sults, if necessary. We believe that this self-evaluation capability will greatly ease the use of our automatic tool in an interactive environment.

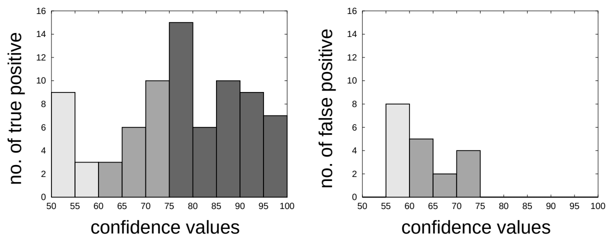
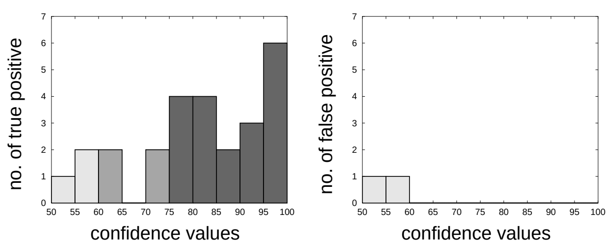


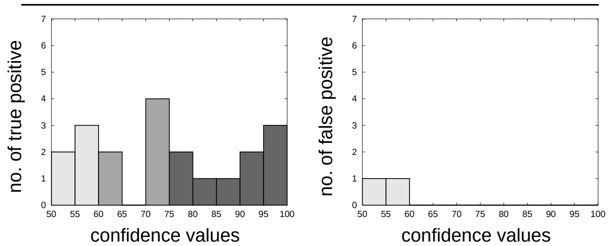
Figure 7 Distribution of Confidence Values

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 6, 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 8 (the histogram of the true positives is skewed towards the higher confidence values). Now, if we set a threshold on the confidence values, the false positives can be eliminated while keeping most of the true positives.

.



(a) Use Shadow and Wall Evidence in Verification



(b) Use Only Shadow Evidence in Verification

Figure 8 Advantage of Using Wall Evidence

6 Conclusions and future work

We have described an automatic system for detection and description of buildings from oblique aerial images. 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

[1] J. Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679-698, Nov 1986.

[2] C. Chung and R. Nevatia, “Recovering Building Structures from Stereo,” IEEE Proceedings of Workshop on Applications of Computer Vision, 64-73, Dec 1992.

[3] M. Herman and T. Kanade, “Incremental Reconstruction of 3D Scenes from Multiple, Complex Images,” Artificial Intelligence, 30(3), 289-341, Dec 1986.

[4] A. Huertas and R. Nevatia, “Detecting Buildings in Aerial Images,” Computer Vision, Graphics and Image Processing, 41(2):131-152, Feb 1988.

[5] 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, SMC 19(6): 1564-1575, Nov/Dec 1989.

[6] C. Jaynes, F. Stolle, and R. Collins, “Task Driven Perceptual Organization for Extraction of Rooftop Polygons,” ARPA IU Workshop 1994, 359-365.

[7] 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.

[8] Y. Liow and T. Pavlidis, “Use of Shadows for Extracting Buildings in Aerial Images,” CVGIP, 49:242-277.

[9] 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.

[10] 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.

[11] R. Nevatia and R. Babu, “Linear Feature Extraction and Description,” CVGIP, 13:257-269, 1980.

[12] 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.

[13] J. Shufelt and D. McKeown, “Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery,” CVGIP, 57(3), 307-330, 1993.

[14] V. Venkateswar and R. Chellappa, “A Framework for Interpretation of Aerial Images,” Proceedings of the International Conference on Pattern Recognition, 204-206, Jun 1990.

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 5.1 below). Image is 1306x1034 pixels and the processing time is about 20 minutes on a SUN Sparcstation 20.

Figure 6 Model Board (J19)

5.1 Detection evaluation

There are many ways to measure the quality of the results [9][13]. We summarize performance on several images in Table 1 using the following four measurements:

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

The first two measurements is calculated by making a comparison of the manually detected buildings and the automated results [9], 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 is calculated by labeling every pixel in the image as either a building pixel or a background pixel [13]. 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.

