The MURI Project for Rapid Feature Extraction in Urban Areas*

R. Nevatia, A. Huertas and Z. Kim
Institute of Robotics and Intelligent Systems
University of Southern California
Los Angeles, CA 90089-0273, USA
nevatia@usc.edu
http://iris.usc.edu
Commission III, Working Group 3

Keywords: Sensor integration, 3-D Modeling, Cartography

Abstract

A summary of research in urban feature extraction being conducted under a MURI (multi-disciplinary research initiative) program sponsored by the U.S. Army Research Office is provided. Most of the paper focuses on extraction of rectilinear building models. Feature extraction from aerial images is difficult due to problems of segmentation, 3-D inference and shape description. A system for building modeling from multiple panchromatic images is described. Enhancements to this system using digital elevation models (DEMs) or thematic maps derived from hyperspectral images are described. Some results and quantitative evaluations are provided. A brief example of work on street grid extraction is also included. The work described here depends on analysis provided by researchers at Purdue University and Marconi Integrated Systems, Inc. who are partners in this multi-disciplinary project.

Zusammenfaßung

In diesem Beitrag wird eine Zusammenfaßung der Forschungsaktivitäten im Bereich der Extraktion von typischen städtischen Strukturen gegeben. Diese Arbeiten stehen im Rahmen des MURI-Programms (Multi-Disciplinary Research Initiative), von das Amt für Forschung der U.S. Streitkräfte. Dieser Beitrag behandelt größtenteils die Extraktion von rechteckigen Gebaeudemodellen. Die Extraktion solcher Strukturen aus Luftbildern gestaltet sich wegen verschiedenster Probleme im Zuge der Segmentierung, 3D-Schlußfolgerung und Formbeschreibung der Objekte als schwierig. Ein System zur Modellierung und Extraktion von Gebaeuden aus multiplen Bildern wird beschrieben. Die erzielten Ergebniße werden quantitativ bewertet. Es wird gezeigt, daß Verbeßerungen der Ergebniße durch die Integration von Digitalen Oberflächenmodellen oder thematischen Karten, die von hyperspektralen Daten abgeleitet wurden, erzielt werden. Außerdem wird in aller Kürze ein Beispiel zur Extraktion von gitterförmigen Straßennetzes beschrieben. Die hier dargelegten Arbeiten stammen aus der Zusammenarbeit mit Forschern der Purdue University und Marconi Integrated Systems Inc., die Partner in diesem fachübergreifenden Projekt sind.

1 Introduction and Overview

Three-D *site models* of cultural features, such as roads and buildings, are required for a number of applications for civilian and military uses. However, construction of such site models from available sensor data remains an expensive, slow and tedious task. A Multidisciplinary University Research Initiative (MURI) project was initiated by the U. S. Army Research Office (ARO) to make advances in methods for rapid and affordable feature extraction in urban areas by increasing the degree of automation in detection and description of the various kinds of features. This paper describes some of the activities in this project.

Our project team is multi-disciplinary. It includes researchers from three institutions: Purdue University, University of Southern California (USC) and Marconi Integrated Systems, Inc. The team includes researchers with background in photogrammetry, in analysis of hyperspectral imagery, in computer vision and in system construction for applications. This paper will focus on computer vision oriented research at USC; however, this research relies critically on other components of the project which will be mentioned in this context. USC projects focus on two main tasks. One is the modeling of rectilinear shaped buildings. The other is extraction of city road grids. As the second project is described in a separate paper in these proceedings, only a brief summary of it is provided in this paper.

The problem of 3-D feature extraction has many sources of difficulties including those of segmentation, 3-D inference and shape description. Segmentation is difficult due to the presence of large number of objects that are not intended to be modeled such as sidewalks, landscaping, trees and shadows near the objects to be modeled. The objects to be modeled may be partially occluded and contain significant surface texture. 3-D information is not explicit in an intensity image; its inference from multiple images requires finding correct corresponding points or features in two or more images. Direct ranging techniques such as those using LIDAR or IFSAR can provide highly useful 3-D data though the data typically has areas of missing elements and may contain some points with grossly erroneous values.

* This research was sponsored in part by the U.S. Army Research Office under grant No. DAAH04-96-1-0444. The views expressed in this paper do not necessarily reflect the views of the U.S. Government or the U.S. Army Research Office.

Once the objects have been segmented and 3-D shape recovered, the task of shape description still remains. This consists of forming complex shapes from simpler shapes that may be detected at earlier stages. For example, a building may have several wings, possibly of different heights, that may be detected as separate parts rather than one structure initially.

The approach used in this effort is to use a combination of tools: reconstruction and reasoning in 3-D, use of multiple sources of data and perceptual grouping. Context and domain knowledge guide the applications of these tools. Context comes from knowledge of camera parameters, geometry of objects to be detected and illumination conditions (primarily the sun position). Some knowledge of the approximate terrain is also utilized. The information from sensors of different modalities is fused not at pixel level but at higher feature levels.

Our building detection system is based on a "hypothesize and verify" paradigm. This system can function with just a pair of panchromatic (PAN) images, but can also utilize more images and information from other modalities. This system also incorporates abilities for Bayesian reasoning and machine learning. We first describe the basic multi-view system, then the Bayesian reasoning and learning methodology followed by a discussion of our methods for achieving sensor fusion

2 Multi-view System

A number of systems that use multiple views have been described in the literature [Jaynes et al., 1997; Collins et al., 1998; Roux & McKeown, 1994]. The system described in this paper, called **MVS**, derives from an earlier system described in [Noronha & Nevatia, 1997]; a block diagram is shown in Figure 1. The approach is basically one of hypothesize and verify. *Hypotheses* for potential roofs are made from fragmented lower level image features. The system is hierarchical and uses evidence from all the views in a non-preferential, order-independent way. Promising hypotheses are *selected* among these by using relatively inexpensive evidence from the rooftops only. The selected hypotheses are then *verified* by using more reliable global evidence such as from walls and shadows. The verified hypotheses are then examined for overlap which may result in either elimination or in merging of them.

This system is designed for rectilinear buildings; complex buildings are decomposed into rectangular parts. Rooftops thus project to parallelograms in the images (the projection is nearly orthographic over the scale of a building). Lines, junctions and parallel lines are the basic features used to form roof hypotheses. Consider the image shown in Figure 2. The images i s from the Ft. Hood, Texas, site that has been in common use by many researchers. The low level features composed of lines, junctions between lines and sets of parallel lines are matched among the available views. In this example, two views were used (only one shown). The set of lines extracted from the image (using a Canny edge detector) to start the process are shown in Figure 3. Roof hypotheses are formed by a pair of matched parallel lines and ***U*** structures (**U**s represent three sides of a parallelogram). A pair of parallel lines may be matched to parallels in more than one view (when more than two views are used) and each matching pair is considered. Closed hypotheses are formed from these features by using the best available image lines if any, else closures are synthesized from the ends of the parallel lines.

