, pages 372-389, April 1989.

[Collins and Weiss 1990a] Collins, R., and Weiss, R., "Deriving Line and Surface Orientation by Statistical Methods on the Unit Sphere", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 433-438. Also COINS TR 90-102.

[Collins and Weiss 1990b] Collins, R., and Weiss, R., "Vanishing Point Calculation as a Statistical Inference on the Unit Sphere", *International Conference on Computer Vision*. Osaka, Japan, December 1990.

[Connolly and Stenstrom 1986] C.I. Connolly and J.R. Stenstrom. Construction of polyhedral models from multiple range views. In *Proceedings of the 8th International Conference on Pattern Recognition*, pages 85–87, October 1986.

[Connolly and Stenstrom 1989] C.I. Connolly and J.R. Stenstrom. Scene reconstruction from multiple sensory images. In *Proceedings of the IEEE Workshop on Interpretation of 3D Scenes*, pages 124–130. IEEE, November 1989.

[Connolly, Kapur et al. 1989] Connolly, C., Kapur, D., Mundy, J., and Weiss, R., "GeoMeter: A System for Modeling and Algebraic Maniupulation", *Proc. DARPA Image Understanding Workshop*, Palo Alto, CA, May 1989, pp. 797-804. Also COINS TR 89-99.

[Davis82] Larry S. Davis. Hierarchical generalized hough transforms and line-segment based generalized hough transforms. *Pattern Recognition*, 15(4):277 – 285, 1982.

[Dolan, Weiss et al. 1988] Dolan, J., Weiss, R., and Kitchen, L., "Perceptual Grouping of Curved Lines", *Proc. SPIE Intelligence Robots and Computer Vision*, pp. 356-364, Cambridge, MA, November 1988.

[Dolan and Weiss 1989] Dolan, J., and Weiss, R., "Perceptual Grouping of Curved Lines", *Proc. DARPA Image Understanding Workshop*, Palo Alto, CA, May 1989, pp. 1135-1145.

[Draper, Brolio et al.1988] Draper, B.A., Brolio, J., Collins, R.T., Hanson, A.R., and Riseman, E.M., "Image Interpretation by Distributed Cooperative Processes", *IEEE Computer Vision and Pattern Recognition Conference*, Ann Arbor, MI, June 1988, pp 129-137.

[Draper, Collins et al. 1988] Draper, B.A., Collins, R.T., Brolio, J., Hanson, A.R., and Riseman, E.R., "Issues in the Development of a Blackboard-Based Schema System for Image Understanding", *Blackboard Systems*, (R.S. Engelmore and A.J. Morgan, Eds.), Addison Wesley, 1988.

[Draper, Collins et al. 1989] Draper, B., Collins, R., Brolio, J., Hanson, A., and Riseman, E., "The Schema System", *International Journal on Computer Vision*, 2, pp. 209-250, 1989.

[Draper, Beveridge et al. 1990] Draper, B., Beveridge, J.R., Brolio, J., Hanson, A., Heller, R., and Williams, L., "ISR2 User's Guide", COINS TR 90-52.

[Ellis 1984] R.E. Ellis. Extraction of tactile features by passive and active sensing. *SPIE Intelligent Robots and Computer Vision*, 521:289–295, 1984.

[Fennema, Riseman et al. 1988] C. Fennema, E. Riseman, and A. Hanson. Planning with perceptual milestones to control uncertainty in robot navigation. In *Proc. of SPIE*, pages 1–16, Cambridge, November 1988. International Society for Photographic and Industrial Engineering.

[Fennema, Hanson et al. 1989] Claude Fennema, Allen Hanson, and Edward Riseman. Towards autonomous mobile robot navigation. In *Proceedings: Image Understanding Workshop, also to appear in Systems Man and Cybernetics 1990*, pages 219 - 231, Los Altos, CA, June 1989. DARPA, Morgan Kaufmann Publishers, Inc.

[Gaston84] P. C. Gaston and T. Lozano-Pérez. Tactile recognition and localization using object models: The case of polyhedra on a plane. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI - 6:721 - 741, May 1984.

