TEC-0111

Research in Model-Based Change Detection and Site Model Updating

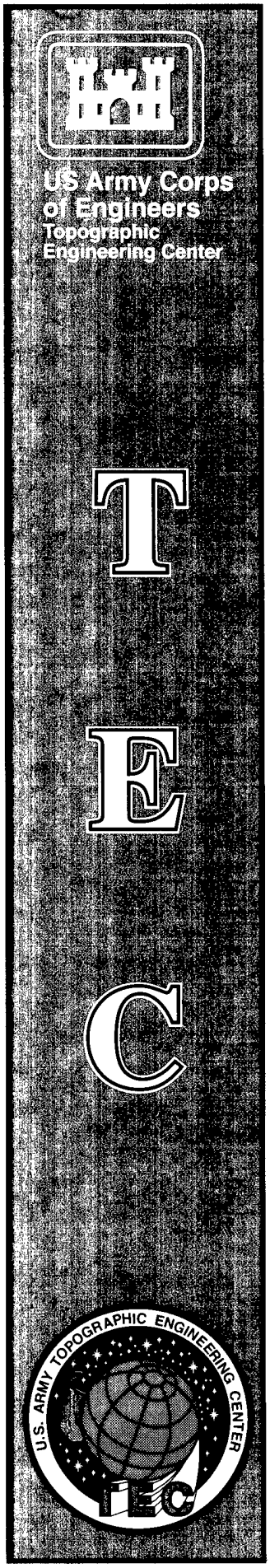
R. Nevatia

University of Southern California
Institute for Robotics and Intelligent Systems
Powell Hall 204
Los Angeles, CA 90089-0273

August 1998

Prepared for:
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road
Alexandria, VA 22315-3864



Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1998	Second Annual July 1994 - July 1995

4. TITLE AND SUBTITLE

Research in Model-Based Change Detection and Site Model Updating

5. FUNDING NUMBERS

DACA76-93-C-0014

6. AUTHOR(S)

R. Nevatia

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern California
Institute for Robotics and Intelligent Systems
Powell Hall 204
Los Angeles, CA 90089-0273

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

U.S. Army Topographic Engineering Center
7701 Telegraph Road, Alexandria, VA 22315-3864

19. SPONSORING / MONITORING AGENCY REPORT NUMBER

TEC-0111

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The major goal of the RADIUS project is to provide tools to assist an image analyst in the analysis of large amounts of imagery. The site models, an essential component of the analysis process, are useful in a variety of ways: projecting the expected structures from the current view point, assisting with the task of change detection, and cueing an analyst to the appropriate parts of an image. Our major RADIUS project deals with change detection and site model updating with particular focus on stationary structures. Some of these techniques also are applicable to automatic site modeling and some of our change detection techniques may apply to detection of larger mobile objects, such as airplanes. We have implemented an interactive modeling system that works in conjunction with our automatic system to minimize the need for tedious interaction.

14. SUBJECT TERMS

Change Detection, Model-to-Image Registration, Matching, Perceptual Grouping, Model Construction

15. NUMBER OF PAGES

61

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE **PAGE**

LIST OF TABLES

TABLE **PAGE**

Preface

This research is sponsored by the Defense Advanced Research Projects Agency (DARPA) and monitored by the U.S. Army Topographic Engineering Center (TEC) under contract DACA76-93-C-0014, titled "Research in Model-Based Change Detection and Model Updating." The DARPA Program Managers were Dr. Oscar Firschein and Dr. Tom Strat, and the TEC Contracting Officer's Representative was Ms. Lauretta Williams.

1 Introduction

This report describes the activities during the period of 28 July, 1994 to 27 July, 1995, the second year of our current three-year effort. The primary focus of this work has been change detection and site model updating. Methods have been developed for detecting changes to fixed structures, such as buildings, and we have studied how they may be modified to function with large mobile objects, such as airplanes. We have continued to develop automated methods for building detection and description, using either monocular or multiple images. These techniques are needed for automated site model construction and for model updating. We have in addition developed a method for interacting with the automatic site modeling system that requires minimal interaction from a human user. These projects are briefly described below; details are given in the following sections.

1.1 Change Detection

Figure 1.1 shows a flowchart of the complete change detection system. It contains five major steps:

- *Site Model to Image Registration*: The first step in change detection is to register the new image(s) to the model(s) contained in the site folder. The system has some capability for performing coarse registration, however, this information is expected to be available from other modules being developed by other contractors under the RADIUS program. The system uses feature matching [1] to compensate globally for translational errors and brings the site model into close correspondence with the observed image.
- *Site Model Validation*: This step verifies the presence of the model objects in the image. A confidence value is computed for each object in the model based on the match information from the previous step. Lower confidence values are likely to represent possible changes to the objects.
- *Change Detection:* In this step we analyze, in more detail, possible changes in the site indicated in the previous step, and determine if the missing correspondences can be explained by techniques that draw attention to significant structures in the image that are not explained by the existing model. The task of finding objects in the image that are not in the model is more difficult, and will require use of *perceptual grouping* operations. Currently the system can detect missing buildings and changes in dimensions.

- *Site Model Updating:* In this step the changes are modeled and incorporated in the new site model.
- *Event Analysis:* In this step the structures indicated by the change detection processes are analyzed in detail. This step requires the development of automated or semi-automated site modeling techniques.

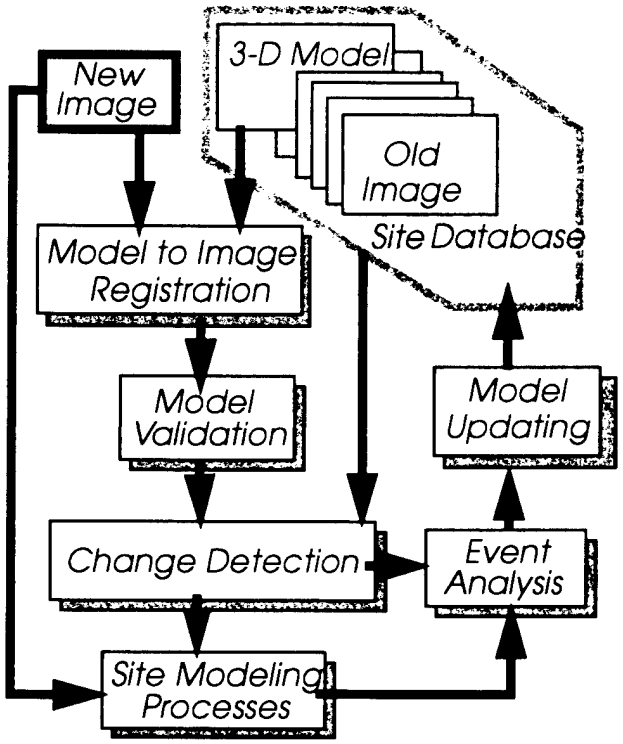


Figure 1.1 Flowchart of change detection system.

Our previous annual report [2] and [3] gave details on the development of a validation system that included fine registration followed by a simple object-by-object verification scheme. The scheme only measured validation by counting the number of object elements matched to image features. During the past year we have continued testing the model registration and validation system. We have made improvements in the validation technique and developed a system to perform preliminary change detection in building structures. A new validation process was incorporated to handle occlusion of objects by other objects, and to calculate confidence values associated with the validation of each object in the model. The confidence values are the result of analyzing the matching elements between the model and the image, and give the initial "clues" of where changes might have occurred. Low confidence values, for instance, may indicate a missing building, may reflect inaccuracies in the model, may result from coincidental alignments, or may indicate actual changes to the structures. Details are given in Section 2. This system, which uses the fast block interpolation projection (FBIP) camera model, is written in LISP (LISt Processor) and runs under

the Radius Common Development Environment (RCDE) [4]. It has been tested favorably with operational imagery at SRI International.

The validation system has been tested for the purpose of verifying the presence of aircraft at a site. The suggested method involves the derivation of simple aircraft models from one or more images, rather than using CAD models. Preliminary results are shown using images of camouflaged aircraft. These results assume that the pose of the aircraft and the sun angles are known *a priori*. Details are given in Section 3.

1.2 Automated Building Detection and Description

We have continued the work in automatic building detection and description. This ability is needed for reliable change detection and site model updating, and is useful for initial site model construction. Two systems are under development: one uses a single intensity image and another uses multiple images. It is, of course, easier to detect and describe buildings using multiple images, however, the ability to at least reliably detect buildings from a single image is needed during the change detection process.

Good progress has been made on both systems. The monocular system now uses both shadows and walls for verification of a building, and for estimating heights. It has been tested extensively on the modelboard images with good results. These are presented in Section 4. We expect to test with the newly available Fort Hood, Texas images in future work. The system using multiple images is in the earlier stages of development. The system uses a hierarchical grouping and matching methodology. The preliminary results are encouraging and we believe this method will lead to robust and reliable building detection and description. This system is described in Section 6.

1.3 Interaction with Automatic Model Construction Systems

Another area of progress, described in Section 5, deals with user interaction with the automated systems to assist the building detection systems in completion of the modeling task. The general idea consists of identifying areas, or cases, where the automated systems fail. The user, by means of a graphical interface, directs the automated systems to make use of the partial results derived automatically. With user guidance, the automated system attemps to complete the task of model construction with minimal user input. This system has been tested in conjunction with our monocular building detection system with encouraging results.

2 Change Detection in Permanent Structures

We have continued the development of a validation mechanism that implements the first step towards a system for detecting changes in images of aerial scenes. Validation seeks to confirm the presence of model objects in the image. An overview of the change detection process was given in Section 1. The following provides details of the various steps.

2.1 Site Model to Image Registration

The first step is to register the site model to an image. Normally, coarse registration should be available from other modules of the RADIUS program (such as a "Model Supported Positioning" module). Our system has the capability to correct translational errors. Our registration method ([2],[3]) consists of the following tasks:

- Calculation of misregistration offsets and compensation for translational errors,
- Establishment of correspondences between the elements of objects in the model, and the supporting features extracted from the image.

The first task is carried out by a matching technique [1] that uses line segments derived from the site model objects, and line segments [5] approximated from the edges extracted [6] from the image. The second task uses the registration offsets to select the matching pairs (model segment, image segment) that bring the model objects and the image features into correspondence.

Figure 2.1 shows an example of the registration step in the system. The site model shown in Figure 2.1a is projected according to the camera model associated with the image. The peak in the matcher accumulator array (Figure 2.1b) gives the global misregistration error. The fine-registered model' (Figure 2.1c) is then used to establish the context needed for further processing. Details may be found in [2] or [3].

2.2 Model Validation

The purpose of model validation is to verify that model objects are present in the image. The system uses the correspondences established in the registration step to assign a confidence value to each object in the model to reflect its image support, and to help select object candidates to analyze likely changes in the site. Some features will be missing because of viewing conditions. These, however, can be predicted and

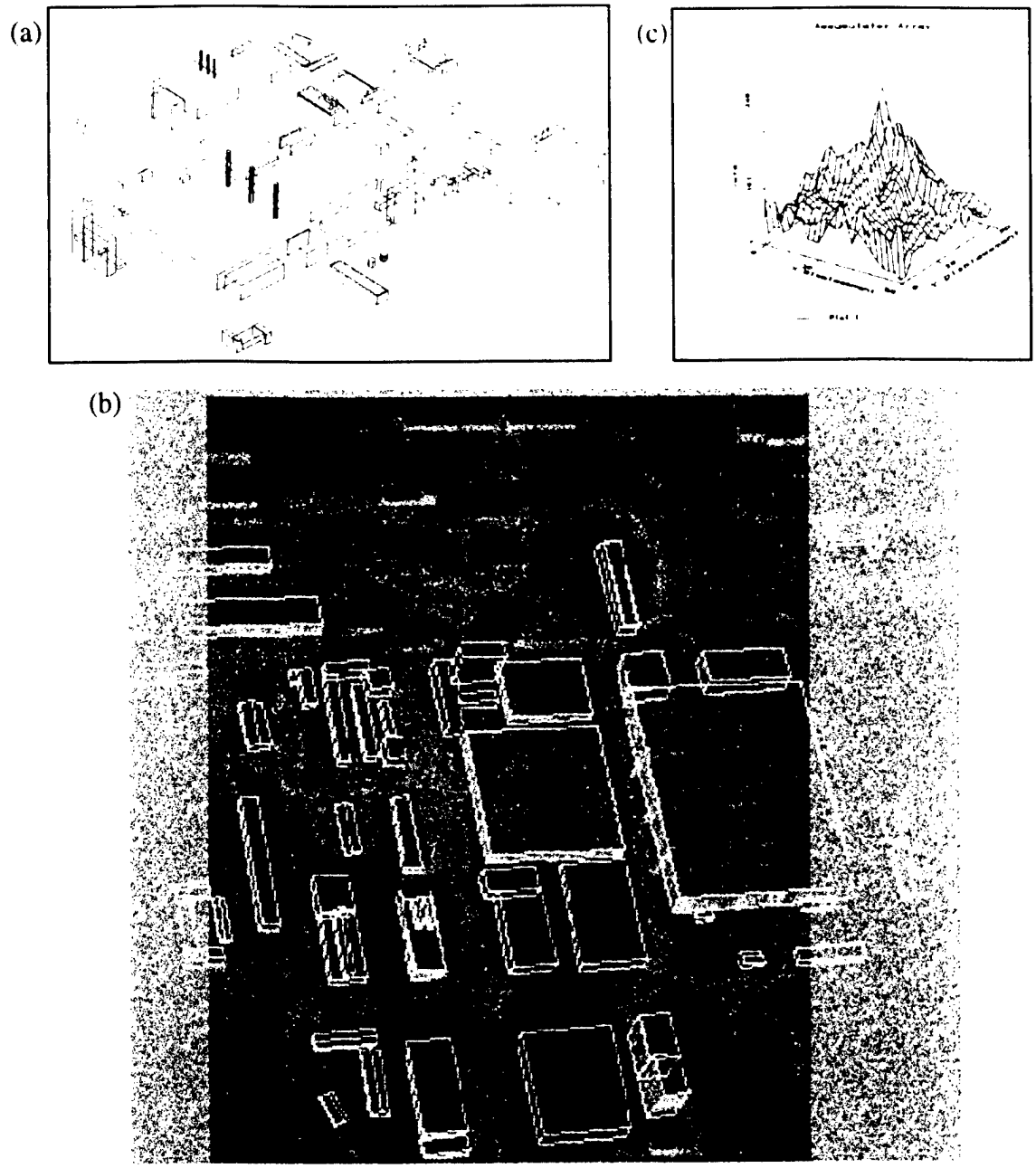


Figure 2.1 Refined model-to-image registration.

explained from the site model itself. The system, at this stage, also deals with ambiguities, such as multiple matches and coincidental alignments.

