Site Model Acquisition and Extension from Aerial Images *

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

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

Abstract

A system has been developed to acquire, extend and refine 3D geometric site models from aerial imagery. This system hypothesize potential building roofs in an image, automatically locates supporting geometric evidence in other images, and determines the precise shape and position of the new buildings via multi-image triangulation. Model-to-image registration techniques are applied to align new, incoming images against the site model. Model extension and refinement procedures are then performed to add previously unseen buildings and to improve the geometric accuracy of the existing 3D building models.

1 Introduction

Acquisition of 3D geometric site models from aerial imagery is currently the subject of an intense research effort, sparked in part by the ARPA/ORD RADIUS project [3, 4, 5, 8]. We have developed a set of image understanding modules to acquire, extend and refine 3D volumetric building models, and to provide a digital elevation map of the surrounding terrain. System features include model-directed processing, rigorous camera geometry, and fusion of information across multiple images for increased accuracy and reliability.

Site *model acquisition* involves processing a set of images to detect buildings and to determine their 3D shape and placement in the scene. The site models produced have obvious applications in areas such as surveying, surveillance and automated cartography. For example, acquired site models can be used for model-to-image registration of incoming images, thus allowing the model to be automatically overlaid on each image as an aid to visual change detection and verification of expected scene features. Two other important site modeling tasks are *model extension* – updating the geometric site model by adding or removing buildings based on the results of change detection – and *model refinement* – iteratively refining the shape, placement and surface structure of building models as more views become available. Model extension and refinement are ongoing processes that are repeated whenever new images become available, each updated model becoming the current site model for the next iteration. Thus, over time, the site model is steadily improved to become more complete and more accurate.

This paper focuses on algorithms for automated building model acquisition and extension. To maintain a tractable goal for our research efforts, we have chosen initially to focus on a single generic class of building models, namely flat-roofed, rectilinear structures. The simplest example of this class is a rectangular box-shape; however other examples include L-shapes, U-shapes, and indeed any arbitrary building shape such that pairs of adjacent roof edges are perpendicular and lie in a horizontal plane. Acquisition of an initial site model is treated in Section 2, followed by model extension in Section 3. This paper concludes with a brief summary and a statement of future work.

2 Site Model Acquisition

The building model acquisition process involves several subtasks: 1) line segment extraction, 2) building detection, 3) multi-image epipolar matching, 4) constrained, multi-image triangulation, and 5) projective intensity mapping. These algorithms will be presented by way of an experimental case study using images J1–J8 of the RADIUS model board 1 data set. Figure 1 shows a sample image from the data set. Each image contains approximately 1320×1035 pixels, with about 11 bits of gray level information per pixel. Unmodeled geometric and photometric distortions have been added to each image to simulate actual operating conditions. The scene is a 1:500 inch scale model of an industrial site. Ground truth measurements are available for roughly 110 points scattered throughout the model, which were used to determine the exterior orientation for each image. The residual resection error for each image is in the 2–3 pixel range, representing the level of unmodeled geometric distortion present in each image. This corresponds to a backprojection error of roughly 3–4.5 feet in (simulated) object space. This is a significant amount of error that presents a good test of system robustness.

*This work was funded by the RADIUS project under ARPA/Army TEC contract number DACA76-92-C-0041 and by ARPA/TACOM contract DAAE07-91-C-R035.

