Use of IFSAR with Intensity Images for Automatic Building Modeling

A. Huertas, Z. Kim and R. Nevatia
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089

Abstract

IFSAR (Interferometric Synthetic Aperture Radar) data contains direct 3-D information about a scene. However, the resolution of such data is typically lower than that of electro-optical (EO) images and the data may have many missing or highly erroneous elements. We propose to use IFSAR data to provide cues for more accurate detection and reconstruction using the EO images. It is shown that such cues can not only greatly improve the efficiency of the automatic system but also improve the quality of the results. Quantitative evaluations are given.

1 Introduction and Overview

Three-D models of man-made structures in urban and sub-urban environments are needed for a variety of tasks. The principal sensors used for this task have been electro-optical (EO) achromatic images acquired from an aircraft [Roux & McKeown, 1994, Noronha & Nevatia, 1997, Collins et al., 1998, Ascona II, 1997, CVIU, 1998]. EO images have many advantages: they are relatively easy to acquire at high resolution (say of the order of 0.5 meters/pixel) and humans find it is easy to visualize them and to extract the needed information from them. However, their use for automatic extraction has proven to be quite difficult. One of the principal causes of this difficulty is the lack of direct 3-D information in the 2-D images. Three-D information can be inferred for features that can be correctly corresponded in multiple images (assuming knowledge of relative camera geometry) but the correspondence problem is a difficult one as the feature appearances can change in different views and other similar features may exist. Also, aerial images may contain large areas that are homogeneous, such as roofs of buildings where few features exist to match in different views and 3-D information must be inferred by interpolation which requires correct surface segmentation.

In recent years, sensors have been developed that can measure 3-D *range* to a point directly. Availability of this information makes the task of building detection much easier as these structures are elevated above the surrounding background. Two classes of such sensors have been developed. The first, called **LI**ght **D**etection **A**nd **R**anging (LIDAR), uses a laser beam for illumination [Stetina et al., 1994]; the distance to a point is determined by the time taken for light to travel to and return from the point (the actual measurement may be done by measuring phase change). The second, called **I**nter**f**erometric **S**ynthetic **A**perture **R**adar (IFSAR), computes 3-D position by interferometry from two SAR images [Curlander & McDonough, 1991; Jakowatz et al., 1996]. Both sensors use active, focused illumination and rely on reflected light to reach back to the sensor. However, many surfaces act like mirrors at the wavelengths of the respective sensors and those points are not well imaged. Thus, data from range sensors typically have many holes or are even completely erroneous. The resolution of such images is typically lower than that of intensity images. Such images can also be difficult for humans to visualize and fuse with the EO images.

The complementary qualities of EO images and IFSAR data provides an opportunity for exploiting them in different ways to make the task of automatic feature modeling easier; in this paper, we focus on the task of building detection and reconstruction. Combining the two data sources at the pixel level is difficult as there is not a one-to-one correspondences between the pixels in the two sources, in general. Instead, we propose to extract information from each which is then combined and perhaps used to guide extraction of additional information. In par-

* This research was supported in part by the Defense Advanced Research Projects Agency of the U.S. Government under contract DACA 76-97-K-0001 and monitored by the Topographic Engineering Center of the U.S. Army.

ticular, we feel that the IFSAR data is suited for *detecting* possible building locations as buildings are easily characterized by being significantly higher than the surround. However, due to the low resolution of range data derived from IFSAR, the derived boundaries are not likely to be precise and it may be difficult to distinguish a building from other raised objects such as a strand of trees. EO data, with much higher resolution and lack of artifacts associated with a radar sensor, can provide precise delineation as well as distinguish a building from other high objects much more reliably. However, features, such as edges in intensity images have inherent ambiguities; it is hard to tell if they belong to object boundaries or arise from discontinuities in illumination (shadows) or surface reflectance (markings). The coarse delineation provided by analysis of IFSAR data can help overcome this ambiguity.

Other approaches to use of range data may be found in [Hoepfner et al., 1997; Chellappa et al. 1997; Haala & Brenner, 1997].

