III. BETWEEN REGIONS AND OBJECTS — SURFACES AND VOLUMES

Kurt Konolige, Bryant York, Allen Hanson , and Edward Riseman
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

Abstract

Several modular processes are used to build relationships between a) two-dimensional regions and boundaries in a segmented image, and b) the hypothesized surfaces, volumes and objects of a three-dimensional world. A variety of bottom-up and top-down strategies coordinate the processes of SHAPE, PERSPECTIVE, SHADOW, and OCCLUSION to construct these relationships in the VISIONS system.

Introduction

There are three major types of processing which lead to model construction at the region, surface, volume, object and frame levels: (i) curve fitting of region boundaries and analysis of contours in terms of primitive 2D shapes, (ii) hypothesizing surfaces and volumes using the 2D shape descriptions, the results of perspective, shadow, and occlusion analyses, and additional semantic information, and (iii) object verification using a spatial processor which matches 2D projections generated from a 3D description of the object against regions and contours from the image [Marr and Nishihara, 1976].

Representation

Objects in the data base are described in terms of primitive volumes. A generalized cylinder description is maintained in object-centered coordinates for each of the component volumes of an object. We are currently investigating the use of Binford's full generalized cylinder description [Agin, 1972; Marr and Nishihara, 1976]. This description allows cross sections which are orientable in the plane (such as triangles, rectangles, and ellipses). Relationships between the primitive volumes, surfaces, contours, and regions are explicitly stored in the data base of long term knowledge. For example, the fact that a cylinder may project as a circle when viewed on end or as a rectangle when viewed from the side is explicitly stored in the knowledge base. These relationships provide some of the pathways between two and three dimensions.

In addition to an object-centered coordinate system, we need to represent partial three-dimensional information available from image data in a viewer-centered coordinate system [Duda and Hart, 1973]. A representation has been developed which allows a simple encoding of hypotheses about viewed object plane angles [Konolige, 1977]. This representation provides a simple relationship between vanishing points and object plane angles, and a separate encoding of scale and angle information. This latter point is important because scale information is available via knowledge about the size of recognized objects, while angle information is derived from hypotheses about image data.

Curve Fitting and Analysis of Contours

The chain-encoded boundary segments are smoothed and fitted with straight lines, conies, and exponentials. Vertices are analyzed in terms of internal angles and classified according to type [Duda and Hart, 1973]. Syntactic and heuristic methods are used to analyze these vertex-segment descriptions to produce region descriptions in terms of stored 2D primitive regions. During the process, syntactic cues allow regions to be split and/or merged to form contours which are more easily analyzed in terms of the primitive 2D shapes.

Specific information for the later use by knowledge sources (KS's) is extracted at this level. Strong geometric forms such as triangles or squares are labelled for the shape KS. Strong straight lines are also labelled and points of convergence of these lines are found for later hypotheses about vanishing points. Non-adjacent regions with similar features are marked for further analysis by the occlusion KS.

Generation and Verification of Hypotheses

All primitive 2D shapes In the data base are associated with the surfaces and volumes which project as those 2D shapes. Once a region or contour in the image has been classified as a primitive 2D shape, we can then index into the knowledge network to obtain possible surfaces and volumes which may be present in the scene. Perspective, shadow, and occlusion information on the orientation and extent of surfaces (such as the ground plane, the horizon line, etc.) may be used to confirm, refute, or refine hypotheses. Finally, objects or parts of objects may be hypothesized on the basis of high-confidence volume hypotheses. It should also be noted that, in addition to shape analysis, the matching of color/texture attributes of regions and objects is a companion path to object hypothesis and verification.

Overall Strategy

Figure 1 shows the results of applying one of the segmentation processes (region planning) of VISIONS to an image. We are currently investigating a variety of volume-surface construction strategies for applying the various KS's to this processed data and using the results of other KS's. Which strategy is actually applicable at a given point in the model construction process depends on a number of factors, including the

This work was supported in part by ONR grant N00014-75-C-0459 and NSF grant DCR75-16098.

2 A. Hanson is affiliated with Hampshire College, Amherst, Massachusetts 01002.

information available at that point and the reliability of that information. One such strategy, which incorporates both data-directed hypotheses and the use of semantic clues, is given below.

Example Strategy

Often regions representing the ground and sky planes can be inferred from color characteristics and knowledge of where they appear in natural scenes. Contour analysis can find the strong straight lines in the image. These could be used by the perspective KS to fix the horizon of the ground plane, and by the shape KS to recognize the triangle and square. Knowledge of the 2D projection of 3D shapes suggest either a pyramid or a wedge for the image triangle; a wedge is hypothesized if the roof line of the house is found. Then the spatial processor would orient the wedge to give the correct projection. At this point the shape KS could index into its data base of objects to suggest a house roof for the wedge ; then a range for the distance of the house could be estimated from stored house size ranges. Finally, since the ground line of the house is above the hypothesized horizon, an upward ground slope to the house can be postulated (which in this case is a correct hypothesis) and computed using the estimated house distance. The desired results of this set of analyses are illustrated in Figure 2.

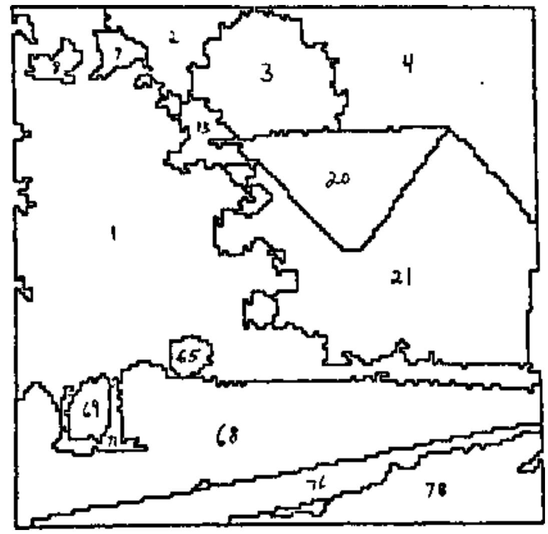

Figure 1 Region plan from original segmentation with labels for regions whose area is greater than 70 pixels.

