Three-Dimensional Grouping and Information Fusion for Site Modeling from Aerial Images*

Christopher O. Jaynes, Frank R. Stolle, Howard Schultz, Robert T. Collins, Allen R. Hanson and Ed M. Riseman

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

1 Introduction

1.1 Motivation

The building reconstruction strategies that have been used in the UMass Ascender system are reasonably effective, but are tuned to extract only one generic building class with single-level, flat roofs bounded by rectilinear polygonal shapes. Extensions to the system must be considered in order to handle other common building types. Examples are multi-level flat roofs (or single-level flat roofs containing significant substructures such as large air conditioner units), peaked-roof buildings, juxtapositions of flat and peaked roofs, curved-roof buildings such as Quonset huts or hangars, as well as buildings with more complex roof structures containing gables, slanted dormers or spires.

To achieve the desired goal of a more general and flexible building extraction system in the ARPA/ORD RADIUS program, a significant research effort is underway at UMass to explore alternative detection and reconstruction strategies that combine a wider set of algorithms for generating and fusing 2D and 3D information. The types of strategies being considered involve generation and grouping of a larger variety of primitive geometric elements (e.g. lines, surfaces, parametrized models) stored as symbolic tokens, as well as techniques for fusing geometric token data with high-resolution digital elevation map (DEM) data. In addition to the 2D line, corner and closed polygons currently extracted by the system, new techniques have been developed to extract 3D lines, corners and surfaces. By verifying geometric consistencies between 2D and 3D tokens associated with building components, larger and more complex 3D structures can be organized using context-sensitive, knowledge-based strategies.

There are alternative ways for organizing the 3D reconstruction mechanisms into a processing taxonomy. From one perspective the algorithms divide into bottom-up (i.e. data-directed) and top-down (i.e. knowledge-directed) approaches.

Many of the bottom-up grouping strategies are obvious, since they are directly derived from the consistency constraints that are embodied in the natural geometric relationships between the relevant tokens. For example, 2D (or 3D) lines whose endpoints are spatially proximate can be grouped into 2D (or 3D) corners. In a more complicated process of grouping geometric tokens of different types, 3D line tokens that bound a planar roof surface could be consistently grouped if the planar surface boundaries are approximately consistent with their corresponding 3D lines. In this case the information would be fused under some spatial optimization criterion. There are many other ways that primitive geometric elements can be brought together to form more complex geometric structures.

In contrast, the RADIUS application domain involves 3D site reconstruction of objects and scene contexts for which a great deal of cultural knowledge is available, implying the use of a set of top-down strategies. An image analyst may know the general classes of buildings that are likely to appear in a geographical area, and therefore the associated geometric constraints that would be present can be predicted. Thus, the general models of peaked-roof and flat-roof buildings would involve a different set of geometric tokens and constraints on the spatial relationships between them. When the information in an image is somewhat degraded, the difficulty

*Funded by the RADIUS project under ARPA/Army TEC contract number DACA76-92-C-0041.

in extracting the relevant geometric details can be compensated for by employing semantic knowledge of context in the form of parametric models. For example, in many images the center ridge-line of a peaked roof can be quite difficult to extract; however, if a parametric model of a peaked-roof building is employed, the grouping and fusion of the geometric primitives that **are** extractable may be able to sufficiently constrain the problem. Consequently, a top-down taxonomy would involve parametric models of the various building classes expected.

In order to construct different building strategies, sequences of bottom-up grouping routines must be composed to extract the necessary 2D and 3D features, followed by application of the appropriate spatial constraints between tokens of the same and different types [6]. These strategies might be supplemented with a context- sensitive set of features that would allow the appropriate model to be invoked.

The reader must understand that the work presented here is still early in development and, unlike the UMass Ascender system [4], these proceedings, (which has been tested under varying conditions), has not yet matured into a well-tested operational system. Furthermore, it should be pointed out that the reconstruction strategies discussed here are at varying stages of development. What follows is a discussion of a general paradigm for 3D grouping and information fusion for the RADIUS application domain, accompanied by a set of case studies with fairly interesting results on a variety of different problems and scene domains. These initial results and approaches are being used to further the development of a more general automated system.

2 The Geometric Elements

Before describing specific strategies for 3D grouping and data fusion, the set of 2D and 3D primitive elements to be grouped/fused are described. Nearly all are geometric elements such as points, lines, curves and planes that are stored and manipulated as symbolic tokens. An exception is the digital elevation map (DEM) data, which is stored in an image-based format. The brief list of geometric elements given below is meant to illustrate the range of data being used – more detailed descriptions of the feature extraction algorithms appear in Section 2.1, or in the referenced papers.

- 2D Elements: line segments, corners, polygons
- 2.5D Elements: line segments, corners, polygons, planes
- 3D Elements: DEM data, lines, curves, corners, surfaces

2.1 2D Geometric Elements

Straight line segments are extracted using the Boldt algorithm [2]. This algorithm hierarchically groups edgels into progressively longer line segments based on proximity and collinearity constraints.