The confidence values derived are based on the following measures:

Object Presence

Each object model consists of a number of edges representing its boundaries. Object presence denotes how many of these boundaries have a corresponding segment or segments in the image (see Figure 2.2). Currently, object presence is measured as a percentage of model edges matched to correspond to image edges. This quantity

takes into account only visible elements, from the particular viewpoint of the image. Both self-occlusion and occlusion by other objects are determined using the range image derived from the model itself, thus non-visible elements are not counted.

Object presence is calculated separately for **roof** elements, **vertical** elements, and **base** elements to allow us to study the relative importance of these components as a function of the viewpoint. Near nadir views, for example, may highlight the contribution of the roof elements. These weights may be set by annotations in the site model; currently, they are given equal weight.

Object Coverage

Object coverage is equivalent to length-weighted object presence. It denotes the percentage of the perimeters of the matched boundaries of the faces of the objects. These quantities represent the amount of boundary evidence detected in the image in support of the validation of a model object. Figure 2.2a shows an object with all sides represented by small supports. Figure 2.2b shows the opposite; a few sides represented with good support. Object coverage measurements take into account occlusion, and are calculated separately for roof, vertical, and base elements.

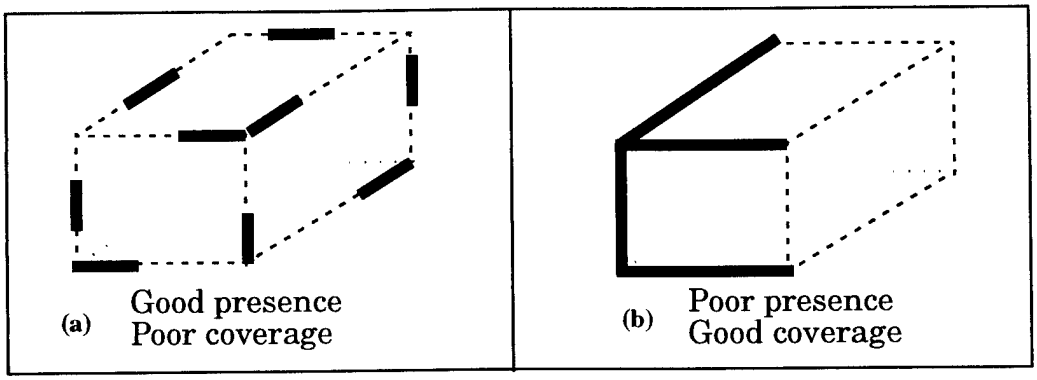


Figure 2.2 Presence and coverage.

Shadow Presence

Shadow presence is inferred from the models and verified in the image. The model information is used to project the shadow boundaries, taking into account their visibility. Note that in situations where reasonable object matches (correspondences) are not found, the absence of shadows helps confirm the absence of the building, but the presence of shadows does not guarantee the presence of a building. The final interpretation is the subject of our current and future work.

The number of shadow elements (boundaries and junctions) derived from the model (see Figure 2.3) is compared with the number of potential shadow elements extracted from the image to give the shadow presence measure. The image segments are labelled as potential shadow segments by noting the consistency of the "dark" side of the segment with respect to the direction of illumination. Segments oriented parallel to the direction of illumination also correspond to possible shadow lines cast by

vertical object edges. Shadow junctions are detected similarly. The L-junctions formed (allowing for gaps) by potential shadow lines are labeled potential shadow junctions. Details on the shadow labeling of segments and junctions may be found in [7] and [8].

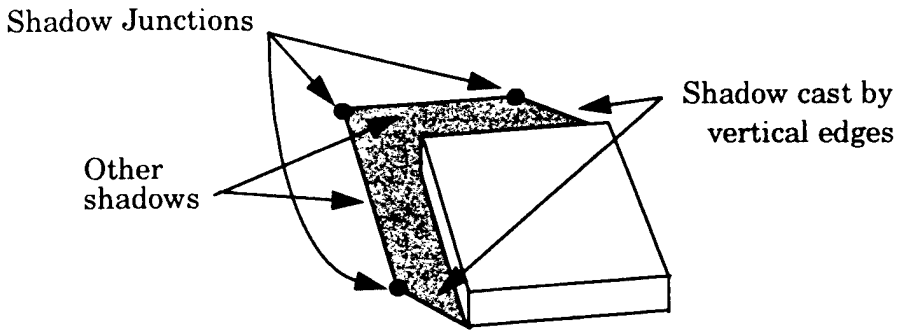


Figure 2.3 Shadows cast by "cubic" building.

Object presence and coverage, and shadow presence are currently combined linearly to give a confidence value interpreted as follows as follows:

Table 1: Confidence Levels.

Percent	> 75%	> 50%	> 40%	> 25%	< 25%
Confidence	Very high	High	Medium	Low	Very low
Color code	Green	Blue	Yellow	Pink	Red

The interpretation of these confidence values drives the change detection procedures that follow. In this report, we describe our progress in dealing with building structures only. Figure 2.4 shows an example of the registration/validation step applied to one of the modelboard 1 images. The colors indicate the confidence level associated with each building structure.

2.3 Change Detection

The confidence values computed in the previous step give the first indication, for each object, of potential changes in the site. High values indicate close correspondence between model and image. Low values signify possible changes to the site. In some cases, however, high values are caused by multiple matches and other ambiguities that may exaggerate or reduce image support for an object. These conditions, however, are isolated. In order to distinguish between apparent and actual changes we first perform an analysis of possible ambiguities and correct the confidence values appropriately. Details are given in the following.

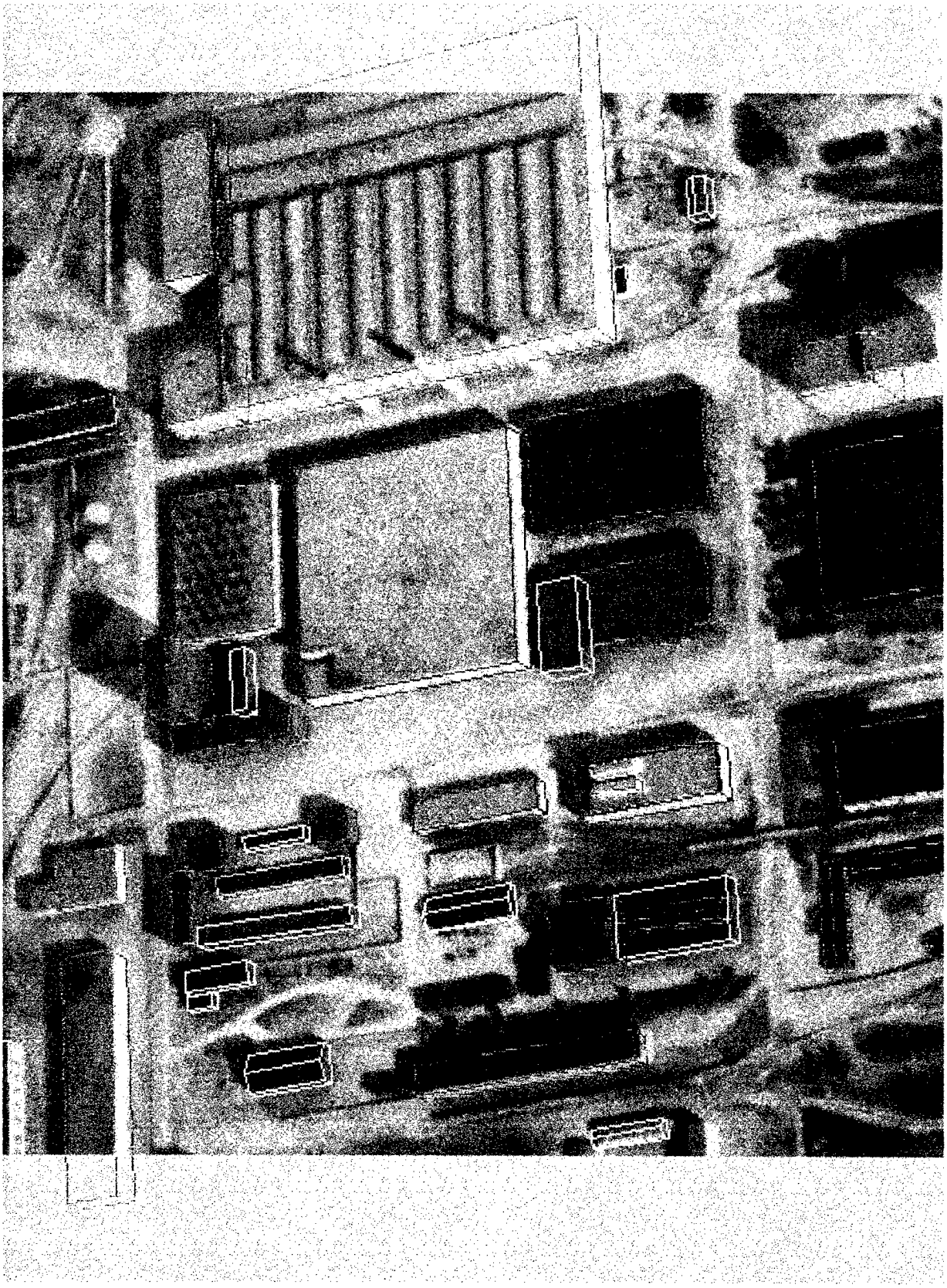
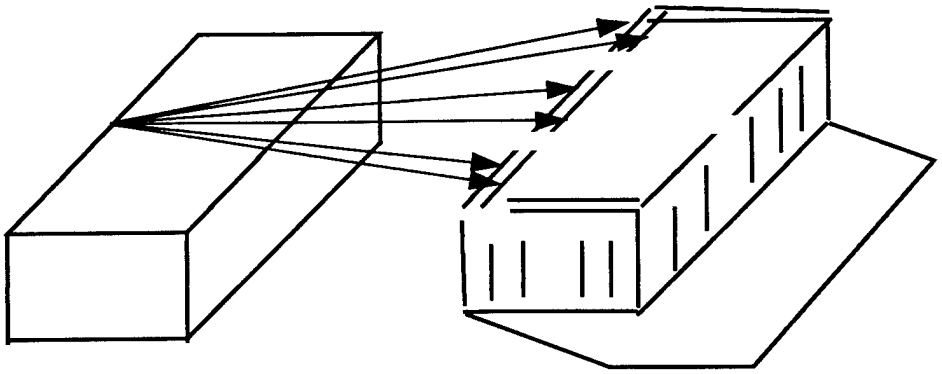


Figure 2.4 Validation results and color-coded confidence levels.

2.3.1 Analysis of Ambiguities

Multiple and Insufficient Matches

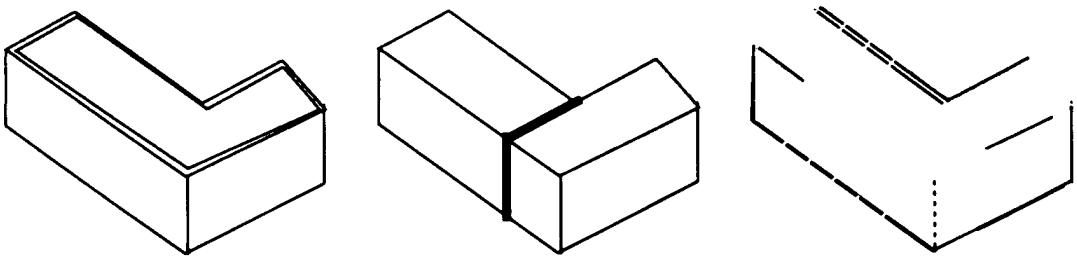
The model-to-image matcher in the system corresponds each model element with one or more image elements. This is necessary in order to deal with expected fragmentation in the image elements. Fragmentation is caused by inadequacies in the feature extraction process and due to actual image content, such as trees occluding buildings, or road boundaries and shadows. This may result in some individual model segments being associated with multiple image building boundaries (Figure 2.5) or with boundaries of other nearby objects. This condition is detected by observing the object coverage measures described above and is handled in the following manner: If the multiple matches include colinear image segments, these are currently taken together. If the multiple matches involve parallel image segments, the one with the closest fit to the model segment is taken to represent the matched boundary (see example below.)



(a) Model segments (b) Image segments

Figure 2.5 Model to image correspondence.

In some cases complex objects are modeled in terms of simpler shapes, thus, may include some elements that do not correspond to physical elements. Figure 2.6 shows an L-shaped building that has been modeled by two rectangle parallelepipeds. The thick lines on the building model do not correspond to physical boundaries, and are impossible to match. The lack of image support results in lower confidence. Figure 2.7 shows two buildings that are likely to be undermodeled because of their complexity. These are likely to require additional search strategies that are designed to look for additional evidence, such as a large number of vertical or horizontal boundaries. The system is not currently capable of determining these conditions, thus, the confidence values may be underestimated. It is assumed that some of these conditions may require annotations in the site model to help the system adjust the weights used to determine confidence values.



(a) Building (b) Building Model (c) Image Segments

Figure 2.6 Impossible match because of overmodeling.