References

Agin, G.J. [1972] "Representation of Curved Objects," AI-MEMO 173, Stanford AT Lab.

Duda, R.O. and Hart, P.E. [1973] Pattern Classification and Scene Analysis, John Wiley and Sons.

Konolige, K. [1977] "Perspective Analyses in VISIONS," in preparation, Univ. of Mass.

Marr, D. and Nishihara, H.K. [1976] "Representation and Recognition of the Spatial Organization of Three Dimensional Shapes," AI-MEMO 377, M.I.T.

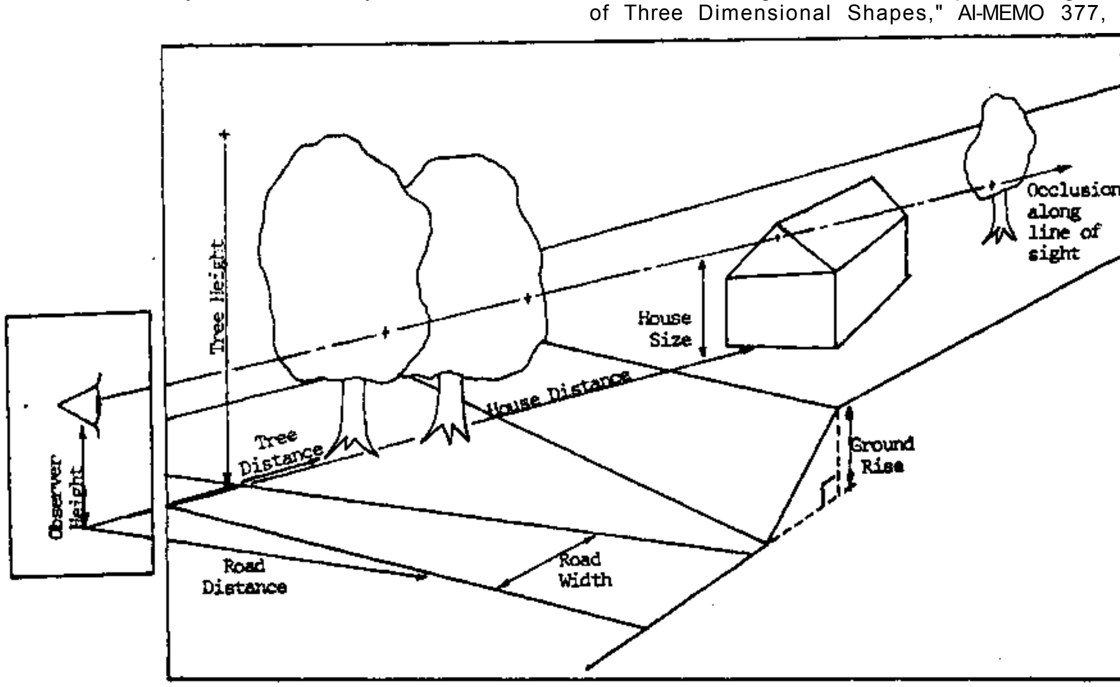

Figure 2 Desired volume-surface plan for the coarse region segmentation in Figure 1. This schematic figure shows house size, road width, ground plane slopes, and other 3D information which might be derived using perspective, shape, and occlusion knowledge sources. The observer's view in this schematic gives the projection shown in Figure 1.

TOWARDS A SCIENCE OF IMAGE UNDERSTANDING

Berthold K. P. Horn
M.I.T. Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

A great deal of work in machine vision has brought us to the point where we can tell that:

a) image understanding is difficult, and
b) image understanding is very useful.

In order to conserve our limited resources, we should pause briefly and consider what new problems to approach and how to tackle them in order to maximize the pay-off. In particular, we might hope to pull together results of efforts in many fragmented areas in order to build the foundations for future work. Perhaps, too, we can see how to structure investigations so that they contribute to basic understanding, rather than just developing special solutions. This may not always be practical under real-world pressures. Nevertheless, I will list four ideas I consider germaine to this topic:

(1) We should pursue vision problems which lead to theories that can be tested by programming a digital computer.

The proof of the pudding is in the eating of it. It is the digital computer that enforces a certain rigour in our discipline and we ought to avoid spending much effort pursuing theories that cannot be tested.

(2) We should look at real images -- in all their natural splendor.

It is unusual not to be surprised by the actual data -- the real world rarely conforms to our expectations. Often the tools for doing this are not available -- an interactive system capable of displaying arbitrary intensity profiles and calculating various simple measures and histograms is the minimum needed. One should be able to browse around in the image in a natural way.

(3) We ought to understand the imaging process -- the way surfaces reflect light .in particular.

It is hard to imagine any serious computer scientist processing information in some interesting fashion when he has no understanding of its origins. We are only now beginning to understand how the physics of image-formation can be used to guide and constrain the process of obtaining a symbolic description of an image.

(4) We need the equivalent of the molecucular biologist's Escherichia coli or the geneticist's Drosophila melanogaster.

Research in machine vision, scene analysis and image understanding has been fragmented because of the enormous domain of potentially interesting problems. If we could agree that one, or a small number, of vision problems or image types can provide enough stimulation for work on several aspects of image understanding we could benefit tremendously from the resultant cross-fertilization amongst experts in different domains.