

# Detection and Modeling of Buildings from Multiple Aerial Images

**S. Noronha\* and R. Nevatia\*\***

Institute for Robotics and Intelligent Systems  
University of Southern California  
Los Angeles, California 90089-0273  
{nevatia}@iris.usc.edu  
Telephone: (213)740-6428  
Fax: (213)740-7877

## Abstract

Automatic detection and description of cultural features, such as buildings, from aerial images is becoming increasingly important for a number of applications. This task also offers an excellent domain for studying the general problems of scene segmentation, 3-D inference and shape description under highly challenging conditions. We describe a system that detects and constructs 3-D models for rectilinear buildings with either flat or symmetric gable roofs from multiple aerial images; the multiple images, however, need not be stereo pairs (*i.e.* they may be acquired at different times). Hypotheses for rectangular roof components are generated by grouping lines in the images hierarchically, the hypotheses are verified by searching for presence of predicted walls and shadows. The hypothesis generation process combines the tasks of hierarchical grouping with matching at successive stages. Overlap and containment relations between 3-D structures are analyzed to resolve conflicts. This system has been tested on a large number of real examples with good result, some of which, and their evaluation, are included in the paper.

---

\* S. Noronha is now with eLance Inc., Sunnyvale, CA.94086.

\*\* This research was supported by Defense Advanced Research Projects Agency under contract No. DACA76-97-K-0001, monitored by the U. S. Army Topographic Research Center, and a sub-grant from MURI grant No. F49620-95-1-0457 from Army Research Office awarded to Purdue University.

## 1 Introduction and Overview

Automatic detection and modeling of cultural features, particularly buildings, from aerial images is becoming of increasing practical importance. Traditional applications are those of cartography and photo-interpretation. Newer applications include mission planning, urban planning, placement of communications resources and virtual tours of cities and sites. Besides the application needs, building detection and description also provides an excellent domain to explore the problems of scene segmentation, 3-D recovery and shape descriptions in a rich, realistic and demanding environment.

In this paper, we describe a system that detects and constructs 3-D wireframe models of buildings by using multiple (two or more) panchromatic aerial images. The individual and relative camera models are assumed to be known (or derived by use of some photogrammetric techniques), however, the image pairs are not necessarily stereo pairs in the usual sense- the different images may be taken at different times and may have different resolutions. We assume that the buildings are rectilinear (compositions of rectangles) in shape and that the roofs are either flat (planar *and* parallel to the ground) or are “symmetric gables” (two symmetric slanted surfaces). Walls are assumed to be vertical and the approximate ground height is assumed to be known. Illumination is assumed to be directly from the sun whose position is computable from the imaging date and time and the known geometric location of the site. Shadows are assumed to fall on flat ground of known height near the buildings. Figure 1 illustrates two views of a small part of the Ft. Hood, Texas site that will be used to illustrate our processing steps.

These assumptions constrain the task of detecting and describing the buildings. However, many of the traditional problems of computer vision must still be overcome. The objects of interest need to be segmented from the background; this is particularly difficult in aerial images due to presence of natural texture of vegetation and variety of detail present on or near the objects of interest. The 3-D structure needs to be inferred from the 2-D images; feature correspondence is made difficult due to presence of many similar features nearby (such as parallel lines near building edges) and relatively low disparity ranges for the low buildings. Even though we deal with rectilinear shapes only, specific 3-D shapes need to be recovered, conflicting descriptions need to be resolved, and in case of non-rectangular shapes, the different rectangular components need to be combined. showing The line segments detected from images in Figure 1, shown in Figure 2, should help visualize some of the difficulties outlined above (the line segments shown are from

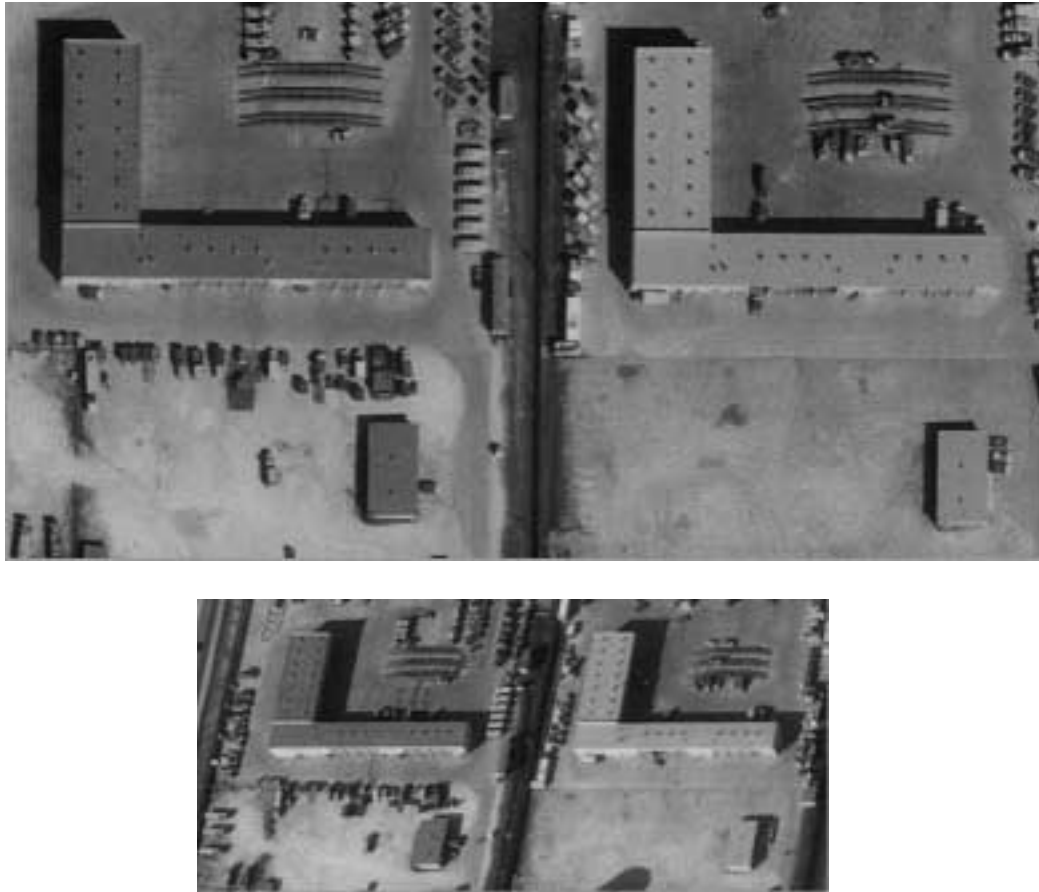


Figure 1 Two views of a portion of Fort Hood ( top image is 800x425 pixels with ground sample distance, GSD, of 0.3 meters, second is 480x220 with GSD of 0.6 meters)

Canny edges [1] after linking, linear approximation, colinearization and folding of nearby parallel segments).

The problem of object detection and description, of course, has been a central one in computer vision over many years. However, the general techniques developed are not powerful enough to handle the complexities inherent in aerial images of urban scenes. There has also been intensive research in systems designed explicitly for building detection and description; [2] contains a collection of some recent efforts.

Some of the approaches have attempted to work from a single image only ([3][4][5] [6] [7] [8]). Restriction to a single image makes the problem much more difficult as feature correspondences can not be used to infer 3-D and some ambiguities are harder to resolve in a single image; consequently, they are typically restricted to the case of flat roofed buildings only. Nonetheless, such systems can provide reasonable performance on relatively simple scenes and some of the techniques developed therein are useful for multiple image analysis as well.

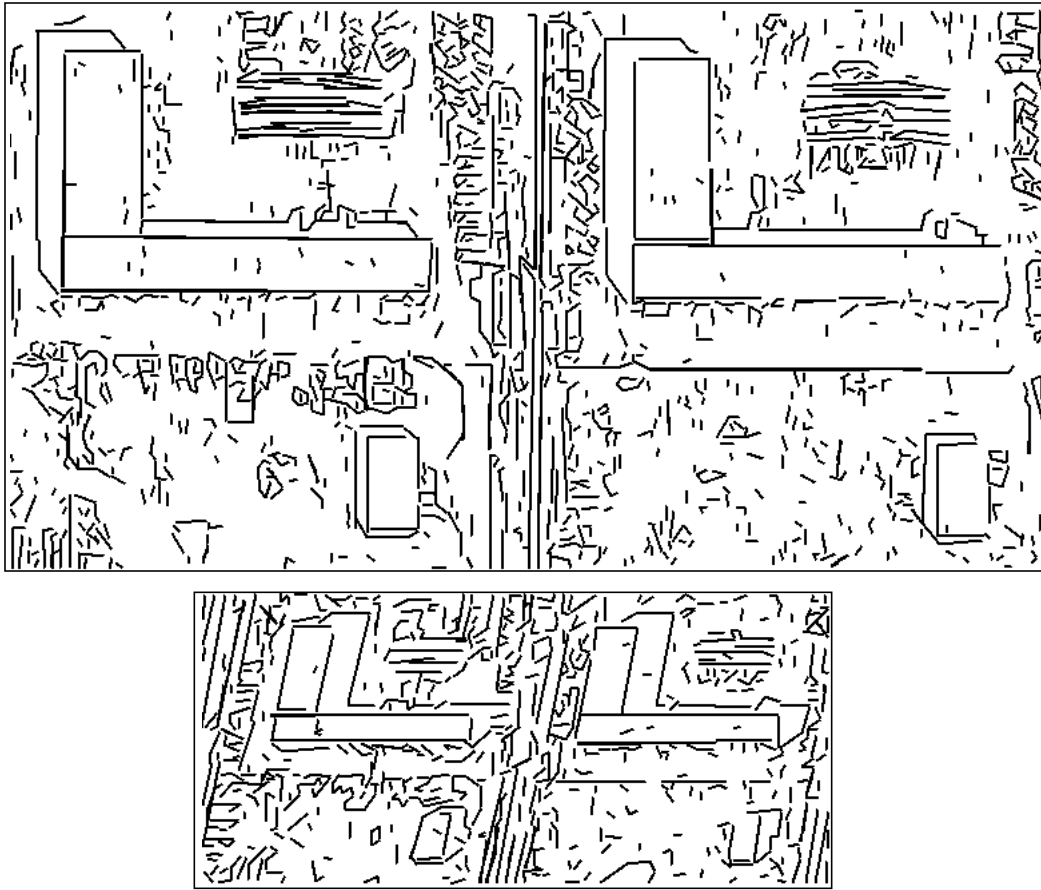


Figure 2 Linear segments extracted

Multiple images have also been used in previous work ([2][9][10][11][12][13][14]). Most of these systems assume that the images are from nadir views and taken at nearly the same time, which simplifies the task of matching features and even allows digital surface elevations to be computed by area correlation and interpolation methods. Jaynes et al. [11] describe a system that works with similar data to ours; it uses an approach to first find roof tops in a single view with the other views being used to determine the 3-D structure.

