Recovery of Building Structure From IFSAR-derived Elevation Maps

K.B. Hoepfner
CMPSCI Technical Report 99-16
July 1999

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

Recovery of Building Structure From IFSAR-derived Elevation Maps

K.B. Hoepfner
E-mail: hoepfner@cs.umass.edu

Abstract

A novel approach to extracting building rooftops from high resolution Interferometric Synthetic Aperture Radar (IFSAR) images is presented here. The presence of noise, missing data and poorly understood radar artifacts in such images necessitates the use of robust and context sensitive techniques. The algorithm presented here exploits knowledge about the geometric structure of buildings and how this geometry interacts with the sensor.

Rooftops are extracted in two stages. In the first, a building's back edge is located by way of the shadow it casts in the image. Once the back edge of a building has been found, its rooftop is extracted through region growing. The region's growth begins at this back edge, and proceeds along the building's boundary. Once growth has terminated, a rectangle is fit to the rooftop region so grown.

1 Introduction

In recent years, IFSAR-derived digital elevation maps (DEMs) have been used in site reconstruction tasks. SAR interferometry has several advantages over the traditional means of generating DEMs, such as stereo photography or the use of laser altimeters. For instance, optical images can only be acquired during the day and under favorable weather conditions. SAR interferometry, on the other hand, is invariant with respect to the weather, and can be used night or day. The IFSAR sensor can also operate at greater altitudes than most laser scanners.

There are several methods of generating IFSAR data, but we shall only consider the two-antenna, single-pass case here. This means that a single aircraft with two antennae, separated by some known baseline, collects all the data from the scene in a single pass. The phase difference between the two returns (one per antenna) generated by a target on the ground is used to determine that target's 3D position [1]. The Kirtland Air Force Base and MOUT data sets were collected in this manner.

IFSAR-derived DEMs are inherently noisy and often have a significant amount of data missing from them. As an example of how inaccuracies arise in the elevation data, consider the effects of layover on the front edge of a building. Layover occurs whenever two or more points are at the same distance from the sensor [2]. In this case, a point P along the front edge of a building will be at the same range as some point G on the ground (see Figure 1, left). Because of this, the elevation measured for P will be the average of P's actual height and the height of the ground

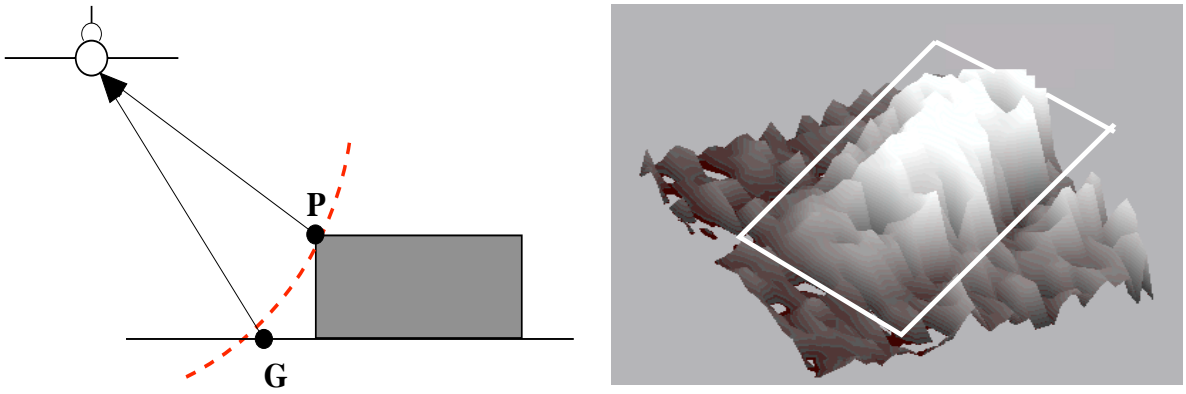


Figure 1: Left: Point G on the ground is at the same range as point P on the rooftop. Right: Height map of a building. The building's boundary is shown in white. The darker values at the building's front edge indicates that it is at a lower elevation than the rest of the rooftop.

point G. This phenomenon gives the front edge of a building a "crumbled" appearance (see Figure 1, right).

Given the inaccurate and incomplete nature of IFSAR-derived DEMs, much of the previous work done on extracting buildings from other types of DEMs - such as those derived from a stereo pair of optical images - may not be applicable here. This would include, for instance, systems that use parametric models to recognize buildings in the scene [3]. Such systems may not recognize a building after it has been imaged by the IFSAR sensor. That is, the sensor may distort a building's height map in ways the models cannot account for. For instance, the crumbled front edge of a building may make it difficult to fit a stored model to that building. As such, the models employed by the system would need additional parameters to account for the effects of layover. Given that the specific effects of layover are dependent on factors that cannot be anticipated - such as the material properties of the surrounding terrain - this may not be a viable solution. Model-based target detection in SAR images, however, has met with some success [4].

As stated earlier, IFSAR-derived DEMs will have data missing from them. Points for which no return was measured are referred to as "drop-outs". Points are dropped-out of the IFSAR image for several different reasons. For instance, a specular target, such as a calm body of water or piece of metal siding, will not have a return measured for it if its surface normal does not point towards the sensor. Point are also dropped-out when the slant range image is converted into a grid of elevation values. It is this orthorectification process that creates "layover holes", which can be found near the front edge of a building [5]. A point can also be dropped-out of

the image because of an occluding object. For instance, a building's rooftop will occlude the terrain behind it. Because of this, a building will cast a shadow in the image (see Figure 2, right). These shadows manifest themselves as large regions of drop-outs in the image, and can be used to detect the presence of buildings in the scene.