Corner features are created by grouping the 2D line segments according to proximity and an expected angle, α. The expected angle can be computed from an assumption of corner orthogonality in the world and the camera position for any particular view of the site. Each corner has a u, v position within the image as well as two "leg" vectors that help direct the search for higher level features.

Using these low level extracted image features, higher level 2D features are computed as possible groupings that are perceptually consistent. Mid-level collated features are sequences of corners and lines that are grouped together to form *chains*. For a chain to be formed, corners must have legs that are approximately oriented towards one another. Corners cannot be grouped unless there is either a straight line segment between them, or a localized, top-down analysis verifies the existence of a supporting edge in the image data.

High-level 2D polygon hypotheses are formed from closed chains. If a chain cannot be closed because a small number of low-level features are missing, a top-down feature hypothesis is made and the image is searched for the missing feature. Because a highly constrained region of the image is searched using top-down knowledge, a more context-specific feature detector can be used to find the missing feature where one was expected. This technique is especially valuable for detecting features in regions of poor contrast or in regions contaminated with clutter.

Because single collated features can be part of several, possibly conflicting closed polygons, the final set of polygons must be searched for as the best independent set of closed chains. This is done using certainty measures that are maintained throughout the entire grouping process. As each feature is extracted it is assigned a certainty; the final 2D grouping choice is then found as the independent set of closed chains that maximizes the global certainty.

2.2 2.5D Geometric Elements

Features in the 3D scene that are horizontal (perpendicular to gravity) can be represented by the corresponding 2D image element plus an associated scene height z_0. We call features represented in this way '2.5D" geometric elements. It is trivial to convert this representation into a true 3D feature by backprojecting the feature out onto the plane $Z = z_0$ to compute the 3D position of the horizontal image feature.

A set of 2.5D features can be derived directly from 2D features by augmenting each with a height value computed via multi-image matching. The height estimate for each feature is formed by histogramming the set of elevations implied by potential corresponding features within epipolar-constrained search regions across multiple images, the disparity of each potential match voting for a possible height in the scene. After the histogram is created, the Z-value associated with the highest peak in the histogram is associated with the original image feature. If multiple peaks of approximately equal magnitude are present, then multiple Z-values, and thus multiple 2.5D features, can be hypothesized for the same 2D image feature.

For example, we currently compute 2.5D lines from the set of 2D Boldt line segments using this strategy. The endpoints of an image line segment are backprojected into three planes at heights $Z = z_i, i = \{1, 2, 3\}$, where z_1 and z_3 are estimated heights that bound the unknown height z_2. This produces 3 known horizontal 3D line segments L_1,L_2 and L_3. Forward projecting these line segments into another image yields three known image line segments l_1, l_2 and l_3, with the endpoints of l_1 and l_3 bounding a quadrilateral search area in which a corresponding image line must lie. For any candidate line segment match lying within this image area and having approximately the right orientation, a simple linear equation can be solved for the scene height z_0 associated with any candidate line match in the quadrilateral search area. The line height histogram tallies z_0 votes for multiple matches over multiple images and a histogram peak yields the final line height.

2.3 3D Geometric Elements

True 3D line segments can be computed using a multi-image matching process related to the method for computing 2.5D lines. 3D line segments provide information about non-horizontal structure and can be used in grouping structures that do not lie in a single plane. For example, a true 3D line segment might represent the sloping edge of a gabled roof. These line line segments can be grouped according to proximity into 3D corner features. Each 3D corner has a true orientation in the world given by the orientation of the 3D lines from which it was created.

Curves and surfaces are represented parametrically and can either be extracted from or fit to the site data. The 3D position and orientation for arbitrary curves and surfaces can be computed by extracting the feature in a single image and matching over several views. On the other hand, a parametric surface can be fit to existing geometric elements (see Section 4).

The UMass terrain reconstruction system [10] produces a digital elevation map (DEM) consisting of a dense array of 3D elevation estimates and registered orthographic image (ortho-image) from two overlapping images that may have been taken from oblique and widely spaced viewpoints, provided that: (1) The approximate viewing geometry is known, (2) surfaces are textured, and (3) most objects have approximately the same base-to-height ratios (e.g., mountains, hills and boulders have approximately the same base-to-heigh ratios). Large, low contrast surfaces, occluded boundaries, and tall objects with a small base (e.g., flag poles) will produce artifacts in the final DEM and ortho-image.

3 Context-Sensitive Grouping Strategies

3.1 Introduction

The extraction of complex structures from images, such as 3D buildings, is a complex process which must take into account variations in the image appearance of the object(s) (e.g. shadows and reflections), variations in generic classes of objects, and variations arising from changes in viewpoint. Given these inherent difficulties, building extraction cannot be a purely bottom up task but must be approached as an image interpretation problem in which a-priori knowledge and general geometric constraints can be brought to bear. Under this paradigm, generality is achieved by combining large numbers of competent special purpose algorithms under appropriate contextual, geometric, spatial, and temporal constraints derived from a-priori knowledge [6].

