TEC-0121

The UMass RADIUS Project: A System for Automated Site Model Acquisition and Extension

Robert T. Collins
Allen R. Hanson
Edward M. Riseman

University of Massachusetts
Computer Science Department
Amherst, MA 01003-4610

May 1999

Approved for public release; distribution is unlimited.

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, Virginia 22315-3864

US Army Corps
of Engineers
Topographic
Engineering Center

T

E

C

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE May 1999	3. REPORT TYPE AND DATES COVERED Final Technical October 1992 - September 1996

4. TITLE AND SUBTITLE

The UMass RADIUS Project: A System for Automated Site Model Acquisition and Extension

5. FUNDING NUMBERS

DACA76-92-C-0041

6. AUTHOR(S)

Robert T. Collins
Allen R. Hanson
Edward M. Riseman

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

University of Massachusetts
Computer Science Department
Amherst, MA 01003-4610

8. PERFORMING ORGANIZATION REPORT NUMBER

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

Defense Advanced Research Projects Agency
3701 N. Fairfax Drive, Arlington, VA 22203

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

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

TEC-0121

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Under the DARPA RADIUS program, the University of Massachusetts (UMass) developed techniques to automatically populate a site model with 3-D building models extracted from multiple, overlapping images. The Automated Site Construction, Extension, Detection and Refinement (ASCENDER) system incorporates several key ideas. First, 3-D reconstruction is based on geometric features that remain stable under a wide range of viewing and lighting conditions. Second, rigorous photogrammetric camera models are used to describe the relationship between pixels in an image and 3-D locations in the scene, so that diverse sensor characteristics and viewpoints can be effectively exploited. Third, information is fused across multiple images for increased accuracy and reliability. Finally, known geometric constraints are applied whenever possible to increase the efficiency and reliability of the reconstruction process. Ascender was the primary deliverable of the UMass RADIUS effort and was delivered in mid-April 1995 to Lockheed-Martin for integration into the RADIUS Testbed System. At the same time, an informal transfer was made to the National Exploitation Laboratory (NEL) for familiarization and additional testing. This report presents an overview of the system, results of an evaluation on an unclassified data set of Fort Hood, TX, and suggestions for future extensions to the system.

14. SUBJECT TERMS

Building reconstruction, image understanding, spatial modeling, site model construction, aerial image analysis, model-based reasoning

15. NUMBER OF PAGES
31

16. PRICE CODE

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

Table of Contents

Table of Contents iii
List of Figures iv
List of Tables v
Preface vii
Acknowledgments vii
1. Introduction 1
2. The Ascender System 1
    2.1. System Overview 2
    2.2. The Building Extraction Process 4
3. Evaluation on Fort Hood Imagery 4
    3.1. Methodology 5
    3.2. Evaluation of 2-D Detection 6
        3.2.1. Detection Rates 7
        3.2.2. Quantitative Accuracy 9
    3.3. Evaluation of 3-D Reconstruction 10
        3.3.1. Baseline Reconstruction Accuracy 10
        3.3.2. Actual Reconstruction Accuracy 12
    3.4. Summary 13
4. Ascender Delivery 14
5. Grouping and Data Fusion 15
    5.1. Extracting/Grouping 2.5-D Lines 15
    5.2. Surface Fitting to DEM Data 17
    5.3. Extracting Surface Structures for Visualization 19
6. Summary and The Future 23
8. Bibliography 24

List of Figures

1. Sample building model automatically generated by the Ascender system 2

2. Some additional samples of building models generated by Ascender................ 3

3. Fort Hood evaluation area with 30 ground truth building models. 6

4. Building detector sensitivity vs. total number of roof hypotheses. 8

5. Building detector sensitivity vs. 2-D polygon accuracy in pixels (see text)......... 9

6. Number of views used vs. 3-D reconstruction accuracy in meters (see text)......... 11

7. Building detector sensitivity vs. 3-D polygon accuracy.................................. 12

8. Using 2.5-D lines in the grouping process helps disambiguate multi-level building roofs. 16

9. Three parametric peaked-roof surfaces that have been fit to DEM.. 18

10. Subimage of the Fort Hood dataset with roof polygons detected..............................18

11. Six reconstructed buildings from the Fort Hood scene..................................20

12. Three dimensional view of the Martin-Marietta (Denver) building constructed using the Terrest terrain reconstruction system..........................20

13. The same scene as shown in Figure 12 after replacing the building with the reconstructed geometric model...21

List of Tables

1. Ground sample distances (GSD) in meters for seven evaluation images. 5

2. Median inter-vertex distances (in pixels) between detected polygon vertices and projected ground truth roof vertices... 10

3. Baseline accuracy of the 3-D reconstruction process. 11

4. Median planimetric and altimetric errors between reconstructed 3-D polygon vertices and ground truth roof vertices. 13

5. Major changes to Ascender system after initial delivery to prime contractor.......... ..14

Preface

This research was sponsored by the Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense and was monitored by the U.S. Army Topographic Engineering Center (TEC) under contract No. DACA76-92-C-0041. The DARPA Program Manager is Dr. Oscar Firschein and the TEC Contracting Officer's Representative is Ms. Lauretta Williams.

Acknowledgments

This report owes much to the dedication and hard work of the UMass RADIUS team members: Prof. Howard Schultz, Chris Jaynes, Frank Stolle, Xiaoguang Wang, and Yong-Qing Cheng, to the software support of our system people: Robert Heller, Jonathan Lim, and Kris Denio, and to the administrative support of Janet Turnbull and Laurie Downey.

The UMass RADIUS Project

1. Introduction

The Research and Development for Image Understanding Systems (RADIUS) project is a national effort to apply image understanding (IU) technology to support model-based aerial image analysis [Gerson and Wood 1994]. Automated construction and management of 3-D geometric site models enables efficient exploitation of the tremendous volume of information collected daily by national sensors. The expected benefits are decreased work-load on human analysts, together with an increase in measurement accuracy because of the introduction of digital IU and photogrammetric techniques. When properly annotated, automatically generated site models can provide the spatial context for specialized IU analysis tasks such as vehicle counting, change detection, and damage assessment, while graphical visualization techniques using 3-D site models are valuable for training and mission planning. Civilian benefits of this technology also are numerous, including automated cartography, land-use surveying, and urban planning.

Over the past three years, the University of Massachusetts (UMass) has developed techniques to automatically populate a site model with 3-D building models extracted from multiple, overlapping images. There are many technical challenges involved in developing a building extraction system that works reliably on the type of images being considered under RADIUS. Multiple images of the scene may be captured by different cameras from arbitrary viewing positions, and images may be collected months or even years apart, under vastly different weather and lighting conditions. There is typically a lot of clutter surrounding buildings (vehicles, pipes, oil drums, shrubbery) and on top of them (roof vents, air conditioner units, ductwork). Buildings often occlude each other in oblique views, and shadows fall across building faces breaking up low-level extracted features such as line segments and regions. To overcome these difficulties, the UMass design philosophy incorporates several key ideas. First, 3-D reconstruction is based on geometric features that remain stable under a wide range of viewing and lighting conditions. Second, rigorous photogrammetric camera models are used to describe the relationship between pixels in an image and 3-D locations in the scene, so that diverse sensor characteristics and viewpoints can be effectively exploited. Third, information is fused across multiple images for increased accuracy and reliability. Finally, known geometric constraints are applied whenever possible to increase the efficiency and reliability of the reconstruction process.

Section 2 presents an overview of the Automated Site Construction, Extension, Detection and Refinement (ASCENDER) system, designed to automatically acquire models of buildings with flat, rectilinear rooftops. Ascender is the primary deliverable of the 3-year UMass RADIUS effort. Section 3 presents results of an evaluation conducted at UMass on an unclassified data set of Fort Hood, Texas. The system is being extended via new strategies for acquiring models of other common building classes, such as peaked and multi-level roof structures, which are described in Section 4. Section 5 outlines recent advances in the symbolic extraction of surface details, such as windows and doors, and their applications to graphical rendering for scene visualization.

2. The Ascender System

The Ascender system has been designed to automatically populate a site model with buildings extracted from multiple, overlapping images exhibiting a variety of viewpoints and sun angles. In mid-April 1995, Version 1.0 of the Ascender system was delivered to Lockheed-Martin for testing on classified imagery and for integration into the RADIUS Testbed System [Gerson and Wood 1994]. At the same time, an informal transfer was made to the National Exploitation Laboratory (NEL) for familiarization and additional testing. This section presents a brief overview of the Ascender system and its approach to extracting building models. More detailed descriptions can be found in [Collins et al., 1995a, b; Collins et al. 94]. Some sample building models automatically generated by the Ascender system are shown in Figures 1 and 2.

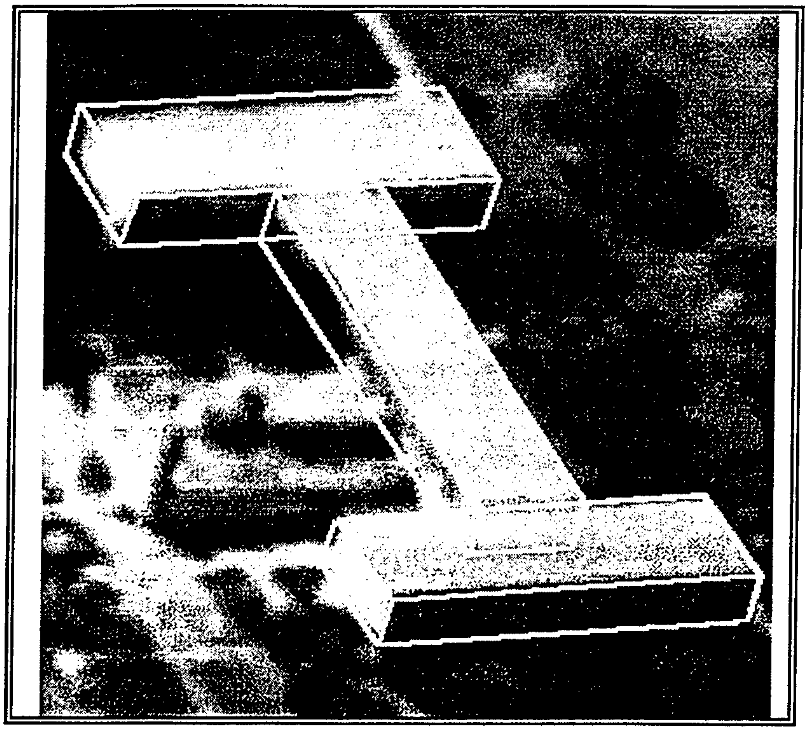

Figure 1. Sample building model automatically generated by the Ascender system.

2.1. System Overview

Ascender was developed on a Sun Sparc 20, using the Radius Common Development Environment (RCDE) [Mundy et al., 1992]. The RCDE is a combined Lisp/C++ system that supports the development of image understanding algorithms for constructing and using site models. The RCDE provides a convenient framework for representing and manipulating images, camera models, object models and terrain models, and for keeping track of their various coordinate systems, inter-object relationships, and transformation/projection equations. To be more specific, the following items needed by Ascender are managed by the RCDE and assumed to be present before the building extraction process begins:

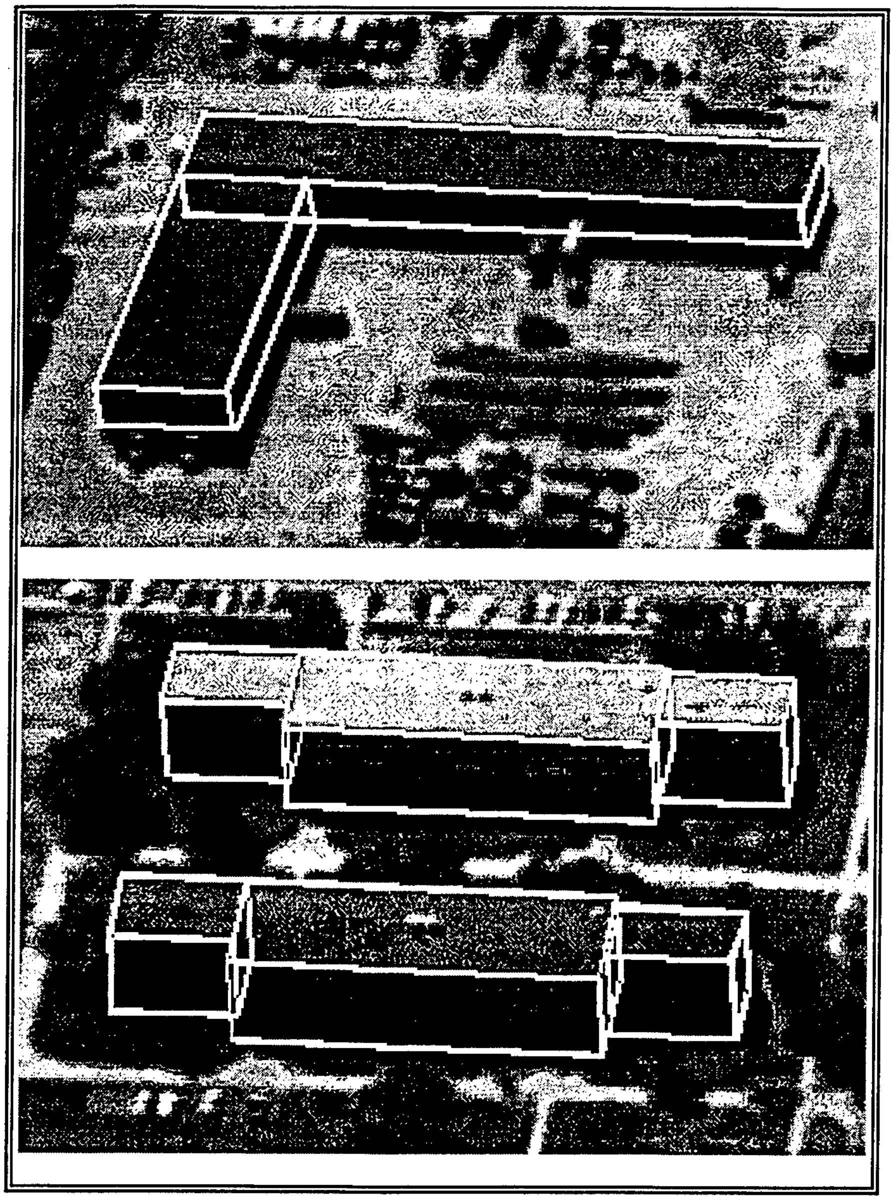

Figure 2. Some additional samples of building models generated by Ascender.

- **Images.** A set of images, both nadir and oblique, that view the same area of the site. Best results are obtained with images exhibiting a variety of viewing and sun angles.

- **Site Coordinate System.** A Euclidean, local-vertical coordinate system (Z-axis points up) for representing building models.

- **Camera Models.** A specification of how 3-D locations in the site coordinate system are related to 2-D image pixels in each image. One common camera representation is a 3 by 4 projective transformation matrix encoding the lens and pose parameters of each perspective camera. Ascender also can handle the fast block interpolation projection (FBIP) camera model used in the RCDE to represent the geometry of non-perspective cameras.

- **Digital Terrain Map.** A specification of the terrain underlying the site. This could be as simple as a plane equation, or could be a full array of elevation values computed via correlation-based stereo.

2.2. The Building Extraction Process

The Ascender system uses a straightforward control strategy to extract building models. The process is described briefly here, with particular attention given to the algorithmic parameters that can be set by the user to vary the number and quality of the resulting building hypotheses.

Building detection begins by extracting straight line segments using the Boldt algorithm [Boldt et al., 1989]. Intensity edges are grouped recursively into longer straight lines with subpixel accuracy via a set of Gestalt perceptual organization criteria. Two user thresholds, minimum line length and minimum contrast (gray-level difference across the line), are available to control the set of lines returned.

Two-dimensional building roof boundaries are hypothesized from extracted image line segments via a graph-based perceptual grouping algorithm [Jaynes et al., 1994]. Line segments are grouped into corners, chains, and eventually into complete closed polygons. A single variable sensitivity parameter ranging from 0.0 (very low sensitivity) to 1.0 (very high) controls the settings of several less-intuitive internal parameters that govern the polygon grouping process.

The recovery of 3-D building information begins by estimating a height for each hypothesized 2-D roof polygon via multiimage epipolar matching. This estimate is chosen as the peak in a height histogram formed by matching the polygon's edges to line segments in multiple images and allowing each potential match to vote for a height range. The size of the epipolar search region in each image is governed by two parameters: the minimum and maximum Z-values that building rooftops could be found at (the minimum value could potentially be determined from an accurate terrain map). A third parameter that governs the search for correspondences is the expected residual error (in pixels) between true and observed 2-D feature locations, roughly summarizing the level of error in image features caused by inaccuracies in the camera resection and feature extraction routines.

After a set of matching line segments for the building roof is found, a rigorous photogrammetric triangulation procedure is performed to determine the precise 3-D size, shape and position of the building rooftop. The optimization criterion simultaneously minimizes the sum-of-squared residual errors between projected 3-D roof polygon edges and corresponding line segment features in all the images. There are no user parameters. The resulting 3-D polygon is then extruded down to the provided terrain to form a complete building wireframe.

3. Evaluation on Fort Hood Imagery

The success of the Ascender system will ultimately be judged by its performance on classified imagery. Such tests have been performed at Lockheed-Martin. In parallel with that effort, UMass is performing an in-depth system evaluation using unclassified data. The set of experiments are designed to address questions such as:

- How is the rooftop detection rate related to system sensitivity settings?

- Is the detection rate affected by viewpoint (nadir vs. oblique)?
- Does 2-D detected polygon accuracy vary by viewpoint?
- Is 2-D accuracy related to sensitivity settings?
- How does 3-D accuracy vary with the number of images used?
- How does 3-D accuracy vary according to 2-D accuracy of the hypothesized polygons?

This section presents evaluation results on a large data set from Fort Hood, Texas. The imagery was collected by Photo Science, Inc., (PSI) in October 1993 and scanned at the Digital Mapping Laboratory at CMU in Jan-Feb. 1995. Camera resections were performed by PSI for the nadir views, and by CMU for the obliques.

3.1. Methodology

An evaluation data set was cropped from the Fort Hood imagery, yielding seven subimages from the views labeled 711, 713, 525, 927, 1025, 1125 and 1325 (images 711 and 713 are nadir views, the rest are obliques). Table 1 summarizes the ground sample distance (GSD) for each image. The region of overlap covers an evaluation area of roughly 760 by 740 meters, containing a good blend of both simple and complex roof structures. Thirty ground truth building models were created by hand using interactive modelling tools provided by the RCDE. Each building is composed of RCDE 'cube', 'house' and/or 'extrusion' objects that were shaped and positioned to project as well as possible (as determined by eye) simultaneously into the set of seven images. The ground truth data set is shown in Figure 3.

Table 1. Ground sample distances (GSD) in meters for the seven evaluation images. A GSD of 0.3 means that a length of 1 pixel in the image roughly corresponds to a distance of 0.3 meters as measured on the ground.

711	713	525	927	1025	1125	1325
0.31	0.31	0.61	0.52	1.10	1.01	1.01

Since the Ascender system explicitly recovers only rooftop polygons (the rest of the building wireframe is formed by vertical extrusion), the evaluation is based on comparing detected 2-D and triangulated 3-D roof polygons vs. their ground truth counterparts. There are 73 ground truth rooftop polygons among the set of 30 buildings. Ground truth 2-D polygons for each image are determined by projecting the ground truth 3-D polygons into that image using the known camera projection equations.

The *Center-Line Distance* measures how well two arbitrary polygons match in terms of size, shape, and location[1]. The procedure is to oversample the boundary of one polygon into a set of equally spaced points (several thousand of them). For each point, measure the minimum distance from that point to the other polygon boundary. Repeat the procedure by oversampling the other polygon and measuring the distance of each point to the first polygon boundary. The center-line distance is taken as the average of all these values. This metric provides a measure of the average distance between the two polygon boundaries, reported in pixels for 2-D polygons, and in meters for 3-D polygons. We prefer the center-line distance to other comparison measures, such as the one used in [Roux et al., 1995], since it is very easy to compute and can be applied to two polygons that do not have the same number of vertices.