Figure 1: A sample image from Model Board 1

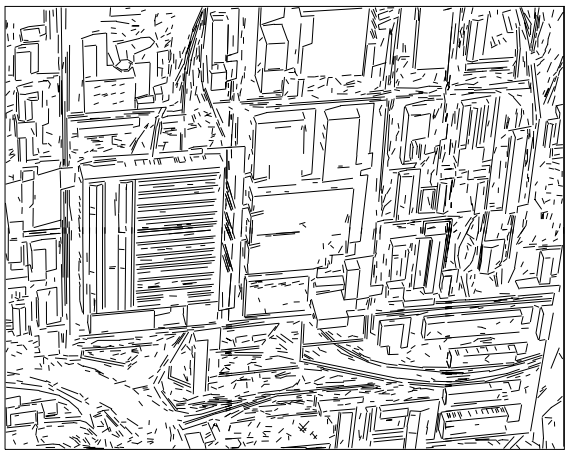

Figure 2: Line segments extracted from Figure 1

2.1 Line Segment Extraction

To help bridge the huge representational gap between pixels and site models, feature extraction routines are applied to produce symbolic, geometric representations of potentially important image features. The algorithms for acquiring building models rely on extracted straight line segments [2]. At the heart of the Boldt algorithm is a hierarchical grouping system inspired by the Gestalt laws of perceptual organization. Zero-crossings of the Laplacian of the intensity image provide an initial set of local intensity edges. Hierarchical grouping then proceeds iteratively; at each iteration edge pairs are linked and replaced by a single longer edge if their end points are close and their orientation and contrast values are similar. Filtering to keep line segments with a length of at least 10 pixels and a contrast of at least 15 gray levels produced roughly 2800 line segments per image. Figure 2 shows a representative set of lines extracted from the image shown in Figure 1.

2.2 Building Detection

The goal of automated building detection is to roughly delineate building boundaries that will later be verified in other images by epipolar feature matching and triangulated to create 3D geometric building models. The building detection algorithm is based on finding image polygons corresponding to the boundaries of flat, rectilinear rooftops in the scene [6]. Briefly, possible roof corners are identified by line intersections. Perceptually compatible corner pairs are linked with surrounding line data, entered into a feature-relation graph, and weighted according to the amount of support they receive from the low-level image data. Potential building roof polygons appear as cycles in the graph; virtual corner features may be hypothesized to complete a cycle, if necessary. Rooftops are finally extracted by partitioning the feature-relation graph into a set of maximally weighted, independent cycles representing closed, high-confidence building roofs.

Figure 3 shows the results of building detection on image J3 of the model board 1 data set. The roof detector generated 40 polygonal rooftop hypotheses. Most of the hypothesized roofs are rectangular, but six are L-shaped. Note that the overall performance is quite good for buildings entirely in view. Most of the major roof boundaries in the scene have been extracted, and in the central cluster of buildings (see area **A** in Fig. 3) the segmentation is nearly perfect.

There were some false positives, i.e. polygons extracted that do not in fact delineate the boundaries of a roof. The most obvious example is the set of overlapping polygonal rooftops detected over the large building with many parallel roof vents (area **B**) Note that the correct outer outline of this building roof is detected, however. There are also some false negatives, which are buildings that should have been detected, but weren't. The most prevalent example of this is a set of buildings (area **C**) that are only partially in view at the edge of the image. Label **D** marks a false negative that is in full view. Two adjacent corners in the rooftop polygon were missed by the corner extraction algorithm. It should be stressed that even though a single image was used here for bottom-up hypotheses, buildings that are not extracted in one image will often be found easily in other images with different viewpoints and sun angles.

There are several cases that cannot be strictly classified as false positives or false negatives. Several split-level buildings appearing along the right edge of the image (area **E**) are outlined with single polygons rather than with one polygon per roof level. Some peaked roof buildings were also outlined, even though they do not conform to the generic assumptions underlying the system.

2.3 Multi-image Epipolar Matching

After detecting a potential rooftop in one image, corroborating geometric evidence is sought in other images (often taken from widely different viewpoints) via epipolar feature matching. Rooftop polygons are matched by searching for each component line segment separately and then fusing the results. For each polygon segment from one image, an epipolar search

