detected. After specifying the problem and a roughly location of the building, the missing part was found and fitted without any manual corrections.

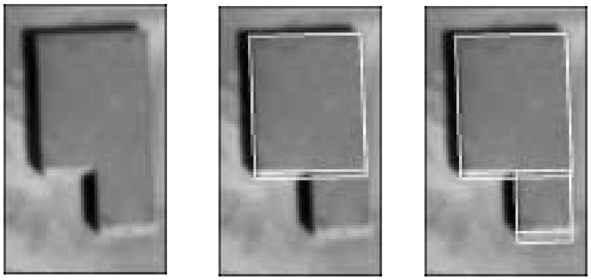

(a) (b) (c)

Figure 7 A partly detected L-shaped building (a), (b) and the manually corrected building.

4.2 Evaluation

Our approach fulfills the requirements proposed earlier: by the initial step, translation and rotation is usually defined by two “qualitative” interactions. In manual or computer assisted manual systems the position is given by more or less accurate measurements in the image. The initial step also gives a first guess of the shape of the building, which might be already the correct hypothesis. In our examples, a correct hypothesis was always found, when it was generated but rejected by USC-MABS.

In nearly all cases only corrections of the sides and height are necessary, because rotation and position is already given by the initial step. Also the number of correction steps in many cases was at most two (see Table 1). A correction step of the height can be saved because of the fitting after each step.

Also the precision of the user’s interaction is decreased: The corrective part uses a fitting process, so that high precision is not needed. Furthermore, by adding already extracted features to the model, like image edges, the quality of those features is undertaken and included in the hypothesis.

Table 1: Distribution of numbers of required interaction steps

initial interaction	1 corrective interaction step	2 corrective interaction steps	≥3 correct. interaction steps
4	9	4	0

4.3 Extensions

The selection of hypotheses after the initial step has to be extended to all known detection problems and be included in the current system.

By an analysis of the results of the USC-MABS, areas with possibly undetected buildings could be found. One way to accomplish that would be to represent each hypothesized parallelogram with a fixed reference point and cluster those points in the image with a hypothesis score derived during the selection process. Comparing the cluster image with already found buildings can give hints for missing buildings. The system would determine probable areas and ask the user whether a building was not detected.

References

[1] Bejanin, M., Huertas, A., Medioni, G., and Nevatia, R., *Model Validation for Change Detection*, Proc. of the ARPA Image Understanding Workshop 1994, Volume I, pages 287-294

[2] Collins, R. et al., *Site Model Acquisition and Extension from Aerial Images*, Proc. of the International Conference on Computer Vision, 1995

[3] Kass, M., Witkin, A.P., and Terzopoulos, D., *Snakes: Active Contour Models*, International Journal of Computer Vision, No. 4, January 1988, pages 321-331

[4] Lang F., Löcherbach T., Schickler W., *One Eye Stereo System for 3-D Semiautomatic Building Extraction*, GIM 1995, Netherlands, to appear

[5] Lin, C., Huertas, A., Nevatia, R., *Detection of Buildings Using Perceptual Groupings and Shadows*, Proc. of IEEE Conference on Computer Vision and Pattern Recognition 1994, pages 62-69

[6] McKeown, D.M., and Denlinger, J.L., *Cooperative Methods for Road Tracking in Aerial Imagery*, Proc. of IEEE Conference on Computer Vision and Pattern Recognition 1988, pages 662-672.

[7] Neuenschwander, W., Fua, P., Szekely, G., Kubler, O., *Making Snakes Converge from Minimal Initialization*, Proceedings of the International Conference on Pattern Recognition 1994, pages 613-615

[8] Roux, M., McKeown, D.M., *Feature Matching for Building Extraction from Multiple Views*, Proc. of IEEE Conference on Computer Vision and Pattern Recognition 1994, pages 46-53

[9] Sarkar, S., Boyer, K.L., *Using Perceptual Inference Networks To Manage Vision Processes*, Proceedings of the International Conference on Pattern Recognition 1994, pages 808-810

[10] Schickler, W., *Feature Matching for Outer Orientation of Single Images Using 3-D Wireframe Controlpoints*, Proc. XVII ISPRS Congress, Vol. 29, Part B3, Washington D.C. 1992, pages 591-598

[11] Strat et al., *The RADIUS Common Development Environment*, Proceedings of the DARPA Image Understanding Workshop 1992, pages 215-226

supposed to be rather precise (otherwise no hypothesis would have been established). The gray-level along the sides of the roof is supposed to change only on the non-shadow sides. The overall average gray-level should be low and the variance rather small.