There has also been work on using data with other attributes, such as color [15], digital surface models derived from intensity images or range sensors directly ([16] [17] [18] [19] [20] [21][22]) or multi-spectral data which can significantly simplify the task of building detection and description ([23], [24]). Usually, digital surface models are derived from stereo pairs taken at nearly the same time; we assume that such models are not available for our conditions.

We provide an overview of our approach in the next section and the details in the subsequent sections. We believe that not only does our system show good results but it introduces several new features. The steps of matching and perceptual grouping are performed at each level before

proceeding to the next level. It uses a hypothesize and verify paradigm and delays decisions until more information is available to make them more reliably. The multiple views are used in an order independent way. Our approach also makes extensive use of monocular analysis even though multiple images are available. We believe that these concepts are also likely to be useful in detection and description of objects in a variety of other domains.

## **2 Overview**

We use a *hypothesize and verify* approach to building detection and description. Hypotheses for rectangular roof components are formed by grouping image lines in each view and verification is conducted by combining evidence in all the available views and by looking for evidence of visible walls and expected shadows. The perceptual grouping process used to generate the hypotheses is a hierarchical one: line pairs are grouped to form parallels or L-junctions which are then used to form U-shapes (i.e. three sides of a rectangle) which are finally used to generate rectangular roof components. The features are matched at each level in the hierarchy. Thus, the processes of grouping and matching are carried out simultaneously. At each matching stage, we allow multiple matches to be present, thus, a feature may appear in multiple grouping hypotheses at the next level. A more conventional approach to hypothesis formation may be to first match low level features, such as lines and junctions, and use only those (now 3-D) features in grouping, but this approach requires reliable matching of low-level features which is difficult without the context of higher level groups. Our approach, however, does allow use of limited low-level feature matching to guide the hypothesis formation process and thus reduce the number of hypotheses that would be generated by a purely monocular approach as in [6]. Figure 3 shows the block diagram of the system.

This approach shares some similarities with an earlier monocular system developed in our group [6]. The perceptual grouping process is similar, however, the multiple system can be more selective due to matching of features across images. The wall and shadow verification processes are also similar but simpler because height of a hypothesis is determined prior to the application of the verification steps. The monocular system is also unable to handle buildings with slanted roofs.

## **3 Generation of Hypotheses**

Hypotheses for building roof components are generated from the detected image features in a hierarchical fashion, matching features across views at each level in the feature hierarchy. All available views are used non-preferentially when forming hypothesis. A selection mechanism

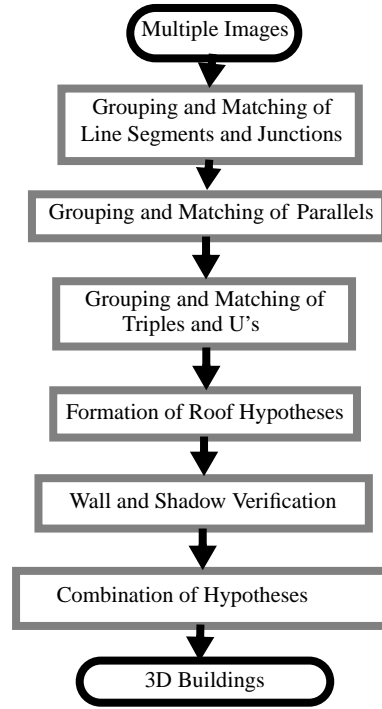


Figure 3 Block diagram of the automatic system

makes a value judgement of the evidence supporting each to eliminate those without sufficient evidence.

We first describe the feature hierarchy and the constraints applied at each stage in grouping and matching of the features. We then describes the method for hypothesizing and selecting flat-roof and gabled roof hypotheses in this order.

### 3.1 Grouping and Matching of Features

The feature hierarchy in order of complexity consists of lines, junctions, parallels and U-contours. These feature sets are then used to hypothesize flat and gable roofs.

#### Lines:

The lines are formed by linking of edgels and colinearization of line segments. Lines are matched by using the following *epipolar constraint*. The match for a line segment in one view must lie at least partially within a quadrilateral defined by the epipolar geometry and the 3D height constraints. Consider Figure 4. Let line  $l$  in  $view_1$  have endpoints  $p_1$  and  $p_2$ . By the epipolar constraint for points, the match for  $p_1$  in  $view_2$  must lie on an epipolar line defined by  $p_1$ , say  $l_1'$ . Similarly a match for  $p_2$  must lie on  $l_2'$ . A particular height in the world coordinate system corresponds to a particular point on the epipolar line. Hence, knowing the local ground height in the world coordinate system, and the maximum height of a building in 3-D, the search space can be reduced to the segment on each epipolar line defined by these z-coordinate values. In general,

there are 4 distinct points, two for each epipolar line. This limits the search for matching segments to a quadrilateral defined by these four points. In Figure 4 these points are denoted by  $p_{11}'$ ,  $p_{12}'$ ,  $p_{21}'$  and  $p_{22}'$ , corresponding to line  $l$ . Each pair of lines ( $l, l'$ ) that satisfy the epipolar (quadrilateral) constraint in any pair of views is determined to form a line match and included in the set of line matches that we will call  $S_{lm}$ , and is passed to the higher levels for further processing.

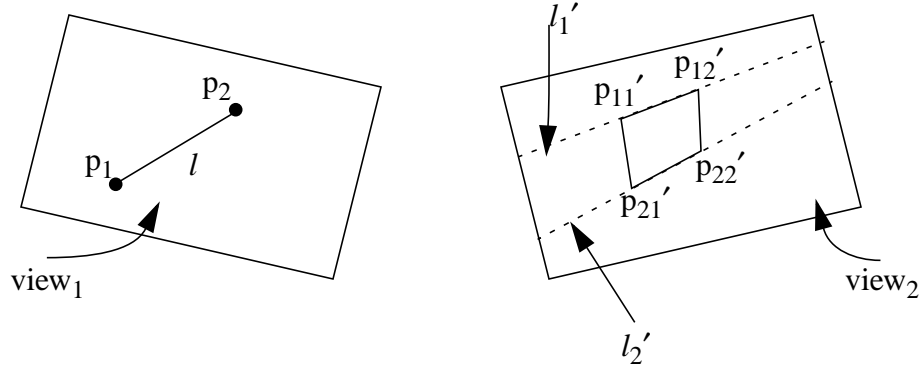


Figure 4 Quadrilateral (epipolar) constraint

Note that a line in one view may match multiple lines in other views, causing multiple matches (e.g. nearby vertical lines in Figure 2). We do not attempt to further disambiguate the matches at this level.

## Junctions

Junctions are formed by intersections of nearby line segments. Let the set of junctions formed in  $view_i$  be denoted by  $S_{J_i}$ . Junctions in the sets  $S_{J_i}$  ( $i = 1..n_{views}$ ) are then matched across the views, when the following constraints are satisfied:

- **Epipolar constraint:** The matching junctions in two views must be on corresponding epipolar lines and within a certain range (depending on the allowed height range in 3D).
- **Line Match Constraint:** The lines that form the two matching junctions must themselves match (*i.e.* the line pairs must be present in the line match set  $S_{lm}$ ).
- **Orthogonality Constraint:** Given a junction match, we can compute the 3D angle between the lines forming it (from the knowledge of the matching lines). This angle is required to be nearly orthogonal (between  $80^\circ$  and  $100^\circ$ ) in 3D as we are looking for rectilinear structures only.
- **Trinocular Constraint:** When there are more than 2 views available, the well-known trinocular constraint may be applied to the locations of the junctions. Functionally, given a point  $p_1$ , in  $view_1$ , and a point,  $p_2$  on the epipolar line of  $p_1$  in  $view_2$ , a point  $p_3$  in  $view_3$ , is determined uniquely (the intersection of the epipolar lines of  $p_1$  and  $p_2$  in  $view_3$ ).

Note that a junction in one view may match multiple junctions in other views, causing multiple junction matches for the same junction. Let  $S_{j_m}$  be the set of junction matches.

## Parallels

Starting with the lines in  $S_{lm}$ , parallel pairs of lines are detected in each the view and are matched across views. Parallels are formed between pairs of lines,  $L_{ij}$  and  $L_{ik}$  in the same  $view_i$ , when the following constraints are satisfied:

- Lines  $L_{ij}$  and  $L_{ik}$  are nearly parallel (make an acute angle of less than  $10^\circ$ ; ideally this angle should be a function of the line lengths and expected errors in point positions).
- The perpendicular distance between  $L_{ij}$  and  $L_{ik}$  is less than the maximum projected width of a building.
- At least 50% of  $L_{ij}$  overlaps with  $L_{ik}$  OR at least 50% of  $L_{ik}$  overlaps with  $L_{ij}$ .

The pair of lines  $(L_{ij}, L_{ik})$  forms a parallel. Along with the lines  $L_{ij}$  and  $L_{ik}$ , an abstract representation in the form of two line segments is stored. Denote these line segments by  $L_{ij}'$  and  $L_{ik}'$ . The angle of  $L_{ij}'$  and  $L_{ik}'$  is the average of the angles of  $L_{ij}$  and  $L_{ik}$  weighted by the lengths of  $L_{ij}$  and  $L_{ik}$ . If necessary,  $L_{ij}$  and  $L_{ik}$  are extended to form  $L_{ij}'$  and  $L_{ik}'$  so that they overlap completely (and are of same length in the image). Figure 5 illustrates the extension process. The representation  $(L_{ij}', L_{ik}')$  is used in the computation of parallel matches.

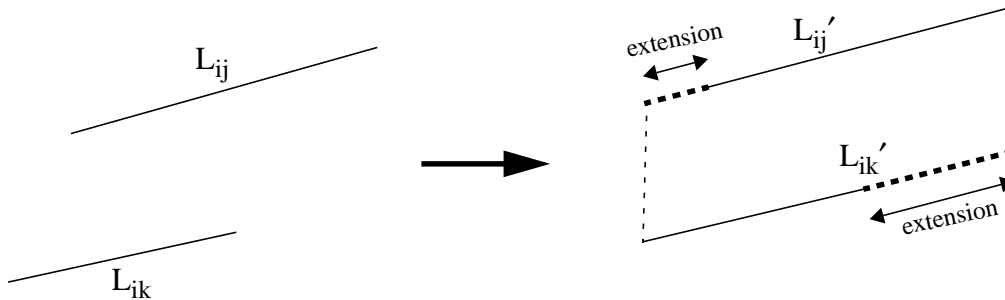


Figure 5 Extension of lines forming a parallel

While the task domain causes a large number of parallels in each view (two to three times the number of lines in that view), because of the alignment of buildings, roads, parking lots and shadows, the number of parallel matches is typically lower than the number of lines in any view. A parallel match is hypothesized if there is evidence in at least two views. When there are matching parallels in more than two views, a single parallel match (called a maximal parallel match) is formed across all the views. The constraint used in matching is the parallel match constraint described below:

- **Parallel match constraint:** Consider parallels  $P_{ik}$  with component segments  $L_{ik1}$  and  $L_{ik2}$  in  $view_i$ , and  $P_{jl}$  with component segments  $L_{jl1}$  and  $L_{jl2}$  in  $view_j$ . The parallel match constraint is satisfied for this pair of parallels if and only if exactly one of the following criteria is met:
  - $(L_{ik1}, L_{jl1})$  and  $(L_{ik2}, L_{jl2})$  are elements of  $S_{lm}$

- $(L_{ik2}, L_{jl1})$  and  $(L_{ik1}, L_{jl2})$  are elements of  $S_{lm}$

Parallel matches may be formed across two or more views. If there are a total of  $n$  views, parallel matches over the views are represented as  $n$ -tuples. The set of parallel matches is denoted by  $S_{pm}$ . Matches across more views dominate the matches across fewer views. Thus, if there are 4 views and  $(P_{11}, P_{21}, \text{nil}, \text{nil})$  and  $(P_{11}, \text{nil}, P_{31}, \text{nil})$  are detected as parallel matches, they are replaced by the maximal parallel match  $(P_{11}, P_{21}, P_{31}, \text{nil})$  in  $S_{pm}$ . Note that a parallel in one view may match multiple parallels in other views, causing multiple parallel matches for the same parallel.