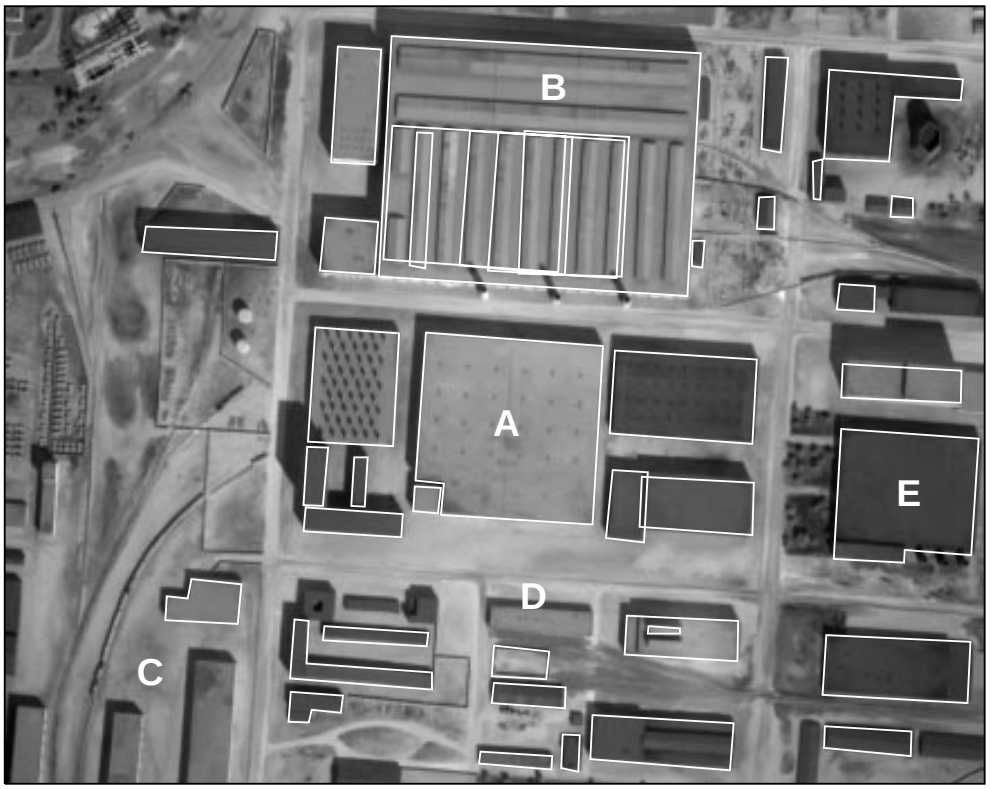

Figure 3: Roof hypotheses extracted from image J3. Alphabetic labels are referred to in the text.

area is formed in each of the other images, based on the known camera transformations and the assumption that the roof is flat. This quadrilateral search area is scanned for possible matching line segments, each potential match implying a different roof height in the scene. Results from each line search are combined in a 1-dimensional histogram, each match voting for a particular roof height, weighted by compatibility of the match in terms of expected line segment orientation and length. A single global histogram accumulates height votes from multiple images, and for multiple edges in a rooftop polygon. After all votes have been tallied, the histogram bucket containing the most votes yields an estimate of the roof height in the scene and a set of correspondences between rooftop edges and image line segments from multiple views.

Epipolar matching of a rooftop hypothesis is considered to have failed when, for any edge in the rooftop polygon, no line segment correspondences are found in any image. Based on this criterion, epipolar matching failed on eight rooftop polygons. Six were either peaked or multi-layer roofs that did not fit the generic flat-roofed building assumption, and the other two were building fragments with some sides shorter than the minimum length threshold on the line segment data. At this stage, six incorrect building hypotheses were removed by hand; detecting and removing such mistakes automatically is being actively investigated.

2.4 Multi-image Line Triangulation

Multi-image triangulation is performed to determine the precise size, shape, and position of a building in the local 3D site coordinate system. A nonlinear estimation algorithm has been developed for simultaneous multi-image, multi-line triangulation of 3D line structures. Object-space constraints are imposed for more reliable results. This algorithm is used for triangulating 3D rooftop polygons from the line segment correspondences determined by epipolar feature matching. Outlines of the final set of triangulated rooftops are shown in Figure 4.

The parameters estimated for each rooftop edge are the Plücker coordinates of the algebraic 3D line coinciding with the edge - specific points of interest, like vertices of the rooftop polygon, are computed as the intersections of these infinite algebraic lines. Plücker coordinates are a way of embedding the 4-dimensional manifold of 3D lines into R^6. Although the Plücker representation requires 6 parameters to be estimated for each line rather than 4, it simplifies the representation of geometric constraints between lines. For the generic flat-roofed rectilinear building class being considered here, we specify a set of constraints to ensure that pairs of adjacent lines in a traversal around the polygon are perpendicular, that all lines are coplanar, and that all lines are perpendicular to the Z-axis of the local site coordinate system. An iterative, nonlinear least-squares procedure determines the Plücker