This analysis leads to an easily derivable set of parameters which are used for the calculation of the most likely hypothesis.

3.2 Manual Feature Extraction

If the building is still not correctly detected, additional information is needed and one has to go one step backwards in the hierarchy of USC-MABS to extract new features, like edges or corners. Two ways of correcting the first hypothesis are offered: first the user can adjust the roof-parallelogram by dragging sides with the mouse, rotating or translating the whole model. Changes can only be made within the constraints of the building model, for example opposite sides remain parallel (see Figure 5). The extraction of a ground corner or -edge (shadow corner or -edge) will determine the building height. These interactions are similar to a completely manual system.

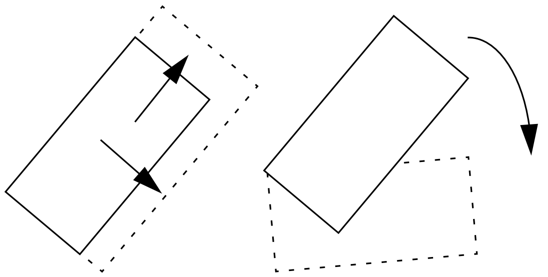


Figure 5 Manual adjustments - sides and rotation

Second, one can choose to extract edges and corners and associate them to a part of the building model. For example, a roof-side of the building can be specified by an edge extracted in the image. Then this edge is added to the current hypothesis. Our systems are implemented to run under the RCDE [11]. This environment allows the use of mouse-sensitive features thus facilitating user selection and manipulation of features.

After each corrective interaction the system forms a new parallelogram-hypothesis, the system looks for new edges, shadow and wall evidence to support the new hypothesis and finally performs a fitting and verification step. These methods are the same as those in the automatic system. This important step of verifying the consistence to the constraints proposed in the USC-MABS can be compared to a fitting process in a computer assisted manual systems, though in our system a fitting is performed after each interaction. Therefore it is possible that, after a manual correction of a roof-boundary, the wrong building height is also corrected automatically.

Without the fitting step the system would perform like a manual system and at least three interaction steps (two corner adjustments and one correction of the building height) would be necessary for adjusting the shape of one building model. Rotation and translation as parameters of the position add another two steps.

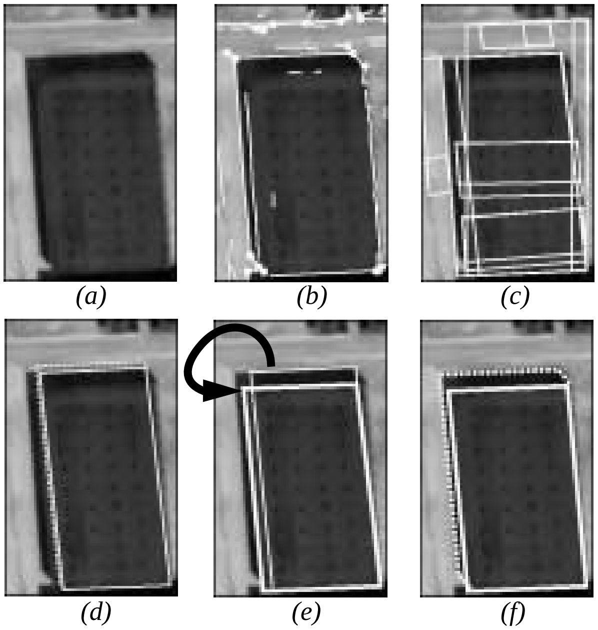
Note, that the manual feature extraction and the following fitting and verification steps can also be applied to buildings that are automatically detected, but partially wrong.

4 Results and Extensions

4.1 Examples

The system was tested on a number of examples provided by the RADIUS [11] program (oblique and nadir views). In Figure 4, an example of using only initial interaction was shown. In Figure 6 the building is not correctly detected due to missing edges. There is no correct parallelogram formed and all roof-hypotheses in *(c)* are rejected by USC-MABS. After the initial interaction, a partly wrong roof-hypothesis *(d)* is found, where the shadow casting roof boundary are missed. The dotted line shows the estimated shadow boundary. The adjustment of one corner *(e)* leads to a new hypothesis. Note that after the correction of the corner, the system found automatically the associated shadow boundary (dotted line) and it corrected the building height.