In the next section, we describe how useful *cues* can be extracted from the IFSAR data. Use of these cues in the building extraction process is then described. Results comparing the effects of these cues are presented in section 4.

2 Cues from IFSAR Data

We aim to extract cues for presence of buildings. IFSAR data is usually given in the form of three images, called the **mag**, **dte** and **cor** images. The **mag** image is like a normal intensity image measuring the amount of reflected signal coming back to the sensor (at the radar wavelength). The **dte** image encodes the 3-D information in the form of a *digital terrain elevation* map where the pixel values define the height of the corresponding scene point. The **cor** image contains the phase correlation information between two images used for the interferometric process; it can be useful in distinguishing among types of materials as the returns associated with objects that remain stationary, such as buildings, are highly correlated.

Certain kinds of IFSAR sensors, such as a searchlight mode sensor developed by Sandia National Laboratories, use several views of an area from different angles and produce a higher resolution (of the order of 0.4 meters per pixel) and more reliable **dte** image. Figure 1 shows a built up area at the Fort Benning, Georgia, site. The derived **dte** image is shown in Figure 2. For such an image, cues for buildings can be detected from the **dte** image alone (in fact, the **mag** and **cor** images are not very meaningful for this sensor mode). As the ground height is not necessarily constant over a significant area, it is not sufficient to simply threshold the **dte** image. We need to find regions that are high *relative* to the surround.

Figure 1 McKenna MOUT at Fort Benning

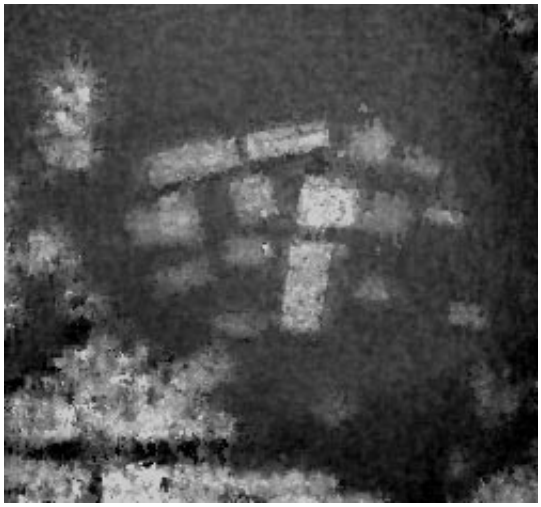


Figure 2. IFSAR derived DTE image.

We extract object cues by convolving the **dte** image with a Laplacian-of-Gaussian filter that smooths the image [Chen et al., 1987]. The space constant of the Gaussian is the only parameter associated with this process, and it is not a critical one. The objective is to apply a reasonable amount of smoothing as a function of the characteristics of the **dte** image.

The pixel values in the **dte** image represent elevation thus, the zero-crossings in the convolution output denote significant elevation changes. The positive-valued regions in the convolution output thus are taken to represent objects above the local ground. We collect these regions by connected component analysis, and select those having a reasonable size to represent raised objects such as buildings and groups of trees. Figure 3 shows the regions that result from this process; as can be seen that not only are the buildings in the center of the

image detected but so are some of the groups of trees on the South and West sides of the building area. It is difficult to further distinguish between these regions or to get more accurate descriptions from the given IFSAR data alone.

