

Recognition and Reconstruction of Buildings from Multiple Aerial Images

Christopher Jaynes[†], Edward Riseman[‡], and Allen Hanson[‡]

[†]Dept. of Computer Science University of Kentucky, Lexington, KY 40506

[‡]Dept. of Computer Science University of Massachusetts, Amherst

October 18, 2001

Abstract

We present a model-based approach to the automatic detection and reconstruction of buildings from aerial imagery. Buildings are first segmented from the scene in an optical image followed by a reconstruction process that makes use of a corresponding digital elevation map (DEM). Initially, each segmented DEM region likely to contain a building rooftop is indexed into a database of parameterized surface models that represent different building shape classes such as peaked, flat, or curved roofs. Given a set of indexed models, each is fit to the elevation data using a robust iterative procedure that determines the precise position and shape of the building rooftop. The indexed model that converges to the data with the lowest residual fit error is then added to the scene by extruding the fit rooftop surfaces to a local ground plane.

The approach is based on the observation that a significant amount of rooftop variation can be modeled as the union of a small set of parameterized models and their combinations. By first recognizing the rooftop as one of several potential rooftop shapes and fitting only these surfaces, the technique remains robust while still capable of reconstructing a wide variety of building types. In contrast to earlier approaches that presuppose a particular class of rooftops to be reconstructed (e.g. flat roofs), the algorithm is capable of reconstructing a variety of building types including peaked, flat, multi-level flat, and curved surfaces. The approach is evaluated on two datasets. Recognition rates for the different building rooftop classes and reconstruction accuracy are reported.

1 Introduction

Automatic reconstruction of buildings from aerial imagery involves several significant problems related to computer vision. Buildings are difficult to locate within urban areas that contain hundreds of other objects in close proximity, such as parking lots, fields, road networks, and vehicles. Trees, power lines, and other buildings often occlude building rooftops, particularly in oblique views. Furthermore, building rooftops exhibit extensive shape variability, are composed of different surface materials with differing reflectance properties, and contain significant surface clutter. These issues make automatic recognition and reconstruction of buildings a challenging problem, as well as a research topic that is relevant to several application domains.

We introduce a model-based technique for the automatic recognition and three-dimensional reconstruction of buildings directly from a single digital elevation map (DEM), derived from optical stereo processing or active sensors, registered to a corresponding optical view of an urban site. Figure 1 depicts a typical dataset used by the reconstruction technique. At least two optical images and accurate camera calibration information are used to compute a DEM of an urban region. Initially, focus-of-attention regions that are likely to contain buildings are segmented from the scene. A perceptual grouping algorithm detects building boundaries as closed polygons in the optical image. Corresponding regions in the DEM can then be processed to determine rooftop shapes. We utilize the key idea of matching a database of shape models against each DEM region using a model-indexing procedure. Surface orientation histograms for each parameterized model in the database are compared to a histogram derived from surface normals fit to small planar patches within the DEM region.

The set of models (surfaces) that most closely match the DEM region are used as the initial estimates for a robust surface fitting technique that refines the model parameters (such as orientation and peak-roof angle) of each hypothesized roof surface. The surface model that converges to the DEM with the lowest residual fit error is retained as the most likely description of the surface.

The database of surface models contains a limited number of canonical shapes common to rooftops, such as planes, peaks, curves, domes, and gables. Reconstruction of complex shapes is achieved through a composition of different parameterizations of the canonical shape models. Prior knowledge about the scene to be reconstructed, such as the presence or absence of particular rooftop classes, can be used by adding or removing rooftop models from the database.

The approach is evaluated on several datasets and we demonstrate that this two-phase reconstruction approach leads to the accurate reconstruction of a wide variety of building types while still retaining robustness.

Figure 1 Approximately Here

2 Related Work

There has been significant progress in several areas relevant to the site modeling task in the United States (largely due to the DARPA RADIUS program (1992–1996) and the Automatic Population of Geospatial Databases (APGD) follow-on program (1996–1999) [19; 6; 4; 53; 25; 40]). In conjunction with this work, there are significant efforts underway in Europe that combine traditional photogrammetric approaches to aerial image reconstruction with object recognition and segmentation to acquire three-dimensional building models [13; 16]. However, general building reconstruction that is not restricted to a small class of models and contexts remains out of reach.

Over the past decade, automated building detection systems have evolved along many lines, but the trend has always been towards greater generality. Research has evolved from restricted nadir views [43; 31; 53] to general oblique viewpoints [8; 37], from single image analysis [31; 53] to multi-image techniques [18; 41; 50], and from purely 2D-hypothesis extraction in image-space [38; 52] to rigorous 3D geometric reconstruction in object-space [19; 15; 51; 8; 16; 33; 20]. Systems have typically

employed increasingly sophisticated scene-based and image-based geometric constraints for the recovery of urban scenes. Although the use of strict geometric constraints can increase system robustness, each system is only capable of reconstructing object classes that satisfy the particular implied geometries. Modification of the constraints in order to increase system generality requires sophisticated information about the system (at a software coding level) and may involve many months of reengineering the system.

Here, we introduce a two-phase approach to the recognition and reconstruction of multiple building types that is capable of exploiting geometric constraints while remaining somewhat flexible. Surface classes can be added or removed from a database of models that then constrain the type of buildings addressed by the system.

Many early building reconstruction systems were based on the nadir viewpoint assumption, in part because most of the available images at that time were borrowed from map-making applications that relied on nadir views. The nadir assumption greatly simplifies building extraction geometry since rectangular building roofs appear as rectangles in the image, and there is little occlusion of one building by another. However, because aerial images typically cover a large-scale area, only the small portion of the image near the image center is truly nadir. Viewing rays that intersect the image plane near its edge will intersect the world at oblique angles. The easiest generalization from nadir views to handle obliquity is to assume weak-perspective or affine views, where rectangular roofs appear as parallelograms [41]. The ultimate generalization is to introduce photogrammetrically rigorous camera projection equations that more accurately model the (typically projective) viewing geometry [6; 36; 37] such as was done in the Ascender II system.

Early systems were dominated by monocular, image-based approaches, since often only a single view of the area was available. However, buildings are raised structures with complex 3D rooftop shapes. It is difficult to disambiguate rooftop hypotheses and determine building height without 3D information.

One source of information in monocular images is shadows, and indeed, systems have been designed that exploit the relationship between shadows and roof hypotheses [42; 33; 34; 21]. Shadow analysis is particularly attractive when combined with nadir viewpoints, as then building height is directly proportional to the length of the building shadow on the ground. Systems that rely on shadow analysis often assume that the sun position (illuminant direction) is known, and always assume that building shadows fall on flat terrain surrounding the building (and not across other buildings or on rocky or hilly terrain). This is obviously problematic in many scenarios, such as dense urban building scenes.

A more general and reliable method of deriving 3D height information is to use stereo triangulation across two or more images [6; 26; 5; 47; 7]. Three-dimensional data can remove the ambiguities present in a single aerial image and allows the interpretation of shadows in the context of local terrain surface shape. Noronha and Nevatia [44] describe a system where hierarchical grouping and matching across multiple images is used to reconstruct 3D building models. Buildings are extracted in hierarchical stages, ranging from line segments, to junctions, parallel pairs, U-shapes, and finally, to complete parallelograms. At each stage in the extraction hierarchy, the following three steps are performed: 1) 2D grouping of features at that level in each image, 2) epipolar matching of the features across pairs of views, and 3) application of geometric constraints to check the consistency of the 3D structures implied by those feature matches. Final building hypotheses are verified by searching for consistent shadows and evidence of vertical walls. Only rectangular building hypotheses are found – arbitrary rectilinear

structures are formed by merging any abutting or overlapping rectangular 3D building hypotheses of similar height. A notable feature of the system is that information from all views is used in a non-preferential way, where each image may be used to extract hypotheses, and the remaining set of views are used to corroborate each hypothesis and compute 3D structure [6]. This approach is also present in the Ascender system [6].

In the MULTIVIEW system, developed at Carnegie-Melon's digital mapping laboratory [39; 47], corner features extracted via vanishing point analysis are matched across a pair of views to get 3D corners. These 3D corners become nodes in a graph, and pairs are linked if image intensity gradient information supports an edge hypothesis. Polygonal surface hypotheses are formed by looking for cycles in the graph that meet certain planarity and perpendicularity constraints. When more views are available, relationships between corners and lines in the graph are updated as each new view is added. A similar approach has been taken by the Institute of Photogrammetry at the University of Bonn, where trihedral corners are matched across views; matched corners imply aggregated features, such as pitch roof angles, and building facades [15]. Using building terminals (building boundary walls and associated trihedral corners) as a primary feature, work by Henricsson [22] attempts to reconstruct buildings using a search for increasingly complex aggregated models. For example, building terminals are grouped together from simpler image features that are then grouped into complete building models based on geometric constraints that require they be both complementary and can fit together with minimal error.

More recent approaches to automated building detection integrate information from ground level photographs to refine models acquired from aerial image processing [29]. Other approaches make exclusive use of ground-level imagery to detect building facades, recover their three-dimensional position, and reconstruct urban scenes [11]. These techniques exploit domain constraints such as the vanishing points induced by parallel facade structure [29; 9] or make use of a large number of calibrated ground images [29; 11]. Using a multitude of calibrated ground-level views, for example, [10], has shown that a consistent and accurate three-dimensional model can be acquired that covers a significant geographic urban area.

Table 6 summarizes the work in automated building detection over the past 10 years. The type of input data used by the system is shown. Typically this is either **Optical Stereo** which refers to a controlled image pair, or **Multi-optical** which refers to a system that utilizes many views of the area under consideration and the corresponding camera parameters. Footprint shape defines the type of building boundary detectable by the system as either a simple rectangle, a rectilinear combination of rectangles called a **compound**, or an arbitrary closed polygon consisting of junctions at any angle, referred to as **complex** in the table. Finally, the type of reconstruction that the system is capable of producing is categorized by either 2D or 3D and the class of rooftop shapes addressed by the system is categorized into a small set of intuitive shapes.

Table 1 Approximately Here

Table 6 shows a research tradition that has moved from 2D to 3D processing, and towards the fusion of more data sources for a more complex reconstruction that includes a wider number of building rooftop types. The work presented here makes significant advances in the types of building models that can

automatically be recovered. Although several other approaches do not preclude the reconstruction of complex rooftop shapes [19; 15; 14], these techniques require that the model class be known a priori [6; 15] (peaked roof, for example) or that the basic building components are known [19]. Our approach separates the recognition and reconstruction phases in order to address a wider variety of building classes while retaining system robustness.

3 Building Detection

Prior to the recognition and reconstruction phases, building regions are segmented in an optical image by detecting and grouping lines corresponding to building rooftop boundaries. The building recognition phase requires regions of interest that are likely to contain building rooftops. In this section we introduce one such method that can be used to generate the required region of interest. Other techniques, such as those used in Ascender [8] and Ascender II [26] can be used as well.

The approach, presented here shows, how the two image sources can provide useful information for the detection and reconstruction process. Typically, elevation data is acquired through stereo optical processing and optical data, and its correspondence to the recovered DEM is available.¹

The goal of automated building detection is to roughly delineate building rooftop boundaries that will later be verified during the recognition and reconstruction phases of the system. The algorithm is based on perceptual grouping of line segments into image polygons for rectilinear polygons in the scene [31]. The algorithm, then, attempts to denote the outermost boundary of rooftop polygons that is coplanar and meets at right angles in the world. Figure 2 depicts the general geometry of the situation.

Figure 2 Approximately Here

The rooftop segmentation algorithm proceeds in three steps: low-level feature extraction, collated feature detection, and hypothesis arbitration. Each module generates features that are used during the next phase in order to progress from simple image features to grouped, closed polygons that represent rooftop boundaries.

Low-level features are straight line segments and image corners. Line segments can be produced by any reasonable straight line extraction from edge operator. We have used the Boldt algorithm [1], or grouping of canny edges [3], there are many candidates that will work as well. Extracted line segments are then grouped according to the domain assumption that the outermost building boundary is planar and rectilinear. Lines with spatially proximate endpoints are grouped into candidate corner hypotheses. Using viewpoint pose information, each potential image corner is backprojected to a nominal Z plane in the scene, and the resulting scene corner is tested for orthogonality. A parameter that describes how aggressively features should be grouped is used to threshold corners based on the angular distance from an orthogonal corner. This parameter is set by the user and can be adjusted according to expected image variance.

Corners and line segments are further grouped into chains of features that represent a polygonal

¹When no optical data is present (e.g. DEMs generated from radar processing, for example) we have suggested techniques [27] that are appropriate.

boundary. Chains are a generalization of collated features [25] and allow final segmented boundaries of arbitrary rectilinear shape to be constructed from low-level features. These chains are stored in a *feature relation graph* which is an encoding of the feature dependencies and perceptual compatibility in the image. Low-level features (corners and line segments) are nodes in the graph, and perceptual groupings are represented as edges between corresponding nodes. Nodes have a certainty measure related to the confidence of the low-level feature extraction routines, and edges are weighted with the certainty that the grouping represents. A chain of collated features inherits an accumulated certainty measure from all the nodes and edges along its length.

