The Visual Components of an Automated Inspection Task

Linda G. Shapiro[1]
Robert M. Haralick[1]
Ting-Chuen Pong[2]

[1]Machine Vision International
Ann Arbor, Michigan

[2]Department of Computer Science
University of Minnesota

ABSTRACT

A large complex machined object (an F-15 bulkhead) is to be inspected by a computer system employing television cameras and robot arms with sensing devices. The system will use vision, planning, and reasoning to guide the robot arms. The vision portion of the system must determine, from a stereo pair of images of a piece of the bulkhead, the set of parts of the bulkhead that are currently being viewed by the cameras and the exact positions of the cameras.

We have designed a hierarchical, relational object model for representing the bulkhead object and other three dimensional objects in the scene, such as the robot arms and tools. We have also designed indexing schemes to go along with the bulkhead model, so that the vision system can quickly determine which parts of the bulkhead intersect a given volume in three-dimensional space. Thus, given an estimate of the location of the cameras, we can select a subset of the bulkhead model to be used in matching.

Because the bulkhead is a shiny object, the images have problems with specular reflectance. We have designed procedures to segment the images and extract both regions and bounding arcs, plus a variety of attributed spatial relations. Using regions and arcs as primitives, we have developed an algorithm to find correspondences between the images of a stereo pair. Primitives in one image that have no corresponding primitive in the second can be discounted in matching to the model. Finally, we are now designing algorithms to find the correspondence between the primitives extracted from the images and the selected subset of parts from the model. Finding this correspondence will enable the computation of the positions of the cameras and robot arms.

I. INTRODUCTION

A vision system is being designed for use in an automated inspection task. The object to be inspected is an F-15 bulkhead, a large, metal, machined object of width 120 inches, height 40 inches, and weight 165 pounds. The scenario includes two robot arms, each mounted on an independent translating base. The bulkhead is mounted on a translating table between the two robot arms. The robot arms will use measurement devices to test the thickness of various parts of the bulkhead. The arms will be guided by a planning system that gets some of its input from a stereo vision system. The vision system will input a pair of images from cameras mounted on the arms and will determine the parts of the bulkhead being viewed and the positions of the cameras and robot arms. The scenario is illustrated in Figure 1.

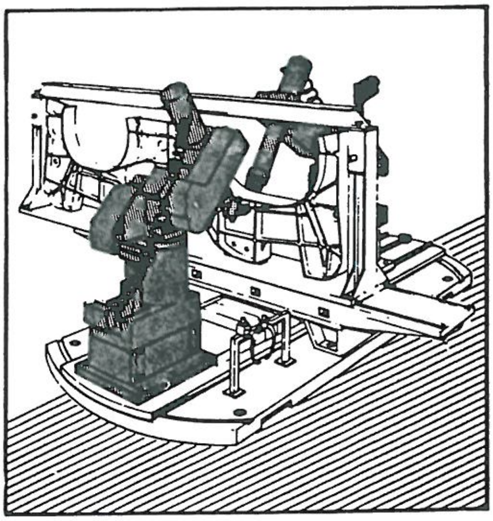

Figure 1 illustrates the F-15 bulkhead and robot arms.

The vision system uses a hierarchical three-dimensional model of the bulkhead. At any given time, it has an estimate of the current position of the cameras. Thus the approximate volume in three-space that the cameras are viewing can be determined. With this information, the vision system must select a subset of the parts of the bulkhead for attempted matching. Regions and arcs extracted from the images will be used first in stereo matching to help determine which really correspond to parts and part boundaries and which are due to illumination or errors in segmentation. After the stereo matching phase, the exact correspondences between the two-dimensional regions and arcs and the three-dimensional surfaces and bounding arcs of the parts of the model must be found.

II. THE HIERARCHICAL RELATIONAL MODEL

The hierarchical relational model used to describe the bulkhead has been described in detail in [1]. To briefly summarize, the bulkhead description is a hierarchic description with four main levels: world level, object level, part level, and surface/arc level. The world level, at the top of the hierarchy is concerned with the arrangement of objects in the world that the camera(s) is(are) looking at. The object level is concerned with the arrangement of parts in each possible object. The parts level describes the relationships between surfaces of a part and also specifies the actual surfaces and arcs, at the physical point level, but represented by functions over bounded domains. At each level are spatial data structures (SDSs), each consisting of a set of relations giving the features and relationships of an entity at that level. Some of these relations contain information that is only stored in one place, while others may contain redundant information organized in a different way for convenient access. Each SDS contains its own attribute-value table relation to store the global properties associated with that entity.