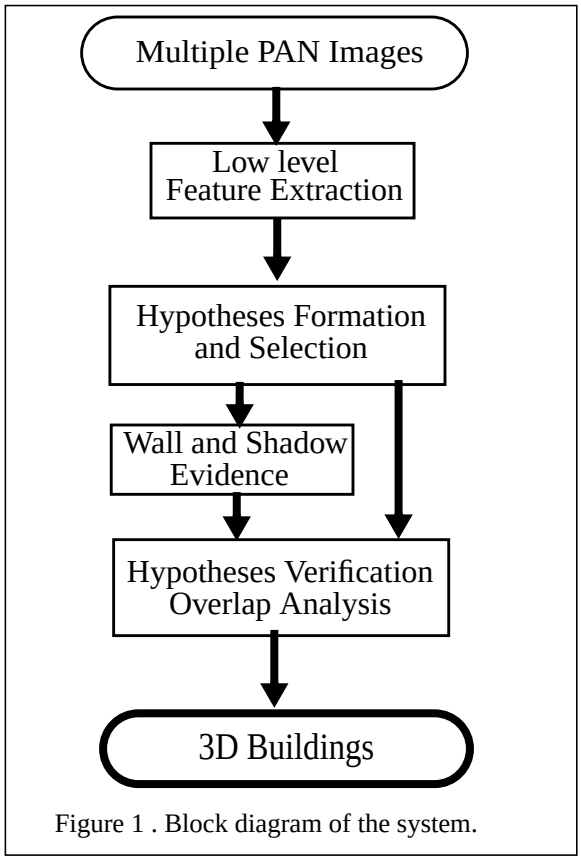


Figure 1 . Block diagram of the system.

Figure 2 A Portion of an image from Ft. Hood, Texas

Three-D roof hypotheses could be inferred from the 2-D hypotheses by using line matches; however, the line matches are often not necessarily unique. Instead, an estimate for the heights of roof lines is made by conducting a search. For a flat roof only a single height needs to be determined, for the symmetric gables we need to find two heights. For each height estimate, the corresponding 3-D hypotheses is projected in each other view and line evidence for each projection is computed. The heights with the best evidence are selected. The hypothesis formation process is rather liberal and a large number of hypotheses are typically formed at this stage. A

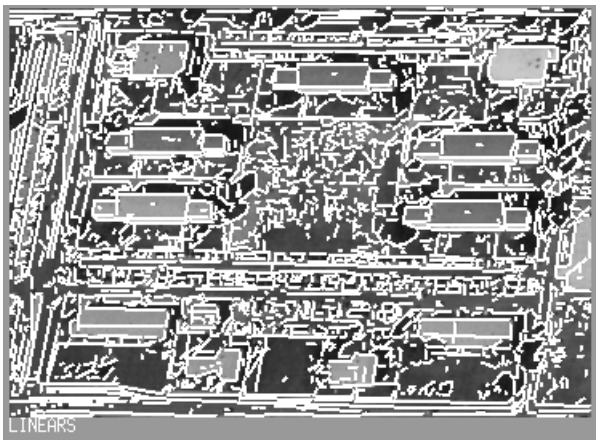


Figure 3 Linear segments from image in Figure 2

smaller set is *selected* using the underlying image evidence for the roof hypotheses. Positive evidence comes from lines near the projected hypotheses, negative evidence comes from lines crossing the hypotheses. The hypotheses selected in our Ft. Hood example are shown in Figure 4.

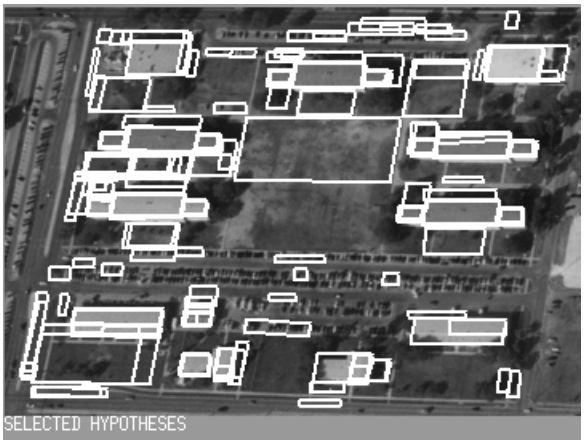


Figure 4 Selected hypotheses from image in Figure 2.

The next step is to *verify* whether the selected hypotheses have additional evidence for corresponding to being buildings. This evidence is collected from the roof, the walls and the shadows that should be cast by the building. Since the hypotheses are represented in 3-D, deriving the projections of the walls and shadows cast, and determining which of these elements are visible from the particular view point is possible. These in turn guide the search procedures that look in the various images for evidence of these elements among the features extracted from the image. A score is computed for each evidence element.

Each of the collected evidence parameters is composed of smaller pieces of evidence. A critical question is how to combine these small pieces of evidence to decide whether a building is present or not and how much confidence should be put in it. Results shown in this paper use a Bayesian reasoning approach as described later in section 3.

After verification, several overlapping verified hypotheses may remain. Only one of the significantly overlapping hypotheses is selected. The overlap analysis procedure examines not only the evidence available for alternatives but also separately the evidence for components that are not common.

Figure 5 shows the wireframes of the detected buildings from the pair of images. Note that while most of the buildings are detected correctly, some are missing. The missing components have some underlying support (as indicated by the hypotheses shown in Figure 4) but lack verifying evidence.

These results can be improved by observing that several of the roofs in the image have very similar *appearance*. We collect roof intensity statistics for the verified rooftops and find other hypotheses that are *similar* to them (by using Kolmogrov-Smirnov similarity measures). Figure 5 shows the results after this process.

Figure 5 3-D wireframe model of detected buildings

The performance of the system can be expressed quantitatively by computing detection and false alarm rates for the building components in comparison to a reference model, such as shown in Figure 6. The performance rates for our Ft. Hood example are shown in table 1. For this analysis, a component is considered detected if any part of it is detected. Evaluation of accuracy of models using the method suggested in [Nevatia, 1999] is given later in section 5.

Figure 6 The reference model for evaluation.

Table 1: MVS Performance

	Components
Reference Model	26
True Positives	20
False Positives	2
False Negatives	6
Detection rate	0.769
False alarm rate	0.09

3 Bayesian Reasoning and Learning

A system such as the one described above, makes many binary decisions at various stages, such as which hypotheses to generate, which to select and which to retain as being verified. These decisions need to be based on evidence that is inherently incomplete and uncertain. Some ad hoc methods, such as a linear combination of evidence followed by thresholding are commonly used. These methods are not necessarily optimal and require considerable effort in adjusting parameters for good performance on a selected set of examples.