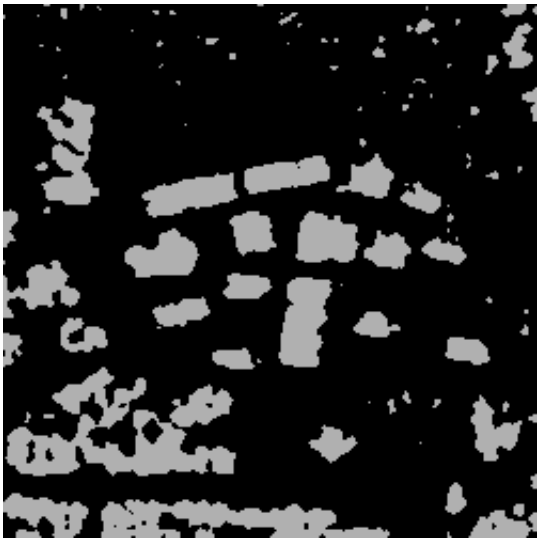
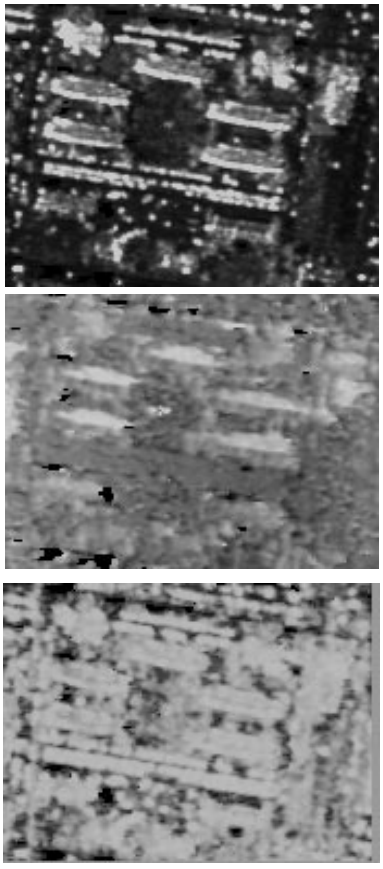


Figure 3. Cues extracted from IFSAR DTE.

Next, we discuss the use of low resolution IFSAR data which is much more common (such as from the IFSARE sensor). Consider the portion of an intensity image from a Fort Hood, Texas site shown in Figure 4; it contains 11 buildings. The corresponding **mag, dte** and **cor** images from an IFSARE sensor with 2.5 meter/pixel resolution are shown in Figure 5. Only some of the buildings appear salient in the **dte** image. The buildings are more apparent to us in the **mag** image but the intensity over a building is hardly constant with the edge towards the sensor (the bottom edge in the image) consisting of much brighter points (as expected from a Radar sensor). The **cor** image also appears to contain useful information corresponding to the buildings. Phase decorrelation values indicate pixels that remain stationary between successive SAR acquisitions and help verify the presence of stationary objects such as buildings.

Figure 4. Portion of Fort Hood EO image.



Mag image

Dte image

Cor image

Figure 5. IFSARE components

Figure 6 shows that it is not sufficient to threshold the **dte** images to obtain cues corresponding to objects of interest. Figure 6 (left) shows the **dte** regions "just above the ground", that is, about 1.5 meters above the ground (mean elevation). Figure 6 (right) shows the thresholded **dte** image at the mean intensity plus one standard deviation. The buildings are somewhat apparent in this image but the presence of many artifacts would be misleading to an automated system beyond a rough indication of possible presence of a building.

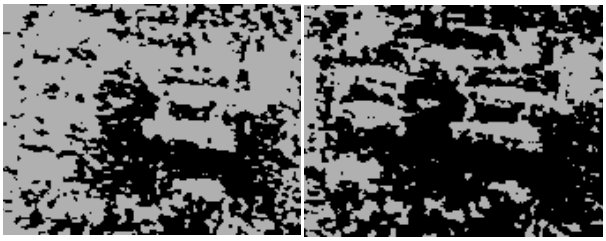


Figure 6. Thresholding of **dte** image. At mean (left); at mean plus one std. dev. (right).

We feel that it is advantageous to use a combination of the **dte**, **mag** and **cor** images to extract reliable cues in such cases. The combination of these images proceeds in three steps as follows:

Step 1: combine the **mag** and **dte** components, Let ***M***, ***D***, and ***C*** represent the registered **mag**, **dte** and **cor** image components respectively. Further, let:

$$M_{LoG} = \boldsymbol{M} \otimes \nabla^2 G$$

$$D_{LoG} = \boldsymbol{D} \otimes \nabla^2 G$$

represent the LoG convolutions. The image:

$$I_{MD} = W_M \cdot M_{LoG} + W_D \cdot D_{LoG}$$

linearly combines the enhanced positive-valued pixels in the smoothed LoG images. W_M and W_D represent the relative weights of the **dte** and **mag** contributions. Our current implementation uses 10.0 and 1.0 respectively, reflecting the increased importance of the **dte** component.