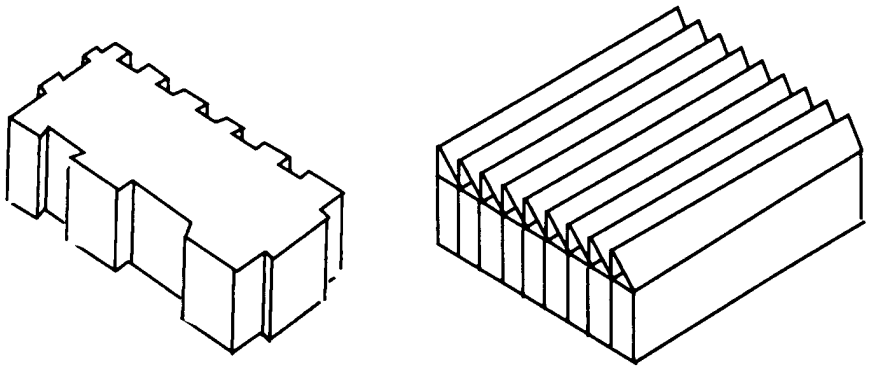


Figure 2.7 Complex buildings may be undermodeled.

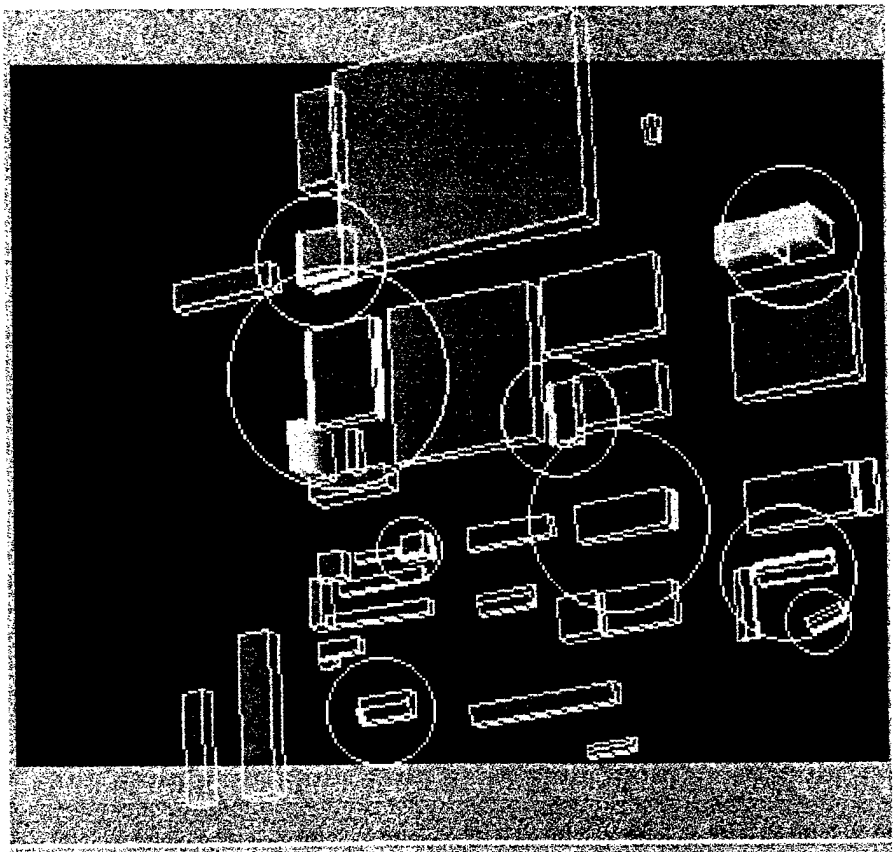
Next, an example from the modelboard is used to illustrate our previous discussion, and helps explain the remaining conditions that the system can handle currently. Figure 2.8a shows the model segments. The model elements that might have changed are shown as thick lines. A number of possible changes are denoted by circles on the structures. The corresponding image segments are shown as thick lines in Figure 2.8b. In Figure 2.9, the thick black and white lines denote ambiguous multiple matches. After resolution of the ambiguity, the white lines denote the image segments chosen to correspond to model segments.

Coincidental Alignments

Some of the multiple matches described in the previous section are caused by coincidental alignments of buildings with other structures. Some of these include roads, and adjacent objects. Nearby objects and shadows sometimes result in image features that have a larger extent than that predicted by the model features. These are explained by examining nearby shadows with knowledge of the direction of illumination, and by examining adjacent structures.

The building on the top right of Figure 2.9 has a vertical edge aligned with the shadow cast by the same edge. Both edges in the image, the vertical edge and its

(a)



(b)

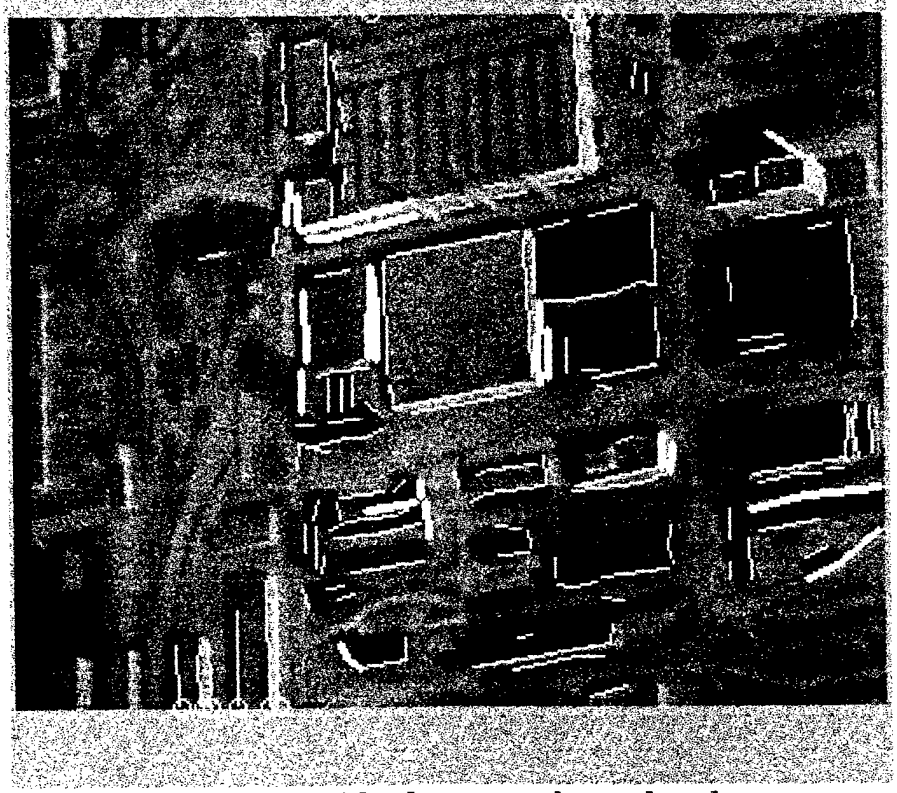


Figure 2.8 Possible changes to be explored.

shadow are good candidates to match the model's vertical edge. The multiple match may indicate an increase in height, but in this case, the situation is identified correctly as a coincidental alignment. The white portion of the edge is then determined to be the portion corresponding to the model edge.

Figure 2.9 Ambiguity because of multiple matches and alignment.

Coincidental alignments caused by nearby and adjacent structures are determined by locating adjacent structures that help explain a possible change. The small building on the top of Figure 2.9 helps illustrate this point. The model roof and base edges are matched to much longer lines in the image. Figure 2.10 shows two buildings (white boundaries) that were found to explain the situation detected, thus dismissing the possibility of determining a change in the small building's (black boundaries) horizontal dimensions.

In this particular example, all possible changes are explained by resolving ambiguities in the matching process, and by detecting coincidental alignments with shadows or nearby structures, therefore, no changes are reported.

2.3.2 Changes in the Site

Changed Objects

Changes in the dimensions of the structures located in the image that are not caused by errors or coincidental alignment signify real change. The changes in dimensions detected by the current system are preliminary in the sense that they are not fully described. The system reports the possibility of change without a full de-

Figure 2.10 Adjacent buildings may introduce ambiguity.

scription. A final determination of change requires that the entire object geometry be analyzed for consistency in view of the possible change. This is one of the subjects of our future work.

Figure 2.11 shows an example, also from the modelboard image set. The models of the two buildings were altered by hand (reduced in size) to the dimensions illustrated by the thin white lines. The matching and fine registration step correctly registers the modified models to the structures in the image. The thick white lines are the image segments that matched the corresponding model edges. The differences then denote the extent of the change found at this preliminary stage.

Figure 2.12 shows a building wing that has been added to an existing structure. The portion of the building in the model is correctly registered to the image. The two thick white lines denote the extent of the match. Because the object presence measure for the roof of this structure indicates that all four sides of the current model were matched, the change is labeled "added" wing.

Missing Buildings

Figure 2.13 shows a large number of object models (in white) added by hand to the site model. The size and location of these objects were determined randomly and added deliberately to the site model to test for "missing" object capability. Note that in spite of the added information, the "legitimate" models are correctly registered with the image, as shown by the black lines. The low confidence values calculated for the added building models indicate that there is no image evidence to support the

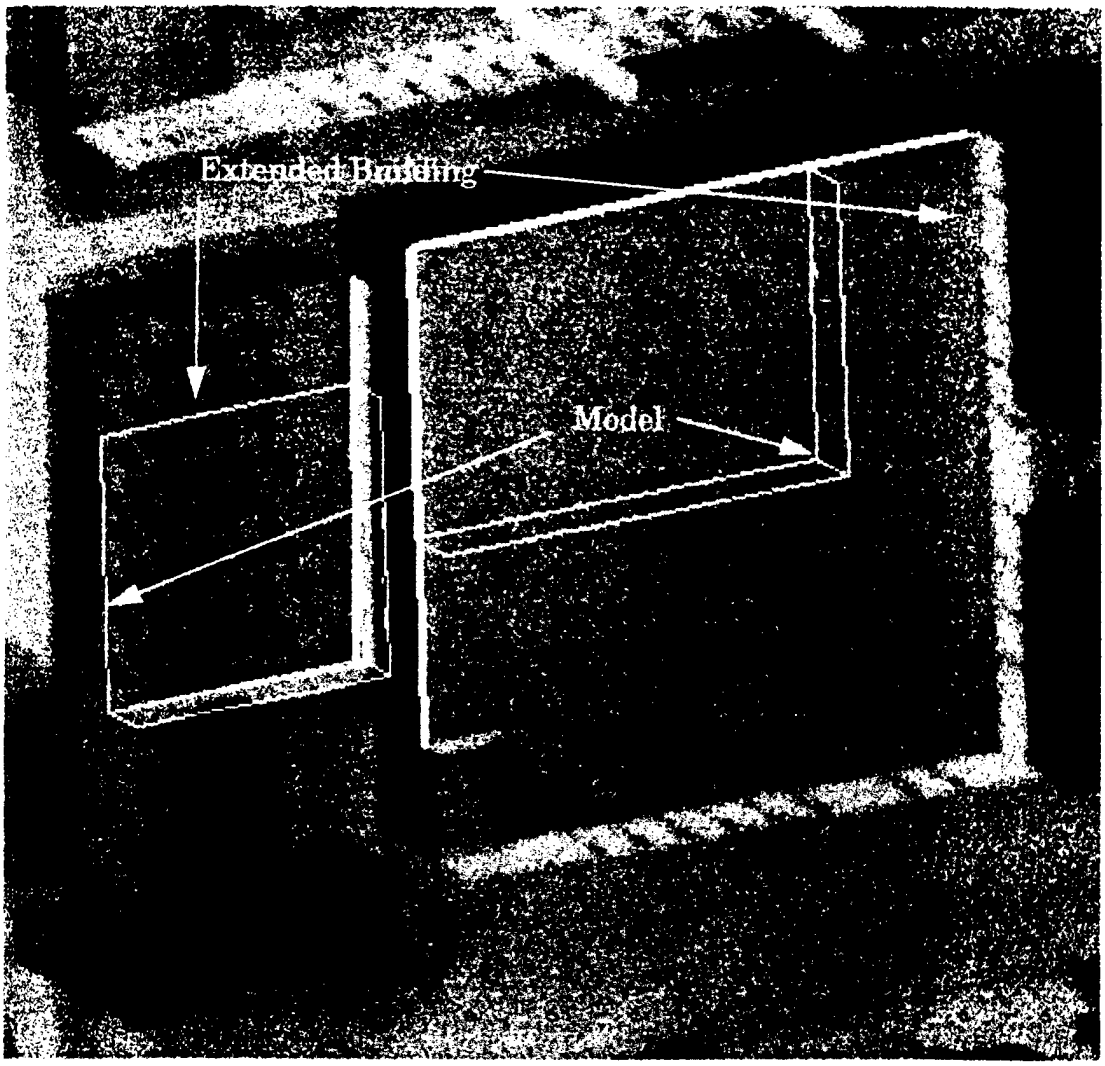


Figure 2.11 Actual change in dimensions.

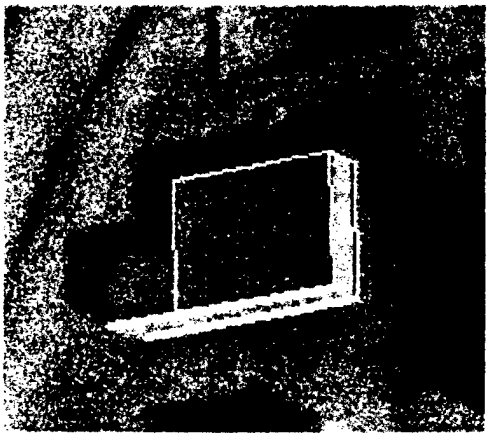


Figure 2.12 Added "wing" is reported in this case.

presence of a building at that location. The two possible causes for this condition are that either the model is incorrect or the building has been removed or destroyed (assuming that images are of sufficient quality). Resolving these ambiguities may require examination of these locations in other images.