In Figure 7 an L-shaped building is not correctly



(a) *(b)* *(c)*

(d) *(e)* *(f)*

Figure 6 An undetected building, which is correctly extracted after one corner-correction, the associated shadow-boundary is corrected automatically (see text for details) match of parallelogram and roof sides, which would lead to a correct guess after the initial interaction step.

This set of patterns has to be established by the designer of the system after an analysis of system failures.

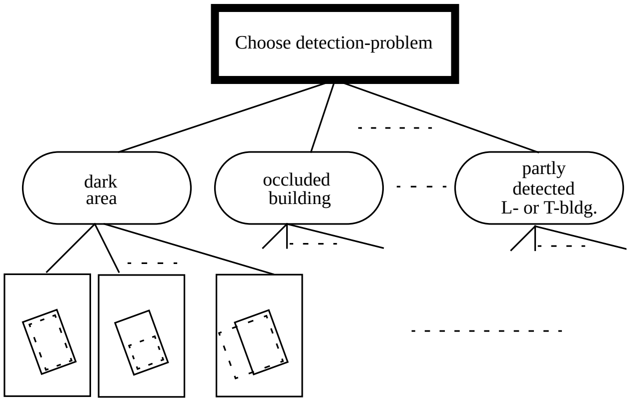


Figure 3 Example for classes of detection-problems and their patterns

Once a class of problems is selected, a probability for being the missed hypothesis is assigned for each parallelogram according to the set of patterns: observation x_i for each pattern j is collected and transformed to a number ω_i which can be related to the associated likelihood. The observation can be represented either as a real number, an integer or a boolean.

$$\omega_i = \frac{1}{2}\left(\frac{x_{ij}-\bar{x}_{ij}}{\sigma_{ij}}\right)^2 \quad \text{for real numbers, or}$$

$$\omega_i = -\ln P\left(x_{ij}\right) \quad \text{for integers/boolean}$$

These formulas are derived by assuming Gaussian distribution. x_{ij}, σ_{ij} or $P(x_{ij})$ (mean value, standard deviation or probability of observation x_{ij}) are parameters which have to be determined either theoretically or empirically.

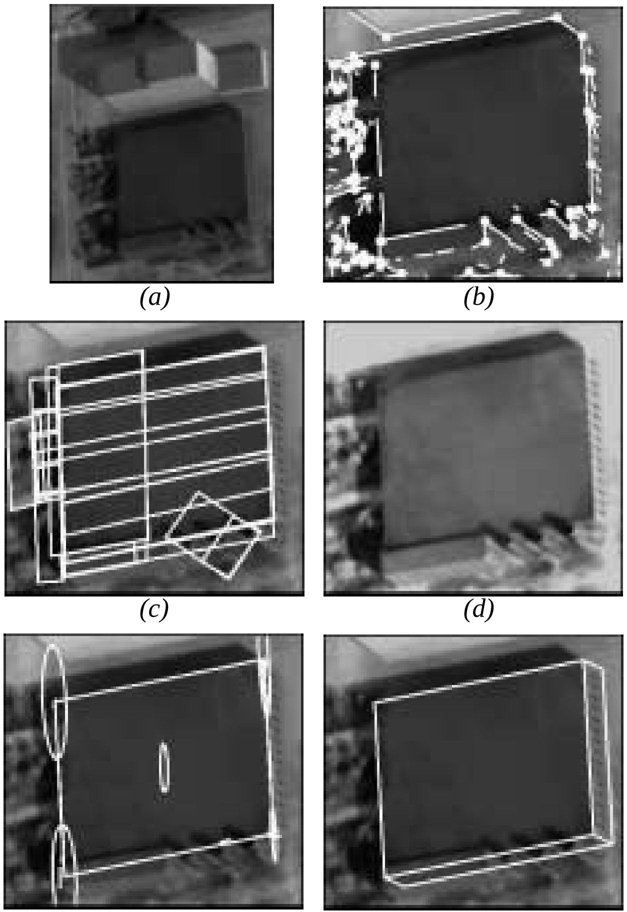
For each pattern $e^{-\Sigma\omega_i}$ is proportional to the likelihood, so that the most likely pattern for each parallelogram can be chosen. And similar the most likely hypothesis for the roof of the missing building is selected by comparing the ω of the most likely pattern of each parallelogram.

