Building Reconstruction from Optical and Range Images*

Christopher O. Jaynes, Allen Hanson, Edward Riseman, and Howard Schultz
Computer Vision Laboratory
Dept. of Computer Science,
Box 34610 University of Massachusetts, Amherst

Abstract

A technique is introduced for extracting and reconstructing a wide class of building types from a registered range image and optical image. An attentional focus stage, followed by model indexing, allows top-down robust surface fittting to reconstruct the 3D nature of the buildings in the data. Because of the effectiveness of model selection, top-down processing of noisy range data still succeeds and the algorithm is capable of detecting and reconstructing several different building roof classes, including flat single level, flat multi-leveled, peaked, and curved rooftops.

The algorithm is applicable to range data that may have been collected from several different range sensor types. We demonstrate reconstructions of different buildings classes in the presence of large amounts of noise. Our results underline the usefuless of range data when processed in the context of a focus-of-attention area derived from the monocular optical image.

1 Introduction

We introduce a solution to the problem of building reconstruction from aerial images. The technique presented here supports the reconstruction of a wide class of building models and is robust to sensor noise. The reconstructed models may be used for urban planning, three dimensional visualization for simulated walk/fly throughs, and as a geometric representation capable of supporting additional image understanding tasks.

The detection and accurate reconstruction of three dimensional scene structures from aerial data proves to be difficult in both optical and range data. Within optical imagery, building boundaries are occluded by other features at the site and rooftop regions contain varying surface structure and surface markings. Problems in range images are similarly complex. Depth discontinuities at the boundaries of buildings cause typical correlation-based stereo optical systems to degrade. Likewise, depth computed from radar depends on surface geometry and material properties of objects in the scene.

A successful automated reconstruction system will make use of both range and optical data in order to overcome the inherent problems in each. The system should be robust with respect to a wide variety of range sensors including Interferometric Synthetic Aperature Radar (IFSAR) and optical stereo. This paper introduces one such system that demonstrates how appropriate use of both the registered optical image and range data increases reconstruction accuracy, reliability and completeness.

There has been a large amount of work in aerial image interpretation, and building reconstruction in particular, using both monocular and multiple image strategies [Collins, Jaynes, et. al '96, Jaynes'94, Herman'94, Matsuyama'85, McKeown'90]. In earlier work, perceptual organization techniques provided the impetus for many building extraction systems. Heurtas and Nevatia [Huertas'80], organize lines and corners into possible rectangles and select the best possible set of groupings from the hypotheses. The ASCENDER system [Collins'95] hypothesized 2D building rooftops through a perceptual grouping scheme [Jaynes'94] and computed a height estimate for each roof through multi-image triangulation. The system assumed flat roofed buildings and extruded the rooftop polygon to a known ground plane.

More recently, the utility of range images for site reconstruction has been recognized. Kim and Muller [Kim'95] extract rooftop boundaries from optical data and use an elevation map to estimate rooftop height. Haala and Hahn [Haala'95] search an elevation map for local maxima, with three dimensional lines computed in these regions used for parametric model fits to the extracted line segments. The approach estimates the initial parameters for model fitting, but assumes that the buildings at the site can be reconstructed using a single parametric model (e.g. peaked roof model).

All these approaches assume a small class of building types (typically flat roofs), and, in the case of the ASCENDER system, where no elevation map is utilized, require several registered optical images to ar-

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

rive at an accurate 3D model. The work presented here makes few assumptions about the shape of the building rooftop, other than it can be composed from a set of surface shapes defined in an existing database.

Our approach to the problem has been threefold: 1) We focused on extraction techniques that were not restricted to a small class of buildings. Instead, automatic classification of the different surface primitives are combined to reconstruct a wide class of building types. Examples are multi-level flat roofs (or single level structures with significant substructures such as air conditioner units), peaked roof buildings, curved-roof buildings, such as Quonset huts or hangars. 2) Our goal is to acquire 3D models even at the limits of feasibility due to noise and the size of the structures being extracted. In order to accomplish this, the use of top-down model application is used to arrive at a correct reconstruction in the presence of significant noise and ambiguities. We make use of a database of surface types that represent geometrically feasible buildings (51 models, in eight classes). 3) The system is fully automatic. Although user intervention can be a valuable source of information, the system attempts to segment, classify, and reconstruct the site completely automatically with as little sensitivity to parameter settings as possible.