This paper presents an algorithm for extracting buildings from an IFSAR DEM. This algorithm operates in two stages. First, the back edges of all of the buildings in the scene are located. These edges are identifiable because of the shadows they cast in the image (see Figure 2). Specifically, points that belong to the back edge of a building can be identified by their proximity to a large region of drop-outs. Once the back edge of a building has been found, its rooftop is extracted through region growing. The region's growth begins at this back edge, and proceeds along the building's boundary. A point is added to the growing region only if it belongs to the building's rooftop. This determination is made by comparing the point's height to a threshold found in an elevation histogram of its neighborhood (ie. through adaptive thresholding). If the point's elevation exceeds that threshold, it is considered to be part of the rooftop and is added to the growing region. Growth terminates once the region encompasses the building's entire rooftop. This algorithm is only suitable for use on buildings with rectilinear boundaries.

Section 2 of this paper details the back edge detection process. Section 3 describes how a building's rooftop is extracted once its back edge is found. The results of applying this algorithm to the Kirtland AFB and MOUT scenes are shown in Section 4.

2 Back Edge Detection

2.1 Properties of a Back Edge

The back edges of a building are along those walls facing away from the sensor. The rooftop of the building occludes the ground adjacent to a back edge from the sensor, causing no return to be measured for that portion of the surrounding terrain. Therefore, points belonging to a building's back edge can be identified by the shadows they cast in the image (see Figure 2). These shadows extend outward from the back edge in the direction of the sensor, where the direction of the sensor is the 2D projection of the axis perpendicular to the flight path (see Figure 2, left). Since the occluded area is part of the terrain surrounding the building, we make the assumption that the shadows cast by a back edge will terminate at some point on the ground. This assumption is reasonable in contexts where the buildings are not too closely spaced or surrounded by trees and other obstructions. Thus, a back edgel (which is part of the building's rooftop) will have an elevation greater than that of the point at the terminating end of its shadow (which is part of the ground). This information allows the formulation of two different constraints that any point E must satisfy before being labeled aback edgel :

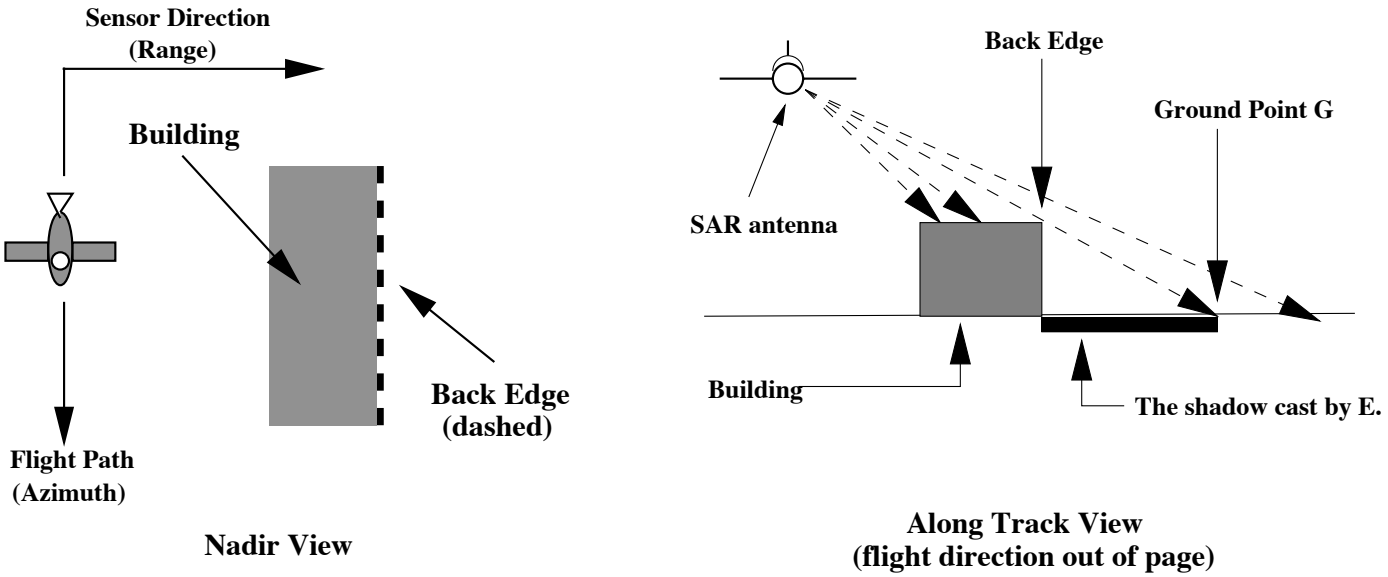


Figure 2: Geometry of SAR data acquisition. The shadow cast by a building's back edge extends from back edgel E to a point G belonging to the surrounding terrain.

1. E must lie on the border between a rooftop and the shadow it casts.

2. There must be a shadow extending from E in the direction of the sensor that terminates at some point G belonging to the surrounding terrain. E must have an elevation greater than that of the surrounding terrain as represented by the point G. The height disparity dH between E and G must be greater than or equal to the minimum height expected of a building (taken here as 3.5 meters).

2.2 Characterizing Points on a Building's Back Edge

The process of finding back edgels begins by identifying those points P in the image that satisfy the first constraint. Because it is not known a priori which points belong to a building's rooftop, this condition must be approximated. For instance, we could require that such a point border a shadowed region (ie. a region in the image for which no returns have been measured). All back edgels will have this property because back edges cast shadows in the image.