Final building rooftops are discovered by parsing the feature relation graph for closed chains that are both independent (have no edges in common) and are of maximum certainty. Standard graph theoretic techniques are employed to discover this maximally weighted independent set of cycles. While searching for closed cycles the collated feature detector may be invoked in order to attempt closure of chains that are missing a single feature. The system then searches top-down for evidence of a missing line in the image based on the geometry of the collated chain. For example, shadows or scene occlusions may cause the corner detector to fail in the image; however, given the presence of a strong, nearly closed boundary, of lines and corners the line detector may be reinvoked with a different set of parameters in the image region implied by the existing features and the rectilinear assumption.

Figure 3 Approximately Here

Figure 3a depicts the lines and corners extracted from an nadir image of Fort Hood, Texas. Detected rooftop regions are shown in Figure 3b. Notice that the corner labeled 'C' was not detected in the initial low-level feature detection phase and was only discovered in the top-down search process while attempting to close the polygonal boundary of polygon 6.

4 Model-Based Recognition

Given a focus-of-attention region in the optical image, and a corresponding set of DEM points, the system next determines the underlying shape of the rooftop. In general, the goal of the model indexing phase is to select a set of appropriate surface models that will then be fit to the DEM (this fitting process is discussed in section 5.1. The surface model that most closely fits the DEM is then considered to be the correct model and is inserted into the scene.

The digital elevation map contains a regular grid of elevation estimates derived from processing multiple optical views, or is captured using active sensors capable of deriving height estimates (e.g Interferometric Synthetic Aperture Radar (IFSAR) or laser rangefinders). Several factors confound the production of an accurate elevation map using either technique. Significant depth discontinuities generate occlusions in the scene that lead to incorrect correlations in the stereo-optical processing systems, or may lead to dropouts (points on the DEM that do not contain an elevation estimate) in remote sensing systems using active sensors such as radar. This is particularly a problem in dense urban areas in which buildings often occlude surrounding terrain and neighboring buildings. Illumination boundaries from shadows and the absence of significant surface markings often produce miscorrelations

in stereo-optical algorithms that lead to incorrect height estimates. Likewise, surface geometry and material properties can produce a changing radar return and incorrect height estimates using active sensors.

Several researchers have studied the accuracy of DEMs as a function of the viewing angle of the sensors and relative height of objects versus the stereo baseline. The interested reader is referred to [48; 49; 45] as an example of these types of studies. Three characteristics of the DEM accuracy are of importance to the work here. First, errors in the DEM increase with perspective distortion due to off-nadir viewing angles. Secondly, both stereo-optical and IFSAR algorithms require a significant amount of numerical processing and small errors that accumulate from the initial image acquisition phase can lead to inconsistencies in the DEM. Finally, DEMs consist of independent elevation estimates that are perturbed by both correlated and uncorrelated noise. In order to interpret these height measurements as coherent surfaces, robust recognition and fitting techniques are required. IFSAR data has similar noise problems due to changing material types and noise related environmental conditions during data capture. We have analyzed the noise characteristics of IFSAR data in order to determine a rough accuracy estimate for IFSAR elevations in a built-up area [93].

In our approach, the role of the model-based indexing algorithm is to classify a segmented array of elevation estimates as an instance of a distinct rooftop surface class (peak, flat, hemi-cylindrical, domed, etc.). Under the assumption that elevation estimates represent a rooftop region, the indexing technique must determine the surface model that most closely resembles the data. In general, each of the models is matched to the elevation data using a correlation scheme that compares the orientation of component surface elements in the model and the DEM. A match-score for each model and parameterization is computed and the models are ranked accordingly. A set of best ranking models (the top three models for our experiments) are passed to a robust fitting algorithm that attempts to fit each model to the data, using the parameters determined during indexing to initialize the algorithm.

The recognition technique must be invariant with respect to translation and scale, and robust with respect to noise characteristics found in DEMs [30]. Model-directed processing of this sort allows the explicit representation of knowledge about rooftop shapes typically found in the scenes under consideration. This immensely reduces the degrees of freedom typically found in purely bottom-up approaches to scene reconstruction. Although the final building models are constrained to be the union of the available surface primitives, this implies a reduction in the search for possible solutions and leads to robustness under significantly noisy conditions.

4.1 Model Indexing

The model-indexing algorithm estimates the surface orientations of small surface patches on the DEM and constructs an orientation histogram of these surface normals. The orientation histogram is then correlated with the a set of histograms associated with predefined building-rooftop surfaces in a Surface Primitive Database (SPD). The surface primitive database is defined by a set of parametric surface classes and several discrete parameterizations for each surface class. For example, the SPD may contain the peak rooftop surface class that is parameterized by the pitch-angle of the surface. Stored within SPD are several specific parameterizations of the general peak-roof class, each with a different pitch angle, that will be used to recognize peak-roof buildings of various shapes. The resolution of the class (pitch-

angle at every 10 degrees, for example), is dependent upon the requirements of the indexing algorithm and the expected noise characteristics of the DEM. Associated with each surface is an orientation histogram that is used to match surface models to regions within the DEM.

Figure 4 Approximately Here

The surface primitive database used for experiments in this paper contains a set of 8 surface classes called surface primitives, such as planes, cylindrical surfaces, peaks, cones, and domes known typically to be part of rooftop surfaces. Associated with each surface primitive are a number of models, representing different parameterizations of each class of surface primitives, forming 119 total models.

For each surface primitive, an orientation histogram is precomputed and stored with the model in the SPD. These orientation histograms, sometimes called the Extended Gaussian Image [Ehler, 1986], are normalized so that they are both scale and translation invariant. In order to compare the model histograms (and corresponding surface models) with regions in the DEM, histograms are derived from the regions in the elevation data as well. We refer to histograms derived from the DEM as image histograms to distinguish them from histograms derived from the SPD.

Both image and model histograms are constructed in largely the same manner. An orientation histogram is computed from the set of three-dimensional points that are either computed by evaluating the parametric SPD model, or are given by the elevation estimates within a region P of the DEM. The set of points \mathbf{X}^P is triangulated into a surface mesh using the well-known Delaunay algorithm [32; 46]. The simple surface mesh is a set of triangular surface patches, $T_i = [x_i^P \ x_j^P \ x_k^P]$ where x_n^P represents a single 3D point vector from the DEM elevation estimates \mathbf{X}^P from polygon region P . The Delaunay tessellation simultaneously maximizes the smallest angle per triangular patch and minimizes the maximum enclosing circle for each triangle. Because the triangular patches will be used to compute surface orientation, these properties are important in that they ensure surface patches are both local estimates of surface orientation and will completely cover the data. Figure ??b shows the triangulated surface fit to the DEM that corresponds to the polygon detected in the optical image in Figure 1a.

Figure 5 Approximately Here

The local orientation of the elevation data is estimated by computing the outward surface normal for each local triangular patch, T_i . X_i^k is defined as the point central to triangle T_i , within region k , such that the sum of the distances from X_i^k to \mathbf{p}_1 , \mathbf{p}_2 , and \mathbf{p}_3 is minimized. The surface normal is computed as the cross product of the direction vectors defined by two points in the surface patch and centerpoint X_i^k :

$$N_i^k = v_1 \times v_2 = \left[\frac{(\mathbf{p}_2 \times X_i^k)}{\|\mathbf{p}_2 \times X_i^k\|} \times \frac{(\mathbf{p}_3 \times X_i^k)}{\|\mathbf{p}_3 \times X_i^k\|} \right] \quad (1)$$

Because it is assumed that 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 histogram that will receive a vote for a particular orientation. All surface normals that are oriented within a discrete direction, referred to as an orientation histogram accumulator, are tallied

and stored in the accumulator cell using a smoothed voting scheme (so that some surrounding cells will receive a weighted vote). The histogram then, is a two dimensional representation of the local orientation over the entire surface. To avoid sensitivity problems with the method by which orientation space is discretized, votes are smoothed over the sphere via a Gaussian function (although other spherical smoothing functions can be used [26; 9]).

For the results shown here, the orientation histogram contains 240 bins, reflecting a tessellation based on the semiregular icosahedron [24]. If the surface normal, N , intersects the Gaussian sphere at (x,y,z) the weighted vote is given by:

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

where D is the angular distance from (x,y,z) to the center of the histogram bin, α , to receive the weighted vote, and σ describes the amount of smoothing in the voting process. Sigma is related linearly through a system parameter to the expected noise in the digital elevation map and was 0.5 for the results presented here. A single surface normal, then, may vote for more than one orientation bin on the Gaussian sphere, to allow a spread of each vote. Voting to a bin for a given surface normal is cut off when the value of $V(x,y,z,\alpha)$ falls below a threshold (0.1 for the results shown here).

Figure 6a below shows the histogram constructed from the range data corresponding to polygon 8 from Figure 12b. For comparison, the histogram corresponding to the SPD that most closely matches the image histogram is shown in Figure 6b. Histograms visualized in this way provide an interesting characterization of the noise within a surface patch. Notice, for example, that the underlying peak structure is discernible, but that number of votes for each of the lobes is unequal.

Figure 6 Approximately Here

For purposes of visualization, these histograms can be orthographically projected onto the X-Y and X-Z planes in order to isolate the planar orientation of the surface normal vectors. Figures 7 and 8 show the histograms from Figure ?? projected to the X-Y and X-Z planes respectively.

Figure 7 Approximately Here

Figure 8 Approximately Here

As can be seen in the example DEM histograms derived from the sensor data, the indexing algorithm must be capable of selecting the correct rooftop model in the presence of significant noise levels. Errors introduced in the construction of the DEM lead to errors in surface normal computation and a highly corrupted orientation histogram for the region. As an example, consider a dense DEM with an inter-sample distance (horizontal posting) of 1 meter. An error in height of 1/4 of a meter for one elevation estimate (and hence one vertex of a triangular patch) will induce an angular error on the corresponding surface normal of approximately 13 degrees. Therefore, the model indexing algorithm must use a robust measure of similarity between orientation histograms to account for outliers. In addition, the

set of models selected by the indexing procedure must contain both the model and initial parameters that are capable of converging to the correct surface through an iterative fitting procedure (discussed in Section 5.1).

4.1.1 Histogram Correlation

Each histogram extracted from the DEM is correlated with the set of predefined histograms in the SPD. Since the building being extracted may be situated at any orientation, the image histogram must be compared with the SPD at discrete relative orientations about the vertical axis. The model histogram remains fixed relative to a rotating image histogram. For each relative rotation about this axis the image histogram is multiplied by the appropriate rotation matrix $R(\theta)$ and a correlation measure is computed by comparing the values of the aligned histogram accumulators. This is performed many times, for small relative rotations between the two histograms around an axis of rotation. The correlation score, for each relative orientation from 0 to 2π radians, is stored for each model in the SPD as an array of values. The overall maximum correlation score represents model similarity over the set of relative rotations. A sorted list, based on this model similarity, can then be presented to a model fitting procedure that makes use of the matched model and the corresponding relative rotation information as initial parameters in an iterative refinement algorithm.

The correlation score at a particular angle θ is given by:

$$C_{\theta}(I, M) = \frac{\sum_{\theta}^{2\pi} \sum_{i,j} [I(i, j) - \mu_I] \cdot [(R(\theta)(M(i, j) - \mu_M)]}{\sigma_I \cdot \sigma_M} \quad (3)$$

where μ and σ represent the mean and the variance of the image and model histograms respectively, and $R(\theta)$ is the 3-by-3 rotation matrix that rotates the model by θ radians about the Z-axis.

For the results shown here, θ was restricted to 0.034 radian increments (one-degree) about the Z-axis assuming that the Z-axis of the rooftop model is parallel to the gravity vector (i.e. building rooftops are not tilted with respect to a local ground). Rotation is measured with respect to a fixed coordinate system aligned with the regular DEM grid so that the X and Y axes align with the rows and columns of the elevation samples.

Figure ?? graphs the correlation score for the top three matching models from the SPD, correlated with the histogram derived from polygon 8 (see Figures ?? and ??). The image histogram was compared to all models in the SPD. The maximum correlation scores returned by the matching algorithm were 0.987, 0.844, and 0.763 for the **Peak-55**, **Peak-45**, and **FlatPeak-45** models respectively.

The indexing procedure is similarly applied to all DEM regions detected as potential rooftops in the scene. For each region, the top three models, and the relative rotation around the normal to the ground plane at which each model correlated maximally is computed. This information is then used by a robust surface-fitting technique that determines a more precise fit of each model to the data. In this way, we are able to determine the approximate rooftop type, using a limited number of models followed by a procedure that is capable of finer resolution model fitting.