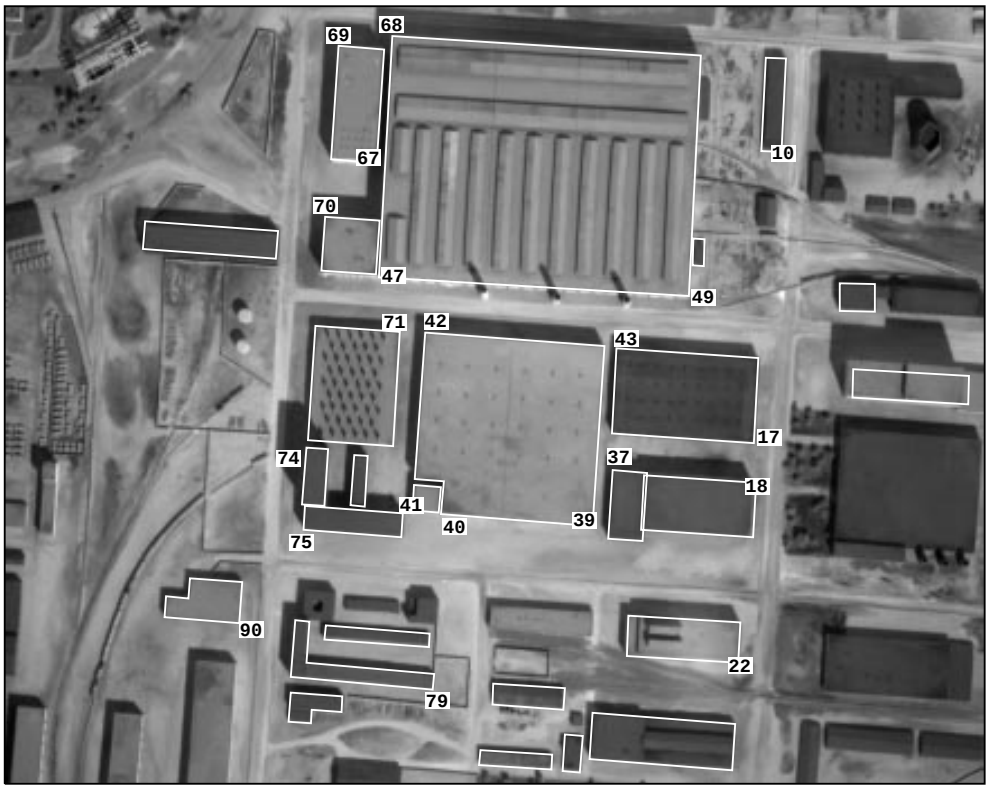

Figure 4: Reprojection of 3D triangulated rooftops back into image J3 (compare with Figure 3).

coordinates for all lines simultaneously such that all the object-level constraints are satisfied and an objective "fit" function is minimized that measures how well each projected algebraic line aligns with the 2D image segments that correspond to it.

After triangulation, each 3D rooftop polygon is extruded down to the ground to form a volumetric model. For the Model Board 1 site, the ground was represented as a horizontal plane with Z-coordinate value determined from the ground truth measurements. More generally, the system will soon be using digital terrain maps produced by the UMass Terrain Reconstruction System[9].

To evaluate the 3D accuracy of the triangulated building polygons, 21 roof vertices were identified where ground truth measurements are known (numbered vertices in Figure 4). The average Euclidean distance between triangulated polygon vertices and their ground truth locations is 4.31 feet, which is reasonable given the level of geometric distortion present in the images. The average horizontal distance error is 3.76 feet, while the average vertical error is only 1.61 feet. This is understandable, since all observed rooftop lines are considered simultaneously when estimating the building height (vertical position), whereas the horizontal position of a rooftop vertex is primarily affected only by its two adjacent edges.

2.5 Projective Intensity Mapping

Backprojection of image intensities onto polygonal building model faces enhances their visual realism and provides a convenient storage mechanism for later symbolic extraction of detailed surface structure. Planar projective transformations provide a locally valid mathematical description of how surface structure from a planar building facet maps into an image. By inverting this transformation using known building position and camera transformations, intensity information from each image is backprojected to "paint" the walls and roof of the building model. Since multiple images are used, intensity information from all faces is available, even though they are not all visible from any single view (see Figure 5). The resulting intensity mapped site model can then be rendered to predict how the scene will appear from a new view, and on high-end workstations realistic real-time "flythroughs" are achievable.

3 Site Model Extension

The goal of site model extension is to find unmodeled buildings in new images and add them into the site model database. The main difference between model extension and model acquisition is that now the camera pose for each image can be determined via model-to-image registration. Our approach to model-to-image registration involves two components: *model matching* and *pose determination*.