A more discriminating approximation requires that one must be able to draw a line through P that divides its local neighborhood into an *occluded side* and a *rooftop side* : all points to one side of the line (the occluded side) should be drop-outs, while all points to the other side (the

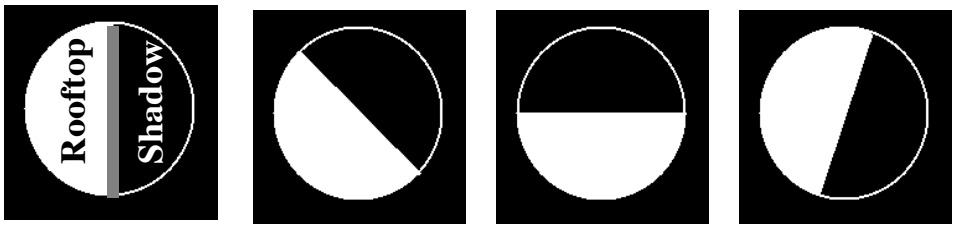


Figure 3: Binary masks at varying orientations. These masks are used to determine if a point borders a shadowed region. The hypothesized border element separating shadow from rooftop is shown in grey.

rooftop side) should have a return measured for them. This dividing line represents a segment of the hypothetical back edge to which P belongs. If P is indeed a back edgel, then, in theory, such a line must exist : it runs between the shadow cast by the building (ie. the occluded side of the dividing line) and the building itself (the rooftop side of the dividing line). Due to the influence of noise, however, the conditions stated above - namely that only dropouts can lie to one side of the edge and only visible points to the other - must be relaxed.

Points that satisfy the above constraint, henceforth known as *shadow edges*, can be identified by applying a series of binary masks to every point in the image for which a return was measured. Each mask M is a disc with a radius of k pixels ($k = 4$ here) and represents the neighborhood of a point on or near a shadow/rooftop border. The dark side of the mask represents the shadow cast by the building's back edge, while the bright side represents the building's rooftop near that back edge. Examples of these masks can found in Figure 3.

The *orientation* of a mask points into the occluded, or shadowed, side of the mask. The dividing line (which passes through the mask's center) has an orientation perpendicular to that of the mask's. For example, a binary mask with an orientation of zero has a dividing line that passes through the mask's origin at an angle of 90 degrees. This dividing line represents the hypothetical back edge that passes through the mask's center. All points to the right of that line belong to the occluded side of the mask, while all points to the left of that line belong to the visible, or rooftop, side of the mask (see Figure 3). The masks have orientations from 0 to 2π, spaced at 10 degree intervals. This gives us a total of 36 different masks.

Each time a mask is applied to a point P in the image, a disc-shaped window of pixels (with a radius of 4) centered at that point is compared to the mask to generate a match score. One way to compute a match score is to cross-correlate the mask with P's neighborhood. However, this metric is inappropriate given that the mask is binary (ie. shadow or rooftop). A better approach is to count the number of mismatches S_M^P between the mask M and P's neighborhood.

Mismatches occur whenever :

- there is a return for a point in the building's shadow (ie. on the dark side of the mask), as occluded points cannot register a return to the sensor, or
- a point falling into the region reserved for the building's rooftop (ie. the bright side of the mask) is a drop-out, since presumably the point is not occluded and should therefore have returned the emitted signal

When determining whether or not a point P is a shadow edge, a set $S^P = (S_0^P, S_1^P, \ldots, S_{35}^P)$ of 36 such scores are generated, one for each mask in the set of all masks $(M_0, \ldots, M_{350})$.

Note that the logic expressed in the second condition is somewhat flawed since there are other situations in which a target on the ground will not produce a return (see the Introduction). These masks instead represent the neighborhood of a border point under ideal conditions (ie. no noise or other distortions).

Figure 4a shows an IFSAR image of one of the buildings in the Kirtland scene. When the binary mask M_0 (Figure 4b) is applied to this image, those points along the building's right-most edge received the best match scores (Figure 4d). This is because the edge bordered a large region of shadowed pixels (ie. drop-outs) and had an orientation perpendicular to that of the mask's. However, when the mask M_{270} (Figure 4c) was applied to the same image, those points along the building's bottom-most edge received the best match scores (Figure 4e). The two masks generated significantly different responses because of their different orientations: M_0 detects vertical back edges while M_{270} detects horizontal back edges.

2.3 Locating Shadow Edges in the Image

The match scores produced by the masks can be used to determine if a point P and its neighborhood (defined earlier as a disc with a radius of 4 pixels) are consistent with the hypothesis that they belong to a shadow/rooftop border. Specifically, they can be used to determine if P is consistent with the hypothesis that a building back edge E_θ passes through it, where the hypothetical back edge is characterized by a single scalar θ (given in radians). The orientation of the hypothesized back edge is perpendicular to θ, while θ itself points into the edge's shadow.

The determination as to whether or not P is consistent with the hypothesis that a back edge similar to E_θ passes through it is made by comparing the set of match scores S^P observed for P to those one would expect to observe for a point along the hypothesized back edge E_θ. That is, the set of match scores observed for P are compared to the set of match scores S^E one would expect to observe for P under the assumption that E_θ passes through P. If the observed scores are similar to the expected scores, then it is plausible that P belongs to a back edge similar to E_θ.

