

Multisensor Integration for Building Modeling¹

Andres Huertas, ZuWhan Kim, and Ramakant Nevatia
Institute for Robotics and Intelligent Systems
University of Southern California
3737 USC-Watt way, PHE 204 Los Angeles, California 90089
{huertas, zuwhan, nevatia}@iris.usc.edu

Abstract

Machine perception can benefit from the use of features extracted from data provided by a variety of sensor modalities. Recent advances in sensor design makes it possible to incorporate multiple sensors into vision systems for increased capability. Two important issues must be considered for the integration task: The sensors must be spatially coregistered and the phenomenologies must be compatible. In this paper we address these issues as they apply to the problem of automatic modeling of building structures from aerial views. We present a methodology to incorporate cues extracted from IFSAR (Interferometric Synthetic Aperture Radar) to significantly improve the performance and the quality of the results of an existing system that relies on electro-optical panchromatic images, while reducing processing time. Quantitative evaluations are given.

1 Introduction

A critical task in computer vision is to derive 3-D models of objects from images. In some domains the scenes can be very difficult to analyze because of increased clutter from various kinds of features and complex detail.

One domain that can clearly benefit from multisensor modalities is the analysis of aerial and satellite images. The number of sensors available for modeling, monitoring and management of objects and conditions on the Earth's surface is large and operate in very different modalities. Our group has many years of experience analyzing aerial images using EO panchromatic (PAN) images to construct 3-D models of man-made structures in urban and suburban environments, and we believe this task can benefit from addressing many of the challenges of integration of

information from different sensors to improve the results and performance of these systems.

The principal sensors used for the building modeling task have been PAN images acquired from an aircraft [1-5]. PAN 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. In this paper we propose the integration of information from a second sensor to help ameliorate the difficulties. In particular, the use of IFSAR images have proved quite useful. Combining the two data sources however 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. We present examples from two different IFSAR imaging methods and evaluate the results with and without their use for two different scenes.

2 Sensor Modalities, Phenomenology and Registration

One of the principal causes of difficulty in analyzing PAN images of semi- and urban environments 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 fea-

1. 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, and in part, by the U.S. Army Research Office under grant No. DAAH04-96-1-0444.

tures 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 **L**ight **D**etection **A**nd **R**anging (LIDAR), uses a laser beam for illumination [6]; 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**nterferometric **S**ynthetic **A**perture **R**adar (IFSAR), computes 3-D position by interferometry from two SAR images [7, 8]. Both sensors use active, focused illumination and rely on reflected radiation 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 PAN images.

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.

The images generated by the processing of radar signals and by interferometry have the geometric characteris-



Figure 1. McKenna MOUT at Fort Benning



Figure 2. IFSAR derived DTE image

tics of an orthographic projection. The estimation of the sensor parameters, or “camera” model, associated with this overhead (nadir) viewpoint is straightforward. The camera model allows us to derive the appropriate 3D-to-2D and 2D-to-image transforms needed to register the available PAN images to the IFSAR images. We use these transforms to project PAN 2-D and 3-D features onto the IFSAR images to assist and support the building detection system at various stages of processing.

3 Features from IFSAR Data

The complementary qualities of PAN 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, using statistical or other methods, 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 particular, 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. PAN 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 [9-11].

Next, 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.

We extract object cues by convolving the **dte** image with a Laplacian-of-Gaussian filter [12]. 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.

Next, we discuss the use of low resolution IFSAR data which is much more common (such as from the IFSARE sensor). A portion of an intensity image from a Fort Hood, Texas site is 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



Figure 3. Cues extracted from IFSAR DTE

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 PAN image

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.

We believe 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:

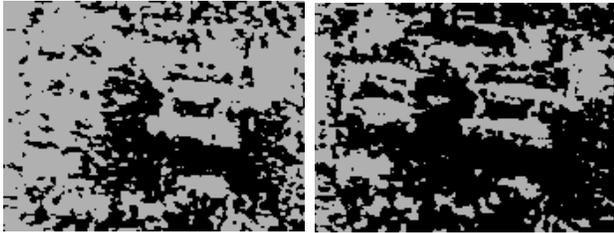
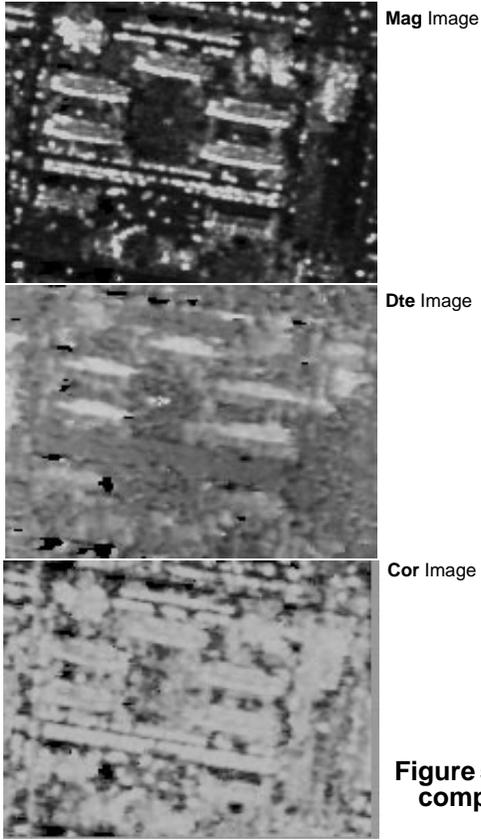


Figure 6. Thresholding of dte image. At mean (left); at mean plus one std. dev. (right)

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:

$$M_{LoG} = M \otimes \nabla^2 G \text{ and } D_{LoG} = D \otimes \nabla^2 G$$

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 filter 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.

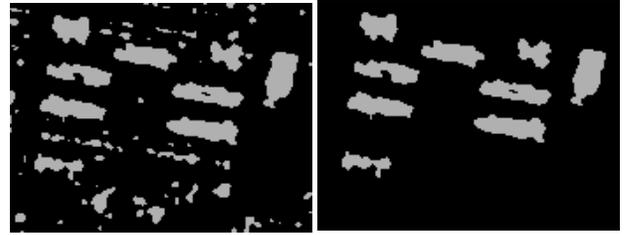


Figure 7. Computed cue regions (left) and cues selected by size (right)

4 Integration of IFSAR 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.

Next, we show that the IFSAR cues help improve the performance of the building description system at each of the three stages described above.

Hypothesis Formation:

Cues can be used to significantly reduce the number of hypotheses that are formed by only considering linear segments that are on or near the cue regions. Figure 8a shows

the line segments detected in the image of Figure 4 (using a Canny edge detector) and Figure 8b shows the lines that are near the IFSAR cues. The number of lines is reduced drastically (95.6%, 95.5% and 96.4% in each of the three PAN images processed) without losing 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 and processing time required but many false hypotheses are eliminated allowing us to be more liberal in the hypotheses formation and thus include 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.

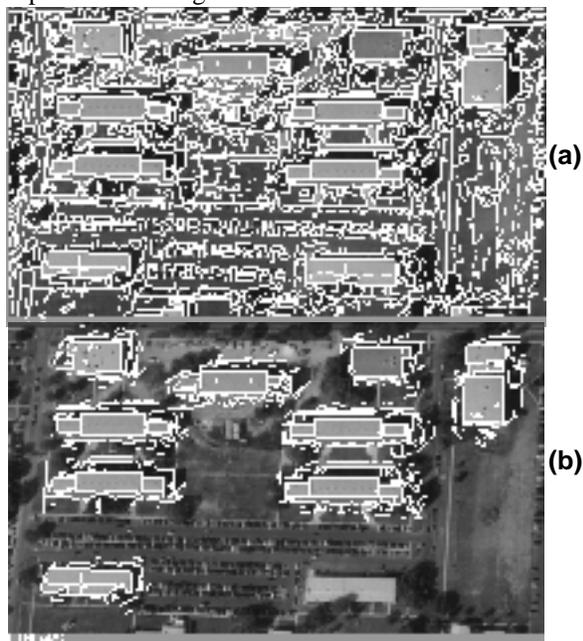


Figure 8. (a) Linear segments in PAN image and (b) those near IFSAR cues



Figure 9. Lines near IFSAR cues

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 is measured in terms of the overlap between the roof hypotheses and the IFSAR cue regions. The hypotheses are projected onto the cues image and the overlap of the projected roof with IFSAR regions is computed. The current system requires that the overlap be at least 50%. Figure 10 shows the selected hypotheses in the Ft. Hood example.



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

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 required presence of shadow evidence and wall evidence. Details of the shadow and wall analysis are given in [hidden]. 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 and are analyzed to give a set of final building

hypotheses. These are shown in Figure 11, with and without 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.

