ADA043418

(12)

Semiannual Technical Report

Covering the period 12 November 1976 through 12 May 1977

INTERACTIVE AIDS FOR CARTOGRAPHY AND PHOTO INTERPRETATION

By: HARRY G. BARROW
Principal Investigator
(415) 326-6200, Ext. 5089

Sponsored by

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
ARLINGTON, VIRGINIA 22209

CONTRACT DAAG29-76-C-0057
ARPA Order No. 2894-5
Program Element Code 61101E

STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

Approved for public release; distribution unlimited.

9 Semiannual Technical Report, 11 May 1977
Covering the period 12 November 1976 through 12 May 1977,
Stanford Research Institute Project 5300

12 37p.

6 INTERACTIVE AIDS FOR CARTOGRAPHY AND PHOTO INTERPRETATION.

Submitted by: 10 HARRY G. BARROW
Principal Investigator
(415) 325-6200, Ext. 5089

15 CONTRACT DAAG29-76-C-0057, ARPA Order No.-2894
Program Element Code 61101E

Effective Date: 12 May 1976
Expiration Date: 9 October 1977
Amount of Contract: $521,363

Prepared for:

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
ARLINGTON, VIRGINIA 22209

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

Approved by:

PETER E. HART, *Associate Director*
Artificial Intelligence Center

EARLE D. JONES, *Associate Executive Director*
Information Science and Engineering Division

ABSTRACT

This report describes the status of the SRI Image Understanding project at the end of twelve months. The central scientific goal of the research program is to investigate and develop ways in which diverse sources of knowledge may be brought to bear on the problem of interpreting images. The research is focused on the specific problems entailed in interpreting aerial photographs for cartographic or intelligence purposes. A key concept is the use of a generalized digital map to guide the process of image interpretation.

CONTENTS

ILLUSTRATIONS

ACKNOWLEDGMENTS

The research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by the U.S. Army Research Office under Contract No. DAAG29-76-C-0057.

I INTRODUCTION

This report describes the ongoing SRI image understanding project. The central scientific goal of this project is to investigate and develop ways in which diverse sources of knowledge may be brought to bear on the problem of interpreting images. The research is focused on the specific problems entailed in interpreting aerial photographs for cartographic or intelligence purposes. Additional details are to be found in two earlier progress reports [1] [2].*

A key concept is the use of a generalized digital map to guide the process of image interpretation. This map is actually a data base containing generic descriptions of objects and situations, available imagery, and techniques, in addition to topographical and cultural information found in conventional maps.

We recognize that within the limitations of the current state of image understanding it is not possible to replace a skilled photo interpreter. It is possible, however, to greatly facilitate his work by providing a number of collaborative aids that relieve him of his more mundane and tedious chores [1].

The substance of this report was presented at the April 1977 Image Understanding Workshop, Minneapolis, as a progress report and a separate technical report on a new technique for matching images to symbolic models. Section II is an amplified version of the progress report, which covers the past year, with emphasis upon the last six months. Section III describes the new matching technique.

----

* All references are listed at the end of the report.

II PROGRESS TO DATE

A. Overview

Our work has been centered on evolutionary development toward an integrated interactive system. The system consists of an interactive display console, a map data base, an image library, general image analysis routines, and task specialist routines. At the present time, the system is not a unified whole, but exists as a collection of programs: we are still working toward their integration. The following scenario illustrates the major capabilities that have been demonstrated to date.

The first task when a new image enters the system is to establish correspondence with the map. This is accomplished automatically, by selecting potentially visible landmarks (using navigational data associated with the image) and then locating them in the image using scene analysis techniques. The next step is to confirm the validity of existing knowledge. The system can automatically verify the presence of certain cartographic features, such as roads and waterways, and can also monitor the status of some typical dynamic situations, such as ships berthed in harbor or boxcars stored in a classification yard. New features are identified and incorporated into the data base through the use of a number of interactive aids for mensuration and tracing. For example, new roads can be traced, or heights of bridge supports can be measured.

The system can now use the data base to answer simple queries, such as (in paraphrase), "show me Pier14", "what is this building?" or "how high is that mountain?". These queries are entered by a photo interpreter via keyboard and display cursor. It also has the capability for responding to a more complex query, such as "how many ships were in

Oakland-Harbor yesterday?", by retrieving the relevant image from the library, and then invoking the appropriate task specialist.

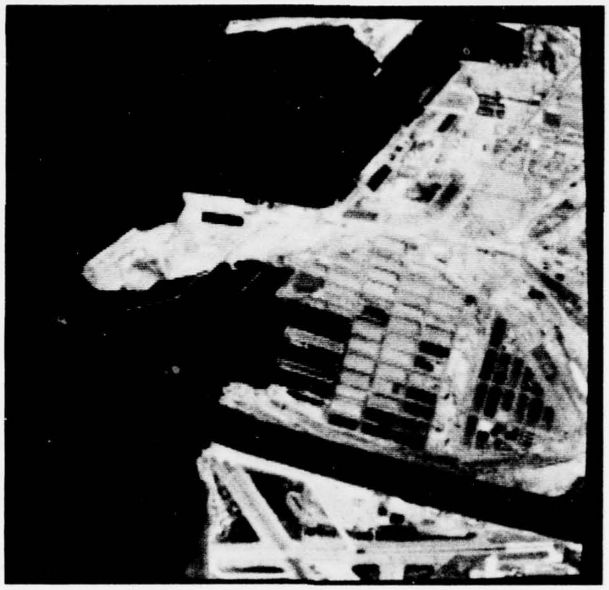
At this time, the questions that can be asked are limited by the small size of the data base and the available specialist routines. The specialists to date are for carefully chosen tasks that could be performed with existing primitive low-level vision capabilities. Moreover, as pointed out earlier, the demonstrated task capabilities do not yet exist as a truly unified system, but rather as a collection of independent programs that share a common data base. They do, however, show the potential of bringing image understanding and artificial intelligence approaches to bear on problems in cartography and photo interpretation.

