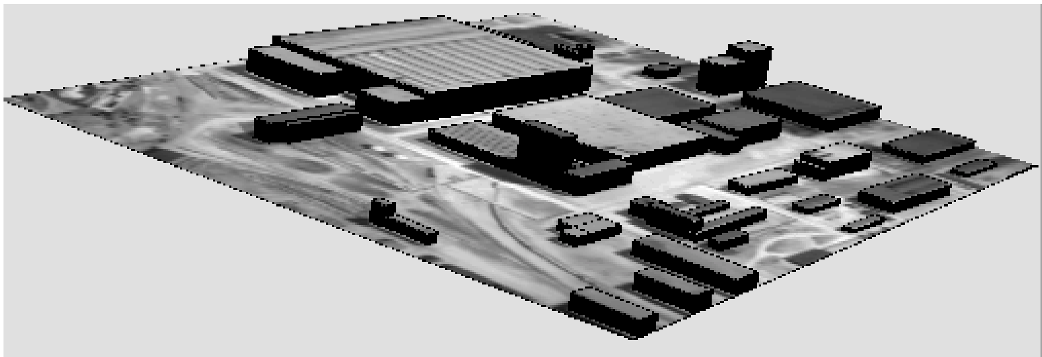

Fig. 8: 3-D rendered view of scene

7 Conclusion and Future Work

We plan to continue testing our system with a variety of scenes, and to incorporate capabilities to detect and describe buildings with a variety of building roofs. Currently we assume that the detected and verified structures lay on the ground. Some structures however are located on top of other structures. That level of refinement of the description requires an additional step in our system and is one of the subjects of our current and future work.

References

Canny J. (1986), "*A Computational Approach to Edge Detection*", IEEE-PAMI, Vol. 8, No. 6, pp 679-698.

Chung C., Nevatia R (1992) *Recovering Building Structures from Stereo*, In Proceedings of the IEEE Workshop on Applications of Computer Vision, pp 64-73, Palm Springs, Calif., December 1992.

Huertas A. (1983), *Using Shadows in the Interpretation of Aerial Images*, University of Southern California USC-ISG Technical Report 104, October 1983.

Huertas A., Nevatia R. (1988), *Detecting Buildings in Aerial Images*, Computer Vision, Graphics and Image Processing, Vol. 41, No. 2, pp 131-152.

Huertas A., Lin C., Nevatia R. (1993), *Detection of Buildings from Monocular Views of Aerial Scenes using Perceptual Organization and Shadows*, in Proceedings of the 1993 ARPA Image Understanding Workshop, Washington D.C., pp 253-260, April 1993.

Irvin R., McKeown D. (1989) *Methods for exploiting the Relationship Between Buildings and their Shadows in Aerial Imagery*, IEEE-SMC, Vol. 19, No. 6, pp 1564-1575.

Liow Y., Pavlidis T. (1990) *Use of Shadows for Extracting Buildings in Aerial Images*, Computer Vision, Graphics and Image Processing, Vol. 49, pp 242-277.

Mohan R., Nevatia R. (1989) *Using Perceptual Organization to Extract 3-D Structures*, IEEE-PAMI, Vol. 11, No. 11, pp 1121-1139.

Nevatia R. Babu R. (1980) *Linear Feature Extraction and Description*, Computer Vision, Graphics and Image Processing, Vol. 13, pp 257-269.

Venkateswar V., Chellappa R. (1990) *A Framework for Interpretation of Aerial Images*, In Proceedings of the International Conference on Pattern Recognition, pp 204-206, Atlantic City, NJ.

computed from the parallelograms verified in our Ft. Hood example is shown in Fig. 6c.

