COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 72, No. 2, November, pp. 101–121, 1998
ARTICLE NO. IV980724

Building Detection and Description from a Single Intensity Image

Chungan Lin* and Ramakant Nevatia†

Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, California 90089-0273
E-mail: chungan@iris.usc.edu, nevatia@iris.usc.edu

Received April 15, 1997; accepted December 20, 1997

We describe a method to detect buildings and construct 3-D shape descriptions of buildings from a monocular aerial image with a general viewpoint. A grouping process is used to generate 2-D roof hypotheses from fragmented linear features extracted from the input image. Good hypotheses are selected based on 2-D evidence and some local 3-D evidence. These are then *verified* by searching for 3-D evidence, consisting of shadows cast by the roof and the walls associated with it. This process also helps construct 3-D models of the verified buildings. Overlap and containment relations between 3-D structures are analyzed to resolve conflicts. This method also allows for the integration of results obtained from multiple images. This system has been tested on a large number of real examples with good results, some of which, along with their evaluation, are included in this paper. © 1998 Academic Press

1. INTRODUCTION AND OVERVIEW

Automatic detection and description of cultural features from aerial images is of great practical interest for a number of applications such as cartography and photointerpretation. Of the many cultural features to be detected, buildings are perhaps the most salient due to their number and complexity. Building detection and description also serves as an excellent test domain for the more general problem of 3-D object detection and description in computer vision. Buildings usually have simple geometry, such as being polyhedral, and the aerial viewpoint reduces the possibilities of interbuilding occlusions. However, the aerial images are typically very complex and contain a large number of objects in the scene. This makes the problem of object detection more difficult, in many ways, than in factory or indoor environments. We believe that the techniques presented here for building detection also help provide some insight into how to approach the more general problems.

* C. Lin is currently with Nichimen Graphics Inc, Los Angeles, CA 90066.

† This research was supported by Defense Advanced Research Projects Agency under Contract DACA76-93-C-0014, monitored by the U.S. Army Topographic Research Center, and by Grant F49620-95-1-0457, monitored by the Air Force Office of Scientific Research.

1.1. Difficulties

There are several difficulties in detecting and inferring the shape of 3-D objects from intensity images that apply to the building detection task as well.

a. Segmentation. The first difficulty is in finding a desired object and separating it from the background in the presence of distractions caused by other features such as surface markings, vegetation, shadows, and highlights. This is an instance of the well-known *figure–ground* problem. The figure–ground problem is illustrated in Fig. 1a, which shows a portion of an aerial image containing an L-shaped building. Figure 1b shows the line segments extracted from this image. It is clear that the building boundary is fragmented, and there are many extraneous boundaries that do not correspond to the building. Note that most of these other boundaries are not caused by sensor "noise" but by other objects that are present in the scene.

b. 3-D inference. The next difficulty is to infer 3-D structure from 2-D images. If multiple images are available, feature correspondences can be used to infer 3-D (assuming calibrated cameras). In this paper, we do describe a technique using multiple images, but most of the paper is devoted to working with a single image. Direct 3-D information is not available in a single intensity image, although the heights of the buildings can be estimated, under certain assumptions, from the shadow cast by them and by the visible walls. Note that having 3-D information can help with the segmentation problem and vice versa.

c. Shape description. Even after an object has been segmented and some 3-D features have been recovered, we need to build a 3-D shape description of the object. This is not so difficult for polyhedral shapes but we must still reason about combinations of wings (such as the two parts of an L-shape) or about the superstructures.

1.2. Previous Work

Segmentation and description of 3-D objects have been major topics of research in computer vision for many years. However, usually the methods deal with a small number of objects in rather sparse environments. Also, many applications can assume that precise, metric models of the objects to be found are given.

1077-3142/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

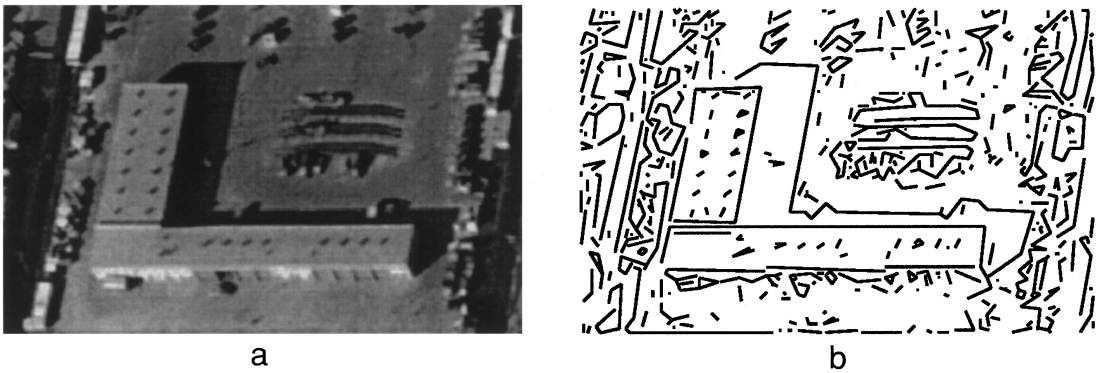


FIG. 1. Segmentation.

Even though such systems may deal with highly complex shaped objects, we do not believe that they can be applied to the task of building detection directly.

There have been many methods proposed to solve the problem of building detection and description in the past [1–10]. Additional techniques have been reported recently, some of which rely on a variety of data sources, such as range data [11–22]. Simple edge-based techniques such as contour tracing [2, 4, 9, 10] encounter the problem of a rapidly growing search space. A more robust edge-based technique is that of perceptual grouping [4, 5, 8]. 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 [2, 3, 5].

Most of the previous systems assume that the images are taken from a "nadir" view where only the roof of a building is visible. The described system is designed to handle images from general viewpoints. An oblique view image 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 as 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 the presence of a building and to estimate height, may be occluded by the building itself.

1.3. Overview

Our basic approach is to use the geometric and projective constraints to make hypotheses for the presence of building roofs from the low-level features and to verify this by using available 3-D cues.

Figure 2 shows the block diagram of the system. Our system design philosophy has been to make only those decisions that can be made confidently at each level. Thus, the hypothesis generation process creates hypotheses that may have rather weak evidence as described in Section 2. A selection process prunes these based on more global evidence. The hypothesis verification process uses the most global information and can therefore make more informed decisions.

Initial hypotheses are generated for 2-D projections of building roofs. The roofs are usually the most distinct features in aerial images, even from oblique views, and focusing on the roofs helps reduce the number of hypotheses to be examined later. Formation of roof hypotheses is facilitated by restricting the building shapes to be rectilinear, (i.e., rectangular or compositions thereof, thus allowing for L, T, and I shapes, for example) and roofs to be flat. We believe that such structures cover a significant fraction of buildings of interest and provide a good test domain and that the approach can be generalized to other known shapes and to other roof types. Note that this method does not require knowledge of the metric shape, such as aspect ratios of the buildings to be detected. Roofs of rectilinear buildings should

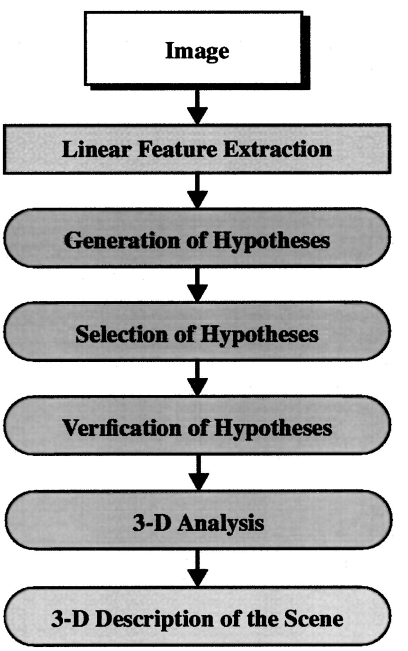


FIG. 2. Block diagram of the automatic system.

project into parallelograms, or their compositions, under weak perspective projection, which adequately models many aerial imaging situations. The hypotheses formation is described in Section 2.

Next, a selection process to choose promising hypotheses based on the detectable evidence for them in the image and on the global relations among them is applied. This process relies primarily on 2-D shape properties, although some local 3-D cues are also used. Thus, it may not distinguish between a rectangular parking lot or a garden from a building roof, for example. This distinction is made by looking for 3-D cues. If multiple images are available, they can provide a reliable method for inferring 3-D. Various cues are also available in a single image but they are indirect. We believe that for the building detection problem, two of the more reliable cues are shadows associated with them and evidence of visible walls, if any. Some specific properties of shadows that should be cast by a hypothesized structure can be obtained from the knowledge of viewing geometry and the sun position (derivable from the time of the day and the lattitude/longitude of the site). Our analysis assumes that shadows are cast on flat ground. In some cases, the shadows may fall on other buildings or be occluded and will be less useful. In an oblique view, some features of the wall should be visible, although these features become unreliable as the viewpoint approaches the nadir. If the combined wall and shadow evidence is strong enough, a hypothesis is considered verified and its 3-D structure is recovered. These 3-D structures are then examined for mutual overlap and containment and a composite 3-D shape description is computed. An analysis for occlusion from nearby structures is also conducted and sometimes helps recover structures that may be missed in the initial analysis. The 3-D shape verification and description process is described in Section 4.

We also describe a version of our system that uses multiple images (in Section 6). In this system, correspondences between the two images are made at the highest level of abstractions, i.e., for the detected buildings themselves rather than at lower levels, such as pixel intensities or line features. We do not necessarily advocate this as a general approach, but it may be advantageous when accurate camera calibration information between the different views are not available or when the two views are taken at different times with very different illumination conditions.

Our system has been tested on several real examples. The experimental results are shown and analyzed in Section 7. Of course, our system is not perfect and can fail to detect some buildings (false detection is much more rare). Because of the hierarchical design of this system, it is possible to go back to a previous layer and correct the results by an interactive process without having to do all the modeling work manually; details of this process are not provided in this paper due to lack of space but a preliminary description may be found in [23]. We believe that further improvements in performance can be obtained by introducing context from other analysis, such as modules that may find roads and vegetation, although we have not experimented with these.

2. GENERATION OF HYPOTHESES

The first step in our system is to make hypotheses for roofs of buildings from a 2-D image. The roofs are hypothesized based on the observation that a planar rectangular feature in 3-D projects to a parallelogram in 2-D under weak perspective projection, which adequately models most aerial imaging situations. The difficulty of hypotheses formation comes from the fact that the low-level features, such as lines, that can be detected from a typical image are highly fragmented and a grouping process is necessary. We can form hypotheses simply by examining all four-tuples of lines in the image; however, this operation will have a time complexity of $O(n^4)$, where n is the number of lines, and will also generate many unnecessary hypotheses. We can reduce this complexity significantly by considering only close, related features and some projective constraints.

We have developed two approaches to grouping. The first approach follows the algorithm described in [8] with appropriate extensions to handle oblique views and the use of strong shadow cues and orthogonal trihedral vertices (OTVs). In this approach, parallelograms are formed hierarchically by first detecting "aligned" parallel pairs (parallel lines whose end points have the desired relationships) and then "U-contours" (three sides of a parallelogram). This approach requires at least one side of a parallelogram to have complete edge evidence and one of the parallel pairs to be aligned for correct hypotheses to be formed; this is not always possible for highly fragmented boundaries or for some shapes that have concavities. A second approach has been developed to form parallelograms more directly, which can overcome this difficulty. The second approach has a slightly higher time complexity and tends to generate more hypotheses but misses fewer of the correct ones. We will only describe the second approach; some details of the first can be found in [24].

