EXPERIMENTS IN
KNOWLEDGE-DRIVEN INTERPRETATION
OF NATURAL SCENES*

Terry E. Weymouth

Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

Among the sources of information that can be used to guide the processing of visual sensory data are the constraints implied by the geometry of the objects being viewed. Experiments have been performed using representations of this type of knowledge to control image processing. They show how such geometric knowledge can be used to aid in the identification of object projections in images of natural outdoor scenes.

Introduction

The general goal of processes operating on visual data in the VISIONS system is the construction of a symbolic representation of the objects, and of the relations among those objects, from images of a given natural outdoor scene. Recent experiments have been designed to determine the effectiveness of the use of world knowledge in the interpretation task and to explore representations of this knowledge [9]. A portion of the knowledge used in the system consists of descriptions of specific objects (e.g. house, tree, grass, etc.) and the representation of three-dimensional spatial relationships among those objects [3]. This paper discusses one way in which such geometric knowledge can be used to control image processing. The use of geometric knowledge to guide image processing has been used with some success by [1],[2], and [5].

In the experiments described here, a three-dimensional scene model which represents the relative location and size of objects is projected to the two-dimensional Image to form a plan. This plan can be used to restrict image analysis to certain portions of the image and to trigger the application of procedures for the recognition of specific objects during image interpretation. The results will be illustrated using the house scene shown in Figure 1.

A Spatial Plan from a Geometric Scene Model

Knowledge about approximate object location and size was used to investigate the effect of geometric constraints on a region-labeling process. The goal of the experiment was to partition a region segmentation into non-disjoint subsets *of* regions, each containing the projection of an object from the scene. The data used is an initial segmentation of the image into regions of similar color and texture [7]. An image region does not necessarily correspond to a single object, nor is it the case that a given object will unfailingly project to a single region. Scene knowledge can be used to partially overcome these difficulties through the formation of a plan which predicts the location of the image projection of actual scene objects. The subimage produced by a matching process which uses this plan can then be further processed to better extract the image location of the object. One possibility is resegmentation, which can be performed with greater accuracy using the local image context [7]. Specific knowledge of the surface features of the object can also be used to constrain the subimage even further.

If all the objects in a scene were of known size, shape, and at a known location and orientation, and if the scene were viewed with a known camera from a known camera position and orientation, then the scene model would be easily projected to the image using the standard techniques of computer graphics: rigid object transformation, perspective projection, and hidden surface removal [8]. In general, however, the scene model must capture uncertainties as to object location and permit projection methods that are tolerant to errors in expectations of camera position and orientation. Consequently, several less constraining assumptions were made.

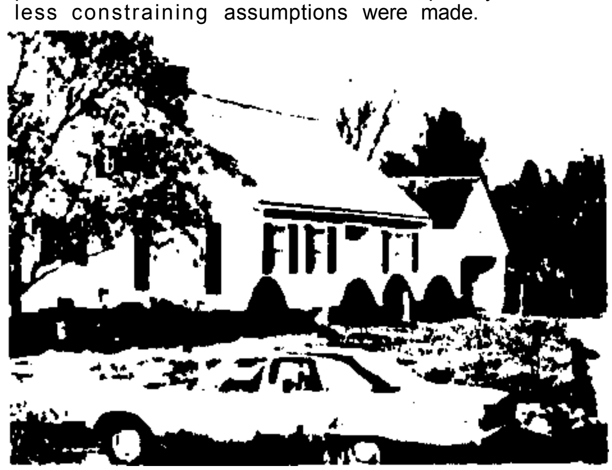

Figure 1. House scene used in experiments.

•This research was supported in part by the National Science Foundation under Grant MCS79-18209.

Scene and Object Model

The three-dimensional geometry of an object in a scene can be characterized by: location, orientation, scale (size), and shape. Of these, shape is obviously the most complex. Easing the constraint on shape by generalization to a sphere permits a rough description of location, location uncertainty, and size. The location of an object, expressed in the coordinate system of the scene model, is approximated by the center of a "location sphere." The radius of this sphere expresses uncertainty as to location of the object center. The object size is also represented by a sphere, centered at the object location. The radius of this "size sphere" is related to the size of the object. Because these spheres are embedded in a scene coordinate system, the addition of a camera model permits the transformation of the scene model (by translation and rotation) to a specific viewpoint and the projection of that model to the image.