[Giblin and Weiss 1987] P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. In *Proceedings of the First International Conference on Computer Vision*, pages 136–144, London, ENGLAND, December 1987.

[Grimson86] W.E.L. Grimson and T. Lozano-Pérez. Model-based recognition and localization from tactile data. *Journal of Robotics Research*, 3(3):3–35, 1986.

[Grimson87] W. E. L. Grimson and T. Lozano-Pérez. Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(3):469 -482, 1987.

[Grimson88] W. E. L. Grimson and D. P. Huttenlocher. On the sensitivity of the hough tranform for object recognition. In *Proc. of the International Conference on Computer Vision*, pages 700 - 706, 1988.

[Grimson89] W. E. L. Grimson. On the recognition of parameterized 2-d objects. *Internation Journal of Computer Vision*, 2(4):353 - 372, April 1989.

[Grupen, Weiss et al. 1990a] R. Grupen, R. Weiss, and D. Oskard. Grasp-oriented sensing and control. In *Symposium on Advances in Intelligent Systems*, Boston, MA, November 5-9 1990. SPIE.

[Grupen, Weiss et al. 1990b] R.A. Grupen and R.S. Weiss. Sensor-based planning for grasping and manipulation with multifingered robot hands. Technical Report COINS Technical Report 90-58, COINS Department, University of Massachusetts, July 1990.

[Hutchinson89] S. Hutchinson and A. Kak. Planning sensing strategies in a robot work cell with multi-sensor capabilities. *IEEE Transactions on Robotics and Automation*, 5(6):765–783, 1989.

[Ikeuchi87] Ikeuchi, K. "Generating an Interpretation Tree from a CAD Model for 3D-Object Recognition in Bin-Picking Tasks," *Int Journal of Computer Vision*, Vol. 1(2), pp. 145-165, 1987.

[Illingworth88] J. Illingworth and J. Kittler. A survey of the hough transform. *Computer Vsion, Graphics, and Image Processing*, 44:87 - 116, 1988.

[Kanizsa76] Kanizsa, G.K., Subjective Contours, *Scientific American*, April 1976.

[Korn87] Korn, M.R. and C.R. Dyer, "3D Multi-view Object Representations for Model-based Object Recognition," *Pattern Recognition*, vol 20(1), pp. 91-103, 1987.

[Kumar 1989] Rakesh Kumar. Determination of camera location and orientation. In *Proceedings: Image Understanding Workshop*, pages 870 - 881, Los Altos, CA, June 1989. DARPA, Morgan Kaufmann Publishers, Inc.

[Kumar and Hanson 1989a] Kumar, R., and Hanson, A., "Robust Estimation of Camera Location and Orientation From Noisy Data Having Outliers", *IEEE Workshop on Interpretation of 3D Scenes*, Austin, TX, November 1989, pp. 52-60.

[Kumar and Hanson 1989b] Kumar, R., and Hanson, A., "Robust Estimation of Camera Location and Orientation From Noisy Data with Outliers," Dept. of Computer and Information Science, University of Massachusetts, Amherst, TR89-120.

[Kumar and Hanson 1990a] Kumar, R., and Hanson, A., "Pose Refinement: Application to Model Extension and Sensitivity to Camera Parameters", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 660-669.

[Kumar and Hanson 1990b] Kumar, R., and Hanson, A., "Analysis of Different Robust Methods for Pose Refinement", *IEEE International Workshop on Robust Computer Vision*, Seattle, WA, October 1990.

[Kumar and Hanson 1990c] Kumar, R., and Hanson, A., "Sensitivity of the Pose Refinement Problem to Accurate Estimation of Camera Parameters", *International Conference on Computer Vision*, Osaka, Japan, December 1990.

[Lamdan88] Y. Lamdan and H. J. Wolfson. Geometric hashing: A gneral and efficient model-based recognition scheme. In *Proc. IEEE Second Int. Conf. on Computer Vision*, pages 238 - 249, Tampa, December 1988.