Initial Model	Initial Fit Error	Final Model	Residual
Peak-55	0.864m	Peak 54.5 ⁰	0.120m
Peak-45	1.185m	Peak 54.5 ⁰	0.120m
FlatPeak-45	3.194m	FlatPeak 54 ⁰ , 57 ⁰	1.842m

Table 1: Initial model, initial fit error, fit model and final residual fit error for the best matching models after surface fitting to polygon 8 (Figure 5). Angles for reconstructed peak models are reported as full pitch angle. In conjunction with a decision method (such as a simple threshold on the residual error), the fit error indicates if a region was recognized as the correct surface class.

5 Parameterized Surface Reconstruction

Regions recognized by the model indexing procedure are fit to the corresponding DEM data using a robust surface fitting procedure. The algorithm requires three pieces of information; M , x^P , and a . The model M , is a surface primitive from the SPD with a vector of free parameters a ; and is written as $M_i(x^P; a)$, where x^P is a vector of points from the elevation data contained in region P . The parameter vector a and the model M are used as initial estimates for the robust surface fitting procedure. In many cases, several candidate models (three in our case) may be fit to the same set of data points, x^P . Typically, the system is able to discriminate correct models from incorrect ones by fitting several alternative models and discarding the models with the higher residual fit error.

Continuing our running example for polygon 8, Table 1 shows the residual fit error, using the fitting algorithm described in this section. For each of the top three models and parameters returned by the model-indexing scheme, a corresponding surface was fit.

Clearly for the example shown, the result of the surface fitting algorithm can be used to discriminate competing object classes for a region of the DEM, as well as recover the three-dimensional model and parameters that most closely resemble the underlying data. Different starting parameters may converge to the same model with the least error as shown by **Peak-55** and **Peak-45** in 1.

5.1 Surface Fitting Details

Model fitting involves a multidimensional optimization scheme for $M_i(x; a)$, where a is the parameter vector associated with the model being fit. The median squared fit error is minimized to determine the best fit of the parameterized model to the data. We use the downhill simplex method [46] to iteratively converge on a least median squared solution for the free parameters defined in a . Because a triangular mesh has already been fit to the range data, the surface normal at each patch N_P is used to compute the distance between the current model and the observed data. Specifically, the median of

$$E_P = \llbracket |x - \hat{x}^P|^2 \rrbracket \quad (4)$$

is minimized, where

$$\hat{x}^P = x^P + t \cdot \overline{N}x^P | x^P \in M(x; a) \quad (5)$$

that is, \hat{x}^P is the point on the fit surface corresponding to data point x^P and is obtained along the computed surface normal of the surface patch center at x^P . Figure 9 depicts how fit error is measured.

Figure 9 Approximately Here

This median squared error function avoids measuring error in an arbitrary way, and uses information from the surface mesh to estimate an appropriate direction from the observed data to the model surface. This is particularly important in the case of models with surface derivative discontinuities (the peak model, for example where surfaces are very steep in Z and adjoined at sharp angles), and error measured near the peak and along the Z-axis may induce an unusually large errors as depicted in Figure 9. In this case, the steep slope of the model could magnify the error when it is not measured in the direction of the surface.

Absolute position in the scene is fixed as the center of the region of data being fit. A vector of free parameters a is initialized by the results of the indexing phase and is then iteratively refined using our minimization algorithm. Given the initial parameter vector, the downhill-simplex technique requires initial estimates of the other $k+1$ points that define the simplex. These are computed from a as:

$$A_i = a + \delta e_i \tag{6}$$

where e_i are k unit vectors and δ is a scalar constant that defines the amount of perturbation in the initial estimates. δ is related to the scale of the model being fit and is derived, for each region, from the region size and the scaling factor that relates the data being fit to the site world coordinate system. At each iteration, the simplex moves a particular vertex A_i , typically the vertex with the highest residual error, E_P , through the opposite edge/face of the simplex to a point with smaller error. Each of these steps, called reflections, maintains the simplex volume and thus, the set of estimates are able to move towards a global solution in a non-degenerate manner. When a stopping condition is reached, the algorithm terminates, returning the set of parameters A that minimize E_P for $M(x; A)$, the median residual error at A , and the set of points that are considered to be outliers with respect to the fit surface.

Outliers are computed as points in the range data that have a relatively high residual error. We have developed techniques that use the outlier measure as the basis for recursive segmentation of rooftop substructure [28]. Therefore it is important that this measure is not computed from a simple error-prone threshold on $E_P(x_i)$. Instead, outliers are computed on-the fly through multiple fits using the simplex method. The surface fitting procedure is run to convergence and DEM points are sorted in decreasing order based on their individual fit error measure. For a set of n points, $\lambda * n$ points with the largest fit error are then discarded as outliers. For the results presented here, $\lambda = 0.03$. After convergence of the fitting algorithm, then, $k = n - \lambda * n$ inlier points remain that may be refit to the model. A large value of λ affects the speed of convergence of the algorithm by removing large sets of inlier points. However, λ determines the resolution at which points can be separated from the dataset as outliers and selection of too large a λ increases the potential for removing inliers. In this case, the fit error will not be reduced and the set of recently removed points will be returned to the dataset and the fitting procedure terminated. This may lead to larger fit error and a larger number of outlier points present in the final dataset.

A χ^2 per degrees-of-freedom measure is used to determine when discarding outlier points no longer improves the surface fit:

$$T_\chi = \frac{\chi^2}{k-1} \quad (7)$$

where χ is the variance of E_P and k is the current number of inlier points. When the value of T_χ does not decrease as more outlier points are removed from the data, the process stops and the data points removed during the previous fit are returned to the inlier pointset. Using this technique, the number of outliers removed at each step can be small and is not dependent on local characteristics of the data, as a simple threshold based on E_P would be.

Figure 10 demonstrates the fitting process for Region 8 (Figure 5) using the Peak-150 model. The number of steps required for each iteration of the fitting algorithm as outliers are removed is shown versus the residual fit error. Initially, the fit surface converges to a fit error of 0.093 meters. After discarding three percent of the points with the largest fit error, the algorithm converges to a residual fit error of 0.026 meters. The third iteration of the fit algorithm led to a fit error of 0.021 meters. Discarding more points based on fit error does not lead to an improved overall fit and the algorithm terminates.

Figure 10 Approximately Here

Each of the top three indexed models are fit to a region's DEM and the best fitting model (in terms of residual fit error), is then inserted into the scene model. The fit rooftop is extruded to the local groundplane to produce a volumetric building model. As an example, the top three models for polygon 8 (see Figures 5 and Table 1) were fit. Figure 11 shows the resulting reconstructed peak roof building.

Figure 11 Approximately Here

6 Experimental Results

We present results of the algorithm on two different datasets. In each, a high-resolution DEM was produced using the Terrest system [49]. This system uses a multiresolution technique to match intensity features at subpixel resolution in overlapping aerial images. Rooftop regions in a corresponding optical image were then generated using the perceptual grouping rooftop detection algorithm discussed in Section 3. Polygons that overlapped by more than 20% were filtered by removing the polygon with the smaller area. After polygon filtering, the corresponding DEM regions were matched against the SPD and the top three ranking models were used to initialize the surface fitting procedure for each potential building region. The best fitting model was then inserted into the reconstructed scene and extruded to the local ground.

6.1 Experimental Approach

Quantitative accuracy of the three-dimensional reconstructions was analyzed using a number of different metrics. Extracted models almost always will not be in complete alignment with a ground truth model.

Therefore, the accuracy of the acquired model must be within some threshold of accuracy in order to be considered a correct detection.

For the purposes our evaluation, the *wireframe distance* between the acquired model and the ground truth model must be less than 0.2 times the cube root of the volume of the ground truth model to be considered a correct detection. This is an empirical threshold that allows 2 pixels total error for a cube with sides 10 pixels long, and varies linearly with the scale of the ground truth model. The wireframe distance measures how well two arbitrary polygons match in terms of size, shape, and location. A similar measure (referred to as the centerline-distance) was used to evaluate the performance of the Ascender I system [7] by comparing polygons.

The three-dimensional wireframe distance between two wireframes is computed by oversampling the boundary of one model into a set of equally spaced points along the wireframe (typically several thousand of them). For each point, the minimum distance from that point to a point on the other wireframe is measured. This process is repeated by oversampling the other model and measuring minimum distances to the first wireframe. The wireframe distance is the average of all these measurements. The metric provides a measure for the average distance between the two wireframe models, reported for each model in meters. The wireframe measure is straightforward to compute and can be applied to wireframes with different numbers of vertices, edges, and implied surfaces.

For wireframes that have the same number of vertices and belonging to the same object class additional metrics can be applied between corresponding pairs of vertices. The distance to the closest vertex from the acquired model to the ground truth model can be measured. These Inter-Vertex Distances are reported in both the planimetric (distance in the X-Y plane) and altimetric (distance in Z) components as well as the full 3D error.

Finally, the volumetric accuracy of the reconstructions can be analyzed by a comparison of the percentage of the model volume inside and outside the 3D-groundtruth model. These comparisons are based on partitioning each of the three-dimensional models into a set of fine-grained voxels and computing true positive, false positive and false negative percentages based on voxel-by-voxel overlap.

6.2 Fort Hood Reconstruction

The “Fort Hood” scene, presented in Figures 1 and 3 earlier in the paper, was reconstructed using the technique described above. The scene contains eight buildings of both peaked and flat rooftops. After polygon detection and filtering, 11 polygons were then used as focus-of-attention regions into a corresponding DEM for reconstruction. For reference, both the DEM and the detected polygons are shown in Figure 12.

Figure 12 Approximately Here

Notice that several polygon regions (Figure 12) do not correspond to buildings but were segmented by the detection algorithm based on line evidence in the image. Polygon 9 corresponds to a grassy field, polygon 10 corresponds to a loading dock surrounded by sidewalks, and polygon 11 corresponds to a truck. For purposes of this experiment, they were indexed and reconstructed in exactly the same manner as the buildings. In general, the fit error measure itself is an indicator that the surfaces being

ID	1st /Error	2nd/Error	3rd /Error	Selected
1	Plane / 0.144m	FlatPeak-65/5.32m	Cylinder / 9.23m	Plane
2	Plane / 0.261m	FlatPeak-65/5.86m	Peak / 11.32m	Plane
3	FlatPeak-65 / 0.452m	Peak-55/0.387m	FlatPeak-55/ 2.42m	Peak-55
4	Peak-65 / 0.133m	Peak-55/0.832m	FlatPeak-55 / 6.41m	Peak-65
5	FlatPeak-65 / 0.295m	Peak-75/0.385m	Cylinder / 0.938m	FlatPeak-65
6	Peak-65 / 0.392m	Peak-55/0.392m	Peak-65 / 0.508m	Peak-55
7	Peak-45 / 0.477m	FlatPeak-75/1.21m	Peak-55 / 0.519m	Peak-45
8	Peak-55 / 0.120m	Peak-45/0.120m	FlatPeak-45 / 1.82m	Peak-55
9	Gable-65 / 1.98m	Plane/2.03m	Gable-55 / 1.98m	Gable-65
10	Plane / 0.393m	Gable-75/0.642m	FlatPeak-75 / 0.497m	Plane
11	Cylinder / 1.320m	Plane/1.98m	FlatPeak-65 / 2.07m	Cylinder

Table 2: Top three models indexed from the SPD for each of the eleven regions in the Fort Hood scene. Model selected is based on the lowest fit error measure in the second pass of robust surface fitting.

fit do not correspond to the any in our model library and these regions may be eliminated. This would allow polygons 9 and 11 to be detected as errors and removed.

Table 2 reports the three indexed models from the SPD and the final fit error that resulted from fitting each of the three models to the DEM region. The model and corresponding free parameters that lead to the lowest fit error was then selected and inserted in the scene. Eight of the eleven regions were reconstructed using the indexed SPD model that correlated most highly with the region’s DEM. However, region 3, a peaked roof building, was indexed with the first choice as a FlatPeak model initially. Using the fit error measure this mistake is detected and corrected using the lower fit error of the 2nd best model. Based on this region 3 was correctly reconstructed as a peaked roof building.

Three regions (numbers 9, 10, and 11) did not correspond to any of the models in the SPD but were nevertheless reconstructed as rooftop models. Models 9 and 11 have unusually high fit errors (1.98m and 1.320m) compared to a mean fit error of 0.319m for the remaining building regions and can be removed on that basis. Although the fit error of region 10 (the loading dock) is comparable to that of the correct surfaces, the reconstructed height of the plane was only 1.42 meters above the ground plane. This information, in conjunction with domain knowledge about the scene, has been shown to be useful in discriminating region such as this from the other buildings in the scene [30]. Figure 13 shows the reconstructed scene after each model has been extruded and inserted into the scene.