At each level of the hierarchy, there are instances of entities at lower levels. Each instance consists of a pointer to the lower level entity plus a transformation which is used to transform the physical points of the lower level entity to the coordinate system being used in the current level. This is a standard technique used in computer graphics that enables sharing of entities at any level.

III. INDEXING SCHEMES

An indexing mechanism is required that can quickly access the set of parts in the hierarchical relational model of the bulkhead that lie in a given volume in three-space. There are a number of possible schemes. One obvious method is to create an oct-tree representation of three-dimensional space. Each octant is either empty, intersects exactly one part, or can be further subdivided. Thus the oct-tree would be an auxilliary structure with pointers to the parts in the hierarchical, relational model. Given a volume in three-space, the access procedure would start at the root of the oct-tree and trace down all paths leading from the root whose nodes intersected the volume until the parts were found. If the volume boundaries did not correspond well to the boundaries in the oct-tree, this could waste a lot of time.

A second possibility is to partition all of three-space with a three-dimensional grid of equal-sized rectangular cells. Each cell would contain pointers to the parts that intersect it. Given a volume in three-space, the access procedure could determine, in constant time, those grid cells that intersect the volume. It would then need to take the union of the sets of parts pointed to by each grid square. Thus the time would be proportional to the number of grid cells involved. Since the boundaries of the grid cells would not correspond exactly to the boundary of the volume in 3-space, some extra parts might be retrieved. An optimal grid cell size could be predetermined based on the above two considerations.

Suppose that the three-dimensional volume is always a rectangular parallelepiped with sides parallel to the three axes. (If it is not, it can be enclosed by one that is.) A third possibility is to think of the volume in terms of the straight line segments that bound the faces of its rectangular surfaces. If the parts are also represented by (bounding) rectangular parallelepiped, then a generalization of some standard techniques in computational geometry can be used to determine which parts intersect the volume. We have developed a technique which can determine whether one rectangular parallelpiped box intersects any of a set of N rectangular parallelpiped in $\log^3 N$ operations.

IV. EXTRACTING PRIMITIVES FROM IMAGES

Haralick [5] proposed a rich and robust representation for all types of two-dimensional intensity variations. This representation is called the topographic primal sketch. We have developed a feature extraction scheme which extracts topographic structures such as edges, ridges, valleys and hillsides. Notice that edge is included in the set of topographic structures. Edges, which usually correspond to sharp changes in gray tone intensities, are used to describe the basic structure of an image. The rest of the topographic structures are used to describe the intensity variations within regions which are extracted as connected sets of non-edge structure.

The feature extraction procedure is divided into six steps:

(1) detect edge elements by using the second directional derivative zero-crossing edge operator,

(2) form arc segments by linking edges of similar orientations,

(3) clear edges by deleting isolated short arc segments,

(4) fill gaps between arc segments,

(5) extract regions by forming maximum connected sets of non-edge pixels, and

(6) compute topographic structures within regions.

V. STEREO MATCHING

We are considering two-different stereo matching procedures using (subsets of) the primitives extracted above. The first is the work of Pong [3]. His matching process can be summarized in three steps:

1. an epipolar line matching process which uses epipolar geometry to reduce the problem to an one-dimensional matching process;

2. a segment matching process which establishes globally optimal matches for region and arc segments; and

3. a global matching process which uses spatial relations among high-level structures to resolve ambiguous matches.

Figure 2 shows an example of Pong's results on a stereo pair of bulkhead images.

A second relational method whose implementation is now in progress can be summarized as follows.

Let I_1 and I_2 be the two images. Each image I_i is segmented into a set of regions R_i and a set of bounding arcs A_i. We will call $U = R_1 \cup A_1$ the set of <u>units</u> and $L = R_2 \cup A_2$ the set of <u>labels</u>. We also extract from each image a set of relationships such as colinearity, adjacency, parallelness, and so on. We would like to find a mapping from U to L that preserves as many relationships as possible -- i.e. a minimum distance mapping. As in Cheng and Huang [4], we want to use the spatial relationships among the units (and those among the labels) to severely constrain the search for the best mapping.