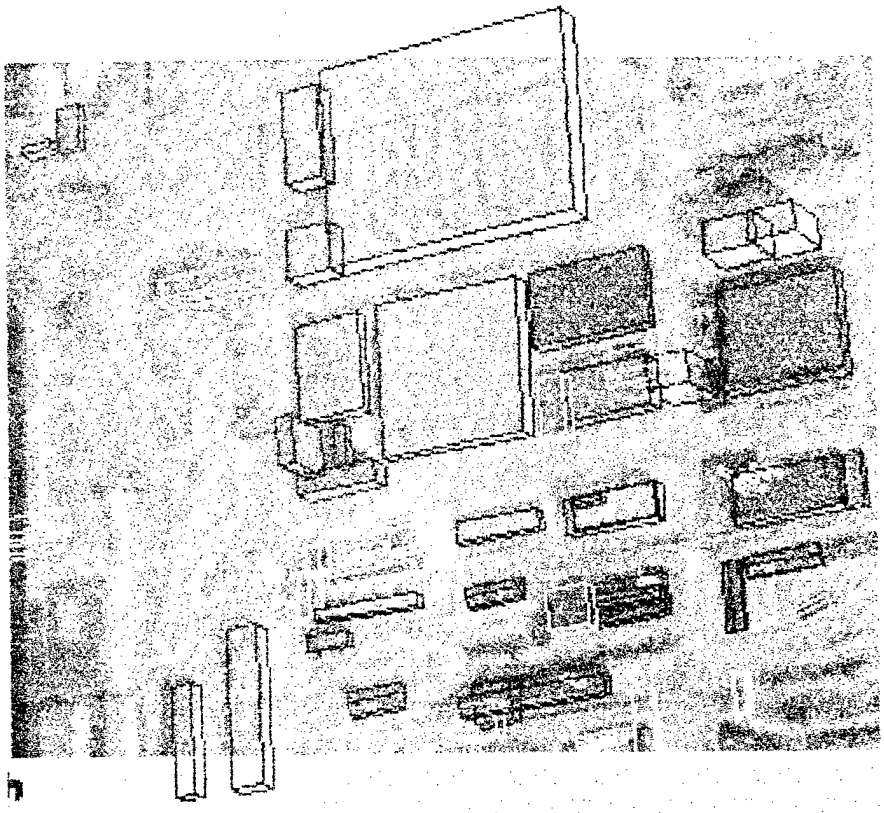


Figure 2.13 Missing buildings because of large change or model error.

2.4 Technology Transfer and Future Work

The model validation software has been ported to SRI in Menlo Park, CA for testing on operational imagery. Preliminary results are promising.

The current system operates in the 2-D domain of projected model structures onto the image viewpoint. We plan to explore the use of the verification mechanism by matching 3-D model features to 3-D features from multiple images or from a range sensor such as IFSAR. Our immediate work will concentrate in giving detailed descriptions of detected changes to building structures, and to other structures of a permanent nature, such as roads and other transportation network objects.

One important type of site change is the introduction of new structures. We plan to incorporate techniques to detect evidence of construction. Together with our capa-

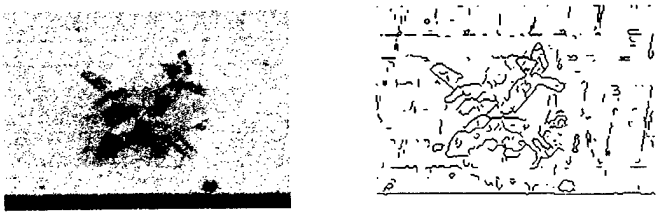
bilities to construct models automatically, we can then proceed to suggest new additions to the site model.

3 Verification of Aircraft Presence

Verifying the presence of aircraft or other mobile objects in expected locations is important for several analysis tasks. This section presents our progress towards a recognition technique based on low-level matching between segments in the images and segments of the projection of a 3-D model of the aircraft. The model is constructed manually from one or more views of the scene. The matching technique is the same as that used for fine registration of the site model with a new image (described in the previous section). The quality of the match is evaluated to determine verified presence.

Aircraft recognition techniques have been reported using a variety of methods. See Subhoved et. al. [9] for example. This elaborate system claims to be capable of detecting aircraft in real-world scenarios that include haze, clutter, and shadows. This system has been reported to be under development and uses a hierarchy of aircraft, models. The model database includes generic aircraft, aircraft classes, specific aircraft and detailed aspects of specific aircraft. The aircraft models consist of two types: edge-based approximations of CAD models, and generalized cylinder-based models. The system we present here "recognizes" aircraft by matching 2-D projections of simple user-derived 3-D models to linear features extracted from the image. In our system, an aircraft is decomposed into its main discernible components: two wings, with possibly two or more engines, the fuselage, two rear wings, and a tail; each of these is described in terms of geometric properties.

The methodology consists of grouping primitives extracted from the image into sets that resemble the chosen geometric properties. These groups represent hypotheses of instances of the objects in the image that are verified by an evaluation criteria. Our approach deals with the expected fragmentation of features caused by poor image quality, cloud cover, noise, clutter, and camouflage. A typical example is shown in Figure 3.1, the image of a camouflaged C-130 transport (a), and the edges [6] extracted from the image (b); it clearly demonstrates the difficulties.



(a) Image (b) Line segments

Figure 3.1 Camouflaged C-130 aircraft.

In some cases, 3-D models of aircraft may be available. We assume, however, that in general they are not, and suggest a mechanism to derive a model that is sufficient for the task. The use of simplified 3-D models of aircraft derived manually from the images available is suggested. Verification can then proceed similarly as it does for buildings.

It is assumed that a camera model is available, and that the sun angles are known in order to make use of the shadow clues available. The system does not currently have a mechanism to estimate the pose, or a range of aircraft poses. To carry out the experiments, the orientation and the position of the aircraft are specified by selecting two points on the aircraft, such as the two extremities of the wing's leading edge. Given the pose of the aircraft in the image, the model segments are projected onto the image to match line segments extracted from the image.

The system is written in LISP and runs under the RCDE [4] on a SUN workstation. The images and camera models used for preliminary testing were supplied by Dr. Joseph Mundy of General Electric Corp.

3.1 Construction of a Simplified Aircraft Model

The model is extracted by hand directly from one or more images. Two orthogonal 2-D planes are constructed; one outlining the wings and the fuselage, and the other, representing the tail (Figure 3.2).

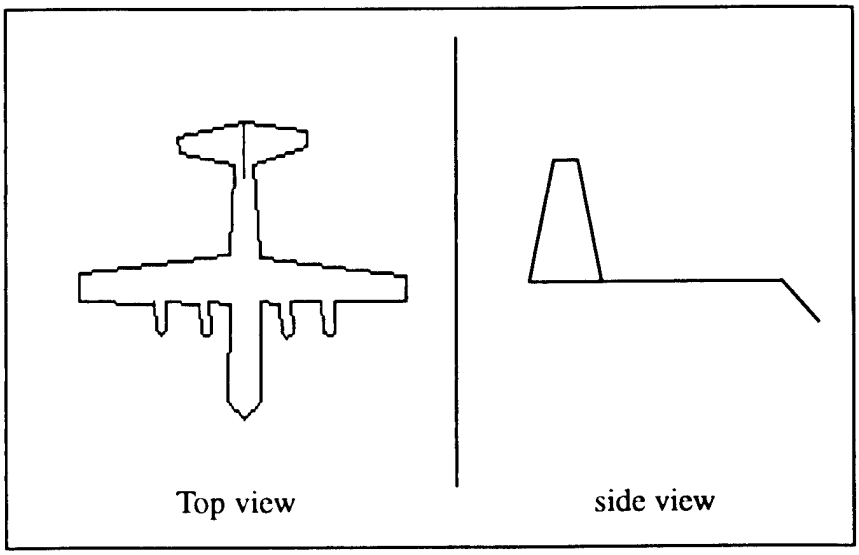


Figure 3.2 3-D simplified model.

The next step is to use the camera model to project the outlines of the two planes into two perpendicular planes in a 3-D coordinate system, using the known camera model. The ambiguity on the z coordinate is resolved by assuming that the aircraft is on the ground, thus, the wings are parallel to the ground plane and are at a given height.

3.2 Model Projection

3.2.1 Translation and Rotation

The system assumes that the aircraft pose (orientation) in the image is known in order to project the model onto the image (see Figure 3.3). In order to carry out the experiments, the orientation and position of the aircraft are specified by selecting, by hand, two points on the aircraft, typically the two extremities of the wing's leading edge.

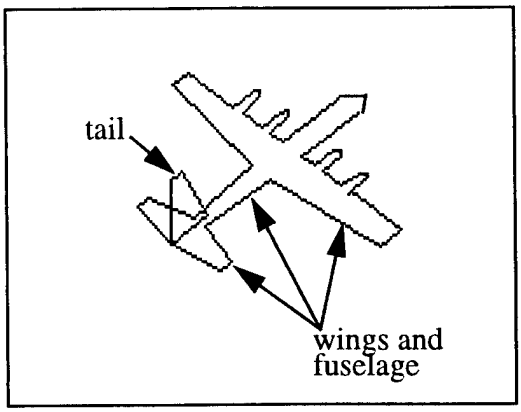


Figure 3.3 Projected 2-D outlines.

3.2.2 Shadow Processing

The shadows cast by 3-D objects are strong clues to the presence of objects, and we have used them extensively in the past. The shadow clues become significant, in particular, when the object appearance has been altered by camouflage. Shadow elements, as described in the previous section for buildings, are calculated from the aircraft model using the camera parameters and the sun angles. Occluded shadows are determined and removed by a simple method (Figure 3.4): The model consists of closed outlines; they form closed general polygons. Occluded segments on the outlines belong to the intersection of those polygons.

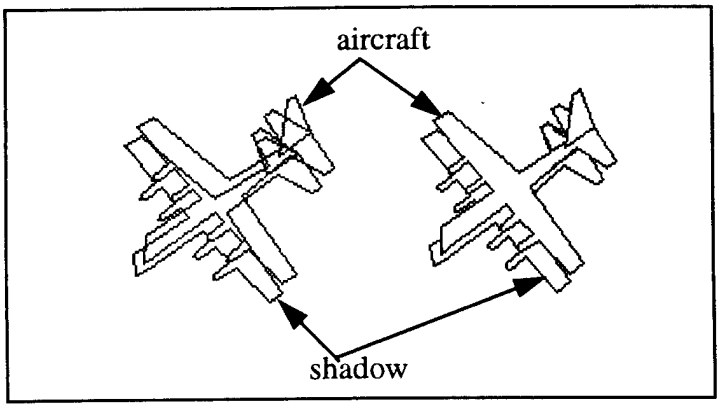


Figure 3.4 An aircraft model and its shadow.

3.3 Matching Algorithm

We use the same matching technique as the one described for building structures in Section 2. Additional details are given in [1], [2], and [3]. After matching, the strength of the match is verified. The following criteria is used: If the rate of matched segments between the model and the image is above 90 percent, then the model is validated at this position and orientation. Otherwise, further validation and evaluation are required. The next section gives the details.

3.4 Validation and Evaluation

The matching algorithm is a global procedure and finds the best translation vector between the model and the image segments regardless of the image content. It is not sufficient, therefore, to require a certain percentage of matched model segments to say that the model is validated and the aircraft recognized. We analyze the results of the match at a higher level to determine the accuracy of the recognition of the model.

The criteria to determine the presence of an aircraft is as follows: the matched image segments have to be well distributed geometrically over the model, i.e., each part of the aircraft wings, tail, etc., must have approximately the same proportion of matched segments in terms of arc length. This criteria is applied separately to the aircraft segments and to the shadow segments. If either the aircraft or its shadow is validated, then we say that the model is verified. Typically the shadows give a better rate of recognition when the aircraft has camouflage applied.

Method:

First, a binary function of the matched segments between image and model along the arc length of the model is computed: the model outline segments are scanned and each corresponding matched image segment is projected onto it. The abscissa maximum is the total arc length of the 2-D model. Then we scale this function modulo 2π in order to map this function onto a circle of radius 1, centered at (0,0). Each point belonging to the perimeter is a matched pixel (see Figure 3.5).

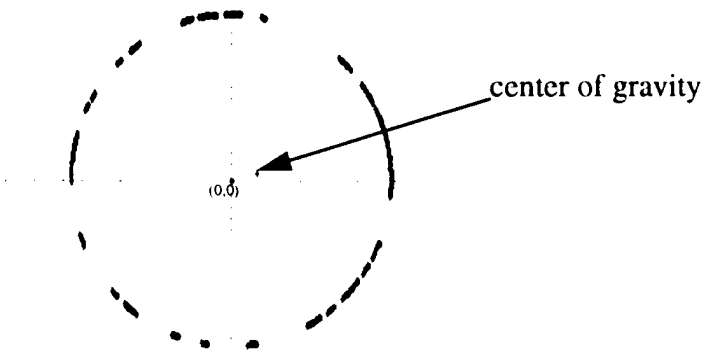


Figure 3.5 Circular distribution of matched pixels.

Second, we compute the moments of inertia of the resulting fragmented "wheel." We compare the distribution of matching pixels along the perimeter of this circle to the distribution of mass on a wheel where each point belonging to the perimeter has a weight contribution of 1.

We determine if this "wheel" is well balanced by computing second order moments of distribution:

$$m_{20} = \sum_C (x_i - x_0)^2$$
$$m_{02} = \sum_C (y_i - y_0)^2$$
$$m_{11} = \sum_C (x_i - x_0) \cdot (y_i - y_0)$$

$$\left.\begin{aligned} x_0 = \sum_C x_i = \sum_0^{2\Pi} r \cdot \cos\alpha \\ y_0 = \sum_C y_i = \sum_0^{2\Pi} r \cdot \sin\alpha \end{aligned}\right\} r \in [0,1]$$

The Hessian matrix represents the distribution of the points along the circle.

$$H = \begin{bmatrix} M_{20} & M_{11} \\ M_{11} & M_{02} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

The two conjugate eigenvalues (λ_1 and λ_2) of H give the two parameters of the distribution:

The eccentricity of the wheel is given by:

$$eccentricity = \lambda_1 / \lambda_2$$

The number of matched pixels (modulo 2π) is given by:

$$\begin{aligned} Length\ of\ match &= \lambda_1 + \lambda_2 \\ &= Tr(H) \end{aligned}$$

Finally, the displacement of the center of gravity gives the spread of matched pixels on the aircraft's outline:

$$displacement = \sqrt{\left(x_0^{\,2} + y_0^{\,2}\right)}$$

The three parameters: eccentricity (< 7%), normalized length of match (> 50%) and displacement of the center of gravity (< 20%), are used to validate the model. Low eccentricities compensate for shorter lengths of match.