Figure 13 Approximately Here

The accuracy of the reconstructed scene was then compared against a hand constructed ground truth model. Using an interactive multi-view modeling system [23], wireframe buildings were constructed by hand aligning their shapes with building boundaries in four images simultaneously. The baseline accuracy of hand constructed models is a function of user accuracy and errors in the known camera parameters. The baseline accuracy for user constructed models for the Fort Hood dataset has been

Error Metric	Peak Roofs	Flat Roofs	Total
Intervertex Plaimetric	0.55	0.491	0.535
Intervertex Alitmetric	0.59	0.402	0.543
Intervertex Absolute	0.806	0.634	0.763
Wireframe Distance	0.631	0.487	0.595
Volumetric Overlap	98.97%	99.5%	99.1%

Table 3: Fort Hood median error results for peak and flat roof building models. Error measured for the eight true positive buildings versus groundtruth wireframes. Intervertex distances reported in meters while the wireframe distance is reported in pixels.

reported in [6] and is approximately less than a third of a meter. The ground truth model was then compared to the recovered scene using the error metrics described in Section 6. Table 3 reports mean reconstruction accuracy for correctly detected buildings by rooftop shape.

6.3 Avenches Reconstruction

The Avenches dataset is a set of aerial photographs collected over Avenches, Switzerland in 1991 for the Institute for Geodesy and Photogrammetry at the Swiss Federal Institute of Technology (ETH), Zurich. The photography has 60% sideward overlap coverage of an urban area at a scale of 1:5000. Prior to scanning, the photographs were handled on many occasions and the effects on the images are apparent with scratches, dust, and hair. Nevertheless, the data is indicative of what can be expected from typical, high detail photogrammetric surveys. The data set was prepared for use within the AMOBE project and was provided for use by the ETH Zurich photogrammetric and computer vision groups, in cooperation with Sarnoff Research International (SRI, California).

The portion of the data set used for testing includes a single downlooking view of an industrial boatyard site, with a ground sample distance of 0.19 meters as well as a high resolution DEM computed from the overlapping imagery that was provided as part of the data set. Figure 14 shows a ground truth CAD model constructed by ETH. The ground truth model includes the 3D shape and position of building rooftops and was used to evaluate our reconstruction results.

Figure 14 Approximately Here

Initially, the polygon detection algorithm was run on the optical image to delineate building rooftop boundaries. In all 18 polygons were detected by the system, eight of these correspond to a building rooftop. Although a significant number of regions were produced by the segmentation phase that are not buildings, there are a number of approaches to detecting and eliminating these regions. Marengoni [35] has shown that straightforward knowledge-based techniques can be used to discriminate building from non-building regions in aerial data. Assuming that non-building regions will have a DEM shape that does not resemble an SPD surface, we can use the correlation score returned by the model-indexing phase to eliminate non-building regions. The results shown here, a threshold of 0.85 was selected. This threshold

ID	Correlation	1st /Error	2nd/Error	3rd /Error	Selected
A	0.987	Peak-65 / 0.101m	Peak-55/2.12m	Peak-45 / 4.83m	Peak-65
B	0.604	Eliminated			
C	0.944	Peak-55 / 0.261m	FlatPeak-65/0.982m	Peak-65 / 3.2m	Peak-55
D	0.812	Eliminated			
E	0.993	Plane / 0.401m	Gable-35/1.297m	FlatPeak-35/ 1.92m	Plane
F	0.894	Plane/ 0.391m	Cylinder/1.238	FlatPeak-45 / 2.83m	Plane
G	0.923	Plane/ 0.362m	Peak-35/1.784m	Cylinder / 3.34m	Plane
H	0.941	Plane / 0.122m	Peak-45/0.293m	FlatPeak-35 / 0.891m	Plane
I	0.964	Peak-65/ 0.378m	Peak-55/0.211m	FlatPeak-55 / 0.915m	Peak-55
J	0.946	FlatPeak-35 / 0.710m	Peak-45/0.204	FlatPeak-45 / 0.710m	Peak-45
K	0.955	Plane / 0.732m	FlatPeak-55/3.76	FlatPeak-35 / 1.33m	Plane
L	0.722	Eliminated			
M	0.893	Cylinder / 1.93m	FlatPeak-75/1.89	Plane/ 1.74m	Plane
N	0.798	Eliminated			
O	0.819	Eliminated			
P	0.848	Eliminated			
Q	0.803	Eliminated			
R	0.795	Eliminated			
S	0.942	Plane / 0.982m	FlatPeak-45/1.08m	Gable-55 / 1.14m	Plane

Table 4: Top three models indexed from the SPD for each of the 11 building regions in the Avenches scene. Model selected based on the lowest fit error measure is also shown.

seeks to eliminate regions based on their maximum correlation score with the surfaces in the SPD. Under this assumption, nonbuilding regions will have DEM surfaces sufficiently different from rooftops to be discriminated with a single threshold which may not always be the case. More sophisticated approaches to fusing data from several sources in order to classify segmented scene regions is discussed in [26] and [35]. Figure 15 shows the results for rooftop detection in the Ascona scene.

Figure 15 Approximately Here

Using the correlation score threshold, 7 polygons were eliminated and the remaining 11 were then fit to the underlying DEM regions. Table 4 reports the correlation score for each region in the scene. For regions that were not eliminated based on the correlation threshold, the top three indexed models from the SPD and the final fit error that resulted from fitting each of the three models to the corresponding DEM region are also shown in Table 4. The model and corresponding free parameters that lead to the lowest fit error was then selected and inserted in the scene.

The reconstructed three-dimensional models are shown in Figure 16. The three-dimensional model can be viewed from new viewpoints for simulated site-flythroughs and visualization purposes.

bf Error Metric	Peak Roofs	Flat Roofs	Total
Intervertex Plaimetric	0.10	0.08	0.09
Intervertex Alitmetric	0.12	0.104	0.112
Intervertex Absolute	0.238	0.212	0.225
Wireframe Distance	0.311	0.253	0.303
Volumetric Overlap	99.47%	99.1%	99.3

Table 5: Avenches median error results for peak and flat roof building models. Error measured for true positive buildings versus groundtruth wireframes. Intervertex distances reported in meters while the wireframe distance is reported in pixels.

Figure 16 Approximately Here

Errors in the initial detection phase led to erroneous reconstructions in two cases (Polygons K and S). In both cases, the constructed image histogram resulted in correlation scores that did not justify their elimination from the scene using a threshold. The top right building, whose rooftop was partially detected as polygon S, was not reconstructed due to errors in the polygon detection algorithm (Figure 15). Accidental alignment of the rooftop and nearby ground clutter (a vehicle in this case) caused the polygon detection algorithm to segment the right half of the peaked roof and part of the surrounding ground. Significant rooftop clutter on the peaked-roof wing of the center building (see polygon K) caused the polygon detection algorithm to again segment only half of the rooftop. Polygon K was then incorrectly reconstructed as a flat roof building because the DEM in this region was particularly poor and was somewhat flat. A small flat roofed wing of the long building (polygon H), near the bottom of the image, was reconstructed as peaked due to noisy elevation data. The remaining regions were reconstructed correctly.

The accuracy measures, discussed in section 6, with respect to the CAD model provided as part of the Avenches data set, were computed using the reconstructed scene. The ground sample distance for the image used in the experiment here was 0.19 meters. Table 5 reports several error metrics for the eight true-positive buildings and the peak and flat roof models independently.

Building rooftop errors are about a third of a meter (in terms of wireframe distance) for all the buildings in the data set. Reconstructed peak-roofed buildings were slightly less accurate than flat roof buildings. Stereo-optical processing of these regions may be sensitive to surface reflectance changes that can occur when observing the angled planes of a peaked-roof structure from different positions and self-shadowing effects.

7 Conclusions

We have introduced a model-directed approach to recognizing and reconstructing a variety of building types from aerial imagery. The technique encodes the differential properties of the DEM surface as an orientation histogram and correlates the histogram with a set of predefined surfaces in the Surface Primitive Database. The primary goal of the algorithm is to index into a set of many possible parametric

surfaces that represent different rooftop shapes. Given the reduced set of surfaces that most closely resemble the DEM, a surface-fitting algorithm then fits each surface to the elevation data in order to select a final surface model and parameters for reconstructing the building rooftop. The approach has been shown to be robust to the errors commonly encountered in automatically derived elevation maps.

Surface reconstruction, using the restricted set of parameters derived from the model indexing process, is more robust by focusing the iterative fitting procedure on fewer parameters and a parametric model that is likely to fit to the DEM. A significant advantage of this approach is that a single complex, parametric description that captures all the variability in rooftop surfaces is not required. Instead, rooftops commonly found in urban areas can be predefined in the SPD and selected according to shape evidence present in the DEM. This restricted class of parameterized rooftops (peaked-roofs, for example) are easier to model, have fewer parameters, and are simpler to fit to the data. Furthermore, the indexing phase computes the initial estimate of the rotation between the parametric model to be fit and the DEM data.

An alternative approach using the model-based fitting methodology presented here addresses buildings that are composed of several different surfaces in the SPD through recursive decomposition. This approach utilizes a different control strategy to cluster outlier points and recursively recognizes and fits surfaces to these new regions [28]. In this way, multilevel buildings and buildings with significant and complex substructure can be reconstructed. For example, buildings with flat chimneys, curved air vents, and peaked dormers can be interpreted with a uniform strategy until there is insufficient resolution in the finest level of substructure.

8 acknowledgment

The authors would like to thank Howard Schultz for generating DEMs with the Terrest system for the experiments presented in this paper. This work was supported by the Defense Advanced Research Projects Agency, Contract Number DACA76-97-K-0005, the DARPA AASERT program (via U.S. Army TEC), Grant Number DAAG55-97-1-0188, the DARPA RADIUS program (via U.S. Army TEC), Contract Number DACA76-92-C-0041, and the Army Research Office, Grant Number DAAH04-96-1-0135.

References

- [1] M Boldt, R. Weiss, and E. Riseman. Token-based extraction of straight lines. *IEEE Transactions SMC*, 19:1581–1593, 1989.
- [2] M. Bro-Nielson. Detecting buildings in aerial images. Technical report, Institute of Mathematical Statistics and Operations Research, 1992. Technical University of Denmark.
- [3] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8:679–698, 1986.
- [4] R. Chellapa, L. Davis, C. Lin, T. Morsse, C. Rodriguez, A. Rosenfeld, X. Zhang, , and Q. Zheng. Site-model-based monitoring of aerial images. *IEEE Conference on Computer Vision and Pattern Recognition, New York*, 1994.
- [5] R. Chung and R. Nevatia. Recovering building structures from stereo. *presented at IEEE Workshop on the Applications of Computer Vision (WACV)*, 1992.
- [6] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, and E. Riseman. The ascender system: Automated site modeling from multiple aerial images. *DARPA Image Understanding Workshop*, 1997. New Orleans, LA.
- [7] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, and E. Riseman. The ascender system: Automated site modeling from multiple aerial images. *Computer Vision and Image Understanding*, 1998.
- [8] R. Collins, C. Jaynes, Y. Q. Cheng, X. G. Wang, F. Stolle, H. Schultz, A. Hanson, and E. Riseman. The umass ascender system for 3d site model construction. In Ed. O. F. a. T. Strat, editor, *ARPA IU RADIUS*. San Francisco, CA: Morgan Kaufmann Publishers, 1997.
- [9] R. Collins and R. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. *presented at IEEE International Conference on Computer Vision*, 1990. Osaka, Japan.
- [10] S. Coorg, N. Master, and S. Teller. Acquisition of a large pose-mosaic dataset. *Computer Vision and Pattern Recognition*, pages 872–878, 1998.
- [11] S. Coorg and S. Teller. Spherical mosaics with quaternions and dense correlation. *IJCV*, 37(3):259–273, June 2000.
- [12] M. Cord, M. Jordan, J. Cocquerez, and N. Paparoditis. Automatic extraction and modeling of urban buildings from high resolution aerial images. *ISPRS GIS'99*, pages 187–192, 1999.
- [13] O. Faugeras, S. Laeau, L. Robert, G. Csurka, and C. Zeller. 3-d reconstructions of urban scenes from sequences of images. *Computer Vision and Image Understanding (CVIU)*, pages 145–168, 1995.
- [14] A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Forstner, L. Pluemer, and V. Steinhage. Extracting buildings from aerial images using hierarchical aggregation in 2d and 3d. *Computer Vision and Image Understanding, vol. 72, pp. 185-203*, 1998.
- [15] A. Fischer, T. H. Kolbe, and F. Lang. Integration of 2d and 3d reasoning for building reconstruction using a generic hierarchical model. *presented at Semantic Modeling for the Acquisition of Topographical Information from Images and Maps*, 1997. Bonn, Germany.
- [16] W. Forstner. Mid-level vision processes for automatic building extraction. *presented at Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, 1995. Monte Vertia, Switzerland.