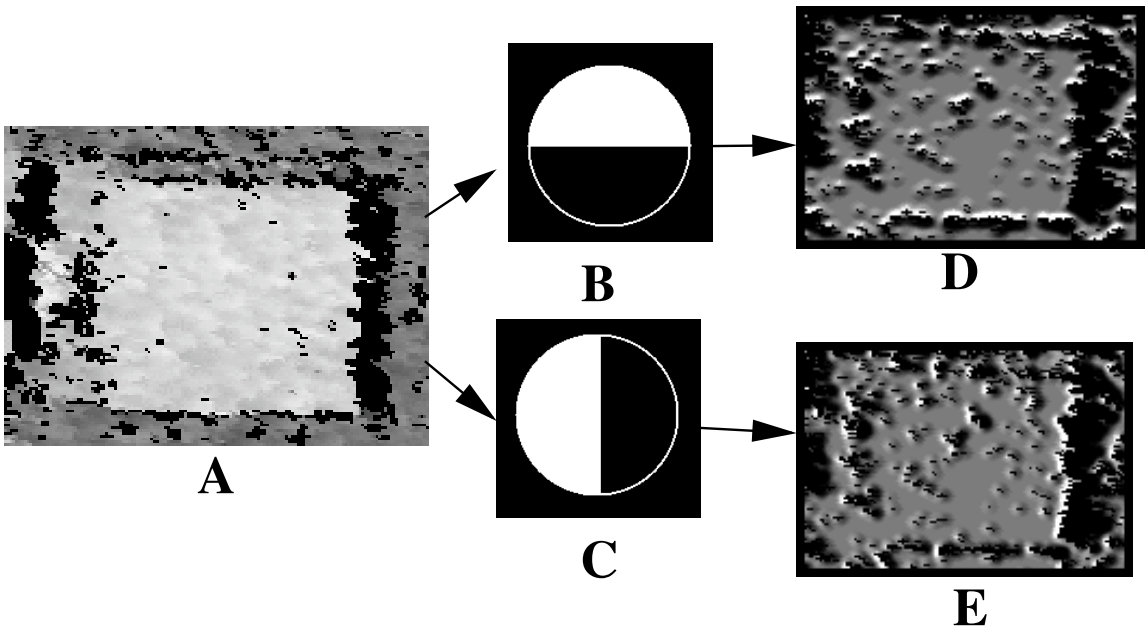


Figure 4: A) A building's height map. B) Binary mask M_{270}. C) Binary mask M_0. D) Match scores resulting from the application of M_{270}. E) Match scores resulting from the application of M_0. A point's grey scale value is inversely proportional to its match score. As such, points receiving the best match scores will be the brightest in the D and E.

Computing the Expected Match Scores S^E

Ideally, the back edge E_θ will neatly bisect P's local neighborhood into an occluded side (ie. shadow) and a visible side (ie. rooftop). That is, all of the points to one side of this edge will lie in the building's shadow. There will therefore be no returns measured for these points. Points on the other side of this edge, however, will belong to the building's rooftop and, as such, be in full view of the sensor. These points will therefore have returns measured for them. It is clear, then, that P's neighborhood will be *identical* to the binary mask M_θ[1] if E_θ does indeed exist. As such, the set of match scores S^E expected for a point along a back edge such as E_θ can be computed by comparing each of 36 masks to M_θ. The match score produced by the application of one mask M_ϕ to another mask M_θ is given by the following equation:

[1]By this we mean that there are no mismatches between M_θ and P's neighborhood

$$S_\phi^E = \left(\frac{2 \cdot |\Delta\theta|}{2\pi}\right)\left(\pi r^2\right) = |\Delta\theta|\ r^2 \tag{1}$$

where r is the radius of the masks (here, r is 4) and $\Delta\theta$ is the difference between the mask's orientation ϕ and θ. This equation will give us the set of expected match scores $S^E = (S_0^E, S_1^E, \ldots, S_{35}^E)$ for a point along the hypothetical back edge E_θ.

Note that the set of expected match scores were computed under the assumption that there was no noise or other distortions present in the image (i.e. S^E is the set of scores we would observe under ideal circumstances). This is quite obviously not the case in a real IFSAR image. In using S^E as our basis of comparison, the issue is whether or not the observed scores are good approximations of the ideal scores. If the approximation is close enough, it is plausible that the observed scores are the ideal scores permuted by noise.

Comparing the Observed Scores to the Expected Scores

The set of match scores S^P derived from the image are compared to the set of ideal match scores S^E using the chi-squared error for binned distributions :

$$\chi^2 = \sum_{i=1} \frac{(N_i - n_i)^2}{n_i} \tag{2}$$

where N_i is the number of events observed in bin i and n_i is the number of events expected to be in bin i. In this case, each binary mask M_ϕ has a corresponding bin, and events occur whenever there is a mismatch between M_ϕ and the neighborhood to which it was applied. (ie the bin count for a mask is equal to its match score - see Section 2.2). The observed matches scores S_ϕ^P are then compared to the expected match scores S_ϕ^E as follows :

$$\chi^2 = \sum_{\phi=0}^{35} \frac{(S_\phi^P - S_\phi^E)^2}{S_\phi^E} \tag{3}$$

A chi-squared distribution with 36 degrees of freedom is used to compute the likelihood p that this large of an error could be generated by chance. If p is greater than 0.05, the two sets are considered to match.

Procedure for Finding Shadow Edges

To find back edges at all possible orientations, our set of back edge hypotheses must have orientations θ that span the range 0° to 360°. This range is broken into ten degree intervals and each interval is represented by a different back edge hypothesis E_θ. This yields a set H of 36 different back edge hypotheses $E_0, E_{10}, E_{20}, \ldots, E_{340}, E_{350}$. However, due to constraints imposed by the geometry of the sensor, only building edges at certain orientations can be back edges. Thus, we do not need to test all 36 hypotheses: if a building edge with an orientation of θ could not possibly be a back edge, then the hypothesis E_θ can be removed from H. Specifically, any hypothesis E_θ whose orientation θ faces towards the sensor can be removed from H. This is because building edges with walls facing the sensor do not cast shadows and are therefore not back edges. For instance, if the the sensor direction is vector $[1, 0]^T$, then H would have hypotheses E_0, E_{10} , ..., E_{80} and E_{270}, ..., E_{350}. This constraint cuts the number of back edge hypotheses we must try in half.