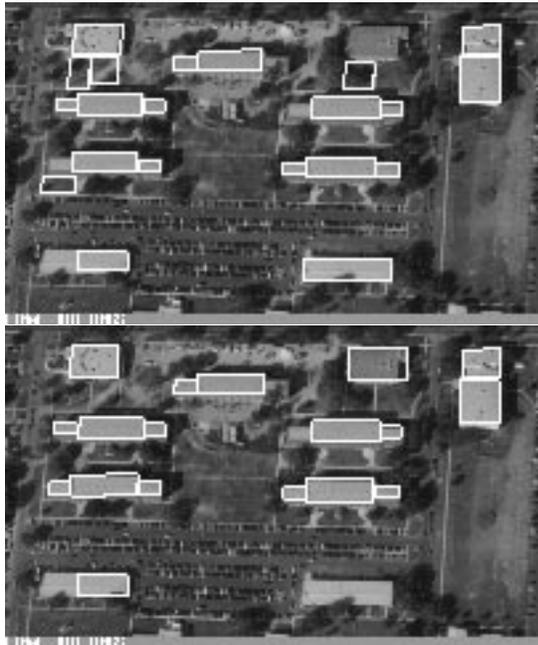


Figure 11. Final hypotheses using PAN only (top) and using IFSAR cues (bottom)

Figure 12 shows the combined detected flat and gable-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.

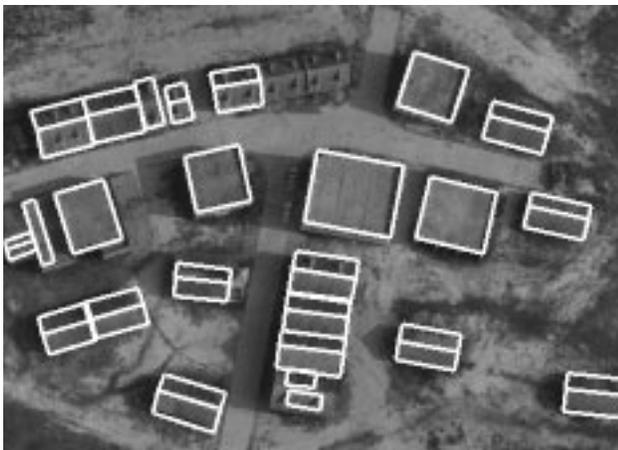


Figure 12. Extracted using IFSAR cueing

5 System Evaluation

Tables 1 and 2 give comparisons of the number of features and final result component counts with and without use of IFSAR cues for the Fort Hood and Ft. Benning examples respectively.

TABLE 1. Ft. Hood Automatic Result

Feature	PAN Only	With IFSAR
Line Segments	7799/7754/5083 (from three images)	
Linear Structures	2959/2963/2042	669/650/552
Flat Hypotheses	2957	1732
Selected Flat	383	296
Verified Flat	215	192
Final Flat	21 (4 false)	18 (0 false)

TABLE 2. Ft. Benning Automatic Result

Feature	PAN Only	With IFSAR
Line Segments	16400/18998 (from two images)	
Linear Structures	55827/6611	1758/2041
Flat Hypotheses	5218	3012
Selected Flat	329	192
Verified Flat	202	116
Final Flat	22	15
Gable Hypotheses	634	240
Selected Gable	181	75
Verified Gable	181	75
Final Gable	19	17
Combined	41	32
Buildings	29 (3 false)	25 (0 false)

To characterize the increase in performance of the system when IFSAR cues are available we use two basic metrics (see [15] for details), detection rate and false alarm rate, as follows:

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

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

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. The first level of evaluation is to measure the detection and false alarm rates at the object levels such as for buildings or wings of a complex building. We consider each rectangular part of a rectilinear building as a separate object. A building object will be considered to be detected, if any part of it has been detected. The reference model for our Ft. Benning example shown in Figure 13. The reference

