The method that deals with rooftops that are inclined with respect to the ground plane in 3D (but are still rectangular) requires exactly 3 mouse clicks. The user clicks near three of the corners of the building being modeled. The method then proceeds in a manner similar to the three junction case described above. Again, two choices are possible for the roof orientation and we select the one that is the least sloping.

In the example shown in Figure 7 below the model was constructed using this tool. While the main purpose of the tool is to augment and correct the model of the site that is generated by the automatic system, we demonstrate its effectiveness in the following example by constructing the model from scratch. To construct the model, the user needed an average of 1.5 clicks near a building junction. The computed heights were very accurate, and required no correction.

Figure 7 Results of the smart interactive system on Fort Benning

Bibliography

[1] A. Gruen and M. Baltsavias, Editors, “Automatic Extraction of Man-Made Objects from Aerial and Space Images”, Birkhauser Verlag, 1997.

[2] S. Heuel, R. Nevatia, “Including Interaction in an Automated Modeling System”, Symposium on Computer Vision, Miami Beach, FL, November 1995(383-388)

[3] M. Ito and A. Ishii, “Three-view stereo analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8:524-532, 1986.

[4] C. Lin, and R. Nevatia, “3D Descriptions of Buildings from an Oblique View Aerial Image”, IEEE International Symposium of Computer Vision, 377-382, 1995.

[5] S. Noronha and R. Nevatia, “Detection and Description of Buildings from Multiple Aerial Images”, IEEE Computer Vision and Pattern Recognition, pp. 588-594, San Juan, PR, June, 1997.

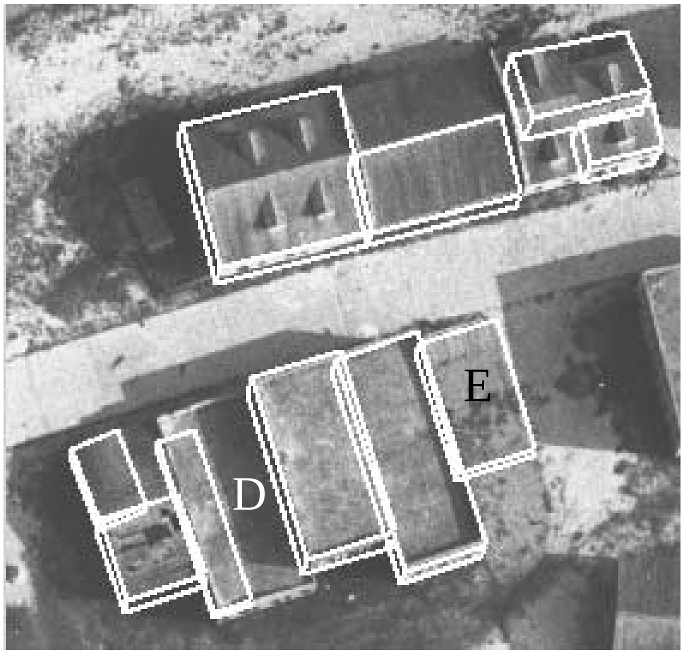

Figure 6 Results on an area of Fort Benning

ther hints.

- If no hypothesis is formed or the user is not satisfied with the generated hypothesis from the first interaction described above, the user needs to specify another junction by clicking near another corner of the desired building. The process of constructing a 2D hypothesis and extending to 3D, is repeated, with the additional constraint of using the line defined by the two junctions near the user specified locations. This interaction usually suffices to complete most buildings that are not modeled after the first interaction.
- If no hypothesis has been generated at this point, or if the generated hypothesis is not satisfactory, the user can choose to click again on a third corner of the desired building. Three points are sufficient to form a roof hypothesis (a parallelogram in the image); the plane of the roof can be determined from just a single view by employing the constraint that the roof hypothesis must form a rectangle in 3-D. Two choices are possible, however. The system chooses the one which forms the smaller angle with the ground plane.
- The system determines the height of the building by searching for a value that maximizes the score for the building hypothesis. Should the height of the hypothesis be unacceptable to the user, it can be corrected by a single click on any “ground” point. The system figures out which “ground” point is being pointed to and reconstructs the 3D hypothesis with this new input. In our experience, this correction has rarely been necessary.

