

Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)

Edited by A. Gruen, E.P. Baltsavias, O. Henricsson Birkhäuser Verlag Basel, Boston, Berlin, 1997

This volume is a collection of 36 papers from the second bi-annual Ascona conference, held in Ascona, Switzerland, in May, 1997. Nearly all major US and European research laboratories which work on applications of computer vision to mapping are represented. As such, this book provides valuable documentation of state-of-the-art results in feature extraction and provides an opportunity for comparison of feature extraction methods.

Space does not permit a detailed paper-by-paper review. Instead, I will attempt to provide an overview of some of the papers within each topic, along with some top-level comparisons and comments.

Building Detection and Reconstruction (17 papers)

The majority of the papers in this volume are concerned with building detection, an extremely hard problem to which a number of approaches have been tried. It is also a problem of great economic importance, since manual delineation is extremely expensive, and there are a number of applications waiting for affordable 3D building data. Although progress has been made, it is evident that no robust general solution for the problem has yet been implemented. The various approaches can be characterized along several axes, including the degree of automation, the features used as algorithm inputs, the number of images used, and the types of knowledge incorporated.

Semi-automated systems must keep the amount of operator input low, while using enough input and the right kind of information to guide the automated processes. The ARUBA system, described by Henricsson and Baltsavias, has the operator delineate the area containing a building on all images; once the search space is thereby reduced, the system uses color and geometric information to form a 3D model. In the TOBAGO system, (Gruen and Dan) the operator measures all the roof points in stereo and the program automatically determines the topology of the point connections to generate the building model. In a system by Gülch, et al., numerous processing options are available;

typically, the operator defines a CSG (constructive solid geometry) building model, then measures the homologous points with optional automated matching and refinement of measurements. Wiman describes a system in which the operator delineates the building roof in one image, then the system matches the delineated edges in the other images using a DSM (digital surface model) for guidance.

To obtain a 3D building model, some systems start with 2D features (image edges and regions) and match them to obtain a 3D model, while others use 3D data from image matching or laser or radar sensors. For instance, Weidner segments the objects above the ground, then fits parametric or prismatic models. Roof segmentation is based on extraction of surface normals and the detection of roof edges. Instead of a DSM

derived from images, Hug describes the use of elevation and reflectance data from an imaging laser altimeter to segment buildings.

2D and 3D data can be profitably combined. For instance, Shi, et al., describe the use of a combination of area- and feature-based matching to derive a DSM. Once the ground plane is determined, regions above the ground plane are verified to be building roofs based on their elevations and on edge-based region segmentation. L  chervy detects roof crest lines in a DSM, then verifies the hypothesized edges in the image. Dissard, et al., analyze objects above the ground plane for evidence of regular geometric structure. Haala and Brenner use a DSM from image matching or from a laser scanner along with 2D building outlines to predict and verify roof structure.

The majority of the papers in this section use standard mapping stereo pairs, sometimes with additional images from extra sidelap. However, Nevatia, et al., describe a system to detect flat-roof rectilinear buildings in single vertical or oblique images. Building hypotheses are formed by looking for parallelogram structures, then verified by examining wall and shadow evidence. Conflicting hypotheses are reconciled by examining the containment and occlusion relationships between them. Muller, et al., describe the effects of image resolution on several monoscopic building detectors.

The concept of "building" is a broad one, ranging from the irregular "shacks" extracted from informal settlements by Mason and Baltsavias to complex residential or commercial buildings. The more precisely the building model is specified, the greater the search space of the automated process can be reduced. Steinhage uses image aspect models to guide the aggregation of features into building components and then into buildings. In TOBAGO (Gruen and Dan), buildings are assumed to have planar polygonal roofs and specific building models are classified by the number of ridge points. In the semi-automated system described by G  lch, et al., the operator selects a specific CSG building model.

One factor complicating the comparison of the various methods presented here is the general lack of discussion of evaluation criteria. For instance, what constitutes "detecting" a building--- when the building hypothesis covers 50, 75, or 90 of the building? How accurately are buildings delineated? One hopes that this important topic is covered more fully in other publications concerning these methods.

Most of the papers use the Avenches data set, which provides a useful basis for comparison of results. The use of such common, well-documented data sets should become more widespread in the research community.

Road Extraction (6 papers)

Nearly all road extraction systems follow a similar flow: Road seeds, or starting points, are detected automatically or provided by an operator. The system starts from these seeds and tracks along the roads as much as possible. Finally, the tracked roads, which may be broken or incomplete, are combined into a complete road network. The papers given here embody different approaches to these basic steps.

Steger, et al. describe the network completion part of their system, which uses grouping based on a graph representation. Fuzzy logic is used to select the most

compatible paths between important vertices, thereby allowing a measure of top-down control. Trinder, et al., apply learning techniques to the road extraction problem. Vosselman and de Gunst use object models for different road classes and types of intersections, thereby supplying additional knowledge and context. Barzohar, et al., apply a Kalman filter to road tracking, while Gruen and Li use 3D LSB snakes for tracking. Baumgartner, et al., define contexts which describe typical scenes. This contextual information then helps control the network completion task by providing possible explanations for image structures.

Map/GIS-Based Methods (4 papers)

Given that most areas of the earth have been mapped to some level of detail, it is logical to utilize existing map or GIS information to help guide the feature extraction process. Absolute image orientation is a necessary first step to image exploitation; Dowman and Ruskoné discuss the matching of polygonal features between satellite images and map data to obtain control information for an orientation solution.

Roux and Maître analyze the structure of urban scenes using images and scanned maps. After the road network is extracted from the map, city blocks are delineated and classified in terms of building types. A disparity image is generated, guided by the road network, then 3D building hypotheses are generated by analyzing the disparity image and correlating it with the map.

Quint has implemented the MOSES system, which generates a semantic model from a digital map, then uses the semantic model to guide interpretation of the image. Objects represented in the map, such as buildings, roads, and parking lots, are found first in the image; this provides a context to guide the recognition of other objects.

Schilling and Vögtle present a fundamentally different approach from most; instead of extracting individual buildings or roads, they use a topographic database to provide training information for extracting settled areas from satellite images.

Visualization (3 papers)

3D visualization is now a main driving force for 3D feature extraction. There are many potential users for 3D databases, the problem being the cost of generating detailed and accurate scene descriptions.

One of the most promising applications is urban planning and design. Danahy gives an overview of the requirements for urban planning, which include the detailed representation of terrain, vegetation, and buildings.

Schickler describes a VRML model of the Atlanta airport, built from manually generated models at four levels of detail. While VRML supports a number of features and VRML viewers are readily

available, the rendering speed for large databases can be problematic. Gruber, Kofler, and Leberl present an architecture for efficiently handling and rendering large 3D urban databases, built around an object oriented database management system and a hierarchical data structure which allows adjustment of rendering detail based on

distance from the viewer. Image data used for texture information is also stored in a hierarchical structure.

Conclusion

The book is well-edited, with only a few minor typographical errors. All of the papers are presented in English. Unfortunately, in some cases the reduction of the papers to fit the book's format has resulted in figures which are too small to see meaningful details or overlaid graphics.

In my opinion, this book contains a well-coordinated group of papers which provide a useful overview of the current state-of-the-art in semi-automated and automated feature extraction. I look forward to the results of the next Ascona conference, to see how many of these problems have been solved!

Reviewed by:

Chris McGlone
Digital Mapping Laboratory,
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA