UMass Progress in 3D Building Model Acquisition*

Robert T. Collins, Allen R. Hanson, Edward M. Riseman
Christopher O. Jaynes, Frank Stolle,
Xiaoguang Wang, and Yong-Qing Cheng

Department of Computer Science
Lederle Graduate Research Center
Box 34610, University of Massachusetts
Amherst, MA. 01003-4610

Abstract

The Automated Site Construction, Extension, Detection and Refinement system (ASCENDER) has been developed to automatically populate a site model with buildings extracted from multiple, overlapping views. Version 1.0 of the system has been delivered for evaluation on classified imagery. Evaluation results on an unclassified Ft.Hood data set are presented here. Extensions to the system that allow it to detect a wide range of building classes, including peaked roof and multi-level flat roofed structures are described. Recent work on symbolic extraction of surface structures such as windows greatly enhances the visual realism of graphical site model displays.

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 [5]. Automated construction and management of 3D 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 due to 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 3D site models are valuable for training and mission planning. Civilian benefits of this technology are also 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 3D 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. To overcome these difficulties, the UMass design philosophy incorporates several key ideas. First, 3D 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 3D 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.

This paper is organized as follows. 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, and is currently being evaluated on classified imagery at Lockheed-Martin. Section 3 presents results of an evaluation

*Funded by the RADIUS project under ARPA/Army TEC contract number DACA76-92-C-0041 and by NSF grant number CDA-8922572.

conducted at UMass on an unclassified data set of Ft.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 and in [9] (these proceedings). 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 [5]. 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 [2, 3, 4]. Some sample building models automatically generated by the Ascender system are shown in Figures 1 and 2.

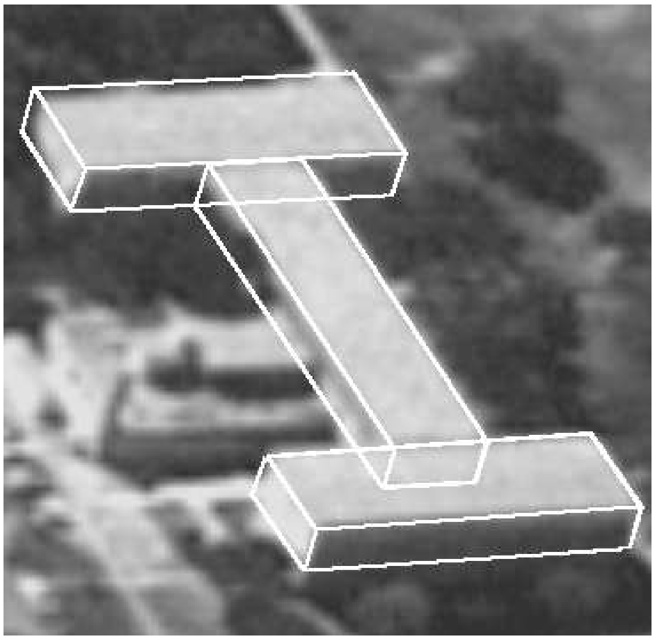

Figure 1: Sample building model automatically generated by the Ascender system.

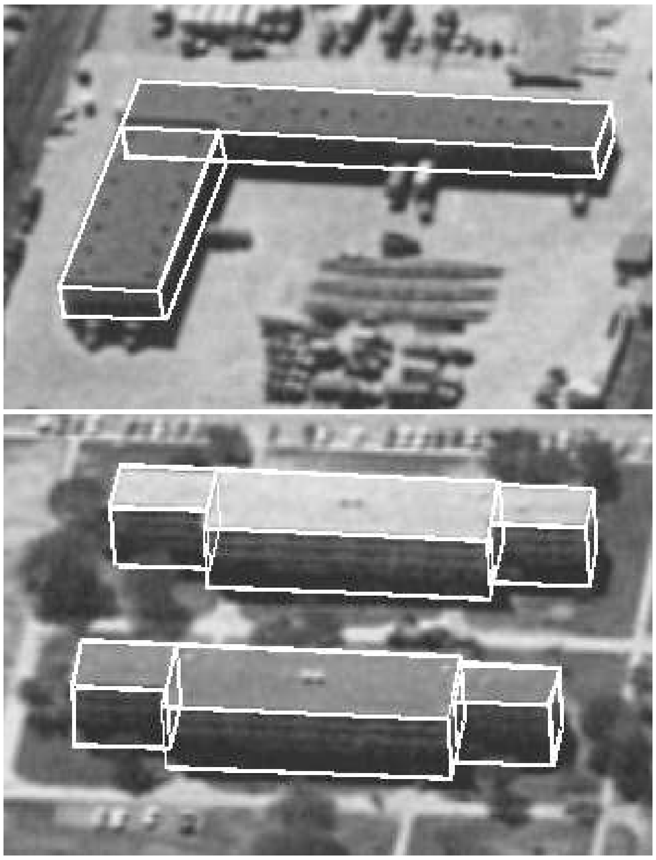

Figure 2: Some additional samples of building models generated by Ascender.

2.1 System Overview

Ascender was developed on a Sun Sparc 20, using the Radius Common Development Environment (RCDE) [7]. 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:

• **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 3D locations in the site coordinate system are related to 2D image pixels in each image. One common camera representation is a 3×4 projective transformation matrix encoding the lens and pose parameters of each perspective camera. Ascender can also 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 [1]. Intensity edgels 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 [6]. Lines 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 3D building information begins by estimating a height for each hypothesized 2D roof polygon via multi-image 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 2D 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 photogrammtric triangulation procedure is performed to determine the precise 3D size, shape and position of the building rooftop. The optimization criterion simultaneously minimizes the sum-of-squared residual errors between projected 3D roof polygon edges and corresponding line segment features in all the images. There are no user parameters. The resulting 3D polygon is then extruded down to the provided terrain to form a complete building wireframe.

3 Evaluation on Ft. Hood Imagery

The success of the Ascender system will ultimately be judged by its performance on classified imagery. Such tests are currently being 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 like:

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

This section presents evaluation results on a large data set from Ft.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 Ft.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 760x740 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.

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

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.

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 2D and triangulated 3D roof polygons vs. their ground truth counterparts. There are 73 ground truth rooftop polygons among the set of 30 buildings. Ground truth 2D polygons for each image are determined by projecting the ground truth 3D 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 polygons boundaries, reported in pixels for 2D polygons, and in meters for 3D polygons.

[1]Robert Haralick, private communication.

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 2D polygons these *Inter-Vertex Distances* are reported in pixels, for 3D 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 2D Detection

One important module of the Ascender system is the 2D 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 2D bounding box out by 20 pixels on each side, then invoking the building detector in that region to hypothesize 2D 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 2D 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. Determining true and false positive detection rates thus 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 2D 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.

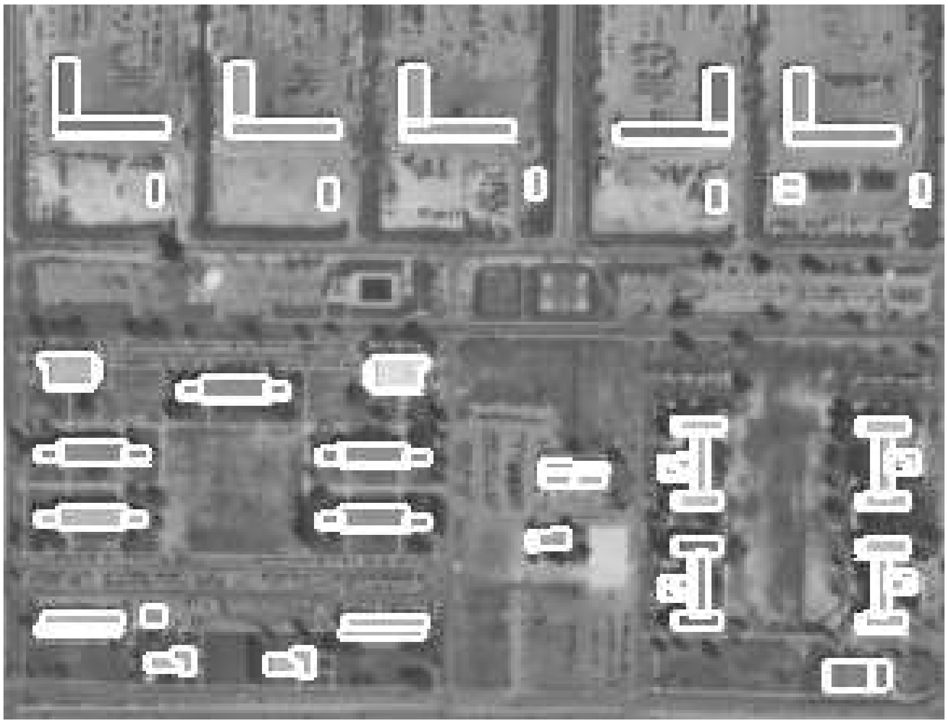

Figure 3: Ft.Hood evaluation area with 30 ground truth building models composed of single- and multi-level flat roofs, and two peaked roofs. There are 73 roof facets in all. The size of the image area shown is 2375x1805 pixels.

The threshold value used was 0.2 times the square root of the area of the ground truth polygon, that is: $\text{Dist(hyp,gt)} \leq 0.2\sqrt{\text{Area(gt)}}$, 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%, 59%, 45% and 47%, 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 [2]. For the highest sensitivity setting, this strategy detects 81% (59 out of 73) of the rooftops in the scene.

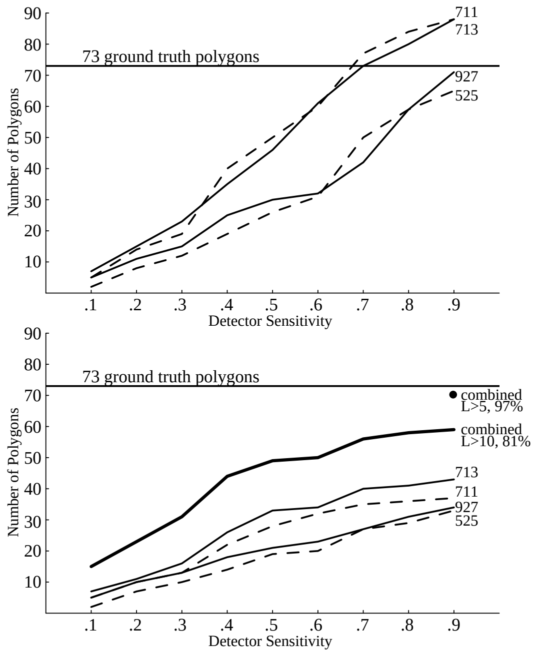

Figure 4: Top: Building detector sensitivity vs. total number of roof hypotheses. Bottom: Sensitivity vs. number of true positives. Horizontal lines show the actual number of ground truth polygons. Combining results from all four views yields a "best" detection rate of 81% with lines of length > 10, and 97% 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 is also likely that most of the discrepancy is due to the difference in GSD of the images for this area (see Table 1). Each building roof simply 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%, 84%, 74%, and 67%, with a combined image detection rate of 97% (71 out of 73).

3.2.2 Quantitative Accuracy

To assess the quantitative accuracy of the true positive 2D roof polygons, each was compared with its corresponding 2D 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 2D 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.

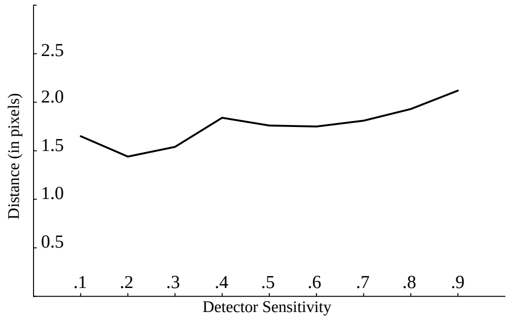

Figure 5: Building detector sensitivity vs. 2D polygon accuracy in pixels (see text).

For pairs of detected and ground truth polygons having the same number of vertices, their set of inter-vertex distances were also computed, and the medians of those measurements are broken down by image in Table 2. The average distance is around 2.7 pixels. Polygons detected in image 927 appear to be a little 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, so the difference in sun angle may be causing it.

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

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

3.3 Evaluation of 3D Reconstruction

The second major subsystem in Ascender takes 2D roof hypotheses detected in one image and reconstructs 3D rooftop polygons via multi-image line segment matching and triangulation. Two different quantitative evaluations were performed on this subsystem. The 3D reconstruction process was first tested in isolation from the 2D detection process by using 2D 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 3D building models would be given perfect 2D rooftop extraction. A second evaluation tested end-to-end system performance by performing 3D reconstruction using the set of automatically detected 2D image polygons from the previous section.

3.3.1 Baseline Reconstruction Accuracy

The baseline measure of reconstruction accuracy was performed using 2D projected ground truth roof polygons. For each of the 7 images in the evaluation test set, all the ground truth 2D polygons from that image were matched and triangulated using the other 6 images as corroborating views. The accuracy of each reconstructed roof polygon was then determined by comparing it with its 3D 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.

	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

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

Another suite of tests was performed to determine how the number of views affects the accuracy of the resulting 3D polygons. These tests were performed using image 711 as the primary image, and all 63 non-empty subsets of the other 6 views as additional views. For each subset of additional views, all 2D 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 3D 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. The distances

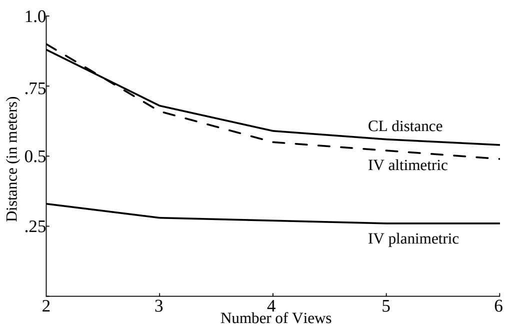

Figure 6: Number of views used vs. 3D reconstruction accuracy in meters. See text.

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 2D 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 2D image accuracy of the hypothesized roof polygons. The typical end-to-end performance of the system was evaluated by taking the 2D polygons detected in Section 3.2.1 and performing matching and triangulation using the other six views. The median center-line distances between reconstructed and ground truth 3D 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 2D polygons at those levels (compare with Figure 5).

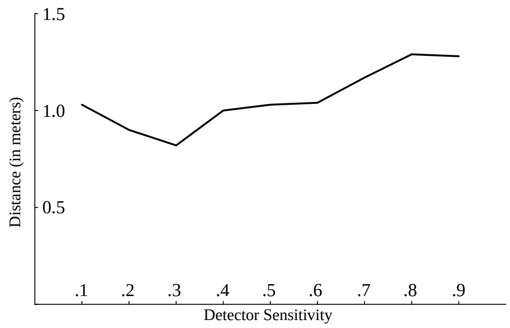

Figure 7: Building detector sensitivity vs. 3D polygon accuracy, computed as the median of center-line distances between reconstructed 3D polygons and ground truth roof polygons.

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 2D 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 2D 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 [4].

	711	713	525	927
IV planimetric	0.68	0.73	1.09	0.89
IV altimetric	0.51	0.55	0.90	0.61

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

3.4 Summary

This section has presented preliminary results of an on-going evaluation of the Ascender system using an unclassified Ft.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% 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% 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 2D 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% using medium-sensitivity Boldt lines, and to 97% using high-sensitivity ones. Matching and triangulation to produce 3D 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 3D 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 3D 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, 3D 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 2D polygons, as one might expect; however horizontal accuracy is more sensitive to 2D polygon errors than vertical accuracy. How 3D accuracy is related to errors in resected camera pose is an issue that is currently under analysis. Also, the version of Ascender tested here using only a simple control strategy for detecting flat-roofed buildings, more complex control strategies under development may yield more robust results.

4 3D 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 multi-level 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.

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 2D and 3D information. The types of strategies being considered involve generation and grouping of 3D 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 2D and 3D tokens associated with building components, larger and more complex 3D 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 [9] (these proceedings). 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 2.5D line segments, and parameteric DEM surface fitting bounded by 2D polygonal roof hypotheses.

4.1 Extracting/Grouping 2.5D Lines

A 3D scene line that is perpendicular to gravity can be represented as a 2D image line segment plus its associated scene elevation. We call this representation "2.5D" line segments. Sets of 2.5D lines are computed by taking 2D Boldt line segments for an image and augmenting each with an elevation value computed via multi-image 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.

The graph-based perceptual organization algorithm used in Ascender for organizing lines and corners into closed 2D polygons [6] has been modified to handle 2.5D lines. An additional set of 3D 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 2D polygons with associated elevation values, which are easily converted into flat 3D roof polygons using the known camera projection equations. The benefit of the 2.5D 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 el-

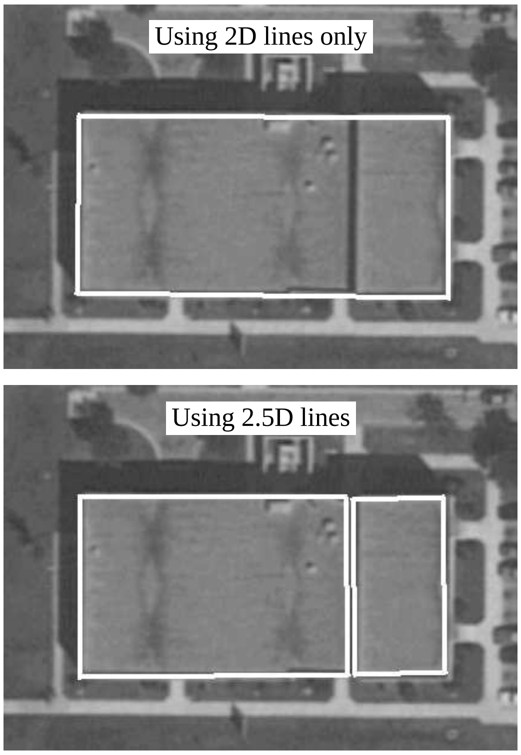

Figure 8: Using 2.5D lines in the grouping process helps disambiguate multi-level building roofs (note the building shadow, which shows two distinct roof levels). The Z-coordinates of vertices on the left and right 2.5D polygon hypotheses are 260.32 and 261.66 meters, respectively, as compared with ground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