This section outlines several grouping/fusion strategies for 3D building detection and reconstruction. These strategies should be viewed as preliminary steps towards the development of special purpose strategies that are intended to provide the next level of abstraction beyond those described in Section 2. In most cases we present a brief description of a general strategy followed by a specific instantiation of that strategy which was used to generate the results in the case studies presented in Section 4.

Some of these strategies have a bottom-up flavor and are meant to be applied to fairly large sets of primitive elements in the absence of strong contextual constraints. For example, the first strategy described below extends the current Ascender 2D polygon grouping strategy to include 3D constraints. It is designed to be applied to relatively large sets of 2D, 2.5D, or 3D lines to search for consistent polygonal hypotheses. Bottom-up strategies may face the problem of a combinatoric explosion of grouping possibilities as the size of the data set increases.

Top-down strategies, on the other hand, control the combinatoric problem by reducing the size of the input set of elements by applying appropriate constraints. The downside of this approach is that some mechanism must be provided for determining when and where the strategy can be applied. For example, some of strategies described below are concerned with fitting parameterized models of several classes of roof structures to digital elevation data. The context sensitive triggering mechanism for these strategies is typically the hypothesis that some form of building is present, which defines spatial limits in the elevation data. Images cues (such as roof shadows) may also limit the class of parameterized models that must be tried.

It is interesting to note that the 2D polygonal extraction algorithm is itself a relatively complex grouping strategy when viewed from the inside. Externally it can be viewed as simply producing a set of 2D polygon hypotheses as input data to other processes. Using this perspective, the grouping strategies offered in the next section can be viewed as preliminary experiments supporting the design of algorithms for extracting more complex and abstract geometric structures, which in turn become the input data for yet another layer of strategies, leading toward systems which can handle a larger variety of generic object classes and objects of increased complexity within generic classes.

3.2 Example Grouping Strategies

Variations on the following example grouping strategies are used in Section 4 to interpret and fuse information in the general context of extracting 3D building models from both optical data and digital elevation maps.

3D feature grouping under 3D spatial and geometric constraints. The various types of feature tokens can be assembled into more complex and coherent structures by applying available constraints. 3D lines and surfaces, as well as combinations of these tokens, can be organized under optimizing strategies for information fusion. A straightforward example that has been implemented is 2.5D polygon line grouping into 3D rooftop polygons. This is a direct extension to the graph-based perceptual organization algorithm used in Ascender for organizing 2D lines and corners into closed 2D polygons [9]. An additional set of consistency checks are introduced into the feature grouping mechanisms to test that compatible lines and corners that are about to be grouped occur at roughly the same elevation in the scene. Individual line heights are combined and propagated into grouped corner, chain, and final polygon features. The results are closed 2D polygons with an associated height value, which is easily converted into a flat 3D roof polygon using the known camera pose information.

Constrained fitting of parametric models. 3D parametric models are fit to a subset of discretely sample elevation data. Spatial constraints, derived from object hypotheses, are used as focus of attention mechanisms to bound the subset of data to be fit and robust procedures are used to determine the fit. To date, the spatial constraints consist of the 2D polygonal rooftop hypotheses derived from visible imagery. These are then mapped onto a digital elevation map created by a stereo algorithm or

IFSAR data. Three types of rooftop models have been built: planar, peaked, and curved (fourth order). Since the DEM data are noisy, least median squares robust techniques are used for surface fitting.

Underconstrained fitting of parametric models. This is a more general version of the previous algorithm in which the subset of the elevation data to be fit is not circumscribed by the object hypothesis. The additional complication in this case is to determine the missing spatial constraints as fitting proceeds. One approach that has been implemented is designed to fit planar roof models to DEM data starting with an area bounded by a partial rooftop boundary hypothesis, for example, a planar surface fit to the area completely contained within a U-shaped hypothesis. The planar fit is extended in the direction of the missing roof edge until the residual error of the planar fit exceeds a threshold value. Other areas of the DEM can be focussed on for fitting planar surfaces by selecting the region between two parallel, overlapping line segments, or the convex triangular area contained between the two line segments forming an L-junction. The resulting plane can extended via planar region growing.

Constructing composite volumetric structures using knowledge. 3D structures are often constructed in pieces, for example the multiple parametric models making up a roof composite. The larger composite volume must then be inferred using heuristic strategies based on general knowledge of the structure. An example of this has been carried out to direct heuristic merging of proximate simple planar roof buildings into complex, multi-level roof structures. Simple flat-roofed models are first clustered into locally adjacent groups, and the roof heights within each group are sorted in decreasing order. Starting from the volumetric block associated with the highest roof polygon and working downward, individual building blocks are extended to and merged with adjacent higher structures. The result, after extrusion to the terrain data, is a complex volumetric building model.