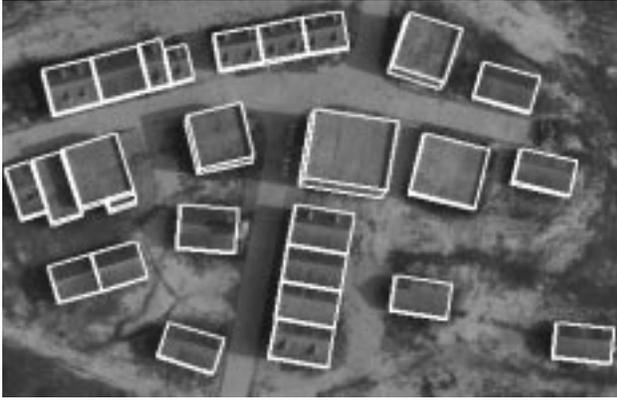


Figure 13. Reference model for evaluation

model for the Ft. Hood example has been omitted for lack of space. Tables 3 and 4 show summaries of detection and false alarm results for the Ft. Hood and Ft. Benning examples, respectively, in terms of object parts. Note that false alarms disappear when IFSAR cues are available.

TABLE 3. Ft. Hood Component Evaluation

	PAN only	With IFSAR
Reference Model		21
Detected Components	21	18
TP	18	18
FP	4	0
FN	2	1
Detection Rate	0.90	0.95
False Alarm Rate	0.18	0.00

TABLE 4. Ft. Benning Component Evaluation

	PAN only	With IFSAR
Reference Model		27
Detected Components	29	25
TP	26	25
FP	2	0
FN	1	2
Detection Rate	0.96	0.92
False Alarm Rate	0.07	0.00

To better reflect the quality of the detected components we also compute the accuracy the overlap between the footprints of the detected and the reference models and in the overlap between the 3-D volume occupied by them.

The area (volume) elements of the reference model that overlap with some area (volume) element of an extracted model can be considered to give the true positive (TP) values for the area (volume) elements of the reference model (the remaining elements of the reference models are the false negatives, FN). The area (volume) elements of the extracted model that do not overlap with any area (volume) element of the reference model give us the false positives (FP) for the area (volume) elements of the extracted model.

One way to combine the results of the above area (or volume) overlap analysis is to consider each area element as an object and count the detection and false alarm rates for all the area elements in the models. Tables 5 and 6 show these results for the Ft. Hood and Ft. Benning examples respectively. Ground detection rate is computed for the ground area elements (all elements that are not part of other objects); ground false alarm rate are not shown.

TABLE 5. Ft. Hood Combined Area Evaluation

	PAN 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

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

To better characterize the accuracy, we compute the detection rates for the area elements of each reference building component and the false alarm rates for each extracted building component separately. To visualize the result we compute a cumulative distribution of the detection and false alarm rates. Specifically, we can compute the percentage of building components of the reference model whose area (volume) elements detection rate (TP) is at a give value or higher. A curve plotting such a distribution is called a CDR curve [15]; Figure 14a shows the CDR curve for area elements of our Ft. Benning example. Similarly, we can compute the percentage of the building components of the extracted model whose false alarm rate (FP) is at a given value or lower. A curve plotting such a distribution is called a CFR curve; Figure 14b shows the CFR curve for the area elements of our Ft. Benning example. We also compute CDR and CFR curves for the volume elements for the reference and extracted building components. These are shown in Figure 15. A CDR curve that is consistently higher than another CDR curve indicates consistently better performance (similarly, a CFR curve that is consistently lower is consistently better).

The CFR and CDR curves for our Ft. Hood example have been omitted due to lack of space. These however, present similar characteristics as the ones shown.