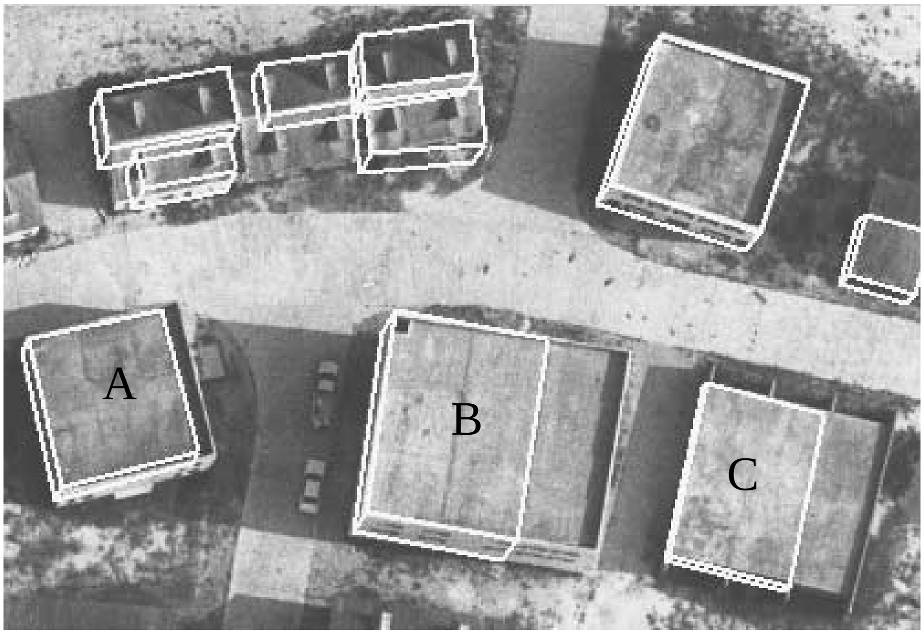

Figure 5 Results on an area of Fort Benning

methodology that can use partial results of the automated analysis. Our method follows the one described in [2], but is made easier due to availability of multiple images.

There are two distinct tools available to the user to interactively construct or refine a model. The first tool requires only an approximate location for a building; it proceeds as follows:

- the user clicks once within the area contained by each building the user would like to model.
- the system searches for and returns the best hypothesis that contains this point. This hypothesis is taken from the set of selected hypotheses from the automated analysis.

The second tool requires a user to specify approximate junctions. It is divided into two sub-methods: one that deals specifically with roofs that are parallel to the ground plane; the other deals with roofs that slope with respect to the ground. The method that deals with roofs parallel to the ground plane operates as follows:

- The user specifies a junction by clicking near one corner of a building. This causes the system to look for junctions in the neighborhood. If one or more junctions is found, the system attempts to construct hypotheses from all the junctions found, by looking for U-completions and lines parallel to the lines forming the junctions. The system chooses the best hypothesis, and then determines the 3D height of the hypothesis by matching the 2D hypothesis with the line evidence in all the other available images. In many cases, this is sufficient to construct a complete 3D hypothesis. If a hypothesis is formed it is displayed that the user may accept or choose to provide furconveniently decomposed into rectangular fragments.

Figure 4 Results on a complex area of Fort Hood, Texas

Figure 5 and Figure 6 show results obtained on the Fort Benning site. In total, 20 buildings (or partial buildings) were correctly detected, with no buildings missed, and 1 false alarm. Buildings labeled A, B, and C in Figure 5 are examples of partial hypotheses. The reason these hypotheses were not completed is that the bounding roof walls are raised above the rest of the roof, causing shadows to be cast on the roof itself. This is a special case of self-occlusion that the system has no knowledge of, currently. A similar problem is encountered in the multi-level complex building in Figure 6 in the area labeled D. However, the system correctly detects and describes other fragments of the same complex building. This building is particularly hard because it has many levels, coupled with gabled roofs. The false alarm is in Figure 6 and is labeled E. This false alarm is very deceptive, and looks like a sloping part of a building even to a human observer. The only cue that it is at ground level is the vegetation that occurs in a part of it.

Fort Benning is a fairly difficult site, with a variety of buildings. The gabled roofs have sub-structures on them. The pathways between the buildings have the same texture as the many of the rooftops. In addition there are confusing details like vegetation and vehicles. Considering the difficulties, the automatic system does a very good job.