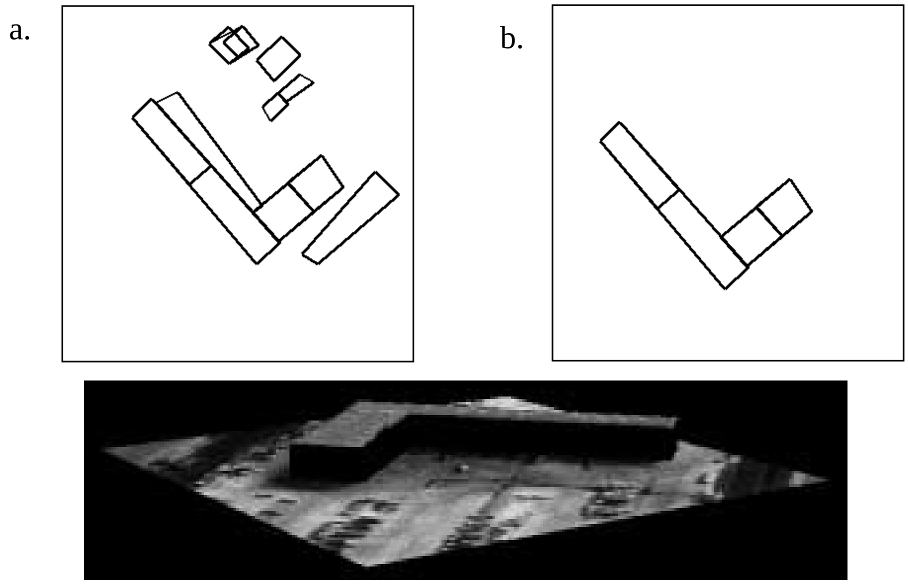

Fig. 6: (a) Selected roof hypotheses. b) Verified roof hypotheses. (c) 3-D rendered view.

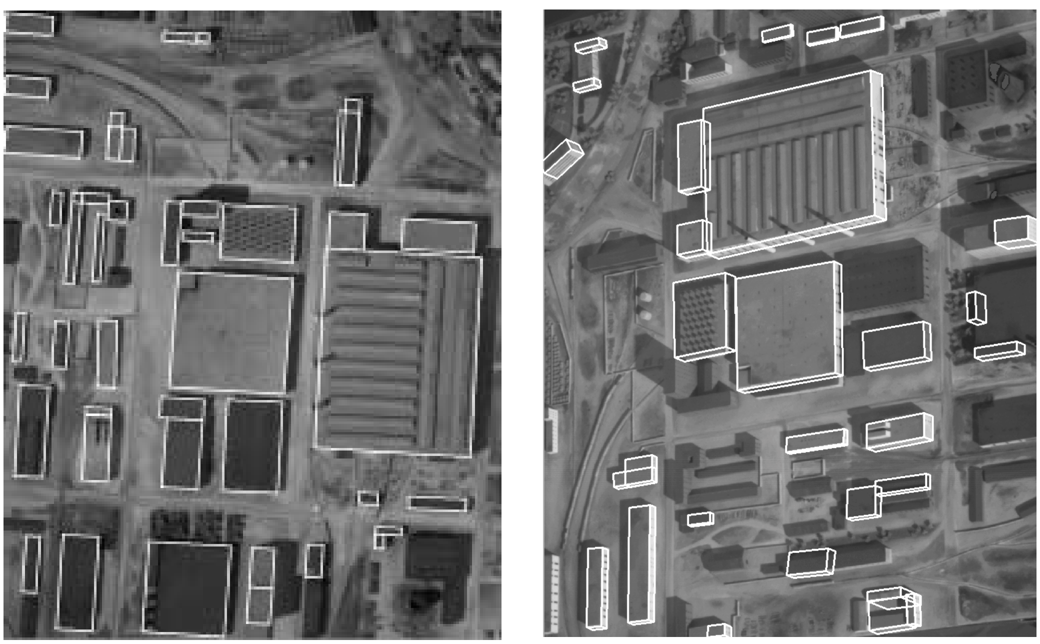

Fig. 7: (a) Nadir view scene. (b) Oblique view scene.