Step 2: use the **cor** image to filer out the pixels with that do not have high decorrelation values (lower than 0.9) from the I_{MD} image. Then filter the I_{MD} image to contain positive-valued pixels only.

Step 3: collect connected components in the filtered I_{MD} image and select those regions having a certain minimum area.

Figure 7 shows the result of applying this process to the Fort Hood example. The *cues* image, shown in Figure 7 (left) is the output of Step 2. The regions denote objects above the ground, including trees, buildings, large vehicles and vehicle formations. Figure 7 (right) shows the connected components that have a certain minimum size (area) and taken to correspond to cues for building structures. Note that the all the buildings are well represented except for the one in the lower right which is difficult to discern in the **dte** component.

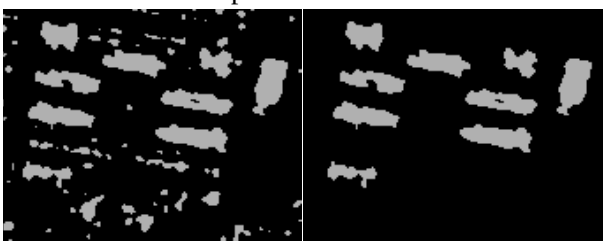


Figure 7. Computed cue regions (left) and cues selected by size (right).

3 Integration of Cues into the Building Detection System

We next describe the use of these cues in the multi-view building detection and description system described in [* hidden *]. This system has three major phases: hypothesis formation, selection and validation. This system assumes that the roofs of buildings are rectilinear though the roofs need not be horizontal (some forms of gables are allowed). Hypotheses are formed by collecting a group of lines that form a parallelogram in an image or a rectangle parallelepiped in 3-D. Multiple images and matches between lines are used in the hypotheses formation stage. As line evidence can be quite fragmented, liberal parameters are used to form hypotheses. Properties of resulting hypotheses are used to select among the competing hypotheses. The selected hypotheses are then subjected to a verification process where further 3-D evidence, such as presence of walls and predicted shadows are examined. Details of this process are not included for lack of space.

The cues extracted from the IFSAR data can help improve the performance of the building description system at each of the three stages described above. We show some details and results of these processes.

Hypothesis Formation:

Cues can be used to significantly reduce the number of hypotheses that are formed by only considering line segments that are within or near the cue regions. Top of Figure 8 shows the line segments detected in the image of Figure 4 (using a Canny edge detector); the bottom of this figure shows the lines that lie near the IFSAR cues. As can be seen, the number of lines is reduced drastically (77.4% and 78.1% respectively in each of the two EO images processed) by without loosing any of the lines needed for forming building hypotheses (except for the one building in the lower right that is not cued by IFSAR processing in this example). This not only results in a significant reduction in computational complexity but many false hypotheses are eliminated allowing us to be more liberal in the hypotheses formation and thus including hypotheses that may have been missed otherwise. Figure 9 shows the linear structures that are near the cue regions for the Ft. Benning example shown in Figure 1 earlier.

Hypothesis Selection:

The building detection system applies a series of filters to the hypotheses formed. The remaining hypotheses are then evaluated in the basis of the geometric evidence (underlying line segments that support the hypothesized roof boundaries), in an attempt to select a set of “strong” hypotheses. With the introduction of IFSAR cueing evidence we can eliminate the initial filtering stages and introduce this evidence into the roof support analysis. The new evidence consists of an indication of the support of a roof hypotheses in terms of the overlap between the roof hypotheses and the cue regions extracted from IFSAR. The hypotheses are projected onto the IFSAR image and the overlap of the pro-

Figure 8. Linear segments in EO image (top) and those near IFSAR cues (bottom).