We have focused on use of Bayesian methods as they are well-founded and give optimal results, given the evidence and the assumptions. However, they require knowledge of joint probability distributions which can be of high dimensionality and hence hard to obtain. One approach to reducing the dimensionality is to assume that the various evidence variables are conditionally independent of each other given a hypothesis, this leads to a so-called *naive* Bayes classifier. The needed probabilities in this case can be simply calculated from direct observation of examples and evidence variable values. Naive Bayes classifiers give surprisingly good results but have limited representational power and are hard to generalize when applied to multi-image and multi-sensor systems.

More accurate models of the joint distribution can be made by loosening the conditional independence assumption given a hypothesis. For example, we can expect conditional independence among each of the shadow evidence given the strong overall shadow evidence but not necessarily given a presence of the building because sometimes the overall shadow evidence of a building can be poor when there is an occlusion for example. Such conditional independence is well represented in a Bayesian network [Pearl, 1988].

Applying a Bayesian network approach to building verification process requires estimating the desired topology of the network and to estimate the parameters associated with it (the conditional probability tables, or CPTs) as well as an augmentation for the multi-image situation.

In dealing with multiple images, some of the evidence variables are view-dependent (such as the shadow evidence for a non-stereo pair) while some are not (evidence from other sensors). This makes it difficult to apply the traditional Bayesian network directly because the number of the view-dependent evidence variables are not fixed but depend on the number of views. Instead, we have developed a dynamic Bayesian network (DBN) which is an augmented Bayesian network with *duplicable* nodes. When an input data set is supplied, a duplicable node is dynamically instantiated into multiple copies according to the number of the respective input data. The implementation is rather simple, because we can easily make an instantiation according to the input values. However, for MVS, instantiation is not necessarily to be done for every input data because we only have two or three different instantiations for hundreds of dataset. Thus, once an instantiation is obtained, we reuse it later for the same configuration of input data. Figure 7 shows an example of DBN.

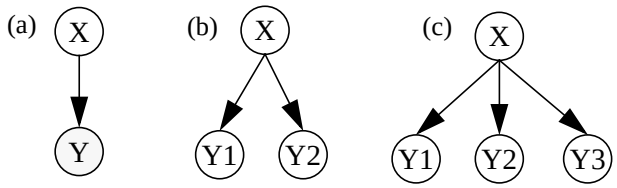


Figure 7 (a) A simple DBN with two nodes. Y is a duplicable node; (b) an instantiation when Y has two values; (c) an instantiation when Y has three values

We infer the topology of the network by a combination of statistical analysis and knowledge of the causal structure of the evidence. First, we obtain the correlations of each evidence variables given buildings. Then, for the set of evidence variables with high correlation, we add a *hidden* (unobservable) node to force the conditional independence assumption. Table 2 summarizes the evidence used in MVS. Some of them have continuous values while the others are either binary or ternary. Figure 8 shows a Bayesian network constructed based on the correlation analysis.

Table 2: Evidence used in MVS.

Category	**Evidence**
Roof	Positive lines (RP), Negative lines (RN), Standard deviation (RS)
Shadow	Strong junctions (SS), Weak junctions (SW), Horizontal lines (SH), Vertical lines (SV), Darkness (SD)
Wall	Vertical lines (WV), Base lines (WB)

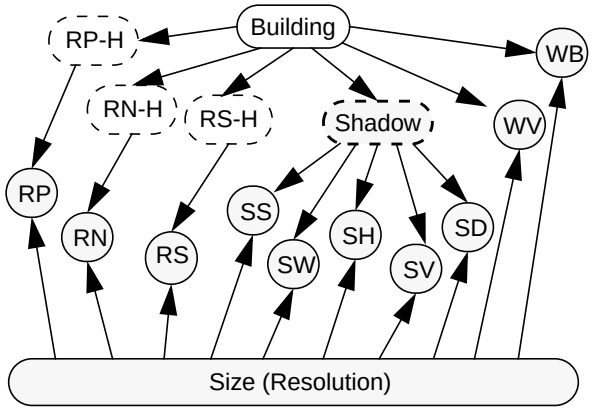
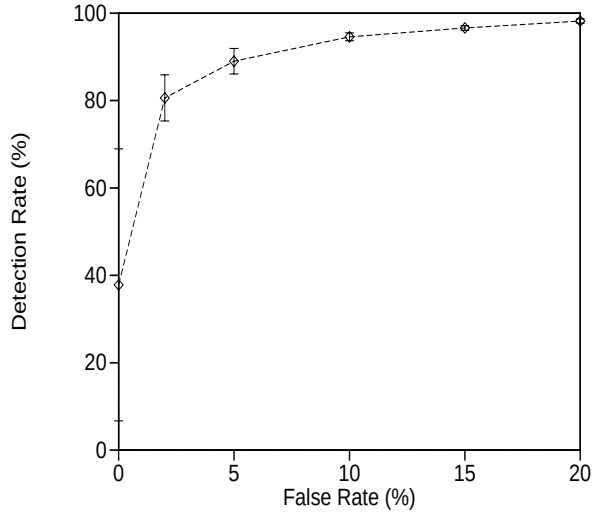


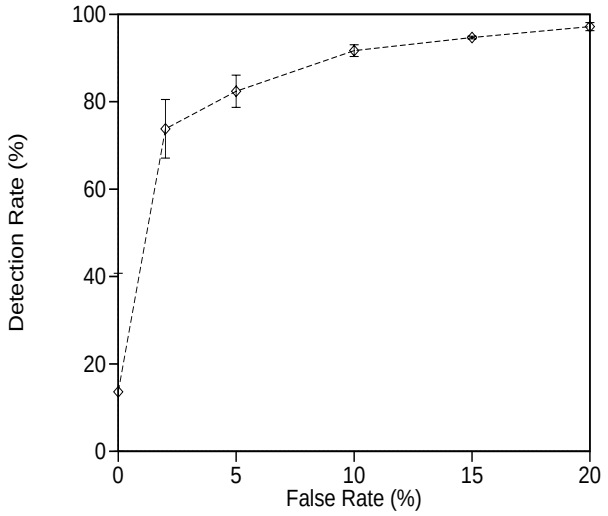
Figure 8 A Bayesian network for MVS constructed based on correlation analysis.

Learning CPTs is much more complex than for a naive Bayes classifier because some of the nodes are hidden. We applied a gradient ascent approach similar to that of [Binder et al., 1997] for the hidden nodes. The resulting learned network demonstrates good performance.

We characterize the results of the Bayes network by plotting an ROC curves illustrating the trade-off between false alarm and detection rate (measured at object component level). We selected 970 hypotheses from several images of two sites and used half of them for training and the other half for testing. Figure 6 shows the ROC curves for the training and test sets. It can be seen that the performance on the two sets is quite similar and that high detection rates (over 80%) can be obtained at relatively low false alarm rates (less than 5%).



(a)



(b)

Figure 9 An ROC curve on (a) training data set, and (b) test data set.