Each unit $u \in U$ has a neighborhood $NBD(u) \subseteq U$ of units that stand in some spatial relationship to u and are also physically near it in the image. Similarly, each label $l \in L$ has a neighborhood $NBD(l) \subseteq L$. Intuitively, when u maps to l, we would expect the neighbors of u to map to the neighbors of l. In fact, there will often be enough information in the extracted relations to determine exactly which neighbor of u must map to a particular neighbor of l when u maps to l. For example, the parallel relation has tuples of the form (a_1,a_2,d,θ) where a_1 and a_2 are parallel arcs, d is the distance between them, and θ is the direction of the vector from a_1 to a_2 that gives the shortest distance. Suppose arc u_1 in I_1 maps to arc l_1 in I_2, and (u_1,u_2,d,θ) is an element of the parallel relation for I_1, $u_2 \in NBD(u_1)$. If there is only one arc $l_2 \in NBD(l_1)$ with (l_1,l_2,d,θ) an element of the parallel relation for I_2, then u_2 must map to l_2. In reality, the distances and angles will not agree precisely, and more than one neighbor of l_1 may have to be considered.

For each $u \in U$ and $l \in L$ satisfying that u and l are either both arcs or both regions, there are zero or more potential mappings from the neighborhood of u to the neighborhood of L that map u to l. The information in the extracted relations can be used to generate for each pair (u,l) the set

$$F_{u,l} = \{f : NBD(u) \rightarrow NBD(l) \mid f(u) = l \text{ and, if } f(u') = l', \text{ the relationships between } u \text{ and } u' \text{ correspond to those between } l \text{ and } l'.\}$$

Let $F_u = \bigcup_{l \in L} F_{u,l}$. Then the exact matching problem of finding a distance $\emptyset$ mapping from U to L can be defined as finding a mapping $h: U \rightarrow L$ satisfying that for each $u \in U$, $h|_{NBD(u)} \in F_{u,h(u)}$, and the inexact matching problem of finding the best mapping can be defined as finding a mapping $h: U \rightarrow L$ that minimizes

$$\sum_{u \in U} \min_{f \in F_{u,h(u)}} Error(f, h|_{NBD(u)}),$$

where $Error(f_1,f_2) = \#\{u \in dom(f_1) \cap dom(f_2) \mid f_1(u) = f_2(u)\}$.

Although this is just an application of finding the relational distance between two structural descriptions, we can derive an efficient tree search procedure using the pregenerated sets F_u for each $u \in U$. The essence of the algorithm is as follows.

Keep an error table that associates with each unit and partial function pair $(u, f_{u,l})$ the error between the mapping being constructed and that pair. Initially there is only an empty mapping and all table entries are $\emptyset$. Select a unit u and a function $f_{u,l}$ that has less accumulated error than any other function $f_{u,l'}$. For each future unit u', add $ERROR(f_{u,l},f_{u',l'})$ to the error of each $f_{u',l'}$. If the sum over all future units u' of the minimal error of any $f_{u,l}$ is larger than the error of the best mapping so far, back up. Otherwise, map each uninstantiated unit u' in $domain(f_{u,l})$ to $f_{u,l}(u')$ and continue.

This algorithm instantiates several mappings of units to labels at each node of the search tree and is thus more efficient than the standard treesearch algorithm. It is a generalization of the forward checking algorithm of [4]. Cheng and Huang [3] present an alternative procedure. be made more efficient for problems of this type.

Summary

We have described the problems that must be solved by the visual component of an automated inspection system. We have developed a three-dimensional model to represent the object being inspected. We have given several alternative indexing schemes to allow fast access to those parts in the model that fall in a specified three-dimensional volume. We have summarized a procedure for extracting features from a pair of stereo images of a portion of the object and two procedures for stereo matching. The algorithm for finding the correspondence between the two-dimensional features and the selected three-dimensional parts is still under development.