Projection of Scene Model to Image

There are two problems in the projection of this three-dimensional model to the image: the projection of the location sphere and the projection of the size sphere. The projection of the location sphere is straightforward. Using a perspective transformation (from a camera model) the center and radius of the sphere are projected to the image as a circle. Figure 2 illustrates the way in which this three-dimensional object model can be projected onto any image (using a camera model) to form an image-specific plan for that object.

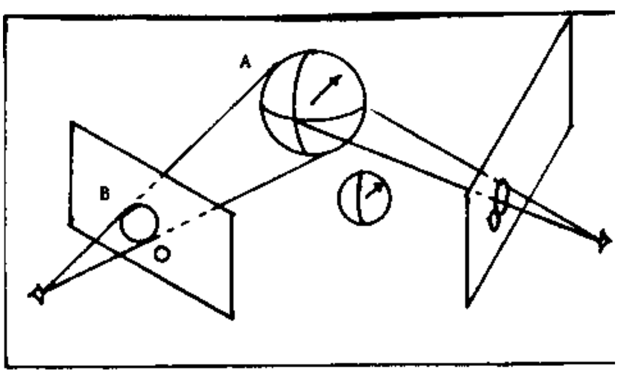

Figure 2. The geometric model of location is a sphere in space (A) which is projected onto the image as a circle (B). Note different viewpoints.

The projection of the size sphere is handled in a slightly different manner. Due to uncertainty of the exact location, Just the radius of the size model is projected to the image. The projected radius is added to the plan as a "size filter;" no region greater than that size is expected as a projection of that object. In the experiments shown here this size radius was added to the location radius to form a composite plan, shown as the outer circle in Figure 3. The other two circles shown are: the projection of the location sphere (the smaller of the two inner circles) and the projection of the size sphere (the next larger circle) arbitrarily positioned at the center of the projection of the location sphere. Note that, although these illustrations show tightly constrained locations (small location spheres), it is possible to represent small objects and/or larger uncertainties in location.

Once the model has been projected to the image, the resulting plan can be used in a matching process. Regions in the image can be matched against the projection circles to form a subimage. All regions that are contained within the composite circle are taken to be candidates for inclusion in the subimage. Further, all regions for which the ratio of region area within the circle to the total region area is above an object-model-specific threshold are accepted as candidates for inclusion. These candidates are then filtered by size, and any region which Is too large is rejected. (Such rejected candidates could be split by resegmentation; this was not done here.) The union of the remaining candidates forms the sub image.

The two spherical representations whose projections are shown in Figure 3 are for "tree-crown" and "house-body." Figure 4 shows the regions selected when those projections are used. Figure 5 shows the way in which geometric matching can be used to limit other recognition procedures. A feature matching process for house-body, based on a color and texture prototype from the object model, was run on an image segmentation to select the set of regions shown in Figure 5a. The same segmentation, together with the plan from the projection of the house-body spherical model, produced the sub image shown in Figure 4c. Using the geometrically based subimage to restrict the feature matching Improves the results. Compare the regions from feature matching alone (Figure 5a) with those selected when the feature matching was constrained by the geometric plan (Figure 5b). The matching using the plan is Improved by the elimination of those regions which *are* inconsistent with size and location information.

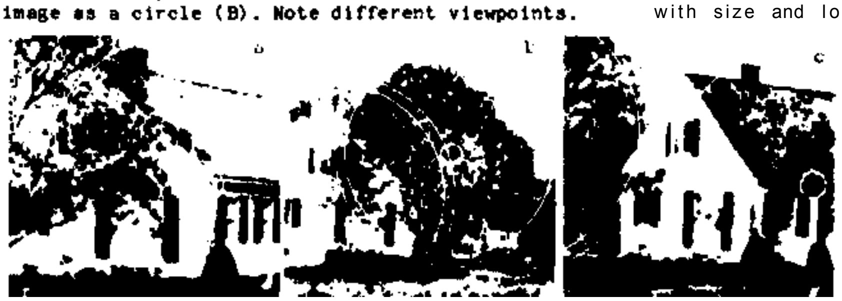

Figure 3. The circles of projection of two different object model spheres in three views of the house scene. For each projection: the inner circle is from the size sphere, the middel circle is from the size radius, and the outer circle is from the sum of the two radii.