4 Integration of Multi-sensor Information

The performance of the building detection and description system can be greatly improved if information from other sensors become available. As described above, our system can take advantage of multiple panchromatic (PAN) images even if they are not acquired at the same time. We consider two other sources of a different modality.

The first source of additional information provides range (*i.e.* height) information in the form of digital elevation models (DEMs). DEMs may be derived from stereo PAN images or acquired directly by active sensors such as LIDAR or IFSAR. The second source of information is from multi- or hyper-spectral imagery, such as from the HYDICE sensor, which is becoming increasingly more available.

DEMs makes the task of building detection much easier as the buildings are significantly higher than the surround and accompanied by sharp depth discontinuities. However, DEM data is not always accurate, particularly near the building boundaries and the active sensors may contain significant artifacts. The spectral information makes it easier to decide if two pixels belong to the same class, and hence to the same object, or not. However, objects, such as building rooftops, are not always homogeneous in material and the hyper-spectral data is usually of a significantly lower resolution than that of PAN images. For these reasons, we have decided to use DEM and spectral sensors to provide *cues* for the presence of buildings but to use PAN images for accurate delineation.

Figure 10 shows a block diagram of our approach. The left most column denotes the multi-view system described in section 2 above. If DEM data is available, object cues are extracted from it and supplied to MVS where this information can be used to aid in the process of hypothesis formation and selection. Similarly, HYDICE data is analyzed to produce thematic maps which again aid in the process of hypothesis formation and selection for MVS. These processes are described in some detail below.

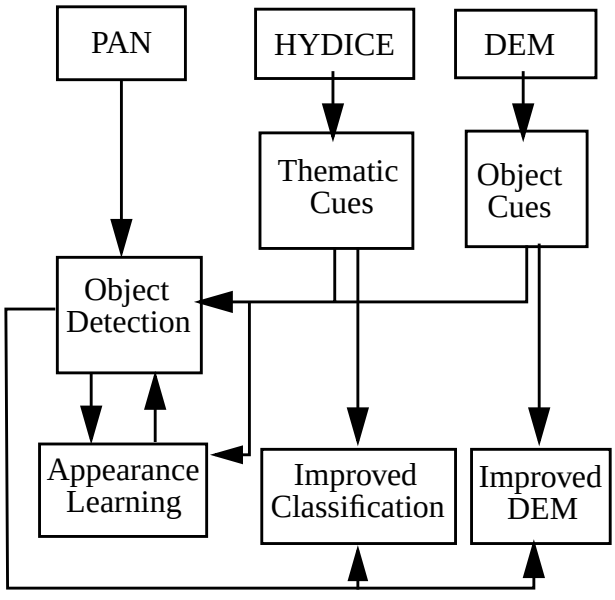


Figure 10 Multi-sensor information integration

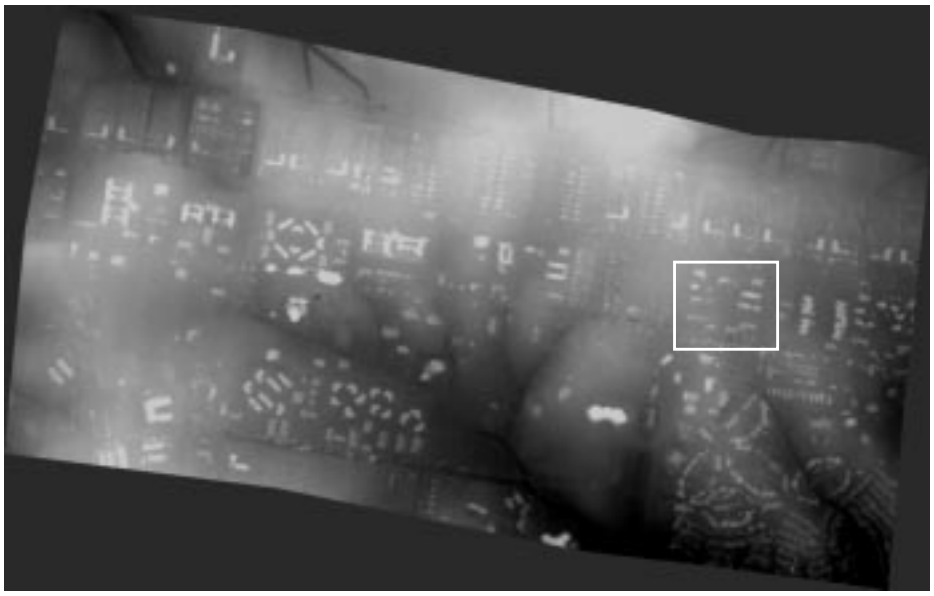


Figure 11 DEM of portion of the Ft. Hood site, constructed at 1 meter postings by Marconi, Inc.

4.1 Use of DEM data

DEM data can be available from active sensors. We have experimented with use of IFSAR images [Nevatia & Huertas, 1998; Huertas et al., 1998]. In this paper, we will focus on use of DEMs derived from PAN images themselves. These DEMs, made from sets of pairs of images, are available to us from Marconi Integrated Systems, Inc., a partner in the MURI project by using methods available in the commercial "SocetSet" product. One such DEM for a large portion of the built up areas in Ft. Hood is shown in Figure 11. A small part of the DEM for the Ft. Hood site, corresponding to the area shown earlier in Figure 2, is shown in Figure 12 (displayed intensity is proportional to elevation.) Note that while the building areas are clearly visible in the DEM, their boundaries are not smooth and not highly accurate. These characteristics prevents direct extraction of buildings from DEM images but clearly can help cue the presence of 3-D objects.

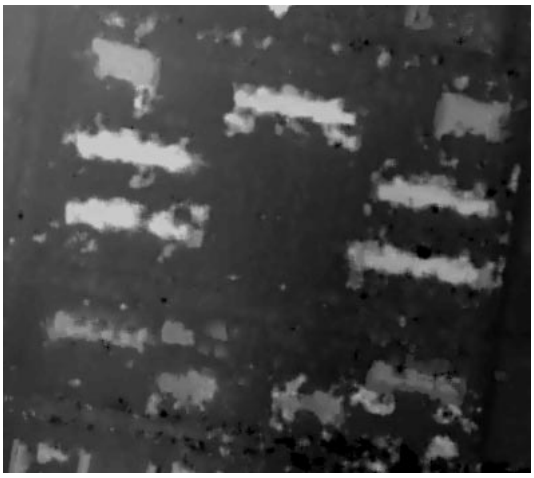


Figure 12 . DEM corresponding to image in Figure 2

The building regions in a DEM are characterized as being higher than the surround. However, simple thresholding of the DEM is not sufficient, as height variations of the magnitude of a single story building can occur even in very flat terrain sites. Our approach is to convolve the image with a Laplacian-of-Gaussian filter that smooths the image and locates the object boundaries by the positive-valued regions bounded by the zero-crossings in the convolution output. Such regions derived from Figure 12 are shown in Figure 13. The cue regions shown correspond to connected components having of a certain minimum size, and can include objects such as trees.