Figure 3.6 shows a typical result. Both the aircraft and its shadow are well represented. Figure 3.7 shows an example of a "missing" aircraft.

3.5 Conclusion and Future Work

This method can be used to verify the presence of an aircraft at a given position, or a number aircraft in an image, if they have the same pose. An additional effort is required to complete automation of the process to include estimation of pose or a range of poses. Derivation of model projections from full CAD models may be incorporated. The use of full CAD models, however, is non-trivial as the models may be too detailed. We may need to process these with a "visibility" and a "sensor" model.

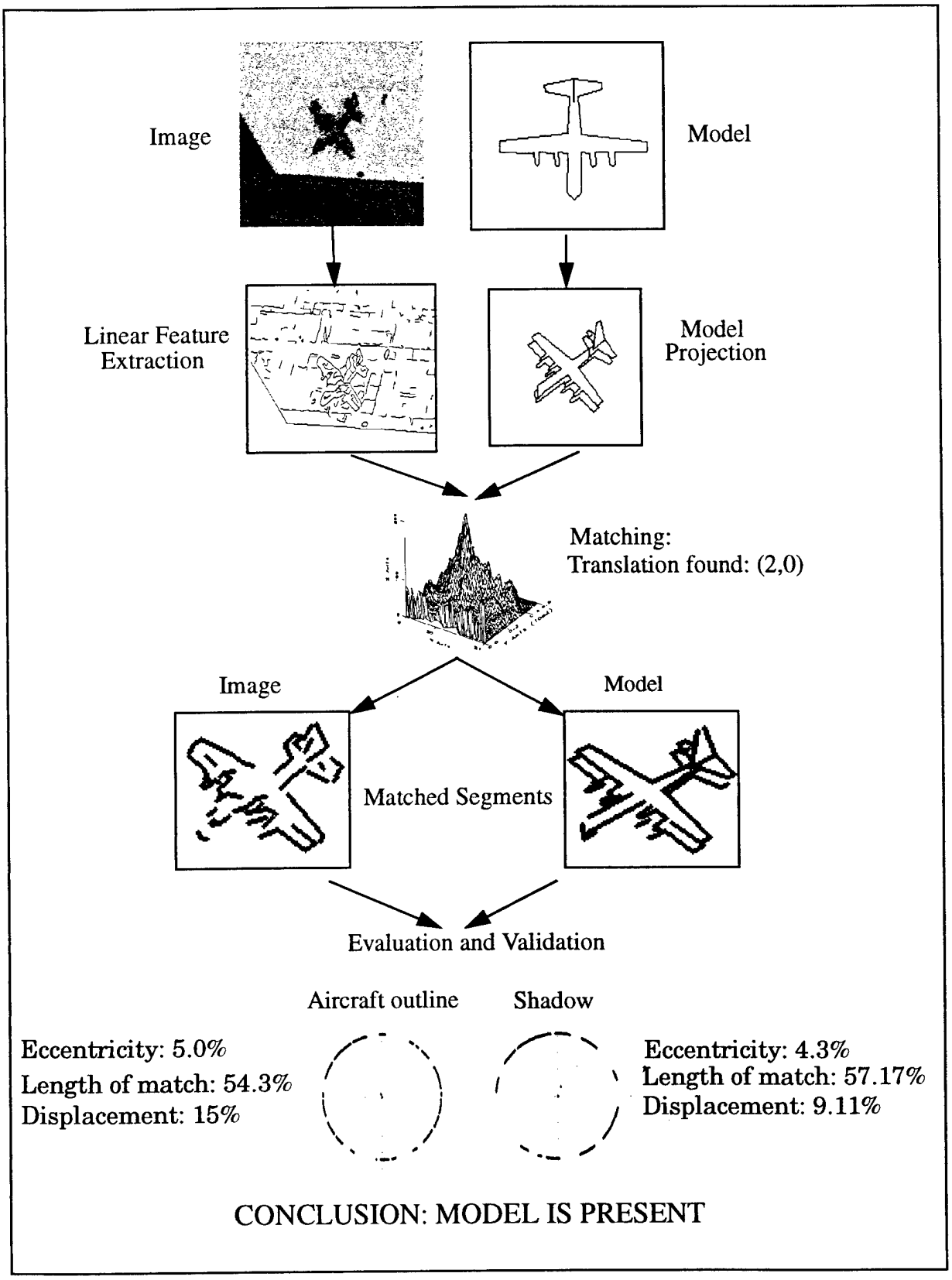


Figure 3.6 True positive example.

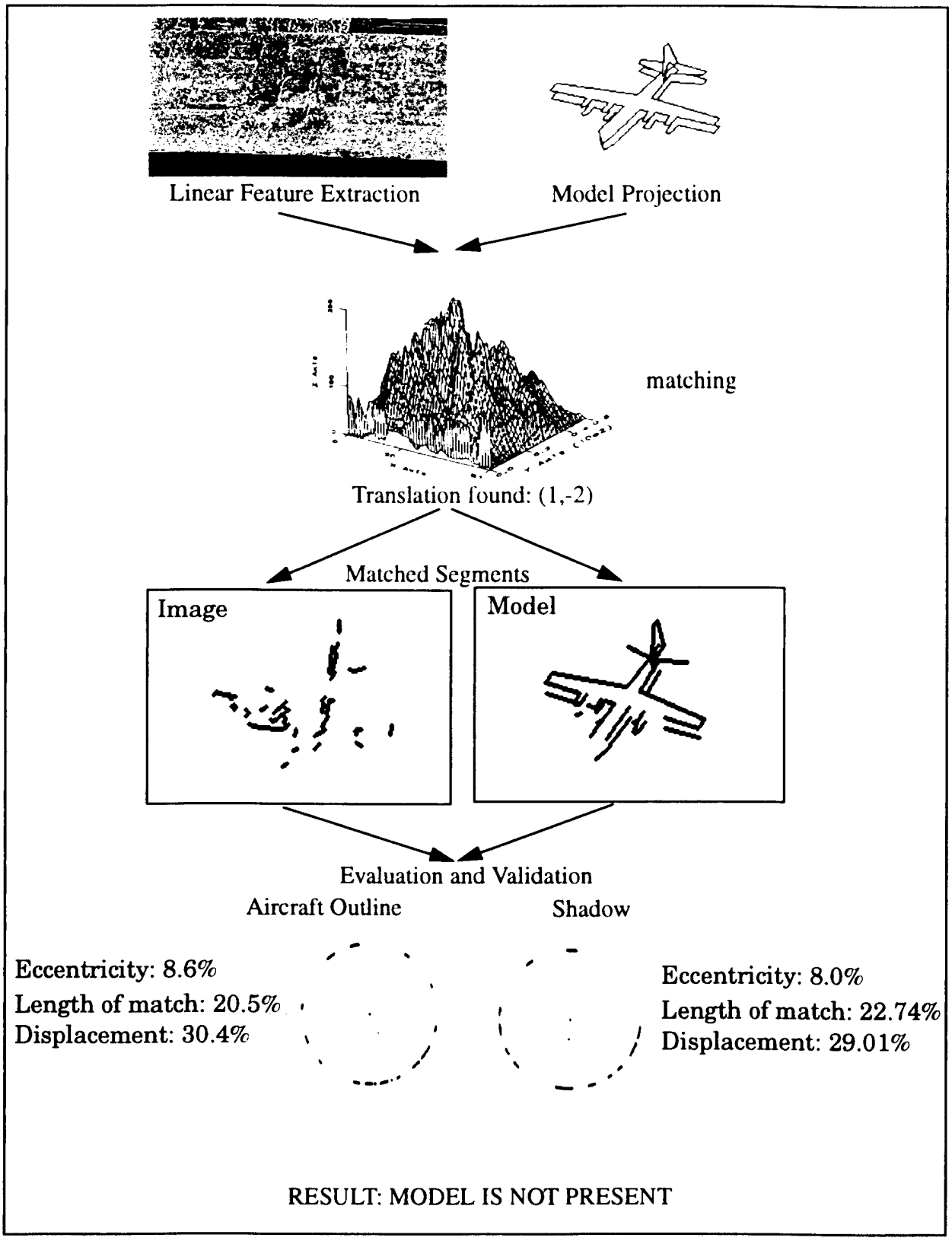


Figure 3.7 True negative example.

4 Building Detection from a Single View

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

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

In previous work ([2],[7]) we described a system that used a perceptual grouping technique to make roof hypotheses from the edges detected from the image. A selection process selects good hypotheses for verification, and shadow evidence is used to verify the selected hypotheses. The 3-D information is inferred from the shadow evidence.

A similar approach is used in the current system, however each step requires many changes to accommodate the problems introduced by the oblique view images. For the hypotheses generation process, the skewness of roof hypotheses has to be handled according to the viewpoints, and the selection process can make use of the 3-D cues such as orthogonal trihedral vertices (OTV). In addition to the shadow evidence, wall evidence is used to verify the hypotheses. The use of both shadow and wall evidence makes the verification process generate more assured results and make the system robust. The corresponding wall evidence of a building also provides another way to infer the 3-D information of the building.

This system makes the following assumptions: that buildings are rectilinear, that the roofs and the surface on which the shadow fall are flat, and that the viewing geometry (camera model) is known. It has been tested on several examples of the modelboard images. Testing has begun on the newly available Fort Hood images. Some results and performance evaluation are given in the following.

4.1 Generation of Hypotheses

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

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

The formation of parallelogram hypotheses is constrained by the following equation:

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

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

Angles α and β are shown in Figure 4.1. θ is the "swing" angle and γ is the "tilt" angle; these are derivable from a camera model.

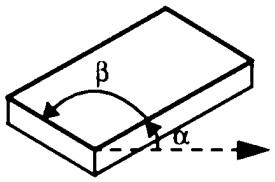


Figure 4.1 Angle constraint of roof hypotheses.

4.2 Selection of Hypotheses

A selection process is applied to choose hypotheses having strong evidence of support and having minimum conflict among them. Based on the local and global supporting evidence of hypotheses, a rule-based selection process selects promising

hypotheses for verification. This process greatly decreases the number of hypotheses to be verified, therefore reduces the run time of the time-consuming verification process.

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

4.3 Verification of Hypotheses

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

4.3.1 Shadow Verification Process

The use of shadow evidence to verify hypotheses is more complicated in oblique views than in nadir views, for the shadow may be occluded by the building itself in oblique view images. See Figure 4.2.

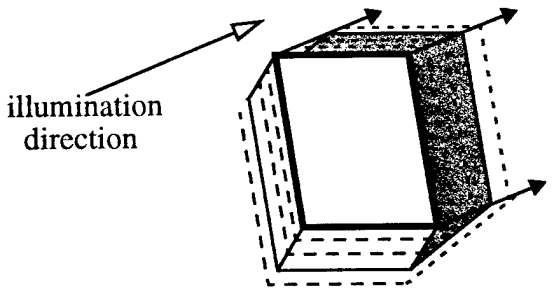


Figure 4.2 Search for shadow evidence.

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

There are a number of difficulties, however, that prevent the accurate establishment of correspondences. Building sides are usually surrounded by a variety of ob-

jects, such as loading ramps and docks, grass areas and sidewalks, trees, plants and shrubs, vehicles, and light and dark areas of various materials. Occlusion of the shadow by the building itself or by nearby buildings may make the shadow region irregular and make the shadow evidence difficult to extract. To deal with these problems we have adopted some geometric and projective constraints and special shadow features.

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

Given the sun and viewpoint angles, the projected shadow region in 2-D can be delineated with appropriate removal of the self occluded shadow region for a given building height. The shadow verification process collects all potential shadow evidence along the expected shadow boundary. For every possible building height, a set of corresponding shadow evidence is collected for evaluation. The range of possible building heights is determined by the knowledge of the maximum building height in the scene. Figure 4.2 shows how the system searches for shadow evidence on several possible building heights.

The shadow evidence associated with each possible building height is evaluated and a score is computed by 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.

4.3.2 Wall Verification Process

Some of the walls of a building should, in general, be visible in oblique view images. Finding wall boundaries provides evidence for the presence of a building. Given the viewing angles and a building height, we can estimate the expected wall boundary for a roof hypothesis. All evidence around the wall boundary is collected and a score is computed for the wall evidence.

Given a roof hypothesis and the viewing angles, the system determines which sides should be visible. The swing angle gives the vertical direction from which building sides are hypothesized. The wall boundary is delineated for a given building height and a search process is activated to collect all evidence around the delineated wall boundary. Figure 4.2 shows the search of wall evidence for several possible building heights.

The estimate of the wall evidence is a weighted sum of the evidence for ground-boundary, vertical-boundary, and corners.

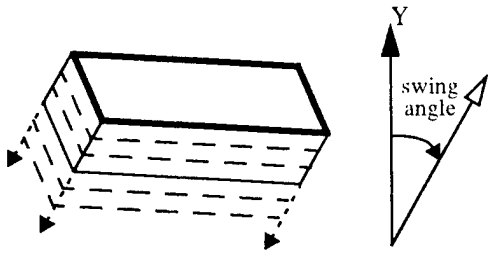


Figure 4.3 Search for Wall Evidence.

4.3.3 Combination of Shadow and Wall Evidence

For each building hypothesis, the previous two steps determine a shadow score and a wall score at every possible building height. The shadow score, S, and the wall score, W, are combined as follows:

$$Confidence = S + W - S \times W$$

$$\text{where} \quad 0 \leq S, W \leq 1$$

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

4.4 3-D Description of Buildings

In this system, the shadow and wall evidence is used not only for verification but for reconstruction of 3-D information (see Figure 4.4). The height of a building can be computed from the projected shadow width and the sun angles (the direction of illumination, the direction of shadow cast by a vertical line, and the sun incidence angle), or from the projected wall height and the viewing angles (the swing angle, and the tilt angle).