B. Technical Details

1. Map/image correspondence

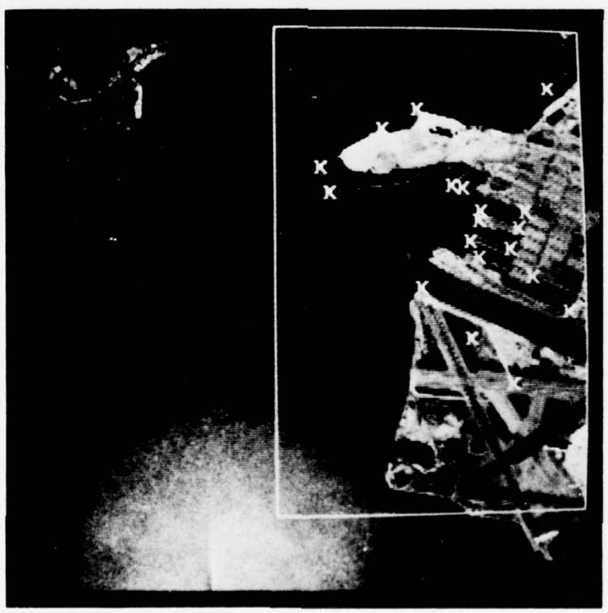
The first task in the scenario is putting the sensed image into geometric correspondence with reference imagery or a map data base. This is fundamental to virtually every military application of imagery. Our initial approach was a modest improvement on conventional image correlation. Given an image, such as Figure 1, and approximate viewpoint, the system determined potentially visible landmarks and then retrieved from the library images containing the landmarks. Figure 2 shows a selected reference image with the area of overlap and the contained landmarks overlaid on it.

For each landmark, an appropriate area of the reference image was extracted and reprojected to make it appear more similar to the sensed image. The reprojection was accomplished using a camera model, calibration data associated with the reference image, and elevation data obtained from the map. The reference image fragment was first projected down onto the ground plane, and thence back up onto the image plane of the sensing camera. Each reprojected image fragment was then correlated in a small predicted area of the sensed image, using Moravec's high-



SA-5300-28

FIGURE 1 A NEW SENSED IMAGE



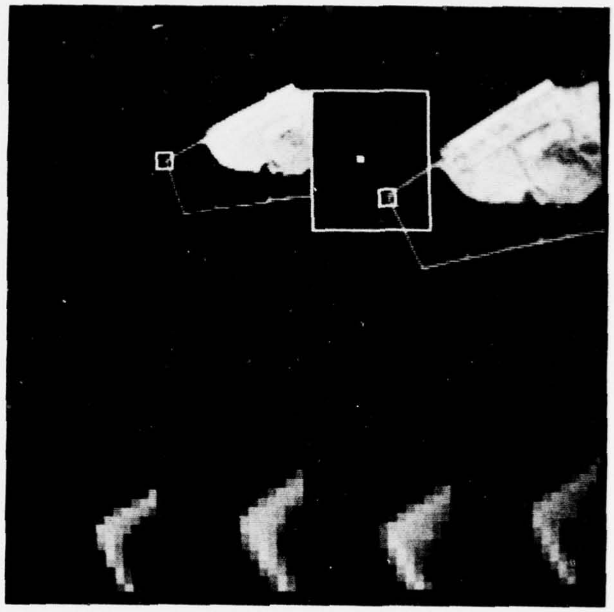
SA-5300-29

FIGURE 2 A SELECTED REFERENCE IMAGE AND LANDMARKS speed algorithm [3]. Figure 3 shows details of the sensed (right top) and reference (left top) images near a landmark. The bottom left detail is the 16x16 image chip surrounding the landmark automatically extracted for use by the system. The landmark is sought in the area delimited by the large square in the sensed image, and the best matching area is shown at bottom midright. The reprojected version of the chip is shown at bottom midleft, and the best matching area at bottom right. Note that the reprojected reference image more closely resembles the sensed image and that the point of correspondence is therefore more precisely located. Figure 4 illustrates improved reliability: without reprojection, the best match is at the wrong location (indicated by X).

The matching process is repeated for all landmarks expected to be visible. This yiellds a set of points in the sensed image, with each point corresponding to a particular landmark (Figure 5). From the pairs of corresponding image and world locations, the exact camera parameters for the sensed image were computed by solving an overconstrained set of equations. We can determine a least-squared-error solution either directly, analytically, or by an iterative parameter optimization process: the latter has the advantage that any known constraints on parameter values can be readily imposed.

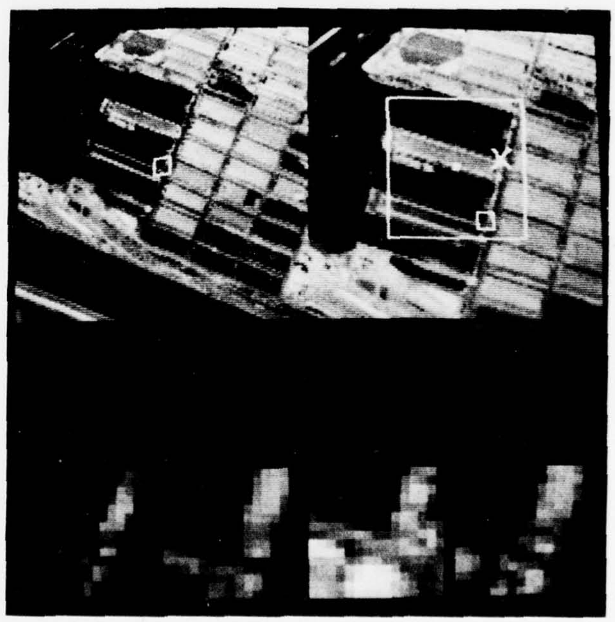
The reprojection technique (unlike currently used techniques) permits the use of reference images that differ radically in viewpoint from the sensed image. Even an oblique image, such as shown in Figure 6, can be matched against the same reference image. Figure 7 shows matching for a single landmark. The views are so different that a meaningful match is impossible without reprojection.