Finding unmodeled structures of surface-volume structures. Once surface fits and reconstruction algorithms have been used to generate 3D building models, unmodelled 3D structures occasionally remain on the roof. (For example, air conditioning units, roof access structures, etc.) These may appear at the limits of resolution of the reconstruction algorithms, and reliable and accurate reconstruction of the fine, detailed structures may not be possible using the original algorithms. However, once a surface has been fit, these structures often appear as spatially coherent groups of outlier points. A strategy for recovering the fine structure based on this observation is described in section 4 (as part of case study 1.)

3.3 Model Selection

A final building model may be a composition of bottom-up and top-down reconstruction strategies as well as many different geometric element types. Knowledge about building properties, or a priori constraints within the building model, should be used to trigger the most appropriate top-down strategies and the choice of appropriate geometric elements for the reconstruction. The wrong choice of reconstruction strategy, or attempt to fit an inappropriate geometric surface, will lead to an incorrect reconstruction or rejection of the building because of large residual error.

Figure 1: Rooftop shape of certain buildings is clearly contained in shadow information.

Currently, building models can be constructed from three different surface elements: planar surfaces, curved surfaces, and peaked roofs. Local image context determines the choice of surface element for each building model. For example, the existence of a centerline within a 2D polygon leads to fitting a peaked roof element. Shadows are also a powerful image feature that can be used to justify the selection of different rooftop elements. Figure 1 shows several buildings from an aerial view. Although a ridge-line is not clear in the peaked roof buildings at the bottom of the image, the shape of the shadows clearly separates them from the buildings at the top of the image that have curved surfaces and require a different reconstruction strategy.

3.4 Fitting Robust Parametric Surfaces to a DEM

Extracting surfaces from the 3D array of data in a digital elevation map (DEM) will be an important component of a variety of information fusion strategies that are being developed. There are many ways that a subset of the DEM might be selected, for example by a 2D polygon that has been identified as a building hypothesis. Using this constraint, previously computed 3D elevation estimates within the polygons can be fit.

Here we outline the procedure to extract primitive surface elements that are used to reconstruct the simple and complex roofs. Since the residuals are not normally distributed due to a variety of complex factors in the formation of the DEM, outliers (large deviations from the correct values) must be assumed to be present. To help design a robust algorithm for fitting surfaces to the types of buildings that are expected, several different roof models are employed for processing regions of the DEM. Three roof models have been implemented so far, a horizontal planar flat roof model, a multi- level flat roof model, and a two-facet peaked roof model.

The elevations data are divided into three categories: (1) good data – data that are normally distributed about the appropriate roof model, (2) outliers that arise from inaccuracies in the roof model, (3) outliers that arise from artifacts in the algorithm that produced the elevations data.

The robust model fitting algorithm has two stages: In the first, the model parameters are estimated by minimizing a cost function that is not sensitive to outliers (provided that outliers account for less than half of data). Stage I processing uses the Downhill Simplex method and a least-median-squared cost function to estimate the model parameters. The residuals computed using this method have the property that outlier residuals are larger than good data residuals.

In the second stage, an iterative process is used to identify residuals and fit the good data to the roof model using an optimal, least squares estimator. At each step in the stage II processing, model parameters are computed for all data with a stage I residual below an adjustable outlier threshold. Starting with an outlier threshold that accepts all data, the threshold is gradually lowered until the chi-squared per degrees-of-freedom stops improving.

Detailed case studies of these techniques are discussed in the following sections. In Section 4.1 a multi-level flat roof building is analyzed, and in Section 4.3 a combination of several peaked roof and single-level flat roofs buildings are reconstructed.

4 Case Studies

4.1 Lockheed-Martin Multi-Level Flat Roof

The following case study describes the reconstruction of a multi-level flat roof building from 2D roof polygons, produced by the UMass Ascender system, and 3D elevation estimates, produced by the UMass terrain reconstruction system. The building is located on the Lockheed-Martin Denver site near the UGV DEMO B test area. The building contains several different levels and is a fairly complex shape. In addition, surface structures clutter the roof tops, and shadows partly occlude several roof sections.

The monocular building detector was run on the test image. A total of nine polygons, shown in Figure 2, were extracted. Note that polygon C was cutoff at the shadow boundary rather than the actual rooftop edge, and the three polygons D, F, and G overlap. These are inherent problems associated with polygon detection, which are addressed in the 3D knowledge-based building composition.

Next, 3D planes were fit to the elevation data within each roof polygon. If the best fit error is larger than a threshold the polygon is rejected. For example, polygon G was rejected and replaced with polygons D and F. Alternatively, because the combined best fit error of polygons D and F is significantly less than the best fit error of G, a two level model consisting of D and F could be accepted.

The final set of planes before and after knowledge-based volumetric composition are shown in Figure 3. Notice that the disjoint planes have been merged into a single building and a separate small structure on the ground (corresponding to the small polygon in the upper right corner of Figure 3).