### U-contours

U-contours correspond to three sides of a rectangle. Suppose a parallel in  $view_i$  is denoted by lines  $(l_{i1}, l_{i3})$ . If there exists a junction  $j$  in  $S_{ji}$  such that  $j$  may be denoted as either  $(l_{i1}, l_{i2})$  or  $(l_{i3}, l_{i2})$ , then  $(l_{i1}, l_{i2}, l_{i3})$  is a potential U-contour, say  $U$ .  $U$  is a valid U-contour iff a major part of  $l_{i2}$  lies between  $l_{i1}$  and  $l_{i3}$ .

## 3.2 Formation of flat-roof rectangular hypotheses

The flat-roof hypotheses are formed by using the feature sets computed above. Two methods are used: one that uses information from multiple images directly and another that uses information from a single image only.

### 3.2.1 Generation of hypotheses from multiple images

Let  $S_{pmj}$  denote a parallel match in the set of parallel matches  $S_{pm}$ . Denote the constituent parallels of  $S_{pmj}$  by  $P_{ij}$  where  $j$  is the image number. For each  $P_{ij}$ , a search is launched, in view  $i$ , to determine the best closure of the parallel match as follows:

- Starting from the center of the parallels in the parallel match, search for possible closure lines towards each end. A closure line is one that lies between the lines forming the parallel, and which forms an acute angle of less than  $10^\circ$  with the projection of the line orthogonal to the parallel in 3D and parallel to the ground plane in 3D.
- The search is also extended beyond ends of the parallel for lines that are elements of the set of matched lines,  $S_{lm}$ , and are thus guaranteed to have matches in at least one other view. We have found a search that extends to 25% of the length of the parallels to be effective.
- Detect possible closures, by scoring the coverage of the gap between the parallel lines. All closures covering greater than half the gap between the parallel lines are considered. The score of the closure is the ratio of the sum of lengths of the lines forming the closure to the distance between the ends of the parallels. Note that the search is performed simultaneously on all the images and hence the closures are matched as they are detected.
- Put qualifying closures into two sets: one for each half of the parallel match.

- Generate all combinations of closures from the two sets to generate 3-D hypotheses.

### 3.2.2 Generation of hypotheses from each image independently

Not all desired hypotheses are typically generated by the above method as some of the features are present (or detected) in only one view. To compensate for this, additional hypotheses are generated by using evidence that is partly visible in a single view only. 2-D parallelograms generated in each view from each U-contour (which come from the matched lines). For each 2-D parallelogram hypothesis a search determines if there is sufficient evidence for it in the other views. This search is performed iteratively through the space determined by epipolar geometry and by the 3-D height constraints on a hypothesis. When an acceptable match is found a 3-D hypothesis is created. The following describes the process in detail.

- For  $view_1$  let the set of U-contours be  $S_{ui}$ . Computation of U-contours is described in Section 3.1.
- For each U-contour  $U$  (say  $(l_1, l_2, l_3)$ ), in  $S_{ui}$ , search for line evidence to support the side that could complete a parallelogram. The direction of the search is the same as the orientation of the  $l_1$  (or  $l_3$ ). The search starts at a distance of 1/4th of the length  $l_1$  (or  $l_3$ ) from the ends of  $l_1$  (or  $l_3$ ) and proceeds the same distance beyond the  $l_1$  (or  $l_3$ ). Line segments that are approximately parallel to  $l_2$  i.e. which form an acute angle less than  $10^\circ$  with  $l_2$ , and which lie in the search space, are collected.
- Group clusters of these line segments into colinear lines if possible. Each of these clustered groups of lines causes a 2-D parallelogram hypothesis to be formed. Denote a representative hypothesis by  $p_u$ .
- For each 2-D parallelogram hypothesis search the other views for matching information. The aim of this search is to determine the best 3-D hypothesis by considering the hypothesis as a whole. The method varies the height parameter from the ground height to the maximum height of a building in small increments (1/10th of the range) and scores the line evidence for the hypothesis in all the views at each height. Each local maxima of the scores forms a 3-D hypotheses.

Figure 6 shows (in one view only) the combined hypotheses generated by the two methods from the lines shown in Figure 2. Note that multiple hypotheses are formed for several roof components as well as several other hypotheses are formed that correspond to no building. Next, a selection procedure is applied to filter out some of these hypotheses.

### 3.3 Selection of flat-roof rectangular hypotheses

It is possible to apply the verification procedure that uses all available evidence to choose among the generated hypotheses. However, this is a relatively expensive process. Instead, the set is first reduced in size by applying a simpler selection process that uses roof evidence only. The nature

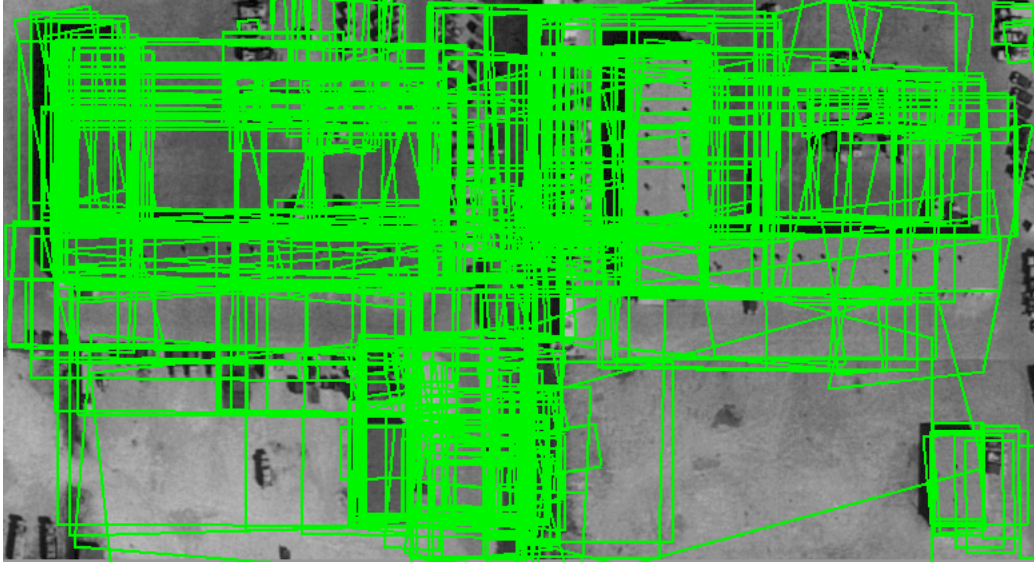


Figure 6 Hypotheses detected in images from Figure 1 (only one view is shown).

of the evidence used is discussed in Section 3.3.1 and method for using this evidence in Section 3.3.2.

### 3.3.1 Accumulating Selection Evidence

- Positive roof evidence:

Positive roof evidence consists of line segments in each view that support a hypothesis. As the hypothesis are 3-D models and camera parameters are known, the models can be projected in each available view. Consider an arbitrary view, say view<sub>*i*</sub>. Let the projected parallelogram of the 3-D hypothesis be  $(p_{i1}, p_{i2}, p_{i3}, p_{i4})$ . A spatially-indexed search is used to detect line segments, such that each line segment  $l_i$  satisfies the following criteria:

- $l_i$  must form a small acute angle (of less than  $10^\circ$ ) with one of the four sides of the parallelogram
- the perpendicular distance of the midpoint of  $l_i$  from any one of the sides of the parallelogram is small (less than 4 pixels).
- a major part ( $> 50\%$ ) of  $l_i$  overlaps with the side of the parallelogram it is closest to

The contribution of a line  $l$  in the set of positive evidence lines to the positive score is the ratio of the length that  $l$  overlaps with the nearest side of the roof hypothesis (in the image) to the perimeter. This automatically weights longer sides (and evidence supporting them) more than it does shorter sides.

- Negative roof evidence

Consider an arbitrary view,  $view_i$ . Let the projected parallelogram of the 3-D hypothesis be  $(p_{i1}, p_{i2}, p_{i3}, p_{i4})$ . A spatially-indexed search is used to detect line segments, such that each line segment  $l_i$  satisfies the following criteria:

- $l_i$  must intersect at least one of the four sides of the parallelogram
- $l_i$  must form a large acute angle (of greater than  $30^\circ$ ) with a side of the parallelogram that it intersects
- a major part ( $> 50\%$ ) of  $l_i$  overlaps with a side of the parallelogram it intersects

The contribution of a line  $l$  in the set of negative evidence lines to the negative score is the ratio of the length of  $l$  to the perimeter (in the image). This operation is performed on all the views available, and the evidence i.e. the line segments, is grouped by view for later evaluation. Figure 7 illustrates the concepts of positive and negative line evidence in one view for flat-roof hypothesis.

- Height inference

A new 3-D height is computed for each roof hypothesis based on optimal matching in all available views. Assume that the parallelogram corresponding to the projection of a 3-D hypothesis in  $view_i$  is represented by a set of 4 lines  $\{l_{ik}\}$  with  $k=1..4$ . The lines in the set  $\{l_{ik}\}$  are matched to lines  $\{l_{jl}\}$  in  $view_j$  subject to the quadrilateral constraint described in Section 3.1. As the height of the 3-D line that projects into  $l_{ik}$  is varied between the minimum and maximum heights, the projection of this line in  $view_j$ ,  $l_{ik}'$ , traverses the quadrilateral defined in the quadrilateral constraint. Parameterize this traversal with parameter  $s=0$  corresponding to the minimum height (possibly the ground plane) and  $s=1$  corresponding to the maximum height.