The overall procedure for determining whether or not a point is a shadow edge is as follows :

1. Apply the masks to P to generate the set of observed match scores S^P.
2. Select a back edge hypothesis E_θ from H that has not already been tried. If there are none left, terminate.
3. Compute the chi-squared error between the observed match scores S^P and the match scores S^E expected for a point on our hypothetical back edge E_θ using Equation 4.
4. If the chi-squared error yields a p value greater than 0.05, label P as a shadow edge and terminate. Otherwise, return to step 2.

The chi-squared error of step 3 is computed as follows :

$$\chi^2 = \sum_{i=0}^{35} \frac{\left(S_i^P - |\Delta\theta|\ r^2\right)^2}{|\Delta\theta|\ r^2} \tag{4}$$

where $\Delta\theta$ is the difference between the mask M_i's orientation and θ.

2.4 Confirming Shadow Edges and Grouping them into Back Edges

We next determine which shadow edges (Figure 5, leftmost) belong to the back edge of a building's boundary as opposed to, say, a layover hole. Such shadow edges can be identified by their compliance with the two constraints given in Section 2.1. If a shadow edge E is indeed a back

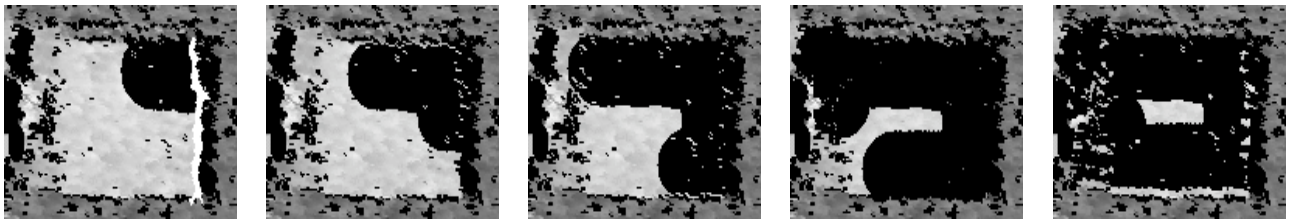
A: Shadow Edges B: Verified Back Edgels C: Final Back Edges

Figure 5: Stages of the Back Edge detection process.

edgel, then, according to the second constraint, E must cast a shadow in the direction of the radar that terminates at some point G on the ground. G is found by moving a small window along a path that begins with E and follows the direction of the sensor. The search terminates when the majority of the pixels within the window have measured returns (i.e. when the window has moved outside of the shadow cast by the building). To overcome the noise inherent in a SAR derived DEM, the median elevation value of the points in that window is selected as the elevation for G. The elevation value for E is selected in a similar fashion. If the candidate E has an elevation sufficiently greater than that of G, the candidate is selected as belonging to a building's back edge. The difference in elevation between E and G must be greater than (or equal to) the minimum height expected of a building in the scene. Here, we expect the height of a building to be at least 3.5 meters. The shadow edges produced earlier (Figure 5, leftmost) will serve as our back edge candidates. These are then verified using the elevation constraint described above (Figure 5, second from the left). Those shadow edges that were not upgraded to back edgels (ie. those shadow edges that could not satisfy the second constraint given in Section 2.1) are stored for later use.

Next, the back edgels are grouped into connected components that represent back edges. This is done in two stages. In the first stage, the system interpolates between verified back edgels. Interpolation occurs along those shadow edges that have met the first criterion but not the second (ie. those shadow edges that were not promoted to back edgels in the prior step - see above). Such edges are promoted if and only if they form a line with back edgels detected in the previous step (shown in Figure 5, second from the left). After interpolation has occurred, a morphological closing is used to bridge small gaps between back edgels. A disc with a radius of two pixels is used as the structuring element in the closing. The resulting back edge regions are shown in the rightmost panel of Figure 5.

Finally, we ascertain the orientation of the back edge by fitting a line to it. This line is fit using a Hough transform. The accumulator array used in the transform has two axes, R and θ:



A B C D E

Figure 6: Extracting the remainder of the building's boundary via region growing. The rooftop region grown so far is shown in black, while the back edge (Figure 5, far right) from whence it began is shown in white. The region's growth progresses panel A to panel E.

R is the perpendicular distance from the line to the origin and θ is the angle that perpendicular ray makes with the x-axis. As such, the line corresponding to the accumulator cell (r, ϕ) is given by the following equation:

$$x\cos(\phi) + y\sin(\phi) = r$$

θ is broken into ten degree intervals, while R is broken into 5 pixel intervals. The orientation of a building's back edge is used when fitting a rectangle to that building's rooftop.

3 Boundary Detection Through Region Growing

3.1 Overall Strategy

After the building's back edge has been detected, the remainder of its boundary is extracted by identifying those points on its rooftop that are near or on one of its bounding edges. This portion of the building's rooftop is located by using a region growing technique classifies points as belonging to either the ground or the rooftop based on an elevation histogram of their local neighborhood. Points that have been labeled as rooftop are added to the growing region only if they are adjacent to points labeled as ground. In this way, the region's growth is restricted to proceed along the building's boundary (see Figure 6). The region growing process is seeded using the back edges extracted earlier.

As mentioned above, classification decisions are based on a threshold found in an elevation histogram of the neighborhood surrounding a point. Since the region's growth is restricted to points near an edge of the building, these neighborhoods will contain points from both the rooftop and the surrounding terrain. Therefore, an elevation histogram of such a neighborhood

should have fairly distinguishable modes corresponding to the rooftop and the ground, allowing a suitable threshold to be found between these modes (see Section 3.2). Note that it is possible for other structures adjacent to building to be included in the rooftop region if these structures are also elevated above the local terrain (e.g. trees).