Clearly, the reconstructed building is greatly improved. However, there still remain valid 3D structures on the rooftop which are not modelled. They are most likely air conditioning and ventilation equipment. These structures were identified and

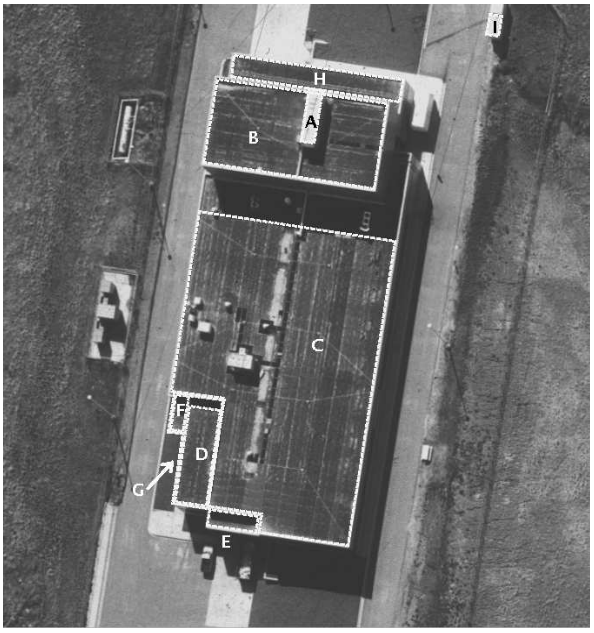

Figure 2: Aerial view of the test site with the detected 2D polygons.

added to the building model using an outlier analysis procedure.

The structural outlier identification procedure generates a series of Boolean masks that identify points on the roof as having qualities consistent with roof top structures. The final selection is then based on the intersection of these masks. For points identified as belonging to a roof top structure, the fitted model elevations are replaced with the original elevation estimates.

Masks are generated based on four selection criteria: (1) Select a data point if the distance from the data point to the fitted plane is greater than the outlier threshold. (2) Because elevation estimates near discontinuities are not reliable, points on a roof polygon are selected if their distance to the boundary is greater than half of the width of the matching correlation window. (3) Because roof top structures usually are built up, points are selected if they were above the roof plane. (4) If a point is not located in shadow.

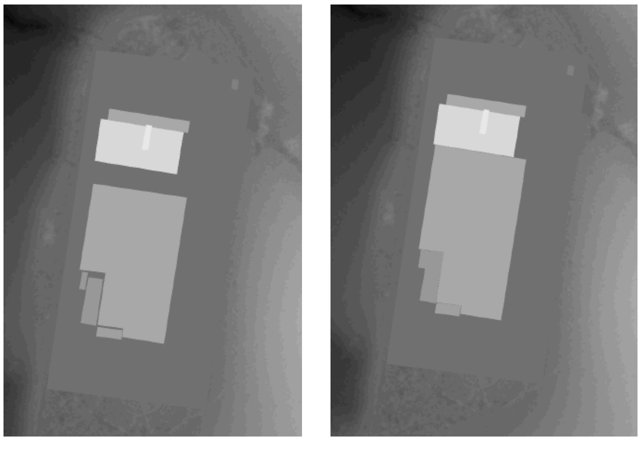

Figure 3: The fit planes before and after knowledge-based volumetric composition.

The histogram of the absolute difference between the fitted plane and the raw elevation data for the main roof is shown in Figure 4. The outlier threshold is 0.57 meters, which rejected 12% of the data. In addition, half the width of the matching window was 14 pixels, and a value of 40 was used for the shadow threshold. To fill in small gaps, the intersection mask was filtered with two passes of a dilation operator. Figure 5 shows these four masks and their intersection for the test building, and a close-up of the main roof section with the identified roof top structures. Of the nine identified objects, seven appear to be real (including the four objects in the shadow of the roof section) and two appear to be false. In addition, two small real objects seem to have been missed. A rendered view of the final reconstruction is shown in Figure 6.

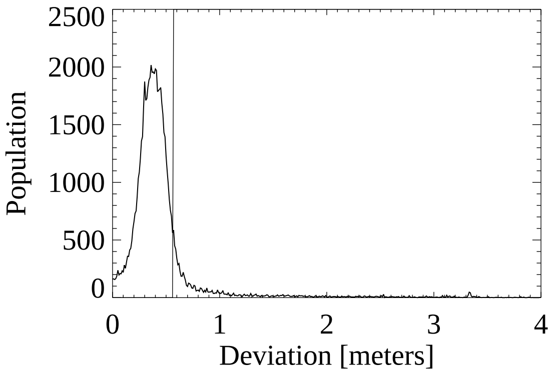

Figure 4: Histogram of the absolute difference between the fitted and raw elevations within the main roof polygon. The outlier threshold (indicated by the vertical line) is 0.57 meters and rejects 12% of the data.