2 Segmentation

Segmentation takes place in the optical image using a perceptual grouping scheme first presented in the ASCENDER I system [Collins'95]. The segmented regions from the optical image provide a focus of attention for surface reconstruction within the range image. Figure 1 shows the 450x450 pixel optical and registered range images for an area located at Fort Hood, Texas that will be used to demonstrate the reconstruction process. The range image was constructed using the UMass TERREST [Schultz'94] system from an optical pair of the region.

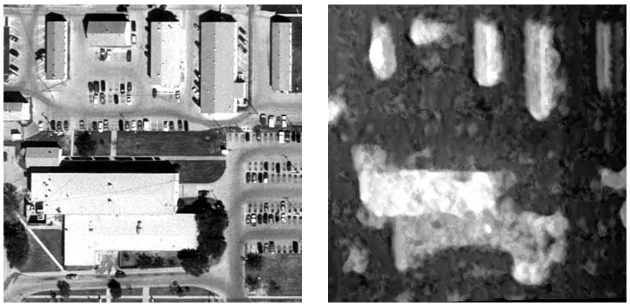


Figure 1: An optical image and corresponding stereo range data for an area of Ft. Hood Texas. Elevation values at right are coded as image brightness.

Segmentation of rooftop boundaries is based on a perceptual grouping scheme [Jaynes'94] that has been shown to be effective in delineating rooftop boundaries in aerial images [Collins, Jaynes, et. al '96]. It is important to note, however, that alternative segmentation schemes can be used to focus the reconstruction process.

Low level features in the segmentation module are straight line segments and corners. We assume that significant rooftop surfaces can be delineated with a flat rectilinear polygons. This implies a search for polygons made up of straight line segments and orthogonal corners (although orthogonal corners in the world are not necessarily orthogonal in the scene when oblique views are processed). To determine a set of relevant corner hypotheses, pairs of line segments with spatially proximate endpoints are grouped together into candidate image corner features. Using the known camera pose, each potential image corner is then back-projected into a nominal Z-plane in the scene, and the hypothetical *scene corner* is tested for orthogonality.

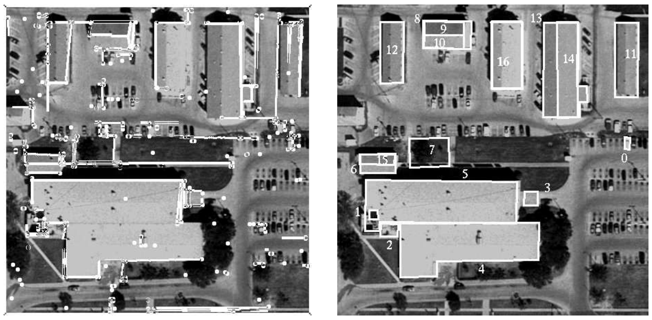
Geometrically, collated features are sequences of geometrically grouped corners and lines that form a chain (Figure 2). Chains are a generalization of the collated features in earlier work [Huertas'80] and allow final polygons of arbitrary rectilinear shape to be constructed from the low level features.

Collated feature chains are represented by paths in a feature relation graph. Low level features (corners and line segments) are nodes in the graph, and perceptual grouping relations between these features are represented by edges in the graph. Nodes have a certainty measure that represents the confidence of the low level feature extraction routines; edges are weighted with the certainty of the grouping that the edge represents. A chain of collated features inherits an accumulated certainty measure from all the nodes and edges along its path.

High Level Polygon hypothesis extraction proceeds in two steps. First, all possible polygons are computed from the collated features. Then, polygon hypotheses are arbitrated in order to arrive at a final set of non-conflicting, high confidence rooftop polygons (Figure 2). All of the cycles in the feature relation graph are searched for in a depth first manner, and stored in a dependency graph where nodes represent complete cycles (rooftop hypotheses). Nodes in the dependency graph contain the certainty of the cycle that the node represents. An edge between two nodes in the dependency graph is created when cycles have low level features in common. The final set of non-overlapping rooftop polygons is the set of nodes in the dependency graph that are both independent (have no edges in common) and are of maximum certainty.