	Detection Percentage	Branch Factor	Correct Building Pixels	Correct Background Pixels
J2	**59.1%**	**0.138**	**86.4%**	**99.6%**
J3	**87.5%**	**0.028**	**96.5%**	**99.5%**
J4	**64.6%**	**0.162**	**90.6%**	**94.1%**
J5	**57.8%**	**0.263**	**68.3%**	**96.4%**
J6	**62.5%**	**0.143**	**67.8%**	**96.9%**
J19	**54.2%**	**0.069**	**80.0%**	**99.3%**

Table 1: Detection Evaluation

Table 1 shows the evaluation on the results of out system on six model board images, all of the same site as shown in Figure 6, but taken from different viewpoints and under different illumination conditions (unfortunately, we are unable to show the results graphically due to lack of space).

Note that our system gives rather consistent results for most images except for J3 which corresponds to a nadir view. 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. 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.

5.2 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 7 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). 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 re-

For each hypothesis, the building height that gives the highest combined score is considered to be the most likely building height of the hypothesis and the combined score is called the confidence value of the hypothesis. If the confidence value of a hypothesis is greater than a given threshold value, the hypothesis is considered verified.

4 3-D description of buildings

In our system, the shadow and wall evidence is used not only for verification but for reconstruction of 3-D information also. The height of a building can be computed from the projected shadow width (S) and the sun angles (the direction of illumination (ϕ), the direction of shadow cast by a vertical line (ψ), and the sun incidence angle (i)) or from the projected wall height (W) and the viewing angles (the swing angle (θ), and the tilt angle (γ)). Equation (5) shows how to compute the building height from shadow width and equation (6) shows the calculation of the building height from wall height. Figure 4 shows projected wall height and shadow width. From equations (5) & (6), we can derive equations to compute the projected shadow width and wall height for a given building height. These equations have been used to help the verification process to compute the expected shadow boundary and wall boundary.

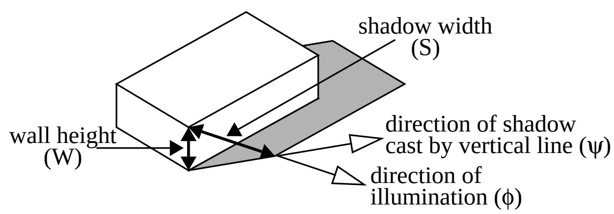


Figure 4 Wall Height and Shadow Width

$$H = \begin{cases} \dfrac{S}{R \cdot \tan i} & \text{when } \gamma = 0 \\ \dfrac{S \cdot \cos i}{R \cdot \sin(i + \gamma)} & \text{when } (\gamma \neq 0 \wedge \psi = \phi = 270° - \theta) \\ \dfrac{S \cdot \cos i}{R \cdot \sin(i - \gamma)} & \text{when } (\gamma \neq 0 \wedge \psi = \phi = 90° - \theta) \\ \dfrac{S \cdot \cos i}{R \cdot \sin(\gamma - i)} & \text{when } (\gamma \neq 0 \wedge \psi = \phi + 180°) \\ \dfrac{S \cdot |\sin(\psi - \phi)|}{R \cdot \sin\gamma \cdot |\cos(\psi + \theta)|} & \text{otherwise} \end{cases} \quad (5)$$

$$H = \frac{W}{R \cdot \sin\gamma} \quad (6)$$

After the verification process, every verified hypothesis will have a building height associated with it. In other words, when a hypothesis is verified it is verified with the best building height of the hypothesis. The building height (H_p) for a hypothesis (p) is the height (H) such that the confidence function $C(p, H)$ is maximized. From the height of the hypothesis the system can generate a description of the shape of the structure and derive a 3-D model.

The 3-D descriptions of the verified buildings together with the camera model and the terrain model of the scene are used to generate a 3-D wire frame model 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.

5 Results and evaluation

Our system has been tested on a number of examples provided by the RADIUS program with very good results. We show a few to demonstrate the performance of our system and point out some of the sources of problems. Figure 5 shows an image of a building from Fort Hood (a), the intermediate results of all major processes (b)(c)(d)(e), and the final results in 3-D wire frame format (f).