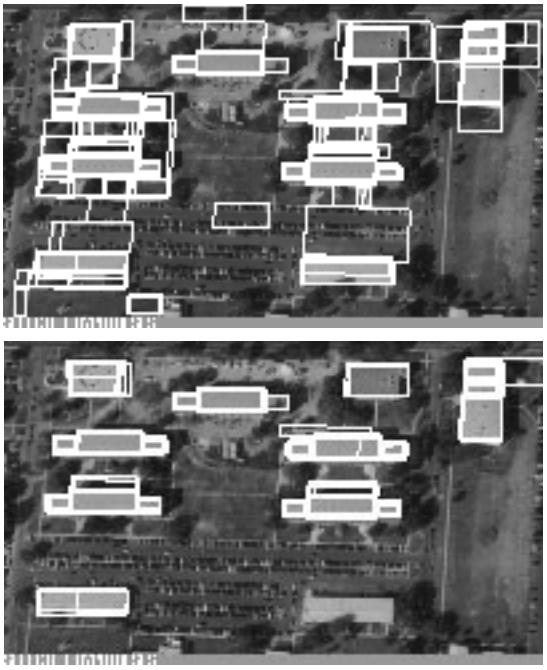


Figure 10. Selected Hypotheses using EO only (top) and using IFSAR cues (bottom).

Figure 9. Lines near IFSAR cues

jected roof with IFSAR regions is computed. The current system requires that the overlap be at least 50% of the projected roof area.

Figure 10 shows the selected hypotheses, with (bottom) and without (top) the use of the IFSAR cues.

Hypotheses Validation:

Just as poor hypotheses can be discarded because they lack IFSAR support, the ones that have a large support see their confidence increase during the verification stage. In this stage, the selected hypotheses are analyzed to verify the presence of shadow evidence and wall evidence. Details of the shadow and wall analysis are given in [* hidden *]. The presence of shadow and wall evidence is required for verification of the selected hypotheses. When no evidence of walls or shadows is found, we require that the IFSAR evidence (overlap) be higher, currently 70%, in order to validate a hypotheses. The validated hypotheses may contain overlapped hypotheses. These are analyzed to give a set of final building hypotheses. These are shown in Figure 11. On top we show the final hypotheses extracted without IFSAR cueing, and on the bottom, those extracted using IFSAR support. Note that false detections are eliminated with IFSAR cueing. Also, the building on the lower right is not found (Figure 11, bottom) as the lack of a cue prevented a hypothesis to be formed there. On the other hand, the building component on the middle left is not found without IFSAR support but found with it.

Table 1 gives a comparison of the number of features and final result component counts with and without use of IFSAR cues for the Fort Hood example. An evaluation of the quality of results is given in section 4.

Table 1: Ft. Hood Automatic Result

Feature	EO Only	With IFSAR
Line Segments	7799/7754/5083	
Linear Struc.	2959/2963/2042	669/650/552
Flat Hypos.	2957	1732
Selected Flat	383	296
Verified Flat	215	192
Final Flat	21 (4 false)	18 (0 false)

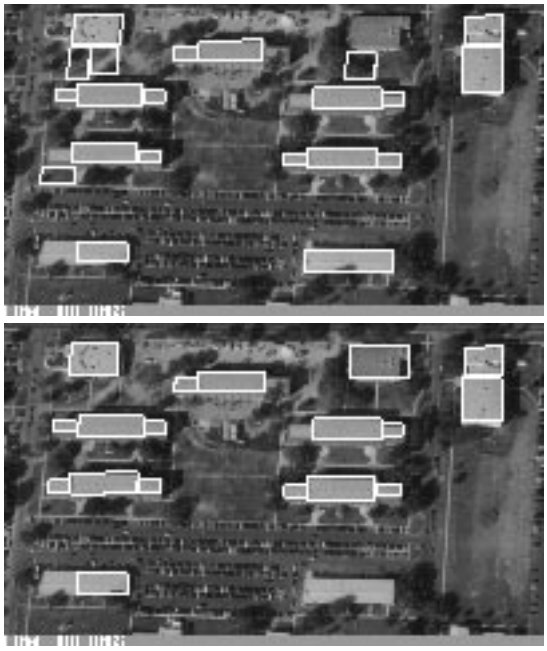


Figure 11. Final hypotheses using EO only (top) and using IFSAR cues (bottom).