The process starts with line segments detected by a line finder (we use the Canny edge detector [25] followed by the USC LINEAR linking and approximation method [26]). Neighboring parallel segments (within 2 pixel distance) are grouped into a single segment whose length and orientation are derived from the contributing segments. L junctions and T junctions are found among these segments, which are also used to break them into smaller segments that we will call *edges*. Next, a colinearization process groups these edges into longer edges to partially overcome the problem of fragmented line segments generated by the low-level vision process. Figure 3 shows the resulting edges and junctions for the image in Fig. 1. These steps are not described in more detail as they are similar to those used in [9]; instead we focus on the step of parallelogram formation.

We can limit our search for roof hypotheses by restricting the parallelogram search to only those that could be projections of a rectangle (general rectilinear shapes are represented as a

composition of some rectangular shapes). Some constraints can be derived from the viewing geometry and the orientation of the parallelogram. Figure 4 shows the viewing angles involved in a simple projection. The tilt angle (γ) is the angle between the vertical direction and the camera optical axis. The swing angle (θ) is the angle between the Y axis of the image and the principal line (projection of a vertical line). These angles are commonly given as part of photogrammetric information associated with an image when a normal "frame" camera is used (for other viewing geometries, equivalent angles can be derived for a specific part of the image).

A 90° corner can be shown to project to an angle, β, given the angle α that one side of the projection makes with the horizontal (an image axis) as shown in Fig. 5, and the viewing angles (θ) and (γ) are given by Eq. (1) below. Any parallelogram we find must satisfy these relationships.[1]

$$\beta = \operatorname{atan}(\mu, v) \tag{1}$$

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

The parallelogram search begins by finding two antiparallel edges (parallel edges of opposite contrast) such that one edge is *contained* by the other. An edge E1 is said to be contained by edge E2 when E1 is inside a region (*containment region* of E2) defined by L1, L2, and E2, where L1 and L2 are the two alignment lines of edge E2. The orientation of the alignment line is calculated by Eq. (1) above. Figure 6 shows an example where E1 is contained by E2. A tolerance distance, which is proportional (multiply by tangent of 2.5°) to the average distance between edge E1 and edge E2, is allowed to decide whether or not E1 is inside the containment region of E2. Therefore, the containment region of an edge is effectively a trapezoidal area.

For each edge that falls between a minimum and maximum length (determined by given or expected limits on building size), a search for antiparallel contained edges is conducted. The edge initiating the search is called an *initial edge*, the others are called *candidate edges* of the initial edge. The search for contained edges can be made significantly faster by using a *spatial index* representation for edges. In this representation, a list of edges passing through each pixel is stored (at a reduced resolution so that a 5×5 pixel area corresponds to one index value). It allows for linear time search of the edges contained in a given region. The candidate edges together with the initial edge are used to trigger the formation of parallelogram features.

Some tests are used to eliminate trigger candidates less likely to correspond to roofs. First, a more distant candidate edge *shadowed* by a more nearby candidate edge is removed. An edge E2

[1] An alternative would be to *rectify* the features by reprojecting to a horizontal plane (suggested by one of the anonymous reviewers).

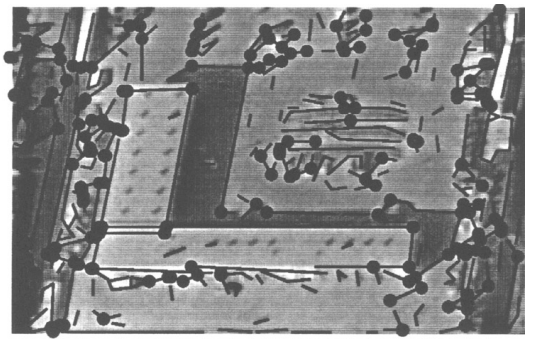


FIG. 3. Lines and junctions.

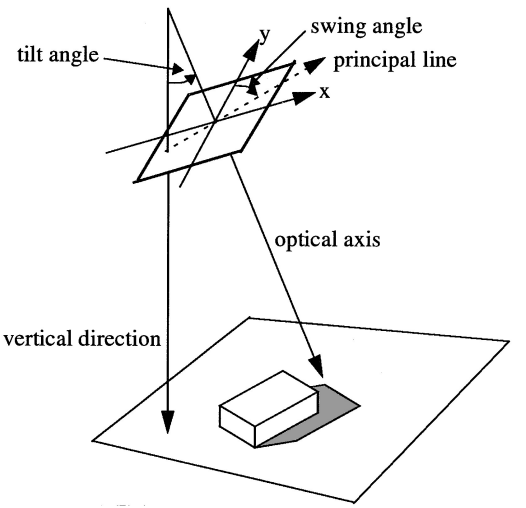


FIG. 4. Viewpoint angles.

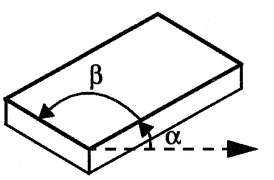


FIG. 5. Right angle constraint.

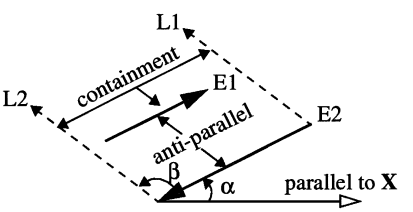


FIG. 6. Antiparallel and containment.

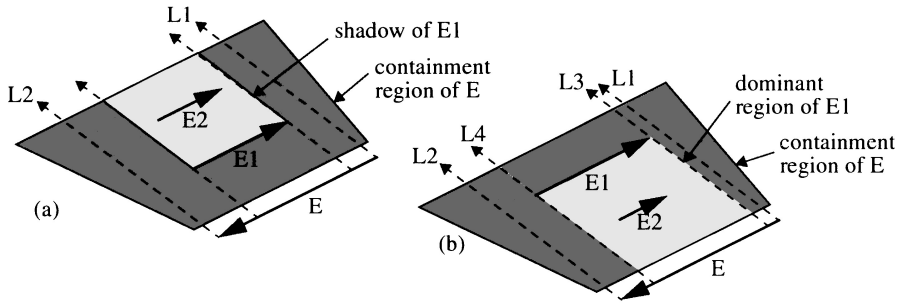


FIG. 7. (a) Shadow of a candidate edge, (b) Dominant region of a candidate edge.

is shadowed by E1 when the projection of E2 to E1 along the direction of alignment line of E is completely inside E1. Figure 7a shows an example where E1 and E2 are two candidate edges of E, and E1 is closer to E than E2 is; L1 and L2 are the two alignment lines of edge E (the shading in the figure is for illustrative purposes and does not represent the image intensities).

Next, if a relatively shorter closer candidate edge is *dominated* by a longer farther candidate edge, the closer candidate edge is removed. Figure 7b shows the dominant region of a candidate edge, E1, defined by L3, L4, and E1. L3 and L4 are two lines passing through the two end points of E1 and parallel to the alignment line of E. The other candidate edge, E2, is inside the dominant region of E1. E2 is relatively shorter than E1, if l_1/d_1 is less than l_2/d_2. l_1 and l_2 are the lengths of E1 and E2 respectively. d_1 and d_2 are the distances from E to E1 and E2 respectively. If E2 is relatively shorter than E1, it is likely that E2 is just a texture edge between E and E1 and accordingly E2 is removed from the list of candidate edges of E.

Each initial edge paired with each of the remaining candidate edges initiates a process of parallelogram feature generation. The trigger edges define two sides of the parallelogram, let us call them the *west side* and the *east side*. The other two sides, say the *north side* and the *south side* (collectively called the *polar sides*), are determined by searching outward (15 piexels) from the end points of the trigger edges as shown in Fig. 8. Here, E1 and E2 are a pair of trigger edges; E1 is the initial edge and E2 is a contained edge. The polar sides must be parallel to the alignment line of the initial edge E1 (so that the resulting parallelogram is a feasible projection of a rectangle). N1, N2, and N3 are the three polar side candidates found around the north side and S1, S2, and S3 are the three polar side candidates found around the south side.

Together with the trigger edges, each candidate of one polar side can be grouped with each candidate on the other polar side to form a parallelogram feature. The west side and the east side of the parallelogram feature are extended, as necessary, to make connections with the polar sides. To reduce the number of hypotheses, polar side candidates that have weak supporting edge evidence are eliminated as are the ones that have lines crossing the polar side candidates (roof boundaries should not have any lines crossing them).

Figure 9 shows the parallelogram hypotheses generated by this method from the line features shown in Fig. 3.

3. SELECTION OF HYPOTHESES

A selection process is applied to the hypothesized parallelograms to choose those that are more likely to correspond to building roofs. We could apply a 3-D reasoning process for this but that is a rather expensive operation. Hence, we first apply a process that uses primarily 2-D evidence, although some local 3-D evidence is also used, followed by a more rigorous 3-D verification process (described in Section 4). Our goal is to reduce the size of the hypotheses set, not to make final decisions. Thus, the method is biased to eliminate only those that can be confidently deleted.

The selection process applies *local* criteria followed by *global* criteria. The local criteria evaluate local supporting evidence,

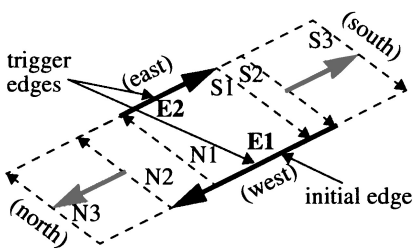


FIG. 8. Search for candidates of polar sides.

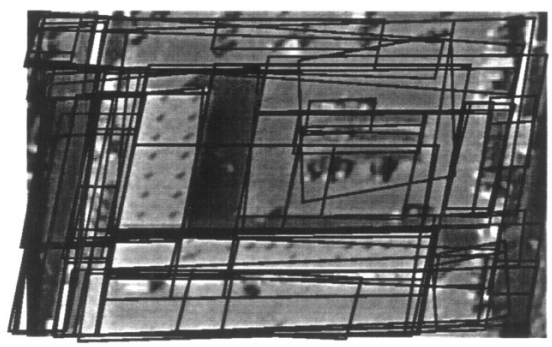


FIG. 9. Parallelogram features: New approach.