4.2 Ft.Hood Peaked Roofs

As a second case study, a set of buildings located at Ft. Hood Texas, were used for reconstruction. As before, the UMass Ascender system generated the 2D roof polygons and the 3D elevation estimates

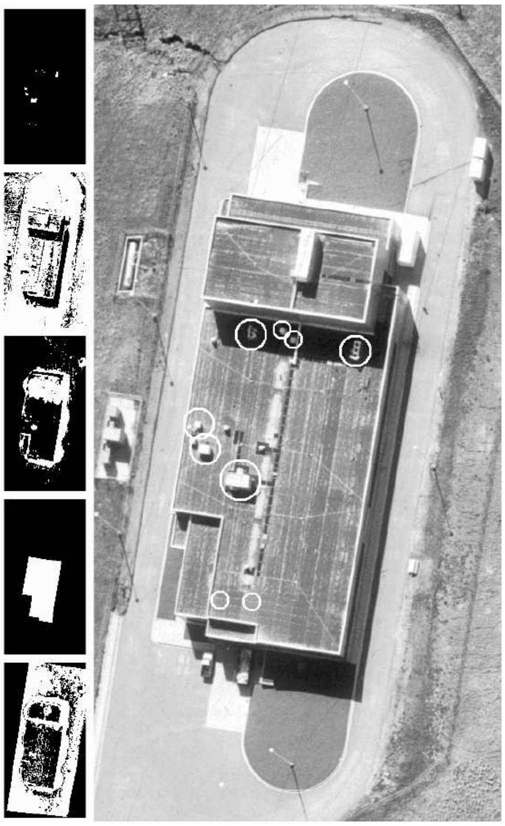

Figure 5: Boolean masks generated during outlier analysis, and detected roof top structures. From top to bottom, along the left edge the Boolean masks are: intersection, shadows, positive differences, main roof, and outliers.

Figure 6: Rendered view of the final reconstruction.

Figure 7: Subimage of the Ft. Hood dataset with roof polygons detected.

were computed from the UMass terrain reconstruction system. An aerial view of the site and the detected roof polygons is shown in Figure 7. The image contained several peaked roof buildings and a large two level flat roof building. Trees, shadows, and the fact that the image is at a much lower resolution than the image of the first case study makes detection and reconstruction of the buildings an interesting task.

The 2D polygon detector was run on the image and the results are shown in Figure 7. Seven polygons were detected. Five of the polygons denote peaked roof buildings while two polygons represent the two sections of the large flat roofed building. The two flat roof polygons were fit to the elevation data using the three dimensional planar roof model as in Section 4.2. To ensure that two pieces of the flat buildings form a coherent single model, a set of rules embodying geometric knowledge of building structures are applied to the fit planes.

A peaked roof model was fit to the elevation data within the five remaining polygons. Currently, the full robust fitting algorithm has not been implemented for peaked roof models. Instead, the initial least-median-squares algorithm brings the model into alignment, determines the best peak angle, position of the ridge line, and height of the peak.

A ground plane was fit to the elevation estimates within a bounding box surrounding the seven polygons. Only elevation points exterior to the polygons are considered for the ground plane fit. For illustration purposes, the plane computed in the bounding box was extended to cover the entire site so that the model fitting results would be clearer. This plane may not actually reflect the actual terrain elevations. Figure 8 shows the site after this reconstruction process.

Figure 8: Six reconstructed buildings from the Ft. Hood scene. Pixels that lay on the ground plane were darkened to highlight the results.

4.3 IFSAR from Kirkland AFB

4.3.1 Introduction

Three-dimensional fusion and grouping strategies can be applied to a combination of Interferometric Synthetic Aperture Radar (IFSAR) data and electro optical (EO) imagery. The following strategies were developed as part of an initial exploration into grouping and fusion of a single optical image and an IFSAR elevation map. Because of several important differences between photogrammetric and IFSAR elevation data, earlier reconstruction strategies had to be adapted to the IFSAR sensor characteristics. Typically, IFSAR DEMs have a higher variance and more dropouts (i.e. no values) than photogrammetrically produced DEMs. In addition, the viewing geometry associated with IFSAR and photogrammetric DEMs vary greatly. IFSAR work best at grazing angles, whereas photogrammetric techniques work best with near nadir views. The grouping strategies discussed below were developed and tested on only a small fraction of the available data and should therefore be considered to be very preliminary.

4.3.2 Strategies for Processing the Sandia Kirkland AFB Data

Figures 9 and 10 show the optical and IFSAR images, respectively, chosen from the Kirkland data set for this study. First, the optical image was registered to the IFSAR DEM by matching building corners and then warping the IFSAR DEM image so that the two images were aligned.

The general reconstruction strategy is similar to that described earlier. The optical images is processed to extract various types of features that can be used in a knowledge-directed control strategy for multi-sensor information fusion. Grouping of lines and analysis of shadow boundaries are clear cues for invoking context-sensitive strategies to effectively process the DEM to extract 3D features in the face of noise that might otherwise obscure the goal, as will be shown via examples.