- [17] D. Frere, M. Hendrickx, J. Vanderkerckhove, T. Moons, and L. VanGool. On the reconstruction of urban roofs from aerial images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, 1997.
- [18] P. Fua and A. J. Hanson. An optimization framework for feature extraction. *Machine Vision and Applications*, vol. 4, pp. 59-87, 1991.
- [19] N. Haala and M. Hahn. Data fusion for the detection and reconstruction of buildings. *Workshop on Automated Extraction of Man-Made Objects from Aerial and Space Images*, 1995.
- [20] D. Harwood, S. Chang, and L. Davis. Interpreting aerial photographs by segmentation and search. *presented at DARPA Image Understanding Workshop*, 1987.
- [21] M. Hatzitheodorou. Shape from shadows: Theoretical and computational aspects. *Computer Science: Columbia University*, 1989.
- [22] O. Henricsson. Analysis of image structures using color attributes and similarity relations. *Institute of Geodesy and Photogrammetry. Zurich: Swiss Federal Institute of Technology*, 1996.
- [23] A. Hoogs, W. Bremner, and D. Hackett. The radius phase ii program. *The Radius Image Understanding Progra,*, 1998.
- [24] B. K. P. Horn. Robot vision. *Cambridge, MA: MIT Press*, 1986.
- [25] A. Huertas and R. Nevatia. Detecting buildings in aerial images. *Computer Vision, Graphics and Image Processing (CVGIP)*, vol. 41, pp. 131-152, 1988.
- [26] C. Jaynes. *Automatic Model Acquisition and Aerial Image Understanding*. PhD thesis, University of Massachusetts, 2000.
- [27] C. Jaynes, A. Hanson, and E. Riseman. Model-based recovery of buildings in optical and range images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Bonn, Germany, 1997.
- [28] C. Jaynes, A. Hanson, and E. Riseman. Recursive recovery of three-dimensional scenes. *DARPA Image Understanding Workshop*, 1998.
- [29] C. Jaynes and M. Partington. Pose calibration using approximately planar urban structure. *Asian Conference on Computer Vision (ACCV'99)*, 1999.
- [30] C. Jaynes, E. Riseman, and A. Hanson. Building reconstruction from optical and range images. *presented at CVPR*, 1997. San Juan, PR.
- [31] C. Jaynes, F. Stolle, and R. Collins. Task-driven perceptual organization for extraction of rooftop polygons. *presented at DARPA Image Understanding Workshop, Monterey, CA*, 1994.
- [32] E. LeBras-Mehlman, M. Schmitt, O. D. Faugeras, and J. D. Boissonnat. How the delaunay triangulation can be used for representing stereo data. *INRIA*, 1988. Sophia-Antipolis.
- [33] C. Lin, A. Huertas, and R. Nervatia. Detection of buildings using perceptual groupings and shadows. *presented at IEEE Computer Vision and Pattern Recognition*, 1994.
- [34] Y. T. Liow and T. Pavlidis. Use of shadows for extracting buildings in aerial images. *Computer Vision, Graphics and Image Processing*, pp. 242-277, 1990.
- [35] M. Marengoni, C. Jaynes, A. Hanson, and E. Riseman. Ascender ii: A visual framework for 3d reconstruction. *DARPA Image Understanding Workshop*, 1998.

- [36] J. McGlone. Bundle adjustment with geometric constraints for hypothesis evaluation. *presented at ISPRS*, 1995.
- [37] J. McGlone and J. Schufelt. Projective and object space geometry for monocular building extraction. *presented at IEEE Computer Vision and Pattern Recognition (CVPR)*, 1994.
- [38] D. McKeown and J. Denlinger. Map-guided feature extraction from aerial imagery. *presented at IEEE Workshop on Computer Vision: Representation and Control*, 1984. Annapolis, MD.
- [39] D. M. McKeown, W. A. Harvey, and J. McDermott. Rule-based interpretation of aerial imagery. *IEEE T-PAMI*, vol. 7, pp. 570-585, 1985.
- [40] D. M. J. McKeown, W. A. Harvey, and L. E. Wixson. Automating knowledge acquisition for aerial image interpretation. *CVGIP*, vol. 46, pp. 37-81, 1989.
- [41] R. Mohan and R. Nevatia. Using perceptual organization to extract 3-d structures. *IEEE T-PAMI*, vol. 11, pp. 1121-1139, 1989.
- [42] J. L. Mundy, A. J. Heller, , and D. W. Thompson. The concept of an effective viewpoint. *presented at DARPA Image Understanding Workshop*, 1988. Los Altos, CA.
- [43] M. Nagao and T. Matsuyama. *A Structural Analysis of Complex Aerial Photographs*. New York: Plenum Press, 1980.
- [44] S. Noronha and R. Nevatia. Detection and description of buildings from multiple aerial images. *presented at IEEE Computer Vision and Pattern Recognition*, 1997.
- [45] M. Polis and D. McKeown. Iterative tin generation from digital elevation models. *presented at Computer Vision and Pattern Recognition*, 1992.
- [46] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. Numerical recipes. *Cambridge, MA: Cambridge University Press*, 1988.
- [47] M. Roux and D. McKeown. Feature matching for building extraction from multiple views. *presented at DARPA Image Understanding Workshop*, 1994.
- [48] H. Schultz. Terrain reconstruction from oblique views. *presented at DARPA Image Understanding Workshop*, 1994. Monterey, CA.
- [49] H. Schultz. Terrain reconstruction from widely separated images. *presented at SPIE*, 1995. Orlando, FL.
- [50] J. Shufelt and D. M. McKeown. Fusion of monocular cues to detect man-made structures in aerial imagery. *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 57, pp. 307-330, 1993.
- [51] T. M. Silberberg, D. Harwood, and L. S. Davis. Three-dimensional object recognition using oriented model points. *Techniques for 3D Machine Perception*, A. Rosenfeld, Ed. Elsevier Science Publishers, P.V. (N. Holland), pp. 271-320, 1986.
- [52] M. Tavakoli and A. Rosenfeld. Building and road extraction from aerial photographs. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 12, pp. 84, 1982.
- [53] V. Venkateswar and R. Chellapa. A framework for interpretation of aerial images. *Tenth International Conference on Pattern Recognition*, 1990.

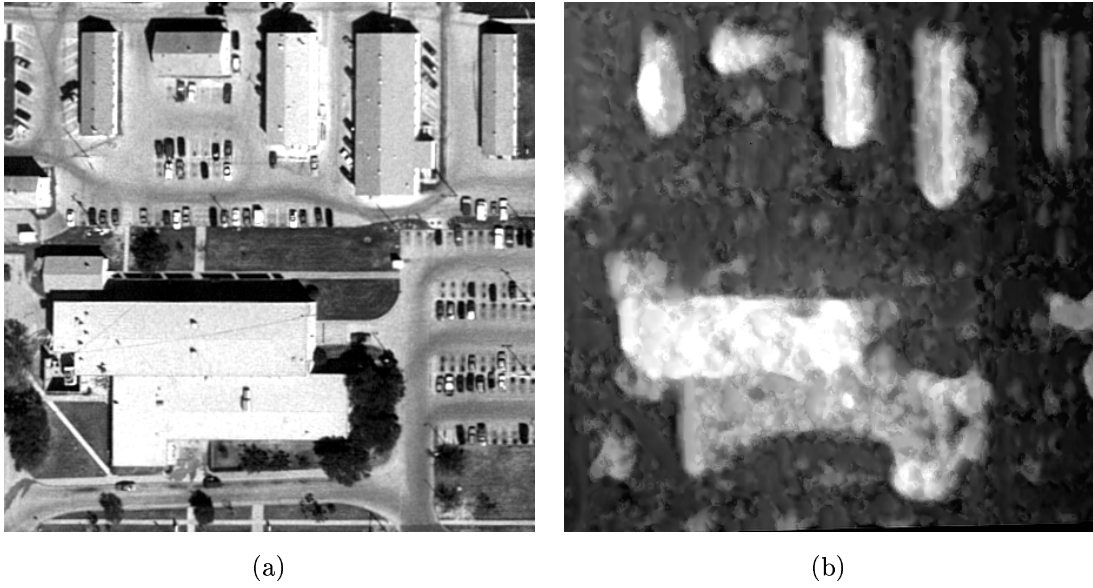


Figure 1: Aerial image data used by the reconstruction system. (a) Optical image containing a number of different buildings. (b) Corresponding digital elevation map (DEM) acquired through stereo-optical processing (brightness is proportional to elevation).

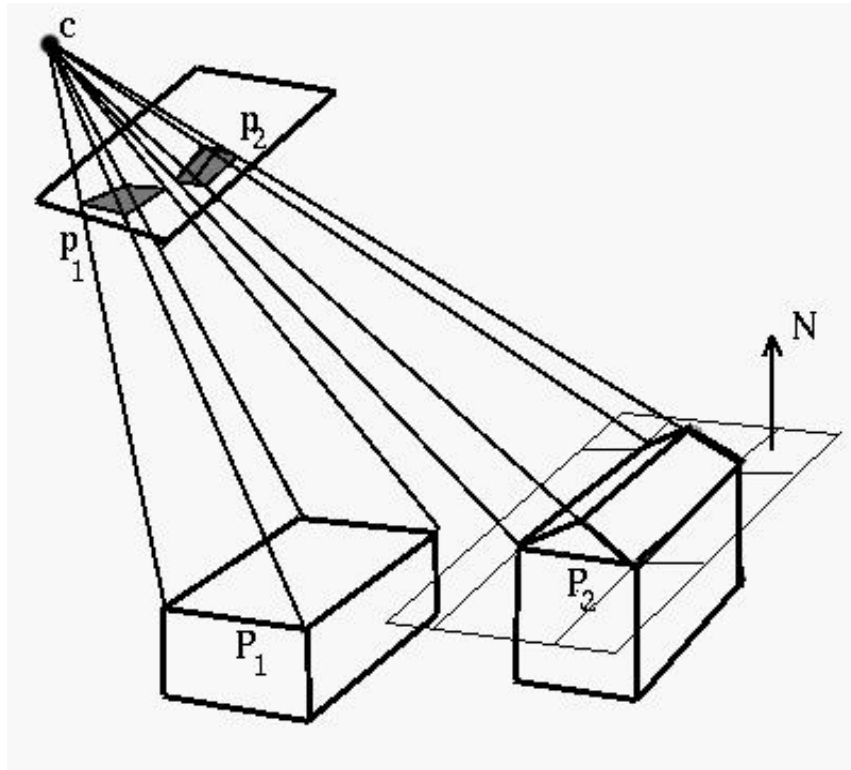


Figure 2: Geometry of the rooftop segmentation problem. The role of the polygon detection algorithm is to detect the outermost, planar polygon that encloses the building rooftop. Rooftop polygons are assumed to lie in a plane whose normal N is parallel to the world Z -axis.

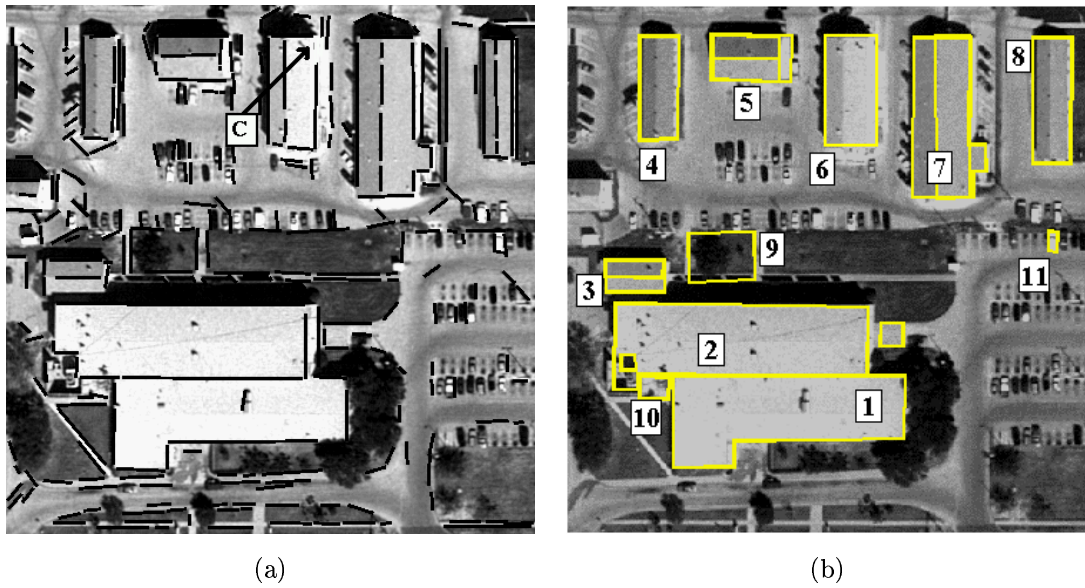


Figure 3: Detection of rooftop regions in the Fort Hood scene. (a) Extracted line segments used to detect orthogonal corners and rooftop boundaries (shown in black). Corner labeled 'C' was not found using feature extraction but is discovered using perceptual completion (see text). (b) Closed polygonal regions.