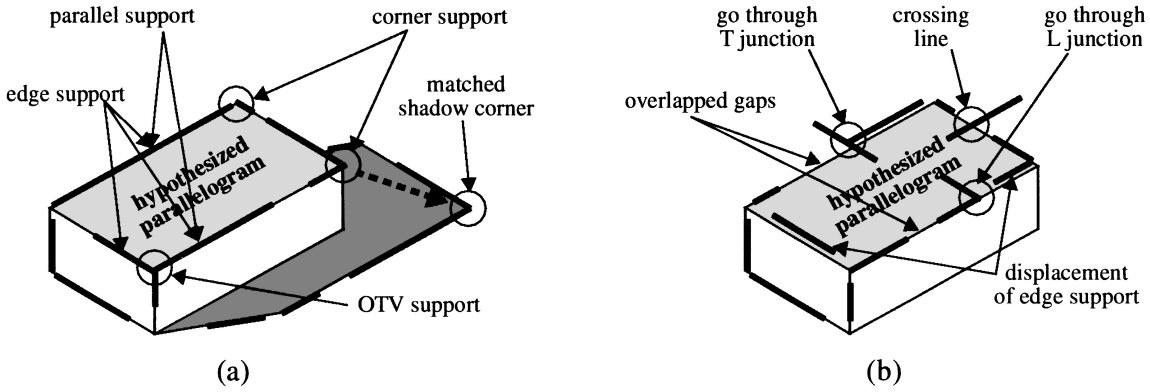


FIG. 10. (a) Positive evidence, (b) negative evidence.

such as lines, corners, and their spatial relations. A score for each parallelogram hypotheses is computed by using all local selection criteria and only those exceeding a threshold are retained for global selection. Global selection criteria evaluate relationships such as containment and overlap among the remaining hypotheses, which allows some of them to be eliminated.

3.1. Local Selection

The local selection criteria are derived from both ***positive evidence*** and ***negative evidence*** of the existence of a roof. The positive evidence includes the presence of edges, corners, parallels, OTVs and matched shadow corners (see Fig. 10a). The negative evidence includesthe presence of lines crossing any side of a parallelogram, existence of L or T junctions in any side of a parallelogram, existence of overlapped gaps on opposite sides of a parallelogram, and displacement between a side of a parallelogram and its corresponding edge support (see Fig. 10b).

Each local evidence, such as edge support, provides a numerical score. The score is typically based on how much of an expected feature is actually observed. The local evidence are multiplied by a weighting factor and added to yield an overall score. The weighting factor is determined by the expected importance of that evidence; corners, for example, are considered more important than edges. Hypotheses with an overall score below a threshold are eliminated. This method of combining local evidence is quite simple but we have found it to be very effective and not highly sensitive to the choice of parameters. All of our examples have been tested with the same parameter setting. Some specific measurement criteria and weights are given below (see also Table 1).

Positive evidence. • Edge support is measured by the percentage of edge coverage along the perimeter of the hypothesized parallelogram.

• Corner support is measured by how close the edge support is to the corner of hypothesized parallelogram.

• Parallel support is an approximate measurement of distribution of edge support on parallel sides of the parallelogram.

• Potential OTV support is computed in oblique views and its strength depends on the distances from the edge support of the OTV to the corresponding corner on the parallelogram.

• Matched potential shadow corner support is proportional to the number of matches from the potential shadow casting corners on the hypothesized parallelogram to the potential shadow corners extracted from the image.

Negative evidence. • Displacement of edge support is measured by how close the edge support is to the perimeter of the hypothesized parallelogram.

• Overlap of parallel gaps is measured by the percentage of the length of overlapping gaps over the perimeter of the parallelogram.

• The effect of T junctions, L junctions, and crossing lines in any side of a hypothesized parallelogram is measured by the relativelengths and strengths of the crossing edge, the angle between the crossing lines, and the length of the gap near the crossing.

TABLE 1
Coefficients of Weighted Sum

Evidence	Range of returned value	Normalized weight
Positive Evidence		
Edge support	[0, 1]	0.60
Corner support	[0, 1]	0.25
Parallel support	[0, 1]	0.15
Potential OTV support	[0, 1]	0.30
Matched shadow corner	[0, 1]	0.15
Negative evidence		
Displacement of edge	[0, 1]	−0.50
Overlap of gaps	[0, 1]	−0.50
Through junctions	$[0, n]$, $n =$ number of junctions	−2.00
Crossing lines	$[0, 2n]$, $n =$ number of crossing lines	−2.00
Local selection threshold = 0.55		

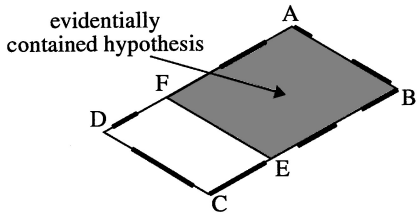


FIG. 11. Evidentially contained hypothesis.

3.2. Global Selection

Parallelograms surviving local selection may overlap with each other. If they do, they are considered to be in competition and a selection is made among them where possible. However, the goal is not to remove overlap completely at this stage. Our method uses the following three global selection criteria:

1. Duplicated hypotheses. It is possible for hypothesis formation to generate parallelograms that cover the same or almost the same area as the search can be triggered in several different ways. In this case, the selection criterion retains the one with the best evaluation score from the local selection process.

2. Evidentially contained hypotheses. If a hypothesis is contained in another, and the supporting evidence of the contained is completely covered by the supporting evidence of the containing hypothesis, we say it is an *evidentially contained hypothesis* and eliminate it. Evidentially contained hypotheses are likely to correspond to a part of a roof structure and the containing hypothesis is likely to represent a more complete roof structure. Figure 11 shows an example where parallelogram (ABEF) is evidentially contained in parallelogram (ABCD).

3. Containment analysis. When a hypothesis but not its entire supporting evidence is contained in another, we need to decide whether the containing hypothesis should be preserved. The containing hypothesis is likely to cover an area beyond a roof; however, the contained hypothesis could just be a superstructure on the base formed by the containing hypotheses. The criteria applied is that the score of the containing hypotheses is higher than the normal threshold by an amount determined by the area of the largest contained hypotheses. Figure 12 shows a hypothesis, H1, that contains two other hypotheses, H2 and H3. The relative selection standard for H1 is proportional to the ratio of the area of H2 (the bigger of H2 and H3) over the area of H1.

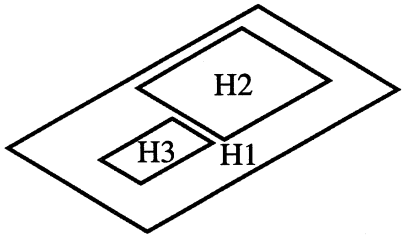


FIG. 12. Containment situation.

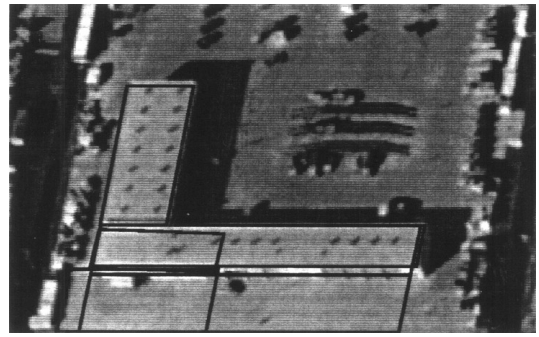


FIG. 13. Selected hypotheses.

The situation where one hypothesis overlaps with another but neither is contained by the others is deferred for resolution until after the verification stage when a more informed decision can be made. Figure 13 shows the selected hypotheses (6) after all the local global selection criteria have been applied to the hypotheses in Fig. 9.

4. VERIFICATION OF HYPOTHESES AND INFERENCE OF 3-D SHAPE

The purpose of verification is to test whether the selected hypotheses correspond to 3-D structures. We use the cues of whether the hypothesized roof casts a shadow and whether there are one or more visible walls associated with it. This process also provides the 3-D information required to create the 3-D model of the structure. A block diagram of the verification process is shown in Fig. 14. First, wall and shadow evidence for a hypothesis are collected at several possible building heights and the height at which the best combined evidence value is obtained is selected. Those hypotheses which have strong enough supporting evidence are analyzed to resolve the ambiguity of containment or overlap situations among them. Then a reasoning

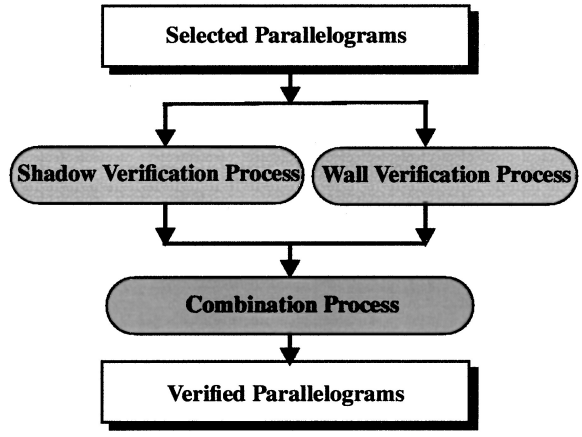


FIG. 14. Verification process.

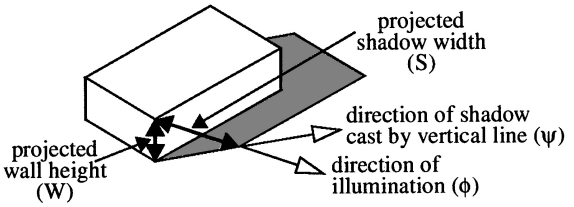


FIG. 15. Wall height and shadow width.

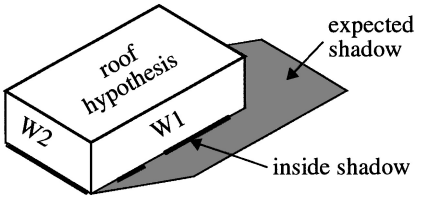


FIG. 17. Evaluation of wall ground evidence.

process is used to analyze spatial interactions, such as occlusion which may help recover some of the occluded hypotheses.

The building height is related to several parameters, shown in Fig. 15, that can be measured from the image features. Let the image resolution be R (pixels/meter) in the neighborhood of the building location. The projected wall height, W (in pixels), can be computed from the building height, H (in meters), and the viewing angles by

$$W = H \cdot R \cdot \sin\gamma. \tag{2}$$

The projected shadow width, S, can be computed from the building height, the viewing angles 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) can be computed from

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

4.1. Wall Verification Process

The wall verification process attempts to find wall evidence for a given building height. Given the viewing angles and a possible building height, the visible sides can be determined and the expected wall boundary can be computed. The verification process collects evidence along this projected boundary. With the knowledge of the minimum and maximum heights of buildings, the search for wall evidence is limited to a certain range. Our method searches the entire range in 1 meter steps (as shown in Fig. 16); an alternative would be to sample the range coarsely first and use these values to perform a finer search.

The following evidence is collected for each hypothesized height:

Ground evidence. Figure 17 shows schematical evidence visible for two walls, W1 and W2. The ground evidence is evaluated according to the percentage of the ground boundary covered by edge evidence and the displacement of the edge evidence. The weight is decided by the length of the ground boundary, the vertical distance between the roof boundary and the ground boundary of the wall, and whether the wall is inside the shadow or not. Ground boundaries that are close to the roof boundary or in shadow are given a lower weight as they tend to be less reliable.

Vertical evidence. The vertical wall evidence is a weighted sum of the score of each vertical wall boundary. Figure 18 shows three vertical wall boundaries. The score of a vertical wall boundary depends on the coverage percentage of edge evidence, the displacement of the edge evidence (from the expected boundary), and the distribution of the edge evidence (shorter gaps are considered better). The weight of a vertical wall boundary is determined by the angle between the boundary and the adjacent roof boundary (lower angle has a smaller weight), whether or not the vertical boundary is inside the expected shadow (boundaries inside a shadow have lower weight). If the expected vertical boundary length is sufficiently large and not sufficient edge evidence is found for it, a negative value is given to this evidence.