As parameter  $s$  is varied from 0 through 1,  $l_{ik}'$  will match different lines,  $\{l_{jl}\}$ , in the quadrilateral. This gives rise to multiple matches at different parameters (and hence different 3-D heights). Define a proximity measure for line  $l_{ik}$ ,  $d_{ijkl}$  as the perpendicular distance from the midpoint of  $l_{ik}$  to any line  $l_{jl}$  in  $\{l_{jl}\}$ . Define a measure of the goodness of the match for a set of lines  $\{l_{ik}\}$  in  $view_i$ ,  $m_s$ , at parameter value  $s$ , over all  $n_{views}$  views by

$$m_s = \sum_{j \neq ik, l} \sum e^{-\left(\sigma^2 \cdot d_{ijkl}^2\right)} \cdot len(l_{jl})$$

where  $\sigma$  is chosen to be 0.5 (hence  $\sigma^2 = 0.25$ ) and  $len(l_{jl})$  is the length of  $l_{jl}$ . The physical significance of this is to weigh the lines in  $\{l_{jl}\}$  by the distance from the projection of  $l_{ik}$  in  $view_j$  ( $l_{ik}'$ ) and have that weight drop off exponentially. This may be done with an arbitrary set of lines  $\{l_{ik}\}$  over any number of views from 1 to  $n_{views}$ . The value of  $s$  at the maximum of the function  $m_s$  is used as the best match of the set of lines  $\{l_{ik}\}$ . This value of  $s$  corresponds to the 3-D height at which the line support across views is the maximum.

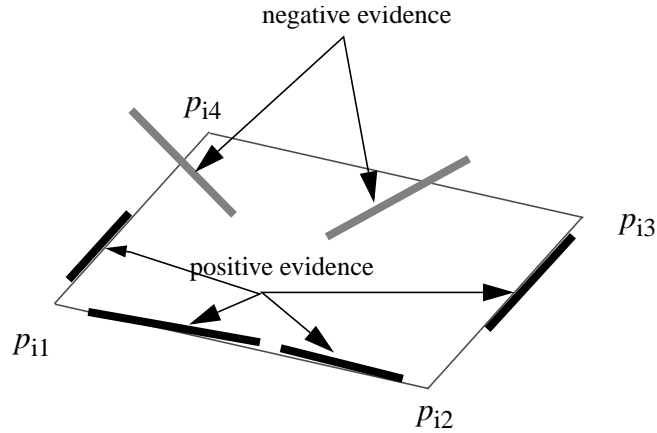


Figure 7 Positive and negative line evidence

### 3.3.2 The selection mechanism

Hypotheses that are not within an acceptable height range (typically between 2m and 20m above ground) are eliminated. Even though, some height constraints have been applied in earlier stages of hypothesis generation, the refined height estimate allow application of a more stringent criterion. Remaining hypotheses are examined for roof evidence.

Two roof quality measures are computed. One is the average positive line score,  $avg\_pos\_score$ , computed from the positive scores for each view (normalized to be between 0 and 1). The second is the average of the difference between positive and negative line scores, called  $avg\_diff\_score$ . A hypothesis is selected if and only if  $posRoofScore / n_{views} > 0.5$  and  $(posRoofScore - negRoofScore) / n_{views} > 0.3$ .

Figure 8 shows the selected hypothesis from Figure 6. Note that the number of hypothesis is greatly reduced. Most of the undesired hypotheses have been eliminated without rejecting the desired ones (though this may not always be the case).

### 3.4 Formation of gable-roof rectangular hypotheses

Figure 9 shows two views of an aerial scene containing symmetric gable-roof buildings. To generate hypotheses for such roofs, a *triple* of parallel lines is computed first (the triples are to correspond to the “ridge” or the top of the roof and the two sides).

#### Triples

Triples are derived from the set of parallels constructed earlier. All lines are considered, regardless of whether they also participate in flat roof hypotheses (overlaps are resolved as described in Section 5 ). Suppose that a parallel  $P_{ik}$  in  $view_i$  comprises line segments  $L_{ik1}$  and  $L_{ik2}$ . A search is launched for line segments  $L_{ik3}$  that lie either between  $L_{ik1}$  and  $L_{ik2}$ , or on the side of

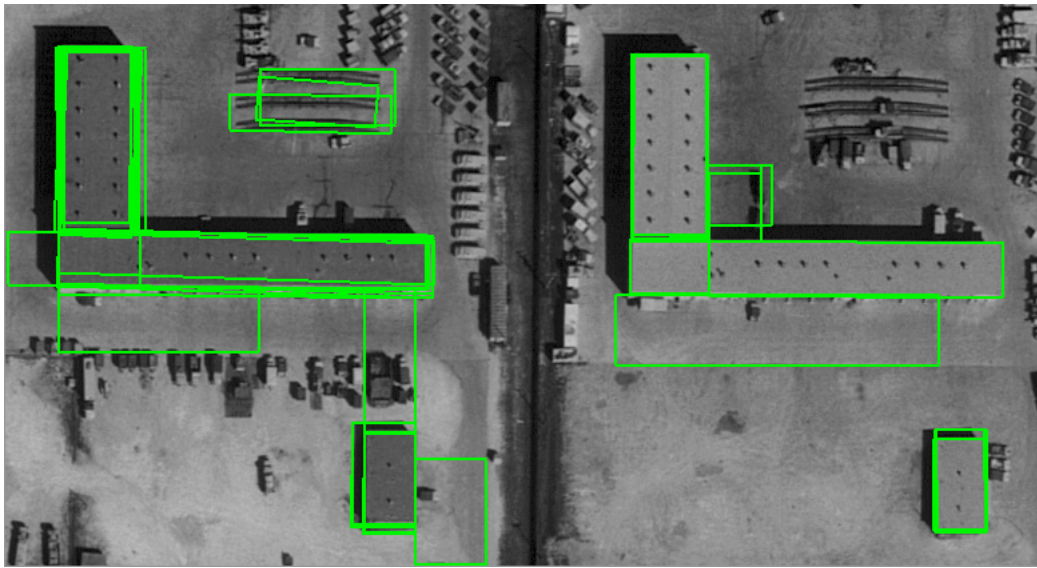


Figure 8 Selected hypotheses from the images in Figure 1

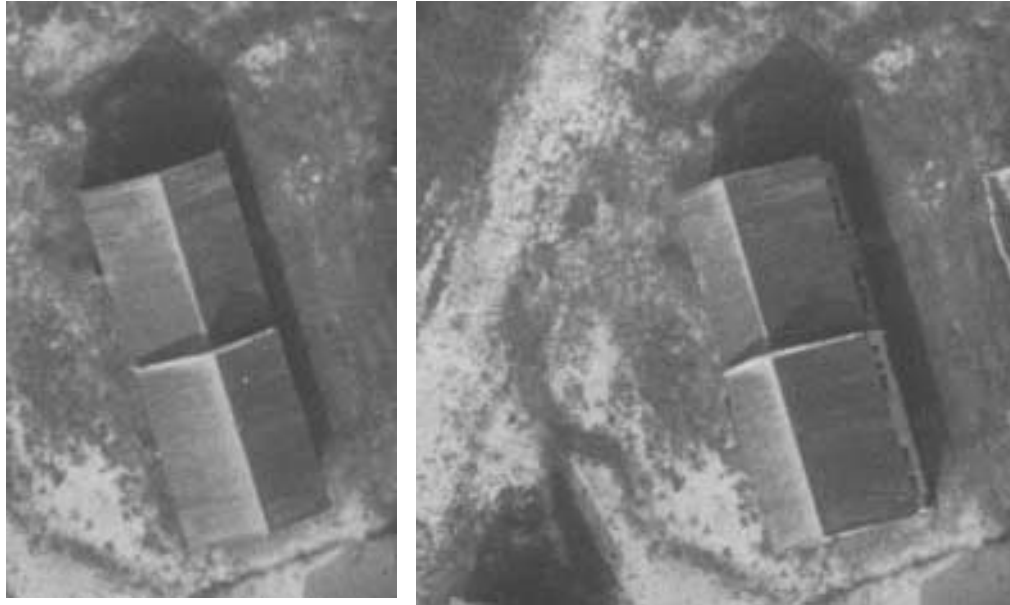


Figure 9 Two views of an aerial scene with symmetric gable-roof buildings. Left image is 155x240 pixels, right image is 240x240 pixels, both have GSD of 0.15 meters  $L_{ik1}$ , or on the side of  $L_{ik2}$ .  $L_{ik3}$  must have an orientation similar to  $L_{ik1}$  and  $L_{ik2}$  i.e. the acute angle between  $L_{ik3}$  and  $L_{ik1}$  (or  $L_{ik2}$ ) is less than  $10^\circ$ .

### 3.4.1 Generation of symmetric gable-roof hypotheses

The generation of a gable-roof hypothesis is initiated by a triple in any image. Triples are examined to see if they could give rise to valid 3-D hypotheses as follows: Given a camera model, and ground height, the relative height in 3-D of the “ridge” of the gable roof with respect to the sides in that view may be derived as follows. Assuming that the gable is symmetrical in 3-D, if the ridge and the sides were at the same height, they would project to equidistant parallel lines in the

image (as the projection is orthographic locally). However, as the ridge is higher than the sides, it must be displaced in the direction of the projection of the vertical (in 3-D) in that image. The extent of displacement from the center, in the image, provides an estimate of the relative height of the ridge with respect to the sides. This constraint is extremely important in filtering out triples that could not give rise to a valid gable hypothesis, as it does not rely on matching information from other views.

For each triple in the set of triples a search is conducted for matching sets of lines in the other views. The height of the ridge is varied through the maximum and minimum allowed heights of a building. At each iteration (height) the search scores the goodness of the supporting line features for a hypothesis over all the views. Maxima of these scores are used to hypothesize gable-roof hypotheses at the heights at which the maxima occur. At this point the system possesses 3-D information for the ridge and the sides, but no information about the extent of the hypothesis *i.e.* we need to determine closures for the hypothesis.

Closures for gable-roof hypotheses are determined by search for terminating junctions on the lines forming the ridge and the sides of the roof in all the views. The system uses binary junctions. If a junction  $j$  is to be a terminator, the lines that form  $j$ , say  $l_1$  and  $l_2$ , must meet the following constraints in order to qualify as a terminator for the hypothesis:

- either  $l_1$  or  $l_2$  should be a component of exactly one of the lines forming the triple
- if  $l_1$  is a component of one of the lines in the triple,  $l_2$  is constrained to be the projection of a 3-D line, say  $L_2$ , that is perpendicular in 3-D to the 3-D ridge and the 3-D side closest to  $L_2$ . The system has computed the 3-D orientations of the sides and the ridge. The location of the 3-D junction  $J$ , corresponding to its projection  $j$ , must lie on the 3-D ridge or on one of the sides of the gable in 3-D. The terminator of the gable is uniquely defined by these constraints. This terminator must correspond to  $L_2$ , and its projection to  $l_2$ . If the acute angle between the projected terminator and  $l_2$  is less than  $10^\circ$ ,  $l_2$  is a valid terminator.