The region growing process is fairly straight-forward. A portion of the building's back edge serves as our initial rooftop region. It does not matter where this segment is located on the back edge, so long as it is 8-connected. These points are then labeled as rooftop and added to the list of available seeds, which is initially empty. Next, the points at the terminating ends of the shadows cast by these back edgels are labeled as ground (ie. the ground points G derived in Sections 2.1 and 2.4). The rest image of the image remains unclassified.

At each iteration, a point is removed from the list of seeds. Unclassified points within an adaptively sized window centered at this seed will be assigned labels (either *rooftop* or *ground*) during this iteration. This window will be made large enough to include several points that have already been labeled as ground. Once the size of the window has been determined, an elevation histogram of the unclassified points within the window are taken and a threshold is selected. This threshold will be used to classify the unlabeled points as either ground or rooftop. The mean elevation of the points within this window that have already been classified as ground will be used to guide the selection of this threshold (see the next paragraph). Points that have been labeled as rooftop are then added to the list of seeds provided they are adjacent to points classified as ground. This process repeats until no suitable seed points remain.

3.2 Threshold Selection Using A Priori Classifications

Since the terrain adjacent to a building is typically flat (at least locally), points belonging to the ground within the same local neighborhood should have similar The points within the classification window that have already been labeled as ground can therefore provide a rough estimate of the elevation of any ground point within that window, including those yet to be labeled as ground. As such, the mean elevation μ_{prior} of those points already labeled as ground can aid us in selecting an appropriate threshold t. Specifically, the mean elevation μ_t of those unclassified points that *would* be labeled as ground by a particular threshold t should be approximately the same as the mean elevation μ_{prior} of those points *already* labeled as ground. Note that the two means μ_t and μ_{prior} need not be identical. This will allow for a small gradient in the elevation of the ground plane.

After the window size has been selected, we generate the elevation histogram and compute the mean elevation μ_{prior} of those points within the window that have already been labeled as ground. Next, any local minima t within the elevation histogram are identified and added to the set of all such minima S_{minima}. These local minima will serve as our set of candidate thresholds. Finally, for each t in S_{minima}, we then identify those unclassified points with elevations less than

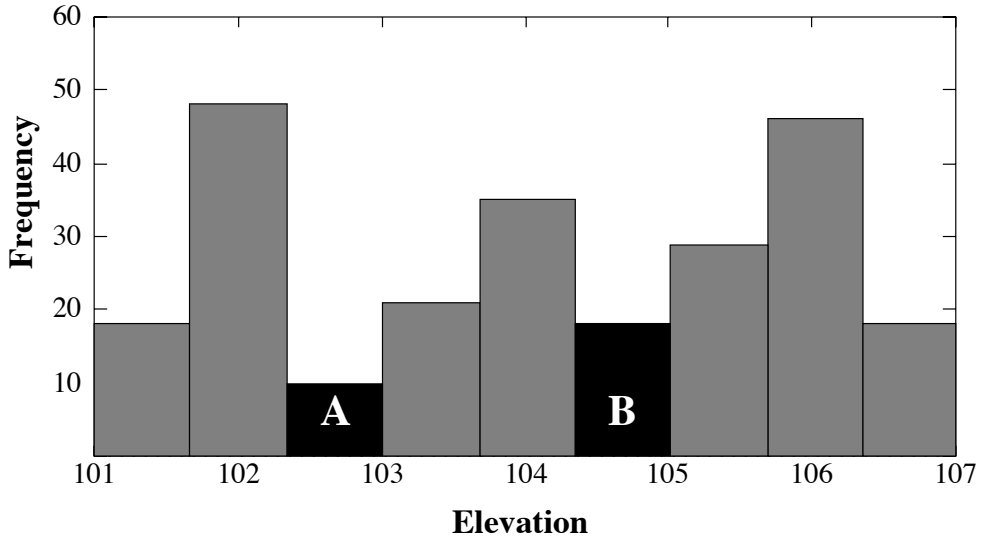


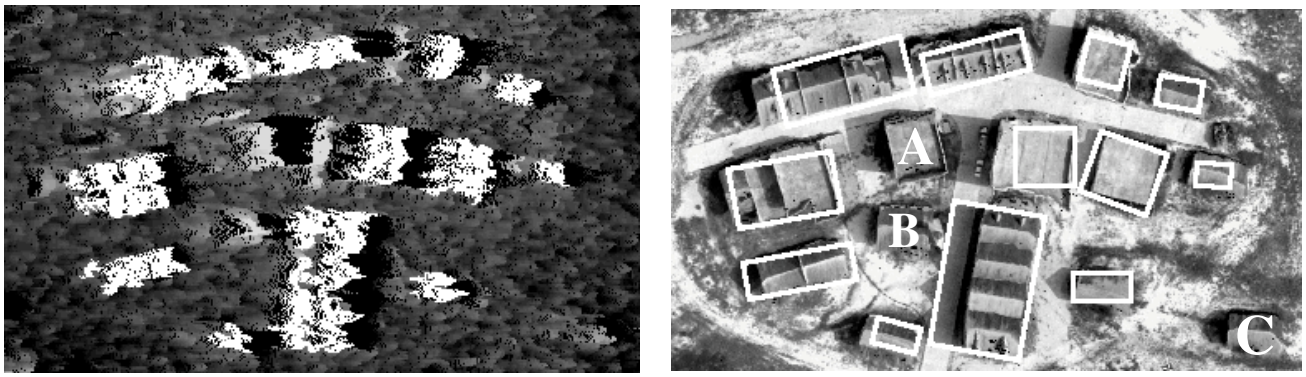
Figure 7: A local elevation histogram used in determining the new classification threshold.

t and compute their mean elevation μ_t. That is, we compute the mean elevation of the points that would be labeled as ground by our candidate threshold t. Once this has been done, our threshold is the elevation t in S_{minima} that minimizes the absolute value $|\mu_{prior} - \mu_i|$. After the threshold has been selected, classification is performed. Note that in the early iterations of t he process, the ground points G used to validate the back edgels will provide our estimates of the ground's elevation.