Corner evidence. We evaluate the corners formed by the vertical evidence and ground evidence of a wall, and between

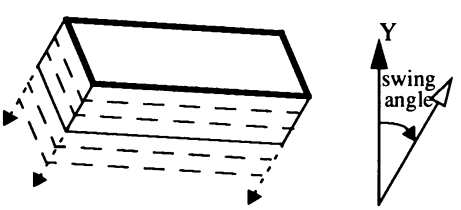


FIG. 16. Search for wall evidence.

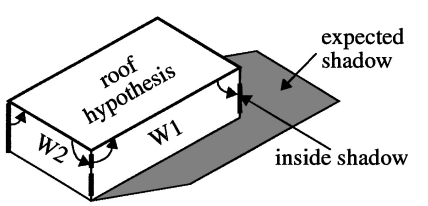


FIG. 18. Evaluation of wall vertical evidence.

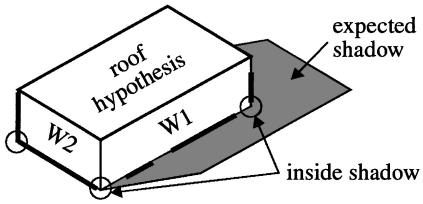


FIG. 19. Evaluation of wall corner evidence.

ground evidence of adjacent walls, possibly by extending them. The score of the corner evidence is the weighted sum of the scores of all corners. A higher weight is given to a corner with more edge support and when it is not in shadow. Figure 19 shows the corner evidence of the walls of a roof hypothesis.

The overall wall evidence for a given hypothesis p at the building height H is given by a weighted sum of its components as given in Eq. (4), where k_i is the score for the ith component of the wall evidence and v_i is the corresponding weight. $W(p, H)$ is normalized to be between -1 and 1.

$$W(p, H) = \sum_i v_i \cdot k_i(p, H). \tag{4}$$

The ranges and values for the three components are given in Table 2.

4.2. Shadow Verification Process

The shadow verification process tries to establish correspondences between shadow casting elements and shadows cast, assuming 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 in establishing shadow correspondences. The shadow intensity cannot be relied on to be uniformly dark. Building sides are usually surrounded by a variety of objects such as loading ramps and docks, grassy areas and sidewalks, trees and shrubs, vehicles, and light and dark areas of various materials. The shadow may be occluded by the building itself or by nearby buildings, particularly in oblique views. To deal with these problems we use 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.

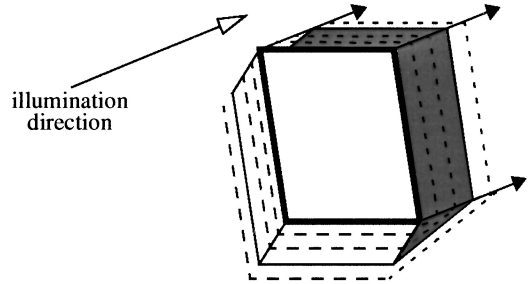


FIG. 20. Search for shadow evidence.

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 can be computed by Eq. (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. A set of corresponding shadow evidence is collected for each height (with 1 meter steps) within the allowed range; Figure 20 illustrates this schematically. We find that this is more reliable than inferring 3-D directly from potential shadow evidence, as it can be quite fragmented.

The shadow evaluation consists of the following components:

Shadow lines cast by a roof. A weighted sum of the scores of all shadow lines cast by a roof is computed. The score of a shadow line cast by a roof boundary is evaluated according to the coverage percentage of edge evidence and the displacement of the edge evidence on the visible part of the shadow line. Figure 21 shows the edge evidence of the shadow lines cast by roof boundaries, B1 and B2; note that a part of the shadow line cast by roof boundary, B2, is occluded by the building itself. The weight of each shadow line is based on the length of the visible part of the shadow line and the distance between the shadow line and the nearest ground boundary or roof boundary of the building.

TABLE 2
Wall Evidence Values and Weights

Wall Evidence	Range of returned value	Normalized weight
Ground evidence	[0, 1]	0.30
Vertical evidence	[−1.33, 1]	0.30
Corner evidence	[0, 1]	0.40

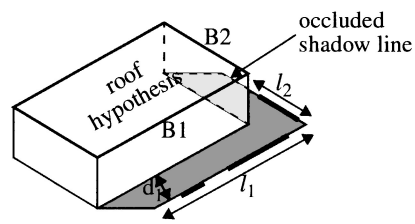


FIG. 21. Evidence of shadow lines cast by a roof.

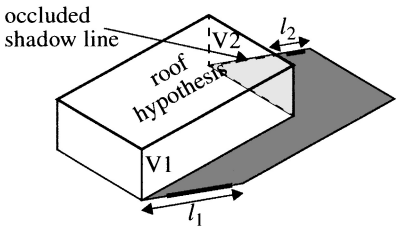


FIG. 22. Evidence of shadow lines cast by a wall.

Shadow lines cast by vertical lines. This evaluation function computes the weighted sum of the scores of all shadow lines cast by vertical lines of the walls. The score of each shadow line is measured by the coverage percentage of edge evidence, the displacement of the edge evidence, and the distribution of the edge evidence on the visible part of the shadow line, as in the case of wall evidence. The evidence of two shadow lines cast by vertical lines, V1 and V2, of walls is shown in Fig. 22. If the length of the visible part of the shadow line is very long and very little edge evidence is found, it is considered negative evidence and a negative score is returned. The score of each shadow line is weighted according to the length of the visible part of the shadow line.

Shadow junctions. A weighted sum of the scores of all visible shadow junctions, is calculated by this evaluation function. Figure 23 shows the evidence of two visible shadow junctions and a shadow junction occluded by the building itself. The score for each junction depends on how close the lines forming the junction are to the junction and by what is the shadow junction formed. A "medium" shadow junction is formed by a shadow line cast by the roof and another shadow line cast by a vertical line, which indicates a match between one side of the roof and the shadow junction. A "strong" shadow junction is formed by the shadow lines cast by the roof, which indicates a match between two sides of the roof and the shadow junction, so a higher weight is assigned to a strong shadow junction.

Shadow region statistics. The intensity information inside the visible shadow region is collected and analyzed by this evaluation function. An approximate distribution of the intensity values of all the pixels in the visible shadow region is used to compute a score for the shadow region. This distribution is compared with an expected one; a big difference suggests a bad shadow region and a negative score is given. The expected distribution is computed from mean and variance values provided by the user. Even though the distribution varies over the image, we find that this measure still gives useful results.

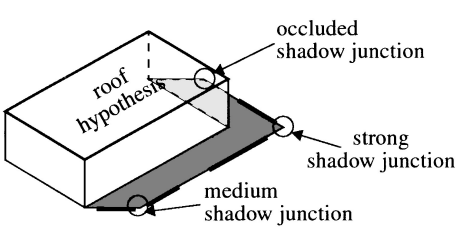


FIG. 23. Evidence of shadow junctions.

TABLE 3
Shadow Evidence Values and Weights

Shadow evidence	Range of returned value	Normalized weight
Roof shadow lines	[0, 1]	0.18
Vertical shadow lines	[−1, 1]	0.24
Shadow junctions	[0, 1]	0.30
Shadow line statistics	[−1, 1]	0.10
Shadow region statistics	[−2, 1]	0.18

An overall score for shadow evidence is computed by a linear weighted sum of the components as given by Eq. (5) for a hypothesis p at the building height H, where h_i is an evaluation function for the ith shadow evidence and u_i is the corresponding weight. $S(p, H)$ is normalized to be between −1 and 1.

$$S(p, H) = \sum_i u_i \cdot h_i(p, H). \tag{5}$$

The range and weights of the different components of the shadow evidence terms are given in Table 3.

4.3. 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 these scores are combined by Eq. (6), where $C = WS(p, H)$ is the combined score, C_1 is $W(p, H)$ and C_2 is $S(p, H)$.

$$C = \begin{cases} C_1 + C_2 - C_1 \times C_2 & \text{when } C_1, C_2 \geq 0 \\ C_1 + C_2 + C_1 \times C_2 & \text{when } C_1, C_2 < 0 \\ \dfrac{C_1 + C_2}{1 - min(|C_1|, |C_2|)} & \text{otherwise.} \end{cases} \tag{6}$$

This combination rule follows the *certainty factor* method used in MYCIN [27] and other systems. It is accurate if the wall and shadow evidence are *conditionally* independent (i.e., independent given the hypothesis). 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 as shown in

$$WS_p = WS(p, H_p) = \text{Max } WS(p, H) \tag{7}$$

$$\text{where} \begin{cases} H_p & : \text{estimated height for hypothesis } p \\ WS_p & : \text{confidence value of hypothesis } p. \end{cases}$$

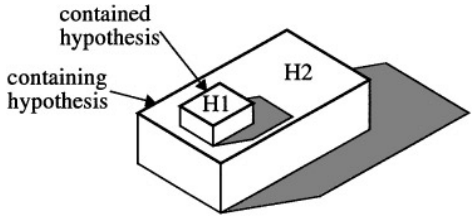


FIG. 24. Super structure relationship.

The wall-shadow confidence value, WS_p is again combined with roof evidence, say R_p, by the same method as in Eq. (6) above and gives the total confidence value C_p for the hypothesis. If this values is greater than a given threshold value, the hypothesis is considered verified. R_p is scaled, before combination, so that the roof evidence alone cannot verify a hypothesis.

5. 3-D ANALYSIS

The verified hypotheses have 3-D information associated with them; the following reasoning takes advantage of this information. The resulting hypotheses may overlap with one another or be contained in another as the earlier verification processes examine each hypothesis individually and not the relationships among them. 3-D analysis now allows us to make further choices among structures that may describe parts of the same 3-D space.

5.1. Containment Analysis

First, consider the case where a contained hypothesis does not share any side with the containing hypothesis. In this case, the contained hypothesis is considered to be a super structure of the containing hypothesis. If all supporting evidence of the contained hypothesis is inside the roof of the containing hypothesis, its height is adjusted to be relative to that of the containing roof. Figure 24 shows an example.

Now, consider the case where the two hypotheses share some common boundaries. If the two have different heights, we consider them to be in conflict and we remove the one with lower confidence. If they have the same height and share boundaries as in Fig. 25, where hypothesis (ABEF) is contained by hypothesis (ABCD) and they share three sides, a more complex analysis becomes necessary. The larger structure is not necessarily the desired one, it may contain elements other than those belonging to a building. It is selected only if there is sufficient evidence for the *nonshared* parts of the roof boundaries.

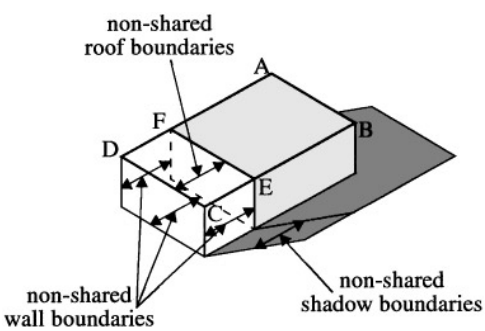


FIG. 25. Containment analysis.