3 Classification

The segmentation module produces a set of two-dimensional closed regions within the optical data that represent high-confidence building rooftop hypothe-



(a) (b)

Figure 2: (a) Perceptually grouped corner and line segments that are represented in a feature relation graph. (b) The final maximally weighted set of closed cycles.

ses. To reconstruct the 3D structure, parameterized models (called surface primitives) are fit to each corresponding region in the registered range data independently. The final buildings are reconstructed as a composition of these primitives. The work here presents the case when each building is modeled as a separate primitive; however, work is underway to merge multiple models into a single complex building. The key to a reconstruction system that is able to address a wide class of building types lies in the system's ability to automatically classify the type of surface primitive associated with each region, and to constrain the parameters of a robust fit (see Section 4) sufficiently to arrive at an accurate solution. Furthermore, the method must be invariant with respect to translation, scale, and noise.

The classification scheme indexes into a database of surface primitives based on an analysis of differential geometry within each region in the range image. The surface orientations of small surface patches are estimated and an orientation histogram is constructed that is then correlated with an existing library of roof models. These orientation histograms, sometimes called the Extended Gaussian Image, are normalized so that they are both scale and translation invariant. A detailed introduction to the Extended Gaussian Image can be found in [Horn'86].

The surface primitive database (SPD) contains a set of surface classes called *surface primitives*, such as planes, cylindrical surfaces, peaks, and spires, known to typically be part of rooftop surfaces. Associated with each surface primitive are a number of models, representing different parameterizations of each class of surface primitives. For example, the "Peak" surface primitive class is the canonical shape for a number of models in the SPD, each with a different peak angle. Corresponding orientation histograms are stored in the SPD for indexing purposes. Figure 3 shows the SPD used for the results shown here. It contains 8 different surface primitives and 51 total models.

For each of the segmented regions, an orientation histogram is constructed and correlated with the set of models stored in the SPD. The set of points within a region are triangulated into a simple surface using the Delauney algorithm [Aurenhammer'91]. The triangulated surface is a set of triangular surface patches, $T_I = (p_1, p_2, p_3)$, where p_1, p_2, p_3 are datapoints from the original pointset. Figure 4 shows the triangulated surface that corresponds to polygon 11 in Figure 2b.

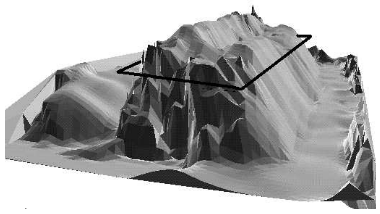


Figure 4: The Delauney representation of the elevation data. The polygon shown corresponds to the upper-rightmost polygon (polygon 11) from figure 2.

The surface normal at each triangular patch is then computed as the cross product of the Vectors $v_1 = \frac{(p_2-p_1)}{||p_2-p_1||}$ and $v_2 = \frac{p_3-p_1}{||p_3-p_1||}$. Because we assume the normal of the plane representing the footprint of the rooftop is aligned with the gravity vector, the surface normal pointing in the positive Z direction is used to determine the cell on the Gaussian sphere that will receive a "vote" for a particular orientation.

To avoid sensitivity problems with the method in which orientation space is discretized, votes are smoothed over the sphere via a Gaussian function. If the surface normal, N, intersects the Gaussian sphere at (x, y, z), the weighted vote at B is given by:

$$V(x, y, z, B) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(D^2/\sigma^2)} \tag{1}$$

where D is the angular distance from (x, y, z) to the center of the histogram bucket, B, to receive the weighted vote. The amount of smoothing is related to the expected noise in the range image, however as σ increases the separability of the model classes degrades. For the results shown here, $\sigma = 0.3$ and the orientation histogram contains 240 buckets, reflecting a tesselation based on the semiregular icosahedron.

A single surface normal may induce a smoothed vote over several buckets, given by equation 1, and voting for a given vector stops when the bucket value of $V(x, y, z, B)$ falls below a threshold (0.1 for the results shown here).

Figure 5 shows the histogram constructed from the range data corresponding to polygon 11 in the Fort Hood example. Histograms visualized in this way provide an interesting characterization of the noise within

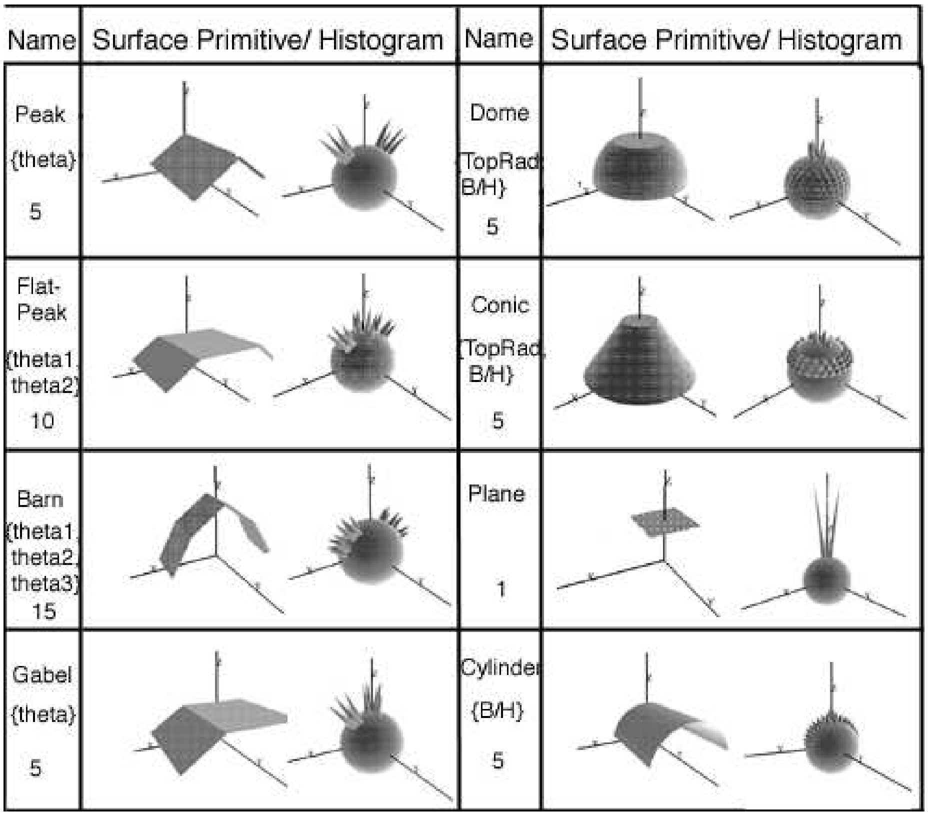


Figure 3: The Surface Primitive Database used for indexing. Each cannonical surface primitive is shown along with the orientation histogram (see text). The parameterization for each model is shown to the left. Several different models, each with unique parameters, are stored in the database. The number of different models is shown to the left.

a surface patch. Notice, for example, that the underlying peak structure is discernable, but that number of votes for each of the "lobes" are unequal.

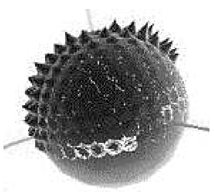


Figure 5: The surface orientation histogram for the surface contained within the polygon shown in figure 4.

To achieve model indexing, the constructed histogram, referred to as the image histogram, is then correlated with each of the model histograms stored in the SPD. The correlation process is thus independent of scale. The normalized cross-correlation score, given by:

$$\mathcal{C}_\theta(I, M) = \frac{\sum^{(i,j)}(I(i,j) - \mu_I)(M(i,j) - \mu_M)}{(\sigma_I * \sigma_M)} \quad (2)$$

where μ and σ represent the mean and the variance respectively.