6 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

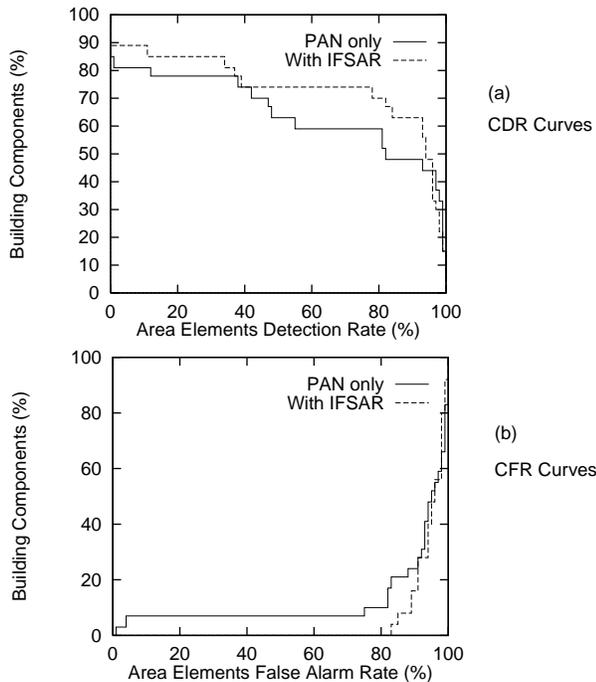


Figure 14. Curves for area elements

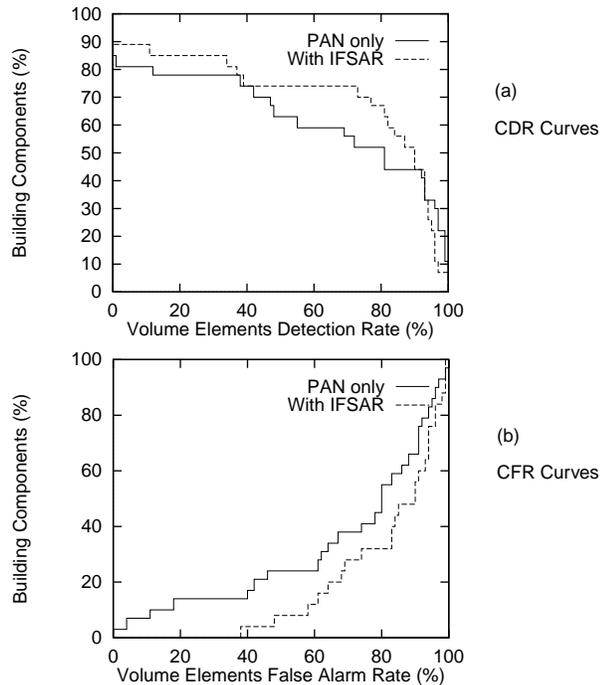


Figure 15. Curves for volume elements

missing elements and artifacts, it has been shown that it can be used to enhance the results of PAN 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 sensors of different modalities become available for more domains.

References

- [1] M. Roux and D. McKeown, "Feature Matching for Building Extraction from Multiple Views". *Proc. DARPA Image Understanding Workshop*, Monterey, CA, pp. 331-349, 1994.
- [2] S. Noronha and R. Nevatia. "Detection and Description of Buildings from Multiple Aerial Images", *Proc. IEEE CVPR*, San Juan, PR, pp. 588-594, 1997.
- [3] 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*, vol 72, no. 2, pp 143-162, 1998.
- [4] A. Gruen, E. Baltsavias, and O. Henricsson (Editors), *Proceedings of the Ascona Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images II*, Brinkhauser Verlag, Switzerland, May, 1997.
- [5] A. Gruen and R. Nevatia (editors), *Computer Vision and Image Understanding: Special Issue on Automatic Building Extraction from Aerial Images*, vol. 72, no. 2, 1998.
- [6] F. Stetina, J. Hill and T. Kunz, "The Development of a Lidar Instrument for Precise Topographic Mapping," *Proc. International Geoscience and Remote Sensing Symposium*, Pasadena, CA, 1994.
- [7] J. Curlander and R. McDonough, "Synthetic Aperture Radar", *Wiley Interscience*, New York, 1991.
- [8] 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.
- [9] K. Hoepfner, C. Jaynes, E. Riseman, A. Hanson, and H. Schultz, "Site Modeling using IFSAR and Electro-Optical Images", *Proc. DARPA Image Understanding Workshop*, New Orleans, LA, pp 983-988, 1997.
- [10] R. Chellapa, Q. Zheng, S. Kuttikkad, C Shekhar, and P. Burlina, "Site Model Construction for the Exploitation of E-O and SAR Images", *Proc. RADIUS97*, Morgan Kaufmann, San Francisco, pp 185-208, 1997.
- [11] N. Haala and C. Brenner, "Interpretation of Urban Surface Models using 2D Building Information", *Automatic Extraction of Man-Made-Objects from Aerial and Space Images II*, Brinkhauser, Basel, pp 213-222, 1997.
- [12] J. Chen, A. Huertas, and G. Medioni, "Fast Convolution with Laplacian-Of-Gaussian Masks," *IEEE Trans. PAMI*, 9(4), pp 584-590, 1987.
- [13] 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*, vol 72, no. 2, pp 185-203, 1998.
- [14] 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*, vol 712, no. 2, pp 122-142, 1998.
- [15] R. Nevatia. "On Evaluation of 3-D Geospatial Modeling Systems," in *ISPRS Proceedings of the International Workshop on 3D Geospatial Data Production*, Paris, France, April, 1999.