evations in the scene (Figure 8).

4.2 Surface-Fitting to DEM Data

A second building detection extension that has proven very effective is to directly fuse 2D rooftop polygon hypotheses with high-resolution DEM data in order to estimate various classes of parametrically modeled 3D rooftop surfaces. The DEM data is produced from a pair of overlapping images by hierarchical, area-based correlation matching along epipolar lines [8]. In order to extract paramteric surfaces, pixels within each detected roof polygon are backprojected onto the DEM data to determine a set of sampled 3D points. Since the DEM data is potentially noisy, due to 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 2D 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 [10].

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.

5 Extracting Surface Structures for Visualization

One of the benefits that a softcopy, 3D model-based approach to site analysis has over the traditional 2D 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 [3, 4]. 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 10.

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) two dimensional, unlike the surface of an actual wall. 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, but instead, 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.

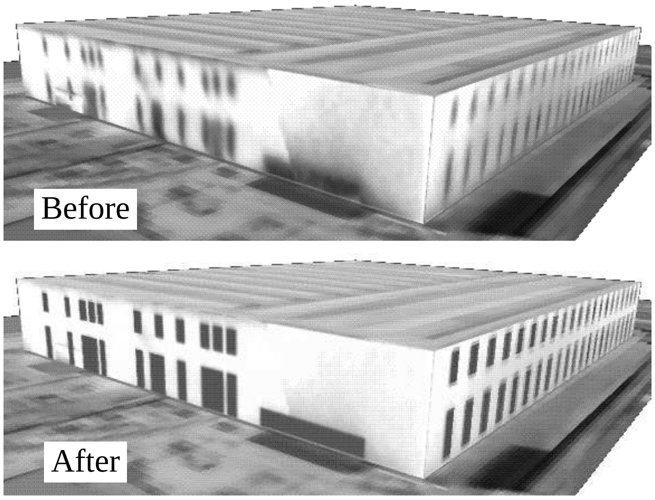

Figure 10: Rendered building model before and after symbolic window extraction.

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 10. 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 On-Going Work

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 Ft.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 is also being performed. Initial results show that the idea is feasible, although challenging, and that the payoff is large in terms of realistic scene rendering.

References

[1] M.Boldt, 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.

[2] R.Collins, Y.Cheng, C.Jaynes, F.Stolle, X.Wang, A.Hanson and E.Riseman, "Site Model Acquisition and Extension from Aerial Images," *International Conference on Computer Vision,* Cambridge, MA, June 1995, pp. 888–893.

[3] R.Collins, C.Jaynes, F.Stolle, X.Wang, Y.Cheng, 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 1995, pp. 244–254.

[4] R.Collins, A.Hanson and E.Riseman, "Site Model Acquisition under the UMass RADIUS Project," *Arpa Image Understanding Workshop,* Monterey, CA, November 1994, pp. 351–358.

[5] D.Gerson and S.Wood, "RADIUS Phase II – The RADIUS Testbed System," *Arpa Image Understanding Workshop,* Monterey, CA, November 1994, pp. 231–237.

[6] C.Jaynes, 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.

[7] J.Mundy, R.Welty, L.Quam, T.Strat, W.Bremner, M.Horwedel, D.Hackett and A.Hughes, "The RADIUS Common Development Environment," *Arpa IUW,* San Diego, CA, Jan 1992, pp. 215–226.

[8] H.Schultz, "Terrain Reconstruction from Oblique Views," *Arpa Image Understanding Workshop,* Monterey, CA, Nov 1994, pp. 1001–1008.

[9] H.Schultz et.al., "Three-Dimensional Grouping and Information Fusion for Site Modeling from Aerial Images," *Arpa IUW, 1996,* these proceedings.

[10] T.Strat, "Employing Contextual Information in Computer Vision," *Arpa Image Understanding Workshop,* Washington, DC, April 1993, pp. 217–229.