The method not only selects the correct surface model class (peak roof, for example) based on the correlation score, but estimates the set of parameters for surface fitting (angle of the peak roof). To select the correct model orientation on the ground plane, the value of $\mathcal{C}_\theta(I, M)$ must be computed for many possible values of θ that represent different alignments between the spherical histograms I and M. For the results in this paper, θ was restricted to rotations about the Z-axis under the gravitational alignment assumption. The plane model, however, was allowed any possible orientation to reflect the possible presence of flat, sloped roofs or a sloping ground plane. The value of θ for which $\mathcal{C}$ is a maximum is stored and compared to the results of other correlations in the database. The highest scoring models are used for robust fitting during the final reconstruction phase. For the results here,

Model Name	Correlation Score	
Peak (130)	0.8813	$\theta = 0.0$
Peak (150)	0.8320	$\theta = 0.0$
Flat Peak (65,65)	0.8054	$\theta = 0.0$
Flat Peak (75,75)	0.7900	$\theta = 0.0$

Table 1: Polygon 11 Correlation Match Scores

the top two models are selected and fit to the data; a residual fit error is used to select the final model.

Table 1 below shows the results of correlating the histogram shown in figure 5 (constructed from polygon 11) with the SPD. The correlation score and the rotation angle of maximum response are shown for the top four models. Note that all are at $\theta = 0.0$, i.e. aligned with the Y axis. A graph depicting the correlation score of the model "Peak (130)" with the histogram produced from the polygon 11 range data is shown in figure 6. Figure 7 shows the model selected for each of the regions in the Fort Hood image.

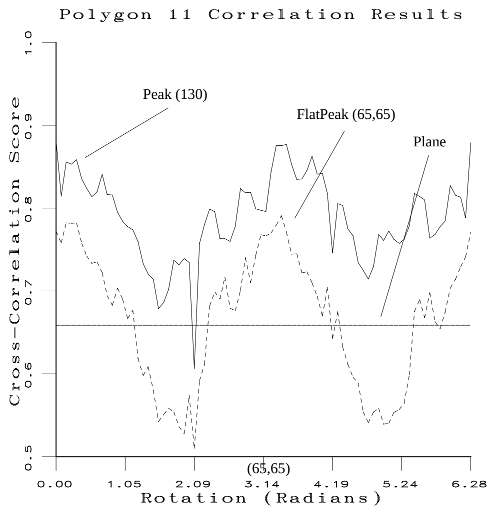


Figure 6: The correlation response of three models (Peak, FlatPeak, and Plane, for comparision) with polygon 11. The highest scoring model other than a peaked roof primitive was the "FlatPeak (65,65)" model. The correlation score, however, clearly separates the two.

4 Reconstruction

Each region within the data has been indexed into the SPD to provide a set of inital models and parameters; these are then fit to the elevation data. The role of the reconstruction module is to use the initial parameters from the SPD match to determine a precise model fit to the range data.

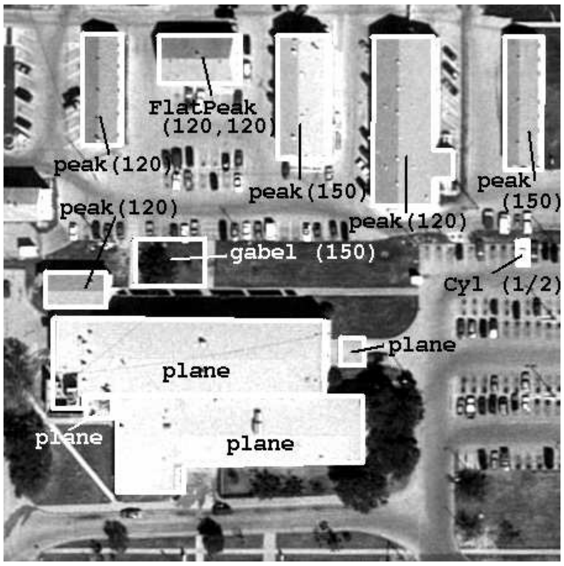


Figure 7: Classification of each region. The model with a maximum response from the library is shown.

For each polygon boundary, the set of elevation points that project within the polygon are extracted from the corresponding range image. All these points are considered to be part of the inital inliers. The model selected from the SPD is fit to the data using a least-median fit technique and the downhill Simplex method.