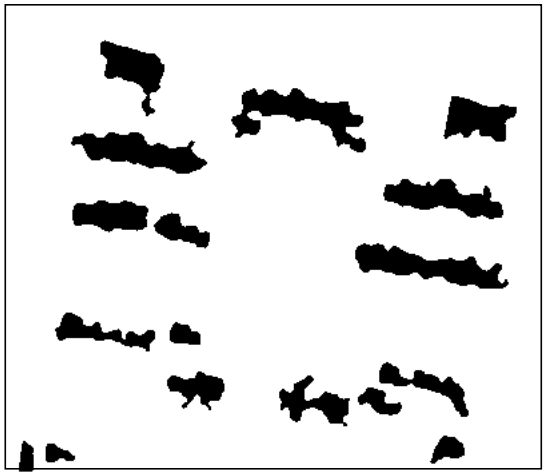


Figure 13 Cues from DEM in Figure 12.

The DEM image has the geometric characteristics of an orthographic projection. The estimation of the sensor parameters, or "camera" model, associated with this overhead (nadir) viewpoint is straightforward. This 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 DEM image. We use these transforms to project PAN 2-D

and 3-D features onto the DEM image to assist and support the building detection system at various stages of processing. We describe in more detail, and illustrate these processes, with our Ft. Hood example, below.

Object cues are used in several ways and at different stages of the hypotheses formation and validation processes; they 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. The 3-D location of a line segment in the 2-D PAN images is not known. To determine whether a line segment is near a DEM cue region we project the line onto the cue image at a range of heights, and determine if the projected line intersects a cue region. Figure 3 earlier showed the line segments detected in the image of Figure 2; Figure 14 shows the lines that lie near the DEM cues. As can be seen, the number of lines is reduced drastically (81.5%) by filtering without loosing any of the lines needed for forming building hypotheses. 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 14 . Lines near DEM cues

We also use these cues to help select and verify promising hypotheses, or conversely, to help disregard hypotheses that may not correspond to objects. Just as poor hypotheses can be discarded because they lack DEM 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. When no evidence of walls or shadows is found, we require that the DEM evidence (overlap) be higher, to validate a hypotheses. The 3-D Models constructed with DEM support from the validated hypotheses are shown in Figure 15. For comparison, the model shown earlier in Figure 5 was derived without DEM support. Note that false detections are eliminated with DEM cueing. Also, the building components on the top left and on the lower part are not found without DEM support but found with it.

Once the buildings have been detected, the DEM can also be

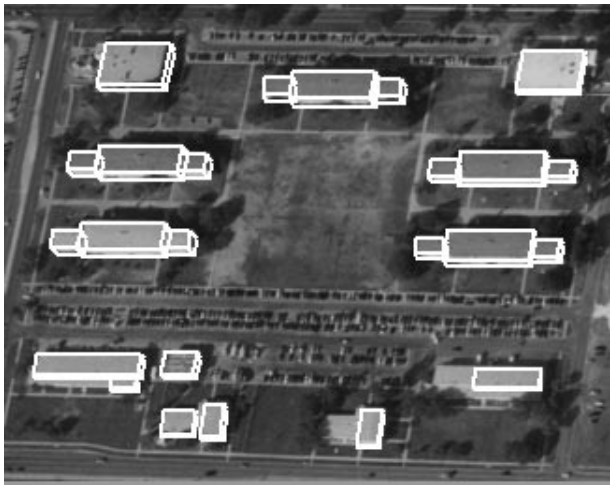


Figure 15 . Building components extracted using DEM cues.

improved by replacing parts of the DEM with building models as shown in Figure 16.

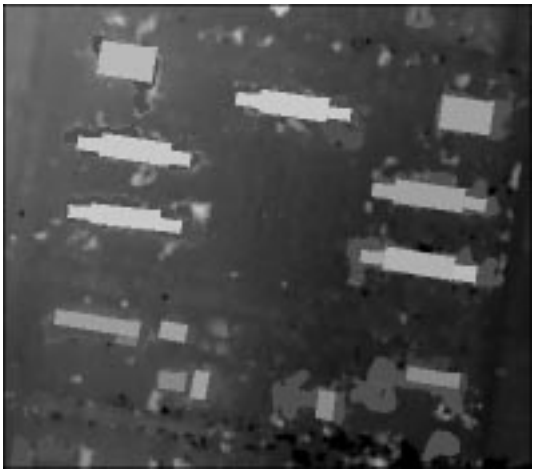


Figure 16 . Corrected DEM.

Table 3 gives a summary of the performance of the system with DEM support. Compared to results summarized in Table 1 earlier, four additional components are detected and there are no false detections. The components still missing are low, small, partially occluded by trees and have limited DEM support.

Table 3: MVS Performance with DEM support

	Components
Reference Model	26
True Positives	24
False Positives	0
False Negatives	2
Detection rate	0.93
False alarm rate	0.0

4.2 Use of HYDICE Data

The HYDICE (HYperspectral Digital Imagery Collection Experiment) sensor collects data of 210 bands over the range 500 to 2500 nm with a field of view of 320 pixels wide at an IFOV (pixel size) of 1 to 4 m depending on the aircraft altitude and ground speed. Rather than merging this data directly with the PAN images, we first derive a thematic map from an HYDICE image which provides cues for building modeling as in the case of DEM processing above.

A thematic map is constructed by associating a ground cover label to each pixel in the image. Given the multivariate nature of HYDICE data, the process of data analysis is one of dividing up the N-dimensional feature space into M exhaustive but non-overlapping regions where M is the number of classes of materials existing in the scene. The process involves defining the M classes of interest in a quantitative fashion, such that each pixel in the scene, which exists as a discrete location in the N-dimensional space, can be uniquely associated with one of the M classes. Frequently, this is done by using a small number of samples in the scene, called design samples or training samples, to define an N-dimensional probability density function for each of the M classes. Then an unknown pixel can be evaluated in terms of the likelihood of each possible class to determine the most likely class membership. This classification procedure has been developed by Prof. Landgrebe and his associates at Purdue University. More details may be found in [Huertas et al., 1999; Lee & Landgrebe, 1993; Jimenez & Landgrebe, 1998; Landgrebe, 1999].

In order to integrate cues extracted from HYDICE data into the building detection and description system we require that the thematic map be rectified and registered to the PAN imagery. Geometric rectification is needed to correct for the oscillations and "waviness" introduced by the nature of the HYDICE pushbroom sensor. Rectification is performed on the thematic map rather than on the HYDICE data directly. The method utilizes ground control points and control linear features typically found in urban scenes together with the pushbroom sensor model and a gauss-markov platform model to yield coordinate relationships between ground and image spaces (see [Huertas et al., 1999; Lee, et al. 1999] for details). The accuracies achieved are in the 0.5 to 1 pixel range. Figure 17 shows a geometrically rectified thematic map corresponding to a portion of the Ft. Hood site. The "waviness" of the boundary of the image gives an indication of the needed correction.