An example can be seen in Figure 7. There are two local minima in this histogram, indicated in black. The first, or minima A, is at 102.672 meters, and the second, or minima B, is at 104.683 meters. The mean elevation μ_{prior} of the points already classified as ground is 102.199 meters. The mean ground elevation if minima A was used as our threshold would be 102 meters, only 0.199 meters away from our μ_{prior} of 102.199 m. The mean ground elevation if minima B was used as our threshold would be 102.96 meters, which is 0.761 meters away from our μ_{prior}. Thus, a threshold of 102.762m is selected.

4 Results

Boundaries were established for each building found in the image by placing a bounding box around the rooftop region grown for that building. These bounding rectangles were at an orientation equal to that of the building's back edge. The resultant fits for selected areas of the



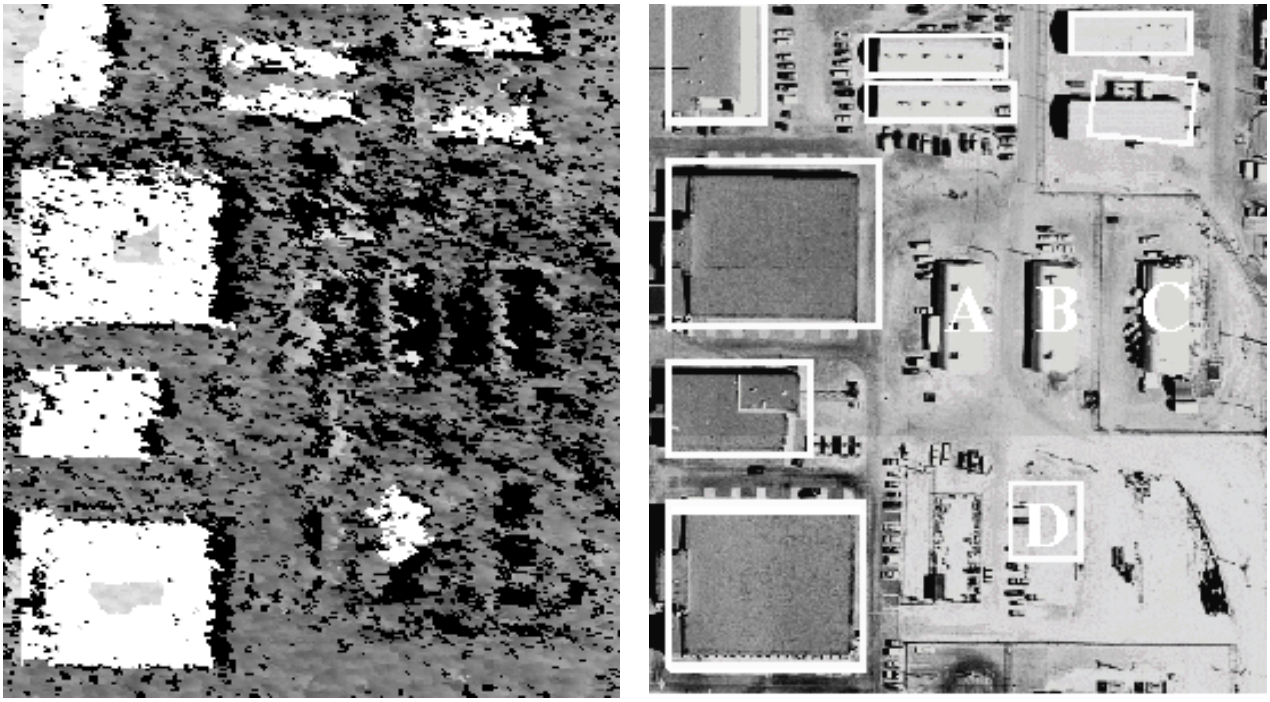
Regions Grown Fitted Rectangles

Figure 8: Buildings extracted from the MOUT DEM. Buildings A, B and C were not detected.

MOUT and Kirtland scenes are shown in Figures 8 and 9. Eight of the eleven buildings in the Kirtland scene were detected along with one false positive. Twelve of the fifteen buildings in the MOUT scene were detected. There were no false positives in the MOUT scene.

Building A and B in the MOUT site were missed because the shadows they cast extended to the front edge of another building. As such, the shadow cast by any point on either buildings' back edge will not terminate on the ground. Instead, it will terminate at a point on another building's rooftop. Because the shadows cast by A and B both terminate on the rooftop of a taller building, any point E on either buildings' back edge will have an elevation less than that of the point G found at the terminating end of its shadow. Therefore, none of the points on either buildings' back edge will meet the second criterion required of a back edgel (see Section 2.1). As such, neither A nor B's back edge was detected.

Buildings A, B and C were not detected because the majority of the points corresponding to those buildings were dropped-out of the IFSAR image. That is, no returns were measured for most of the points corresponding to buildings A, B and C. As such, the height of their rooftops could not be determined, making it impossible to detect their back edges. Figure 10 shows an optical image of the buildings A, B and C. Those points that were dropped-out of the IFSAR image are indicated in black. It is evident from this figure that buildings A, B and C were simply not detected by the IFSAR sensor. The only false positive of both data sets occurred at site D of the Kirtland scene, and is shown in Figure 9. From the optical image of that scene, it appears that D may be some sort of construction site. If this is the case, then it is possible that a foundation erected at that site cast a shadow in the IFSAR image. This would lead to the detection of a back edge at site D. This back edge would then serve as the seed of the false positive.