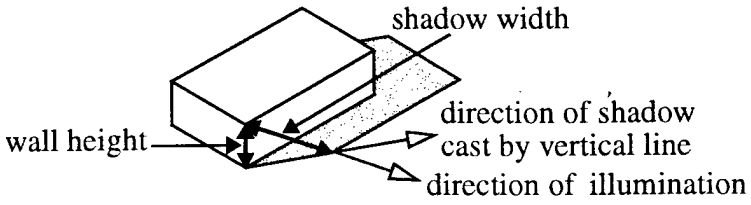


Figure 4.4 Three-dimensional model.

After the verification process, every verified hypothesis will have a height associated with it. From the height of the hypothesis the system can generate a description of the shape of the structure and derive a 3-D model.

4.5 Results and Evaluation

The system has been tested on a number of modelboard images. Some pictorial examples and a summary of results is given here. Figure 4.5 shows the result on an image (J19) from the RADIUS model board set containing a large number of structures (about 48). The system forms 2,247 hypotheses and selects 106. Of these, 29 are verified and all but two are correct (in conformity with the human judgement). The false positives are from very small and low contrast structures. The missing structures also are mostly very small and of very low contrast. We feel that the results are very good given the complexity of the image. Our system computes a confidence measure (not shown graphically), and the two false positives are of low confidence. The image is 1306x1034 pixels, and the processing time is about 20 minutes on a SUN Sparcstation 20.

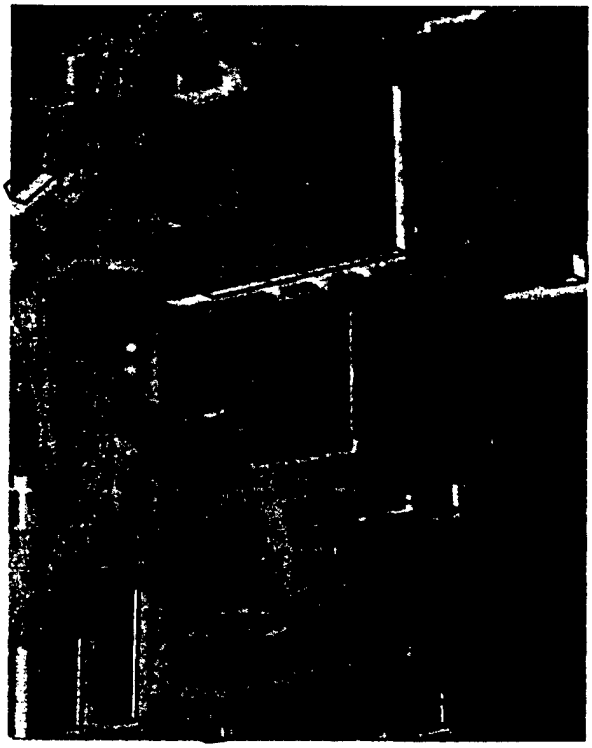


Figure 4.5 Model board (J19).

4.5.1 Detection Evaluation

There are many ways to measure the quality of the results [11][18]. We summarize performance on several images in Table 1 using the following four measurements:

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

The first two measurements are calculated by making a comparison of the manually detected buildings and the automated results [11], where True Positive (TP) is a building detected by both the human and program, False Positive (FP) is a building detected by the program but not human, and True Negative (TN) is a building detected by human and not the program.

The other two measurements are calculated as follows: Using the spatial extent of the buildings detected, we label every pixel in the image as either a building pixel or a background pixel [12]. "Correct Building Pixels", expressed as a percentage, is the ratio of the number of pixels correctly labeled as building pixels and the number of actual building pixels in the image. A similar measure for the background pixels is derived from the ratio the number of pixels correctly labeled as background pixels and the number of actual background pixels in the image.

Table 2: Detection Evaluation.

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

Table 2 summarizes the evaluation of results of the system on six model board images, all of the same site as shown in Figure 4.5, however, taken from different viewpoints and under different illumination conditions.

Note that the system gives rather consistent results for most images, except for J3, which corresponds to a nadir view. Also note that the measure for correct building pixels is considerably higher than for detection percentage indicating that the missed buildings are rather small. The number for correct background pixels is even higher indicating that false positives are rare and correspond to very small structures. We find that most errors of our system are associated with buildings with dark roofs where the boundary between the roof and the shadow is difficult to detect.

4.5.2 Confidence Evaluation

The system associates a confidence value with each hypothesis which can further be used to evaluate the performance of the system and guide a user on how to interpret the results. Figure 4.6 shows a histogram of the number of true and false

positives corresponding to certain confidence levels (ranging between 50 and 100, in increments of 5). Note that there are few false positives with high confidence values. In fact, if we set a confidence threshold of 75, we detect no false positives at all, and more than half of the true positives also are above this threshold. This indicates that the confidence values can be used profitably by an end-user or by another program. Results given with high confidence can be taken to be reliable and further attention for improving the results can focus on the lower confidence results, if necessary. We believe that this self-evaluation capability will greatly ease the use of our automatic tool in an interactive environment.

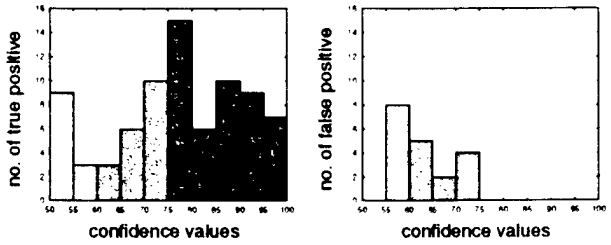
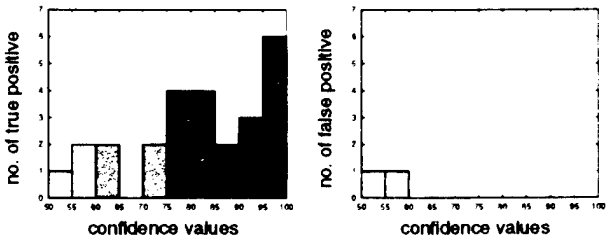
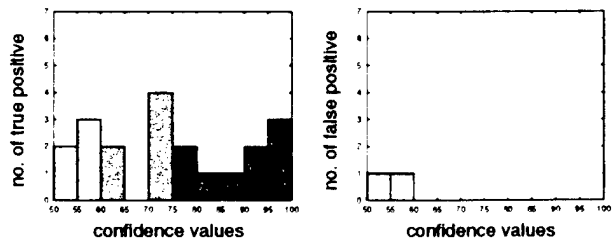


Figure 4.6 Distribution of confidence values.

Confidence analysis gives us a tool for evaluating the effectiveness of using various kinds of evidence. For example, on the J19 image shown in Figure 4.5, our system finds more true positives when the wall evidence is used. Moreover, if the wall evidence is used, the confidence of the correct hypotheses is increased substantially as shown in Figure 4.6 (the histogram of the true positives is skewed towards the higher confidence values). Now, if we set a threshold on the confidence values, the false positives can be eliminated while keeping most of the true positives. .



(a) Use Shadow and Wall Evidence in Verification



(b) Use Only Shadow Evidence in Verification

Figure 4.7 Advantage of using wall evidence

4.6 Conclusions and Future Work

We have described an automatic system for detection and description of buildings from oblique aerial images. We believe that the results show that the system gives good performance, particularly on large buildings with reasonable contrast and shadows. We believe that the confidence measures offer a tool that can help use the results even when they are not perfect. In future work, we plan to test extensively on real data, such as the images of the Fort Hood site. We plan to port the system to the RADIUS contractor for further evaluation and integration into the RTS.

5 Including Interaction in an Automated Modeling System

The results of the automatic building detection system described in the previous section ([2],[7]) are good, however, not perfect. We have developed tools for correcting some of the results by an interactive process such that manual intervention is minimized.

A variety of interactive systems have been built for site model construction ([4] [13]). The amount of interaction required of the human operator is typically of two kinds: Some systems require almost complete manual construction with an operator locating all the significant features. Others require the operator to select parametric models or rough outlines which are then fit to image data under operator control. In all such cases, the machine's task is limited to that of bookkeeping, simple geometric calculations, or some form of error minimization. No perceptual capability of the machine is used; and the operator is required to provide a large number of inputs, in some cases, very accurately. While such systems can aid in constructing site models from aerial images, they are quite tedious to use as the number of structures to be extracted is typically large.

We suggest an alternative strategy for combining the activities of the operator and the machine by taking advantage of what perceptual abilities a machine does have. Our goal is to provide a minimum amount of input to the machine and let the machine make the decisions that it can. Our approach is based on the observation that the automatic system often works quite reliably under certain conditions, and the operator should not need to do this work. Also, when the automatic system fails, it does so because of some salient difficulties. In such cases, the operator may be able to supply an indication of the difficulty or the desired result which may suffice for the machine to finish the computation. One such situation is when the building has a dark roof and the boundary of the roof with the shadow is not detected. In this case, the automatic system fails to confirm the presence of a building because of the lack of sufficient evidence. However, simple guidance from the operator, can indicate that a dark building is present in the vicinity, which suffices for the automated system to find one on its own.

The methodology allows for more detailed interaction with the system, in stages, and as necessary. In the worst case, the system reduces to the user having to provide

all the information as is the case for most manual systems. However, we find this capability is seldom needed in the system.

The design goals for the system can be summarized as follows:

- The complexity of the interaction process should be minimized, and in the worst case, should not exceed the complexity of the interaction process required by a manual system
- The type of information called up in each step should be easy for the user to determine.

"Easy" information for the user would be qualitative information without the need of precision, such as answering the question *"In the indicated area, is a building visible but not detected?."* The last requirement also could be stated as: the precision required by the user should be minimized.

5.1 Interacting with an Automated System

The approach combines aspects of the automatic system [7] with user interaction. Figure 5.1 shows the steps in building detection by the automatic system: The image (a) contains three buildings. The segments and junctions extracted (b) are used to form roof hypotheses (c). Promising hypotheses (d) are selected automatically for verification as described in [7]. The verified hypotheses (e) and the 3-D model (f) are computed automatically. After an image is processed automatically, the user interaction with the system starts. The process of interaction can be divided in two parts, *initial interaction* and *corrective interaction* (see Figure 5.2).

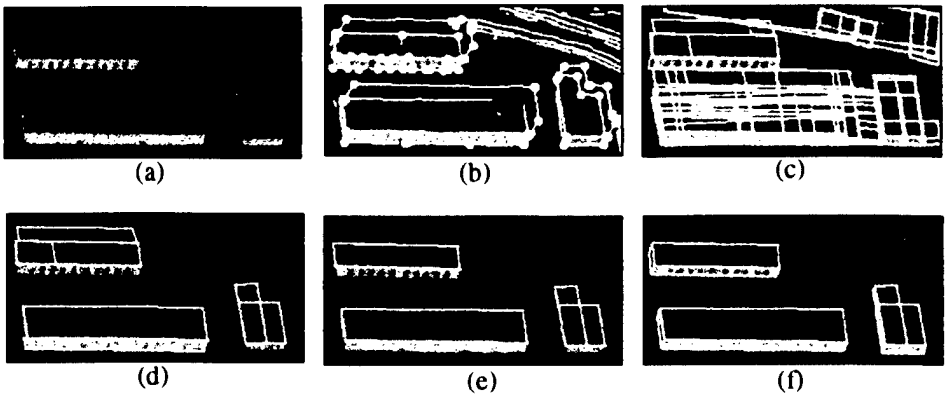


Figure 5.1 Automatic building detection.

5.1.1 Initial Interaction (qualitative)

First, the user classifies the detection problem. Classes of problems are, for example, dark areas, poor contrast, occluded buildings, occluded shadows, or partly detected L- or T-shaped buildings. This selection helps constrain the search for candidate hypotheses. While a particular situation may belong to more than one class of problems, it is unlikely that a correct hypotheses will be rejected as long as the classification is correct.

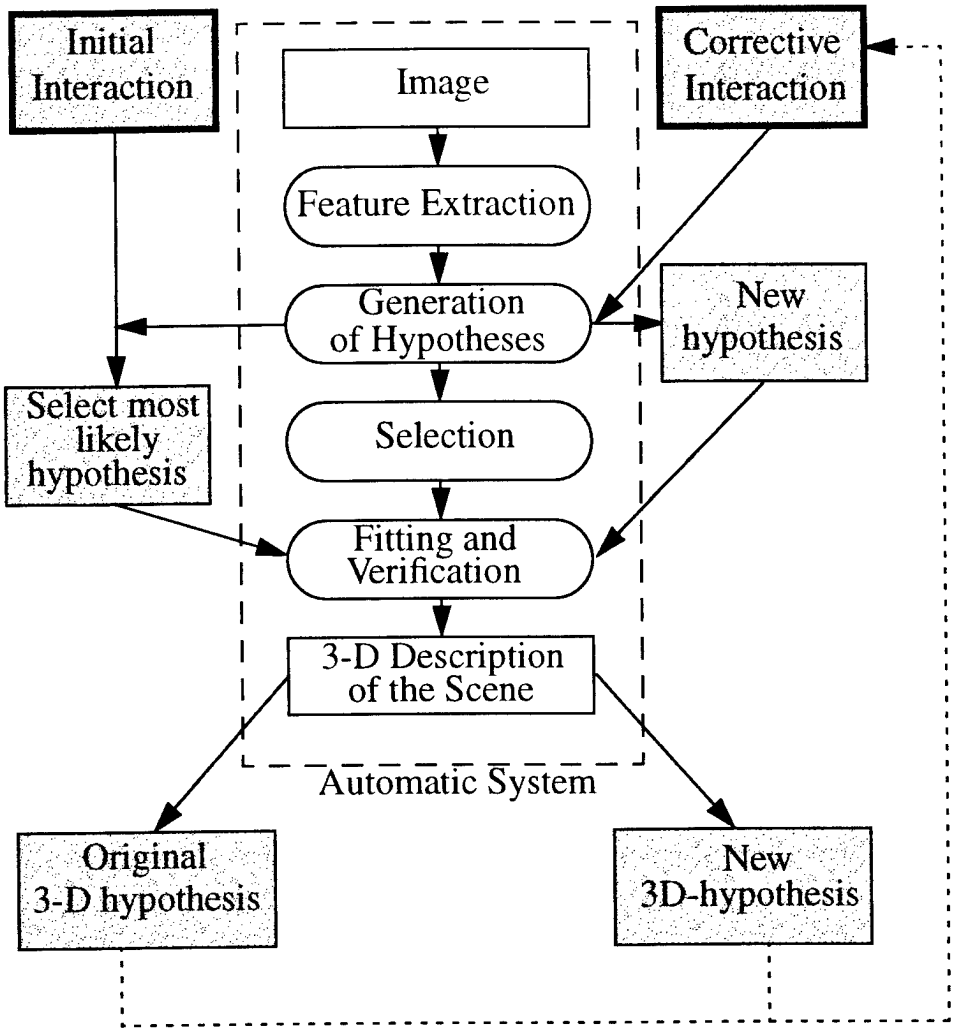


Figure 5.2 Interaction system embedded in the automatic system.