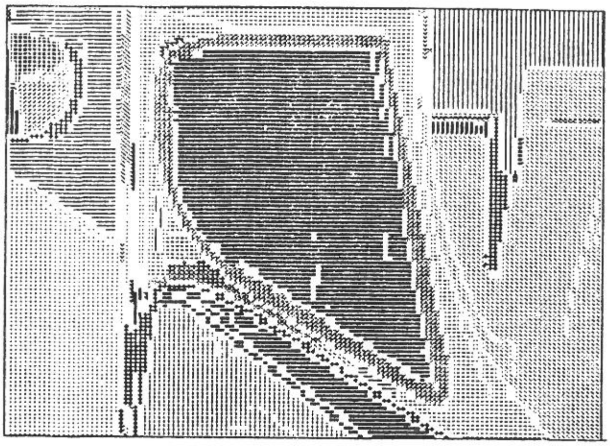

Figure 2a illustrates regions of the left image.

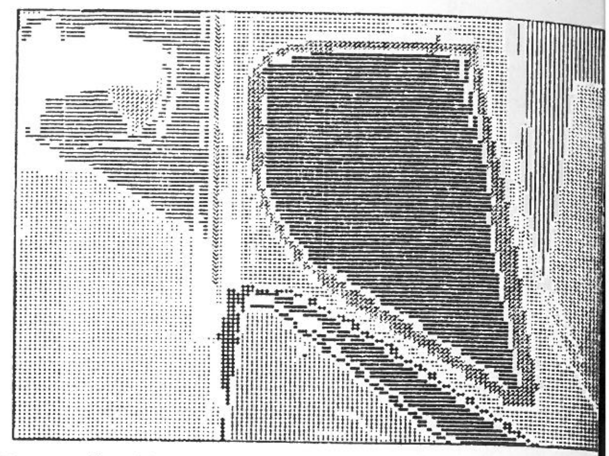

Figure 2b illustrates corresponding regions of the right image.

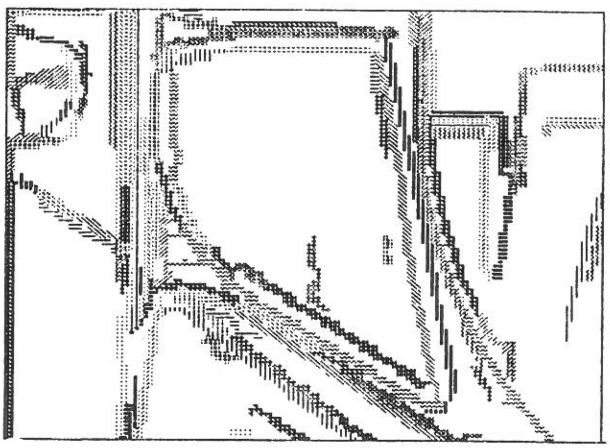

Figure 2c illustrates arcs of the left image.

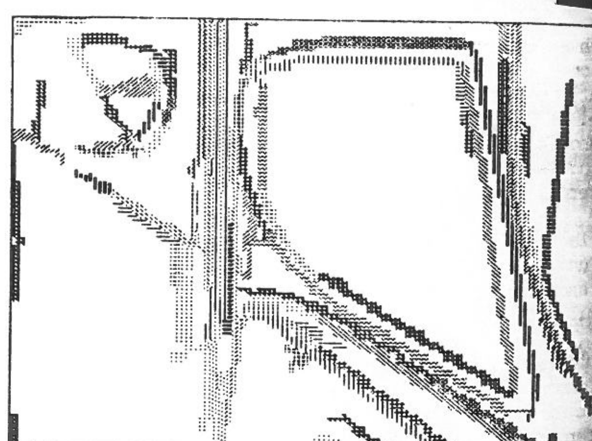

Figure 2d illustrates corresponding arcs of the right image.

References

[1] Shapiro, L.G. and R.M. Haralick, "A Hierarchical Relational Model for Automated Inspection Tasks", Proceedings of the International Conference on Robotics, Atlanta, March 1984, pp. 70-77.

[2] Pong, T.C., "Determining Intrinsic Scene Characteristics from Images", Ph.D. Dissertation, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, August 1984.

[3] Cheng, J.K. and T.S. Huang, "Image Recognition by Matching Relational Structures", Proceedings of PRIP 81, August 1981, pp. 542-547.

[4] Shapiro, L.G. and R.M. Haralick, "Structural Descriptions and Inexact Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No. 5, Sept. 1981, pp. 504-519.

[5] Haralick R.M., L.T. Watson and T. Laffey, "The Topographic Primal Sketch", International Journal of Robotics Research, Vol. 2, No. 1, 1983, pp. 50-72.