The downhill Simplex method requires a set of inital parameters to initalize the search; these are provided directly from the indexed SPD model. In order to avoid unusually high residual errors in the case of models with steep surfaces, residual errors are computed as the distance along the approximate surface normal from the elevation data (derived from the Delauney triangulation) to the current model estimate.

Figure 8 shows the final reconstruction after the SPD model "Peak (130)" was fit to the data. The final peak angle (measured from plane to plane) converged to 134 degrees with a median residual error of 0.065 meters for a roof whose height was by 5.9 meters above the groundplane.

Each of the regions was fit to the appropriate SPD model from the database. All remaining points in the range image were assumed to lie in a ground plane and a plane was fit to determine the correct model. Figure 9 shows a rendered view of the entire site model.

Using the registered optical image, a texture map can be wrapped onto each of the rooftop surfaces for better visualization of the site. This was applied to the two flat roof models in Figure 9.

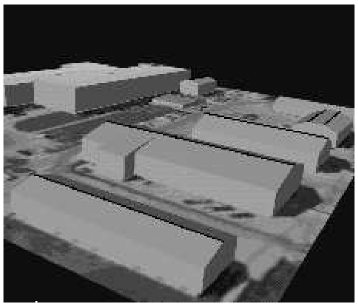


Figure 8: Closeup view of reconstructed polygon 11 (building in foreground).

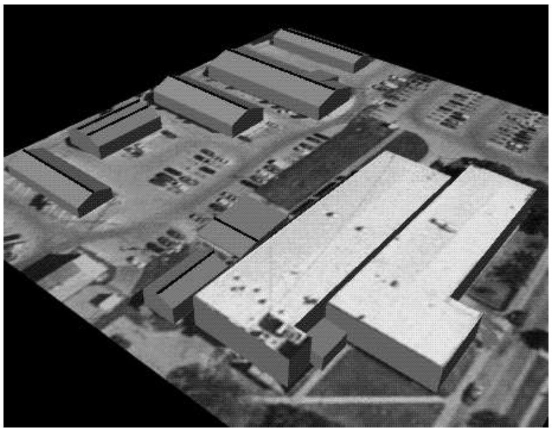


Figure 9: Rendered view of the Fort Hood Scene.

5 Test Results

The technique has also been applied to the Ascona/ISPRS "Flat Scene"; this scene contains several peaked roof buildings of different slopes (see Figure 10). In the face of space restrictions, only key steps in the reconstruction process are shown.

The buildings detected are shown in figure 11. Overlapping polygons were eliminated leaving sixty four percent of the building rooftops to be passed to the indexing module for classification. Note that polygon 21 was detected but was eliminated from the reconstruction process because corresponding range data was not available. The remaining twenty regions were classified using the SPD shown in section 2. The results of the classification are shown in table 2.

The top two models selected for each region were fit to the data and the best fit was chosen for the final reconstruction. All points outside the twenty polygons are used to fit a local ground plane. Planar regions that were classified as planes and were close to the local groundplane (for example polygon 3) were eliminated as false positives and adjusted to lay within the groundplane. Two errors not eliminated by this process are represented by polygons 5 and 7, both being

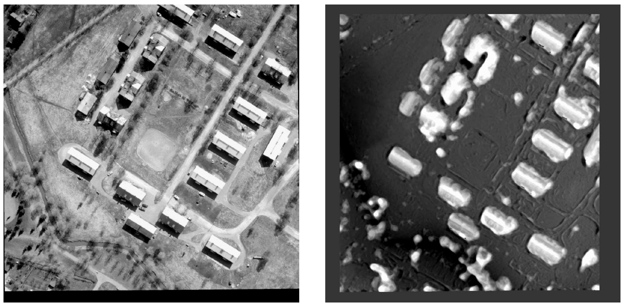


Figure 10: Optical and range image for the Ascona/ISPRS scene.

sloping portions of the rooftops, and reconstructed as extruded planes. Figure 12 shows the final site model acquired as a result of the reconstruction process.

References

[Aurenhammer'91] F. Aurenhammer, "Voronoi diagrams - a survey of a fundamental geometric data structure". *ACM Computing Surveys* 1991. vol 23. pp. 345-405.