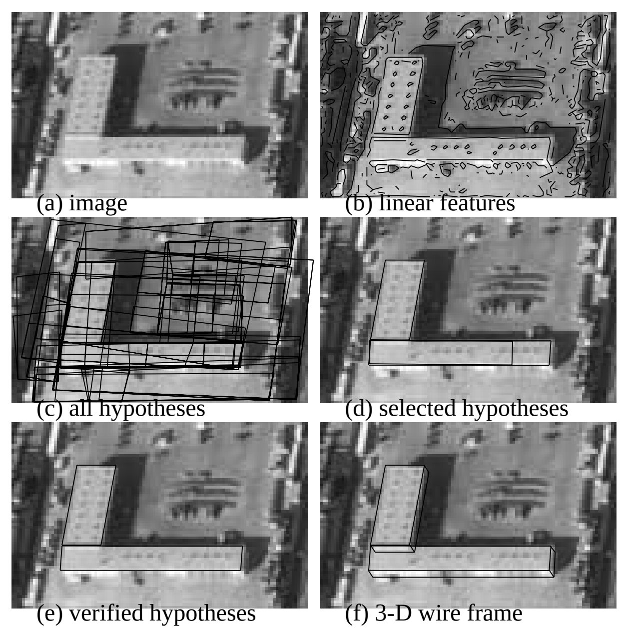


Figure 5 Ft. Hood, Texas (fhov927)

Figure 6 shows the result on an image (J19) from the RADIUS model board 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 (in conformity with the human judgement). The false positives are from very small and low contrast structures. The missing structures are also mostly very small and

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 image elements and 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 can delineate the projected shadow region in 2-D with appropriate removal of self occluded shadow region for a given building height. The shadow verification process collects all potential shadow evidence along the expected shadow boundary. For every possible building height, a set of corresponding shadow evidence is collected for evaluation. The range of possible building height is determined by the knowledge of the maximum building height in the scene. Figure 2

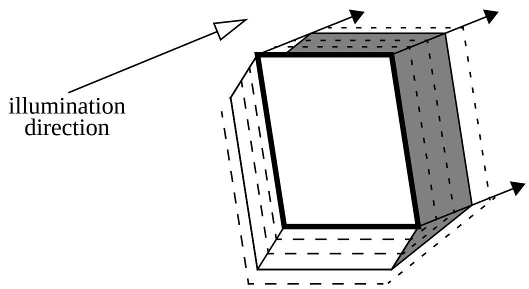


Figure 2 Search for Shadow Evidence

shows how the system searches for shadow evidence on several possible building heights.

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 (2) is used to compute a score for the shadow evidence of a hypothesis p at the building height H. h_i is one of the evaluation functions of different shadow evidence and u_i is the corresponding weight.

$$S(p, H) = \sum_i u_i \cdot h_i(p, H) \qquad (2)$$

3.2 Wall verification Process

Generally speaking, some walls of buildings should be visible in oblique view images. As obliqueness increases wall information becomes more useful and shadow information becomes more difficult to handle, if available at all. Walls, after all, are part of the structures and some strong assumptions can be made, such as that they are vertical.

With the knowledge of the minimum and maximum heights of buildings, we can limit the search for wall evidence to a certain range. The system can either do an exhaustive search over the range or look for some evidence first and do a smart search just on some positions where the chances of having wall evidence are high.

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

Given a roof hypothesis and the viewing angles we determine which sides should be visible. The swing angle gives the vertical direction from which building sides are hypothesized. We delineate the wall boundary for a given building height and activate a search process to collect all evidence around the delineated wall boundary. Figure 3

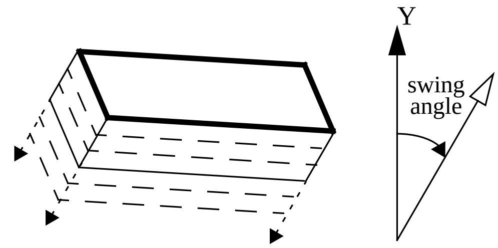


Figure 3 Search for Wall Evidence

shows the search of wall evidence for several possible building heights.