6 Results

Our system has been tested on a number of examples provided by the RADIUS program with very good results. Due to lack of space, we show only the final result for two images of a modelboard scene. Fig. 7a shows the result for a nadir view and Fig. 7b the result for an oblique view. These 1320x1100 images with about 40 structures takes about one hour on a Sparc10/30 to process. In Fig. 9 we show an automatically generated 3-D view of the scene, from an arbitrary viewpoint.

Collecting and evaluating wall evidence proceeds similarly to the shadow processing described above. We search within an arbitrary window, or within minimum and maximum weight values, if these are known. The system can either do an exhausted search over the search range or look for some evidence first and do a smart search just over the range with higher probability of existence of wall evidence.

The purpose of wall process is to find all wall evidence at every possible building height for each and every roof hypothesis. Given a building height, we can compute the height of walls in 2-D and estimate the expected wall boundary for a hypothesis. All evidence around the wall boundary will be collected and a score will be evaluated for the wall evidence.

Search and Evaluation of Wall Evidence

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

The evaluation process evaluates the wall evidence collected from the previous step. Basically the score is a weighted sum of the evidence of ground-boundary, vertical-boundary, and corners. Equation (2) is the evaluation function of the wall evidence of a hypothesis p given a building height H. k_i is one of the evaluation functions of different wall evidence and v_i is the corresponding weight.

$$W(p, H) = \sum_i v_i \cdot k_i(p, H) \tag{2}$$

4.3 Combination of Shadow and Wall Evidence

Next we determine at what building height can each hypothesis find the best combined shadow and wall evidence. For this, we formulate a function to evaluate a score for the shadow and wall evidence of a hypothesis (p) for a given building height (H). The corresponding shadow and wall scores are given from equations (1) and (2). We use the equation (3) from certainty theory to combine these two scores.

$$C(p, H) = S(p, H) + W(p, H) - S(p, H) \times W(p, H) \quad \text{where} \quad 0 \le S(p, H), W(p, H) \le 1 \tag{3}$$

The possible building heights are limited by the range between the minimum building height and the maximum building height. For each hypothesis, the building height gives the highest combined score will be considered the most possible building height of the hypothesis and the score will be called the confidence value of the hypothesis. If the confidence value of a hypothesis is greater than a given threshold value, the hypothesis will then be verified. The parallelograms verified in our Ft. Hood example are shown in Fig. 6b.

5 3-D Description of the Scene

Verified roof hypotheses have a height associated with them, allowing us to derive a 3-D model. The verified roofs are used to generate an “elevation map” with the pixels encoding structure height. The map can be rendered and viewed from an arbitrary viewpoint. Other 3-D representations such as wire frame models, are easily derived also. A 3-D arbitrary view lected in our Ft. Hood example after both the local and global selection criteria have been applied are shown in Fig. 6a.

4 Verification of Hypotheses

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

4.1 Shadow Verification Process

The purpose of shadow verification process is to establish the correspondences between shadow casting elements and shadows cast, and the use of these correspondences to verify a hypothesis. We assume that the ground surface in the immediate neighborhood of the structure is fairly flat and level. The shadow casting elements are given by the sides and junctions of the selected parallelogram hypotheses. The shadow boundaries are searched for among the lines and junctions extracted from the image.

There are a number of difficulties that prevent the accurate establishment of correspondences however. Building sides are usually surrounded by a variety of objects such as loading ramps and docks, grass areas and sidewalks, trees, plants and shrubs, vehicles, light and dark areas of various materials. Nearby structures may reflect light into the shadowed areas making the objects in it more visible, and so on. To deal with these problems we have adopted some geometric and projective constraints and special shadow features.

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

The shadow verification process finds all shadow evidence from the extracted potential shadow evidence within a given distance of a roof hypothesis. This distance can be set to be an arbitrary value, a range of values, or derived from building height information if available.

