

to eliminate the false positives is to use negative evidence. For example, if the system can not find any wall evidence when the hypothesis is supposed to have a large area of visible walls, this would be considered as a negative evidence and the system might reject the hypothesis based on this observation.

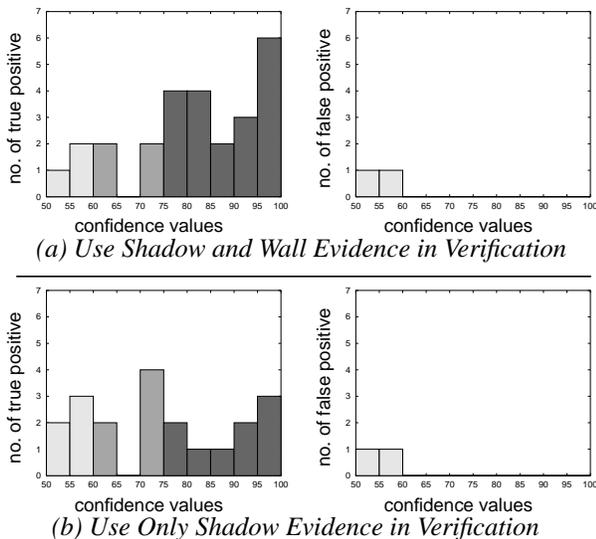


Figure 11 Advantage of Using Wall Evidence

## 5 Conclusions and future work

We have described an automatic system for detection and description of buildings from oblique aerial images. Oblique views provide more evidence for verification, but hypotheses are harder to create. The building height estimation is more reliable with the use of both wall and shadow evidence. We believe that the results show that the system gives good performance, particularly on large buildings with reasonable contrast and shadows. We also believe that the confidence measures offers a tool that can help utilize the results even when they are not perfect. In future work, we intend to work on extending the range of imaging conditions and complexities of shapes that our system can handle.

## References

- [Canny, 1986] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679-698, Nov 1986.
- [Herman & Kanade, 1986] M. Herman and T. Kanade, "Incremental Reconstruction of 3D Scenes from Multiple, Complex Images," *Artificial Intelligence*, 30(3): 289-341, Dec 1986.
- [Huertas & Nevatia, 1988] A. Huertas and R. Nevatia, "Detecting Buildings in Aerial Images," *Computer Vision, Graphics and Image Processing*, 41(2): 131-152, Feb 1988.
- [Irving & McKeown, 1989] R. Irving and D. McKeown, "Methods for exploiting the Relationship Between Buildings and their Shadows in Aerial Imagery," *IEEE Transactions on Systems, Man and Cybernetics*, 19(6): 1564-1575, Nov/Dec 1989.
- [Jaynes, et al., 1994] C. Jaynes, F. Stolle, and R. Collins, "Task Driven Perceptual Organization for Extraction of Rooftop Polygons," *Proceedings of the 1994 ARPA Image Understanding Workshop*, 359-365, 1994.
- [Lin, et al., 1994] C. Lin, A. Huertas, and R. Nevatia, "Detection of Buildings Using Perceptual Grouping and Shadows," *IEEE Proceedings of Computer Vision and Pattern Recognition*, 62-69, 1994.
- [Liow & Pavlidis, 1990] Y. Liow and T. Pavlidis, "Use of Shadows for Extracting Buildings in Aerial Images," *Computer Vision, Graphics and Image Processing*, 49: 242-277, 1990.
- [McGlone & Shufelt, 1994] J. McGlone and J. Shufelt, "Projective and Object Space Geometry for Monocular Building Extraction," *IEEE Proceedings of Computer Vision and Pattern Recognition*, 54-61, 1994.
- [Mohan & Nevatia, 1989] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3-D Structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11): 1121-1139, Nov 1989.
- [Nevatia & Babu, 1980] R. Nevatia and R. Babu, "Linear Feature Extraction and Description," *Computer Vision, Graphics and Image Processing*, 13: 257-269, 1980.
- [Noronha & Nevatia, 1996] S. Noronha and R. Nevatia, "Detection and Description of Buildings from Multiple Aerial Images," *Proceedings of the 1996 ARPA Image Understanding Workshop*, 1996.
- [Roux & McKeown, 1994] M. Roux and D.M. McKeown, "Feature Matching for Building Extraction from Multiple Views," *IEEE Proceedings of Computer Vision and Pattern Recognition*, 46-53, 1994.
- [Shufelt & McKeown, 1993] J. Shufelt and D. McKeown, "Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery," *Computer Vision, Graphics and Image Processing*, 57(3): 307-330, 1993.
- [Venkateswar & Chellappa, 1990] V. Venkateswar and R. Chellappa, "A Framework for Interpretation of Aerial Images," *Proceedings of the International Conference on Pattern Recognition*, 204-206, Jun 1990.

rameters over a set of training examples. All the results shown here use the same parameters.