The second qualitative step consists of giving a rough localization of the missing building. This can be, for example, any point on the roof (it is possible to automate this step by clustering rejected hypotheses, see below). The initial interaction step results in the most likely hypothesis and can be established from the set of all hypotheses formed.

5.1.2 Corrective Interaction (quantitative)

If the hypothesis established in the first step is (partly) wrong, the user manually adjusts the sides or corners of the building model. For example, if one roof-side is incorrect, the user can either drag the line to the desired location or select an underlying image segment that best describes the location of the roof-side. After one or more adjustments, the verification and parallelogram fitting steps are activated automatically to recompute the building height and adjust the 3-D model. In the worst

case, the complexity of interaction is equivalent to that of adjusting the shape of a building model in a manual system.

5.2 Selecting the Most Likely Hypothesis

The input for this step is the result of the initial interaction and the set of roof hypotheses generated by perceptual organization. The initial interaction constrains the search in the set of hypotheses. According to the specified area, a local subset of hypothesized parallelograms is established, from which the most likely hypothesis, according to the detection problem, is computed. When no detection problem is specified and, therefore, no specific knowledge of the scene is known, the system uses the confidence values assigned to the hypotheses during the selection process of the automatic system.

A set of parallelogram patterns is assigned to each detection problem, which classifies the parallelogram hypotheses that can occur for a certain problem. An example of a pattern is a parallelogram, in which one roof side is wrong by a translation (because there were no edges detected at this roof side), and all other sides and the angles are correct (see Figure 5.3). Another example is a complete match of parallelogram and roof sides, which would lead to a correct guess after the initial interaction step.

This set of patterns has to be established by the designer of the system after an analysis of system failures.

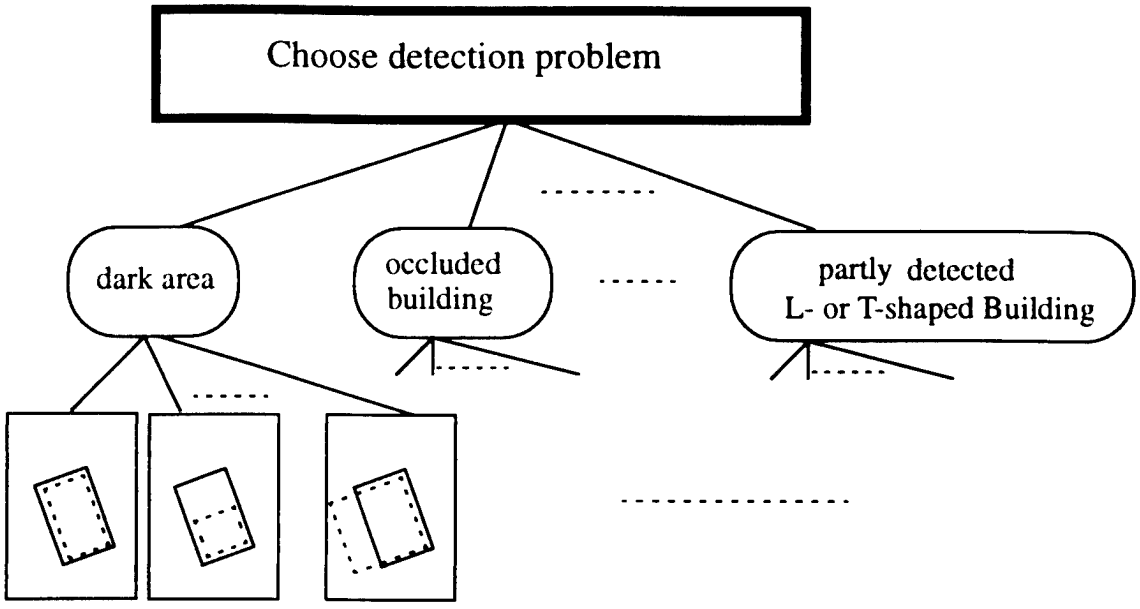


Figure 5.3 Classes of problems and their patterns.

Once a class of problems is selected, a probability for being the missed hypothesis is assigned to each parallelogram according to the set of patterns: observation x_i

for each pattern j is collected and transformed to a number ω_i which can be related to the associated likelihood. The observation can be represented either as a real number, an integer, or a boolean.

$$\omega_i = \frac{1}{2}\left(\frac{x_{ij} - \bar{x}_{ij}}{\sigma_{ij}}\right)^2 \quad \text{for real numbers, or}$$

$$\omega_i = -\ln P(x_{ij}) \quad \text{for integers/boolean}$$

These formulas are derived by assuming a Gaussian distribution. x_{ij}, σ_{ij} or $P(x_{ij})$ (mean value, standard deviation or probability of observation x_{ij}) are parameters that have to be determined either theoretically or empirically.

For each pattern, $e^{-\Sigma\omega_i}$ is proportional to the likelihood, so that the most likely pattern for each parallelogram can be chosen. Similarly, the most likely hypothesis for the roof of the missing building is selected by comparing the ω of the most likely pattern associated with each parallelogram.

An advantage of this selection method is that the system can --- because of the selected pattern --- give a prediction, with a certain probability, as to whether a corrective interaction is necessary, and where it has to be made. Also, note that the selection process described here is not suitable for the automatic selection step, because too many hypotheses would be accepted --- the automated system does not know for sure that there is a certain building at this location.

Example: dark buildings

Consider the problem class of "dark buildings." The boundary between the shadow and the roof is typically difficult to detect. The image edges of two sides of the roof are, at best, only partly visible. Three observations are sufficient to select the best hypothesis available after the perceptual organization step in the automatic system: evaluation of the parallelogram-corners, the grayvalue changes at the roof boundaries, and the overall average gray level. Two patterns are used, one where all sides are correct, and one where one or two sides nearby the shadow are incorrect.

It is possible to calculate the roof boundaries and corners that cast the shadow; the corner formed by these roof sides is likely to be very inaccurate, while the corner formed by the other two sides (non-shadow casting) is supposed to be rather precise (otherwise no hypothesis would have been established). The gray-level along the sides of the roof is supposed to change only on the non-shadow sides. The overall average gray-level should be low and the variance rather small.

This analysis leads to an easily derivable set of parameters which are used for the calculation of the most likely hypothesis.

Figure 5.4 shows an example of a missed dark building: (a) an image, (b) the line segments and junctions extracted, and (c) the roof hypotheses. After specifying the detection problem, the image-contrast is enhanced for display, (d). A roof hypothesis with error ellipses of corners and center of gravity is shown in (e). The 3-D building model found *just after* initial interaction is shown in (f).

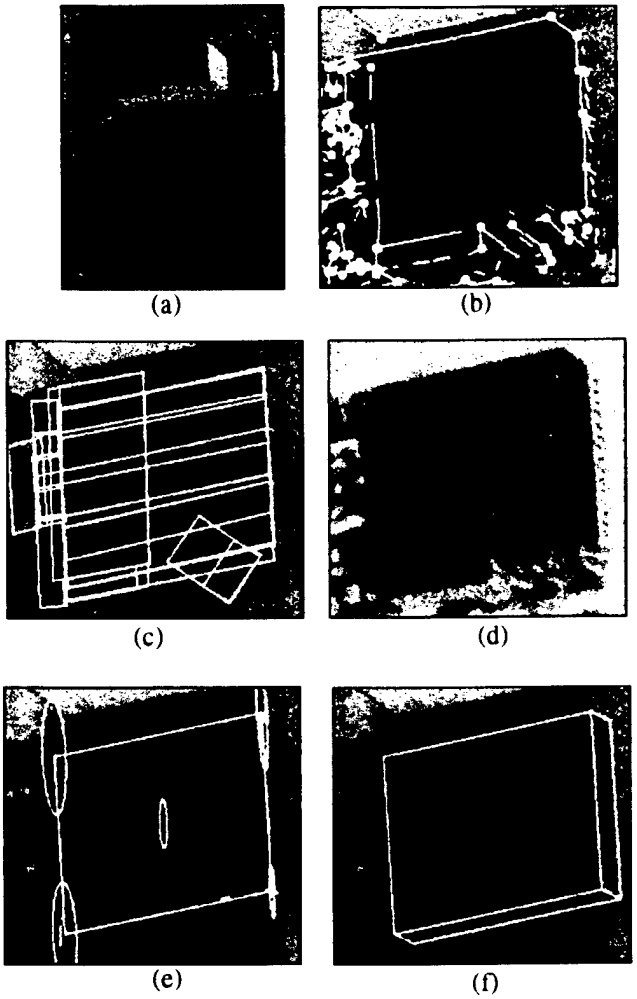


Figure 5.4 An example of a missed dark building.

5.3 Manual Feature Extraction

If the building is still not correctly detected, additional information is needed and one has to go one step backwards in the hierarchy of the automatic system to extract new features, such as edges or corners. Two ways of correcting the first hypothesis are offered: first the user can adjust the roof parallelogram by dragging sides with the mouse, and rotating or translating the whole model. Changes can only be made within the constraints of the building model, for example, opposite sides remain parallel (see Figure 5.5). The extraction of a ground corner or edge (shadow corner or edge) will determine the building height. These interactions are similar to those with an entirely manual system.

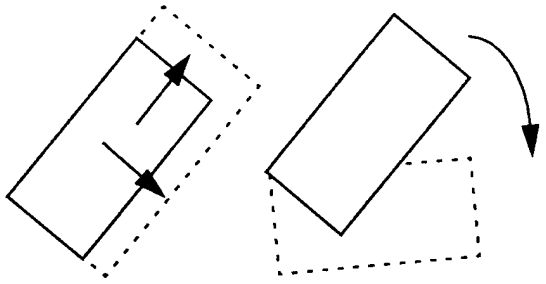


Figure 5.5 Manual adjustments - sides and rotation.

Second, one can choose to extract edges and corners and associate them to a part of the building model. For example, a roof-side of the building can be specified by an edge extracted in the image. Then this edge is added to the current hypothesis. Our systems are implemented to run under the RCDE [4]. This environment allows the use of mouse-sensitive features, thus facilitating user selection and manipulation of features.

After each corrective interaction the system forms a new parallelogram hypothesis. The system looks for new edges, shadow and wall evidence to support the new hypothesis, and finally, performs a fitting and verification step. These methods are the same as those in the automatic system. This important step of verifying the consistence to the constraints proposed in the automatic system can be compared to a fitting process in a computer assisted manual system, though in our system, a fitting is performed after each interaction. Therefore, it is possible that after a manual correction of a roof boundary, the wrong building height also corrected automatically.

Without the fitting step the system would perform like a manual system and at least three interaction steps (two corner adjustments and one correction of the building height) would be necessary for adjusting the shape of one building model. Rotation and translation as parameters of the position add another two steps.

Note, that the manual feature extraction and the following fitting and verification steps can be applied to buildings that are automatically detected, but partially wrong.

5.4 Results and Extensions

5.4.1 Examples

The system was tested on a number of examples provided by the RADIUS program (oblique and nadir views). In Figure 5.4, an example of using only initial interaction was shown. In Figure 5.6 the building (a) is not correctly detected because of missing edges, (b). There is no correct parallelogram formed and all roof hypotheses in (c) are rejected by the automatic system. After the initial interaction, a partly wrong roof hypothesis, (d), is found, where the shadow casting roof boundary is missed. The dotted line shows the estimated shadow boundary. The adjustment of one corner (e) leads to a new hypothesis (f). Note that after the correction of the cor-

ner, the system automatically found the associated shadow boundary (dotted line) and it corrected the building height.

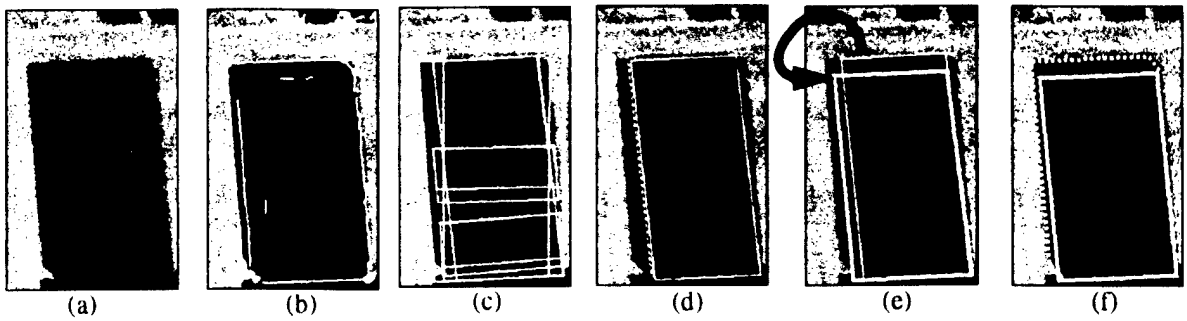


Figure 5.6 Undetected building extracted after one corner correction.

In Figure 5.7 an L-shaped building, (a), is only partly detected, (b). After specifying the problem and giving a rough location of the building, the missing part was found and fitted without any manual corrections, (c).