Future Work

In addition to controlling region grouping procedures, knowledge of the geometric constraints implied by object subpart relations can be used to guide the extraction of image features [10]. The three-dimensional object geometry, when combined with information about the location of object features in the image, can be used to build a three-dimensional model of the objects in the image [4]. This model, in turn, can be projected to the image to provide more specific information about the location of object projections in the image. Such specific information will eventually be used to control the invocation of both feature extraction procedures and procedures for the recognition of objects from two-dimensional image characteristics [6]. Work is currently underway to examine how such control can best be implemented.

Acknowledgments

Edward M. Riseman, Allen R. Hanson, Clif McCormick, and Leonard P. Wesley helped by critiquing the general concepts and specific implementations discussed in this paper.

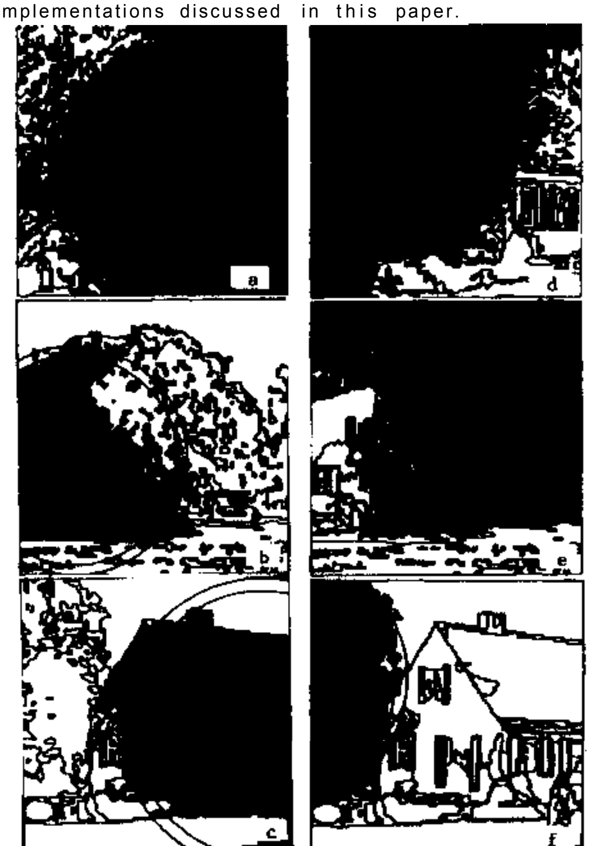

Figure 4. Image regions selected by the circles of projection shown in Figure 3: house-body (4a-4c) and tree-crown (4d-4f).

References

[1] D.H. Ballard, CM. Brown, and J.A. Feldman, "An Approach to Knowledge-Directed Image Analysis" in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[2] Thomas 0. Binford, Rodney A. Brooks, and David G. Lowe, "Image Understanding via Geometric Models," in Proceedings of Fifth ICPR, 1980, pp. 364-9.

[3] A.R. Hanson and E.M. Riseman, "VISIONS: A Computer System for Interpreting Scenes", in Computer Vision Systems, A. Hanson and E. Riseman, Eds., Academic Press, New York, 1978.

[4] D.L. Lawton, "Constraint-based Inference from Image Motion," in Proc. AAAI, Stanford University, August 1980, 31-34.

[5] Alan K. Mackworth, "Vision Research Strategy: Black Magic, Metaphors, Miniworlds and Maps" in Computer Vision Systems (A. Hanson and E. Riseman, eds.). Academic Press, New York, 1978.

[6] C.H. McCormick, "Strategies as Knowledge-Based Control in Image Interpretation", Technical Report in preparation. Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts.

[7] Paul A. Nagin, "Studies in Image Segmentation Algorithms Based on Histogram Clustering and Relaxation", Dissertation published as COINS Technical Report 79-15, September 1979

[8] William M. Newman and Robert F. Sproull, Principles of Interactive Computer Graphics, second ed '., McGraw Hill, 1980.

[9] C.C. Parma, A.R. Hanson, and E.M. Riseman, "Experiments in Schema-Driven Interpretation of a Natural Scene", COINS Technical Report No. 80-10, University of Massachusetts, Amherst, Massachusetts.

[10] T.E. Weymouth, "Experiments in Knowledge-Driven Interpretation of Natural Scenes," Technical Report in preparation, Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts.

Figure 5. Images showing use of the geometric plan (refer to Figure 4c) on a feature matching process: feature matching without geometric constraints (5a), and with geometric constraints (5b).