The corrected thematic map, just as the DEM images, has the geometric characteristics of an orthographic projection and the estimation of the sensor parameters associated with this 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 thematic map. As with the DEM assisted processes, we use these transforms to project PAN 2-D and 3-D features onto the thematic map to assist and support the building detection system at various stages of processing.

To extract cues from the thematic map we first extract the roof pixels from the thematic map. These are shown in Figure 18. Many pixels in small regions are misclassified or correspond to objects made of similar materials as the roofs. The building cues extracted from this image are the connected components of certain minimum size. These components are shown in Figure 19; Except for one region, these components correspond to building roofs.

HYDICE cues are used, in ways similar to those for the DEM cues described above, at different stages of the hypotheses formation and validation processes. The linear segments near HYDICE cues, are very similar to those shown earlier in Figure 14 with an increased reduction in the number of lines (84%). As with the DEMs the HYDYCE evidence helps simplify the hypotheses selection process. The evidence consists of support of a roof hypotheses in terms of the overlap between the roof hypotheses and the HYDICE cue regions. The hypotheses are constructed from matching features in multi-

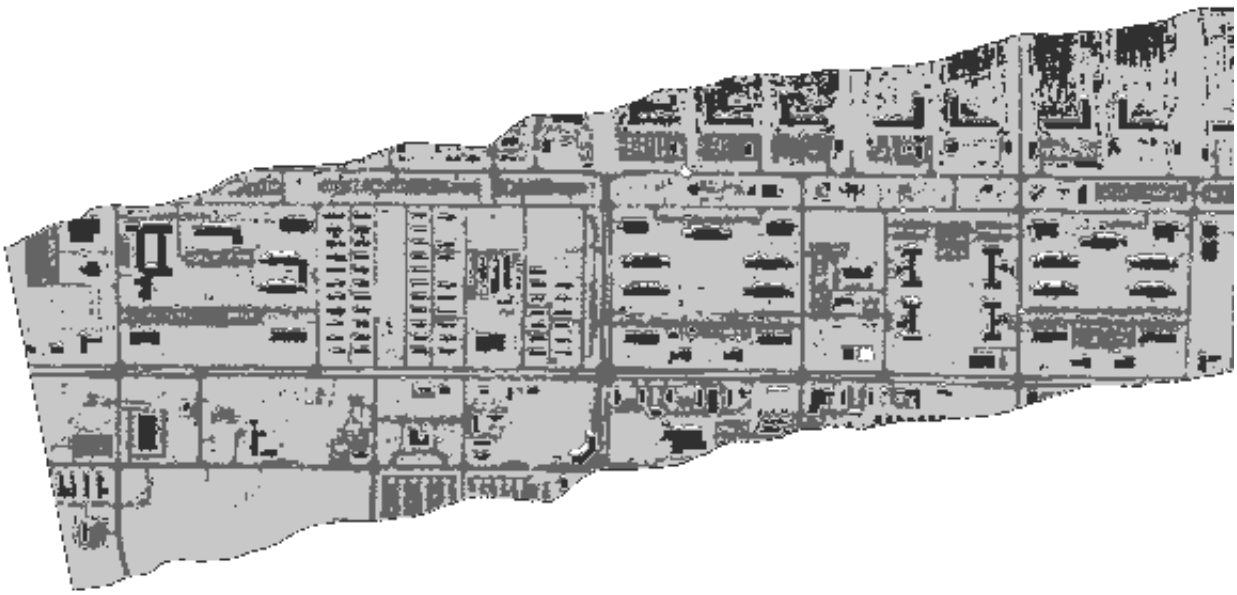


Figure 17 Geometrically corrected thematic map of Ft. Hood. Thematic classification and correction provided by Purdue University

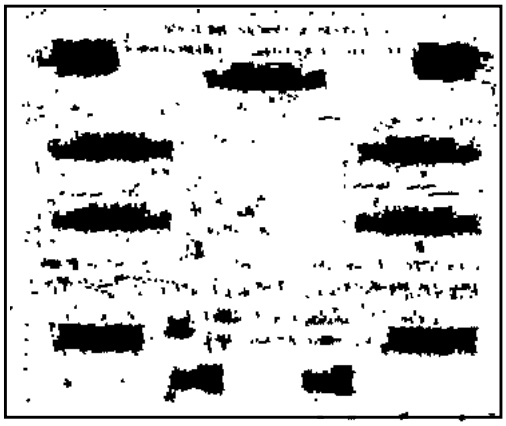


Figure 18 Roof class from thematic map.

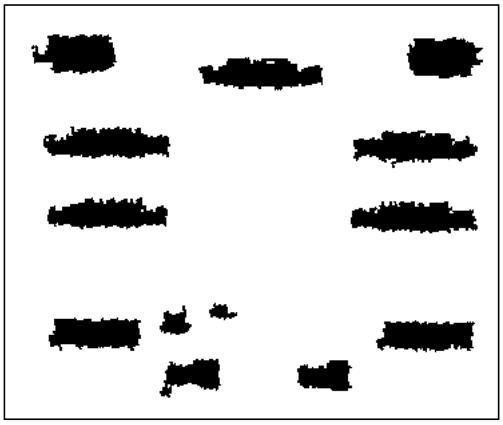


Figure 19 . Hood Cues.

ple (two in our Ft. Hood example) images and are represented by 3-D rectilinear components in 3-D world coordinates. We can therefore project them directly onto the HYDICE cues image to compute roof overlap (See Figure 20). The system requires that the overlap be at least 50% of the projected roof area.

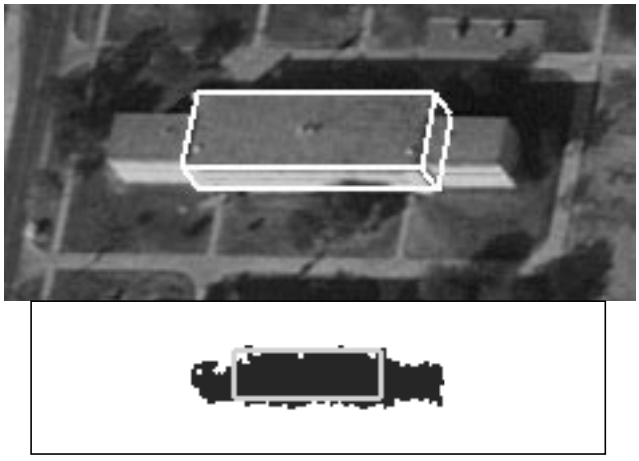


Figure 20 . A 3-D hypotheses projected on PAN image (top) and on cue image.