We evaluate the shadow evidence and give a score as a weighted sum of the evidence of shadow lines cast by roof, shadow lines cast by vertical lines, shadow junctions and the shadow region statistics. Equation (1) is used to compute a score for the shadow evidence of a hypothesis p given a building height H. h_i is one of the evaluation functions of different shadow evidence and u_i is the corresponding weight.

$$S(p, H) = \sum_i u_i \cdot h_i(p, H) \quad (1)$$

4.2 Wall Verification Process

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

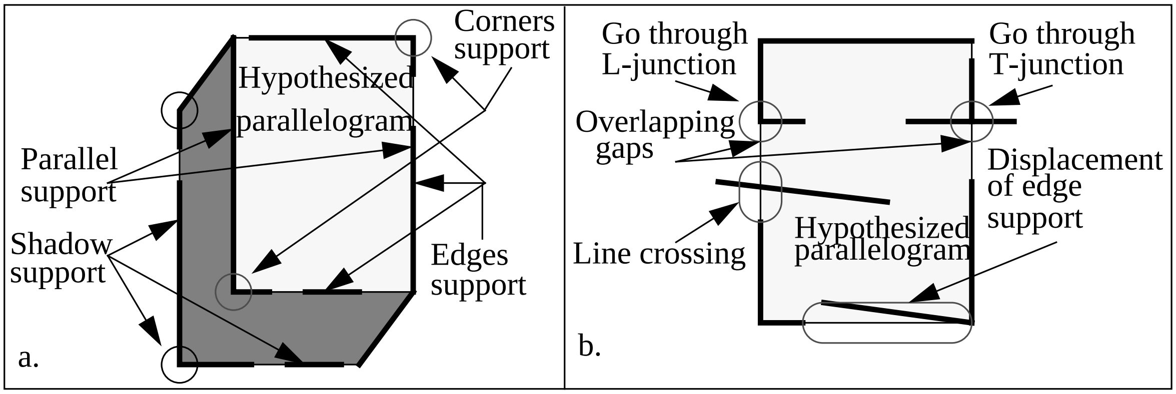

Fig. 5: (a) Positive evidence. (b) Negative evidence.

positive evidence, because it helps us to remove those parallelograms which are less likely to be part of buildings.

Each kind of evidence of support is formulated into an evaluation criterion. There are no formal definition of goodness of a parallelogram thus, our evaluation criteria formulated from evidence of support are based on analysis of likely and unlikely events. For example, four junctions are very unlikely to fall on the four corners of a parallelogram accidentally. Also, from the Gestalt Laws of Perceptual Grouping, the Law of Closure suggests that the existence of L-junctions or T-junctions on a side of a parallelogram will make a closure on part of the parallelogram and it means that the hypothesized parallelogram is not good. Some evidence of support are not always available such as the shadow evidence and the OTV evidence, but they are very important because it is very unlikely that some shadow features will appear around the hypothesized parallelogram by chance and the probability for three lines to form an OTV corner by chance is very small. We emphasize the importance of an evaluation criterion by assigning a higher weight to it.

Positive weights are assigned to those evaluation criteria formulated from positive evidence of support while negative weights are assigned to those evaluation criteria formulated from negative evidence of support. A weight should be assigned to each evaluation criterion according to the probability of existence of buildings under the condition of presence of the evidence of support from which the evaluation criterion is derived. However, we do not have the probabilistic analysis of goodness of a parallelogram, but the problem of optimal weights assignment for a given set of examples could be formulated into a search problem.

Good parallelograms surviving local selection may compete with each other. For example, some parallelograms could share the same edges or corners support and some parallelograms might overlap with each other. The goal of global selection criteria is to select a minimum set of parallelograms which best describe the composition of the scene. There are four global selection criteria in our system. They deal with duplicated, mutually contained, fully contained, and overlapping parallelograms. They examine competing parallelograms and choose one if appropriate. The selection is based on relative properties of each parallelogram, the amount and kind of overlap, and whether they share support or not. Note that a parallelogram fully contained in another is not necessarily removed. If a parallelogram does not overlap with any other parallelogram then it is not in competition, and is kept.