4.3.3 Fitting planar roof models to IFSAR DEM data.

The 2D building extraction algorithm produced an initial set of 2D polygons that represent building roof hypotheses. These polygons provide focus of attention regions that are likely to be planar surfaces in the fairly noisy IFSAR data. For example, the elevation of rooftop polygon A in Figure 9 is 7.33 meters while the standard deviation (excluding dropouts) of the IFSAR DEM points within the polygon is 13.4 meters! Although the elevation data are noisy, there is a clear separation between roof and ground level elevations. Next, a robust least-median-squared (LMS) fitting technique provides an estimate of the elevation of the roof polygons. The elevation data within the roof polygons then can be replaced with the recovered plane. Assuming that the buildings lay on a planar surface, all

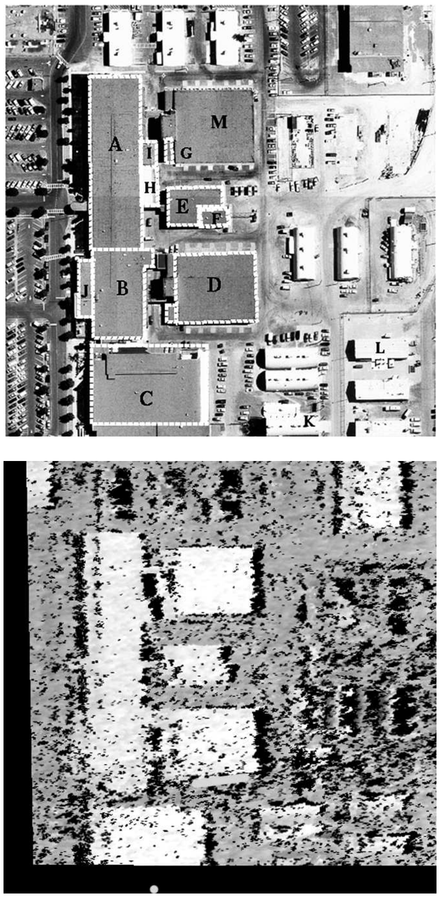

Figure 9: Optical subimage and the corresponding IFSAR data of Sandia/Kirkland dataset. Dotted lines in the optical image are described in the text.

non-building elevation points can be fit to a median ground plane. Finally, the optical image is draped over the refined elevation map to produce a phototextured 3D model, as shown in Figure 10.

Note that some of the buildings in the test region are not detected by this strategy. Examination of the missed building provides clues for developing additional strategies, as well as the means to invoke them. Because line data may not be reliable, it is imperative that several complementary approaches be employed when extracting buildings.

The quality of the IFSAR image depends upon the orientation and position of the sensor with respect to objects within the site. Thus, certain regions of the IFSAR DEM may not contain reliable data. For example, in Figure 9, the Quonset huts to the right of building D are degraded by radar dropouts and shadows. In contrast, the same type of huts to the right of building C are clearly visible in the IFSAR elevation (see Figure ??). This demonstrates the effect of object orientation with respect to the radar.

4.3.4 Region growing.

Consider the building near the center of the image labeled M in Figure 9. All the boundaries lines are not extracted, and one of the 2D roof polygons was not completed. In this case, the polygon detection algorithm failed because the line extraction algorithm was unable to detect roof edges along the top of the building. In contrast, region segmentation strategies within the radar data can find rooftop planes that were missed in the search of the optical image. To extract the missing building (bldg. M), we make use of the following observations:

- In the optical data, only 2 or 3 sides of the building have detected; and a closed building polygon has not been detected. Nevertheless, the number of sides is sufficient to hypothesize a building.
- In the IFSAR DEM there are a significant number of pixels above the surrounding ground plane. In this case, the roof plane is extracted by selecting pixels associated with the distinctive cluster and fitting a planar model to the extracted elevations.
- In the SAR images some building edges and corners are clearly visible. These cues may compliment the optically detected building features and can be used to improve the accuracy of IFSAR processing by localizing and filling in missing building boundaries.

The missing sides were hypothesized using geometric constraints and symmetry. In this example, a bounding box within the IFSAR data was computed using the extracted optical edges of building M. A median plane was fit to the region and the improved site model is shown in Figure 10.

4.4 Determination of building models from optical cues.

Other optical cues, such as shadow information, can help determine the form of the parametric model. Consider, for example, the Quonset huts in area K (see Figure 9). The shadows clearly imply that the buildings have curved roof. The structure of the two Quonset huts were found by first segmenting the

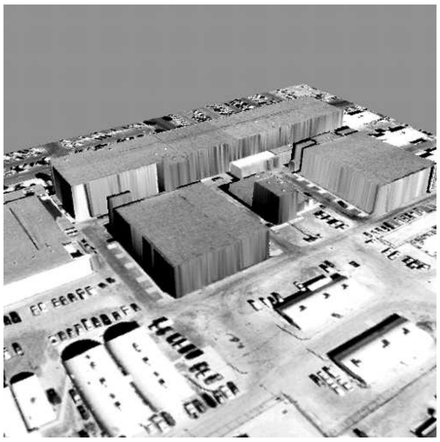

Figure 10: Improved site model after extraction of building M (see Figure 9 using focussed detection mechanisms

footprint of the building using the EO image, then smoothing the IFSAR DEM with a median filter, and finally fitting the smoothed elevation estimates to a fourth order polynomial. Figure 11 shows the EO image draped over the recovered shape of one of the Quonset huts. The shadow of similar shaped building in the optical image allows the peaked roof buildings to be easily and clearly distinguished from the Quonset hut buildings as shown in Figure 1. This allows invocation of a peaked roof parametric model for processing the DEM.