#### 4.2 Detection evaluation

There are many ways to measure the quality of the results [McGlone & Shufelt, 1994; Shufelt & McKeown, 1993]. We use the following four measurements:

- Detection Percentage =  $100 \times TP / (TP + TN)$
- Branch Factor =  $FP / (TP + FP)$
- Correct Building Pixels Percentage.
- Correct Background Pixels Percentage.

The first two measurements are calculated by making a comparison of the manually detected buildings and the automated results, where TP (True Positive) is a building detected by both human and the program, FP (False Positive) is a building detected by the program but not human, and TN (True Negative) is a building detected by human but not the program.

The other two measurements are calculated by labeling every pixel in the image as either a building pixel or a background pixel [McGlone & Shufelt, 1994; Shufelt & McKeown, 1993]. We calculate the percentage of the number of pixels correctly labeled as building pixels over the number of building pixels in the image and the percentage of the number of pixels correctly labeled as background pixels over the number of background pixels in the image.

Table 1 shows the evaluation on the results of our system on six modelboard images, all of the same site as shown in Figure 8, but taken from different viewpoints and under different illumination conditions (unfortunately, we are unable to show the results graphically due to lack of space).

	Detection Percentage $tp/(tp+tn)$	Branch Factor $fp/(tp+fp)$	Correct Building Pixels	Correct Background Pixels
<b>J2</b>	59.1%	0.138	86.4%	99.6%
<b>J3</b>	87.5%	0.028	96.5%	99.5%
<b>J4</b>	64.6%	0.162	90.6%	94.1%
<b>J5</b>	57.8%	0.263	68.3%	96.4%
<b>J6</b>	62.5%	0.143	67.8%	96.9%
<b>J19</b>	54.2%	0.069	80.0%	99.3%

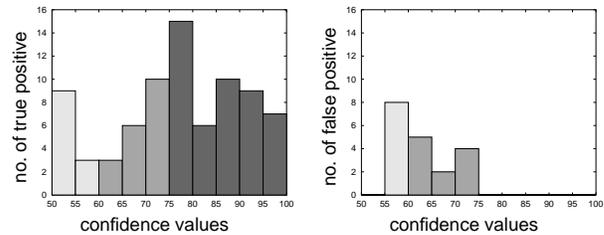
**Table 1** Detection Evaluation

Note that our system gives rather consistent results for most images except for J3 (see Figure 9) which corresponds to a nadir view. In oblique views the hypotheses generation is harder because there are more interference which could make the roof boundaries fragmented.

We believe that this is the reason why the system has lower detection rate on oblique view images. Also note that the measure for correct building pixels is considerably higher than for detection percentage indicating that the missed buildings are rather small. The number for correct background pixels is even higher indicating that false positives are rare and correspond to very small structures. Note that we can make the claim above because there are many buildings in the scene and the ratio of building pixels over background pixels is high. We find that most errors of our system are associated with buildings with dark roofs where the boundary between the roof and the shadow is difficult to detect.

#### 4.3 Confidence evaluation

Our system associates a confidence value with each hypothesis which can further be used to evaluate the performance of the system and guide a user on how to interpret the results. Figure 10 shows a histogram of the number of true and false positives corresponding to certain confidence levels (ranging between 50 and 100, in increments of 5). The confidence values which are between 0 and 1 have been scaled to the range of 0 and 100 for display purpose.



**Figure 10** Distribution of Confidence Values

Note that there are few false positives with high confidence values. In fact, if we set a confidence threshold of 75, we detect no false positives at all and that more than half of the true positives are also above this threshold. This indicates that the confidence values can be used profitably by an end-user or by another program. Results given with high confidence can be taken to be reliable and further attention for improving the results can focus on the lower confidence results, if necessary. We believe that this self-evaluation capability will greatly ease the use of our automatic tool in an interactive environment.

Confidence analysis also gives us a tool for evaluating the effectiveness of using various kinds of evidence. For example, on the J19 image shown in Figure 8, our system finds more true positives when the wall evidence is used. Moreover, if the wall evidence is used, the confidence of the correct hypotheses is increased substantially as shown in Figure 11 (the histogram of the true positives is shifted towards the higher confidence values). Now, if we set a threshold on the appropriate confidence value, the false positives can be eliminated while keeping most of the true positives. Another way

building in Figure 7 (c) occlude the shadow and wall of the other parts. It makes the detection of such building more difficult. Such occlusion can be predicated from partial analysis of the building, however, we have not implemented this capability yet.