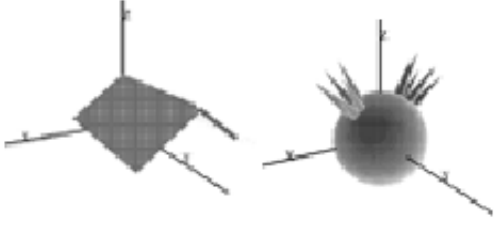
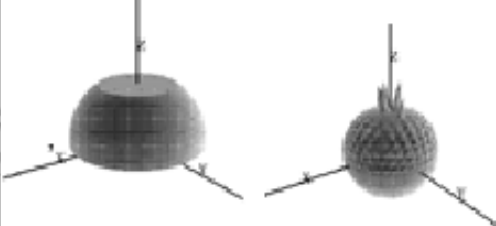

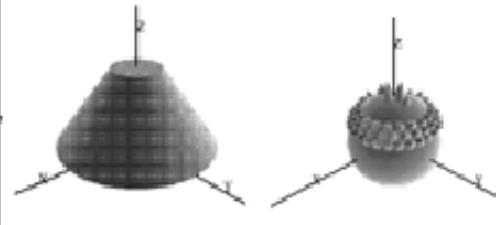
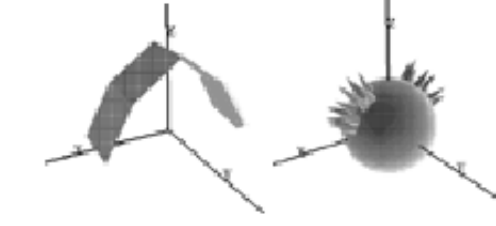
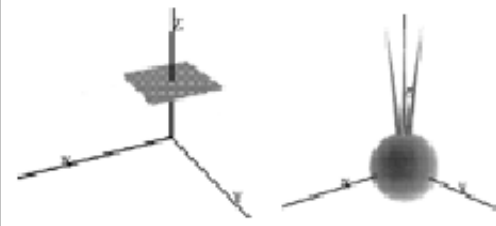
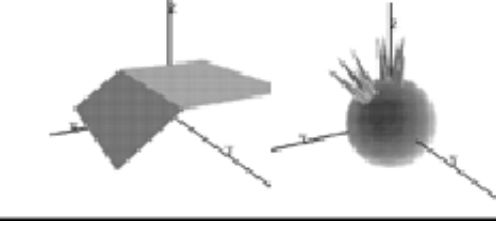
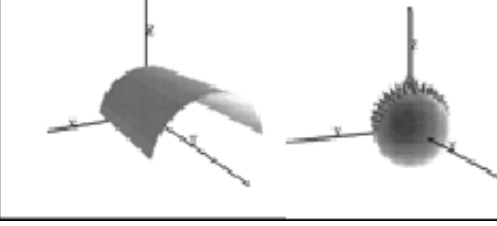
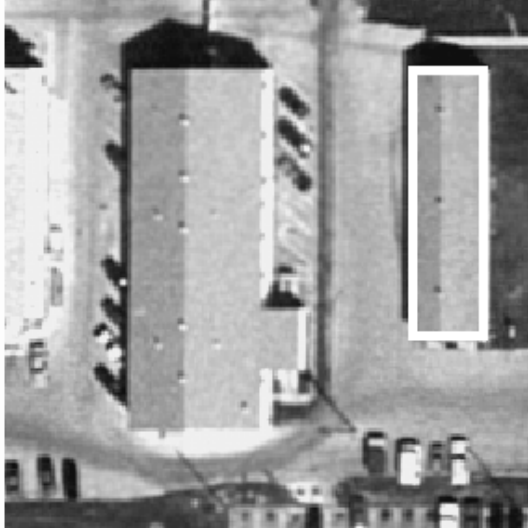
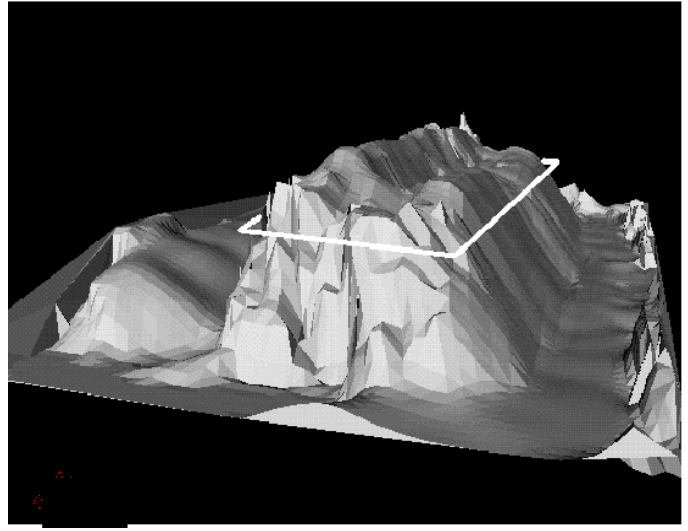
Name	Surface Primitive/ Histogram	Name	Surface Primitive/ Histogram
Peak {theta}		Dome {TopRad, B/H}	
Flat-Peak {theta1, theta2}		Conic {TopRad, B/H}	
Barn {theta1, theta2, theta3}		Plane	
Gable {theta}		Cylinder {B/H}	

Figure 4: The Surface Primitive Database used for indexing. Each surface primitive is shown along with the orientation histogram (see text). The parameterization for each model is shown to the left, representing several different submodels that are stored in the database for discrete parameter values.

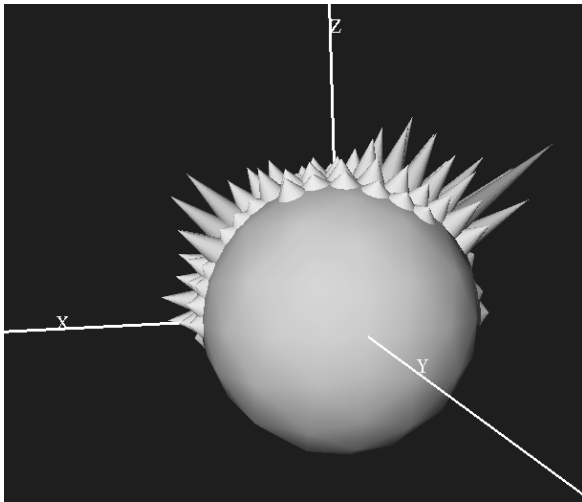


(a)

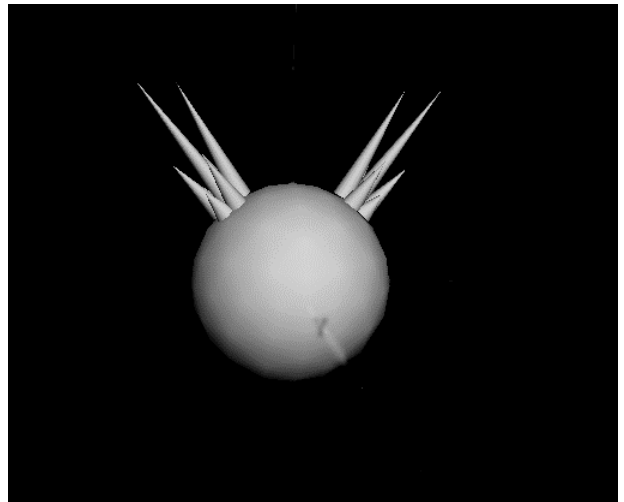


(b)

Figure 5: Building rooftop and corresponding DEM surface mesh. (a) Close-up view of Polygon 8 from Figure 1a. (b) Resulting surface after Deluanay triangulation has been applied to the DEM points.



(a)



(b)

Figure 6: Histogram derived from DEM compared to precomputed SPD histogram. (a) Histogram corresponding to polygon 8 (Figure 5). (b) Histogram from the SPD that most closely matched polygon 8 (see Text).

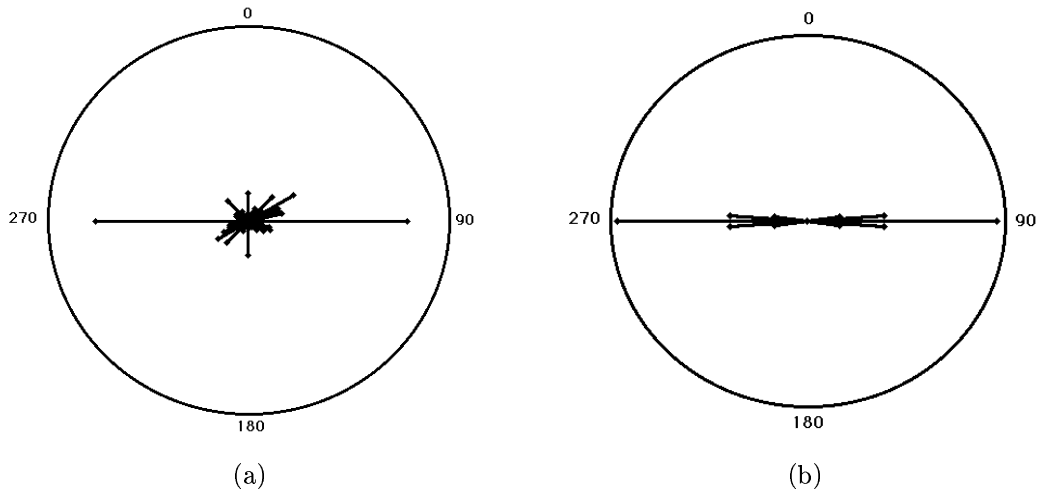


Figure 7: Orthographic projection of two histograms to the X-Y plane for comparison. (a) Histogram corresponding to polygon 8 (Figure ??). (b) Histogram from the SPD that most closely matched polygon 8 (see Section 4.1).

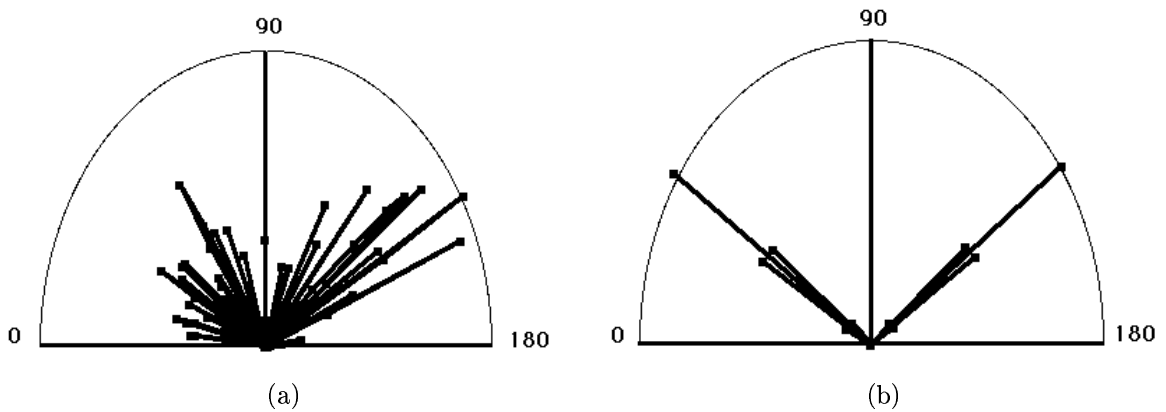


Figure 8: Orthographic projection of two histograms to the Y-Z plane for comparison. (a) Histogram corresponding to polygon 8 (Figure 5). (b) Histogram from the SPD that most closely matched polygon 8 (see Section 4.1).

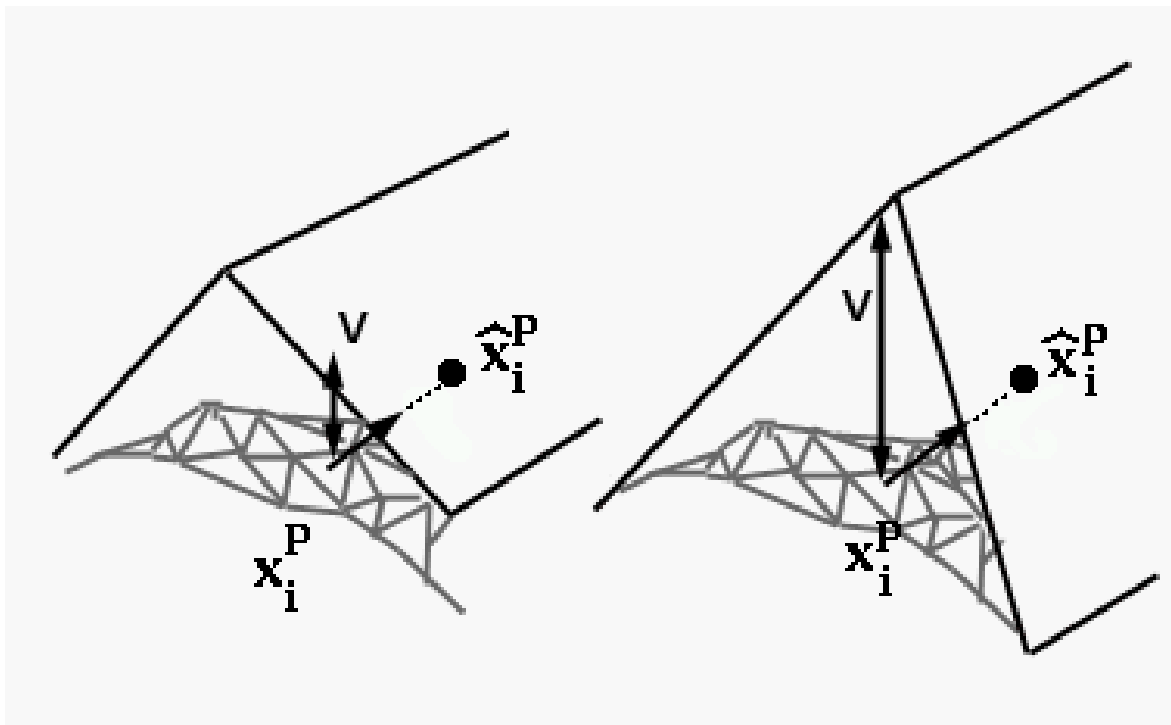


Figure 9: Residual error is measured along the surface normal for each mesh patch to the current model. Error measured in a more arbitrary way, along the Z-axis, for example, can introduce high error along discontinuities or where the slope of the surface is high in Z, and leads to unstable convergence properties during optimization.