Although reprojection prior to matching is an improvement on conventional image correlation, the fundamental limitation of the correlation approach, namely sensitivity to viewing conditions, remains. In particular, it still cannot match images obtained from radically different viewpoints when the three-dimensional scene structure is complex, from different sensors, or under different illumination or



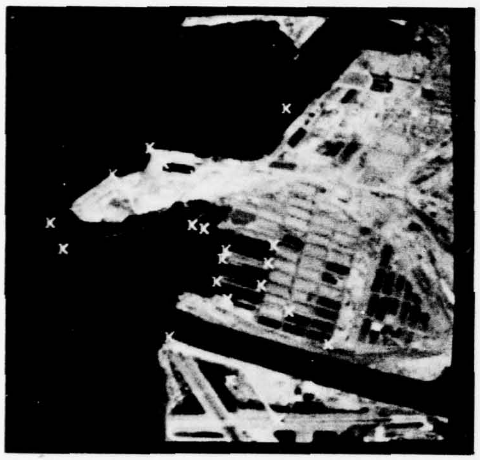
SA-5300-30

FIGURE 3 CORRELATION MATCHING OF AN IMAGE CHIP



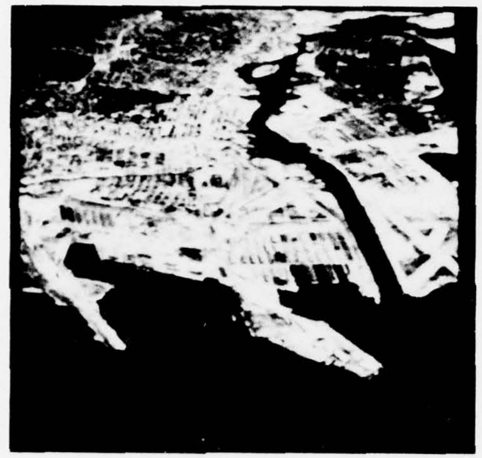
SA-5300-31

FIGURE 4 A MISMATCH WITH AN UNREPROJECTED CHIP



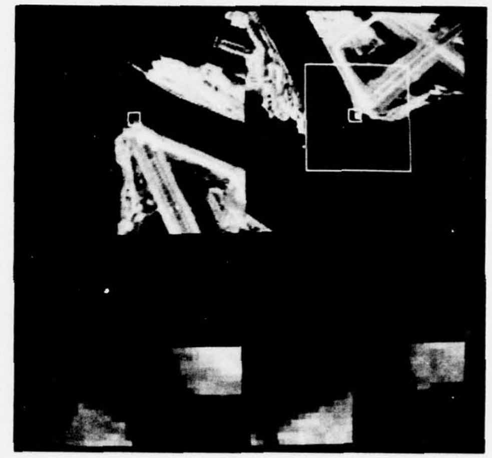
SA-5300-32

FIGURE 5 LANDMARKS LOCATED IN THE SENSED IMAGE



SA-5300-33

FIGURE 6 AN OBLIQUE SENSED IMAGE



SA-5300-34

FIGURE 7 MATCHING VERTICAL AND OBLIQUE VIEWS climatic conditions; and it cannot match images against symbolic maps. To overcome these limitations, we developed a new approach, which we call parametric correspondence, for matching images directly to a three-dimensional symbolic reference map.

The map contains a compact three-dimensional representation of the shape of major landmarks, such as coastlines, buildings, and roads. An analytic camera model is used to predict the location and appearance of landmarks in the image, generating a projection for an assumed viewpoint. Correspondence is achieved by adjusting the parameters of the camera model until the predicted appearances of the landmarks optimally match a symbolic description extracted from the image. The matching of image and map features is performed rapidly by a new technique, called "chamfer matching", that compares the shapes of two collections of shape fragments, at a cost proportional to linear dimension, rather than area. These two new techniques permit the matching of spatially extensive features on the basis of shape, which reduces the risk of ambiguous matches and the dependence on viewing conditions inherent in the conventional correlation based approach. The techniques are described in more detail in Section III. They have obvious application to navigation as well as photo interpretation.

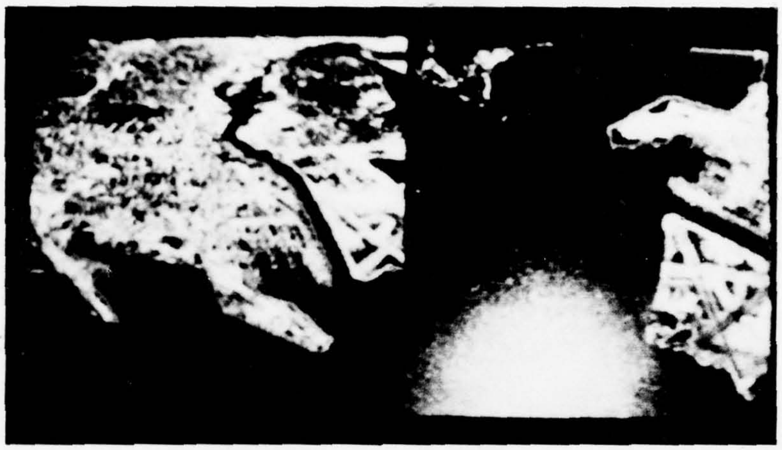
Having placed the image into parametric correspondence with the three-dimensional map, we are now in a position to predict the image coordinates of any feature in the map. Figure 8 shows two pictures with the same section of coastline from the map superimposed on each. This facility is used in monitoring to indicate exactly where in the picture to look. Conversely, we can predict the map features corresponding to any point in the image. This can be used to facilitate interactive graphical communication between the photo interpreter and the data base. In Figure 9, the user has two images displayed simultaneously and can point with a cursor at a location in one image and have the system indicate the corresponding point in the other. (To perform the latter function accurately, the system needs to know the three-dimensional nature of the terrain. We are still in the process of setting up terrain data in the map data base, so in these examples the user supplied the fact that the area in question has roughly constant elevation.)

Using the camera model and image calibration permits many photo interpretation mensuration tasks to be accomplished simply. Routines exist for determining location, length, height, or straight-line distance for features indicated interactively in the image. In Figure 10, the user is measuring the height of a bridge support. Velocity of objects (e.g. ships or cars) indicated in two images can also be determined. In Figure 11, the user indicated a ship in one image, and the system used the landmark finding process to locate the same ship in the other image and hence to determine speed from the deduced distance and the known time delay between the pictures.

The camera model provides a unifying theoretical foundation that subsumes what would otherwise be a collection of ad hoc trigonometric techniques [4]. Combining the map and calibrated image, the system can also, for example, determine alternative routes and travel distances along roads between indicated points.

2. Map-guided monitoring

Having a map and image in correspondence makes many monitoring tasks simpler, because the map can indicate where to look and what to look for in the image. It is important, however, to keep in mind that a map is only an approximation to reality: it may be incomplete, be out of date, suppress details, or contain errors. In order to monitor or to make a detailed interpretation of an image, it is necessary to locate image coordinates of objects more precisely than can be predicted using the map and calibration. In other words, we need routines which can take predictions and verify them in the image. As a first step in that direction, we developed a guided line tracing routine that accepts a rough approximation to the path of linear features, such as rivers or roads, and extracts a best estimate of the precise path in the image.