The flexibility of our scheme allows for straightforward incorporation evaluation of goodness and addition of filters expressing additional global relationships. The parallelograms se-

Skewed Rectangles or Parallelograms

Parallelogram structures are generated from the U-structures. The parallelograms formed in our example are shown in Fig. 4b. Thu number of parallelograms formed can be reduced by restricting their formation on the basis of size, as a function of image resolution, for example. Parallelograms are also generated from matching junctions along the direction of illumination (see Huertas et al., 1993.) We hypothesize the missing portions of a parallelogram having a corner with a matching shadow corner or evidence of an OTV.

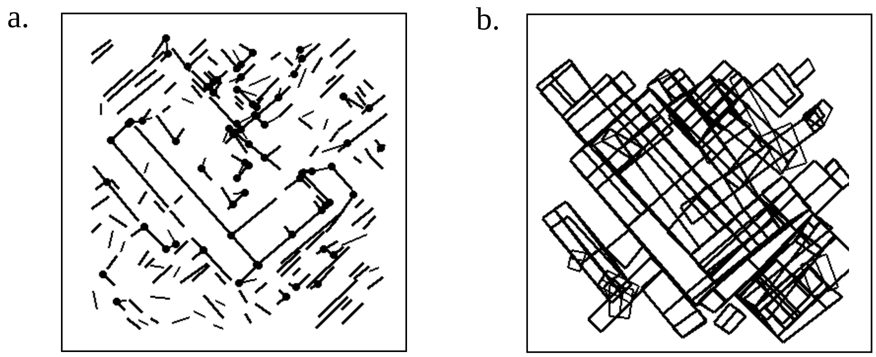

Fig. 4: (a) Linear Structures and Junctions. (b) Parallelogram generated

3 Selection of Hypotheses

After the formation of all reasonable parallelograms, a selection process is applied to choose those having strong evidence of support and minimum conflict among them. We use a criteria-based method which uses two kinds, *local* and *global* selection criteria. Local selection criteria determine whether or not a parallelogram is "good" on the basis of local supporting evidence. Good parallelograms are retained for global selection even if they overlap. Global selection criteria select the best consistent parallelograms from good parallelograms.

We apply local selection (or evaluation) criteria and global selection criteria differently. Local selection criteria work together to evaluate the goodness of a parallelogram, while global selection criteria work separately. Each global selection criterion acts like a filter. The set of retained parallelograms pass through all filters and the set of parallelograms coming out from the last filter will be the set of parallelograms selected by the selection process.

The local selection criteria are used to compute, for each parallelogram, a goodness value to be compared to a threshold value. Every evaluation criterion is weighted according to its importance. The goodness value is then measured by the sum of the weighted values calculated by the evaluation criteria. The problem of measuring the goodness of a parallelogram now becomes a problem of finding and formulating good evaluation criteria, and assigning appropriate weights. This allows us to remove parallelograms formed using weak evidence.

Whether a parallelogram is good or not depends on the evidence of support. We distinguish between positive evidence and negative evidence of support. The positive evidence of support includes the presence of edges, corners, parallels, OTV's and shadows (Fig. 5).

The negative evidence of support includes the presence of lines crossing any side of a parallelogram, existence of L-junctions or T-junctions in any side of a parallelogram, existence of overlapping gaps on opposite sides of a parallelogram, and displacement between four sides of a parallelogram and its corresponding edges support. Negative evidence is as important as

We construct a feature hierarchy which encodes the structural relationships specific to oblique views of rectangular shapes, presumably corresponding to the visible roof surfaces: Lines, skewed parallels, skewed U-contours, and skewed rectangles or parallelograms. Fig. 3 shows the viewing and sun angles involved. We expect that images from aerial scenes have a camera model and other data associated with them from which these angles can be derived. Next, we describe the hierarchy of features in our system:

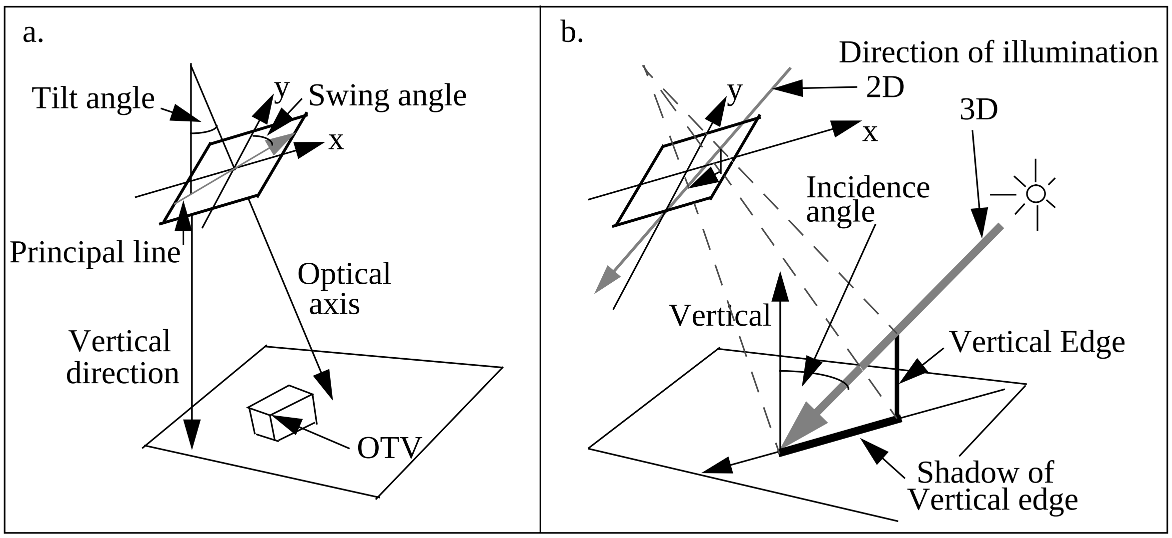

Fig. 3: (a) 3-D Viewpoint angles. (b) Sun angles and shadow geometry

Lines and Junctions

A group of close parallel lines represent a linear structure at a higher granularity level than the edges (see the common boundary between the building wings in Fig. 1a.) The resulting lines have a length and an orientation derived from the contributing elements. Fig. 4a shows the lines obtained from grouping the segments in Fig. 1b. We use these lines to detect L-junctions and T-junctions, also shown in Fig. 4a. For oblique views we also look for evidence of vertical edges in the immediate neighborhood of the L and T-junctions, thus allowing us to detect potential OTV's (Orthogonal Trihedral Vertices.)

Parallels and Skewed U-structures

Structures in urban scenes like buildings, roads and parking lots are often organized in regular grid-like patterns. These structures are composed of parallel sides. Thus, for each significant line-structure detected in the scene, there is not one but many lines parallel to it. For each line, we find lines that are parallel and satisfy a number of reasonable constraints. Note that the formation of a parallel structure also aids in the formation of new lines, as they suggest extension and contraction of the parallels to achieve full skewed overlap.

When the two lines in a parallel structure have their ends aligned as a function of the viewing angles, they strongly suggest the presence of a line with which the parallel structure would form a skewed U-structure. Even if the third line does not exist in the set of lines, we hypothesize it and generate the U-structure.

use them to help validate roof hypotheses. We believe that this approach leads to many fewer hypotheses than would be generated by a complete contour tracing scheme.

Our approach combines several of the techniques from previous work. Our perceptual grouping approach comes from the stereo analysis work described in (Mohan, Nevatia 1989). Here we use monocular cues and a very different hypotheses selection technique. Our shadow analysis method is an extension of the approach first described in (Huertas, 1983, Huertas, Nevatia, 1988, Huertas et al., 1993.)