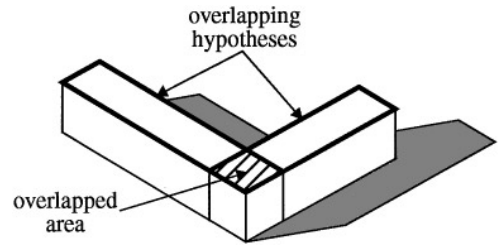


FIG. 26. Overlapping hypotheses with the same height.

5.2. Overlap Analysis

The idea of the overlap analysis is to examine the evidence of the overlapped part of the hypotheses and decide to which hypothesis the overlapped part belongs. The other hypothesis is deleted or modified according to the evidence of the nonoverlapped part of the hypothesis. If the two overlapping hypotheses have the same estimated building heights, they are taken to be two parts of a complex shape building, as shown in Fig. 26, and both are retained.

When two roof hypotheses with different building heights overlap, they conflict in 3-D space. It is possible for two building hypotheses to have overlapping footprints even if the roof hypotheses do not overlap as shown in Fig. 27. In our method, if the overlap is relatively large, the structure with the lower confidence is removed, else both are retained. A more sophisticated analysis is required to find the best combination, if there are several overlapping structures; we have not implemented this step.

Figure 28 shows the two hypotheses retained after the wall and shadow verification process for the image shown in Fig. 13. The upper part of the structure is verified because it has a clear shadow boundary. The lower part of the structure has fragmented wall and shadow boundaries, but the system is able to spot the small pieces of evidence and verify it. Since there is no containment or overlap between them, both hypotheses are retained. The confidence value of the upper part of the structure is 0.61,

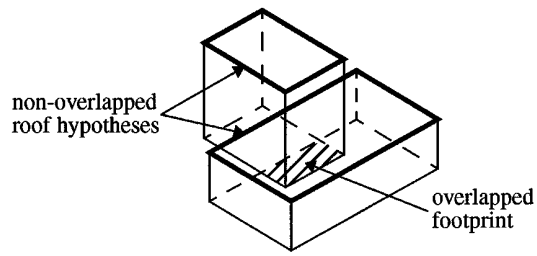


FIG. 27. Overlapped footprint.

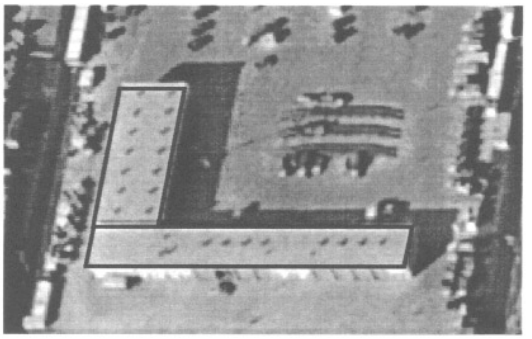


FIG. 28. Verified hypotheses.

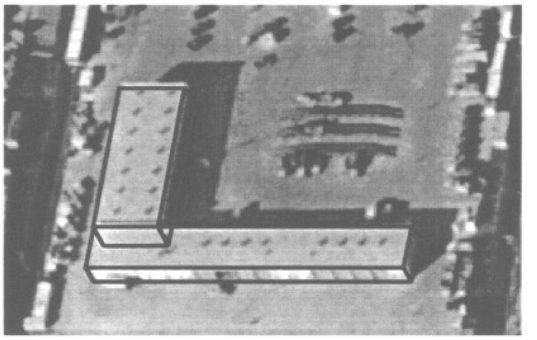


FIG. 30. 3-D wire frame.

and the confidence value of the lower part of the structure is 0.67. Note that there are some mutual occlusions between them. The following interaction analysis can detect and handle the mutual occlusion situation and increase the confidence of both hypotheses.

5.3. Building Interaction Analysis

The wall and shadow verification processes assume that the evidence of a building is not occluded by other buildings. When such occlusions are present, some correct hypotheses may be rejected due to insufficient visible evidence. Analysis of spatial interaction between various hypotheses can help recover some of these. We reevaluate all *selected* (but not verified) hypotheses by evaluating only the evidence on the parts of the wall and shadow not occluded by the already verified hypotheses. A newly verified hypothesis could change the confidence of its surrounding hypotheses, so the process is iterated until no more new hypotheses are verified. The confidence of a hypothesis is computed as the combination of wall and shadow scores, which are evaluated by the same evaluation functions used previously except that they consider only that evidence in the nonoccluded parts of the wall and shadow boundaries.

Figure 29 shows two hypotheses, H1 and H2. A part of shadow evidence of hypothesis H2 is not visible because of the existence of hypothesis H1. On the other hand, a part of wall evidence of hypothesis H1 is blocked by hypothesis H2. In this case, say there is enough evidence to support hypothesis H2, and H2 is retained after the wall and shadow verification process. However, a large part of the wall and shadow of H1 is occluded by H2, so it might be rejected by the earlier verification process. Knowing that H2 has been verified, the interaction analysis process can make a better evaluation on H1 by examining the evidence in the nonoccluded area and verify it. Once H1 is verified, the process can go back to reassess the supporting evidence of H2 and increase the confidence of H2 as well.

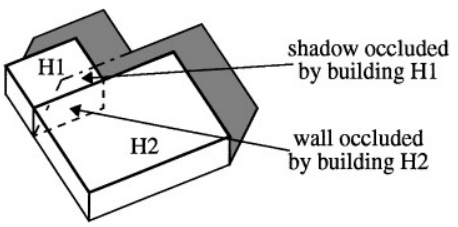


FIG. 29. Evidence occluded by other buildings.

There may be cases where none of the hypotheses involved in a severe occlusion situation have enough confidence to be retained by the wall and shadow verification process. A more complex interaction analysis that considers certain combinations is needed; we have not implemented this.

Figure 30 shows the 3-D wire frames of the two verified hypotheses shown in Fig. 28. There are some occlusions between these two structures. The confidence of the upper part of the structure is 0.61 and the confidence of the lower part of the structure is 0.67 before the interaction analysis. After the interaction analysis, the confidences of the upper and lower parts of the L-shape building have been increased to 0.68 and 0.88, respectively. Obviously, the visible evidence of the lower part is better than the upper part, because it has better wall evidence. With the interaction analysis, the confidence calculated by the system is more consistent with our observation.

Figure 31 shows an example where two of the three structures of the building in the scene are detected initially. Parts of the wall and shadow boundaries of the structure on the left are occluded by the structure in the middle, and therefore the left structure cannot be verified without consideration of the context. After the occlusion analysis, our system recovers the left structure by examining the situations of mutual occlusions. Note that the confidence of the middle structure is increased also, because it is occluded partly by the right structure.

5.4. 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 are used to generate a 3-D model of the scene(see Fig. 30). The textures inside the roofs and visible walls of verified buildings are painted onto the corresponding surfaces in the 3-D model. The textures of the ground surface in the input image are painted onto the ground surface of the 3-D model also.

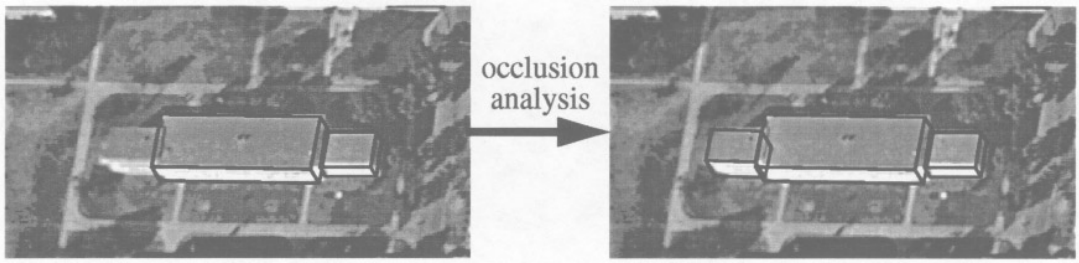


FIG. 31. An example of occlusion analysis.

This 3-D 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 (see Fig. 32).

6. INTEGRATION OF RESULTS FROM MULTIPLE VIEWS

The described method can also be used to integrate results of analysis from multiple views. If more than one view of a scene is available, the system can integrate the results from multiple views. The evidence of a building may be more clear in one view than in the others, depending on several conditions, such as the viewing direction, the illumination direction, and the building orientation. The traditional approach to use multiple images is to match features at a low level and use the results to hypothesize higher level structures. It has many advantages, but matching of low-level features can be highly ambiguous. If a system has the abilities to recover good high-level descriptions from each image, matching at the high-level becomes viable and is much simpler; this is the approach described below.

The basic approach is to project hypotheses of one view into the other views and verify them in all views. If a building is correctly detected in one view, supporting evidence for it should be found in other views. On the other hand, if an incorrect hypothesis has been made, it should be unlikely to find much supporting evidence from other views. Based on this observation, a better decision can be made by integrating all evidence from all available views.

Given that the relative camera geometries are known, we project the 3-D wire frame of a verified hypothesis in one view into another view. All evidence around the projected wire frame of the verified hypothesis in the second view is collected and then the evaluation function of the hypothesis verification process is applied on the collected evidence to compute the confidence of the hypothesis in this view. The confidence values of a hypothesis in all views are combined using the certainty factor method shown earlier by Eq. (6). It is possible that a verified hypothesis in one view has negative confidence in another view. A threshold value depending on the number of views is set to remove those unsatisfactory hypotheses (threshold is 0.45 for two views vs 0.5 for a single view).

A building could be verified individually in more than one viewresulting in multiple hypotheses for the same structure. An overlap analysis is performed and the hypothesis with the highest combined confidence among the overlapping hypotheses is retained. A set of 3-D models is created from the list of retained hypotheses, which can be projected into any view for visualization.

Our algorithm does not handle the case when none of the hypotheses of a building from any of the views is correct. Here, one would need to fuse the evidence by using parts of the hypotheses

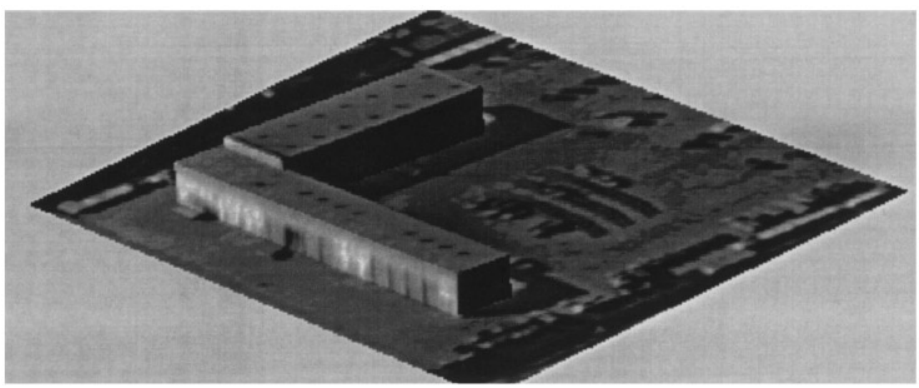


FIG. 32. Rendered image from another viewpoint.