[Collins'95] 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.

[Collins, Jaynes, et. al '96] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, H. Schultz, A. Hanson, E. Riseman. "The UMass Ascender System for 3D Site Model Construction", chapter to appear in forthcoming DARPA IU RADIUS book. Oscar Firschien (ed), 1996.

[Haala'95] N. Haala and M. Hahn, "Data fusion for the detection and reconstruction of buildings," *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, April 1995.

[Herman'94] M. Herman and T. Kanade. "3D Mosaic Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images". *Proc. ARPA Image Understanding Workshop*, 1994.

[Horn'86] B. Horn. *Robot Vision*, MIT Press, McGraw-Hill, Cambridge, Massachusetts, 1986.

[Huertas'80] A. Huertas and R. Nevatia. "Detecting Buildings in Aerial Images" *Computer Vision, Graphics, Image Processing.* vol. 13, 1980.

[Jaynes'94] C.Jaynes, F.Stolle and R.Collins, "Task Driven Perceptual Organization for Extraction of

Figure 11: Detected polygons in the Ascona scene. Polygons labeled with a subscript were eliminated due to overlap.

Rooftop Polygons," *IEEE Workshop on Applications of Computer Vision,* Sarasota, FL, December 1994, pp. 152–159.

[Kim'95] "Building Extraction and Verification from Spaceborne and Aerial Imagery using Image Understanding Fusion Techniques," *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, April 1995.

[Matsuyama'85] T. Matsuyama and V. Hwang, "SIGMA: A Framework for Image Understanding: Integration of Bottom-Up and Top-Do wn Processes," *Proceedings of the Ninth IJCAI*, Los Angeles, CA, pp. 908-915, 1985.

[McKeown'90] D. McKeown, "Toward Automatic Cartographic Feature Extraction," *Mapping and Spatial Modelling for Navigation,* Nato ASI Series, Vol. F65, p p. 149–180, 1990.

[Schultz'94] H. Schultz, "Terrain Reconstruction from Oblique Views," *Arpa Image Understanding Workshop,* Monterey, CA, Nov 1994, pp. 1001–1008.

Polygon #	Model Name	Score	θ
1	Plane	0.9322	
2	Peak (120)	0.9104	$\theta = 121.0$
3	Plane	0.9051	
4	Plane	0.9462	
5	Plane	0.8933	
7	Plane	0.7543	
8	Peak (110)	0.8963	$\theta = 29.0$
9	Peak (140)	0.9502	$\theta = 118.0$
10	Plane	0.9353	
11	Peak (140)	0.9017	$\theta = 119.0$
12	Peak (120)	0.8991	$\theta = 120.0$
13	Peak (150)	0.9136	$\theta = 118.0$
14	Plane	0.8773	
15	Peak (120)	0.9114	$\theta = 119.0$
16	Peak (130)	0.9276	$\theta = 116.0$
17	Plane	0.8994	
18	Peak (130)	0.9212	$\theta = 117.0$
19	Plane	0.9102	
20	Peak (120)	0.9532	$\theta = 118.0$

Table 2: Model indexing results for each region in the "Flat" scene.

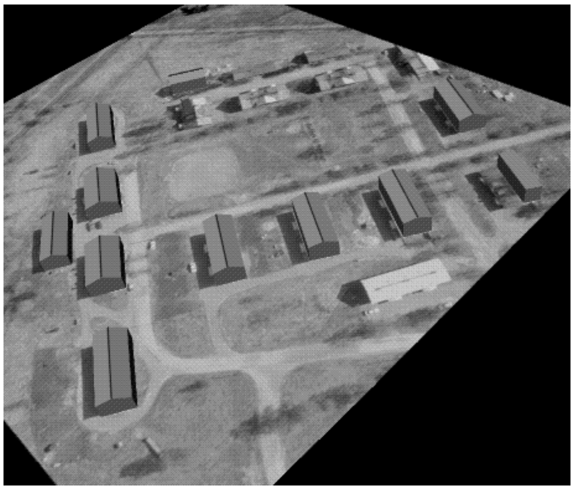


Figure 12: The reconstructed site elevation image.