---

[1] Robert Haralick, private communication, 1996.

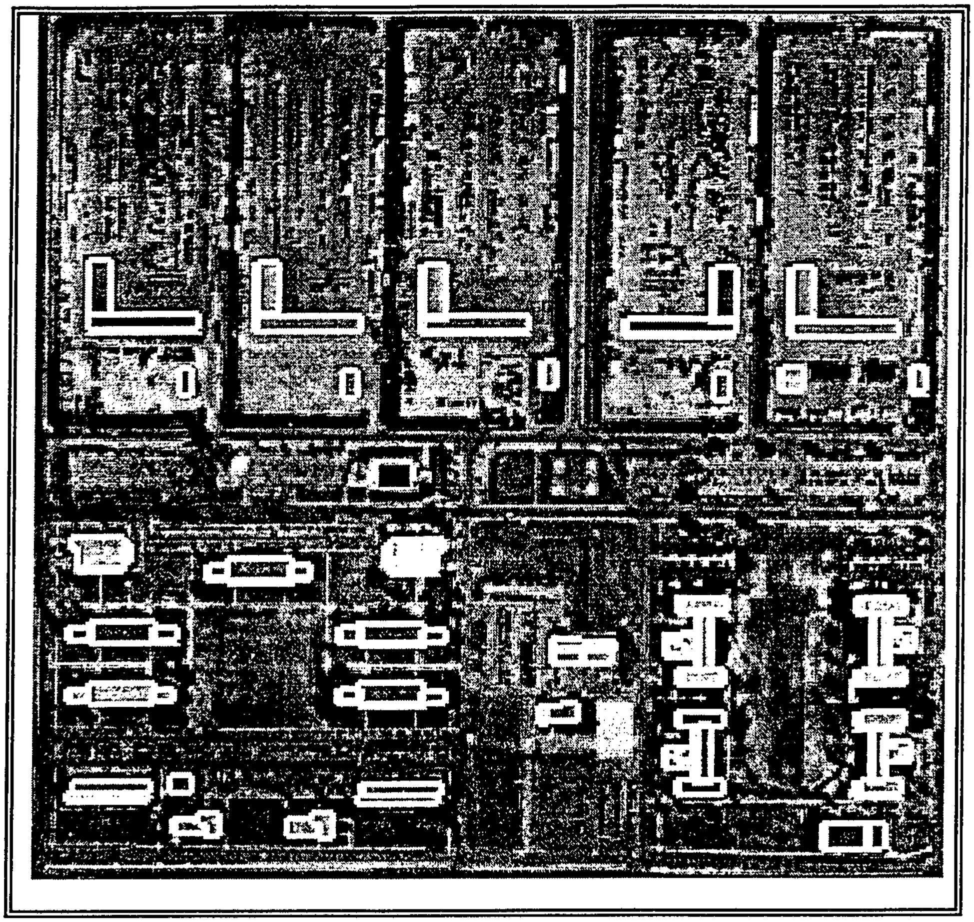

Figure 3. Fort Hood evaluation area with 30 ground truth building models composed of single- and multilevel flat roofs, and two peaked roofs. There are 73 roof facets in all. The size of the image area shown is 2375 by 1805 pixels.

For polygons that have the same number of vertices, and are fairly close to each other in terms of center-line distance, an additional distance measure is computed between corresponding pairs of vertices between the two polygons. That is, for each polygon vertex, the distance to the closest vertex on the other polygon is measured. For 2-D polygons these *Inter-Vertex Distances* are reported in pixels, for 3-D polygons the units are meters, and the distances are broken into their planimetric (distance parallel to the X-Y plane) vs. altimetric (distance in Z) components.

3.2. Evaluation of 2-D Detection

One important module of the Ascender system is the 2-D polygonal rooftop detector. The detector was tested on images 711, 713, 525, and 927 to see how well it performed at different grouping sensitivity settings, and with different length and contrast settings of the Boldt line extraction algorithm. The detector was tested by projecting each ground truth roof polygon into an image, growing its 2-D bounding box out by 20 pixels on each side, then invoking the building detector in that region to hypothesize 2-D rooftop polygons. The evaluation goals were to determine both true and false positive detection rates *when the building detector was invoked on an area containing a building*, and to measure the 2-D accuracy of the true positives.

3.2.1. Detection Rates

The polygon detector typically produces several roof hypotheses within a given image area, particularly when run at the higher sensitivity settings. Thus, determining true and false positive detection rates involves determining whether or not each hypothesized image polygon is a good match with some ground truth projected roof polygon. To automate the process of counting true positives, each hypothesized polygon was ranked by its center-line distance from the known ground truth 2-D polygon that was supposed to be detected. Of all hypotheses with distances less than a threshold (i.e.polygons that were reasonably good matches to the ground truth), the one with the smallest distance was counted as a true positive; all other hypotheses were considered to be false positives. The threshold value used was 0.2 times the square root of the area of the ground truth polygon, that is:

$$\mathrm{Dist}(\mathrm{hyp}, \mathrm{gt}) \leq .2\sqrt{\mathrm{Area}(\mathrm{gt})}$$

where 'hyp' and 'gt' are hypothesized and ground truth polygons, respectively. This empirical threshold allows 2 pixels total error for a square with sides 10 pixels long, and varies linearly with the scale of the polygon.

The total numbers of roof hypotheses generated for images 711, 713, 525, and 927 are shown at the top of Figure 4 for nine different sensitivity settings of the building detector ranging from 0.1 to 0.9 (very low to very high). The line segments used for each image were computed by the Boldt algorithm using length and contrast thresholds of 10. The second graph in Figure 4 plots the number of true positive hypotheses. For the highest sensitivity setting, the percentage of rooftops detected in 711, 713, 525, and 927 were 51 percent, 59 percent, 45 percent and 47 percent, respectively. The graph also shows the number of true positives achieved by combining the hypotheses from all four images, either by pooling hypotheses computed separately for each image, or by recursively masking out previously detected buildings and focusing on the unmodeled areas in each new image [Collins et al., 1995a]. For the highest sensitivity setting, this strategy detects 81 percent (59 out of 73) of the rooftops in the scene.

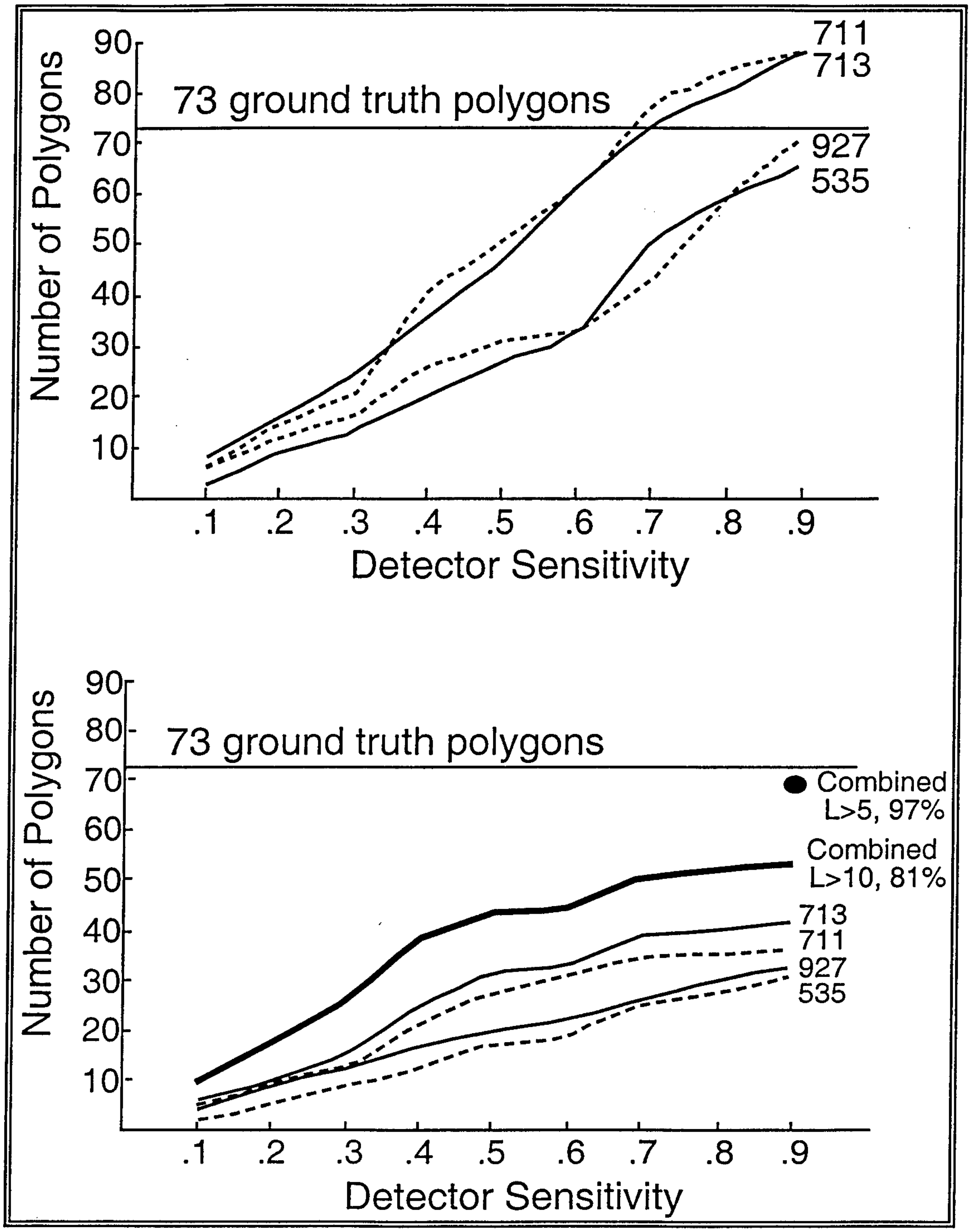

Figure 4. Top: Building detector sensitivity vs. total number of roof hypotheses. Bottom: Sensitivity vs. number of true positives. Horizontal lines show the actual number of ground truth polygons. Combining results from all four views yields a 'best' detection rate of 81 percent with lines of length > 10, and 97 percent with lines of length > 5.

The detection rates seem to be sensitive to viewpoint. More total hypotheses and more true positives were detected in the nadir views than in the obliques. This may represent a property of the building detector, but it also is likely that most of the discrepancy is caused by the difference in GSD of the images for this area (see Table 1). Each building roof occupies a larger set of pixels in the nadir views than in the obliques, for this data set.

To measure the best possible performance of the rooftop detector on this data, it was run on all four images at sensitivity level 0.9, using Boldt line data computed with length and contrast thresholds of 5. These were judged to be the highest sensitivity levels for both line extractor and building detector that were feasible, and the results represent the best job that the building detector can possibly do with each image. The percentages of rooftops detected in each of the four images under these conditions were 86 percent, 84 percent, 74 percent, and 67 percent, with a combined image detection rate of 97 percent (71 out of 73).

3.2.2. Quantitative Accuracy

To assess the quantitative accuracy of the true positive 2-D roof polygons, each was compared with its corresponding 2-D projected ground truth polygon in terms of center-line distance. Figure 5 plots the median of the center-line polygon distances between detected and ground truth 2-D polygons, for different sensitivity settings. Polygons detected at low sensitivity levels seem to be slightly more accurate than those detected at the high sensitivity settings. This is so because the detector only finds clearly delineated rooftop boundaries at the lower settings, and is more forgiving in its grouping criteria at the higher settings.

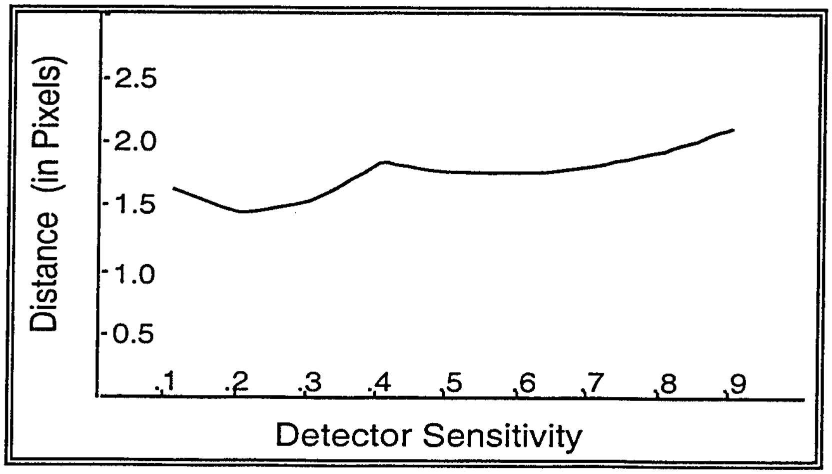

Figure 5. Building detector sensitivity vs. 2-D polygon accuracy in pixels (see text).

For pairs of detected and ground truth polygons having the same number of vertices, their set of inter-vertex distances also were computed, and the medians of those measurements are broken down by image in Table 2. The average distance is approximately 2.7 pixels. Polygons detected in image 927 appear to be slightly more accurate. This difference may or may not be significant; however, image 927 was taken in the afternoon, and all the other images were taken in the morning, therefore the difference in sun angle may be the cause.

3.3. Evaluation of 3-D Reconstruction

The second major subsystem in Ascender takes 2-D roof hypotheses detected in one image and reconstructs 3-D rooftop polygons via multiimage line segment matching and triangulation. Two different quantitative evaluations were performed on this subsystem. The 3-D reconstruction process was first tested in isolation from the 2-D detection process by using 2-D projected ground truth polygons as input. This initial evaluation was done to establish a baseline measure of reconstruction accuracy, that is, to see how accurate the final 3-D building models would be given perfect 2-D rooftop extraction. A second evaluation tested end-to-end system performance by performing 3-D reconstruction using the set of automatically detected 2-D image polygons from the previous section.

3.3.1. Baseline Reconstruction Accuracy

The baseline measure of reconstruction accuracy was performed using 2-D projected ground truth roof polygons. For each of the seven images in the evaluation test set, all the ground truth 2-D polygons from that image were matched and triangulated using the other six images as corroborating views. The accuracy of each reconstructed roof polygon was then determined by comparing it with its 3-D ground truth counterpart in terms of center-line distance and inter-vertex distances. Table 3 reports, for each image, the median of the center-line polygon distances between reconstructed and ground truth polygons for that image. Also reported are the medians of the planimetric (horizontal) and altimetric (vertical) components of the inter-vertex distances between reconstructed and ground truth polygon vertices. Horizontal placement accuracy was about 0.3 meters, which is in accordance with the resolution of the images.

Another suite of tests was performed to determine how the number of views affects the accuracy of the resulting 3-D polygons. These tests were performed using image 711 as the primary image, and all 63 non-empty subsets of the other six views as additional views. For each subset of additional views, all 2-D projected ground truth polygons in image 711 were matched and triangulated, and the median center-line and inter-vertex distances between reconstructed and ground truth 3-D polygons were recorded. Figure 6 graphs the results, organized by number of images used (including 711), ranging from only two views up to six views.

Table 2. Median inter-vertex distances (in pixels) between detected polygon vertices and projected ground truth roof vertices, for four images.

	711	**713**	**525**	**927**
IV Distance	2.75	2.82	2.71	2.22

Table 3. Baseline accuracy of the 3-D reconstruction process. Median center-line distances as well as inter-vertex planimetric and altimetric errors are shown (in meters) for four images. See text.

	711	713	525	927
CL Distance	0.57	0.46	0.45	0.53
IV Planimetric	0.29	0.25	0.33	0.35
IV Altimetric	0.49	0.42	0.37	0.43

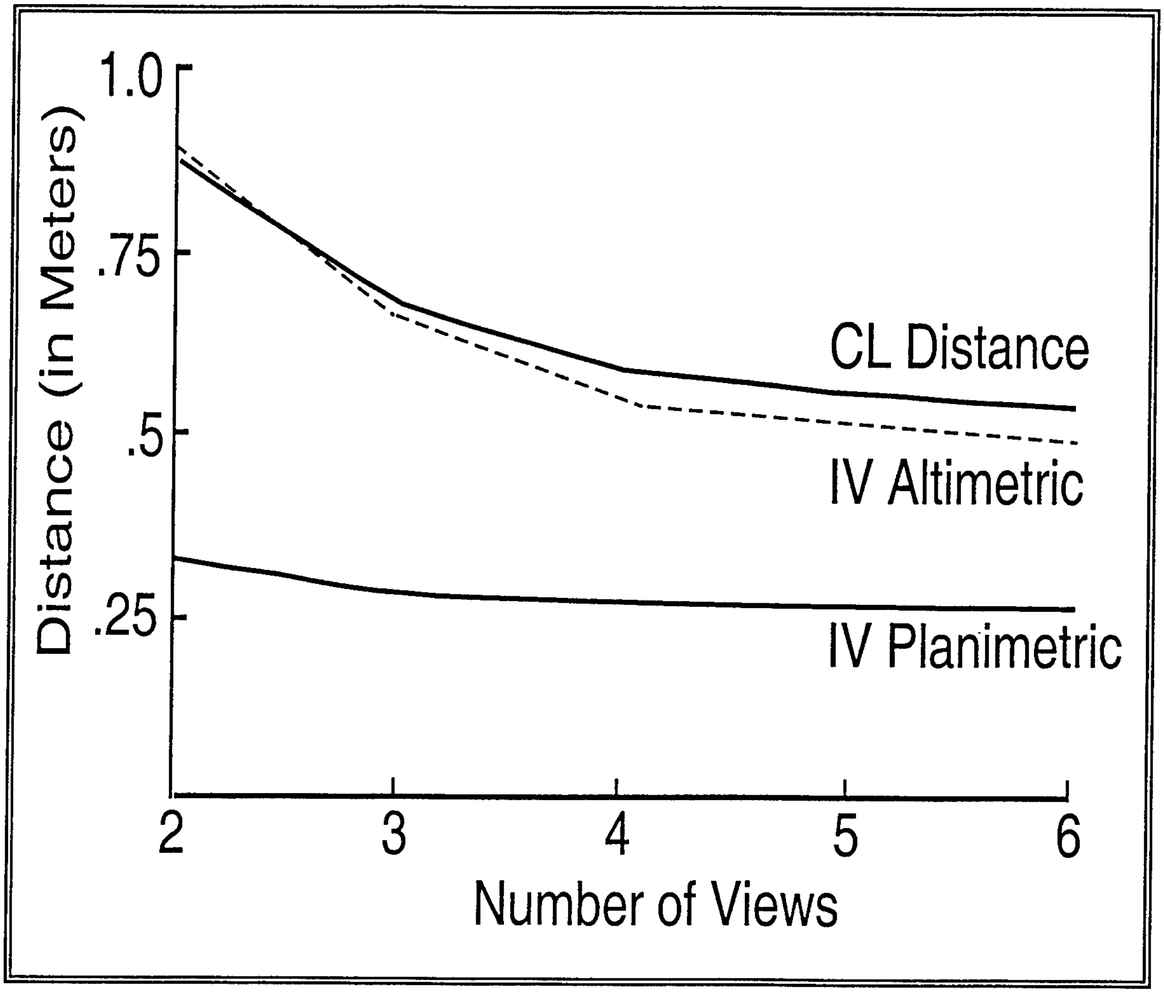

Figure 6. Number of views used vs. 3-D reconstruction accuracy in meters (see text).

The distances reported under label '2' are averaged over the 6 possible image sets containing 711 and one other image, distances reported under '3' are averaged over all 15 possible image sets containing 711 and two other images, and so on. There is a noticeable improvement in accuracy when using three views instead of two, but the curves flatten out after that, and there is little improvement in accuracy gained by taking image sets larger than four.