The building in Figure 7 (e) is composed of three parts. The part on the lower right corner has different height from the other two parts. Although our system makes a hypothesis corresponding to this part, it can barely find wall or shadow evidence to support the hypothesis; it is difficult for humans to determine the height of this part as well (the height would be easier to infer in an oblique view if a vertical side was visible). There are four gable roof buildings in Figure 7 (f). Our system does not currently model gable roofs; however, these examples are from a nadir view and hence it is able to detect three of these correctly. The fourth building, on the upper right corner, is also detected partially.

Most of our testing has been with the RADIUS modelboard images. These contain no vehicles or vegetation, however, the site is crowded and thus there are many occlusions of wall or shadows between the buildings which make detection difficult. The rest of this section describes the results on the modelboard images which are also used for a quantitative evaluation of our system.

Figure 8 shows the result on an oblique image (J19) from the RADIUS modelboard set containing a large number of structures (about 48). The system forms 2,247 hypotheses and selects 106. Of these, 29 are verified and all but two are correct (i.e. in conformity with the human judgement). The false positives are from small and low contrast structures. The missing structures are also mostly very small and of very low contrast. We feel that the results are very good given the complexity of the image.

Our system also computes a confidence measure (not shown graphically) and the false positives are both of low confidence (confidence evaluation is further discussed in section 4.3 below). The image size is 1306x1034 pixels and the processing time is about 490 seconds on a SUN Sparcstation 20.

Figure 9 shows another modelboard image with a nadir viewpoint. In this example, we can see that the result is much better than the previous one, because the roof boundaries and the shadow boundaries are much clearer in this image. The system forms 2,347 hypotheses, selects 181 hypotheses and verifies 46 hypotheses. Comparisons of the detection rate between J3 and J19 can be found later in section 4.2.

The system uses several parameters in generation, selection and verification of hypotheses. Some parameters, such as the search range of wall and shadow evidence, can be set as a function of the image resolution. Some parameters, such as the weights in the wall and

shadow evaluation functions, are chosen based on our experiences on several test examples. It is also possible that we can have a learning program to find the best pa-

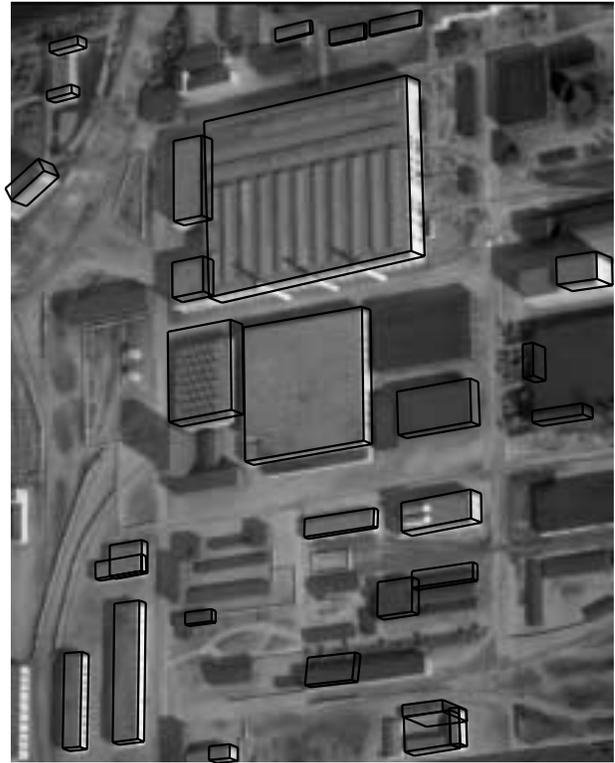


Figure 8 Modelboard (J19)

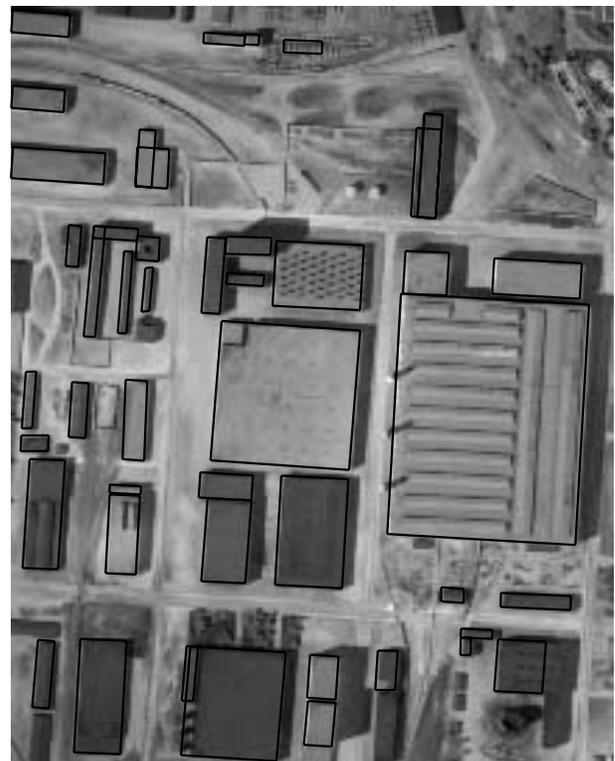


Figure 9 Modelboard (J3)

always the preferred one; sometimes, it may contain elements other than those belonging to a building. Our system makes a choice by examining the evidence of the roof, the wall and the shadow boundaries on the *non-shared* sides of the conflicting hypotheses. In the example shown in Figure 5, the outer hypothesis has more supporting evidence on the non-shared side (AB), so the inner hypothesis is discarded.