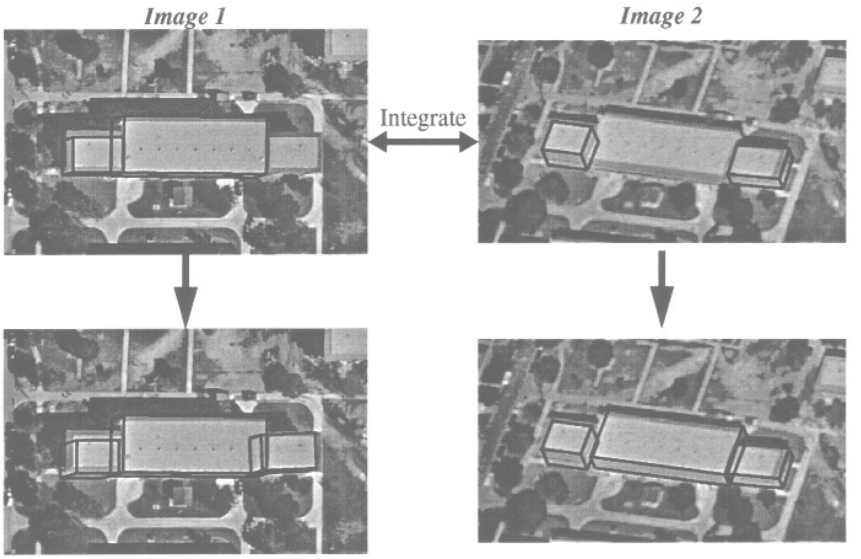


FIG. 33. Integration of results from multiple views.

from different views and create a more accurate composite. Some part of the evidence of a building will be stronger in one view than the other depending on the situations, such as viewpoint direction and illumination direction.

Figure 33 shows an example of integrating the results from two views of a building. The building is composed of three structures. The main structure in the middle is detected in the left image only, the right wing is detected in the right image only and the left wing is detected in both. After the integration process, all three parts of the building are verified (one of the two choices for the left wing is selected). Projections of the resulting 3-D models are shown at the bottom of Fig. 33.

7. RESULTS AND EVALUATION

This system has been tested on a number of examples with good results. Some results are shown in Section 7.1 and sources of some problems are outlined. Typical execution times are given in Section 7.2; detection evaluation is provided in Section 7.3, use of confidence values is discussed in Section 7.4. In Section 7.5, we discuss the issue of the parameter selection for various decision steps in this system.

7.1. Examples

Figure 34 shows the results of several examples on small windows. Figures 34a and 34b show two L-shape buildings. Note that parts of the shadows fall on the nearby vehicles. Although this makes the shadow boundaries highly fragmented, the system still successfully locates the correct shadow boundaries. Also, in Fig. 34b the building is dark and the 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. In Figs. 34c and 34d, note that there are some rectangular shape surface markings on the ground. These do result in roof hypotheses but no shadow or wall evidence is found for them. The building in Fig. 34e is composed of three parts. The part on the lower right corner has a different height from the other two parts. Although a hypothesis is made for this part, no confirming 3-D evidence can be found (humans find it difficult to estimate its height as well). There are four gabled-roof buildings in Fig. 34f. This system

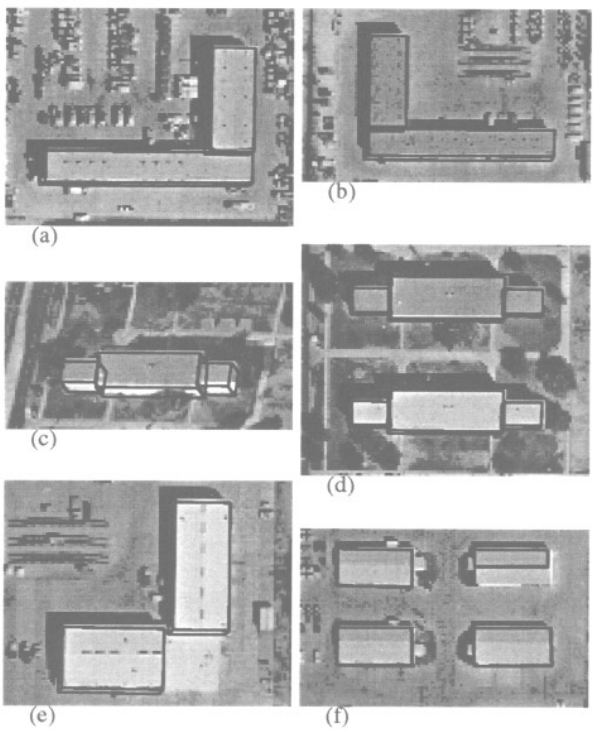


FIG. 34. Small window examples.

does not currently model gabled 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 but only half of the roof is detected. Figure 35 shows a larger window from the Ft. Hood site and the results at various stages of processing. This image is taken from an oblique view and represents a difficult case as the building's shapes are rather complex, many roof boundaries are occluded by vegetation and there is significant texture on the ground. Figure 35b shows the line segments extracted from the image, Fig. 35c shows the parallelogram hypotheses, Fig. 35d shows the selected hypotheses, and Fig. 35e shows the verified hypotheses. The wire frames of the resulting buildings are shown in Fig. 35f. Some quantitative data for this example are given later in Section 7.2.

Figure 36 shows an expanded view of the final results for the same example. The estimation of the building heights is quite accurate. Only one of the verified hypothesis has an obvious error of the building height. No false alarm is verified by the system. Two buildings are not detected (ignoring the complex building at the right edge, which we consider outside the range of our system). One building is near the bottom-left corner of the image; it is not detected because of severe occlusions by nearby trees. The other is the white building with two structures located on the right side of the C-shaped building in the middle. The mutual occlusions between the two structures of the white building prevent either one of them to be verified. The two C-shaped buildings are detected but the descriptions are not accurate. The middle parts of the C-shaped buildings are not hypothesized, because there is no other evidence besides a pair of parallel lines. The occlusion analysis process recovers a structure for the building in the top-middle area of this image. There is another structure in this building that is not detected because of severe occlusions. There are also some structures attached to the four buildings on the left side of this image that are not detected. The system does not generate hypotheses for some of them due to the broken roof boundaries caused by the surrounding trees and buildings.

Figure 37 shows the results on the same area as in Fig. 36 but from an image taken from a different viewpoint. There are 14 buildings in this window (counting all connected parts as one.) Four of these are not detected; three of these are rather small and the fourth has one of its long sides heavily occluded. There is one false alarm at the top left but the size of the falsely detected structure is quite small. Most of the detected buildings are described accurately. The middle part of the middle C-shaped building is missing and the L-shaped structures attached to the four buildings on the left-hand side of this window are not correctly described, because shadows fall on them, their shadows are not clear, and the visible part of the walls is very small.

Figure 39 shows the results on four selected windows in a large image. The two windows on the right are shown in detail in Figs. 37 and 39. Figure 39 shows the details of the area in the upper right corner of the image in Fig. 38. There are 14 buildings in this area; 11 of these are detected correctly. One false alarm is detected (at the right hand side of the left most L-shape building). It comes from a truck next to a dark region which appears similar to a small building.

Figure 40 shows the result of integrating the outputs from the two views shown in Figs. 36 and 37 (the results shown are from the viewpoint of Fig. 36 but can be projected from any viewpoint as we have now constructed 3-D models). Note that the integrated results are more complete and better than for each of the individual views. The one false alarm in Fig. 37 survives because it is small and small pieces of edge evidence can be found for it in the second view.

7.2. *Execution Time Evaluation*

Table 4 shows some quantitative data for the example of Fig. 35. It takes 1314.5 s (21 min and 54.5 s) to process this image on a SUN Sparcstation 10 (using the RCDE environment [29] with all code being written in COMMONLISP). The most time consuming process, at 75.4% of the total, is that of parallelogram formation. The total run time for the hypothesis generation process, which includes segment folding, junction detection, colinearization, and parallelogram formation, is 1116.9 s. Note that the "higher level" processes of hypothesis selection and verification take only a small fraction of the total time. This example is one of the more complex we process. The execution times are generally linearly proportional to the number of lines that are found in an image.

7.3. *Detection Evaluation*

There are many ways to measure the quality of the results of an automatic system. We use the following five measures:

- Detection Percentage $= 100 \times \mathrm{TP}/(\mathrm{TP} + \mathrm{TN})$.
- Branch Factor $= 100 \times \mathrm{FP}/(\mathrm{TP} + \mathrm{FP})$.
- Correct Building Pixels Percentage.
- Incorrect Building Pixels Percentage.
- Correct Nonbuilding 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 a person and the program, FP (False Positive) is a building detected by the program but not a person, and TN (True Negative) is a building detected by a person but not the program. A building is considered detected if *any* part of the building is detected; an alternative could be to require that a certain fraction of the building be detected. These measures are similar to those suggested in [7, 28], but we use a different definition of branching factor and compute the metrics on a building basis rather than a pixel/voxel basis.

The accuracy of shape is determined by counting correct building and nonbuilding pixels. We calculate the percentage of the number of pixels correctly labeled as building pixels over the number of building pixels in the image, the percentage of the number of pixels incorrectly labeled as building pixels over the number of pixels labeled as building pixels, and the percentage