Fig. 2 shows the components in our system. The system uses the line segments approximating the intensity boundaries to compute linear structures junctions among them. A hierarchy of features is constructed, leading to the formation of parallelogram hypotheses. These consist of instances of shapes that potentially correspond to building roofs and parts of building roofs. Next, promising hypotheses are selected and verified to correspond to roofs of buildings. Shadow and information, if available, is used to help form, select and verify hypotheses. Evidence of building walls in oblique views is also used to aid verification and to help derive estimates of the height of the structures, leading to a 3-D description of the scene.

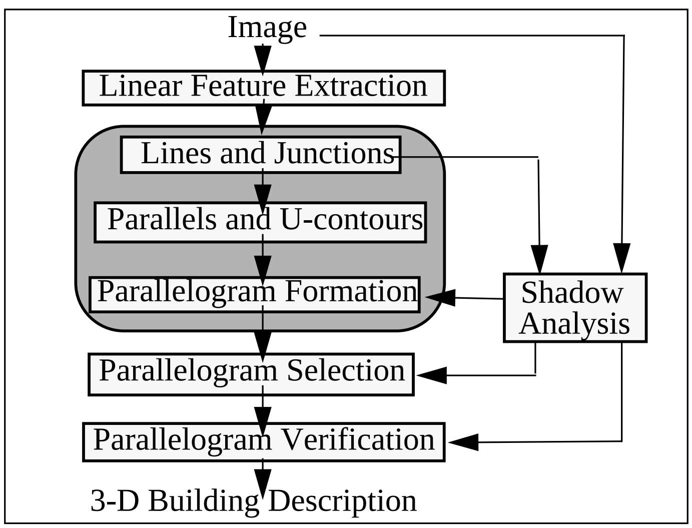

Fig. 2: Block Diagram of the System.

Our design philosophy has been to make only those decisions that can be made confidently at each level. Thus, we choose to generate as many hypotheses as seem feasible at the first level. Our selection process too is conservative and favors keeping hypotheses that may be viable. The verification process has the most global information and can make stronger decisions. Even here, if our system is to be embedded in a larger system, some of the decisions would be deferred to that system where more context is available for decision making.

The technique we describe in this paper, we believe, significantly extends the range of scenes that can be analyzed though many problems remain. Lastly, we show examples of scenes with many buildings, provided by the U.S. Government sponsored RADIUS program to demonstrate the effectiveness of our technique.

2 Generation of Hypotheses

The process of hypotheses formation is similar to the one described in (Mohan, Nevatia, 1989) with the appropriate extensions to oblique views and the use of strong shadow cues.

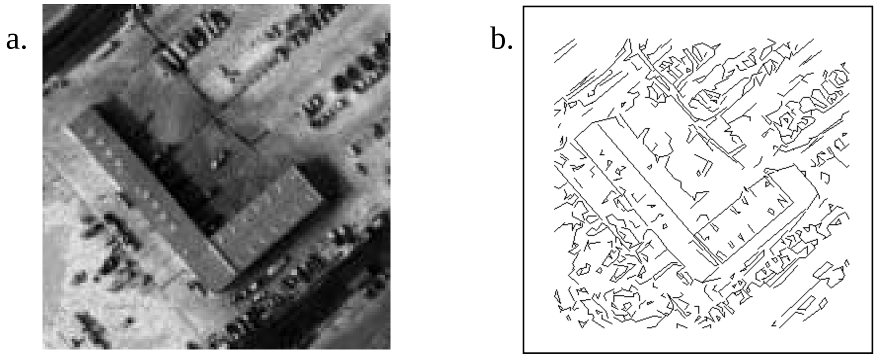

Fig. 1: (a) A building from Ft. Hood, Texas. (b) Line segments extracted from the image.

view as a running example, for simplicity. The building is easy for humans to see and describe, but it is difficult for computer vision systems. Fig. 1b shows the line segments detected in the image, using LINEAR, our linear feature extraction software (Nevatia, Babu, 1980, Canny, 1986). We are still able to see the roof structures of the buildings readily and easily, but the complexity of the task now becomes more apparent. The building boundary is fragmented, there are gaps and missing segments. There are also many extraneous boundaries caused by other structures in the scene.