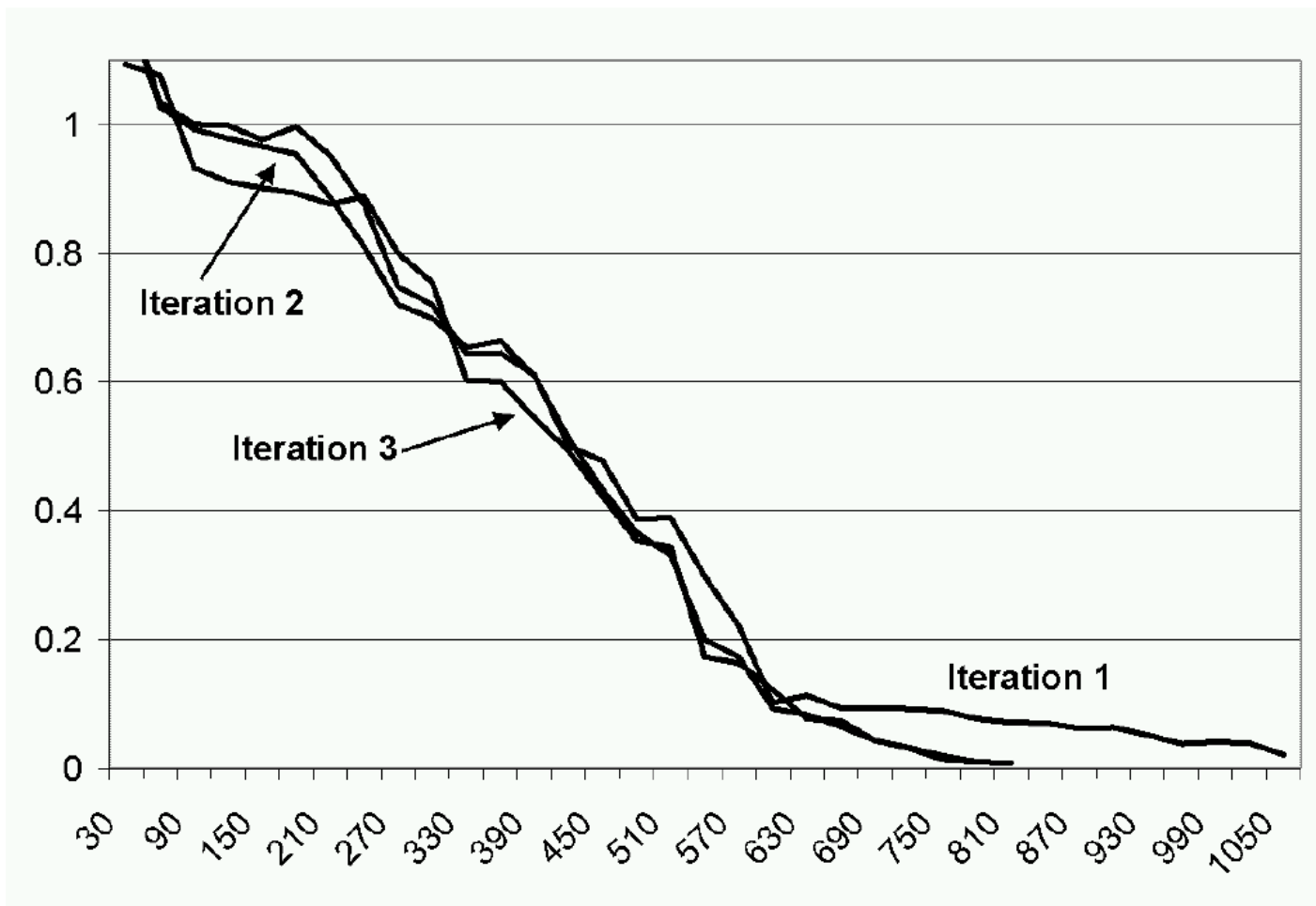


Figure 10: Number of iterations versus fit error for each pass of the surface-fitting algorithm for Region 8 using the Peak-75 model. Fitting required three iterations until the chi-squared measure no longer improved and the algorithm terminated.

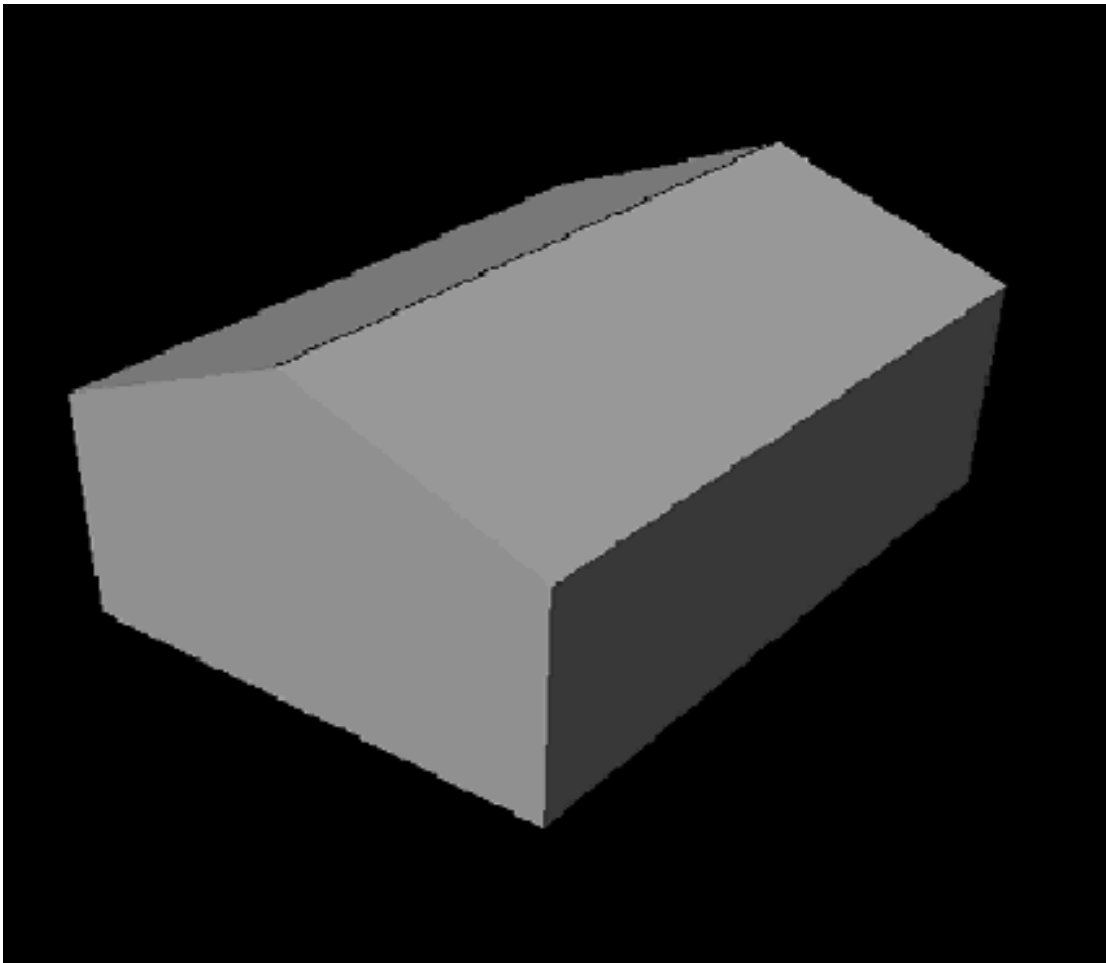


Figure 11: Reconstructed peaked roof building corresponding to region 8 from Figure 5. The model converged with an overall fit error of 0.021 meters.

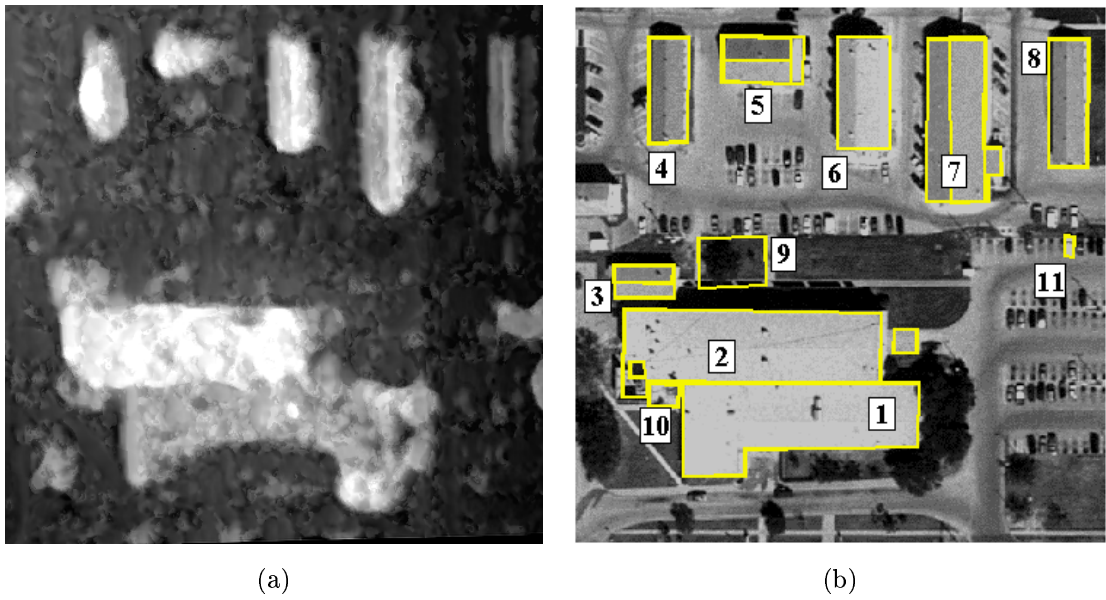


Figure 12: Data used to reconstruct the Fort Hood scene. (a) Elevation map (DEM) acquired through stereo-optical processing. (b) Eleven detected regions, partly corresponding to rooftop boundaries.