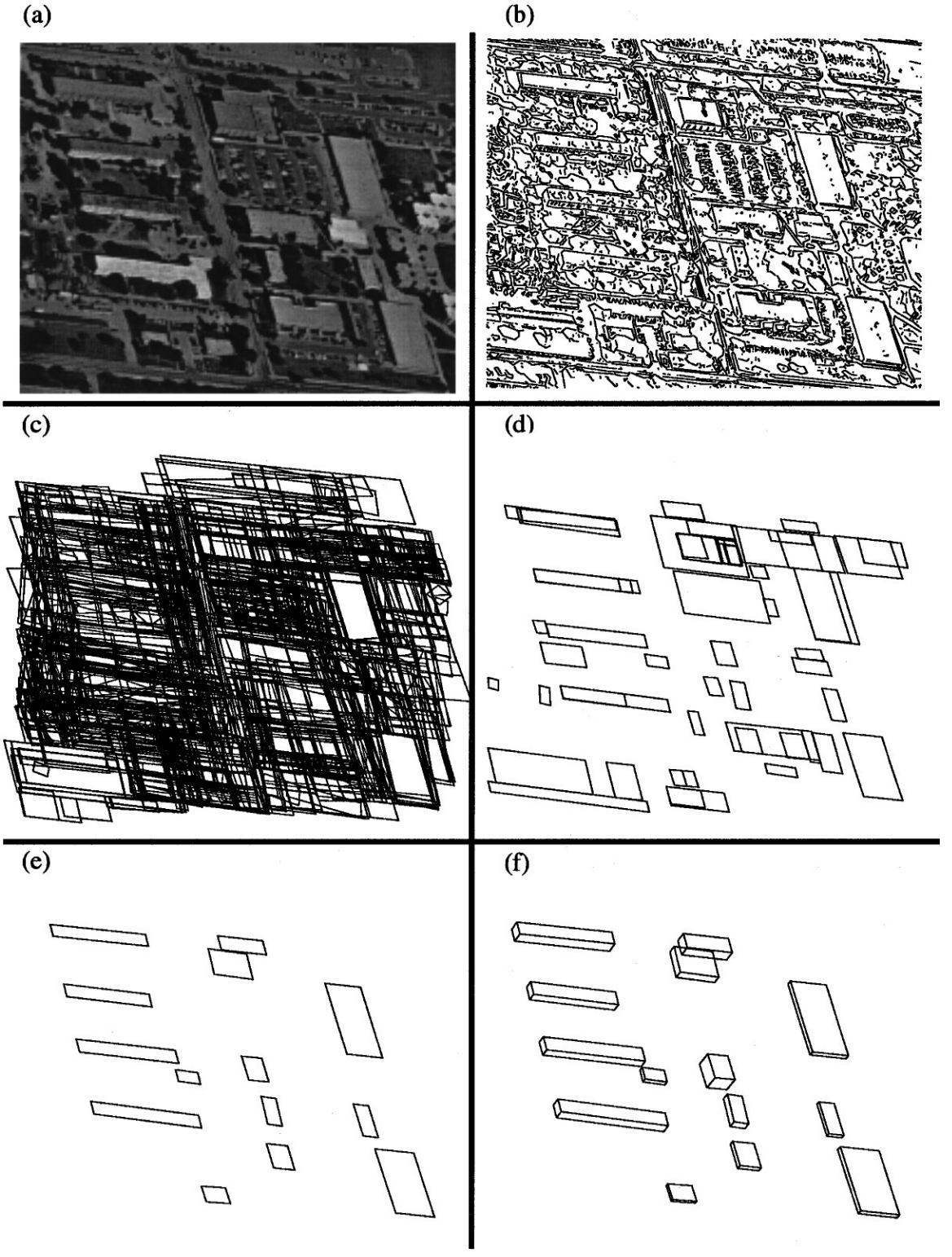
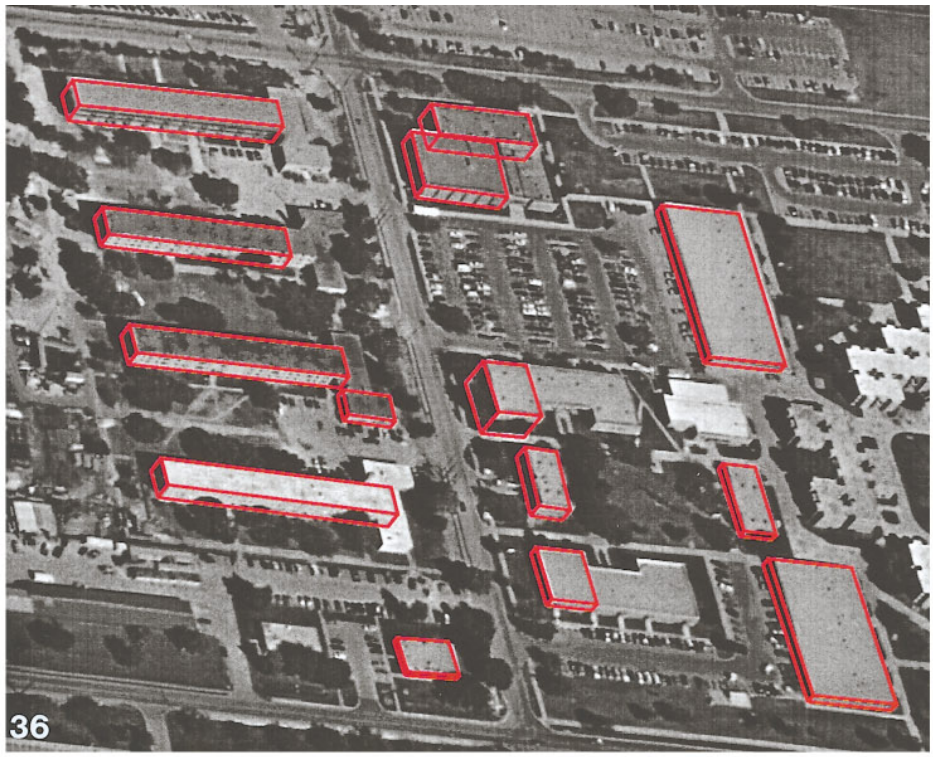


FIG. 35. Results at various processing stages for a complex image.



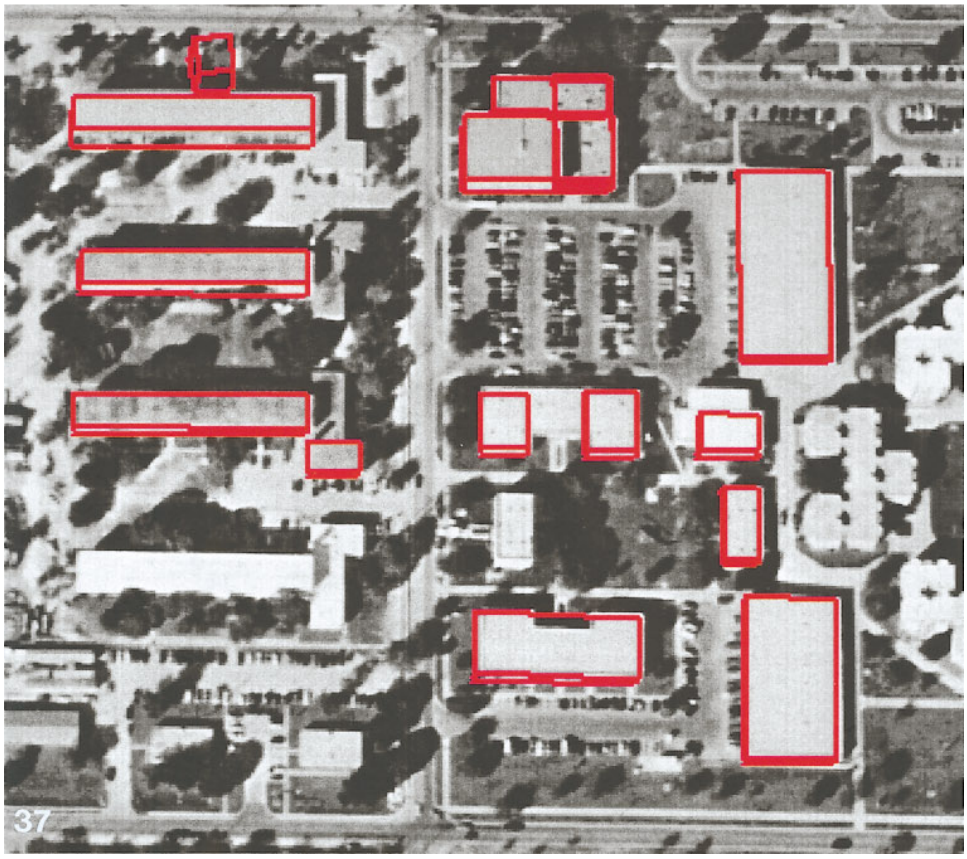


FIG. 36. Complex image final results.

FIG. 37. Another view of the scene shown in Fig. 36.

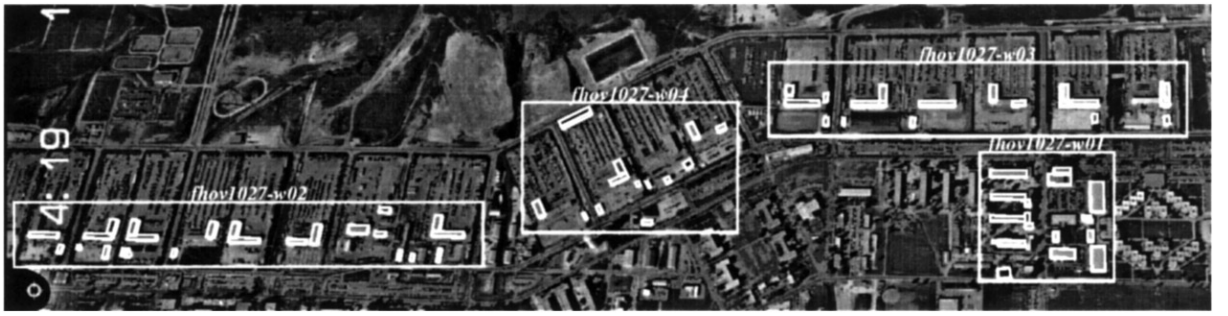


FIG. 38. Results in four windows of a large image.

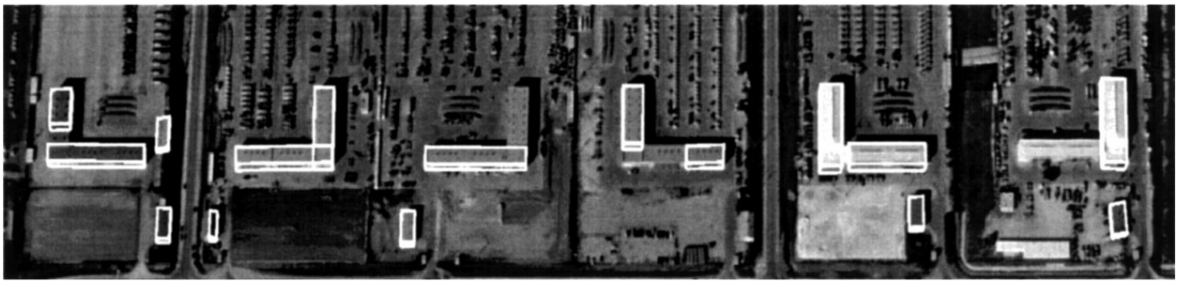


FIG. 39. A detailed view of the upper right window in Fig. 38.

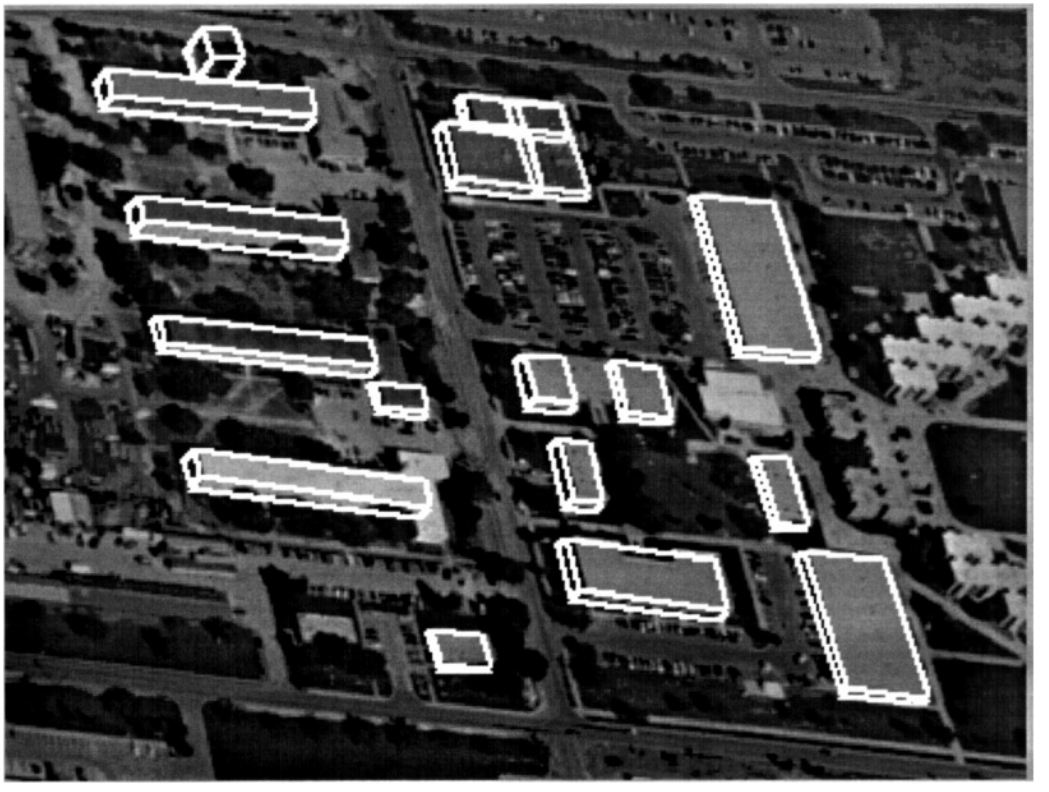


FIG. 40. Combination of results from Figs. 36 and 37.