In general, more than one termination may be found on each end of the roof. The junctions with the highest scores at each end that are valid terminators are selected as the terminators for this hypothesis. If no terminators are found, or if terminators for only one side are found the hypothesis is rejected. Figure 10 shows the results of gable-roof formation on the images in Figure 9. Note that this process is much more selective than is the case for flat roofs as many more constraints apply to the gable roof case.

### 3.5 Selection of gable-roof hypotheses

The gable roof hypotheses selection procedure is very similar to that for flat roof hypotheses; however, the selection can be more liberal as the number of false hypotheses generated is much

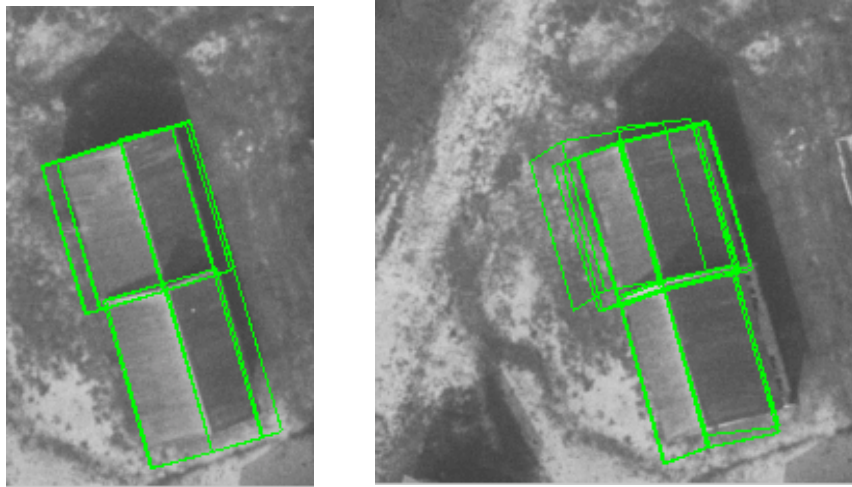


Figure 10 Gable-roof hypotheses for the images in Figure 9

smaller. Positive and negative evidence for roof is computed as before by using image lines that overlap the projected roof lines and negative evidence by the intersecting lines. This evidence is encapsulated into a “roof score”, which allows the hypotheses to be thresholded. The roof score is computed using the positive and negative roof evidence including the positive and negative contributions to the ridge of the gable-roof.

A hypothesis is selected if  $posRoofScore / n_{views} > 0.4$  and  $(posRoofScore - negRoofScore) / n_{views} > 0.25$ . Selected gable-roof hypotheses for the images in Figure 9 are shown in Figure 11.

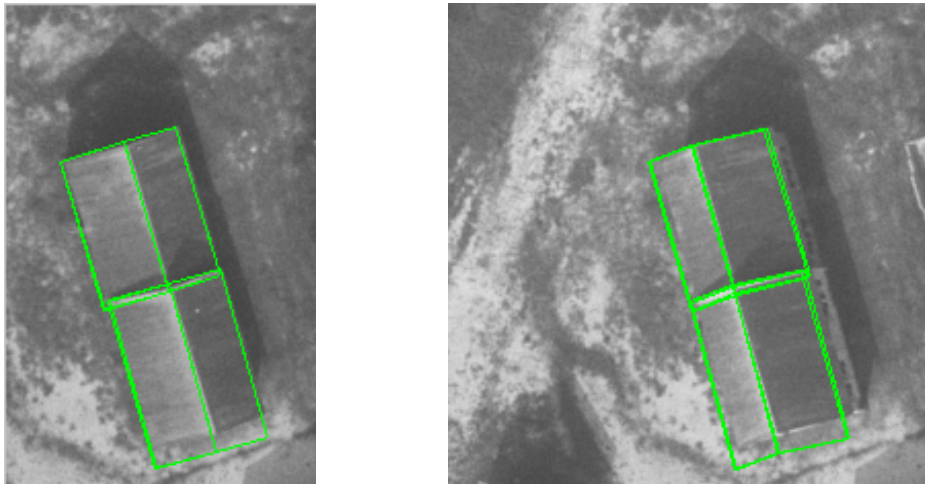


Figure 11 Selected gable-roof hypotheses for the images in Figure 9

#### 4 Verification of Building Hypotheses

The selection mechanism, covered in Section 3, demonstrates that the use of roof evidence and the predicted 3-D height provide good filters for removing hypotheses that do not satisfy the geometric constraints necessary for being declared a building. Additional evidence, consisting of cues for presence of visible walls and shadows cast by a hypothesized building are used to make

further discriminations. Statistical properties of the regions of the hypothesized roof and the shadows cast are also used.

In Section 4.1 the verification process for flat-roof building hypotheses is described. Section 4.2 outlines the verification process for gable-roof hypotheses. At this stage mutually exclusive overlapping or contained hypotheses (both flat-roof and gable-roof) may exist. Section 5 outlines the overlap disambiguation process.

#### **4.1 Verification of the flat-roof hypotheses**

In the verification process, the system uses the available geometric, photometric and domain-specific constraints such as expected shadow and wall lines, to determine whether a selected hypothesis is a building or not as described below.

##### **4.1.1 Collecting the verification Evidence**

###### **Roof Evidence**

Roof evidence for a hypothesis for verification is the same as used in the selection stage described in section 3.3.1 earlier.

###### **Wall Evidence**

In each view which is not nadir at least one and not more than two of the side walls of the buildings should be visible. The walls are assumed to be vertical in 3-D. The verification for walls involves looking for the projections of the horizontal bottom of the wall (the interface of the vertical wall and the ground). Wall evidence is deemed to be found if there is evidence of parallel lines at the distance from the top of the building that is predictable from its height in 3D. A score is computed based on how much of the predicted wall line is covered by an image line. Wall evidence is summed over the multiple available views.

###### **Shadow Evidence**

A 3-D building structure should cast shadows under suitable imaging conditions. The system has knowledge of the direction of illumination from the sun, which allows it to predict the location and orientation of shadows (on flat ground) from the 3-D hypotheses. Shadows have previously been used in monocular detection of buildings [6]. In this system, the analysis is made easier as an estimate of the height of the building is available. Even though the 3-D height of the building is known, a search for shadows is conducted within a small window (5 pixel wide) about the predicted shadow to account for various inaccuracies in estimations of the height, the sun position and the camera models. Figure 12 shows the search for shadows. We assume that the shadow falls on a flat ground of known height and that at least part of it is not occluded by other buildings.

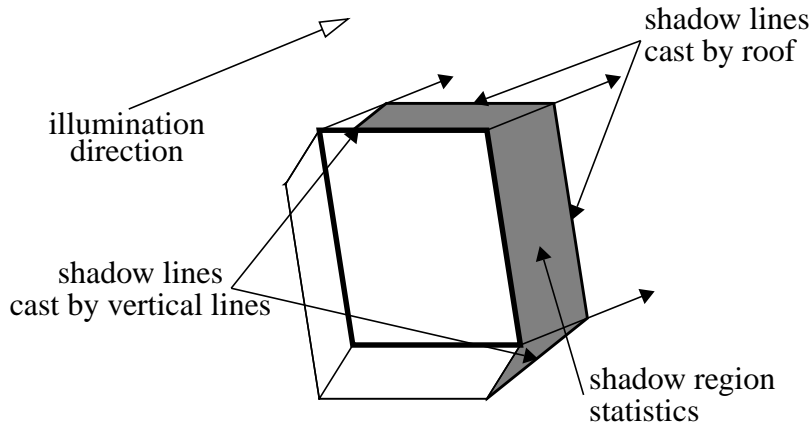


Figure 12 Search for shadow evidence

The evidence searched for includes the shadows cast by the horizontal roof lines, and the shadows cast by the vertical walls of the building. The evidence is in the form of detected lines at or near the outline of the predicted shadow. A shadow score is computed from the shadow evidence. This score is the fraction of the visible shadow outline that has supporting line evidence. Occlusion of shadows by the building itself is taken into consideration when searching for shadows. Score for multiple views is added to obtain a total score.

#### 4.1.2 Use of Roof, Wall and Shadow Evidence in Verification

The roof, wall and shadow evidence are collectively used to verify a selected flat-roof hypothesis. The verification process uses a decision tree involving the evidence available. Suppose that a 3-D hypothesis has a roof score of  $r$  (as computed in Section 3.3.1), a shadow score equal to  $s$  and a wall score that equals  $w$ . If there are  $n_{\text{views}}$  views being considered then

- if  $r \geq 0.75 * n_{\text{views}}$  the hypothesis is verified and the verification process exits. Experimental evidence indicates that not more than 30% of the (eventually verified) buildings satisfy this criterion.
- if  $r < 0.75 * n_{\text{views}}$ , shadow evidence is considered.
  - if  $s \geq 0.25 * n_{\text{views}}$ , or if the shadow evidence  $s_i \geq 0.5$  in some  $view_i$  the hypothesis is verified and the verification process exits.
  - if  $s < 0.25 * n_{\text{views}}$  and  $s_i < 0.5$  for all  $i=1..n_{\text{views}}$  and  $r \geq 0.5 * n_{\text{views}}$  and if  $w \geq 0.5 * n_{\text{views}}$ , then the hypothesis is verified and the verification process exits.
  - if  $s < 0.25 * n_{\text{views}}$  and  $s_i < 0.5$  for all  $i=1..n_{\text{views}}$  and  $r < 0.5 * n_{\text{views}}$  the hypothesis is not verified and the verification process exits.
- if none of the conditions enumerated above are satisfied the hypothesis is not verified and the verification process exits.

The results of verification of hypotheses generated from the pair of images in Figure 8, using the above rules, are shown in Figure 13.