[Lin73] S. Lin and B. Kernighan. An effective heuristic algorithm for the traveling salesman problem. *Operations Research*, 21:498 - 516, 1973.

[Lowe85] David G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.

[Marr82] Marr, D., *Vision*, Freeman Press, San Francisco, Cal., 1982.

[Oliensis 1989] Oliensis, J., "Existence and Uniqueness in Shape From Shading", COINS TR 89-109.

[Oliensis 1990a] Oliensis, J., "Existence and Uniqueness in Shape From Shading", *IEEE International Conference on Pattern Recognition*, Atlantic City, NJ, June 1990, Vol. 1, pp. 342-345.

[Oliensis 1990b] Oliensis, J., "New Results in Shape From Shading", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 145-153.

[Oliensis 1990c] Oliensis, J., "Shape From Shading as a Partially Ill-Posed Problem", COINS TR 90-50.

[Oliensis 1991] Oliensis, J., "Uniqueness in Shape From Shading", *International Journal of Computer Vision*, to appear, 1991.

[Papadimitriou82] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*, chapter Local Search, pages 454 - 480. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[Pentland] A.P. Pentland. Recognition by parts. In *First International Conf. on Computer Vision*, pages 612–620, London, Dec 1987.

[Rock83] Rock, I., *The Logic of Perception*, MIT Press, Cambridge, Mass., 1983.

[Rock84] Rock, I., *Perception*, Scientific American Books, New York, 1984.

[Rock87] Rock, I., A Problem Solving Approach to Illusory Contours, *The Perception of Illusory Contours*, Petry and Meyer (eds.), Springer-Verlag, New York, 1987.

[Shirai71] Y. Shirai and M. Suwa. Recognition of polyhedrons with a range finder. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 80–87, 1971.

[Shmuel90] A. Shmuel and M. Werman. Active vision: 3d from an image sequence. In *Proceedings of the International Conference on Pattern Recognition*, pages 48–53, June 1990.

[Thompson87] D. Thompson and J. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *Proc. IEEE International Conference on Robotics and Automation*, pages 208 - 220, 1987.

[Tucker88] Lewis W. Tucker, C. R. Feynman, and D. M. Fritzsche. Object recognition using the connection machine. In *Proc. CVPR 88*, pages 871 - 878, Ann Arbor, June 1988.

[Vaillant90] R. Vaillant. Using occluding contours for 3-d object modeling. In *Proceedings of the European Conference on Computer Vision*, Antibes, FRANCE, 1990.

[Williams and Hanson 1988] Williams, L.R., and Hanson, A.R., "Depth From Looming Structure", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988, pp. 1047-1051.

[Williams and Hanson 1988] Williams, L.R., and Hanson, A.R., "Translating Optical Flow Into Token Matches", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988, pp. 970-980.

[Williams and Hanson 1988] Williams, L. and Hanson, A., "Translating Optical Flow Into Token Matches and the Recovery of Depth From Looming", *Proc. ICCV*, Tarpon Springs, FL, December 1988. Also COINS TR88-68, University of Massachusetts at Amherst, August 1988.

[Williams 1990] Williams, L., "Perceptual Organization of Occluding Contours", *International Conference on Computer Vision*, Osaka, Japan, December 1990. Also in *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, September 1990, pp. 639-649.

[Witkin83] Witkin, A. P. and Tenenbaum, J.M., On the Role of Structure in Vision, *Human and Machine Vision*, Beck, Hope and Rosenfeld (eds.), Academic Press, 1983, pp. 481-543.

DISTRIBUTION LIST

RL/IRRE 5
ATTN: Lee A. Uvanni
Griffiss AFB NY 13441-5700

University of Massachusettes 1
Department of Computer and
Information Science
Amherst, MA 01003

RL/DOVL 1
Technical Library
Griffiss AFB NY 13441-5700

Administrator 2
Defense Technical Info Center
DTIC-FDAC
Cameron Station Building 5
Alexandria VA 22304-6145

Defense Advanced Research Projects 2
Agency
1400 Wilson Blvd
Arlington VA 22209-2308

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.