Regions Grown Fitted Rectangles

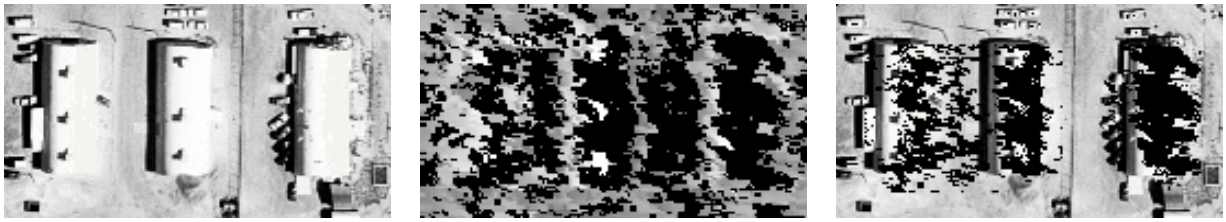
Figure 9: Buildings extracted from the Kirtland AFB scene. Buildings A, B and C were not detected. Building D was a false positive.

The building hypotheses generated for the MOUT site (Figure 11, left) were compared to the set of reference polygons shown in Figure 11. These polygons were extracted by hand and represent the true boundaries of the buildings in the scene. A rooftop hypothesis extracted form the IFSAR image is only valid to the extent that it overlaps one of these polygons. Two metrics were used to evaluate the boundaries extracted by the system:

$$\begin{aligned} \textit{Detection Rate} &= \frac{TP}{TP+FN} \\ \textit{False Alarm Rate} &= \frac{FP}{TP+FP} \end{aligned}$$

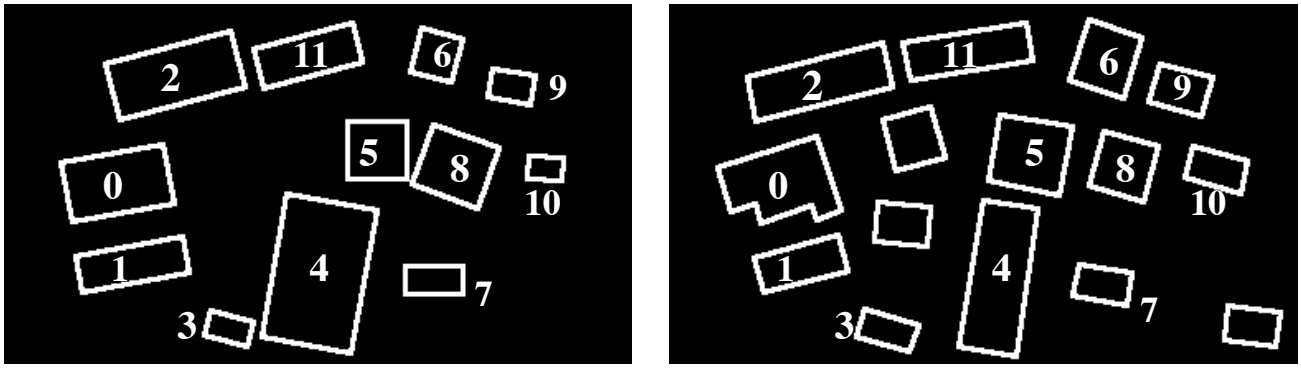
where TP is the total number of true positives, FP is the total number of false positives and FN is the total number of false negatives.

A point inside the bounding box established for a building (Figure 11, left) is considered



Optical IFSAR Optical + IFSAR

Figure 10: Buildings dropped-out of the Kirtland DEM.



System Results Reference

Figure 11: Left: Boundaries extracted by the system. Right: Reference polygons hand-extracted from an orthorectified optical image.

to be a true positive if it is also within that building's reference polygon (Figure 11, right). Otherwise, that point is labeled as a false positive. A point is a false negative if it is interior to the reference polygon but outside of the boundary extracted by the system. Table 1 shows the false alarm and detection rates for the buildings found by the system. Those buildings that went undetected by the system were not evaluated in this fashion.

References

[1] F.W. Leberl. *Radargrammetric Image Processing*, p. 472 - 473, Artech House, 1990.

[2] F.W. Leberl. *Radargrammetric Image Processing*, p. 84 - 88, Artech House, 1990.

Building	TP	FP	FN	Detection Rate	False Alarm Rate
0	2229	218	344	0.86	0.09
1	1130	348	48	0.95	0.23
2	1187	877	455	0.72	0.42
3	374	44	214	0.63	0.1
4	3036	1086	84	0.97	0.32
5	1145	115	698	0.62	0.09
6	708	0	426	0.62	0
7	480	115	174	0.74	0.19
8	1191	459	0	1.0	0.28
9	467	0	341	0.58	0
10	295	0	139	0.68	0
11	1310	226	458	0.74	0.15
Mean				0.73	0.16

Table 1: Detection and False Alarm rates for the MOUT site.

[3] U. Weidner, W. Forstner. Towards Automatic Building Extraction from High Resolution Digital Elevation Models. *ISPRS Journal*, p. 38 - 49, 1995.

[4] R. Chellappa, S. Kuttikkad, R. Meth, P. Burlina, K. Ome, C. Shekkar. Model Supported Exploitation of SAR Imagery. *Proc. ARPA Image Understanding Workshop*, p. 389–408, 1996.

[5] Vexcel Corporation. Building Extraction from IFSAR Data. *Private Communication*, 1998.