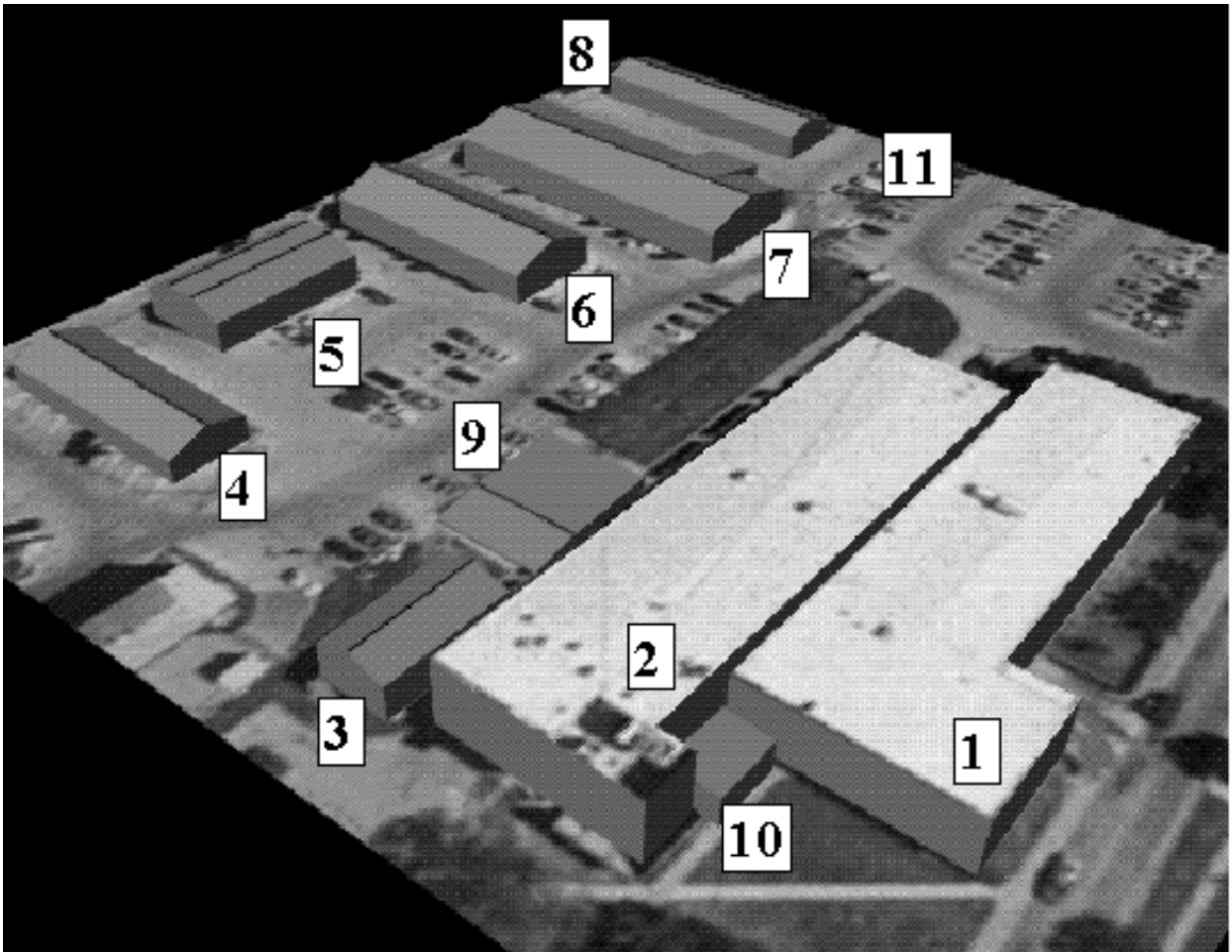


Figure 13: Reconstruction results of the Fort Hood scene. Building labels correspond to detected polygon regions in Figure 12

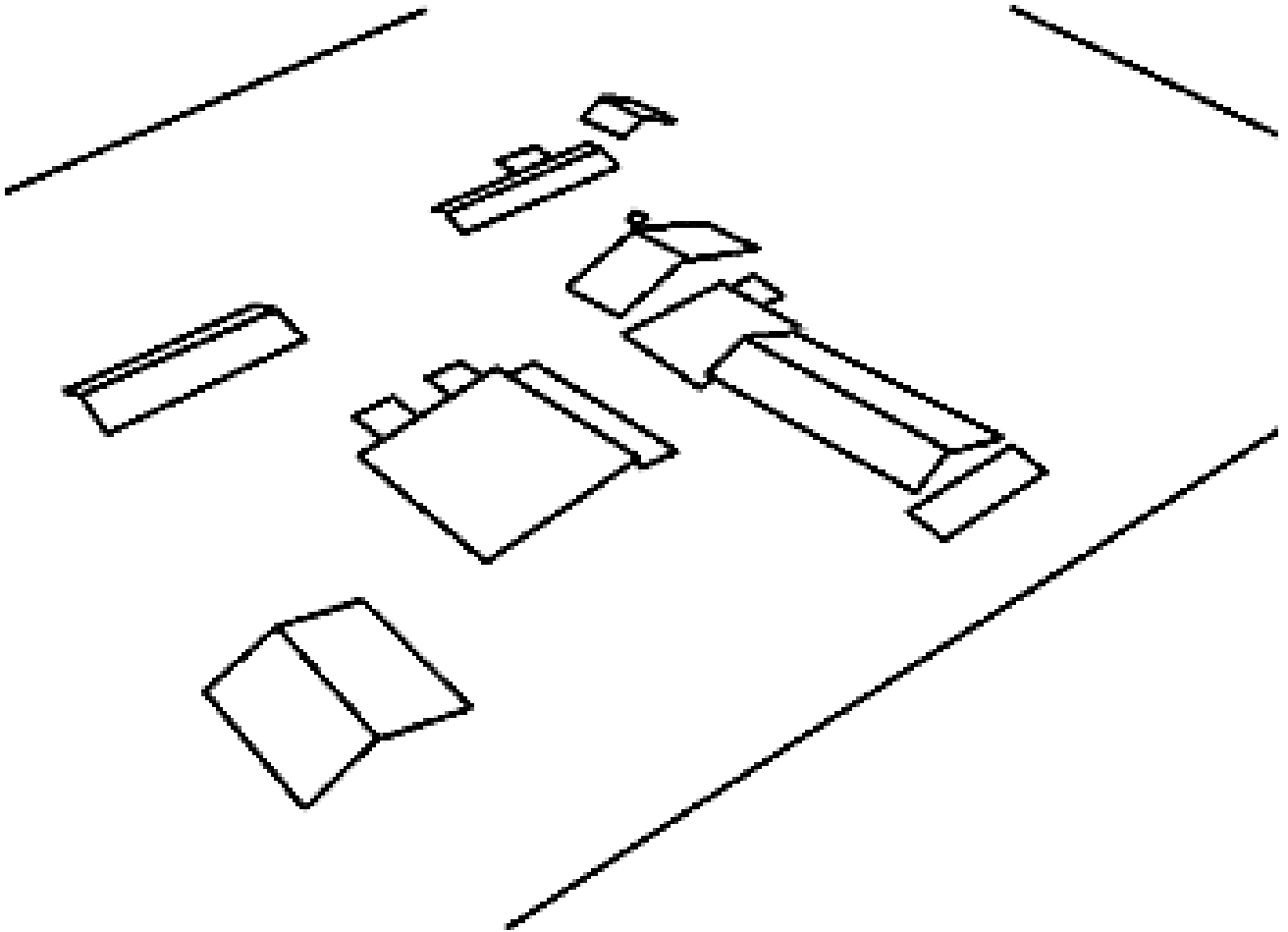


Figure 14: Groundtruth model provided as part of the Avenches dataset. Figure shows a perspective view of the 3D groundtruth building wireframes.

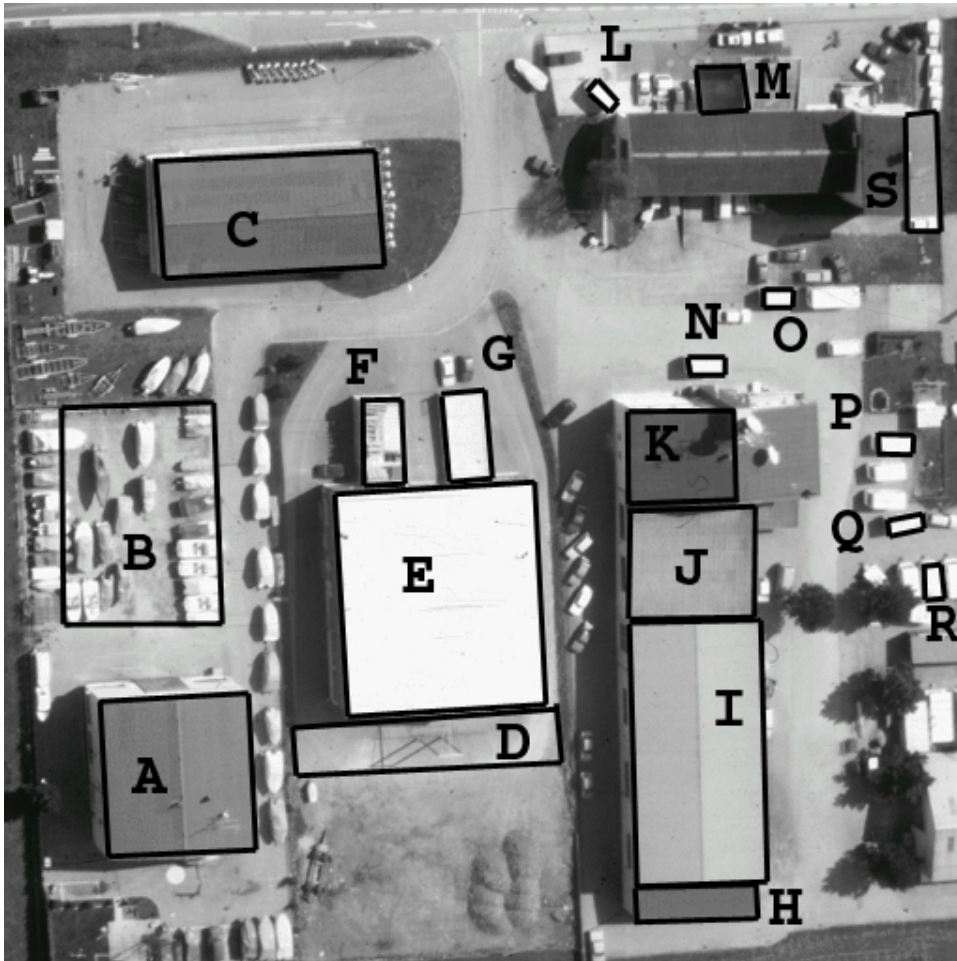


Figure 15: Polygons detected in the Ascona scene. A total of 19 polygons were extracted using the perceptual grouping algorithm. Of these, eight were eliminated based on the fit error threshold.

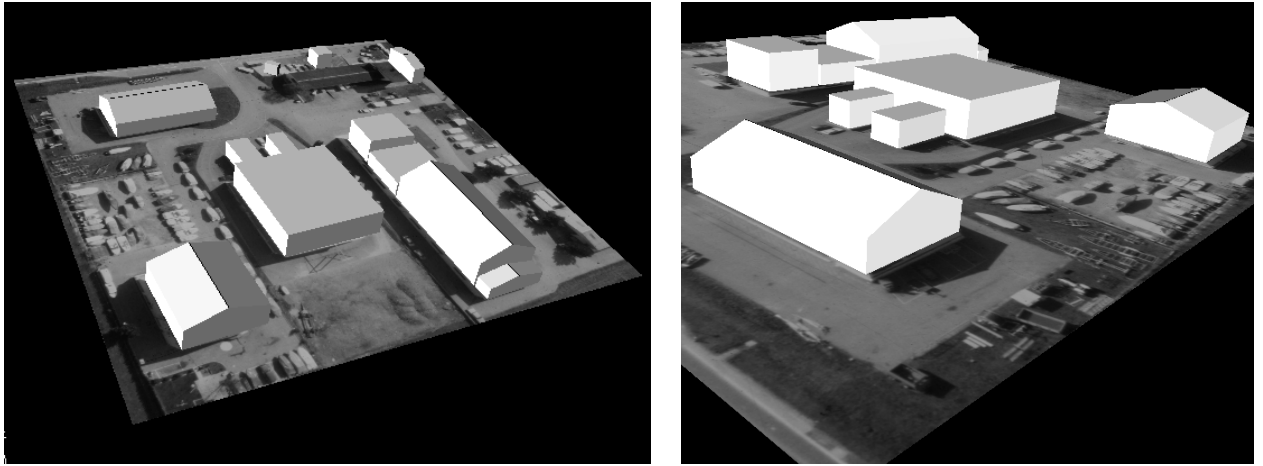


Figure 16: Reconstruction results for Avenches site displayed as three-dimensional models. Entire scene, shown at left, and close-up view of a peaked roof building, at right.

Research	Citation/Year	Input Data	Footprint Shape	Rooftop Shape
Liow, Pavlidis	[34], 1990	Multi-optical	compound	flat
Venkateswar, Chellapa	[53],1990	Multi-optical	compound	flat
Fua, Hanson	[18],1991	Multi-optical	compound	flat
Bro-Nielson	[2],1992	Multi-optical Color	complex polygon	flat,peak
Shufelt, McKeown	[50],1993	Multi-optical	complex polygon	flat,peak
Lin,Huertas Nevatia	[33],1994	Multi-optical	compound	flat
McGlone, Shufelt	[37],1994	Multi-optical	compound	flat,peak
Jaynes, Stolle Collins	[31],1994	Monocular	compound	flat
Haala,Hahn	[19], 1995	Multi-optical DEM	complex	polyhedra
Lin,Nevatia	[33],1994	Optical	complex	polyhedra
Foerstner	[16],1995	DEM	complex	polyhedra
Hennricsson, Bignone, et. al.	[22], 1996	Optical Stereo	compound	flat,peak gabled
Collins, Jaynes, et. al.	[8], 1997	Multi-optical	compound	flat
Fischer, Kolbe, Lang	[15]	Multi-optical	complex	polyhedra
Moons, Frere, et. al.	[17] Stereo	Optical	complex	polyhedra
Cord	[12]	Multi-optical DEM	complex	polyhedra
Jaynes, Riseman, Hanson	This work	Multi-optical DEM IFSAR	complex polygon	parametric surfaces: flat,peak gabled, curved dome, etc.

Table 6: Summary of building detection and reconstruction work in the previous ten years.