There have been many previous attempts to solve this problem (Huertas, 1983, Huertas, Nevatia, 1988, Mohan, Nevatia, 1989, Huertas et al. 1993, Irvin, McKeown 1989, Liow, Pavlidis, 1990, Venkateswar, Chellappa, 1990). Building detection requires robust segmentation techniques and methods to infer the 3-D structure. These methods rely on edges or regions extracted from the image. Simple edge-based methods attempt to collect linked edge curves into the desired object boundaries, and succeed only for relatively simple scenes. Some edge-based methods have used some form of a contour tracing technique, see for example (Huertas, 1983, Huertas, Nevatia, 1988, Venkateswar, Chellappa, 1990). These are essentially local techniques that must make a decision of which path to trace at each local junction. Of course, all paths could be traced using backtracking but then the search space may become prohibitively large. Region based techniques construct closed curves that often do not correspond to the objects of interest.

Model based techniques can deal with fragmentation but require a-priori shape models. These systems have shown interesting performance but on limited examples. None of these systems can generate a description of the buildings at the level of shape descriptions of the different wings.

We have proposed, instead, to use a perceptual grouping approach. Cultural features such as buildings represent structures that are not random but have specific geometric properties. In this we restrict the shapes of buildings to be a single or a composition of rectangular parallelepipeds (thus allowing L, T and I shapes for example).

Previous systems have assumed that the viewpoint is more or less overhead. The system described here uses the viewpoint angles to deal with images acquired from an oblique viewpoint. The geometric constraints relevant to shape take into consideration, as a function of the viewpoint angles, the expected skewness of the rectangular surfaces that most buildings are expected to have. This property is used to organize the detected line segments into roof hypotheses. While the visible building sides (walls) can be hypothesized similarly, we only

Detection of Buildings from Monocular Images*

Chungan Lin, Andres Huertas and Ramakant Nevatia

Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273, U.S.A.
Ph.:(213) 740-6428, Fax: (213) 740-7877
e-mail: {chungan,huertas,nevatia}@iris.usc.edu

Abstract

We describe a system for detection and description of buildings in aerial scenes. This is a difficult task as the aerial images contain a variety of objects. Low-level segmentation processes give highly fragmented segments due to a number of reasons. We use a perceptual grouping approach to collect these fragments and discard those that come from other sources. We use shape properties of the buildings for this. We use shadows and walls to help form and verify the hypotheses generated by the grouping process. This latter step also provides 3-D descriptions of the buildings. Our system has been tested on a number of examples and is able to work with overhead or oblique views.

1 Introduction

The goal of this work is to detect and describe buildings from monocular views of arbitrary aerial scenes. This is a difficult but important task for many applications such as photo-interpretation, cartography and surveillance. Building detection is difficult for several reasons. The contrast between the roof of a building and surrounding structures such as curbs, parking lots, and walkways can be low. The contrast between the roofs of various wings, typically made of the same material, may be even lower. Low contrast alone is likely to cause low-level segmentation to be fragmented. In addition, small structures on the roof and objects, such as cars and trees adjacent to the building will cause further fragmentation and give rise to "noise" boundaries. Roofs may also have markings on them caused by dirt or variations in material. Shadows and other surface markings on the roof cause similar problems.

There are other characteristics of these images which may cause problems. Roofs have raised borders which sometimes cast shadows on the roof. This results in multiple close parallel edges along the roof boundaries and often these edges are broken and disjoint. At roof corners and at junctions of two roofs, multiple lines meet leading to a number of corners making it difficult to choose a corner for tracking. A roof cast a shadow along its side and often there are objects on the ground such as grass, trees, trucks, pavement, etc., which lead to changes in the contrast along the roof sides.

Consider the building in Fig. 1a (from a scene of Ft. Hood in Texas.) We use an overhead

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