Figure 12 show the combined detected flat and gabled roofed buildings using the IFSAR cues for the Ft. Benning example. This result shows no false alarms. Also, the roofs of the gabled buildings are detected correctly. However, parts of the gabled buildings in the upper center have not been detected.

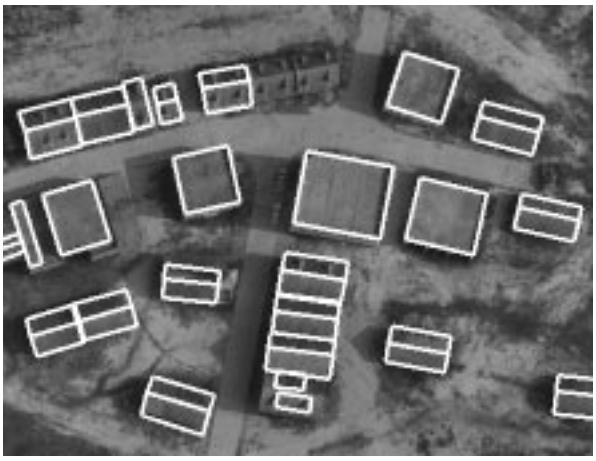


Figure 12. Extracted using IFSAR cueing.

Table 2 gives a comparison of the number of features and final result component counts with and without use of IFSAR cues for the Fort Benning example.

Table 2: Ft. Benning Automatic Result

Feature	EO Only	With IFSAR
Line Segments	116400/18998	16400/18998
Linear Struc.	55827/6611	1758/2041
Flat Hypos	5218	3012
Selected Flat	329	192
Verified Flat	202	116
Final Flat	22	15
Gable Hypos	634	240
Selected Gab.	181	75
Verified Gab.	181	75
Final Gab.	19	17
Combined	41	32
Buildings	29 (3 false)	25 (0 false)

4 System Evaluation

We define detection rate and false alarm rates as follows for quantitative evaluation of our results.

$$\textit{Detection Rate} = \frac{TP}{(TP + FN)}$$

$$\textit{False Alarm Rate} = \frac{FP}{(TP + FP)}$$

TP, FP and FN stand for true positives, false positives and false negatives. Note that with these definitions, the detection rate is computed as a fraction of the reference features whereas the false alarm rate is computed as a fraction of the detected features.

In the definitions given above, a feature could be an object, an area element or a volume element. We consider a building to have been detected if *any* part of it has been detected. The amount by which a building has been correctly detected is computed by the number of points inside that overlap with the reference.

Table 3 shows a summary of detection results for our Ft. Hood example in terms of object components. The reference model is shown in Figure 13. Note that false alarms disappear when IFSAR cues are available.

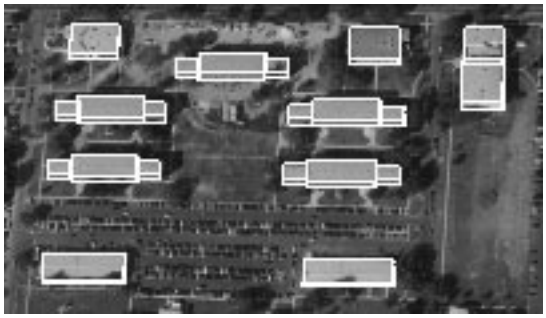


Figure 13. Reference model for evaluation.

Table 3: Ft. Hood Evaluation

	EO only	With IFSAR
Model	21	
Total	21	18
TP	18	18
FP	4	0
FN	2	1
Detection Rate	0.90	0.95
False Alarm Rate	0.18	0.00

Similar results for the Ft. Benning example are shown in Table 4. The reference model is not shown.