An advantage of this selection method is that the system can - because of the selected pattern - give a prediction with a certain probability wether a corrective interaction is necessary and also where they have to be made. Also note that the selection process could not be implemented in the automatic selection step, because too many wrong hypotheses would be accepted - the automated system does not know for sure that there is a certain building at this location.

Example: dark buildings

As an example we want to introduce possible evidence for the problem-class of dark buildings. Usually the boundary to detect between the shadow and the roof will be difficult. The image edges of two sides of the roof are at best only partly visible. Three observations are sufficient to select the best hypothesis available after the Perceptual Organization: evaluation of the parallelogram-corners, of grayvalue-changes at the roof boundaries and of the overall average gray-level. Two patterns are used, one where all sides are correct and one where one or two sides nearby the shadow are incorrect.

It is possible to calculate the roof boundaries and corners which cast the shadow; the corner formed by these roof-sides is likely to be very inaccurate, while the corner formed by the other two sides (non-shadow casting) is



(a) (b)

(c) (d)

(e) (f)

Figure 4 an example of a dark building which is not found by MABS (a), its image-segments and junctions found (b), the roof-hypotheses (c). After specifying the detection-problem, the image-contrast is enhanced for the display (d). A roof-hypothesis with error-ellipses of corners and center of gravity(e). The 3-D building-model found just after initial interaction (f) shows an example of the performance of this system. In Figure 2 we show a diagram of the USC-MABS enhanced for interactive assistance.

The system performs reliably on images where the buildings are separated, wall-, shadow- or roof-lines are at least partially well detected. Otherwise, it may fail to derive a three-dimensional description of a building, even if there is information extracted that at least partly contain the desired result. A rejection of a correct hypothesis can occur during verification or during selection. A correction here would just be to specify this particular parallelogram. It is also possible that the correct hypothesis is not generated due to lack of existing features in the image. Still partly correct hypotheses are likely to exist and the missing feature(s) have to be added to obtain a correct description of a building.

3 Interacting with an Automated System

Our approach combines aspects of the automatic system with user interaction. After an image is processed by MABS, the interaction system starts. The process of interaction can be divided in two parts, *initial interaction* and *corrective interaction* (see Figure 2).

- Initial Interaction (qualitative)
  First the user classifies the detection problem: dark areas, poor contrast, occluded buildings, occluded shadows, partly detected L or T-buildings (this list can be extended). The classification scheme depends on the performance of the automated system. This qualitative information is very useful to constraint the search for new hypotheses. Although the classification is in general not unique (usually several problems occur at the same time), it is unlikely that a correct hypotheses will be rejected as long as the classification is correct.
  The second qualitative step is a very rough localization of the missing building. This can be, for example, any point on the roof (it is possible to automate this step by clustering rejected hypotheses, see below). After the initial interaction the most likely hypothesis can be established by the additional information provided by the user.
- Corrective Interaction (quantitative)
  If the hypothesis established in the first step is (partly) wrong, the user has to correct the sides or corners of the building model. For example, if one roof-side is incorrect, the user can either drag the line to the desired location or select a line in the image which refers to the roof-side. After each single correction, both the verification and fitting of the parallelogram and the determination of the building height will be started. This has to be repeated until the model is fitted to the image. In worst case the complexity of interaction here will be the same as in adjusting the shape of a building model in a manual system

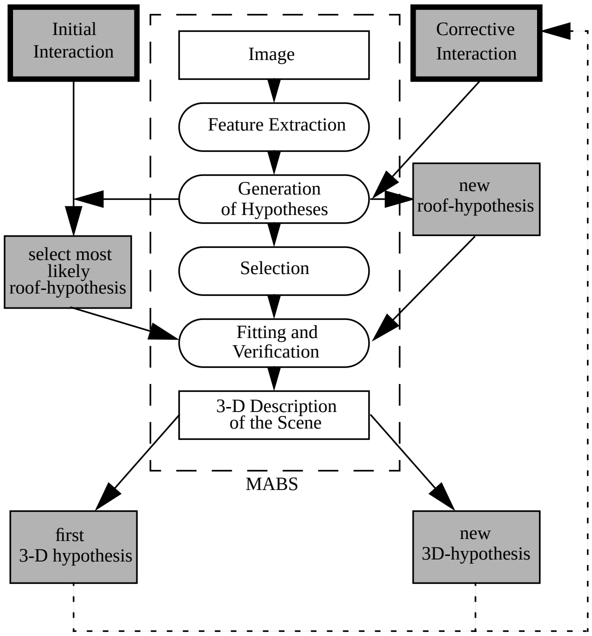


Figure 2 Interaction system embedded in MABS

As mentioned above in order to use intermediate results and computations of an automated system one has to backtrack the steps until given user inputs can be used. Analyzing the existing building detection system, the earliest stage would be before the selection of hypothesized parallelograms. At this stage, the information, which is computed by the feature extraction and perceptual grouping is still available and a unique representation for a possible building can be obtained already.