Currently our system does not resolve the cases where partial overlap occurs; we expect to do this in the near future.

### 3.5 3-D description of buildings

The 3-D information of the verified buildings, that is the roof hypothesis and the estimated building height, together with the camera model and the terrain model of the scene are used to generate a 3-D wire frame model (a 3-D descriptions) of the scene. The textures inside the roofs and visible walls of verified buildings are painted onto the corresponding surfaces in the 3-D wire frame model. The textures of the ground surface in the input image are painted onto the ground surface of the 3-D wire frame model also. This 3-D wire frame model can be viewed from an arbitrary viewpoint. The transformation that projects the 3-D scene onto a 2-D screen for viewing can then be used to collect the pixel values from the 3-D wire frame model and use them to render the projected image.

## 4 Results and evaluation

Our system has been tested on a number of examples provided by the RADIUS program with encouraging results. We show a few to demonstrate the performance of our system and point out some of the sources of problems. We first show some examples and then give a partial evaluation.

### 4.1 Some examples

Figure 6 shows the image of an L-shape building from Fort Hood, Figure 6 (a), the intermediate results of all major processes, Figure 6 (b)(c)(d)(e), and the final results in 3-D wire frame format, Figure 6 (f). There are 1,049 line segments extracted from the image. The system generates 74 hypotheses, selects 3 hypotheses and verifies 2 hypotheses. The upper part of the structure is verified because it has a clear shadow boundary, although no wall evidence can be found. The lower part of the structure has fragmented wall boundaries and imperfect shadow boundaries, but the system is able to spot the small pieces of evidence and verify it.

Figure 7 shows the results of several examples from the Ft. Hood images. Figure 7 (a) and Figure 7 (b) show two L-shape buildings in Ft. Hood image (fhn713). Note that parts of the shadows fall on the nearby vehicles. Although this makes the shadow boundaries highly fragmented, our system still successfully locates the correct shadow boundaries. Also, in Figure 7 (b) the

building is dark and wall on the left side of the building is inside the shadow and invisible. In this case, the building is verified by the strong shadow evidence.

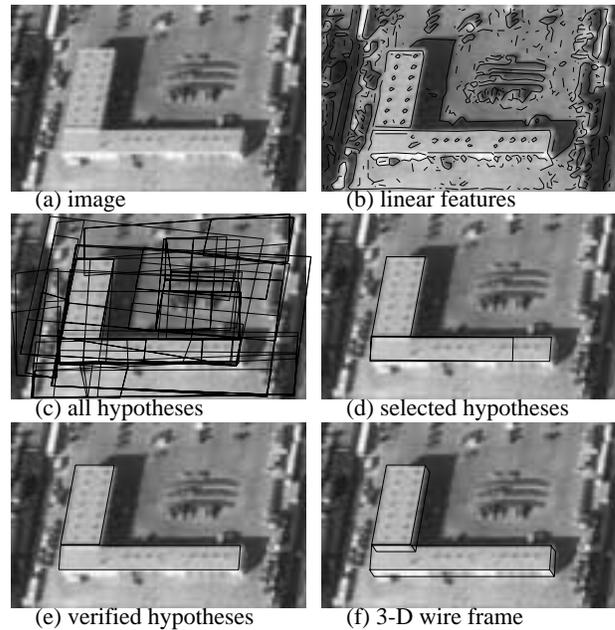


Figure 6 Ft. Hood, Texas (fhov927)

In Figure 7 (c) and Figure 7 (d), note that there are some rectangular shaped surface markings on the ground. The system actually makes roof hypotheses out of these surface markings. However, the system rejects these false hypotheses because no shadow or wall evidence can be found around them. Some parts of the