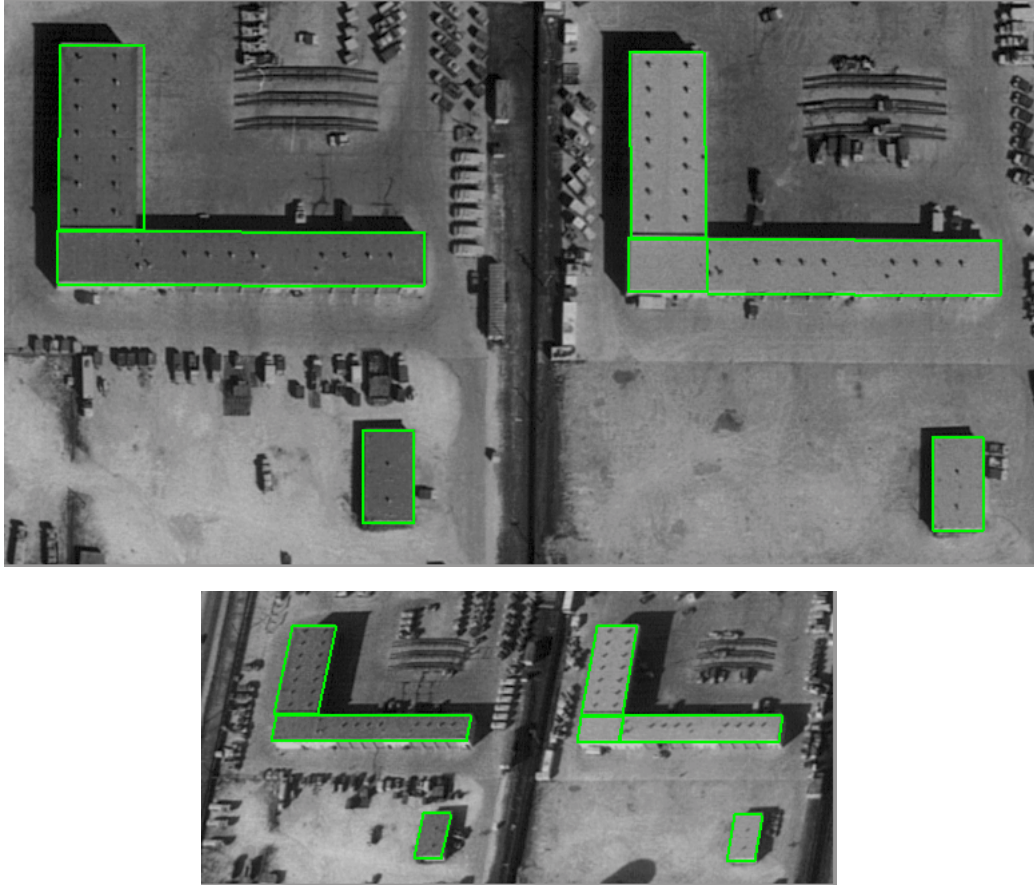


Figure 13 Verified hypotheses for the images in Figure 1

The roof, shadow and wall scores are combined using a sum of weighted averages to yield a confidence measure. If  $r$  is the roof score,  $s$  is the shadow score, and  $w$  is the wall score of a building hypothesis, the equation used to obtain a verification score (or confidence measure),  $v$ , for a verified hypothesis is  $v = (r_{wt} * r) + (s_{wt} * s) + (w_{wt} * w)$ , where  $r_{wt} = 0.6$ ,  $s_{wt} = 0.3$  and  $w_{wt} = 0.1$ . Also  $r_{wt} + s_{wt} + w_{wt} = 1$ . The reason for weighting wall evidence lower than roof and shadow evidence is that it is usually less reliable than roof evidence and shadow evidence. This confidence measure is used in overlap disambiguation that is covered in Section 5.

#### 4.2 Verification of the gable-roof hypotheses

The verification process for gable roof buildings is similar to that of flat roof buildings but the evidence used is more complex because of the complexity of the roof shapes. Collection and use of such evidence is described next.

## 4.2.1 Collecting the verification Evidence

### Roof Evidence

Roof evidence for a hypothesis from the selection process for gable-roof hypotheses, described in Section 4.1.2, is used in the verification process as well.

### Wall Evidence

The wall evidence for a gable-roof hypothesis is similar to that of an equivalent flat-roof hypothesis because the only difference in the generic model used for the flat-roof hypothesis and that used for the gable-roof hypothesis is the shape of the roof. As the wall evidence is independent of the shape of the roof, it is collected by treating a gable-roof hypothesis as a flat-roof hypothesis (by ignoring the “ridge”).

### Shadow Evidence

The search for shadows for gable-roof hypotheses differs from the search for shadows for flat-roof hypotheses. The shape of the gable-roof, coupled with the differing height of the “ridge” of the gable, as compared to the sides of the gable, cause shadow lines that are not parallel to the sides of the building causing them. However, knowing the heights of the sides and the “ridge” of the gable in 3-D, and the direction of illumination, the system predicts the shape before launching a search for supporting line evidence. This search includes the shadow cast by the roof, and the shadow cast by the vertical walls of the building. Following is an algorithmic description of the search for shadows of gable-roof hypotheses:

- Figure 14 shows a typical shadow cast by a gable-roof hypothesis (in view  $i$ ). Suppose the 3-D extremities of the gable-roof  $G$  are denoted by  $P_1$  through  $P_6$ . Denote the projection of  $P_1$  through  $P_6$  in  $view_i$  by  $p_{1i}$  through  $p_{6i}$  (refer to Figure 14).
- The extremities of the visible shadow are computed using the sun angles to be  $ps_{1i}$  through  $ps_{4i}$ .
- Search for line evidence that supports the outline of the shadow in a window of 5 pixels on either side of the predicted shadow outline.
- The shadow score,  $s$ , is computed using the same method as that used for flat-roofs in Section .

## 4.2.2 Use of Roof, Wall and Shadow Evidence in gable roof verification

The method for determining whether a selected gable-roof hypothesis should be verified or not is very similar to that for flat-roof hypotheses described earlier. Suppose that a 3-D gable-roof hypothesis has a roof score of  $r$  (as computed in Section 4.1.2), a shadow score equal to  $s$  and a wall score that equals  $w$ . If there are  $n_{views}$  views being considered then

- if  $r \geq 0.5 * n_{views}$  the hypothesis is verified and the verification process exits.

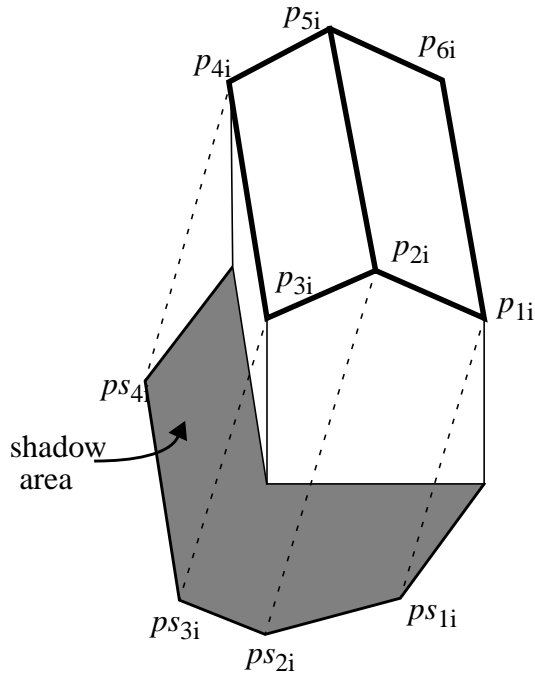


Figure 14 Shadow cast by a gable-roof hypothesis

- if  $r < 0.5 * n_{\text{views}}$ , shadow evidence is considered.
  - if  $s \geq 0.1 * n_{\text{views}}$ , or if the shadow evidence  $s_i \geq 0.2$  in some  $view_i$  the hypothesis is verified and the verification process exits.
  - if  $s < 0.1 * n_{\text{views}}$  and  $s_i < 0.2$  for all  $i=1..n_{\text{views}}$  and  $r \geq 0.25 * n_{\text{views}}$  and if  $w \geq 0.1 * n_{\text{views}}$ , then the hypothesis is verified and the verification process exits.
  - if  $s < 0.1 * n_{\text{views}}$  and  $s_i < 0.2$  for all  $i=1..n_{\text{views}}$  and  $r < 0.25 * n_{\text{views}}$  the hypothesis is not verified and the verification process exits.
- if none of the conditions enumerated above are satisfied the hypothesis is not verified and the verification process exits

When compared to the algorithm for flat-roof hypotheses, the thresholds used for gable-roof hypotheses are low. The reason for this is that a gable-roof hypothesis needs a greater number of lines to form its roof and its shadows, thus raising the possibility that some of these components will either not be found, or be found with less feature support, and hence lower confidence. This is compensated for by the fact that there are a greater number of applicable constraints to gable-roof hypothesis in the hypothesis formation stage, than there are to flat-roof hypothesis, and hence fewer gable-roof hypotheses than flat-roof hypotheses are usually formed for identical numbers of basic features (lines). The results of verification on the selected gable-roof hypotheses shown in Figure 10 are depicted in Figure 15.

The roof, shadow and wall scores are combined using a sum of weighted averages to yield a confidence measure. If  $r$  is the roof score,  $s$  is the shadow score, and  $w$  is the wall score of a building hypothesis, the equation used to obtain a verification score (or confidence measure),  $v$ , for a verified hypothesis is:  $v = (r_{wt} * r) + (s_{wt} * s) + (w_{wt} * w)$  where  $r_{wt} = 0.85$ ,  $s_{wt} = 0.05$  and  $w_{wt} = 0.05$ . Also  $r_{wt} + s_{wt} + w_{wt} = 1$ . The reason for weighting wall evidence and shadow evidence lower than roof evidence is that their expected shapes are complex, and usually less reliable than roof evidence. This confidence measure is used in overlap disambiguation that is covered in Section 5.

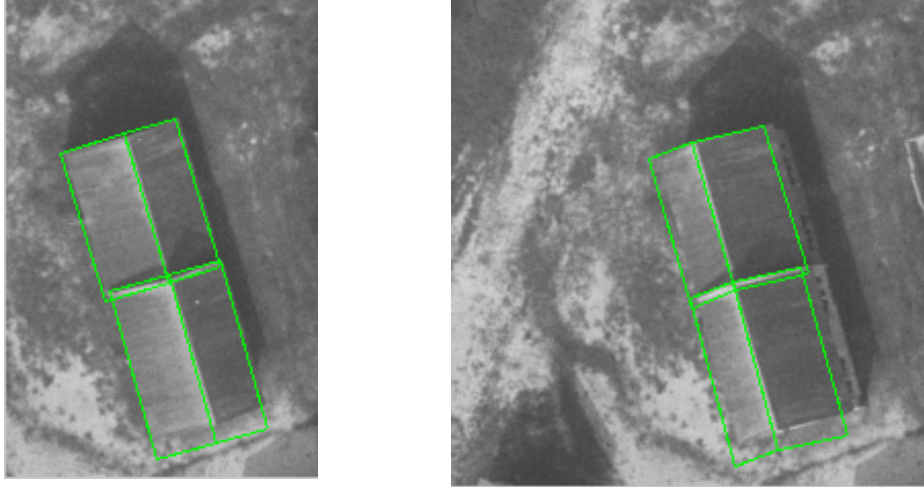


Figure 15 Verified gable-roof hypotheses for the images in Figure 9

## 5 Overlap analysis

The system verifies selected flat-roof and gable-roof hypotheses one at a time. Hence it is possible that verified hypotheses may overlap in 3-D. Selection among overlapping hypothesis is conducted for all verified rectangular components by dropping the models to the ground plane, and testing whether their projections overlap on the ground plane. This effectively reduces overlap testing to a 2-D problem. Note that more than two hypotheses may overlap. In these cases hypotheses are handled in pairs, recursively, till there are no conflicts. The use of verification scores as a discriminator ensures that the process is order-independent, as it orders a set of overlapping hypotheses in an unambiguous way. There are three distinct cases where overlap occurs (the first two cases are similar to those described in [6]).

### 5.1 Complete overlap or containment

If verified hypothesis  $H_2$  is completely contained in verified hypothesis  $H_1$  (in 3-D), and the difference in 3-D heights of  $H_1$  and  $H_2$  is small (say less than 1m), then  $H_2$  is removed from the set of verified 3-D hypotheses. If the difference in height is larger, then both  $H_1$  and  $H_2$  are included in the verified hypotheses. This case implies that  $H_2$  is a superstructure of  $H_1$ .

