Automatic and Interactive Modeling of Buildings in Urban Environments from Aerial Images

R. Nevatia, K. Price
Institute of Robotics and Intelligent Systems
PHE 204, MC-0273
University of Southern California
Los Angeles, CA 90089-0273, USA
{nevatia, price}@usc.edu
http://iris.usc.edu

Abstract

Automatically extracting object models from images is a complex task. We describe research in extracting 3-D models of buildings from aerial images. This work has resulted in several related systems including assisted extraction (minimal manual interaction to guide automatic processing), automatic extraction with limited imagery and limited building models, and automatic extraction with very good imagery and digital elevation models and more complex building models. Some results are provided for the assisted and one of the automatic systems.

1. Introduction

Good three-dimensional models of buildings are important for a variety of tasks including simulation, urban planning, disaster recovery, etc. The task of automatic extraction of object models from images is an extremely complex one. There are several fundamental sources of difficulty. The objects need to be segmented from the background but are not distinguished by a simple characteristic such as having different colors; instead, the lower level features used for segmentation (such as lines) provide highly fragmented boundaries and the lines and edges belonging to the object boundaries are not readily separated from those coming from shadows, texture and noise. Images are 2-D projections of a 3-D scene and hence inherently ambiguous. When two or more images are available from calibrated cameras, 3-D inferences can be made by triangulation but this requires finding feature correspondences in multiple views which is a difficult process. Even when 3-D data is available, either from a stereo DEM extraction process, or by direct measurement by an active sensor, many problems of segmentation and scene description remain. For the geospatial database to be useful, it needs to contain compact, abstract representations of objects and not just a surface description providing height of each point in the scene.

To overcome these difficulties, we have adopted an approach that consists of the following:

- We build high level descriptions of objects by a hierarchical perceptual grouping of lower level features, starting with lines and junctions computed from images.
- To the extent possible, we reason in 3-D object space since features extracted from 2-D images are subject to variations due to viewpoint, illumination, occlusion and other imaging conditions. 3-D is inferred by matching features in multiple views which requires accurate camera models are available from photogrammetric controls. This approach allows us to fuse the results of analysis from multiple images and data from a variety of sensor modalities (such as panchromatic, multispectral and range images) and yields topologically consistent and geo-referenced object models.
- Our approach uses a *hypothesize and verify* paradigm. Simpler properties are used to hypothesize prominent structures, more complex properties are used to select among, and verify the resulting hypotheses. For example, we may first construct hypotheses for the roof of a building and then use walls (and shadows) for verification.
- Our approach uses rigorous Bayesian techniques for making decisions from uncertain evidence in the critical stages. These techniques also allow us to learn the important parameters automatically from examples rather than setting the parameters manually.

2. User-Assisted Modeling Systems

Even though the ultimate goal is a totally automatic system, we have developed an assisted approach where an automatic system uses limited operator input for guidance in extracting buildings. Additionally, the assisted system can be used as an intelligent editing system for initial, simpler models. Our approach is to provide hints to an automatic system which then completes the remainder of the modeling task, thus leading to a highly efficient method for modeling complex structures.

The conventional approach to user-assisted modeling is to provide a set of generic models that are then fit to the image data by changing model and viewing parameters [8]. In this approach, the system provides geometric computations but substantial time and effort are required from the user. Newer approaches have attempted to combine user input with varying amounts of automatic processing. In [6], we have suggested providing just an approximate building location to extract a building but the quality of the final result is completely dependent on the automatic analysis. In [4], other interactive tools are described including methods for replicating model buildings that are identical or very similar to others. In [2], an automatic system constructs topological relations among 3D roof points collected by a user for each roof; this system can work with several types of complex roofs. In [3], the system handles complex building structures by using constructive solid geometry. This system also uses an image correlation method to fit a primitive to the image; however, this method is computationally expensive when modeling urban sites, where many building have complex shapes.

In our approach, basic modeling tasks are performed by an underlying automatic system but this system receives critical assists from the user. The underlying system is a multi-view building detecting system that uses much of the initial processing of the one described in the next section[7]. The user interaction consists of selecting points near a feature (junction or line) while viewing a single image of the scene, and does not require fusion of stereo images by the user. The underlying automatic processing, however, uses at least two images to match image features and hypothesize 3-D models. Multiple hypotheses are possible at each step and the system selects among them by analyzing the matching evidence in all the available views. As illustrated in Figure 1, the user selects a point near a vertex. Often this single point is sufficient to generate a satisfactory 3-D model of a rectangular building. Up to a total of three points may be required for this initial seed building. The seed buildings may also come from an automatic process or any other source.