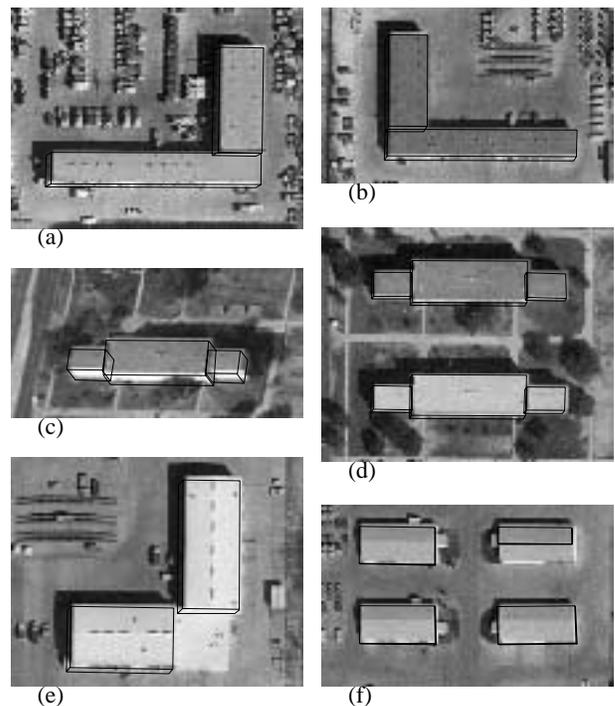


Figure 7 Ft. Hood Examples

The shadow verification process tries to establish correspondences between shadow casting elements and shadows cast. We assume that shadows fall on flat ground. The shadow casting elements are given by the sides and junctions of the selected roof hypotheses. The shadow boundaries are searched for among the lines and junctions extracted from the image.

There are a number of difficulties that prevent the accurate establishment of correspondences. Building sides are usually surrounded by a variety of objects such as loading ramps and docks, grass areas and sidewalks, trees, plants and shrubs, vehicles, and light and dark areas of various materials. Occlusion of the shadow by the building itself or by nearby buildings may make the shadow region irregular and make the shadow evidence difficult to be extracted. To deal with these problems we have adopted some geometric and projective constraints and special shadow features.

The potential shadow evidence is extracted from the linear features of the image and the knowledge of the sun angles: lines parallel to the projected sun rays in the image may represent potential shadow lines cast by vertical edges of 3-D structures, lines having their dark side on the side of the illumination source are potential shadow lines. Junctions among the potential shadow lines are potential shadow junctions, and neighborhood pixel statistics give relative brightness.

Given the sun angles and viewpoint angles, we know which sides of a roof will cast shadow and which part of the shadow will be occluded by the building itself. The shadow is cast along the direction of illumination. The projected shadow width (see Figure 2) can be computed by equation (3) given a possible building height,  $H$ . We can then delineate the projected shadow region in 2-D with the appropriate removal of the self occluded shadow region. The shadow verification process collects all potential shadow evidence along the delineated shadow boundary. For each possible building height, a set of corresponding shadow evidence is collected for evaluation. Figure 4 shows that the system searches for shadow evidence at several possible building heights.

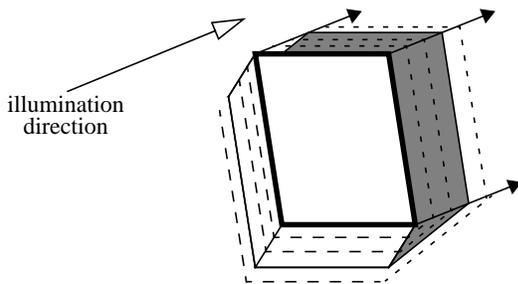


Figure 4 Search for Shadow Evidence

We evaluate the shadow evidence associated with each possible building height and give a score as a

weighted sum of the evidence of shadow lines cast by roof, shadow lines cast by vertical lines, shadow junctions and the shadow region statistics. Equation (5) is used to compute a score for the shadow evidence of a hypothesis  $p$  at the building height  $H$ , where  $h_i$  is an evaluation function for the  $i$ th shadow evidence and  $u_i$  is the corresponding weight.

$$S(p, H) = \sum_i u_i \cdot h_i(p, H) \quad (5)$$

### 3.3 Combination of shadow and wall evidence

For each hypothesis,  $p$ , the previous two steps calculate a shadow score,  $S(p, H)$ , and a wall score,  $W(p, H)$ , for the building height,  $H$ . Next we use equation (6), from certainty theory, to combine these two scores.

$$C(p, H) = S(p, H) + W(p, H) - S(p, H) \times W(p, H) \quad (6)$$

note that  $0 \leq S(p, H), W(p, H) \leq 1$

For each hypothesis,  $p$ , the building height that gives the highest combined score is considered to be the estimated building height of the hypothesis and the corresponding score is called the confidence value of the hypothesis. Equation (7) shows the definitions of estimated building height and confidence value.

$$C_p = C(p, H_p) = \text{Max } C(p, H)$$

where  $\begin{cases} H_p : \text{estimated building height for hypothesis } p \\ C_p : \text{confidence value of hypothesis } p \end{cases}$  (7)

If the confidence value of a hypothesis is greater than a given threshold value, the hypothesis is considered verified. The use of certainty theory in equation (6) allows our system to verify a hypothesis based solely on the wall evidence or shadow evidence. This makes it possible to handle the cases of imperfect wall or shadow evidence.