## 5.2 Partial overlap with no shared evidence

When two hypotheses partially overlap but neither is completely contained by the other, and they share no common roof evidence, then the one with the higher verified score is retained.

## 5.3 Partial overlap with shared evidence

When hypotheses  $H_1$  and  $H_2$  partially overlap and neither is completely contained by the other, but they share common roof evidence, a differential process determines which of  $H_1$  and  $H_2$  survives. The rationale for using a differential process in this case is that the hypotheses will have very similar scores, as they share evidence, and that the differences must be highlighted in order to make a finer judgement about which hypothesis is better.

Suppose in  $view_i$ , hypothesis  $H_1$  is projected to  $h_1$  and  $H_2$  is projected to  $h_2$ . Computation of the differential evidence is done thus:

- without loss of generality say side  $l_{11}$  of  $h_1$  shares evidence with side  $l_{21}$  of  $h_2$ .
- search for line segments for the section of  $l_{11}$  that do not overlap with  $l_{21}$ , and for  $l_{21}$  that does not overlap with  $l_{11}$ .
- score this roof evidence as the fraction of the searched segment that is covered by actual line evidence, using the method for scoring roof evidence described in Section 3.3.2.
- repeat the process for all sides of  $h_1$  and  $h_2$  that share evidence in  $view_i$
- repeat the steps described above on all views.
- compute the line evidence for the part of the predicted wall lines of  $h_1$  that is not shared with  $h_2$ , and for the part of the predicted wall lines of  $h_2$  that is not shared by  $h_1$  (differential wall evidence)
- compute the line evidence for the part of the predicted shadow lines of  $h_1$  that is not shared with  $h_2$ , and for the part of the predicted shadow lines of  $h_2$  that is not shared by  $h_1$  (differential shadow evidence)

The differential score for  $H_1$  is the verification score computed using Equation in Section 4.1.2 if  $H_1$  is a flat-roof hypothesis or using Equation in Section 4.2.2 if  $H_1$  is a gable-roof hypothesis, using the differential roof, shadow and wall scores as input. The differential score for  $H_2$  is computed by the same method. The hypothesis with a higher differential score survives, while the one with a lower differential score is eliminated from the set of verified hypotheses. It is interesting to note that the architecture of the system allows different models (like flat-roof and gable-roof buildings) to be disambiguated using the same methodology.

The verified buildings, both flat-roofed and gable-roofed, are 3-D structures. The 3-D information of the verified buildings coupled with the camera model and the terrain model of the scene can be used to generate the 3-D wire frame model of the scene. At this point, texture-mapping

may be applied to reconstruct the scene from viewpoints other than those captured in the views. This ability to generate realistic views from arbitrary viewpoints has interesting applications such as fly-by simulations.

## **6 Results and Evaluation**

This section shows results on some examples. We have used sections of the Fort Hood, Texas and Fort Benning, Georgia sites. These datasets have been commonly available and allow for comparisons among different systems. We first show graphical results in Section 6.1 and then present an evaluation in Section 6.2.

### **6.1 Results**

#### **6.1.1 Results on Fort Hood, Texas**

Fort Hood, Texas is a site with several hundred 100 buildings. The site is characterized by low buildings of varying shape, size, orientation and roof intensity. It has foliage and trees that clutter the background, as well as a number of man-made structures such as car park areas, vehicles and vehicle ports that create accidental alignment of features that sometimes qualify as buildings. The views are acquired from nadir as well as oblique angles, and at resolutions varying from 0.3 m/pixel to 1.3 m/pixel. They are acquired at different times of the day (at different days in the year) leading to widely varying image characteristics in the different views for the same area on the ground. Only selected examples are shown here due to lack of space.

Figure 16 shows results on some gable roof buildings, however, the roof slope is quite low and hence the buildings are described as flat-roof buildings. One half of the building labeled A is included in the set of verified hypotheses. The whole building (detected as a flat-roof) exists in the set of selected hypotheses. However, that hypothesis is eliminated by the one labeled A, as this hypothesis has much stronger roof evidence, and the differential evidence in favor of the other hypothesis is not sufficient to raise its confidence beyond that of the hypothesis labeled A. Note that the shadow evidence is common, and the wall evidence is almost negligible.

Figure 17 shows a set of multi-level buildings. The roofs in this example, are of varying intensities within each view. In addition, the intensities vary significantly across views, for corresponding areas in each view. Examining the building labeled B in both images shows that the system is able to detect buildings even if they have small protrusions, if a sufficiently large segment of the roof exists to be able it to be pieced together using perceptual grouping.

Figure 18 shows two views of complex buildings in a relatively uncluttered background. The part of the building labeled C is an example of a building that exists but is not detected (a true



Verified hypotheses

Figure 16 Gable-roof buildings detected as flat-roof buildings. Left image is 456x250 with GSD of 0.5m, right image is 450x215, GSD 0.4m



Figure 17 Example of multi-level buildings. Top view is 775x500 pixels, GSD of 0.3m, bottom view is 465x280 with GSD of 0.6m.

negative). The building exists in the set of all hypothesized buildings, but is not selected. The reason it is not selected is that the other part of the building, labeled D, occludes different sections of C because it is taller and the viewpoint causes occlusion, and also because it casts a shadow on

C. This causes a mismatch along an entire side of C, and its consequent elimination in the selection process.

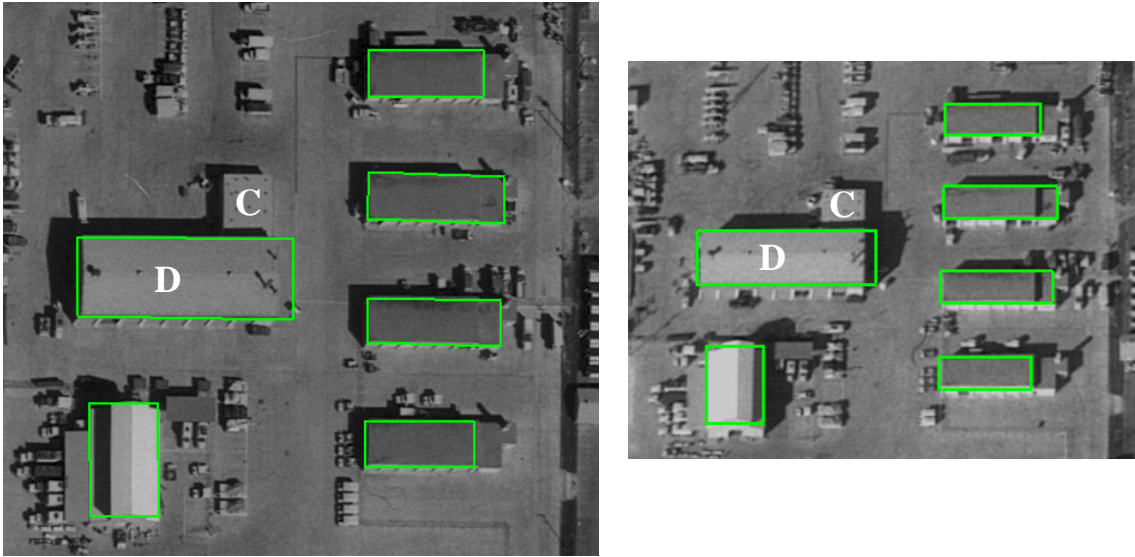


Figure 18 Complex buildings in a relatively uncluttered background. Left view is 500x460 pixels, GSD of 0.33m, right view is 325x260 pixels, GSD of 0.5 m

Figure 19 shows a model constructed by the automatic system for large parts of the Fort Hood site (results shown are on a nadir view but two oblique views are also used for processing). These results were obtained by breaking the nadir image into consecutive sections for efficiency; each section contained several buildings as in examples shown earlier. The sections were selected simple to limit the size of the images to be processed. Only a pair of views was used for each section, however, different pairs of views may have been used in each example as no single pair overlapped completely in the area shown in each of the examples.



Figure 19 Model constructed from several sub-images using three views, displayed on a nadir image. This image is 7600x2700 pixels with GSD of 0.3 m, the oblique views have a smaller GSD (similar to other images of the same site shown elsewhere).

### 6.1.2 Results for Fort Benning, Georgia

The Fort Benning dataset differs from the Ft. Hood dataset in that buildings have varying shape, size, orientation and roof intensity. The buildings have very distinct markings on the roofs that complicate the task of delineating them and there are a number of gable-roof buildings. The site has as a small number of prominent roads that create distinct features in the images. Both view are nadir. Figure 20 shows the combined hypotheses on buildings in this site (the overlap between flat and gabled roof hypotheses is not resolved).

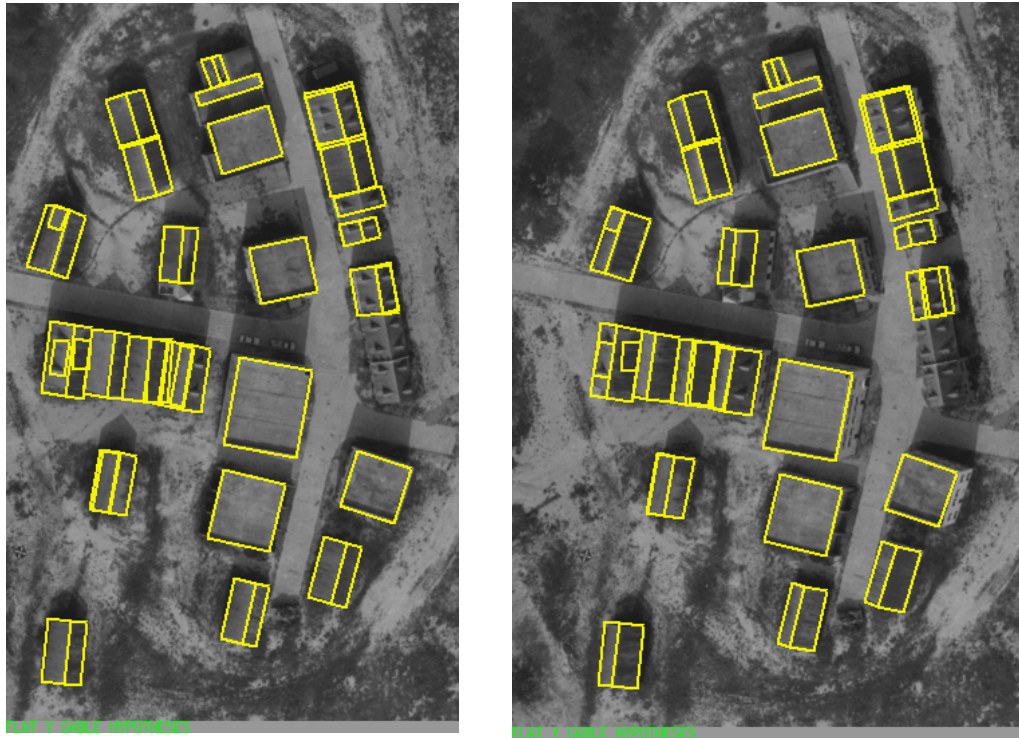


Figure 20 Combined hypotheses (flat-roof and gable-roof buildings). Left image is 345x540 pixels, right image is 400x540 pixels. GSD for both is 0.28 m.