3.1 Selecting the most likely hypothesis

Input for this step is the result of the initial interaction and the set of roof-hypotheses generated by perceptual organization. The initial interaction constrains the search in the set of hypotheses. According to the specified area a local subset of hypothesized parallelograms is established, from which the most likely hypothesis according to the detection problem is computed. When no detection problem is specified and therefore no specific knowledge of the scene is known, the system uses the score computed during the selection process of USC-MABS.

A set of parallelogram-patterns is assigned to each detection problem, which classifies the parallelogram hypotheses that can occur for a certain problem. An example of a pattern is a parallelogram, in which one roof-side is wrong by a translation (because there were no edges detected at this roof side), all other sides and the angles are correct (see Figure 3). Another example is a complete

The design goals for our system can be summarized as follows:

- the complexity of the interaction-process should be minimized and in worst case not exceed the complexity of the interaction-process required by a manual system.
- the type of information called up in each step should be easy for the user to determine.

"Easy" information for the user would be qualitative information without the need of precision, like answering the question *"Is in the indicated area a building visible but not detected?"*. The last requirement could also be stated as: the precision required by the user should be minimized.

2 Previous work

Obtaining three-dimensional object information can be achieved in several ways. In a completely manual system each 3-D point of an object has to be measured in order to calculate the shape. These points can be measured with a photogrammetric stereoplotter by corresponding image points of an overlapping pair of images.

In computer assisted manual systems an object-model is selected and the parameters are extracted by adjusting the model according to the image. The manual extraction is supported by fitting the model either to the image or to extracted features such as image-edges. The model can be parametrized or generic, like active contour models.

There also exist a few automated approaches [2][5][8], where [5] is the only monocular system. In this system low-level features are grouped and used to hypothesize simple geometric models, which are selected and verified by shadow and wall evidence.

2.1 Computer Assisted Manual Systems

Lang et al. [4] use parametrized models which inherently contain geometric and logical constraints. For example a rectangular building can be expressed by three shape-parameters (length, width, height) and the position of the model according to a reference coordinate system (translation, rotation). A specific building-model is selected from a database and projected into the image. The user adjusts the position and the shape of the model in order to give an initial guess for the final fit. The adjustment is a sequence of interaction steps; in each step one has to choose a reference point and a flexible point in the model which determines the parameters to be adjusted. Depending on the choice of these points, up to two parameters can be changed in one step by moving the mouse cursor. The model is updated and projected in real time, so that the operator can adjust the model according to the given image. Thereafter the fitting consists of a robust estimation of the model parameters [10] by matching image-edges with model-edges. This can be applied to one or more images, The user however has to be more precise if only one image is available.

Neuenschwander et al. [7] suggest a different way using active contours [3]: an initial curve given by the user is optimized by propagating edge-information from the end points to the center of the curve. Still a set of curve end points has to be entered interactively, where the choice of end points is not unique. A number of parameters that regulate the curve's tension and rigidity are needed and are not necessarily easy to obtain. There is also no guarantee that the process will converge to the desired result but in a local minimum.

2.2 An Automatic System.

Our work is based on USC-MABS, which in the following is described briefly. It requires only one image with tilt angle, swing angle and sun direction as input and assumes weak perspective. The building model which is used here consists of rectangular parallelepipeds that can be composed to, for example, L,T or I-shapes. After a linear feature extraction roof-hypotheses are formed by an hierarchical perceptual grouping (lines, folded lines, parallels, U-contours, parallelograms). This step creates a large number of parallelograms, allowing the establishment of also weak hypotheses. Next the parallelograms are selected by local and global criteria using domain knowledge and shadow- and wall information. Finally a verification is applied to validate the selected hypotheses using shadow and wall evidence. This step results in a three-dimensional model of the building, which is determined by the associated shadow and wall evidence. Figure 1