### 3.4 Containment analysis

The wall and shadow verification processes examine each hypothesis individually and do not analyze any interaction among them. Thus, some verified hypotheses might contain others or they may overlap with each other. A containment analysis of the verified hypotheses is used to resolve the problem of having more than one building in the same 3-D space.

For example, in Figure 5, hypothesis (EFCD) is contained by hypothesis (ABCD). They both have

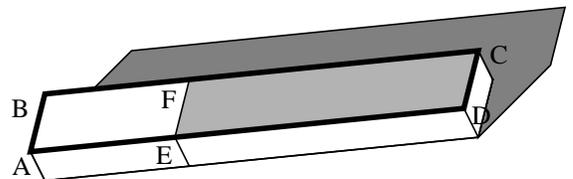


Figure 5 Containment Analysis

enough wall and shadow evidence to allow them to be verified. It is not necessary that the outer hypothesis is

these evidences provide our system the 3-D information to create the 3-D model of the structure.

Figure 2 shows the projection of a typical building and illustrates some of the parameters used by our system. Given a roof hypothesis, we do not know if the hypothesis actually corresponds to a building roof and even if it does, the height of the building is unknown so far. However, we know that the building height,  $H$ , is within a certain range.

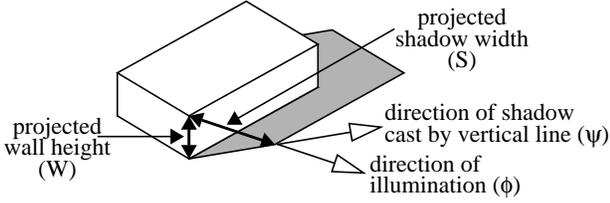


Figure 2 Wall Height and Shadow Width

Assume that the image resolution is  $R$  (pixels/meter). The projected wall height,  $W$ , can be computed from building height and the viewing angles (the swing angle,  $\theta$ , and the tilt angle,  $\gamma$ ) by equation (2).

$$W = H \cdot R \cdot \sin \gamma \quad (2)$$

Also, the projected shadow width,  $S$ , can be computed from the building height, the viewing angles and the sun angles (the direction of illumination,  $\phi$ , the direction of shadow cast by a vertical line,  $\psi$ , and the sun incidence angle,  $i$ ) by equation (3).

$$S = \begin{cases} H \cdot R \cdot \tan i & \text{when } \gamma = 0 \\ \frac{H \cdot R \cdot \sin(i + \gamma)}{\cos i} & \text{when } \begin{cases} \gamma \neq 0 \\ \psi = \phi = 270^\circ - \theta \end{cases} \\ \frac{H \cdot R \cdot \sin(i - \gamma)}{\cos i} & \text{when } \begin{cases} \gamma \neq 0 \\ \psi = \phi = 90^\circ - \theta \end{cases} \quad (3) \\ \frac{H \cdot R \cdot \sin(\gamma - i)}{\cos i} & \text{when } \begin{cases} \gamma \neq 0 \\ \psi = \phi + 180^\circ \end{cases} \\ \frac{H \cdot R \cdot \sin \gamma \cdot |\cos(\psi + \theta)|}{|\sin(\psi - \phi)|} & \text{otherwise} \end{cases}$$

Therefore, we can search for wall and shadow evidence, such as lines and corners, in a certain neighborhood of a given roof hypothesis. Each evidence contributes to the confidence of a hypothesis as explained below (in sections 3.1 through 3.3). Hypotheses with high confidence are considered to be verified.

A containment analysis process is then applied to resolve some of the remaining ambiguities, such as when one verified hypothesis contains another verified hypothesis. This process is explained in section 3.4.

### 3.1 Wall verification process

Generally, some walls of buildings should be visible in an oblique view. As obliqueness increases wall information becomes more useful and shadow information becomes more difficult to handle, if it is available at all. We assume that walls are vertical.

The purpose of the wall process is to find wall evidence at every possible building height for each roof hypothesis. Given the viewing angles and a possible building height, we can estimate wall boundary for a roof hypothesis. All evidence around the wall boundary is collected and a score is computed for the wall evidence.

With the knowledge of the minimum and maximum heights of buildings, the search for wall evidence is limited to a certain range. The system can either do an exhaustive search over the range or do a smart search. The smart search is done by taking samples within the search range to locate some evidence first and then doing a search only on those positions where the chance of finding wall evidence is high. The smart search does not always find the best solution, but with appropriate sampling it could be fast and find the best solution most of the time. Currently we are using an exhaustive search algorithm and the smart search algorithm will be implemented in the near future.