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 (3) is the evaluation function of the wall evidence of a hypothesis p at the building height H. k_i is one of the evaluation functions of different wall evidence and v_i is the corresponding weight.

$$W(p, H) = \sum_i v_i \cdot k_i(p, H) \qquad (3)$$

3.3 Combination of shadow and wall evidence

For each hypothesis, the previous two steps determine a shadow score and a wall score at every possible building height. Next we combine the two. We formulate a function to evaluate a score for the shadow and wall evidence of a hypothesis (p) at a given building height (H). The corresponding shadow and wall scores are given from equations (2) and (3). We use the certainty theory to combine these two scores in equation (4).

$$C(p, H) = S(p, H) + W(p, H) - S(p, H) \times W(p, H) \quad \text{where} \quad 0 \le S(p, H), W(p, H) \le 1 \qquad (4)$$

The possible building heights are limited by the minimum building height and the maximum building height.

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

2 Generation and selection of hypotheses

In this section the process of hypotheses generation and selection will be briefly described. The process is similar to our previous work [7] with the appropriate extensions to oblique views and the use of strong shadow and wall cues.

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 oblique views 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 previous one, for it has to generate parallelogram hypotheses instead of rectangular hypotheses which were used by our previous system. The degree of skewness of a hypothesis is computed as a function of the swing angle (θ) and tilt angle (γ) which are available for the system by our assumptions. Figure 1 and equation (1) show the angle constraint of roof hypotheses.

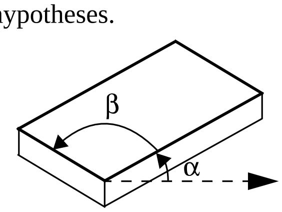


Figure 1 Angle Constraint of Roof Hypotheses

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

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

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, therefore reduce 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. 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 some other good parallelograms. Global selection criteria will select the best consistent parallelograms from *good* parallelograms.

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

3 Verification of hypotheses

The purpose of verification is to validate the selected hypotheses to correspond to buildings. For a roof hypothesis, the existence of shadow evidence or wall evidence strongly suggests that the roof hypothesis is a part of a 3-D structure. Our validation step therefore includes **shadow verification process** and **wall verification process**. A hypothesis could be validated by either shadow and/or wall evidence. Also, these evidence provide our system the 3-D information to create a 3-D model of the structures.

3.1 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 in oblique view images. See Figure 2.

The shadow verification process will try to establish the correspondences between shadow casting elements and shadows cast, and use these correspondences to verify a hypothesis. We assume that the ground surface in the immediate neighborhood of the structure is fairly flat and level. 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 however. 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, 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

3-D Descriptions of Buildings from an Oblique View Aerial Image

Chungan Lin and Ramakant Nevatia*
Institute for Robotics and Intelligent Systems
Computer Science Department
University of Southern California
Los Angeles, California 90089-0273

Abstract

In this paper we describe a method to reconstruct the 3-D descriptions of buildings from an oblique view aerial image. Oblique views present new difficulties over the more conventional nadir views that are now handled by our system. A hierarchical perceptual grouping process is used to generate 2-D roof hypotheses from fragmented linear features of the input image. Good hypotheses are selected and then verified by the corresponding shadow and walls. The 3-D description of a building is reconstructed from the roof hypothesis and the associated shadow and wall evidence. Results on several oblique view images are shown and the evaluation of the results is presented also.

1 Introduction

The goal of this work is to generate the 3-D shape descriptions of buildings from an oblique view aerial image. The work is important for many applications such as photo-interpretation, cartography 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 which provides the 3-D shape 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 reconstruct 3-D from 2-D, because no direct 3-D information is provided by a single intensity image though the heights of the buildings can be estimated from the shadow cast by them and by the visible walls under certain assumptions.

Use of an oblique view can provide us with more 3-D cues than the nadir view aerial image, but many additional difficulties arise in the analysis process. First, the contrast between the roofs and the walls may be lower than the contrast between the roofs and the ground causing boundaries to be even more fragmented. 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 buildings detection and descriptions [3-10] [12][14]. 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 technique such as contour tracing [4][14] usually faces the problem of fast growing search space. A more robust edge based technique is the perceptual grouping technique [6][7][10]. On reconstruction of the 3-D information, most of the monocular systems use the corresponding shadow evidence of a building to infer the building height [4][5][7][8]. Our previous system [7] uses the perceptual grouping technique to make 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.

We use a similar approach in our current system, but each step requires many changes to accommodate the problems 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 assured results and make the system 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 buildings are rectilinear, that the roofs and the surface on which the shadow fall are flat, and that the viewing geometry (camera model) is known.

* This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Army Topographic Engineering Center under contract DACA76-93-C-0014.