Figure 21 shows the detected buildings using the HYDICE cues. This result shows no false alarms. Once the buildings have been detected, the roof class can also be updated as shown in Figure 22. The performance of the MVS system is very similar using DEM or HYDICE cues. There will be many cases where the quality of the cues from one sensor may be higher. It is appropriate to characterize this quality and combine the support from various sensor modalities. This is the subject of our current work.

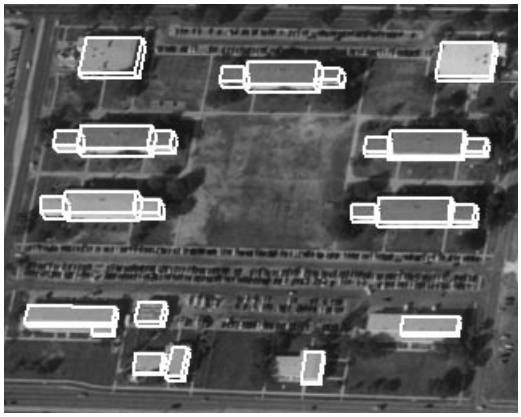


Figure 21 . Building components extracted using HYDICE cues.

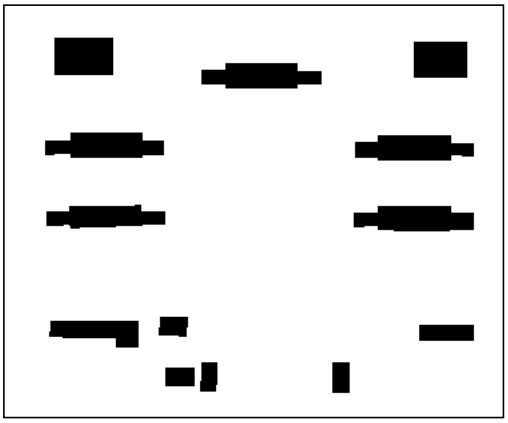


Figure 22 . Updated “roof” class.

5 System Evaluation

We have developed a methodology for evaluation of 3-D geospatial building modeling systems [Nevatia, 1999]. To characterize the performance of the system detection and false alarm rates are computed in terms of true positives, false positives and false negatives. 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. Features could be objects, area elements or volume elements. 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, such as shown earlier in Table 2 and Table 3.

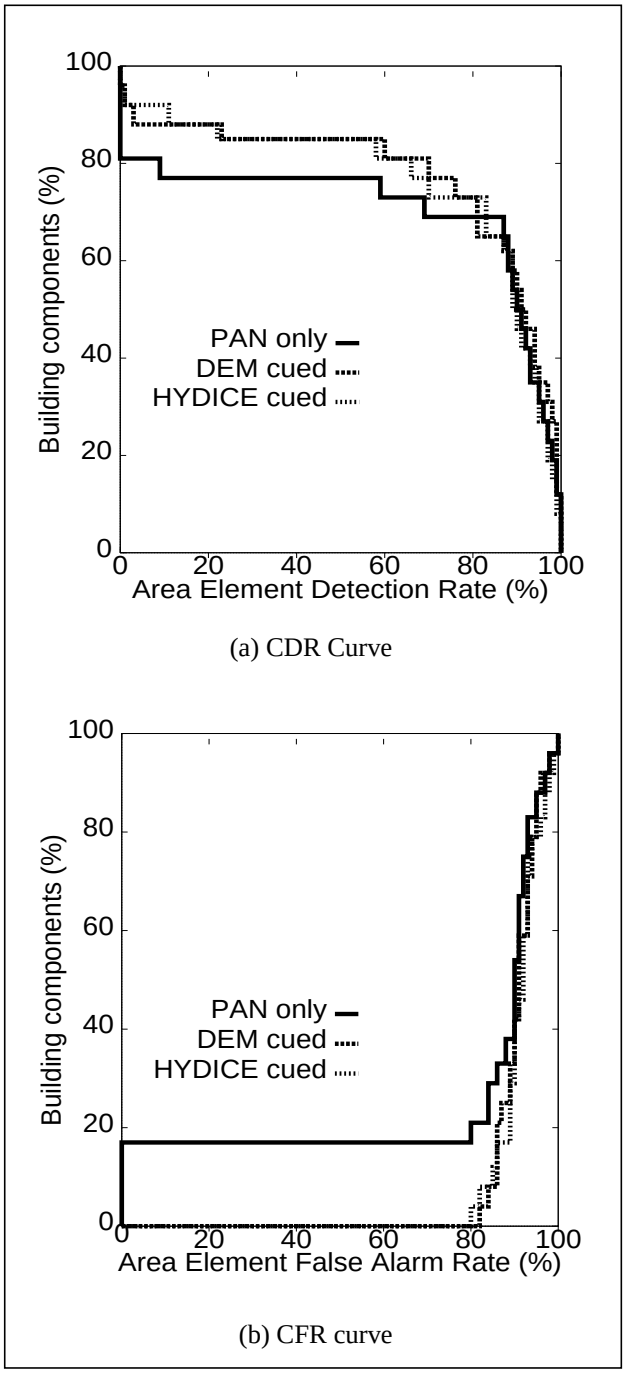
To asses geometric accuracy of the detected components we also compute the accuracy the overlap between the footprints of the detected and the reference models and in the overlap between the 3-D volume occupied by them. The area (volume) elements of the reference model that overlap with some area (volume) element of an extracted model can be considered to give the true positive values for the area (volume) elements of the reference model (the remaining elements of the reference models are the false negatives. 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 (false positive) 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. Table 4 shows these results for our Ft. Hood example. Ground detection rate is computed for the ground area elements (all elements that are not part of other objects); ground false alarm rate is not shown.

Table 4: Combined Area Evaluation

	PAN Only	with DEM	with HYDICE
Detection rate	0.7116	0.8436	0.8453
False Alarm rate	0.1510	0.0929	0.0768
Ground Detection rate	0.9819	0.9888	0.9907

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 (true positive) is at a give value or *higher*. A curve plotting such a distribution is called a CDR curve [Nevatia, 1999]. We compared the reference model shown earlier in Figure 6 with results for our Ft. Hood example with and without using DEM or HYDICE cues. Figure 23a shows the CDR curve for area elements of our Ft. Hood example. Similarly, we can compute the percentage of the building components of the extracted model whose false alarm rate (false positives) is at a given value or *lower*. A curve plotting such a distribution is called a CFR curve; Figure 23b shows the CFR curve for the area elements of our Ft. Hood example. We also compute CDR and CFR curves for the volume elements for the reference and extracted building components. 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).



(a) CDR Curve

(b) CFR curve

Figure 23 . Evaluation curves for *area* analysis.

6 Road Grid Extraction and Verification

Another area of research in our MURI project deals with he the extraction of street grids. A detailed paper describing this work [Price, 1999] is available in these proceedings, and thus we only provide a summary here.