4 Interactive Corrections:

While the performance of our automatic system, we believe, is an advance over previously available techniques, the building models need to be refined further to meet the needs of most application tasks. For this purpose, we have developed an interactive

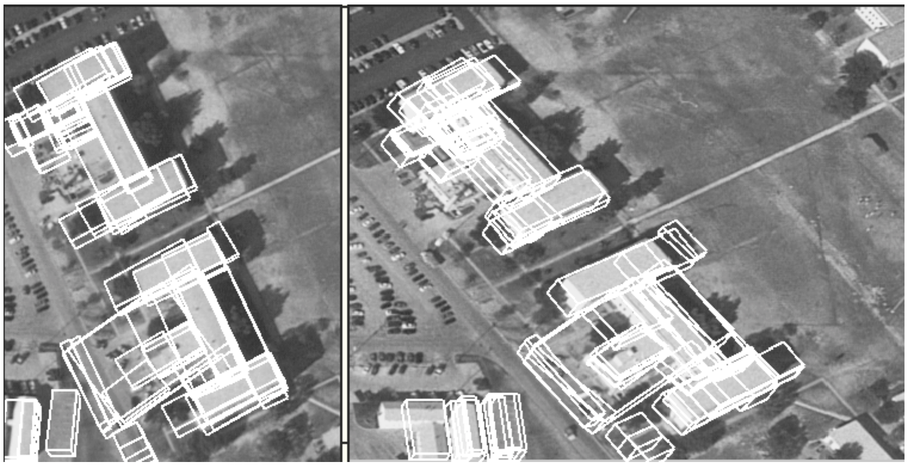

Figure 2 Selected hypotheses from Figure 1

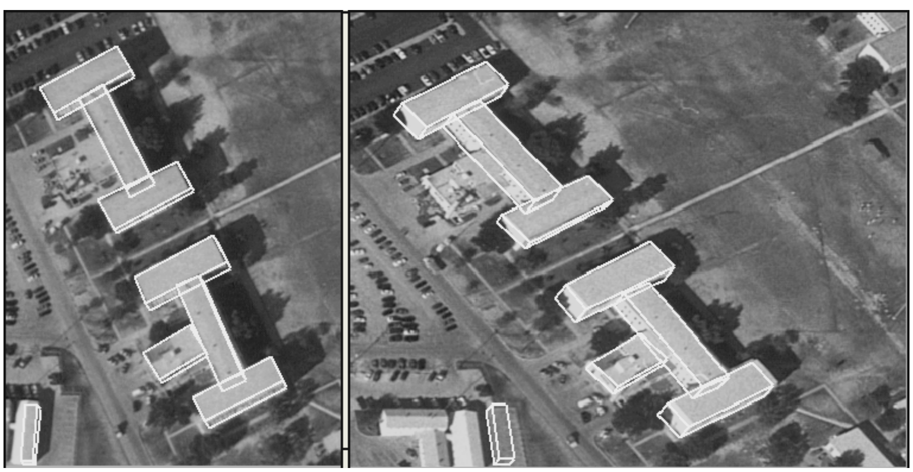

Figure 3 Results are verification and combination

were correctly detected, 3 buildings were not detected, and 4 rectangular building fragments were missed. The buildings in the image are not very high (typically less than 10m tall), and there are numerous instances buildings occluding themselves. This phenomenon caused the 2 rectilinear fragments labeled A and B, in Figure 4 to be missed completely. The building labeled C in Figure 4 was missed because it was too low to verify based on the height evidence alone, and is partially occluded by surrounding foliage. The building labeled D in Figure 4 was missed because of its complex shape that cannot be available, the usual trinocular constraint is applied.

Next, we search for parallels and their matches. Parallels are formed between pairs of lines in the same view that are separated by the less than the maximum width of a building. For two parallels to match, each line forming a parallel in one view should match with a line forming the matching parallel in the other view. While the task domain causes a large number of parallels in each image because of the alignment of buildings, roads, parking lots and shadows, the number of parallel matches is typically lower than the number of lines in any image.

Then, U's are formed by an alignment of two junctions, or by a parallel that has closure evidence near one of its ends. In the case of U match being formed from a parallel match coupled with evidence of closure, no other constraint are applied. In the case that a U match is formed from two aligned junctions, these junction matches should be approximately coplanar.

Finally, parallelograms are formed as a basis for roof hypotheses. To hypothesize a 3D rectangle the minimal requirement is a U match (a match may be in 2 or more images), with additional evidence arising from an "opposing" U match, an "opposing" U in a single image or a parallel match. The components of a parallelogram match (line matches and junction matches) must be coplanar in 3D.

2.2 Selection of Building Hypotheses

The parallelogram matches serve as roof hypotheses, and are equivalent to having a 3D model of the buildings. They satisfy the constraints of being rectangular in 3D and nearly coplanar. In addition, the height and orientation with respect to the ground is known. However, additional processing is still necessary to distinguish building parts from rectangular areas on the ground. This selection is done by using the actual line and junction evidence that is present in forming the roof hypotheses. Some kinds of lines and junctions can actually reduce the confidence in the hypotheses. The selected hypotheses for an example are shown in Figure 2.