## 6.2 Evaluation of the automatic system

We describe the run times, the detection accuracy and effect of multiple views followed by a discussion of the effects of system parameters in this section.

### 6.2.1 Run-time Evaluation

The run-times for the system depend on the complexity of the images and the number of features in them. For the example shown in Figure 1 , the total run time was approximately 225 seconds. Edge and line detection take 55 seconds. Hypothesis formation, selection and verification required 70, 65 and 35 seconds respectively. The runtimes grow approximately linearly with the size of the image for scenes of similar complexity.

### 6.2.2 Detection Evaluation

The following measures proposed in [6] are computed by making a comparison of the automated results with a reference model constructed manually:

$T_p$  (True Positive) is a building in the reference model and detected by the program

$F_p$  (False Positive) is a building detected by the program but not present in the reference model

$T_n$  (True Negative) is a building in the reference model but not detected by the program.

These are combined to give the following two numbers:

- Detection Percentage =  $100 \times T_p / (T_p + T_n)$   
Branch Factor =  $100 \times F_p / (T_p + F_p)$

In the above computations, a building is considered to be detected, if a part of the building is detected by the system. The description of the detected building may not necessarily be correct. To compute accuracy of the delineation, three other measures consisting of the correct building pixel, incorrect building pixel and correct non-building pixel percentages, are calculated by labeling every pixel in the image as either a building pixel or a non-building pixel as proposed by Shufelt and McKeown in [7]. We did not evaluate height accuracies explicitly though displays on multiple images show that the heights for correctly detected buildings are highly accurate. More sophisticated approaches to delineation accuracy are described in [25].

A large section of the Motor Pool area of the Fort Hood site is selected for evaluation using the parameters described above. The area was run in sections with between 3 and 14 buildings in each section. Each section was run using two overlapping views. Figure 19 shows the model depicted in 3 different views. The buildings vary considerably in size, shape and image characteristics. Many rectilinear buildings are composed of rectangular parts. In order to characterize the performance of the system, the parameters  $T_p$ ,  $T_n$  and  $F_p$  are computed for complete (possibly multi-part) buildings, as well as for individual rectangular building fragments. The derived measures, namely the detection percentage and the branch factor, are computed independently in each case. The results are summarized in Table 6.1. The pixel-based measures show 97.1% correct building pixels, 4.3% incorrect building pixels and 99.8% correct non-building pixels. It may be noted that the pixel-based measures indicate better performance than the

measures for individual buildings or building fragments because the large buildings are detected, with false positives and true negatives being usually small.

**Table 6.1**

	$T_p$	$T_n$	$F_p$	Detection Percentage	Branch Factor
<b>Complete (rectilinear) buildings</b>	84	4	14	95.45%	14.2%
<b>Rectangular building fragments</b>	134	11	14	92.41%	9.46%

### 6.2.3 Effect of Multiple Views

The system presented here handles two or more views non-preferentially. Comparison of results using three views and using two views demonstrates that the additional view sometimes aids the building detection and description process with respect to detecting buildings (or parts of buildings) that were not detected using two views (*i.e.*  $T_p$  is increased,  $T_n$  is decreased). More importantly, the additional view causes the number of false positives ( $F_p$ ) to decrease. Table 6.2 shows  $T_p$ ,  $T_n$  and  $F_p$  using three views, for the sections of Fort Hood that have at least three overlapping views. Comparison of Table 6.2 with Table 6.1 shows that the third view increases detection percentage and reduces the branch factor. 97.4% correct building pixels, 3.9% incorrect building pixels and 99.8% correct non-building pixel measures are found. Tests run with four views do not increase  $T_p$ , or decrease  $T_n$  in the areas of Fort Hood that have at least four overlapping views. In these results, the areas with three views are a subset of the areas with only two views. However, the scene is quite homogeneous and the performance does not vary greatly in different parts and hence we believe that the given comparison is meaningful, particularly due to the large reduction in the false positives. In general, the value added by a view will depend on several factors such as its viewing angle, the direction of illumination and occlusions from the viewpoint. We were not able to conduct such evaluations due to the difficulty of acquiring sufficiently varied datasets.

### 6.2.4 Choice of Parameters

Our system needs to make several decisions based on incomplete evidence at various stages in the selection and verification processes. Evidence from several components, such as lines is combined to compute a single score. Evidence in various categories such as from roofs, walls and shadows is combined to make decisions on keeping or discarding hypotheses. Due to the

**Table 6.2**

	$T_p$	$T_n$	$F_p$	<b>Detection Percentage</b>	<b>Branch Factor</b>
<b>Complete (rectilinear) buildings</b>	27	1	0	96.43%	0.00%
<b>Rectangular building fragments</b>	38	3	0	92.68%	0.00%

complexity of the processes involved, and lack of formal models for the contents of an image, it is difficult to find theoretically optimal solutions. Our procedures for such decision making are based on simplicity and intuitive judgements (such as weighting the evidence for various lines in proportion to their length). There is also a new effort in our group to use Bayesian models learned from examples for improved decision making [26].

Our system has been tested on a large number of examples with most of the parameters being either fixed or automatically adjusted based on available input information such as the image resolution with the user supplying the minimum and maximum dimensions of the acceptable buildings only. Some internal parameters could have been chosen from photogrammetric information such as errors in the bundle adjustment process but this was not done in the described experiments.

## 7 Conclusions

We have presented a system for detecting rectilinear buildings with flat or symmetric gable roofs from multiple intensity images. We believe that our results are encouraging and are similar to or better than results shown from other contemporary systems. It is now common for many systems to use hypothesize and verify approach to object detection and modeling. However, we believe that our approach of interleaving feature grouping and matching and delaying decisions until more information is available for a better decision is novel. Our system also treats all images in a non-preferential, order independent way.

We believe that the approach presented here can be easily generalized for more complex, but specific shapes (such as for regular polygon shaped roofs) but would require major modifications to handle more general cases. The system is also limited in its ability to handle buildings which are rectilinear but highly complex such as consisting of many wings, or cases where the buildings are

very close to each other (such as in the “downtown” district of a major city). Modeling of more complex buildings in more complex surrounds remains a topic for future research.

## Acknowledgments

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

## References

- [1] J. Canny, “A Computational Approach to Edge Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698, November 1986.
- [2] A. Grün and R. Nevatia, editors, Special Issue on Automatic Building Extraction from Aerial Images, *Computer Vision and Image Understanding*, Vol. 72, No. 2, November 1998.
- [3] A. Huertas and R. Nevatia, “Detecting Buildings in Aerial Images,” *Computer Vision, Graphics and Image Processing*, 41(2), pp. 131-152, February 1988.
- [4] R. Irving and D. McKeown, “Methods for exploiting the Relationship Between Buildings and their Shadows in Aerial Imagery,” *IEEE Transactions on Systems, Man and Cybernetics*, 19(6), pp. 1564-1575, November/December 1989.
- [5] J. Shufelt and D. McKeown, “Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery,” *Computer Vision, Graphics and Image Processing*, 57(3), pp. 307-330, May 1993.
- [6] C. A. Lin and R. Nevatia, “Building Detection and Description from a Single Intensity Image,” *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 101-121, November 1998.
- [7] J. Shufelt and D. McKeown, Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery, *Computer Vision, Graphics and Image Processing*, 57(3): 307-330, 1993.
- [8] J.C. McGlone and J. Shufelt, “Projective and Object Space Geometry for Monocular Building Extraction,” In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 54-61, Seattle, WA., June 1994.
- [9] M. Roux and D. McKeown, “Feature Matching for Building Extraction from Multiple Views,” In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 46-53, Seattle, WA., June 1994.
- [10] M. Herman and T. Kanade, “Incremental Reconstruction of 3-D Scenes from Multiple, Complex Images,” *Artificial Intelligence*, 30(3), pp. 289-341, Dec. 1986.
- [11] R. Collins, C. Jaynes, Y.Q. Cheng, X. Wang, F. Stolle, E. Riseman and A. Hanson, “The Ascender System: Automated Site Modeling from multiple Aerial Images,” *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 143-162, November 1998.
- [12] P. Fua, Model-Based Optimization: Accurate and Consistent Site Modeling, in proceedings, *18th ISPRS Congress*, Comm. III, WG 2, Vienna, Austria, 1996, pp. 222-233.

- [13] 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*, Virkhauer Verlag, Basel, 1995, pp. 145-168.
- [14] A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Förstner, L. Plümer and V. Steinhage, "Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D," *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 185-203, November 1998.
- [15] O. Henricsson, "The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction," *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 163-184, November 1998.
- [16] 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*, Virkhauer Verlag, Basel, 1995, pp. 211-220.
- [17] 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*, Virkhauer Verlag, Basel, 1995, pp. 199-210.
- [18] U. Weidner, "An Approach to Building Extraction from Digital Surface Models," in *Proceedings of the 18th ISPRS Congress*, Comm. III, WG 2, Vienna, Austria, 1996, pp. 924-929.
- [19] T. Kim and J. Müller, "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*, Virkhauer Verlag, Basel, 1995, pp. 221-230.
- [20] A. Huertas, Z. Kim and R. Nevatia, "Multisensor Integration for Building Modeling," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, SC, pp. 203-210, June, 2000.
- [21] 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*, Virkhauer Verlag, Basel, 1995, pp.135-144.
- [22] N. Paparoditis, M. Cord, M. Jordan and J.P. Cocquerez, "Building Detection and Reconstruction from Mid- and High Resolution Aerial Images," *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 122-142, November 1998.
- [23] D. McKeown, C. McGlone, S. Cochran, W. Harvey, J. Shufelt and D. Yocum, "Automated Cartographic Feature Attribution using Panchromatic and Hyperspectral Imagery," in *Proceedings of the DARPA Image Understanding Workshop*, Monterey, CA, pp. 517-536, November, 1998.
- [24] A. Huertas, R. Nevatia and D. Landgrebe, "Use of Hyperspectral Data with Intensity Images for Automatic Building Modeling," in *Proceedings of the Second International Conference on Information Fusion*, Sunnyvale, CA, pp. 680-687, July, 1999.
- [25] R. Nevatia, "On Evaluation of 3-D Geospatial Modeling Systems," *Societe Francaise de Photogrammetrie et Teledetection*, Bulletin No. 153, SIPT/ISPRS-WG II/6, Paris, France, April 1999, pp. 15-21.
- [26] Z. Kim and R. Nevatia, "Uncertain Reasoning and Learning for Feature Grouping," in *Computer Vision and Image Understanding*, Vol. 76, No. 3, December 1999, pp. 278-288.