TABLE 4
Quantitative Analysis for Fig. 35 Results

Steps	*Number of features*	*Run time (seconds)*	*Percentage of time*
Edge detection	7159	102.0	7.8%
Segment folding	3605	22.7	1.7%
Junction detection	2249	44.6	3.4%
Breaking and colinearization	8803	58.0	4.4%
Parallelogram formation	1890	991.6	75.4%
Hypothesis selection	58	48.9	3.7%
Hypothesis verification	14	46.7	3.6%
Total		1314.5	100.0%

of the number of pixels correctly labeled as nonbuilding pixels over the number of nonbuilding pixels in the image (the latter can be expected to be high as most pixels in an image are nonbuilding pixels).

Table 5 shows the evaluation on the results of our system on four image windows shown in Fig. 38. Note that our system gives rather consistent results for most images except for window 4, an area where the orientation of the L-shaped buildings in the image is almost parallel to the direction of illumination and the other orientation of the L-shape buildings is also almost parallel to the projection of the vertical line. Therefore, only one side of the roof casts a shadow and only one side of the wall is visible; it is difficult for the verification process to find enough evidence to verify these L-shape buildings. Also note that the gray level of three of the L-shape buildings is very similar to the gray level of the surrounding ground and this makes the roof boundaries of these buildings very fragmented.

7.4. 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 41 shows a histogram of the number of true and false positives of 12 (10 of which have been discussed in Section 7.1) 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.

TABLE 5
Detection Evaluation

	Detection percentage tp/(tp + tn)	**Branch factor fp/(tp + fp)**	**Correct building pixels**	**Incorrect building pixels**	**Correct nonbuilding pixels**
Window1	71.4%	0.09%	78.8%	8.71%	98.6%
Window2	72.0%	0.00%	74.4%	1.45%	99.9%
Window3	78.6%	8.33%	78.4%	0.81%	99.9%
Window4	64.3%	18.18%	42.7%	29.71%	99.1%
Average	71.9%	6.65%	71.7%	7.40%	99.5%

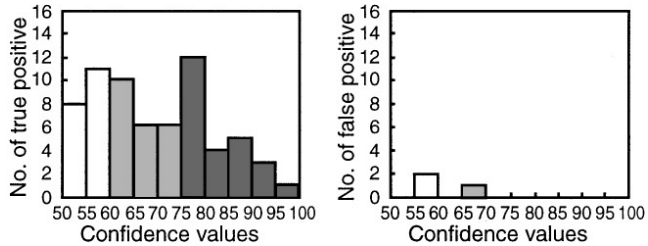


FIG. 41. Distribution of confidence values.

Note that there are only three false positives and they all have low confidence values. In fact, if we set a confidence threshold of 70, 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.

7.5. Choice of Parameters

Our system needs to make decisions based on incomplete evidence at various stages, such as in selection and verification. This is done typically by using evaluation functions which compute scores for individual pieces of evidence and then combine them in some ways (usually by a linear weighted sum). One can question whether each of these steps is computed in an optimal way and whether the results are combined in optimal ways. Due to the complexity of the processes involved, and lack of formal models for the contents of an image, we are unable to find theoretically optimal solutions. For each stage, we have tried to use what reasoning we could. Some decisions are fairly intuitive; for example, the score given to edge evidence for the side of a roof is given by the length of the actual edge divided by the length of the roof side (of course, one could argue that linear proportionality may not be optimal). The rules for combining them have been chosen mostly on the basis of simplicity. The weights given to different evidence have been determined empirically.

Our system has been tested on a large number of examples, in our laboratory as well as at other institutions. All examples are run with all but a few of the parameters being fixed. Some of the parameters are functions of image resolution, this information must be supplied by the user or be part of the image data. We also require a camera model and a sun model (computatble from geographical location, time, and date). The user also supplies minimum and maximum dimensions of acceptable building size, otherwise, default parameters are used (minimum length and width are set to 8 meters, maximum length and width are set to 225 meters, minimum height is set to 3 meters, maximum height to 30 meters). The only significant parameter for a user to choose is the confidence level at which a building description should be output (or displayed). As we have indicated earlier, a

trade-off is involved here between increasing detection rate and reducing the false alarm rate. A default value is used when no explicit input from a user is provided.

In an independent project, we are investigating alternative means of combining the individual evidence and of learning the needed parameters automatically. We are examining alternatives, such as the use of decision trees and simplified Bayes reasoning. Early results indicate that our initial choices are comparable to these alternatives. In future work, we intend to explore more rigorous methods such as Bayes Networks, and we believe that our representation methodology makes it feasible to do so.

8. CONCLUSIONS

We have presented a system for detecting buildings from a single intensity image, although it is also capable of integrating results from several such images. The current system is limited to detecting rectilinear buildings with flat roofs. We believe that such shapes cover a significant fraction of the building types, particularly in industrial and military areas. The roof types can be generalized by computing projected constraints for various shapes; we have derived the constraints for peaked roofs but have not implemented them. Our grouping methodology can also be generalized to other *known* shapes; working with an arbitrary and unknown set of shapes would require more general grouping techniques.

An interactive editing tool that can be used to efficiently correct the errors of this system has also been developed. In this method, simply pointing to a building or correcting a corner, for example, can result in complete recovery of a missed or incorrect building. This is made possible by preserving the hierarchy of descriptions that the automatic system computes and correcting only the parts that are indicated by a user. This approach is further described in [23].

Our system has been ported to some industrial and U.S. Government laboratories for possible use in current applications.

ACKNOWLEDGMENTS

The authors thank Andres Huertas for his help in the preparation of this paper and for providing systems software support for our research.

REFERENCES

1. M. Herman and T. Kanade, Incremental reconstruction of 3D scenes from multiple, complex images, *Artificial Intell.* **30**(3), 1986, 289–341.
2. A. Huertas and R. Nevatia, Detecting buildings in aerial images, *Comput. Vision Graphics Image Process.* **41**(2), 1988, 131–152.
3. R. Irvin and D. McKeown, Methods for exploiting the relationship between buildings and their shadows in aerial imagery, *IEEE Trans. Systems Man Cybernet.* **19**(6), 1989, 1564–1575.
4. C. Jaynes, F. Stolle, and R. Collins, Task driven perceptual organization for extraction of rooftop polygons, in *Proceedings, ARPA Image Understanding Workshop, 1994*, pp. 359–365.
5. C. Lin, A. Huertas, and R. Nevatia, Detection of buildings using perceptual grouping and shadows, in *Proceedings, IEEE Computer Vision and Pattern Recognition Conference 1994*, pp. 62–69.
6. C. Lin, A. Huertas, and R. Nevatia, Detection of buildings from monocular images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 125–134, Birkhauser Verlag, Basel, 1995.
7. J. McGlone and J. Shufelt, Projective and object space geometry for monocular building extraction, in *Proceedings, IEEE Computer Vision and Pattern Recognition Conference, 1994*, pp. 54–61.
8. R. Mohan and R. Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Anal. Machine Intell.* **11**(11), 1989, 1121–1139.
9. M. Roux and D. M. McKeown, Feature matching for building extraction from multiple views, in *Proceedings, IEEE Computer Vision and Pattern Recognition Conference, 1994*, pp. 46–53.
10. V. Venkateswar and R. Chellappa, A framework for interpretation of aerial images, in *Proceedings, International Conference on Pattern Recognition, June 1990*, pp. 204–206.
11. M. Berthod, L. Gabet, G Giraudon, and J. Lotti, High resolution stereo for the detection of buildings, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 135–144, Birkhauser Verlag, Basel, 1995.
12. O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller, 3-D reconstruction of urban scenes from sequences of images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 145–168, Birkhauser Verlag, Basel, 1995.
13. R. Collins, A. Hanson, E. Riseman, and H. Schultz, Automatic extraction of buildings from aerial images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 169–178, Birkhauser Verlag, Basel, 1995.
14. M. Gruber, M. Pasko, and F. Leberl, Geometric versus texture detail in 3-D models of real world buildings, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 189–198, Birkhauser Verlag, Basel, 1995.
15. E. Baltsavias, S. Mason, and D. Stallmann, Use of DTMs/DSMs and orthoimages to support building extraction, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 199–210, Birkhauser Verlag, Basel, 1995.
16. W. Forstner, Mid-level vision processes for automatic building extraction, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 145–168, Birkhauser Verlag, Basel, 1995.
17. N. Haala and M. Hahn, Data fusion for the detection and reconstruction of buildings, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 211–220, Birkhauser Verlag, Basel, 1995.
18. T. Kim and J. Muller, Building extraction and verification from spaceborne and aerial imagery using image understanding data fusion techniques, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 221–230, Birkhauser Verlag, Basel, 1995.
19. O. Jamet, O. Dissard, and S. Airault, Building extraction from stereo pairs of aerial images: Accuracy and productivity constraint of a topographic production line, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 231–240, Birkhauser Verlag, Basel, 1995.
20. P. Fua, Model-based optimization: Accurate and consistent site modeling, in *Proceedings, 18th ISPRS Congress, Vienna, Austria, 1996*, Comm. III, WG 2, pp. 222–233.
21. O. Henricsson, F. Bignone, W. Willuhn, F. Ade, O. Kubler, E. Baltsavias, S. Mason, and A. Grun, Project AMOBE: Strategies, current status and future work, in *Proceedings, 18th ISPRS Congress, Vienna, Austria, 1996*, Comm. III, WG 2, pp. 321–330.

22. U. Weidner, An approach to building extraction from digital surface models, in *Proceedings of the 18th ISPRS Congress, Vienna, Austria, 1996*, Comm. III, WG 2, pp. 924–929.
23. S. Heuel and R. Nevatia, Including interaction in an automated modeling system, in *Proceedings, IEEE Symposium on Computer Vision, Coral Gables, FL, Nov. 1995*, pp. 383–388.
24. C. Lin and R. Nevatia, Building detection and description from monocular aerial images, in *Proceedings, DARPA Image Understanding Workshop, Palm Springs, CA, Feb. 1996*, pp. 461–468.
25. J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Machine Intell*. **8**(6), 1986, 679–698.
26. R. Nevatia and R. Babu, Linear feature extraction and description, *Comput. Vision Graphics Image Process*. **13**, 257–269.
27. B. Buchanan and E. Shortliffe (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison–Wesley, Reading, MA, 1984.
28. J. Shufelt and D. McKeown, Fusion of monocular cues to detect man-made structures in aerial imagery, *Comput. Vision Graphics Image Process*. **57**(3), 1993, 307–330.
29. Strat, *et al*., The RADIUS common development environment, in *Proceedings, DARPA Image Understanding Workshop, 1992*, pp. 215–226.