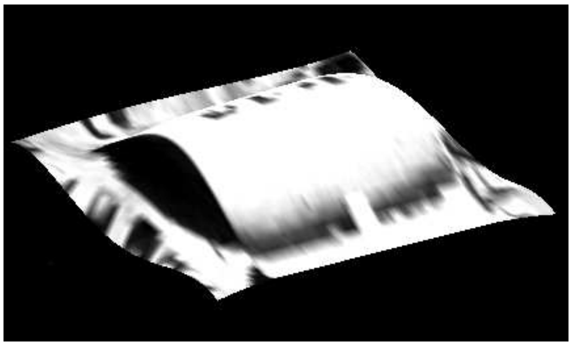

Figure 11: Recovered surface of Quonset hut

These three strategies represent an exemplar of a general approach that makes use of multiple image cues, possibly obtained from the different sources of data, to select the appropriate IU strategy and object model. These results demonstrate a reasonably straightforward extension of the grouping strategies developed in the UMass RADIUS effort. They imply the feasibility of the development of a set of robust complementary strategies to infer the presence and extract 3D models of various objects, including buildings, parking lots, vehicles, etc.

4.5 2.5D Line Grouping

The graph-based perceptual organization algorithm used in Ascender for organizing lines and corners into closed 2D polygons [9] has been modified to handle 2.5D lines (see Section 4.5). An additional set of 3D consistency checks have been introduced to ensure that compatible lines and corners are roughly at the same elevation in the scene. Individual line heights are combined and propagated into grouped corner, chain, and polygon hypotheses. The results are closed 2D polygons with associated elevation values, which are easily converted into flat 3D roof polygons using the known camera projection equations. The benefit of the 2.5D approach to roof polygon detection is that image line segments caused by shadows and ground-level features are automatically ignored, and there is less chance of overgrouping multiple roof levels into a single polygon hypothesis containing edges that actually occur at different elevations in the scene (Figure 12).

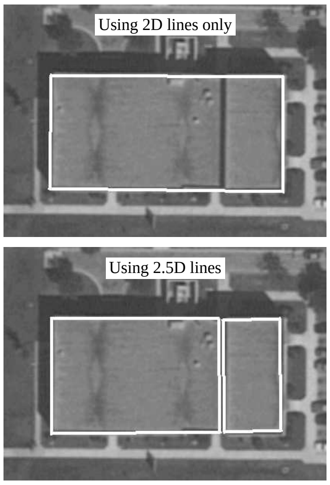

Figure 12: Using 2.5D lines in the grouping process helps disambiguate multi-level building roofs (note the building shadow, which shows two distinct roof levels). The Z-coordinates of vertices on the left and right 2.5D polygon hypotheses are 260.32 and 261.66 meters, respectively, as compared with ground truth Z-values of 260.65 and 262.31.

5 Conclusions and Future Development

This paper demonstrates the utility of data fusion when applied to the problem of site model reconstruction. We combined the results from hierarchical image matching, feature based building detection, robust plane fitting, and heuristic assemble algorithms to form an accurate, robust site model reconstruction system. In the future, the techniques described here will be extended to more complex buildings, including gabled and curved roofs, by fitting the elevation data to a wider variety of geometric models. An additional goal is to use the partially closed chains as focus of attention mechanisms and to explore approaches for recovering surface structure between arbitrary sequences of corners and lines.

Overall, we expect to continue the investigation of plausible strategies for grouping generic elements into complex structures and for simultaneously fusing information from multiple sources into coherent models. The results shown here are encouraging. The accuracy of the final reconstructions can be observed from the visually consistent renderings. Through the careful combination of primitive elements and special purpose strategies, we have the beginnings of an automatic, accurate, and functional system.

References

[1] E. Baltsavias, S. Mason, and D. Stallman. "Use of DTMs/DSMs and Orthoimages to support Building Extraction", *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, April 1995.

[2] M.Boldt, R.Weiss and E.Riseman, "Token-Based Extraction of Straight Lines," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 19(6), 1989, pp. 1581–1594.

[3] 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.

[4] R. T. Collins, A. R. Hanson, E. M. Riseman, C. O. Jaynes, F. Stolle, X. Wang, Y-Q. Cheng. "UMass Progress in 3D Building Model Acquisition.", *Arpa Image Understanding Workshop,* Palm Springs, CA, Feb. 1996.

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

[6] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. "The Schema System", *IJCV*, 1989. vol. 2(3), pp. 209-250.

[7] M. Herman and T. Kanade. "3D Mosaic Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images". *Proc. ARPA Image Understanding Workshop*, 1994.

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

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

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