Efficient segmentation method for range data

Thomas C. Henderson
Computer Science Department, University of Utah
Salt Lake City, Utah 84112

Abstract

A method for finding planar faces in sets of range data images is described. First, the object points are extracted from the range data image (or images). These are then organized in a k-d tree using the x, y and z values as keys. From the k-d tree a spatial proximity graph is efficiently constructed. Finally, a set of seed points for a possible face is chosen, and the spatial proximity graph is used to guide the search for neighboring points lying on this face.

Introduction

Three-dimensional objects are modeled as a set of planar faces and the relations between the faces. Thus, in order to analyze the objects in a scene, the planar faces of those objects must first be determined. Many methods have been proposed to solve this problem for intensity images; however, the availability of a new class of images (called range data images) which directly provides depth information makes it possible to develop more reliable and efficient segmentation methods. One such method is described here.

The segmentation is achieved in three steps: (1) extract the object points from one or more range data images, (2) determine the "neighborhood" of each point, and (3) perform a region-growing step to determine the faces. That is, the original set of points is divided into sets of points which correspond to the planar faces in the scene. Each of these sets may then be approximated by a polygon. Note that a particular point may lie in several faces, e.g., an edge point or a vertex point. Moreover, we do not assume that all the objects in a scene are necessarily composed of planar faces; however, we do feel that a description in terms of planar faces can form the basis for representing curved surfaces.

Extraction of object points

A range data image, $I(x, y)$, is an image whose value at each picture element gives the z-distance to the point projecting onto the image at (x, y). Such images are one type of intrinsic image of a scene.[1] Most segmentation methods are defined directly in terms of the topology defined by the image; however, we have found that a wider class of problems can be handled if the (x, y, z) points of interest are first extracted from the image, and then a neighborhood graph constructed for these points. For example, when modeling an object, one can merge the data from several range data images into one object-centered coordinate system.[2] Even in single image analysis, objects may be easily separable due to depth.

Thus, the result of the object extraction step is merely a set (or possibly several sets) of points of interest. However, any topological information from the image has been lost.

The underlying graph

Given a set of points in 3-space, a description of the nearest neighbors (within a given threshold) of each point is specified by the underlying graph.[3] This graph can be efficiently constructed by building a 3-D tree of the set of points. This tree can be built in Order (n log n) time, where n is the number of points. Basically, a 3-D tree is a binary tree with a multi-dimensional key that splits the data according to maximum spread along an axis at the median. In our case, the key is just the x, y and z values of point.

The underlying graph can be constructed in Order (n log n) time in the following way: for each point in the set, use the 3-D tree to find the nearest neighbors (e.g., 4 or 8); this can be accomplished in Order (log n) time. Simply record the neighbors. This then constitutes the underlying graph: a graph giving the m nearest neighbors in 3-space.

Face formation

The underlying graph can be used in a region-growing scheme to accumulate points into faces. Once a likely starting point has been found, the region around that point is

examined in roughly spiral order. Points that lie in the same plane as the starting point are kept, and cause further exploration in that area. The order of exploration is controlled by a FIFO queue.

The input to the method is the underlying graph. Each point is examined separately. If the point belongs to one or more faces already, then it is not examined. Otherwise, if the point and its neighbors lie on a plane, then the coefficients of that plane are recorded, and both the point and its neighbors are marked. Next, the point's neighbors are put on the queue. While the queue is not empty, the head of the queue is removed, and its neighbors are examined. If the neighbor lies in the plane, then it is marked and put on the queue. Finally, once the queue is empty, if the set of marked points satisfies all necessary conditions, e.g., enough points to ensure a reasonable size, then the face is recorded.

Conclusion

This method is Order (n^2) in the worst case where the number of edge points and the number of faces are of the same order as n; however, in general, only Order (n) time is required. Finally, it may be necessary to do some post-processing to eliminate narrow regions as faces or parts of faces.

References

1. Barrow, H. G. and Tennenbaum, J. M., "Recovering Intrinsic Scene Characteristics from Images," SRI Tech. Rept. 157, April 1978.
2. Henderson, T. C. and Bhanu, B., "Three-Point Seed Method for the Extraction of Planar Faces from Range Data," IEEE International Workshop on the Applications of Industrial Vision, Raleigh, N.C., May 3-5, 1982.
3. Friedman, J. H., Bentley, J. L., and Finkel, R. A., "An Algorithm for Finding Best Matches in Logarithmic Expected Time," ACM Trans. on Math. Soft., Vol. 3, No. 3, pp. 209-226, Sept. 1977.