A second aspect of the general methodology to reduce interaction is the ability to subtract or remove components. An increasingly complex boundary is created in a precise and consistent manner. A complex model starts with a *seed*. Next, the user adds or subtracts *blocks* by selecting roof locations on one of the images. By enforcing parallel relationships, the user is freed from the need for accurate point selection. With this property, only two clicks are needed to add or subtract a rectangular block. Additions and subtractions can be carried out in any order and without specific ordering of the sequence of interactions. The system automatically determines whether a block is being added or subtracted from the current configuration of the model and from the position of the user interactions with respect to the model. To construct models having non-rectangular roofs, the system provides additional triangular block operations that use located appropriate matching edges in the images to reduce user interactions.

Figure 1 illustrates the interactions needed to add a rectangular block **B** to block **A**. One point must be selected on the existing outline of block **A** and the other diagonally across block **B**. As a result a new polygon **C** is computed. These operations can be applied repeatedly until the resulting polygon correctly models the building roof. .

For multi-layered buildings, the height is computed relative to that of the next lower level. Each level is processed independently in the same manner. No additional steps or clicks are needed to specify multiple layers; these are determined automatically. The example in Figure 2 shows the results on some complex, multi-level buildings in Washington, D.C. By basing our assisted system on an automatic approach, improvements in the underlying systems will lead to improved performance of the assisted procedure. This system has been tested by several different user and a large variety of buildings.

3. Automatic Building Modeling

A variety of approaches have been used for automatic extraction of building models from aerial images[1]. These approaches include extraction from high resolution 3-D data, such as LIDAR, using many views, using video, etc.

Our earlier work [7] assumed rectangular buildings with flat roofs. We will describe in more detail our recent approach for describing buildings having multi-layered planar, polygonal rooftops[5]. Figure 3 shows a flowchart for this method. First, 2-D features, such as lines, junctions and parallel relationships, are extracted from the multiple (5-6) images. Next, we derive 3-D features from groups of 2-D features matched over multiple views. A rough DEM, such as that automatically extracted from these images by stereo processing, is used to aid in feature matching. Grouped 2-D features may not be present in all images, but must appear in at least 2 images. Matching across all views is performed first, followed by grouping of matched pairs.

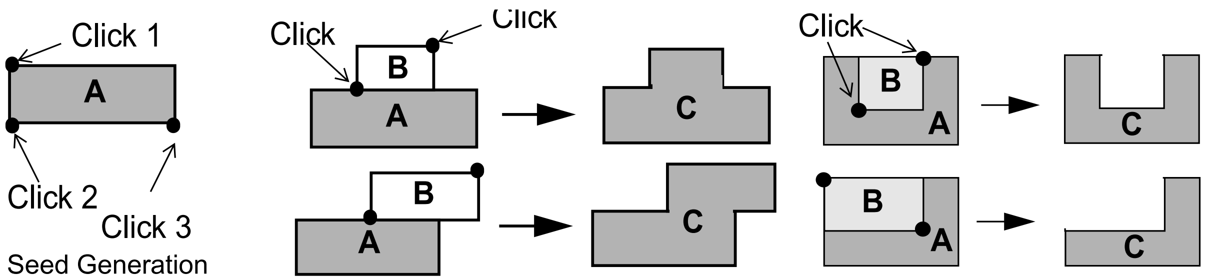

Figure 1 Initial Seed Generation and Protrusions and indentations added by 2 clicks

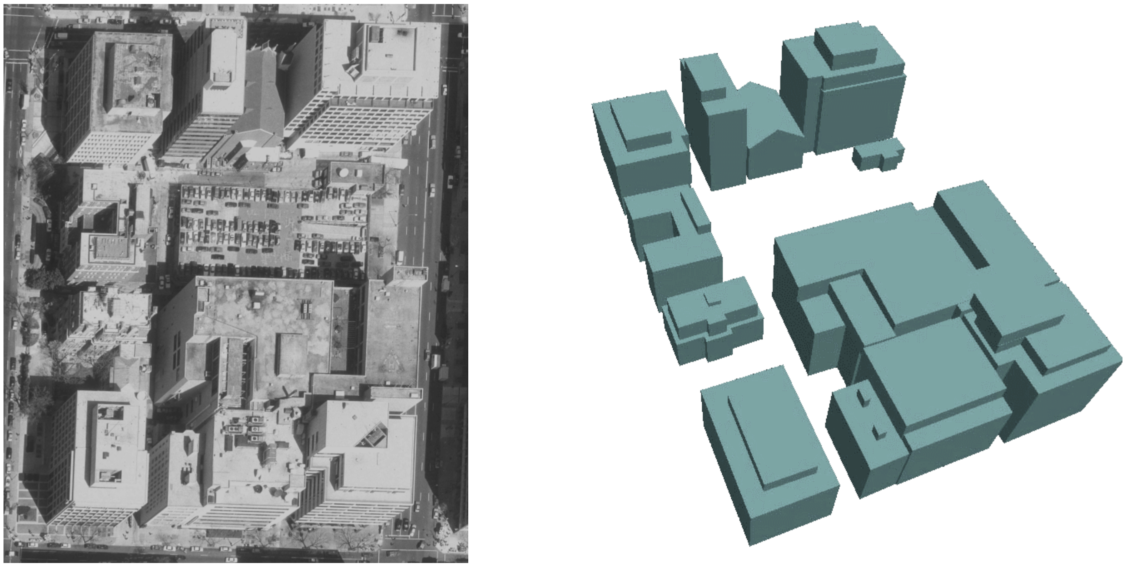

Figure 2 Building models extracted from portion of D.C. using assisted system.