Given a roof hypothesis, view angle information allows us to determine which sides of the building should be visible. The swing angle gives the vertical direction from which building sides are hypothesized. The projected wall height (see Figure 2) can be computed by equation (2) given a possible building height,  $H$ . We delineate the wall boundary and activate a search process to collect all evidence along the delineated wall boundary. For each possible building height, a set of corresponding wall evidence is collected for evaluation. Figure 3 shows the search of wall evidence at several possible building heights.

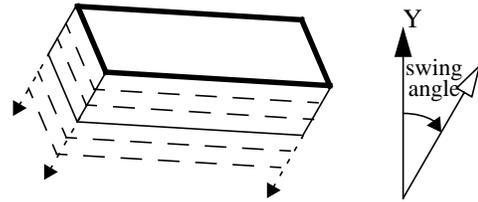


Figure 3 Search for Wall Evidence

The evaluation process evaluates the wall evidence collected from the previous step. Basically the score is a weighted sum of the evidence of ground-boundary, vertical-boundary, and corners. Equation (4) is the evaluation function of the wall evidence of a hypothesis  $p$  at the building height  $H$ .  $k_i$  is an evaluation function for the  $i$ th component of the wall evidence and  $v_i$  is the corresponding weight.

$$W(p, H) = \sum_i v_i \cdot k_i(p, H) \quad (4)$$

### 3.2 Shadow verification process

The use of shadow evidence to verify hypotheses is more complicated in oblique views than in nadir views, for the shadow may be occluded by the building itself (see Figure 2).

& Nevatia, 1988; Irving & McKeown, 1989; Lin, *et al.*, 1994; Liow & Pavlidis, 1990]. However, 3-D cues, such as wall evidence, in monocular images are not fully utilized. Our previous system [Lin, *et al.*, 1994] working on nadir views uses a perceptual grouping technique to generate roof hypotheses from the edges detected from the image. A selection process selects good hypotheses for verification and shadow evidence is used to verify the selected hypotheses. The 3-D information is inferred from the shadow evidence.

In this current work we extend our system to handle oblique view images. Each step requires many changes to accommodate the new difficulties introduced by the oblique view images. For the hypotheses generation process, the skewness of roof hypotheses has to be handled according to the viewpoints and the selection process can make use of the 3-D cues such as OTVs (Orthogonal Trihedral Vertex). In addition to the shadow evidence, wall evidence is used to verify the hypotheses. The use of both shadow and wall evidence make the verification process generate more reliable results and make the system more robust. The corresponding wall evidence of a building also provides another way to infer the 3-D information of the building.

Our system makes the following assumptions: that projection is locally weak perspective, that viewing angles and sun angles are known, that roofs are flat and rectilinear, walls are vertical, and that shadows fall on flat ground.

The system has been tested on several examples of the modelboard images and Fort Hood images provided by the RADIUS program. Some results are shown in this paper and an evaluation of the results is also presented.

## 2 Generation and selection of hypotheses

In this section the process of hypotheses generation and selection is described briefly. The process is similar to our previous work [Lin, *et al.*, 1994] with the appropriate extensions to oblique views and the use of wall and shadow clues.

### 2.1 Generation of hypotheses

First of all, the system use an edge detector to extract intensity linear features from the image. Next, a perceptual grouping process is used to generate roof hypotheses by constructing a feature hierarchy from the linear features.

The feature hierarchy, which includes linear, parallel, U-contour (portions of parallelogram) and parallelogram features, encodes the structural relationships specific to the projection of rectangular shapes, presumably corresponding to the visible flat roof surfaces. A perceptual grouping process is used to group low-level features into high-level features to form the feature hierarchy where linear features are grouped into parallel

features, linear features and parallel features are grouped into U-contour features, and U-contour features are grouped into parallelogram features which are the roof hypotheses.

The hypotheses generation process is more complicated than the one for nadir views as the roofs now may project to parallelograms rather than just rectangles. The degree of skewness of a hypothesis is computed as a function of the swing angle ( $\theta$ ) and tilt angle ( $\gamma$ ) which are available from a camera model. Figure 1 and equation (1) show the angle constraint of roof hypotheses.

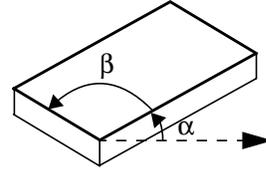


Figure 1 Angle Constraint of Roof Hypotheses

$$\beta = \text{atan}(\mu, \nu)$$

$$\text{where } \begin{cases} \mu = \cos^2(\alpha + \theta) \cos(\gamma) + \frac{\sin^2(\alpha + \theta)}{\cos(\gamma)} \\ \nu = \sin(\alpha + \theta) \cos(\alpha + \theta) \left( \cos(\gamma) - \frac{1}{\cos(\gamma)} \right) \end{cases} \quad (1)$$

### 2.2 Selection of hypotheses

After the formation of all reasonable roof hypotheses, a selection process is applied to choose hypotheses having strong evidence of support and having minimum conflict among them. Based on the local and global supporting evidence of hypotheses, a rule-based selection process selects promising hypotheses for verification. This process greatly decreases the number of hypotheses to be verified, and therefore reduces the run time of the time-consuming verification process.