We describe a system for extracting regular street grids in a moderately dense urban environment (streets are not significantly obscured by buildings). Rather than just marking road pixels in the image, we generate descriptions of extended streets including the intersections connecting the streets. This complements both the work in low resolution road extraction and high resolution road following, since that work usually ignores intersections and is not applied in urban areas. See for example [Fischler & Heller, 1998; McKeown & Denlinger, 1988].

The street grid is rarely present in isolation. Roads may have trees occluding them or buildings and other structures adjacent to them. Thus low level segmentation techniques such as edge detection or region segmentation, give fragmented results with many extraneous boundaries and regions. We address these difficulties by using model based extraction and

Figure 24 Streets detected in the Mall area of Washington D.C

grouping procedures, global and local context from extended roads and intersections, high resolution imagery, and fusion of multiple images or multiple data sources.

A result shown in Figure 24. The detected streets are shown projected onto an orthophoto of the Mall area in Washington D.C. This orthophoto only covers most of the area of three of the images used for the extraction so some extracted roads are displayed off the image. The topology of the network is also available. For each street, all intersecting streets can de identified. Approximately 46 Km of streets are extracted.

7 Summary and Future Work

We have provided an overview of the research being conducted in the MURI project on rapid and affordable extraction of urban feature data. A central theme of this effort is integration of information from different sources and from different sensors. As we have shown, this can result in significant improvements in the quality of the resulting models while also reducing the computational requirements.

We have illustrated how our methods can be used for building modeling where the buildings are relatively simple in shape and are not very close to each other as may be the case in a denser urban environment. To deal with such complexities, we plan to exploit the multi-sensor information in much richer ways. For example, DEMs can provide not only cues for the presence of a building component but should also allow for rough segmentation of surfaces in it. Multi-spectral data may not provide such segmentation but should allow for inference of surface material properties.

8 Acknowledgements

The research described in tis paper is made possible by the efforts of a multi-disciplinary team. We acknowledge the help of Professors J. Bethel, D. Landgrebe and E. Mikhail from Purdue University, and of J. Spann, M. Vriesenga and B. Zhang from Marconi Integrated Systems in conduct of this research and in providing us with several of the data sources and analysis procedures used in this work.

References

[Binder et al., 1997] J. Binder, D. Koller, S. Russell, and K. Kanazawa, “Adaptive Probabilistic Networks with Hidden Variables,” Machine Learning, Vol. 29, pp. 213-244, 1997.

[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”, To appear in Computer Vision and Image Understanding Journal, special issue on Building Detection and Reconstruction. R. Nevatia and A. Gruen, Editors, 1998.

[Fischler et al., 1998] M. Fischler, B. Bolles and A. Heller. “APGD Evaluation Metrics, Methodology, Rationale” SRI International, May 1998, http://www.ai.sri.com/~apgd

[Firshler & Heller, 1998] M. Fischler and A. Heller, “Techniques for Road Network Modeling,”, Pro-

ceedings DARPA Image Understanding Workshop, Monterey, CA, Nov. 1998, pp. 501-516.

[Grün & Nevatia, 1998] Computer Vision and Image Understanding Journal, Special Issue on Automatic Building Extraction from Aerial Images. A. Grun and R. Nevatia, editors. Academic Press, Vol. 72, No. 2, November.

[Höpfner et al., 1997] K. Höpfner, 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.

[Huertas et al., 1998] A. Huertas, Z Kim and R. Nevatia, "Use of Cues from Range Data for Building Modeling", Proceedings DARPA Image Understanding Workshop, Monterey, CA, Nov. 1998, pp. 577-582.

[Huertas et al., 1999] A. Huertas, R. Nevatia and D. Landgrebe, "Use of Hyperspectral Data with Intensity Images for Automatic Building Modeling," Proceedings 2nd IEEE International Conference on Information Fusion, July 1999, Sunnyvale, CA, to appear.

[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.

[Jaynes et al., 1997] C. Jaynes, M. Marengoni, A. Hanson, E. Riseman and H. Schultz, "Knowledge Directed Reconstruction from Multiple Aerial Images", Proceedings of the DARPA Image Understanding Workshop, New Orleans, LA, May 1997, pp 971-976.

[Jimenez & Landgrebe, 1998] L. Jimenez and D. Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," IEEE Transactions on System, Man, and Cybernetics, Volume 28, Part C, No. 1, pp. 39-54, February 1998.

[Landgrebe, 1999] D. Landgrebe, Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data, Chapter 1 of Information Processing for Remote Sensing, edited by C. H. Chen, World Scientific Publishing Co., Inc., River Edge, NJ

[Lee et al., 1999] C. Lee, H. Theiss, J. Bethel, and E. Mikhail, "Rigorous Mathematical Modeling of Airborne Pushbroom Imaging Systems." to appear in the Photogrammetric Engineering and Remote Sensing Journal, 1999.

[Lee & Landgrebe, 1993] C. Lee and D. Landgrebe, "Analyzing High Dimensional Multispectral Data," IEEE Transactions on Geoscience and Remote Sensing, Volume 31, No. 4, pp 792-800, July 1993.

[Li et al., 1999] J. Li, S. Noronha and R. Nevatia, "User Assisted Modeling of Buildings", Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, to appear.

[Maloof et al., 1998] M. Maloof, P. Langley, A. Kamal, A. S. Sage and T. Binford (1998) "Improving Rooftop Detection with Interactive Visual Learning," Proceedings of the DARPA Image Understanding Workshop, Monterey, California, pp. 479-492.

[Maloof et al., 1998] M. Maloof, P. Langlay and R. Nevatia, "Generalizing over Aspect and Location for Rooftop Location", IEEE Workshop on Applications of Computer Vision. Princeton, NJ. October, 1998.

[McKeown et al., 1997] D. McKeown, et. al. "Research in the Automated Analysis of Remotely Sensed Imagery: 1995:1996", Proceedings DARPA Image Understanding Workshop, New Orleans, LA, May 1997, pp. 779-812

[McKeown & Denlinger, 1988] D. McKeown & J. Denlinger, "Cooperative Methods for Road Tracking in Aerial Imagery," Proceedings of the Conference on Computer Vision and Pattern Recognition, 1988, pp. 662-672.

[Nevatia, 1999] 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.

[Nevatia & Huertas, 1998] R. Nevatia and A. Huertas, "Knowledge-Based Building Detection and Description: 1997-1998," Proceedings DARPA Image Understanding Workshop, Monterey, CA, Nov. 1998, pp. 469-478.

[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.

[Pearl, 1988] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc., 1988.

[Price, 1999] K. Price, "Road Grid Extraction and Verification," these proceedings.

[Roux & McKeown, 1994] M. Roux and D. McKeown, "Feature Matching for Building Extraction from Multiple Views", Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 46-53.