2.3 Verification and Combination of Building Hypotheses

It may be noted that so far the evidence that was used was concerning the roof only. The presence of lighting causes shadows to be cast. When the view is oblique, some vertical sides of the walls of the building may be visible. These cues are used to verify the selected hypothesis.,The combined score from the wall evidence and shadow evidence is thresholded to obtain rectangular building (or building component, in the case of non-rectangular rectilinear buildings) hypotheses.

Rectilinear buildings can be decomposed into rectangular components. Verified rectangular hypotheses are examined for combination according to two criteria of proximity, and overlap. Results after this step of processing are shown in Figure 3.

3 Some Results of Automated Processing

We show results on two windows from images taken over Ft. Hood, Texas, and Fort Benning, Georgia. These sites have building shapes that are largely restricted to being rectilinear but images are otherwise very complex. These datasets are commonly used for evaluation of building detection systems.

Figure 4 shows the results obtained on a complex area of Fort Hood. 17 buildings of feature levels (consisting of lines, parallels, “U” shapes and parallelograms). Matching at one level is used to form group hypotheses at the next level. We maintain multiple matches at each level and resolve them only when sufficient information becomes available at the higher levels. We briefly describe our system and show some recent results; more details may be found in [5]. We also describe an interactive editing module of our system that uses partial results from automatic analysis to help correct the errors in a highly efficient way.

2 Description of the Automatic System

Our system can be considered to have three major stages: first, features are extracted and grouped to form roof *hypotheses*, stronger candidates are then *selected* from this set, and finally further reduction is made by *verifying* hypotheses by examining other evidence such as presence of walls and shadow. These three steps are described briefly below.

2.1 Grouping Features to form Roof Hypotheses

Initially, lines are detected using the Canny edge-detector and grouped on the basis of colinearity and proximity. These lines are matched across all views; multiple matches are retained. The constraint used in matching is that the matching lines fall in a quadrilateral defined by the epipolar lines from ends of one line and the allowed disparity range. Figure 1 shows an example with images and matched lines.

Figure 1 Images with matched lines overlaid

Next, binary junctions (formed by the intersection of exactly two lines) are formed and matched. The matching junctions should be on a corresponding epipolar line, the lines forming the junctions should match across the views and be orthogonal in 3-D and the computed height should be within the allowed range. In addition, if a third view is

Recent Advances in Detection and Description of Buildings from Multiple Aerial Images

Sanjay Noronha and Ram Nevatia*
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273
Keywords: Stereo, Building Detection, Multiple Images

Abstract

A brief description of a method for detection and description of rectangular buildings from two or more registered aerial intensity images is provided. A new interactive editing module that can use partial results of the automated process to efficiently correct errors and derive complete descriptions is also described.The automated system operates by grouping features hierarchically to form roof hypotheses which are then verified by using wall and shadow evidence. Grouping and matching steps are interleaved and multiple descriptions are preserved when clear choices are not available. Some recent results are given.

1 Introduction

3-D models of buildings in urban environments are important for a variety of applications and several systems to extract them have been developed (*e.g.* [1]). This is a highly challenging task. The desired object boundaries are typically highly fragmented due to low contrast, occlusion caused by nearby vegetation and by smaller structures on the roofs, and need to be grouped to yield the desired objects. In our work, we limit the buildings shapes to be rectilinear (i.e. rectangular or compositions of rectangular shapes) to aid the task of organization. However, many other structures such as roads, sidewalks and parking lots can also give rise to rectilinear organizations and need to be distinguished from the building structures. The second major problem is to infer 3-D shapes of the objects, while this is possible with single images (*e.g*. [4]), the task is made easier by using multiple images. We now need to make correspondences between the features in different images. This task too is difficult in the aerial image domain. Area correlation methods are likely to have difficulty as the viewpoints can be widely separated, the images may be taken at different times and the building roofs have limited texture.

Our approach is to first form hypotheses for building roofs and to verify the hypotheses by using evidence from cast shadows and visible walls (if any). We use a hierarchy

* This research was supported, in part, by the Defense Advanced Research Projects Agency of the United States Department of Defense under grant No. DACA76-97K-0001, monitored by the Topographic Engineering Center and in part by a subgrant from Purdue University under Army Research Office grant No. DAAH04-96-10444.