3.3.2. Actual Reconstruction Accuracy

In actual practice, Ascender reconstruction techniques are applied to the 2-D image polygons hypothesized by its automated building detector. Thus, the final reconstruction accuracy depends not only on the number and geometry of the additional views used, but also on the 2-D image accuracy of the hypothesized roof polygons. The typical end-to-end performance of the system was evaluated by taking the 2-D polygons detected in Section 2 and performing matching and triangulation using the other six views. The median center-line distances between reconstructed and ground truth 3-D polygons are plotted in Figure 7 for different sensitivity settings of the polygon detector. The accuracy is slightly better when using polygons detected at the lower sensitivity settings, mirroring the better accuracy of the 2-D polygons at those levels (compare with Figure 5).

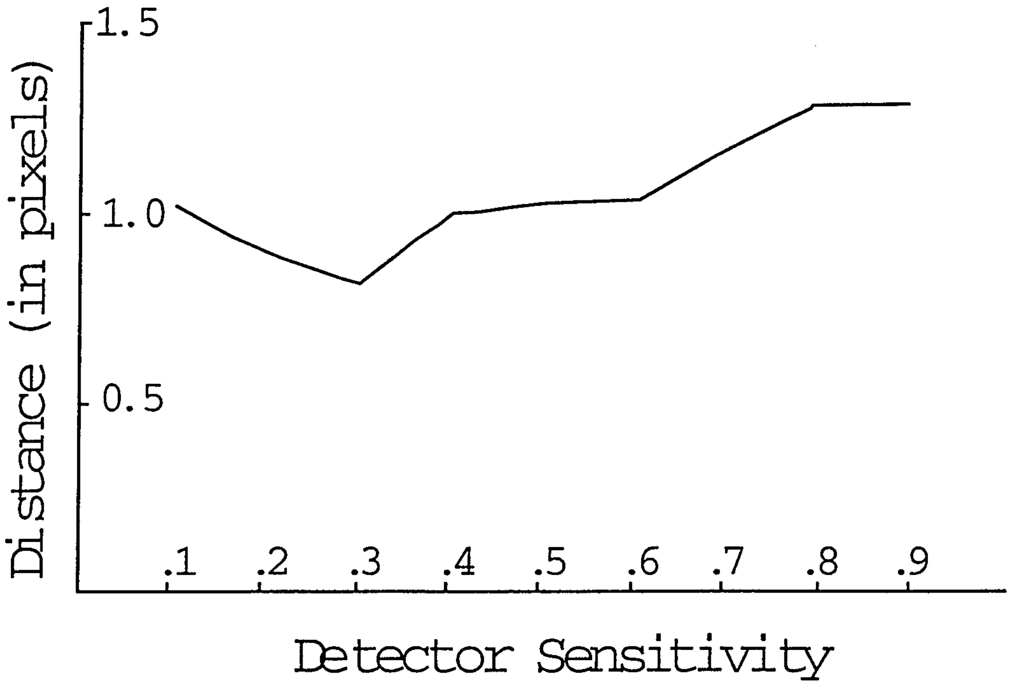

Figure 7. Building detector sensitivity vs. 3-D polygon accuracy, computed as the median of center-line distances between reconstructed 3-D polygons and ground truth roof polygons.

For pairs of detected and ground truth polygons having the same number of vertices, the set of inter-vertex planimetric and altimetric errors were computed, and the medians of those measurements are shown in Table 4, broken down by the image in which the 2-D polygons feeding the reconstruction process were hypothesized. Unlike the baseline error data from Table 3, where the horizontal accuracy of reconstructed polygon vertices was better than their vertical accuracy, here the situation is reversed, strongly suggesting that the planimetric component of reconstructed vertices is more sensitive to inaccuracies in the 2-D polygon detection process than the altimetric component. This result is consistent with previous observations that the corners of Ascender's reconstructed building models are more accurate in height than in horizontal position [Collins et al., 1994].

Table 4. Median planimetric and altimetric errors (in meters) between reconstructed 3-D polygon vertices and ground truth roof vertices.

	711	713	525	927
IV Planimetric	0.68	0.73	1.09	0.89
IV Altimetric	0.51	0.55	0.90	0.61

3.4. Summary

This section has presented preliminary results of an on-going evaluation of the Ascender system using an unclassified Fort Hood data set. While the results of the analysis are inevitably tied to this specific data set, they give us some indication of how the system should be expected to perform under different scenarios.

Single-Image Performance: The building detection rate varies roughly linearly with the sensitivity setting of the polygon detector. At the high sensitivity level, roughly 50 percent of the buildings are detected in each image using Boldt lines extracted at a medium level of sensitivity (length and contrast > 10), and about 75-80 percent when using Boldt lines extracted at a high level of sensitivity (length and contrast > 5). Although line segments and corner hypotheses are localized to subpixel accuracy, the median localization error of 2-D rooftop polygon vertices is around 2-3 pixels, due in part to grouping errors, but also in part to errors in resected camera pose (even a perfectly segmented polygon boundary will not align with the projected ground truth roof if the camera projection parameters are incorrect).

Multiple-Image Performance: One of our underlying research hypotheses is that the use of multiple images increases the accuracy and reliability of the building extraction process. Rooftops that are missed in one image are often found in another, so combining results from multiple images typically increases the building detection rate. By combining detected polygons from four images, the total building detection rate increased to 81 percent using medium-sensitivity Boldt lines, and to 97 percent using high-sensitivity ones. Matching and triangulation to produce 3-D roof polygons, and thus the full building wireframe by extrusion, can perform at satisfactory levels of accuracy given only a pair of images, but using three views gives noticeably better results. After four images, only a modest increase in 3-D accuracy is gained.

Of course, any of these general statements depends critically on the particular configuration of views used. Futher testing is needed to elucidate how different camera positions and orientations affect 3-D accuracy. Nadir views appear to produce better detection rates than obliques, but this can be explained by large differences in GSD for this image set and may not be characteristic of system performance in general -- again, more experimentation is needed. For this data set, 3-D building corner positions were recovered to well within a meter of accuracy, with height being estimated more accurately than horizontal position. The accuracy of the final reconstruction depends on the accuracy of the detected 2-D polygons, as one might expect; however horizontal accuracy is more sensitive to 2-D polygon errors than vertical accuracy. How 3-D accuracy is related to errors in resected camera pose is an issue that is currently under analysis. Also, the version of Ascender tested here uses only a simple control strategy for detecting flat-roofed buildings, more complex control strategies under development may yield more robust results.

4. Ascender Delivery

In mid-April 1995, Version 1.0 of the Ascender system was delivered to Lockheed-Martin for testing on classified imagery; at the same time, an informal transfer was made to the NEL for familiarization and additional testing. Feedback from both of these groups has resulted in several system improvements and an overall "hardening" of the code.

Based on initial experience in the evaluation at the NEL, major changes have been made to Ascender's control system. The original system used a single reference image to generate roof hypotheses in the form of polygons, and then used the remaining images to verify/reject buildings by constructing a 3-D model. If a building hypothesis was not found in the reference image, the building would not be constructed *even though* it might be clearly visible in one or more of the other images. A new control strategy has been implemented under which all images are processed uniformly; polygons found in *any* image are used as the set of initial rooftop hypotheses from which the 3-D reconstruction begins. Table 5 summarizes the changes made to Ascender during the last six months of the contract.

Table 5. Major Changes to Ascender System After Initial Delivery to Prime Contractor (Lockheed-Martin).

Change	Effect
New control strategy uses all images for rooftop detection	Greatly improved detection of rootops and buildings
Improved graph search for rooftops	Improved detection
No longer detects multiple rooftop corners	Reduced graph size, increased speed
Fixed non-detection of rotated buildings	Improved detection
Fixed default parameters and system interface	absolute distance unit is the meter, 2.5D lines now default
Fixed detection of self-intersecting lines	Improved detection
Added Lockheed-Martin code changes	Facilitates transition to operational scenarios

Tests have been performed on a subregion of the Fort Hood dataset. Polygons were detected in seven images and redundant polygons eliminated on the basis of overlap. Each of the remaining polygons was then used to construct a 3-D building model. Models that had a side or height of less than 5 meters were eliminated. Using this scheme 92 percent of the 76 rooftop polygons were detected, leaving six polygons missed in all seven images. An additional 45 polygons represented false positives from either errors in the 2-D grouping process that survived verification or the reconstruction of a cultural feature other than a building (parking areas, playing fields, etc.) that had errors in height because of limited support from the image set.

None of these changes were reflected in the experimental evaluation described in the previous section. An attempt was made to deliver an updated Ascender system containing these changes to Lockheed-Martin and informally to the NEL just after the RADIUS contract expired. However, since no funds were available to cover the transition, installation, and evaluation costs, the improved version of the system was never delivered.

5. Grouping and Data Fusion

The building reconstruction strategies used in the Ascender system provide an elegant solution to extracting flat-roofed rectilinear buildings, but extensions are necessary in order to handle other common building types. Examples are multilevel flat roofs (or single-level flat roofs containing significant substructures, such as large air conditioner units), peaked-roof buildings, juxtapositions of flat and peaked roofs, curved-roof buildings such as Quonset huts or hangars, as well as buildings with more complex roof structures containing gables, slanted dormers or spires.

The building reconstruction strategies used in Ascender are reasonably effective, but are tuned to extract only one generic building class with single-level, flat roofs bounded by rectilinear polygonal shapes. As a result, polygonal rooftop detection strategies can easily be carried out entirely in 2-D. For example, verifying that a hypothesized 2-D image corner could be the projection of a horizontal roof corner in the 3-D scene can be performed based only on the orientation and angle of the 2-D corner in the image, together with the known camera pose information. Furthermore, determination of the probable height of a hypothesized rooftop polygon can be achieved using a simple one-dimensional histogram-based technique where the disparities of potential polygon line segment matches within epipolar search regions across multiple images vote directly for a consensus 3-D roof height in the scene.

To develop more general and flexible building extraction systems, a significant research effort is underway at UMass to explore alternative detection and reconstruction strategies that combine a wider range of 2-D and 3-D information. The types of strategies being considered involve generation and grouping of 3-D geometric tokens, such as lines, corners, and surfaces, as well as techniques for fusing geometric token data with high-resolution digital elevation map (DEM) data. By verifying geometric consistencies between 2-D and 3-D tokens associated with building components, larger and more complex 3-D structures are being organized using context-sensitive, knowledge-based strategies.