Table 4: Ft. Benning Evaluation

	Flat	Gable	Combined
Model	6	20	26
EO Total	22	19	41
TP	6	13	19
FP	2	1	3
FN	0	1	1
Detection Rate for EO Only			0.95
False Alarm Rate for EO Only			0.13
IFSAR Total	15	17	32
TP	6	16	22
FP	0	0	0
FN	0	2	2
Detection Rate for EO plus IFSAR			0.92
False Alarm Rate for EO plus IFSAR			0.00

Table 5 and table 6 show the combined global rates for the Ft. Hood and Ft. Benning scenes shown in Figure 11 and Figure 12 respectively. The ground detection rate is the percent correct classification of ground pixels. When IFSAR cues are available, the performance can be improved significantly

.

Table 5: Ft. Hood Combined Area Evaluation

	EO Only	with IFSAR
Detection rate	0.7461	0.7545
False Alarm rate	0.2023	0.0883
Ground Detection rate	0.9688	0.9863

Table 6: Ft. Benning Combined Area Evaluation

	EO Only	with IFSAR
Detection rate	0.8219	0.8341
False Alarm rate	0.1196	0.0407
Ground Detection rate	0.9814	0.9937

The combined result is visualized by a histogram of buildings which are detected to a certain accuracy. Figure 14 shows four curves for the *area* analyses of the Ft. Hood detection result. The graphs show detection rates vs. percent of buildings detected at that rate or below, with and without IFSAR cueing. The (ref) curves show the detection rate vs. percent of **reference** buildings detected at that rate or below and the (det) curves show the corresponding rates for the **detected** buildings.

Figure 15 show a similar graph for the *volumetric* analyses.

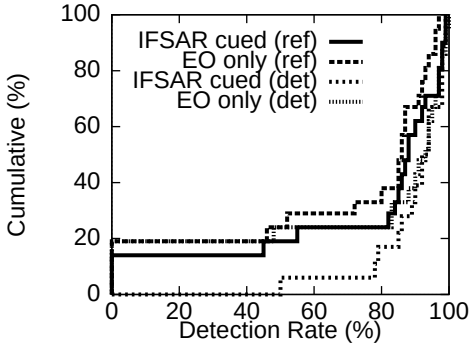


Figure 14. Evaluation curve for *area* analysis of **reference** and **detected** buildings in Fort Hood Example.

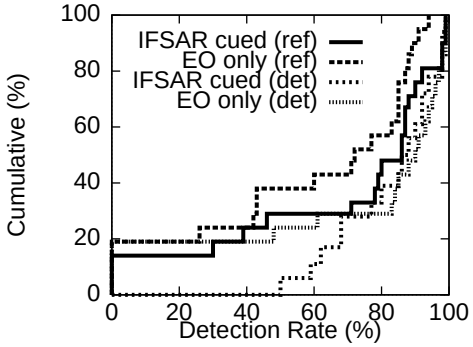


Figure 15. Evaluation curves for *volumetric* analysis of **reference** and **detected** buildings in Ft. Hood Example.

Figure 16 and Figure 17 show similar graphs for the area and volumetric analyses of the Ft. Benning result.

5 Conclusions

We have presented a methodology for detection and reconstruction of building structures by using conventional intensity images with magnitude, elevation and correlation data derived from IFSAR sensors. Even though the IFSAR data is of a lower resolution and contains many missing elements and

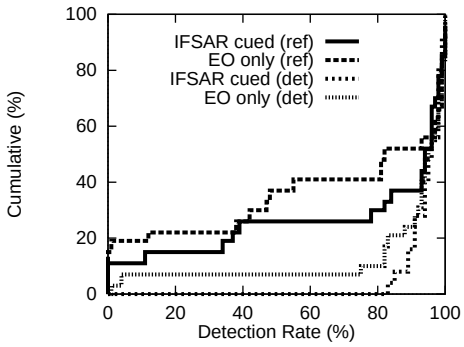


Figure 16. Evaluation curve for *area* analysis of **reference** and **detected** buildings in Ft. Benning example.