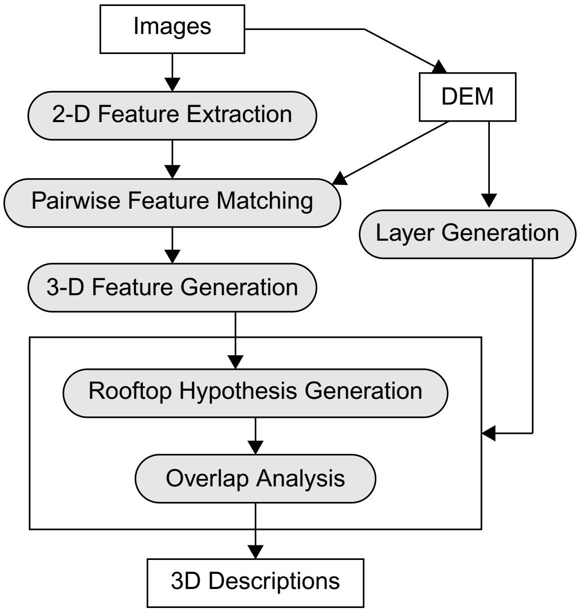

Figure 3 Flow diagram for Automatic Building Modeling

The next step is rooftop hypothesis generation. Flat polygonal rooftop hypotheses are generated by neighborhood searches on 3-D features. To reduce the search space and generate multi-layered rooftops, rough cues for each building component are generated by segmenting the DEM image into connected, flat regions in space called *layers*. For each layer, a search originating from 3-D junctions located near the boundary of the layer and linking "neighborhood features," consisting of other junctions and parallel pairs which are also consistent with the layer, is conducted. Once a rooftop hypothesis is obtained from a node, supporting evidence is collected for the hypothesis. The evidence consists of basic topological properties, line support, and DEM layer cue coverage. Nodes with good evidence are collected as verified hypotheses. Finally, overlap analysis is performed on the generated hypotheses to give final building descriptions. The verification and overlap analysis processes uses Bayesian Networks to make decisions from uncertain evidence.

The use of multiple (more than 3) images has several advantages for 3-D object description over the use of only two stereo images. The multiple images help compensate for missing evidence and accidental alignments (caused by image lines falling along the *epipola*r lines of a stereo pair. However, combining evidence from multiple images is non-trivial as many possible hypotheses can be formed. We use a statistical test to determine compatible matches and to compute height from multiple images. Figure 4 shows 2 results using six views of the same scene.

Figure 4 Automatic extraction results

4. Conclusions

We have briefly presented results of our approaches for assisted and automated extraction of building models from aerial images. Our assisted approach requires limited hints from the user to guide the automatic processing to generate 3-D descriptions. The automatic approaches use rigorous Bayesian learning to select from many potential models in a hypothesize and test paradigm, and uses a perceptual grouping approach with grouping and reasoning in 3-D. These systems combine the work of many students and researchers over a number of years with primary recent contributions from Sung Chun Lee and ZuWhan Kim..

5. References

[1] E. P. Baltsavias, A. Gruen, L. Van Gool, "Automatic Extraction of Man-Made Objects from Aerial and Space Image, III," A.A. Balkema Publishers, 2001.

[2] A. Gruen and H. Dan, "A Topology Builder for Automated Building Model Generation," in *Automatic Extraction of Man Made Objects from Aerial and Space Images (II),* pp. 149-160, 1997.

[3] E. Gülch, H. Müller, and T. Läbe, "Integration of Automatic Processes Into Semi-Automatic Building Extraction," *Proceedings of ISPRS Conference Automatic Extraction Of GIS Objects From Digital Imagery*, September 1999.

[4] Y. Hsieh, "SiteCity: A Semi-Automated Site Modeling System," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 18-20 1996, pp. 499-506.

[5] Z. Kim, A. Huertas, R. Nevatia, "A Model-Based Approach for Multi-View Complex Building Description," in [1], pp. 181-193.

[6] J. Li, R. Nevatia, S. Noronha, "User Assisted Modeling of Buildings from Aerial Images," Proc. IEEE CVPR, June 1999, pp. II:274-279.

[7] S. Noronha, R. Nevatia, "Detection and Modeling of Buildings from Multiple Aerial Images," IEEE Trans. PAMI, Vol. 23, No. 5, May 2001, pp. 501-518.

[8] T. Strat, L. Quam, J. Mundy, R. Welty, W. Bremner, M. Horwedel, D. Hackett, and A. Hoogs, "The RADIUS Common Development Environment," *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, 1992, pp. 215-226.