SA-5300-35

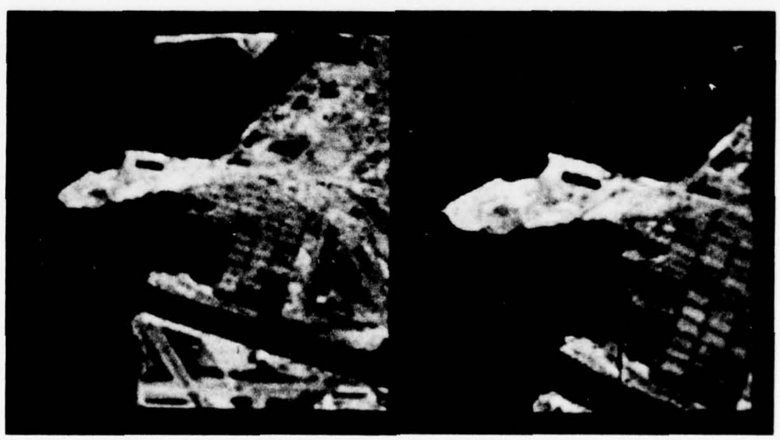
FIGURE 8 THE MAP PROJECTED ONTO TWO PICTURES

SA-5300-36

FIGURE 9 INDICATING CORRESPONDING POINTS IN TWO PICTURES

SA-5300-37

FIGURE 10 MEASURING THE HEIGHT OF A BRIDGE SUPPORT



SA-5300-38

FIGURE 11 MEASURING THE SPEED OF A SHIP

It operates by applying a specially developed line detector in the vicinity of the approximate path and then finding a globally optimal path based on the local feature values [2]. Figure 12 shows the predicted course of a road in a rural area (darker line). The same road has also been predicted without making use of the elevation information in the map (lighter line): note that this prediction is considerably in error. Figure 13 shows the result of the tracing process, obtained fully automatically.

The tracing routine can be used in two ways: to verify the presence of known cartographic features, using prediction from the map and to interactively trace new features for incorporation into the map, using a guideline sketched by the user. The tracing of linear features is currently a tedious manual process that constitutes a major bottleneck in map production [1] [5].

Having a map and image in correspondence makes the automation of many monitoring tasks feasible. Keeping track of boxcars in a railyard, for example, is a typical tedious photo interpretation task. Knowing the layout of the tracks, makes the task essentially a one-dimensional template matching problem. A routine has been developed which flies statistical operators along a track line to hypothesize possible ends of boxcars. These hypotheses are used with knowledge of standard boxcar lengths and characteristics of empty track to locate the gaps between boxcars. The program then reports the number of cars, classified by length [2].

Estimating highway traffic is a similar problem which could be approached by flying car and truck templates along the path determined by the guided road tracer. Recent work at Stanford University could be applied here [6].

Monitoring the presence of ships in a harbor is particularly easy to automate when the map contains details of berths. Given a question about the status of a particular harbor at a particular time, the appropriate image is retrieved from the data base. The ship

SA-5300-39

FIGURE 12 A ROAD PREDICTED FROM THE MAP

SA-5300-40

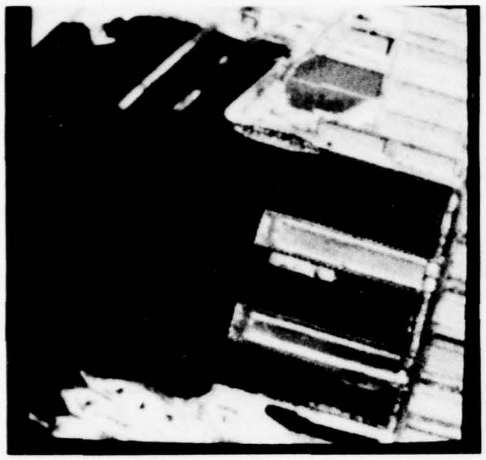
FIGURE 13 THE ROAD AFTER AUTOMATIC TRACING monitoring routine then projects berth locations from the map onto the image (Figure 14) and uses an edge histogram of that region to determine whether the berth is occupied (Figure 15). The same process works equally well for vertical or oblique imagery as shown in Figure 16.

The key to automatic monitoring lies in having the capability to place the image into correspondence with the map, which then accurately specifies where to look. A relatively simple test may then be used in that limited context. We have implemented three representative demonstrations of this approach and believe that many others are possible, especially in remote sensing [7]. In a production environment, such monitoring could be performed automatically on a continuing basis as new imagery arrived.

3. Map data base

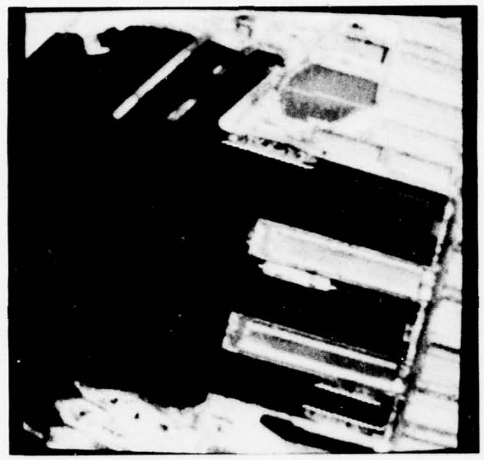
The underlying foundation on which much of the foregoing rests is the map data base. We have implemented a disk-based semantic net data structure that can contain realistic quantities of data represented in a way which permits efficient access. Entities are represented by LISP atoms (e.g. English words), and information associated with the entity is stored in a property list format. Relationships to other entities are also stored on the property lists, thus establishing a network structure in the data base. When information concerning a particular entity is sought, the property list is retrieved from disk and established in core. A "paging" scheme limits the amount of data in core (to, say, 1000 entities) and writes entities back out to disk, if necessary, the least recently used ones first [2]. Indexing of the information is by means of a hash table on disk, which means that access time is constant and independent of data base size.

We are in the process of setting up a map of the San Francisco Bay Area, containing major features, coastlines, bridges, and highways. Figure 17 is a portion of a U.S. Geological Survey (USGS) map of the area; Figure 18 shows the portion of the map currently in the data base.



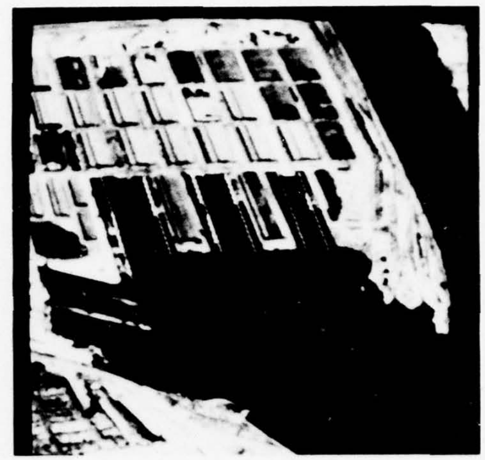
SA-5300-41

FIGURE 14 HARBOR PIERS PREDICTED FROM THE MAP



SA-5300-42

FIGURE 15 BERTHED SHIPS DETECTED

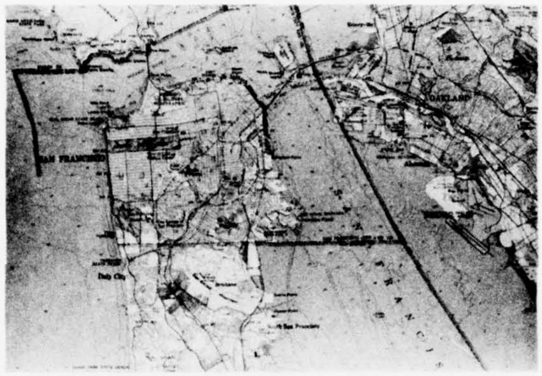


SA-5300-43

FIGURE 16 BERTHED SHIPS IN AN OBLIQUE IMAGE

Figure 19 shows part of the map at higher resolution. The map consists of about 4000 points, plus various semantic relationships, totaling about three-quarters of a million bytes of disk storage. (Access to a particular item of information takes less than a millisecond if it is paged in, and fifteen to thirty milliseconds plus disc access time if it has to be read in). The types of feature currently recorded in the data base include coastlines, major roads, lakes, bridges, airfield runways, oil storage tanks, and harbor lights. The information was derived by manually tracing features on a USGS map using a digitizing table: map data in digital form are not available, and the problem of digitizing printed maps has rather different constraints from the problem of making maps from photographs, so we could not exploit our guided tracing techniques.

We are still in the process of digitizing data for the map data base and of setting up the higher-level concepts, such as harbors, towns and so forth, above the level of basic geometry and topology. The geometric data are indexed (the index structure is part of the data base) via a K-D tree [8] to enable fast retrieval of information relevant to a particular area. In addition to the three dimensional description of cartographic and cultural features, the map contains a partial taxonomy of world entitie¬, with relevant general semantics, a catalogue of available imagery, and descriptions of data structures used by the system. The descriptions of the data structures enable the system to construct automatically new entities of the correct structure for inclusion in the data base.

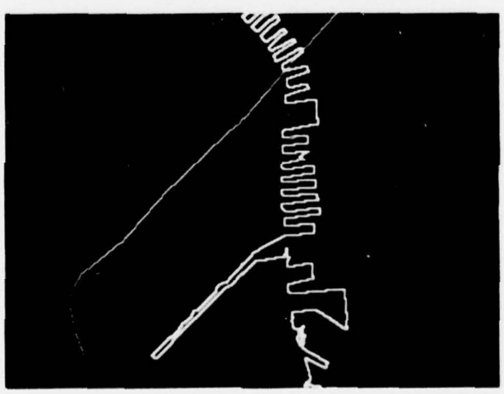


SA-5300-44

FIGURE 17 A USGS MAP OF THE SAN FRANCISCO BAY AREA

SA-5300-45

FIGURE 18 DISPLAY OF THE DIGITAL MAP DATA BASE



SA-5300-46

FIGURE 19 PART OF THE MAP AT HIGHER RESOLUTION

III PARAMETRIC CORRESPONDENCE AND CHAMFER MATCHING

A. Introduction

Many military tasks involving pictures require the ability to put a sensed image into correspondence with a reference image or map. Examples include vehicle guidance, photo interpretation (change detection and monitoring) , and cartography (map updating). The conventional approach is to determine a large number of points of correspondence by correlating small patches of the reference image with the sensed image. A polynomial interpolation is then used to estimate correspondence for arbitrary intermediate points [9]. This approach is computationally expensive and limited to cases where the reference and sensed images were obtained under similar viewing conditions. In particular, it cannot match images obtained from radically different viewpoints, sensors, or seasonal or climatic conditions, and it cannot match images against symbolic maps.

Parametric correspondence matches images to a symbolic reference map rather than to a reference image. The map contains a compact three-dimensional representation of the shape of major landmarks, such as coastlines, buildings, and roads. An analytic camera model is used to predict the location and appearance of landmarks in the image, generating a projection for an assumed viewpoint. Correspondence is achieved by adjusting the parameters of the camera model (i.e. the assumed viewpoint) until the appearances of the landmarks optimally match a symbolic description extracted from the image.

The success of this approach requires the ability to rapidly match predicted and sensed appearances after each projection. The matching of image and map features is performed by a new technique, called "chamfer matching", that compares the shapes of two collections of curve fragments at a cost proportional to linear dimension rather than area.

In principle, this approach should be superior, since it exploits more knowledge of the invariant three dimensional structure of the world and of the imaging process. At a practical level, this permits matching of spatially extensive features on the basis of shape, which reduces the risk of ambiguous matches and dependence on viewing conditions.

B. Chamfer Matching

Point landmarks, such as road intersections or promontories, are represented in the map with their associated three-dimensional world coordinates. Linear landmarks, such as roads or coastlines, are represented as curve fragments with associated ordered lists of world coordinates. Volumetric structures, such as buildings or bridges, can be represented as wire-frame models.

From a knowledge of the expected viewpoint, a prediction of the image can be made by projecting world coordinates into corresponding image coordinates, suppressing hidden lines. The problem in matching is to determine how well the predicted features correspond with image features, such as edges and lines.

The first step is to extract image features by applying edge and line operators or tracing boundaries. Edge fragment linking [15], [11] or relaxation enhancement [12], [2] is optional. The net result is a feature array; each element of the array records whether or not a line fragment passes through it. This process preserves shape information and discards greyscale information, which is less invariant.