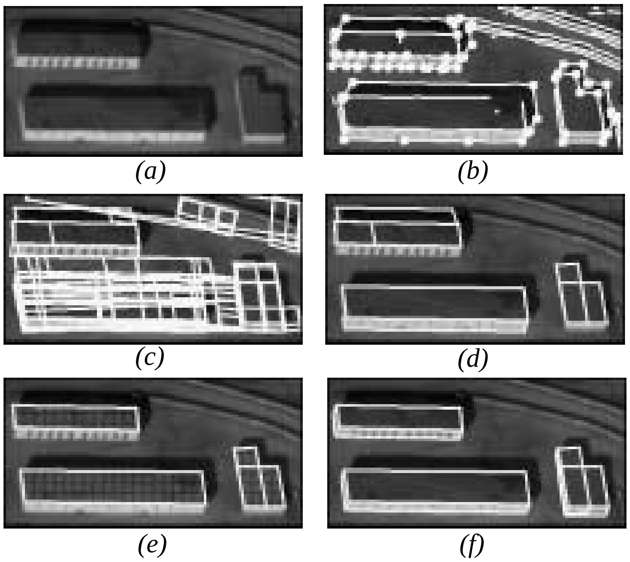


Figure 1 Automatic detection of a building scene (a): extracted segments and corners (b), generated hypotheses (c), selected hypotheses (d), verified roof-hypotheses (e) and their wire frame (f)

Including Interaction in an Automated Modelling System*

Stephan Heuel† and Ramakant Nevatia
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273
Ph.: (213) 740-6428, Fax: (213) 740-7877

Abstract

An approach for including interaction in an automatic building detection and description system is described. It uses intermediate results and computations of the automatic analysis to add or correct the 3-D description of a scene. The proposed method requires a minimum amount of easily obtainable information from the user. The number of interaction steps is less or equal to those of computer assisted manual systems with a final fitting step. The system is built on top of a a monocular automatic building system developed at USC and has been tested on a number of examples with good results.

1 Introduction

The extraction of 3-D site-models from aerial images is an important task and has its application in numerous fields like cartography, city planning, photo interpretation and change detection [1]. Also with the development of GIS (geographical information systems) the need of updated three dimensional information is very high. However, the problem of 3-D site modelling is difficult due to many reasons. The images are quite complex, contain significant texture and the 3-D structure must be inferred from one or more 2-D images. There has been significant progress in automated systems for extracting cultural features from aerial images ([2][5][6][8][9]), but the results are not perfect and some manual editing is still required.

A variety of interactive systems have been built for site model construction ([4][7][11]). The amount of interaction varies from almost complete manual operation with an operator locating all the significant features to where the operator selects a parametric model or a rough outline which is then fit to image data under operator control. In all such cases, the machine's task is limited to that of bookkeeping, simple geometric calculations or some form of error minimization. No perceptual capability of the machine is utilized and the operator is required to provide a large number of inputs, in some cases very accurately. We will refer to such systems as *computer assisted manual systems*. While such systems can aid in constructing site models from aerial images, they are quite tedious to use as the number of structures to be extracted is typically large.

We propose an alternative strategy for combining the activities of the operator and the machine by taking advantage of what perceptual abilities a machine does have. Our goal is to provide a minimum amount of input to the machine and let the machine make the decisions that it can. Our approach is based on the observation that the automatic systems often work quite reliably under certain conditions and the operator should not need to do this work. Also, when automatic systems fail, they fail due to some salient difficulties. In such cases, the operator may be able to supply an indication of the difficulty or the desired result which may suffice for the machine to finish the computation.

We use a monocular automatic building system developed at USC (called USC-MABS from now on) as the underlying system [5]. This system is designed to find rectilinear building structures from a monocular image. It uses perceptual grouping techniques to find likely candidates for building roofs and uses shadows to confirm them and to estimate their height. The system's performance is generally quite good when sufficient parts of the roof boundaries can be extracted. However, in some cases, such as when the roof is dark, the boundary of the roof with the shadow is not detected and the system fails to confirm the presence of such buildings due to lack of sufficient evidence. In such cases, a very simple guidance from the operator, just indicating that in fact a dark building is present in the vicinity suffices for the automated system to find one on its own!

Our methodology does allow for more detailed interaction with the system, in stages, and as necessary. In the worst case, the system reduces to the user having to provide all the information as is the case for most manual systems. However, we find this capability is seldom needed in our system.

---

* This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Army Topographic Engineering Center under Contract No. DACA76-93-C-0014.

† Stephan Heuel was a Visiting Researcher at USC from the Institute of Photogrammetry in Bonn, Germany.