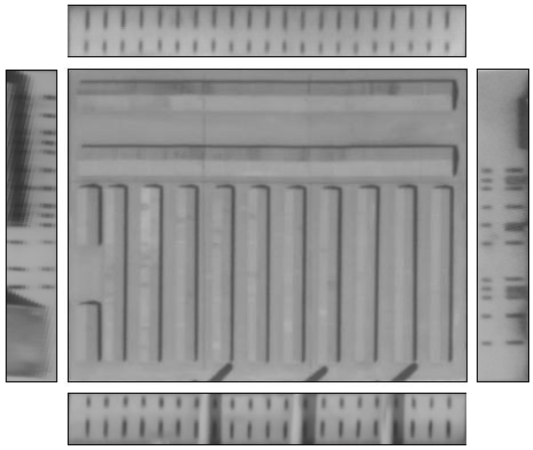

Figure 5: Intensity map information is stored with the planar facets of a building model.

The goal of **model matching** is to find the correspondence between 3D features in a site model and 2D features that have been extracted from an image; in this case determining correspondences between edges in a 3D building wireframe and 2D extracted line segments from the image. The model matching algorithm described in [1] is being used. Based on a *local search* approach to combinatorial optimization, this algorithm searches the discrete space of correspondence mappings between model and image features for one that minimizes a match error function. The match error depends upon how well the projected model geometrically aligns with the data, as well as how much of the model is accounted for by the data. The result of model matching is a set of correspondences between model edges and image line segments, and an estimate of the transformation that brings the projected model into the best possible geometric alignment with the underlying image data.

The second aspect of model-to-image registration is precise **pose determination**. It is important to note that since model-to-image correspondences are being found automatically, the pose determination routine needs to take into account the possibility of mistakes or *outliers* in the set of correspondences found. The robust pose estimation procedure described in [7] is being used. At the heart of this code is an iterative, weighted least-squares algorithm for computing pose from a set of correspondences that are assumed to be free from outliers. The pose parameters are found by minimizing an objective function that measures how closely projected model features fall to their corresponding image features. Since it is well known that least squares optimization techniques can fail catastrophically when outliers are present in the data, this basic pose algorithm is embedded inside a least median squares (LMS) procedure that repeatedly samples subsets of correspondences to find one devoid of outliers. LMS is robust over data sets containing up to 50% outliers. The final results of pose determination are a set of camera pose parameters and a covariance matrix that estimates the accuracy of the solution.

3.1 Model Extension Example

The model extension process involves registering a current geometric site model with a new image, and then focusing on unmodeled areas to recover previously unmodeled buildings. This process is illustrated using the partial site model constructed in Section 2, and image J8 from the Radius Model Board 1 dataset.

Results of model-to-image registration of image J8 with the partial site model can be seen in Figure 6, which shows projected building rooftops from the site model (thin) overlaid on the image. Image areas containing buildings already in the site model were masked off, and the building rooftop detector was run on the unmodeled areas. The multi-image epipolar matching and constrained multi-image triangulation procedures from Section 2 were then applied to verify the hypotheses and construct 3D volumetric building models. These were added to the site model database, to produce the extended model shown in Figure 6 (thick lines). The main reason for failure among building hypotheses that were not verified was that they represented buildings located at the periphery of the site, in an area which is not visible in very many of the eight views. If more images were used with greater site coverage, more of these buildings would have been included in the site model.

4 Summary and Future Work

A set of IU algorithms for automated site model acquisition and extension have been presented. The algorithms currently assume a generic class of flat roofed, rectilinear buildings. To acquire a new site model, an automated building detector is run on one image to hypothesize potential building rooftops. Supporting evidence is located in other images via epipolar line segment matching, and the precise 3D shape and location of each building is determine by multi-image triangulation. Projective mapping of image intensity information onto these polyhedral building models results in a realistic site model that can be rendered using virtual "fly-through" graphics. To perform model extension, the acquired site model is registered to a new image, and model acquisition procedures are focused on previously unmodeled areas. In an operational scenario, this process would be repeated as new images become available, gradually accumulating evidence over time to make the site model database more complete and more accurate.

Several avenues for system improvement are open. One high priority is to add capabilities for detecting and triangulating peaked roof buildings. Another significant improvement would be extending the epipolar matching and triangulation portions of the system to analyze why a particular building roof hypothesis failed to be verified. There are many cases where the rooftop detector has outlined split-level buildings with a single roof polygon; automatic detection of these situations, followed by splitting of the rooftop hypothesis into two separate hypotheses, would result in an improvement in system performance.