To correlate the extracted feature array directly with the predicted feature array would encounter several problems: The correlation peak for two identical curves is very sharp and therefore intolerant of slight misalignment or distortions [13]: A sharply peaked correlation surface is an inappropriate optimization criterion because it provides little indication of closeness to the true match or of the proper direction in which to proceed: Computational cost is heavy with large feature arrays. A more robust measure of similarity between the two sets of feature points is the sum of the distances between each predicted feature point and the nearest image point. This can be computed efficiently by transforming the image feature array into an array of numbers representing distance to the nearest image feature point. The similarity measure is then easily computed by stepping through the list of predicted features and simply summing the distance array values at the predicted locations. The distance values can be determined by a process known as "chamfering", in two passes through the image feature array [14], [16]. Note that this determination is made only once, after image feature extraction.

Chamfer matching provides an efficient way of computing the integral distance (i.e. area), or integral squared distance, between two curve fragments, two commonly used measures of shape similarity.

C. Parametric Correspondence

Parametric correspondence puts an image into correspondence with a three-dimensional reference map by determining the parameters of an analytic camera model (three position and three orientation parameters).

The traditional method of calibrating the camera model takes place in two stages: first, a number of known landmarks are independently located in the image; second, the camera parameters are computed from the pairs of corresponding world and image locations, by solving an overconstrained set of equations [17], [18], [19].

The failings of the traditional method stem from the first stage. The landmarks are found individually, using only very local context (e.g. a small patch of surrounding image) and with no mutual constraints. Thus, local false matches commonly occur. The restriction to small features is mandated by the high cost of area correlation, and by the fact that large image features correlate poorly over small changes in viewpoint.

Parametric correspondence overcomes these failings by integrating the landmark-matching and camera-calibration stages. It operates by hill-climbing on the camera parameters. A transformation matrix is constructed for each set of parameters considered, and it is used to project landmark descriptions from the map onto the image at a particular translation, rotation, scale, and perspective. A similarity score is computed with chamfer matching and used to update parameter values. Initial parameter values are estimated from navigational data.

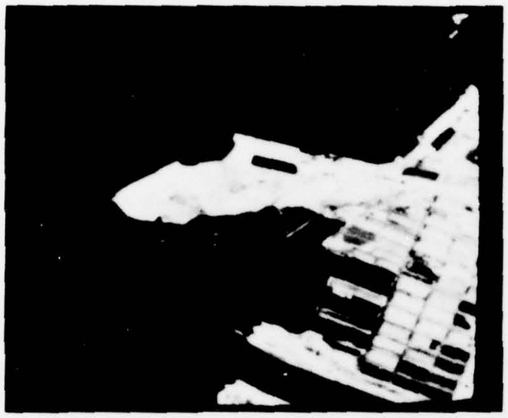
Integrating the two stages allows the simultaneous matching of all landmarks in their correct spatial relationships. Viewpoint problems with extended features are avoided because features are precisely projected by the camera model before matching. Parametric correspondence has the same advantages as rubber-sheet template matching [20], [21] in that it obtains the best embedding of a map in an image, but avoids the combinatorics of trying arbitrary distortions by only considering those corresponding to some possible viewpoint.

D. An Example

The following example illustrates the major concepts in chamfer matching and parametric correspondence. A sensed image (Figure 20) was input along with manually derived initial estimates of the camera parameters. A reference map of the coastline was obtained by using a digitizing tablet to encode coordinates of a set of 51 sample points on a USGS map. Elevations for the points were entered manually. Figure 21 is an orthographic projection of this three-dimensional map.

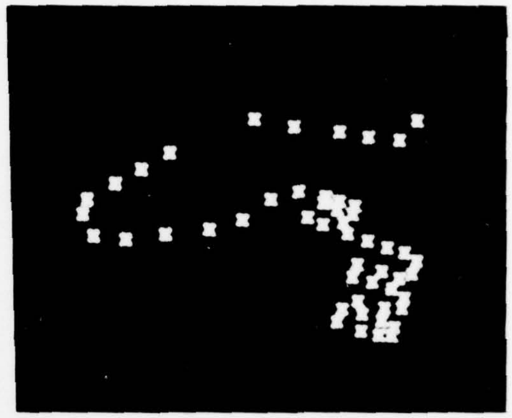
A simple edge follower traced the high contrast boundary of the harbor, producing the edge picture shown in Figure 22. The chamfering algorithm was applied to this edge array to obtain a distance array. Figure 23 depicts this distance array; distance is encoded by brightness with maximum brightness corresponding to zero distance from an edge point.

Using the initial camera parameter estimates, the map was projected onto the sensed image (Figure 24). The average distance between projected points and the nearest edge point, as determined by chamfer matching, was 25.8 pixels.