A more comprehensive description of the new types of extracted geometric features, and methods for grouping/fusing them, is given in [Jaynes et al., 1996]. Here, we briefly outline two of the new reconstruction strategies that have been developed as direct, incremental extensions to current Ascender technology: computation and grouping of 2.5-D line segments, and parameteric DEM surface fitting bounded by 2-D polygonal roof hypotheses.

5.1. Extracting/Grouping 2.5-D Lines

A 3-D scene line that is perpendicular to gravity can be represented as a 2-D image line segment Sets of 2.5-D lines are computed by taking 2-D Boldt line segments for an image and augmenting each with an elevation value computed via multiimage matching. The elevation estimate for each line segment is formed by histogramming the set of elevations implied by potential corresponding segments within epipolar-constrained search regions across multiple images. This is essentially the same algorithm that is used in Ascender to estimate the height of flat roof polygons in the scene, except it is applied to an individual line segment rather than to the set of edges bounding a polygonal roof hypothesis.

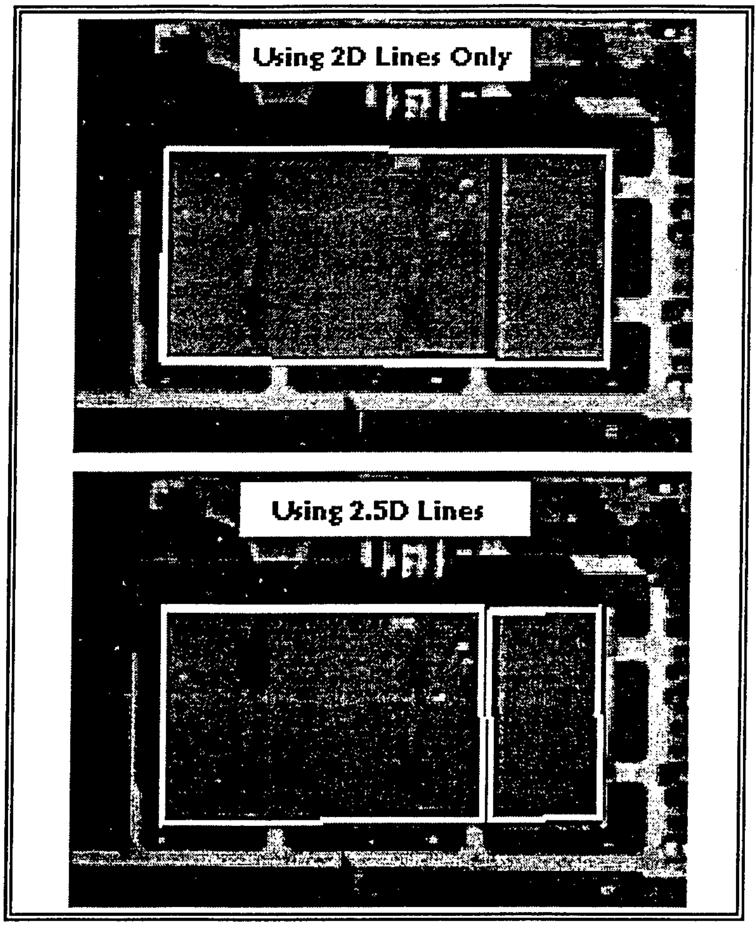

Figure 8. Using 2.5-D lines in the grouping process helps disambiguate multilevel building roofs (note the building shadow, which shows two distinct roof levels). The Z-coordinates of vertices on the left and right 2.5-D polygon hypotheses are 260.32 and 261.66 meters, respectively, as compared with ground truth Z-values of 260.65 and 262.31.

The graph-based perceptual organization algorithm used in Ascender for organizing lines and corners into closed 2-D polygons [Jaynes et al., 1996] has been modified to handle 2.5-D lines. An additional set of 3-D consistency checks have been introduced to ensure that compatible lines and corners are roughly at the same elevation in the scene. Individual line heights are combined and propagated into grouped corner, chain, and polygon hypotheses. The results are closed 2-D polygons with associated elevation values, which are easily converted into flat 3-D roof polygons using the known camera projection equations. The benefit of the 2.5-D approach to roof polygon detection is that image line segments caused by shadows and ground-level features are automatically ignored, and there is less chance of overgrouping multiple roof levels into a single polygon hypothesis containing edges that actually occur at different elevations in the scene (Figure 8).

5.2. Surface Fitting to DEM Data

A second building detection extension that has proven very effective is to directly fuse 2-D rooftop polygon hypotheses with high-resolution DEM data in order to estimate various classes of parametrically modeled 3-D rooftop surfaces. The DEM data is produced from a pair of overlapping images by hierarchical, area-based correlation matching along epipolar lines [Schultz 94]. In order to extract parametric surfaces, pixels within each detected roof polygon are backprojected onto the DEM data to determine a set of sampled 3-D points. Since the DEM data are potentially noisy, because of rooftop clutter and mismatches, robust statistical estimation techniques are used to do the fitting.

Three types of surface fits have been used to date: planar, peaked, and curved. An important issue is how to decide which parametric model to use for fitting the DEM data associated with a given rooftop hypothesis. In some cases building shadows can provide information about the profile of the rooftop. An alternative approach is to fit a number of different parametric classes simultaneously, and simply choose the one that best fits the data.

Figure 9 shows an example of three parametric peaked-roof surfaces that have been fit to the DEM data within local areas defined by building hypotheses generated by Ascender. It is important to run Ascender on nadir views in this case, since the goal is to make the system hypothesize a 2-D flat-roofed polygon that completely surrounds the peaked roof. Encoding this type of knowledge about how and when to apply such context-specific building extraction strategies is an important issue to consider when designing an operational vision system [Strat 93].

As a second case study, a set of buildings located at Fort Hood Texas were used for reconstruction. As before, the Ascender system generated the 2-D roof polygons and the 3-D elevation estimates were computed from the UMass terrain reconstruction system. An aerial view of the site and the detected roof polygons is shown in Figure 10. The image contained several peaked roof buildings and a large two-level flat roof building. Trees, shadows, and the fact that the image is at a much lower resolution than the image of the first case study makes detection and reconstruction of the buildings an interesting task.

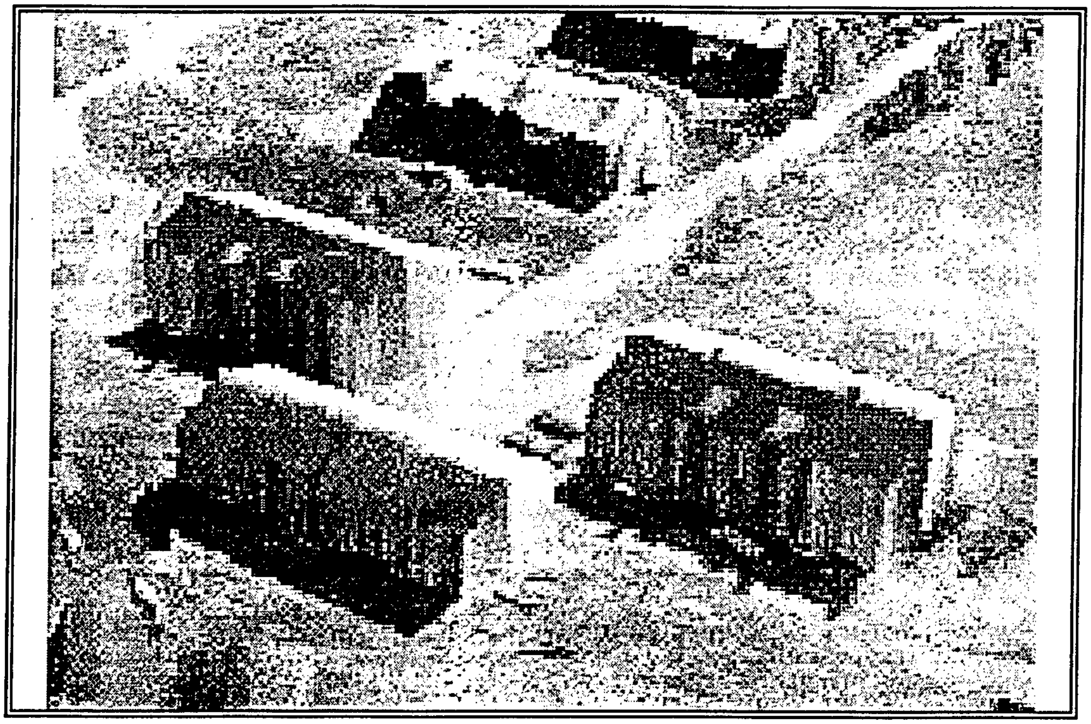

Figure 9. Three parametric peaked-roof surfaces that have been fit to DEM data within building boundaries hypothesized by Ascender. Compare with the raw DEM building data at the top of the image.