Our system uses two kinds of criteria: **local selection criteria** and **global selection criteria**. Local selection criteria determine whether or not a parallelogram is *good* based on the local supporting evidence, such as lines, corners, and their spatial relations. Only *good* parallelograms are retained for global selection. It is possible that some of the good parallelograms retained after the local selection are mutually contained or duplicated or overlapped with each other. Global selection criteria will select the best consistent parallelograms from the *good* parallelograms.

Owing to the use of oblique view images, new supporting evidence, such as OTVs, is incorporated into the selection process.

## 3 Verification of hypotheses

The purpose of verification is to confirm that the selected hypotheses correspond to buildings. The existence of wall or shadow evidence increases our confidence that the hypothesis is actually a part of a 3-D structure. Also,

# Buildings Detection and Description from Monocular Aerial Images

Chungan Lin and Ramakant Nevatia\*

Institute for Robotics and Intelligent Systems

University of Southern California

Los Angeles, California 90089-0273

<http://iris.usc.edu/home/seer-00/chungan/building.html>

## Abstract

We describe a method to construct 3-D shape descriptions of buildings from monocular aerial images of general viewpoints. A hierarchical perceptual grouping process is used to generate 2-D roof hypotheses from fragmented linear features extracted from the input image. Good hypotheses are selected and then verified by the corresponding wall and shadow evidence, which also provide the height information for the roofs. Results on several monocular images are shown and the evaluation of the results is presented.

## 1 Introduction

The goal of this work is to detect buildings and generate their 3-D shape descriptions from monocular aerial images of general viewpoints. This work is important for many applications such as automated site model generation, photo-interpretation, cartography, change detection and surveillance. Given the image, the camera model, and the terrain model of the scene, our system can help to create the site model of the scene by providing the 3-D descriptions of buildings in the scene.

There are two major difficulties in inferring 3-D shape descriptions from a single intensity image. First of all, given an image, the system must know how to find and separate objects from the background. This is the well-known “figure-ground” problem. For several reasons, the low-level process usually produces highly fragmented segments, which makes the problem even worse. The other difficulty is to construct 3-D from 2-D, because no direct 3-D information is provided by a single intensity image, although the heights of the buildings can be estimated from the shadow cast by them and by the visible walls under certain assumptions.

Using multiple images can help with both the 3-D inference and grouping processes. However, the 3-D measurements from multiple images are not very accurate for most buildings which are not very tall and grouping still needs to rely on 2-D relations. Also, multiple images are not always available, such as for initial detection, and their use requires accurate camera models. An approach using multiple images in our group is described in a separate paper in these proceedings [Noronha & Nevatia, 1996].

Use of an oblique view provides more 3-D cues than a nadir view, but many additional difficulties arise in the analysis process. First, the contrast between the roof and walls may be lower than the contrast between the roof and the ground causing more fragmented boundaries. Second, small structures such as windows and doors on walls tend to interfere with the completeness of roof boundaries. Third, the projected shape of a building changes with the change of viewpoint. Fourth, the shadow of a building, which we use to verify roof hypotheses, may be occluded by the building itself.

There have been many methods proposed to solve the problem of building detection and description [Herman & Kanade, 1986; Huertas & Nevatia, 1988; Irving & McKeown, 1989; Jaynes, *et al.*, 1994; Lin, *et al.*, 1994; Liow & Pavlidis, 1990; McGlone & Shufelt, 1994; Mohan & Nevatia, 1989; Roux & McKeown, 1994; Venkateswar & Chellappa, 1990]. The segmentation techniques usually rely on regions or edges extracted from the image. Region based techniques construct closed curves that often do not correspond to the objects of interest. Simple edge based techniques such as contour tracing [Huertas & Nevatia, 1988; Jaynes, *et al.*, 1994; Roux & McKeown, 1994; Venkateswar & Chellappa, 1990] encounter the problem of a rapidly growing search space. A more robust edge based technique is the perceptual grouping technique [Jaynes, *et al.*, 1994; Lin, *et al.*, 1994; Mohan & Nevatia, 1989]. For reconstruction of the 3-D information, most of the monocular systems use the corresponding shadow evidence of a building to infer the building height [Huertas

---

\* This research was supported mostly by Contract No. DACA-76-93-C-0014 from the Advanced Research Projects Agency (ARPA) of the Department of Defense and monitored by the Topographic Engineering Research Center of the U.S. Army. Additional support was provided from other grants and contracts from ARPA.