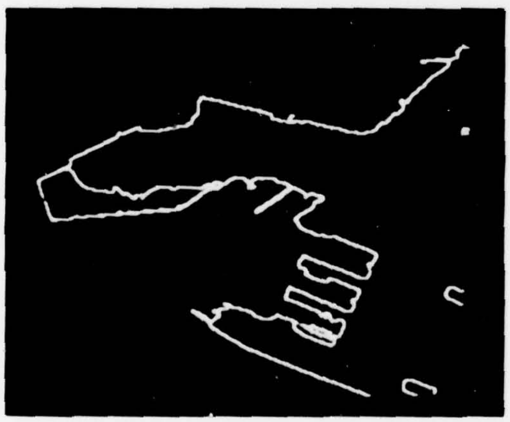
SA-5300-47

FIGURE 20 AN AERIAL IMAGE OF A SECTION OF COASTLINE



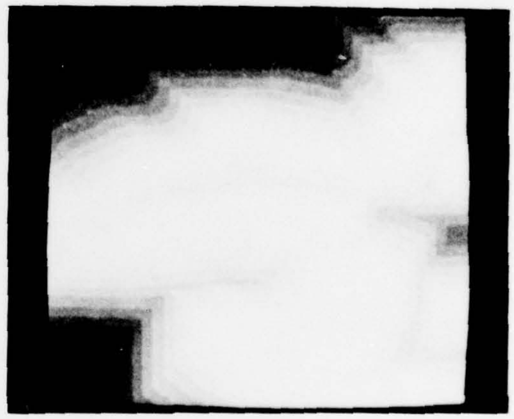
SA-5300-48

FIGURE 21 A SET OF SAMPLE POINTS TAKEN FROM A USGS MAP



SA-5300-49

FIGURE 22 THE TRACED BOUNDARY OF THE COASTLINE

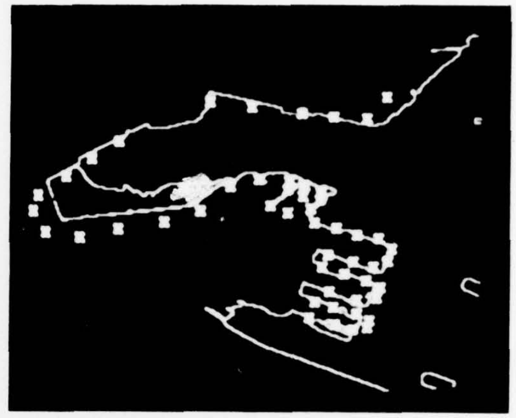


SA-5300-50

FIGURE 23 THE DISTANCE ARRAY PRODUCED BY CHAMFERING

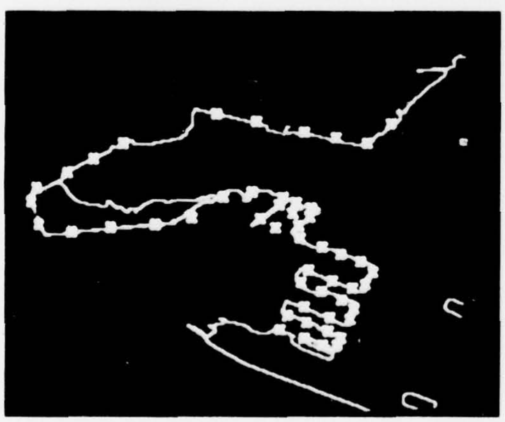
SA-5300-51

FIGURE 24 INITIAL PROJECTION OF MAP POINTS ONTO THE IMAGE



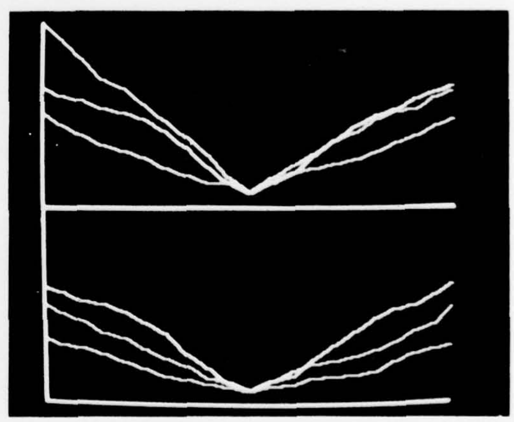
SA-5300-52

FIGURE 25 PROJECTION AFTER SOME ADJUSTMENT OF PARAMETERS



SA-5300-53

FIGURE 26 PROJECTION AFTER OPTIMIZATION OF PARAMETERS



SA-5300-54

FIGURE 27 VARIATION OF DISTANCE SCORE NEAR THE OPTIMUM

A straightforward optimization algorithm adjusted the camera parameters to minimize the average distance. Figure 25 and Figure 26 show an intermediate state and the final state, in which the average distance has been reduced to 0.8 pixels. This result, obtained with 51 sample points, compares favorably with a 1.1 pixel average distance for 19 sample points obtained using conventional image chip correlation followed by camera calibration. The curves in Figure 27 characterize the local behavior of this minimum, showing how average distance varies with variation of each parameter from its optimal value.

E. Discussion

We have developed a scheme for establishing correspondence between an image and a reference map that integrates the processes of landmark matching and camera calibration. The potential advantages of this approach stem from (1) matching shape rather than brightness; (2) matching spatially extensive features rather than small patches of image; (3) matc`ing simultaneously to all features, rather than searching the combinatorial space of alternative local matches; and (4) using a compact three dimensional model rather than many two-dimensional templates.

Shape has proved to be much easier to model and predict than brightness. Shape is a relatively invariant geometric property whose appearance from arbitrary viewpoints can be precisely predicted by the camera model. This eliminates the need for multiple descriptions, corresponding to different viewing conditions, and overcomes difficulties of matching large features over small changes of viewpoint.

The ability to treat the entirety of the relevant portion of the reference map as a single extensive feature reduces significantly the risk of ambiguous matches. It also avoids the combinatorial complexity of finding the optimal embedding of multiple local features.

A number of obstacles have been encountered in reducing the above ideas to practice. The distance metric used in chamfer matching provides a smooth, monotonic measure near the correct correspondence and nicely interpolates over gaps in curves. However, scores can be unreliable when image and reference are badly out of alignment. In particular, discrimination is poor in textured areas, aliasing can occur with parallel linear features, and a single isolated image feature can support multiple reference features.

The main problem is that edge position is not a distinguishing feature; consequently, many alternative matches receive equal weight. One way of overcoming this problem, therefore, is to use more descriptive features: brightness discontinuities can be classified, for example, by orientation, by edge or line, and by local spatial context (texture versus isolated boundary). Each type of feature would be separately chamfered and map features would be matched in the appropriate array. Similarly, features at a much higher level could be used, such as promontory or bay, area features having particular internal textures or structures, and even specific landmarks, such as the top of the Transamerica pyramid. Ideally, with a few highly differentiated features distributed widely over the image, the parametric correspondence process would be able to home in directly on the solution regardless of initial conditions.

Another dimension for possible improvement is the chamfering process itself. Determining for each point of the array a weighted sum of distances to many features (e.g. a convolution with the feature array), instead of the distance to the nearest feature, would provide more immunity from isolated noise points. Alternatively, propagating the coordinates of the nearest point instead of merely the distance to it enables the use of characteristics of features, such as local slope or curvature, in evaluating the goodness of match. Further, since corresponding pairs of points are now known, it makes possible a more directed search, and an improved set of parameter estimates can be analytically determined.

Chamfer matching and parametric correspondence are separable techniques. Conceptually, parametric correspondence can be performed by reprojecting image chips and evaluating the match with correlation. However, the cost of projection and matching grows with the square of the template size: The cost for chamfer matching grows linearly with the number of feature points. Chamfer matching is an alternative to other shape-matching techniques, such as chain-code correlation [22], Fourier matching [23], and graph matching (e.g. [24]). Also, the smoothing obtained by transforming two edge arrays to distance arrays via chamfering can be used to improve the robustness of conventional area-based edge correlation.