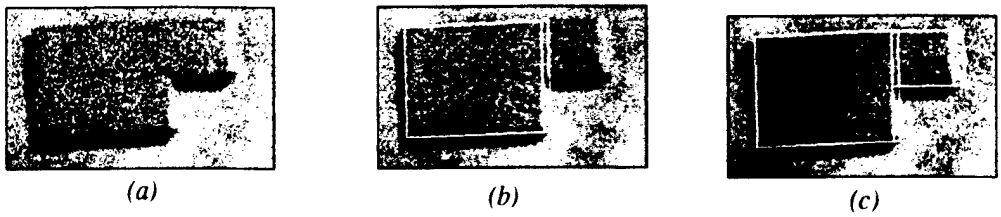


Figure 5.7 A partly detected L-shaped building easily detected.

5.4.2 Evaluation

This approach fulfills the requirements proposed earlier: by the initial step, translation and rotation is usually defined by two "qualitative" interactions. In manual or computer assisted manual systems, the position is given by more or less accurate measurements in the image. The initial step also gives a first guess of the shape of the building, which might already be the correct hypothesis. In our examples, a correct hypothesis was always found, when it was generated but rejected by the automatic system.

In nearly all cases, only corrections of the sides and height are necessary because rotation and position are already given by the initial step. Also, the number of correction steps in many cases was, at most two (see Table 3). A correction step of the height can be saved because of the fitting after each step.

Also the precision of the user's interaction is decreased. The corrective part uses a fitting process so that high precision is not needed. Furthermore, by adding already

extracted features to the model, like image edges, the quality of those features is undertaken and included in the hypothesis.

Table 3: Distribution of numbers of required interaction steps

initial interaction	1 corrective interaction step	2 corrective interaction steps	≥3 correct. interaction steps
4	9	4	0

5.4.3 Extensions

Currently, the interactive system has knowledge about a limited set of problems that the automatic system may encounter. Future extensions can extend this set.

6 Detecting Building Structures from Multiple Aerial Images

Section 3 described a system for building detection and description from a single image [7]. While this system shows good performance on many examples, it needs to rely strongly on presence of detectable shadows or vertical lines. The task can be made easier if multiple views are available as is likely to be the case for initial site model construction or detailed analysis of a change detected in monocular analysis. However, the multiple images are not necessarily taken at the same time: hence imaging conditions, including the sun position, the atmospheric conditions, and the environmental conditions, may be quite different.

Problems of segmentation and 3-D recovery are simplified by presence of multiple views, however do not disappear completely. A simplistic view of multiple view processing would be that we could first recover a dense 3-D map by matching across the different views and then segment the desired structures in 3-D. However, this is rarely possible in stereo processing and is particularly difficult for the problem being considered here. We cannot directly compute a dense 3-D map of the scene as there are large homogeneous areas whose interiors can not be matched directly, and we cannot match intensity values across images as they are not invariant with the changing viewing conditions. Instead, what we can attempt to do is match features, such as object boundaries, that are invariant across the images. However, the set of such features will likely be sparse and fragmented and we must group them to infer coherent objects.

To illustrate the nature of the problem, consider three images of a scene shown in Figure 6.1, with line segments overlaid. Note that the sides of the buildings that are visible are not the same in all views and that the shadows cast on the ground are quite different. The line segments were extracted from the images using an edge detector [6] and LINEAR line finder [5]. Note that not all of these boundaries have correspondences in more than one view. Also, it is unlikely that we can find unambiguous matches even for those lines that do correspond just by looking at the lines individually. Many parallel lines are likely to be present nearby in an urban scene, where buildings are often parallel to each other, as are ancillary structures, such as roads, sidewalks and landscaping.

For such a problem, we suggest that the problems of matching and grouping (i.e. 3-D recovery and object segmentation) not be separated but solved simultaneously.

The difficulty with matching lower level features is that it is difficult to disambiguate the matches correctly; the difficulty at the higher levels is that the correct groupings may not be formed in the first place. We propose a hierarchical grouping scheme where lower level features are grouped into successively higher level features. At each level, the grouped structures are matched across the different views and only the consistent ones are retained. We attempt to recover roof structures first, as they form the dominant regions of the buildings in the projected images. However, final selection of roof hypotheses needs to take advantage of the context provided by the visible walls (which may be different in different views), and by shadows cast by them.

To simplify our task, we restrict the domain of buildings that we work with to rectilinear structures (i.e. those consisting of rectangular components). Further, we assume that the roofs are planar and that the walls are vertical. This allows us to make some predictions about the expected properties of the projected boundaries in the image. Also we assume that the "camera models" are given, that is we can infer the epipolar geometry between the views and know the orientation with respect to a ground frame. Note that we do not require the different views be such that the epipolar lines are parallel, nor do we "rectify" the images to parallelize the epipolar lines.

There have been a few previous attempts to detect buildings from multiple views, though most assume stereo images taken at the same time ([10],[14] and [16]). It is common to match low-level features, such as lines and junctions, and to attempt to infer buildings from the matches by some kind of tracing or grouping method. The system described by Mohan and Nevatia matches higher level hypotheses (rectangles) however, does not use stereo information to form the hypotheses themselves. A recent system by Jaynes et. al. [17] does deal with the same kinds of imagery that we do (in fact, we use the same test data). However, the approach in this system is dif-

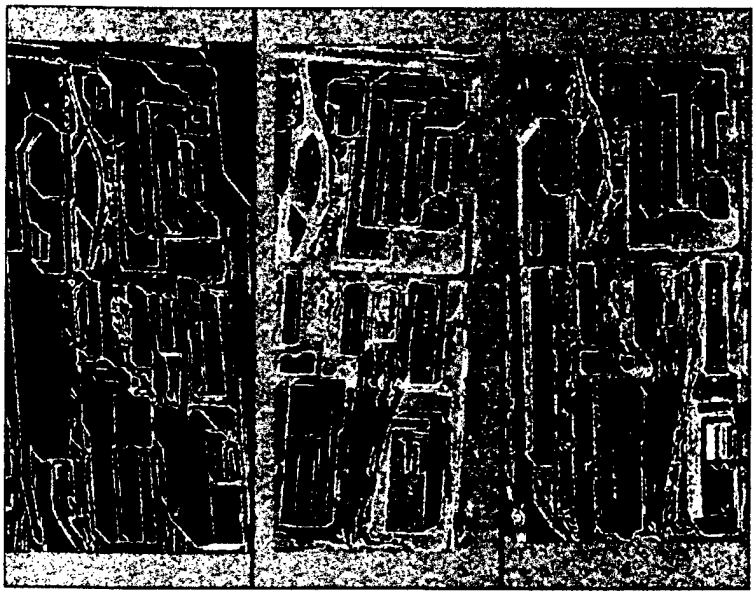


Figure 6.1 Views of modelboard scene

ferent in several ways. This system first uses a single view to determine roof outlines. Matches for these roof outlines are found in other views, and heights are determined by peaks in a histogram of heights from different pairs of views. This method has demonstrated very good results on one set of views. However, its performance may be critically dependent on the ability to generate good hypotheses from a single "seed" view (apparently a nadir view). This system assumes that the orientation of the sides of the roofs in the image is known in advance.

6.1 Overview of the System

This system uses a hypothesize-and-verify paradigm. Roof hypotheses are formed by a hierarchical grouping and matching scheme and verified by using wall and shadow evidence. A block diagram is given in Figure 6.2. With the restrictions of rectilinearity in the shapes of the buildings our system is designed for, the roofs can be expected to project into parallelograms or a combination of them (we assume that projection is either truly orthographic or is approximately orthographic over the extent of a building; this is generally true of aerial images taken from a height substantially larger than the heights of the buildings). We form hypotheses for parallelograms in a hierarchical way, by forming lines, junctions, parallels, "U"s, and finally, the parallelograms themselves. Evidence from all the views is used to generate the groupings and the process is not dependent on the order in which the views are examined. Matching takes place at various levels, and the results of matching at one stage are used for grouping at the higher levels. At each stage, some selections are made but the process is only intended to remove the hypotheses that become unviable with the increasing availability of context; at each stage, multiple hypotheses may remain even after selection.

Each hypotheses that is selected as being a candidate for being a roof, based on the evidence formed by features in the multiple views, is then "verified" by looking for supporting evidence from the walls and the shadows. Since we know the roof hypotheses in 3-D, we can predict the locations of the lines forming the wall boundaries as well as the shadows on ground (ground is assumed to be flat, though other kinds of known terrain could be included). Hypotheses with sufficient combined evidence form the output descriptions of our system. Our system does have the ability of providing confidence values for each object which may be useful for subsequent processes or humans that need to exploit the results. The confidence values are calculated based on the extent and accuracy of detected vertical walls and shadows cast by the roof, compared to their predicted locations.

6.2 Results and Future Work

Figure 6.3 shows the results obtained on the images shown in figure 6.1. Our system is able to correctly detect 13 of the 16 buildings in this scene. The missed buildings have dark roofs whose boundaries are not distinguishable from the shadows they cast. We are in the process of further testing, evaluation and enhancement of

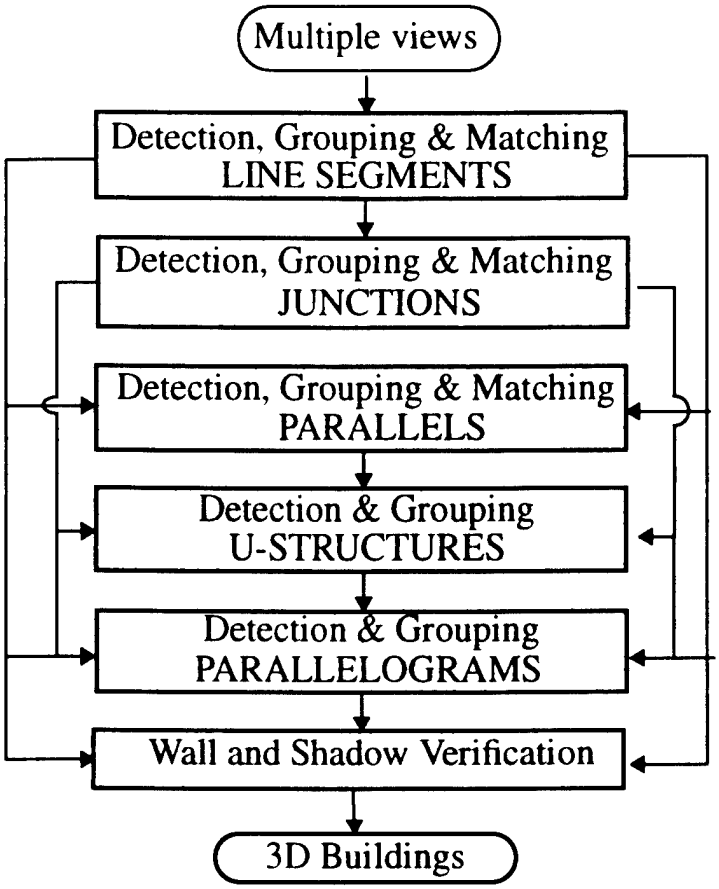


Figure 6.2 Flowchart of the system.

this system. We believe that our hierarchical approach has strong advantages and a potential for providing a highly robust and reliable system. We expect to show more extensive results in our next report.

Figure 6.3 Verified buildings.

7 References

[1] Médioni G., A. Huertas and M. Wilson (1991) *Automatic Registration of Color Separation Films*, Machine Vision and Applications, Springer-Verlag, New York, NY. Vol. 4, pp 33-51.

[2] Nevatia, R. and A. Huertas (1994) "Annual Technical Report," RADIUS Program. USC-IRIS Technical Report No. 340.

[3] Huertas, A., M. Bejanin and R. Nevatia (1995), *"Model Registration and Validation,"* In Automatic Extraction of Man-Made Objects from Aerial and Space Images, Gruen, A., Kuebler, O., Agouris, P. Editors. Birkhauser Verlag, Switzerland, pp 33-42.

[4] Strat , T.M. et al. (1992) *The RADIUS Common Development Environment*, Proceedings of the DARPA Image Understanding Workshop, San Diego, CA, Morgan Kaufman, Publisher, Jan., pp 215-226.

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

[6] Canny, J. (1986) A *Computational Approach to Edge Detection*. IEEE transactions on Pattern Analysis and Machine Intelligence 8(6), Nov., pp 679-698.

[7] Lin C., A. Huertas and R. Nevatia (1994) *Detection of Buildings using Perceptual Grouping and Shadows.* Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, Seattle, WA, Jun., pp 62-69.

[8] Huertas A. and R. Nevatia (1988) *Detecting Buildings in Aerial Images*, Computer Vision, Graphics and Image Processing Vol. 41, pp 131-152.

[9] Subhodev, D., B. Bhanu, X. Wu and R. Braithwaite (1994) *A System for Aircraft Recognition in Perspective Aerial Images*, Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, pp 169-175.

[10] Mohan, R. and R. Nevatia (1989) *Using Perceptual Organization to Extract 3-D Structures,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(11), Nov., pp 1121-1139.

[11] Shufelt, J. and D. McKeown (1993) *Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery,* Computer Vision, Graphics and Image Processing, 57(3), pp 307-330.

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

[13] Neuenschwander, W., P. Fua, G. Szekely and O. Kubler (1994), *Making Snakes Converge from Minimal Initialization*, Proceedings of the International Conference on Pattern Recognition, Jerusalem, Israel, Oct., pp 613-615.

[14] Chung, C. and R. Nevatia (1992), *Recovering Building Structures from Stereo,* IEEE Proceedings of Workshop on Applications of Computer Vision, Palm Springs, CA, Dec., pp 64-73.

[15] Liow, Y. and T. Pavlidis (1990), *Use of Shadows for Extracting Buildings in Aerial Images,* Computer Vision, Graphics and Image Processing, Vol. 49, No. 2, Feb., pp 242-277.

[16] Roux, M. and D. McKeown (1994), *Feature Matching for Building Extraction from Multiple Views*, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, Jun., pp 46-53.

[17] Jaynes, C., F. Stolle and R. Collins (1994), *Task Driven Perceptual Organization for Extraction of Rooftop Polygons,* ARPA Image Uunderstanding Workshop, Monterey, CA, Nov., pp 359-365.

[18] McGlone, J. and J. Shufelt (1994) *Projective and Object Space Geometry for Monocular Building Extraction,* IEEE Proceedings of the Computer Vision and Pattern Recognition Conference, Seattle, WA, Jun., pp 54-61.