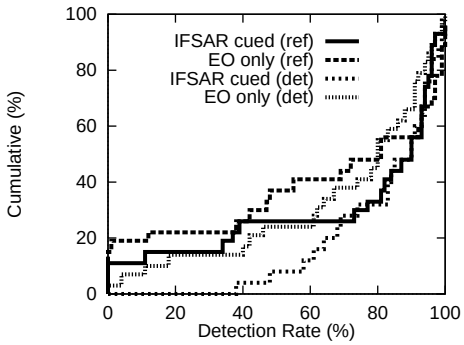


Figure 17. Evaluation curve for *volumetric* analysis of **reference** and **detected** buildings in Ft. Benning example

artifacts, it has been shown that it can be used to enhance the results of EO image analysis while substantially reducing the computational complexity. This was accomplished not by combining the information at the sensor level but rather by using analysis of one to guide the analysis of the other. We believe that this paradigm will be suitable for other tasks as well as sensors of different modalities become available for more domains.

References

[Ascona II, 1997] Proceedings of the Ascona Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images II, A. Gruen, E. Baltsavias & O. Henricsson, Ed., Brinkhauser Verlag, Switzerland, May.

[Chellapa et al., 1997] R. Chellapa, Q. Zheng, S. Kuttikkad, C Shekhar and P. Burlina, "Site Model Construction for the Exploitation of E-O and SAR Images", in RADIUS97, O. Frischein and T. Start, Ed., Morgan Kaufmann, San Francisco, pp 185-208.

[Chen et al., 1987] J. Chen, A. Huertas and G. Medioni, "Fast Convolution with Laplacian-Of-Gaussian Masks", IEEE T-PAMI, 9(4), pp 584-590.

[Collins et al., 1998] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, and E. Riseman, "The ASCENDER System: Automatic Site Modeling from Multiple Aerial Images", Computer Vision and Image Understanding Journal, Vol 72, No. 2, November, pp 143-162.

[Curlander & McDonough, 1991] J. Curlander and R. McDonough, "Synthetic Aperture Radar", Wiley Interscience, New York, 1991.

[CVIU, 1998] Computer Vision and Image Understanding Journal, Special Issue on Automatic Building Extraction from Aerial Images. Academic Press, Vol. 72, No. 2, November.

[Fischer et al., 1998] A. Fischer, T, Kolbe, F. Lang, A. Cremers, W. Foerstner, L. Pluemer and V. Steinhage. "Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D", Computer Vision and Image Understanding Journal, Vol 72, No. 2, November, pp 185-203.

[Hoepfner et al., 1997] K. Hoepfner, C. Jaynes, E. Riseman, A. Hanson and H. Schultz, "Site Modeling using IFSAR and Electro-Optical Images", Proceedings of the DARPA Image Understanding Workshop, New Orleans, LA, May 1997, pp 983-988.

[Haala & Brenner, 1997] N. Haala and C. Brenner, "Interpretation of Urban Surface Models using 2D Building Information", in Automatic Extraction of Man-Made-Objects from Aerial and Space Images (II), A. Gruen, E, Baltsavias and O, Henricsson, Ed., Brinkauser, Basel, pp 213-222.

[Jakowatz et al., 1996] C. Jakowatz, D. Wahl, P. Eichel, D Ghiglia and P. Thompson, "Spot-Light Mode Synthetic Aperture Radar: A Signal Processing Approach", Kluwer Academic, Boston, 1996.

[Noronha & Nevatia, 1997] S. Noronha and R. Nevatia. "Detection and Description of Buildings from Multiple Aerial Images", Proceedings IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, June 1997, pp. 588-594.

[Paparoditis et al., 1998] N. Paparoditis, M. Cord, M. Jordan and J.P. Cocquerez, "Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery", Computer Vision and Image Understanding Journal, Vol 712, No. 2, November, pp 122-142.

[Roux & McKeown, 1994] M. Roux and D. McKeown, "Feature Matching for Building Extraction from Multiple Views". Proceedings of the DARPA Image Understanding Workshop, Monterey, CA, Nov. pp. 331-349.

[Stetina et al., 1994] F. Stetina, J. Hill and T. Kunz, "The Development of a Lidar Instrument for Precise Topographic Mapping", Proceedings of International Geoscience and Remote Sensing Symposium, Pasadena, CA, August 1994.