Parametric correspondence, in its most general form, is a technique for matching two parametrically related representations of the same geometric structure. The representations can be two- or three-dimensional, iconic or symbolic; the parametric relation can be perspective projection, a simple similarity transformation, a polynomial warp, and so forth. This view is similar to rubber-sheet template matching as conceived by Fischler and Widrow [20], [21]. The feasibility of the approach in any application, as Widrow points out, depends on efficient algorithms for "pattern stretching, hypothesis testing, and pattern memory", corresponding to our camera model, chamfer matching, and three-dimensional map.

As an illustration of its versatility, the technique can be used with a known camera location to find a known object whose position and orientation are known only approximately. In this case, the object's position and orientation are the parameters; the object is translated and rotated until its projection best matches the image data. Such an application has a more iconic flavor, as advocated by Shepard [25], and is more integrated than the traditional feature extraction and graph matching approach ([26], [27] and [28]).

As a final consideration, the approach is amenable to efficient hardware implementation. Commercially available hardware already exists for generating parametrically specified perspective views of wire frame models at video rates, complete with hidden line suppression. The chamfering process itself requires only two passes through an array by a local operator, and match scoring requires only summing table lookups in the resulting distance array.

F. Conclusion

Iconic matching techniques, such as correlation, are known for efficiency and precision obtained by exploiting all available pictorial information, especially geometry. However, these techniques are overly sensitive to changes in viewing conditions and cannot make use of non pictorial information. Symbolic matching techniques, on the other hand, are more robust because they rely on invariant abstractions, but are less precise and less efficient in handling geometrical relationships. Their applicability in real scenes is limited by the difficulty of reliably extracting the invariant description. The techniques we have put forward offer a way of combining the best features of iconic and symbolic approaches.

REFERENCES

1. Barrow, H. G., et al., "Interactive Aids for Cartography and Photo Interpretation", in Artificial Intelligence--Research and Applications, Annual Progress Report to ARPA, Contract DAAG29-76-C-0012, Artificial Intellligence Center, Stanford Research Institute, Menlo Park, CA 94025 (June 1976).

2. Barrow, H. G., "Interactive Aids for Cartography and Photo Interpretation", Semiannual Technical Report to ARPA, Contract DAAG29-76-C-0012, Artificial Intellgence Center, Stanford Research Institute, Menlo Park, CA 94025 (November 1976).

3. Moravec, H. P., "Techniques Towards Automatic Visual Obstacle Avoidance", submitted to 5IJCAI.

4. "The Manual of Photogrammetry", Ed. R. N. Colwell, published by American Society of Photogrammetry, Falls Church, Virginia (1972).

5. Garvey, T. D., and J. M. Tenenbaum, "Application of Interactive Scene Analysis Techniques to Cartography", Tech Note No. 127, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, CA 94025 (November 1976)(also published in Proc. 3IJCAI).

6. Binford, T. O., In Proceedings of the ARPA Image Understanding Workshop, University of Southern California, April 1976.

7. Tenenbaum, J. M., et al., "Research in Interactive Scene Analysis", Final Report to NASA, Contract No. NASW-2865, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, CA 94025 (December 1976).

8. Bentley, J. L., "Multidimensional Binary Search Trees Used for Assoiative Searching", CACM, Vol. 18, No. 9 (September 1975).

9. Bernstein R., "Digital Image Processing of Earch Observation Sensor Data", IBM Journal of Research and Development, Vol. 20, No. 1, 1976.

10. Nevatia, R., In Proceedings of the ARPA Image Understanding Workship, University of Southern Calfornia, April 1976.

11. Perkins, W. P., "Multi-level Vision Recognition System", Proc. Third IJCPR, pp. 739-744 (November 1976).

12. Zucker, S., R. Hummel and A. Rosenfeld, "An Application of Realization Labeling to Curve and Line Enhancement", IEEE Trans. Comp, No. 25 (1976).

13. Andrus, J. F., C. W. Campbell and R. R. Jayroe, "A Digital Image Registration Method Using Boundary Maps", IEEE Trans. Comp., pp. 935-939 (September 1975)

14. Munson, J., Internal Memo, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, 94025 (December 1973).

15. Nevatia, R., In Proceedings of the ARPA Image Understanding Workshop, University of Southern Calfornia, April 1976.

16. Rosenfeld, A. and J. L. Pfaltz, "Distance Functions on Digital Pictures", Pattern Recognition, Vol. 1 No. 1, pp. 33-62 (July 1968).

17. Sobel, I., "On Calibrating Computer Controlled Cameras for Perceiving 3-D Scenes", Artificial Intelligence, Vol. 5, p.185-198 (1974).

18. Quam, L .H., "Computer Comparison of Pictures", Stanford AI Project Memo AIM-144 (May 1971, Ph.D. Thesis).

19. Hannah, M. J., "Computer Matching of Areas in Stereo Images", Stanford AI Memo AIM-239, (July 1974, Ph.D. Thesis).

20. Fischler, M. and R. Elschlager, "The Representation and Matching of Pictorial Structures", IEEE Trans. Comp., No. 22, pp.67-92 (1973).

21. Widrow, B., "The 'Rubber-Mask' Technique", Pattern Recognition, Vol. 5, pp.175-211 (1973).

22. Feder, J. and H. Freeman, "Segment Fitting of Curves in Pattern Analysis Using Chain Correlation", AD619525 (March 1965).

23. Zahn, C. and R. Roskies, "Fourier Descriptors for Plane Closed Curves", IEEE Trans. Comp., No. 21, pp.269-281 (1972).

24. Davis, L., "Shape Matching Using Relaxation Techniques", TR-480, Computer Science Dept., University of Maryland, September 1976.

25. Shepard, R. N. and J. Metzler, "Mental Rotation of Three-Dimensional Objects", Science, pp.701-703 (1971).

26. Roberts, L. G., "Machine Perception of Three-Dimensional Objects", in: Tippet, J. T., et al. (Eds.), "Optical and Electro-Optical Information Processing", MIT Press, pp.159-197 (1965).

27. Falk, G., "Interpretation of Imperfect Line Data as a Three-Dimensional Scene," Artificial Intelligence, Vol. 3, pp.101-144 (1972).

28. Grape, G. R., "Model Based (Intermediate Level) Computer Vision", Stanford AI Memo AIM-201, (May 1973, Ph.D. Thesis).

29. Brown, C. M., and K. Lantz, "Representation and Use of Knowledge in a Goal-Directed Vision System", In Proceedings of the ARPA Image Understanding Workshop, Minneapolis, April 1977.