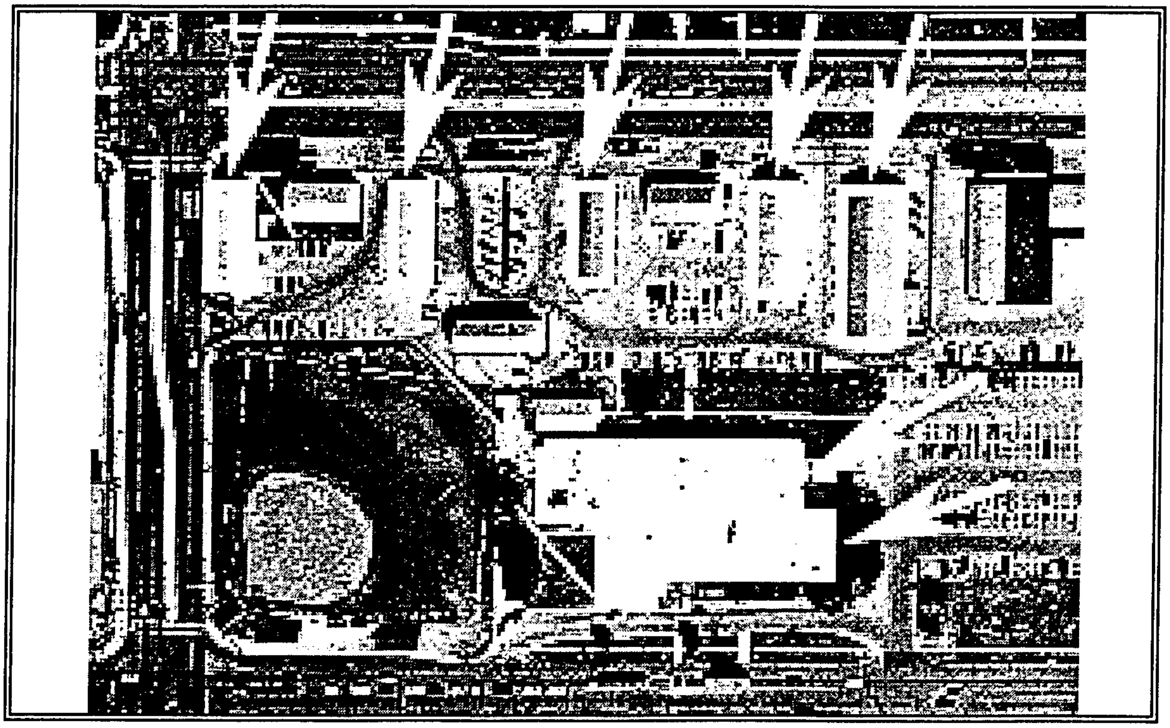

Figure 10. Subimage of the Fort Hood dataset with roof polygons detected.

The 2-D polygon detector was run on the image and seven polygons were detected (shown in Figure 10 by the arrows). Five of the polygons denote peaked roof buildings while two polygons represent the two sections of the large flat roofed building. The two flat roof polygons were fit to the elevation data using a 3-D planar roof model and a least-median squares robust. To ensure that two pieces of the flat buildings form a coherent single model, a set of heuristic rules embodying geometric knowledge of building structures are used to merge the fit planes into a single roof structure. A peaked roof model was fit to the elevation data within the five remaining polygons. In this case, the least-median-squares algorithm brings the model into alignment, determines the best peak angle, position of the ridge line, and height of the peak.

A ground plane was fit to the elevation estimates within a bounding box surrounding the seven polygons. Only elevation points exterior to the polygons are considered for the ground plane fit. For illustration purposes, the plane computed in the bounding box was extended to cover the entire site so that the model fitting results would be clearer. This plane may not actually reflect the actual terrain elevations. Figure 11 shows the site after this reconstruction process.

Figure 13 shows essentially the same techniques applied to the Martin-Marietta dataset. The Ascender system was used to detect rooftop polygons in one of the pair of nadir images. The individual polygons comprising the rooftop of the building were combined using geometric knowledge of rooftops. The digital elevation data shown in Figure 12 was used to fit planar surfaces within the Ascender polygons, and the resulting building model was then inserted into the digital elevation map. Note that Figure 13 shows the rounding of sharp geometric corners that is typical of correlation-based stereo reconstruction algorithms; the effect is caused by the fairly large correlation windows used for matching.

5.3. Extracting Surface Structures for Visualization

One of the benefits that a softcopy, 3-D model-based approach to site analysis has over the traditional 2-D image-based approach is that the image analyst can generate interactive, visual displays of the site from any viewpoint. Rapid improvements in the capability of low-end to medium-end graphics hardware makes the use of intensity mapping an attractive option for visualizing geometric site models, with near real-time virtual reality displays achievable on high-end workstations. These graphics capabilities have resulted in a demand for algorithms that can automatically acquire the necessary surface intensity maps from available digital photographs. Under the RADIUS project, UMass has previously developed routines for acquiring image intensity maps for the planar facets (walls and roof surfaces) of each recovered building model [Collins et al., 1995b, Collins et al., 1994]. Each surface intensity map is a composite formed from the best available views of that building face, processed to remove perspective distortion caused by obliquity and visual artifacts caused by shadows and occlusions. An example of a building from RADIUS Model Board 1 rendered using automatically acquired intensity maps is shown at the top of Figure 14.

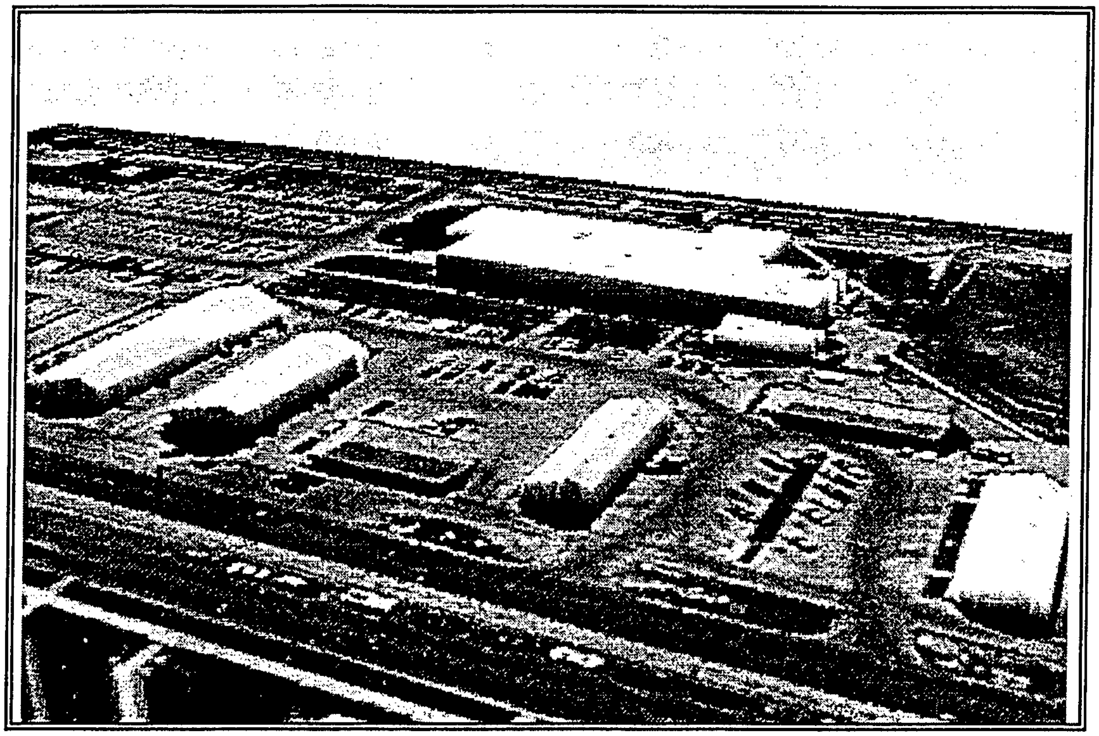

Figure 11. Six reconstructed buildings from the Fort Hood scene. Pixels that lay on the ground plane were darkened to highlight the results.

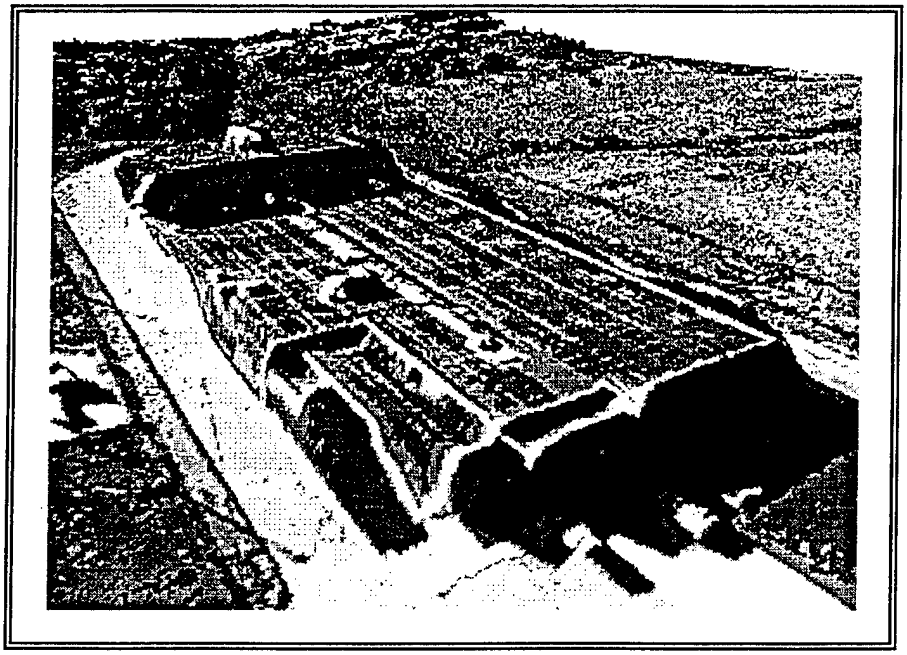

Figure 12. Three dimensional view of the Martin-Marietta (Denver) building constructed using the Terrest terrain reconstruction system.

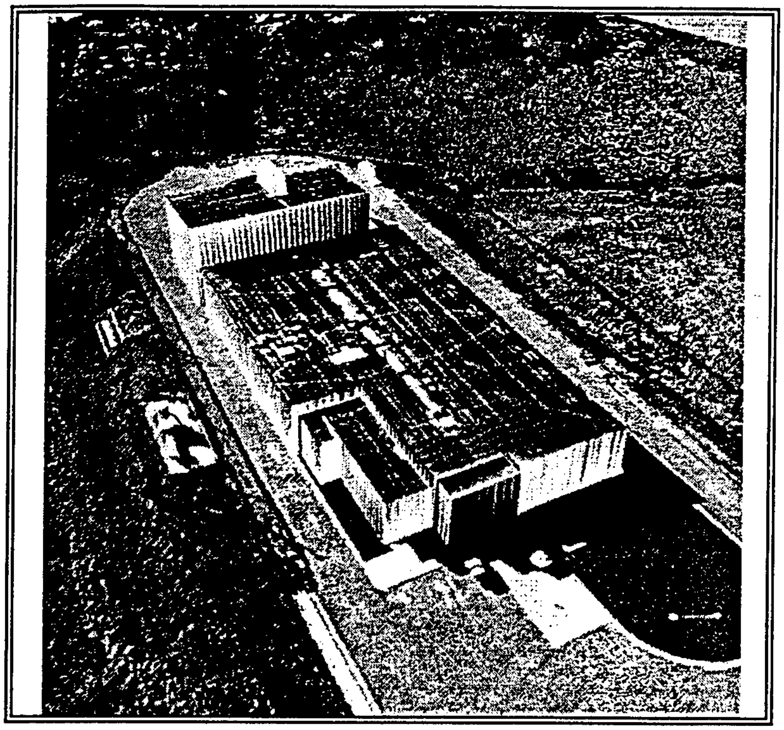

Figure 13. The same scene as shown in Figure 12 (although from a slightly different view) after replacing the building with a model.