These symbolic building extraction procedures will soon be combined with a correlation-based terrain extraction system [9]. The two techniques clearly com-

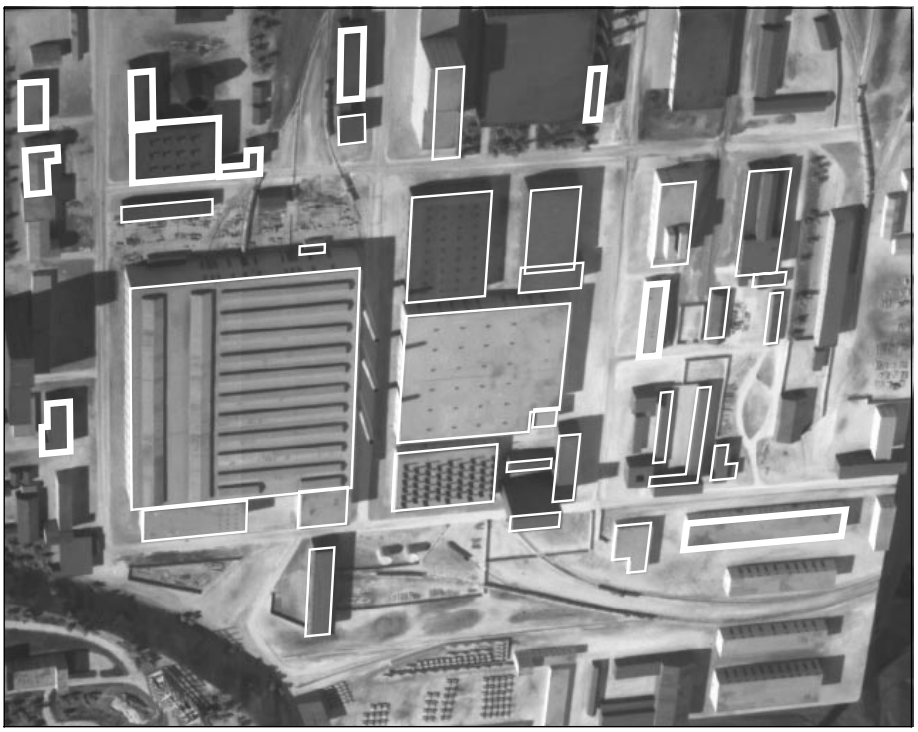

Figure 6: Updated site model projected onto image J8.

plement each other: the terrain extraction system can determine a digital elevation map upon which the volumetric building models rest, and the symbolic building extraction procedures can identify building occlusion boundaries where correlation-based terrain recovery is expected to behave poorly. A tighter coupling of the two systems, where an initial digital elevation map is used to focus attention on distinctive humps that may be buildings, or where correlation-based reconstruction techniques are applied to building rooftop regions to identify fine surface structure like roof vents and air conditioner units, may also be investigated.

Acknowledgements

We would like to acknowledge the technical and administrative support of Robert Heller, Jonathan Lim, Janet Turnbull, Laurie Waskiewicz and Fred Weiss.

References

[1] J. Beveridge and E. Riseman, "Hybrid Weak-Perspective and Full-Perspective Matching," *Proceedings IEEE Computer Vision and Pattern Recognition*, Champaign, IL, 1992, pp. 432–438.

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

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

[4] D. Gerson, "RADIUS : The Government Viewpoint," *Proceedings DARPA Image Understanding Workshop,* San Diego, CA, January 1992, pp. 173–175.

[5] A. Huertas, C. Lin and R. Nevatia, "Detection of Buildings from Monocular Views of Aerial Scenes using Perceptual Grouping and Shadows," *Proceedings ARPA Image Understanding Workshop,* Washington, DC, April 1993, pp. 253–260.

[6] C. Jaynes, F. Stolle and R. Collins, "Task Driven Perceptual Organization for Extraction of Rooftop Polygons," *Proceedings ARPA Image Understanding Workshop,* Monterey, CA, Nov. 1994, pp. 359–365.

[7] R. Kumar and A. Hanson, Robust Methods for Estimating Pose and Sensitivity Analysis," *CVGIP: Image Understanding,* Vol. 60, No. 3, November 1994, pp. 313–342.

[8] M. Roux and D. McKeown, "Feature Matching for Building Extraction from Multiple Views," *Proceedings ARPA Image Understanding Workshop,* Monterey, CA, November 1994, pp. 331–349.

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