Although intensity mapping enhances the virtual realism of graphic displays, this illusion of realism is greatly reduced as the observer's viewpoint comes closer to the rendered object surface. For example, straightforward mapping of an image intensity map onto a flat wall surface looks (and is) 2-D, unlike the surface of an actual wall. Windows and doors on a real wall surface are typically inset into the wall surface, and are surrounded by framing material that extends out beyond the wall surface. While these effects are barely noticeable from a distance, they are quite pronounced close up. A further problem is that the resolution of the surface texture map is limited by the resolution of the original image. As you move closer to the surface, more detail should become apparent, however, the graphics surface begins to look 'pixelated,' and features become blurry. In particular, some of the window features on the building models we have produced are near the limits of the available image resolution.

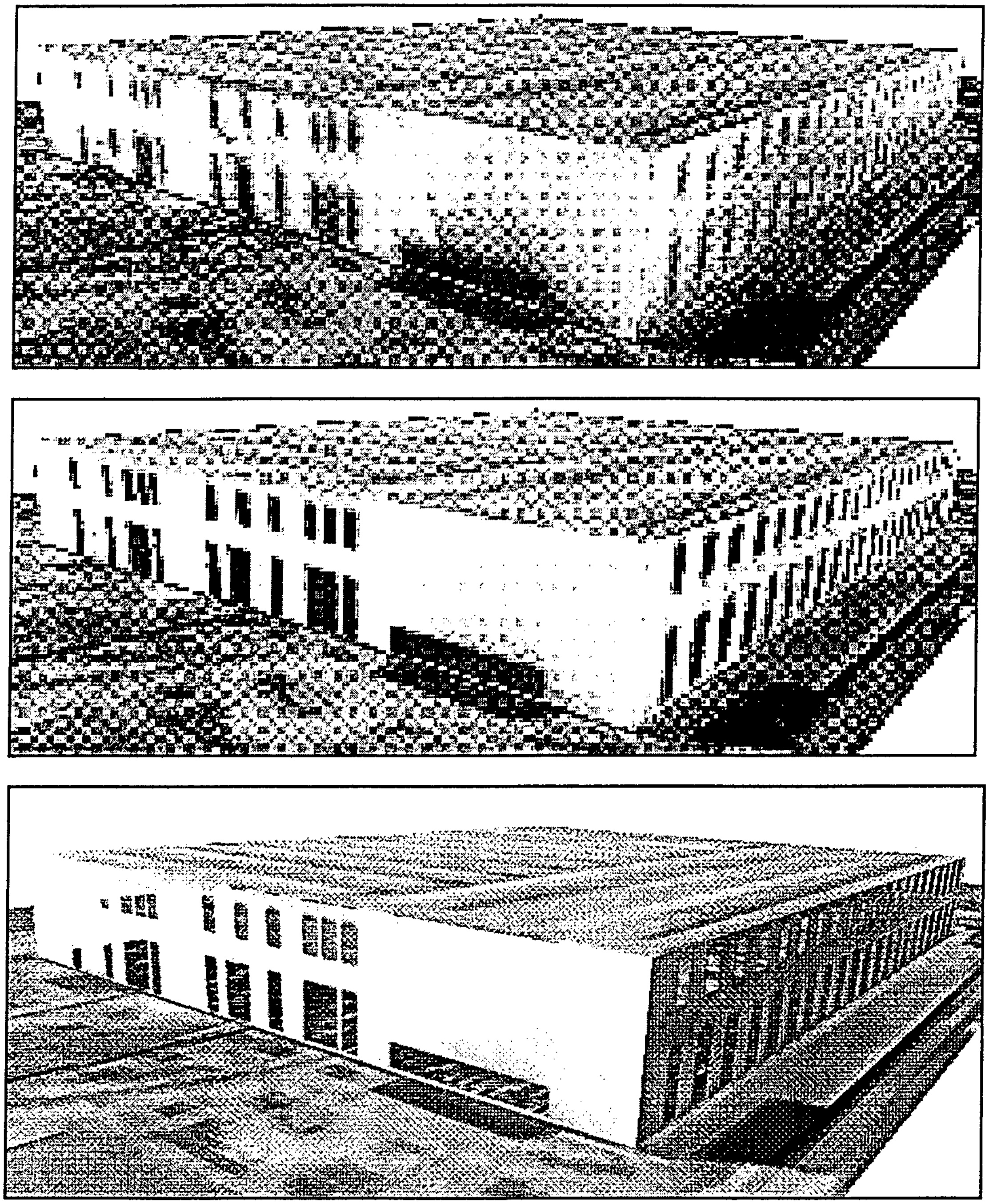

Figure 14. Rendered building model before symbolic window extraction (top), after extraction and modeling (middle), and after attachment of surface properties such as transparency (bottom).

What is needed to go beyond simple intensity mapping is explicit extraction and rendering of detailed surface structures, such as windows, doors, and roof vents. UMass' current intensity map extraction technology provides a convenient starting point, since rectangular lattices of windows or roof vents can be searched for without complication from the effects of perspective distortion, and specific surface structure extraction techniques can be applied only where relevant, i.e., window and door extraction can be focused on wall intensity maps, while roof vent computations are performed only on roofs. As one example, a generic algorithm has been developed for extracting windows and doors on wall surfaces, based on a rectangular region growing method applied at local intensity minima in the unwarped intensity map. Extracted window and door hypotheses are used to compose a refined building model that explicitly represents those architectural details. An example is shown in Figure 14. The windows and doors have been rendered as dark and opaque, but since they are now symbolicly represented, it would be possible to render the windows with glass-like properties, such as transparency and reflectivity.

Future work on extraction of surface structures will concentrate on roof features, such as pipes and vents, that appear as 'bumps' on an otherwise planar surface area. Visual cues for this reconstruction include shadows from monocular imagery, as well as disparity information between multiple images. This is a challenging problem given the resolution of available aerial imagery.

6. Summary and The Future

A large research effort is underway at UMass to develop capabilities for automated site modeling from aerial images. The Ascender system has been developed to extract and model flat-roofed, rectilinear buildings from multiple views. Version 1.0 of Ascender has been delivered to Lockheed-Martin for testing on classified imagery and for integration into the RADIUS Testbed. An evaluation of Ascender on an unclassified data set of Fort Hood has been performed at UMass. The results suggest that the system performs reasonably well in terms of detection rate and accuracy, and that performance degrades gracefully when the number of images used is small. Much more testing will be needed to determine how the system performs under various weather and viewing conditions, in order to formulate a set of recommendations as to how and when to use the system.

Algorithms and strategies for extracting other common building classes with peaked, curved, and multi-level flat roofs are being developed and tested in the lab for eventual inclusion into Ascender. Moving beyond a single control strategy for detecting a single class of buildings brings to the forefront issues of context-sensitive model class selection, data fusion, and hypothesis arbitration, and these topics are the focus of our current research efforts. Research on symbolic extraction of small surface features, such as windows and doors, also is being performed. Initial results show that the idea is feasible, although challenging, and the payoff is large in terms of realistic scene rendering.

8. Bibliography

Boldt, M., R.Weiss and E.Riseman, “Token-Based Extraction of Straight Lines,” *IEEE Trans. on Systems, Man and Cybernetics*, vol. 19(6), 1989, pp. 1581-1594.

Collins, R., Y. Cheng, C. Jaynes, F. Stolle, X. G. Wang, A. Hanson and E. Riseman, “Site Model Acquisition and Extension from Aerial Images,” *International Conference on Computer Vision*, Cambridge, MA, June 1995a, pp. 888-893.

Collins, R., Y. Cheng, C. Jaynes, F. Stolle, X. Wang, A. Hanson and E. Riseman, “A System for Automated Site Model Acquisition,” *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, SPIE vol. 7617, Orlando, FL, April 1995b, pp. 244-254.

Collins, R., A. Hanson and E. Riseman, “Site Model Acquisition under the UMass RADIUS Project,” *DARPA Image Understanding Workshop*, Monterey, CA, November 1994, pp.351-358.

Gerson, D. and S. Wood, “RADIUS Phase II - The RADIUS Testbed System,” *DARPA Image Understanding Workshop*, Monterey, CA, November 1994, pp. 231-237.

Jaynes, C., F. Stolle, H., Schultz, R., Collins, A., Hanson, and E., Riseman,, “Three Dimensional Grouping and Information Fusion for Site Modeling from Aerial Images,” *Proc. DARPA Image Understanding Workshop*, Palm Springs, CA, February 1996, vol. I, pp. 479-490.

Jaynes, C., F. Stolle and R. Collins, “Task Driven Perceptual Organization for Extraction of Rooftop Polygons,” *IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, December 1994, pp. 152-159.

Mundy, J., R. Welty, L. Quam, T. Strat, W. Bremner, M. Horwedel, D. Hackett and A. Hughes, “The RADIUS Common Development Environment,” *DARPA Image Understanding Workshop*, San Diego, CA, January 1992, pp. 215-226.

Roux, M., Y. Hsieh, D.McKeown, “Performance Analysis of Object Space Matching for Building Extraction using Several Images,” in Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, D. McKeown and I. Dowman, Editors, *Proc. SPIE 2486*, pp. 277-297, 1995.

Schultz, H., “Terrain Reconstruction from Oblique Views,” *DARPA Image Understanding Workshop*, Monterey, CA, November 1994, pp. 1001-1008.

Strat, T., “Employing Contextual Information in Computer Vision," *DARPA Image Understanding Workshop*, Washington, D.C., April 1993, pp. 217-229.