

need to be characterized as a function of these variables.

3.3 Demonstration Plan

Our demonstrations will consist of taking input images, and any available collateral data, and displaying the results of our APGD algorithms and to compare them, in some cases, with hand provided *ground-truth* results. We expect to demonstrate results on a variety of objects using a variety of imagery sources. The system will be designed to run autonomously but some human interaction, either to initiate the tasks, or to edit the results may be allowed.

We will also aid in integrating our system with the system to be developed by the APGD IFD contractor and demonstrate our systems in a larger context. We intend to develop our software using the RCDE environment which should simplify integration with the IFD contractor.

References

- [Chung & Nevatia, 1992] C.-K. R. Chung and R. Nevatia, "Recovering LSHGCs and SHGCs from Stereo," In *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, January 1992, pp. 401–407.
- [Heuel & Nevatia, 1996] S. Heuel and R. Nevatia, "Including Interaction in an Automated Modeling System," in *Proceedings of Image Understanding Workshop*, Palm Springs, CA, February 1996, pp. 429-434.
- [Huertas & Nevatia, 1996] A. Huertas and R. Nevatia, "Including Interaction in an Automated Modeling System," in *Proceedings of Image Understanding Workshop*, New Orleans, LA, May 1997.
- [Huertas, *et al.*, 1990] A. Huertas, W. Cole, and R. Nevatia, "Detecting Runways in Complex Airport Scenes," *Computer Vision, Graphics, and Image Processing*, 51(2):107–145, August 1990.
- [Huertas, *et al.*, 1995] A. Huertas, M. Bejanin and R. Nevatia. "Model Registration and Validation", in *Proceedings of the Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, Birkhauser, March 1995, pp 33-44.
- [Lin *et al.*, 1995] C. Lin, A. Huertas, and R. Nevatia, "Detection of Buildings from Monocular Images", in *Proceedings of the Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, pp 125-134, March 1995.
- [Medioni & Nevatia, 1984] G. Medioni and R. Nevatia, "Matching Images Using Linear Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):675–685, November 1984.
- [Medioni & Nevatia, 1985] G. Medioni and R. Nevatia, "Segment-Based Stereo Matching," *Computer Graphics and Image Processing*, 31(1):2–18, July 1985.
- [Mohan & Nevatia, 1989a] R. Mohan and R. Nevatia, "Perceptual Organization for Segmentation and Description," in *Proceedings of the DARPA Image Understanding Workshop*, Palo Alto, California, May 1989. Morgan Kaufmann Publishers, Inc.
- [Mohan & Nevatia, 1989b] R. Mohan and R. Nevatia, "Segmentation and Description Based on Perceptual Organization," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 333–341, San Diego, California, June 1989.
- [Mohan & Nevatia, 1989c] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3-D Structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139, November 1989.
- [Noronha & Nevatia, 1997] S. Noronha and R. Nevatia, "Detection and Description of Buildings from Multiple Aerial Images," in *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, May 1997

cation of a certain building and in predicting locations of other buildings. Extraction of a road and river provides cues to the location of a bridge and so on.

Domain knowledge also helps us choose the tools that are appropriate for the task and in choosing parameters or rules for the algorithms. Different settings or methods may be appropriate for processing in rural or urban areas, or in areas with or without heavy vegetation, or in presence or absence of snow.

e) Use of Previous Models and Maps

In many cases of interest, previous maps and models may exist and the task may be one of updating or extending them. In this case, we need to register the new images with the existing model, find the differences between the two and update the models with descriptions of the new and changed features. The process of finding the differences consists of computing the expected visible features (for the current imaging environment) and verifying whether these features are present in the image. Regions of significant difference can invoke the model construction process. Some capabilities of image to model registration and model validation have been developed as part of our RADIUS effort [Huertas, *et al.*, 1995; Huertas & Nevatia, 1997]

f) Human Interaction

Even though the goal of this effort is complete automation, it is likely that the systems that can be developed in the near term will not be perfect and will miss some objects or find incorrect ones. A mechanism is needed to edit and correct them. This should not require a user to invoke completely manual procedures; in many cases, it is sufficient for the user to provide some hints to the automatic system to recompute and correct the problems. We have some experience with such an approach where in some cases a missed building can be found simply by the operator indicating the approximate location of the building and a possible cause of the failure [Heuel & Nevatia 1996]. Sometimes, more precise interaction may be needed, but it should still not be necessary to revert to a complete manual system. For example, if the size of the roof is corrected by the user, the height can be recomputed automatically using the same procedures as the automatic extraction system. This work is reported on in more detail in [Huertas & Nevatia 1997].

3 Evaluation Plan

We are developing relatively complete, end-to-end systems that start with images (and some collateral data when available) and produce 3-D object models. This makes it easier to establish evaluation metrics and to test the systems. We describe some metrics and an evaluation methodology below.

3.1 Metrics

The following metrics capture issues in evaluating extraction results:

- 1) **Detection rate:** How often are the desired features detected? This can be in terms of the absolute number of detected objects or may be by some weighting (such as by size or by importance). We consider a feature to be detected, if there is any overlap between a detected feature and a desired feature (this could be modified to include a certain amount of overlap or a certain minimum accuracy of the model).
- 2) **False Alarm rate:** This measures the frequency of mistaken detection. A feature is considered a false alarm if it does not overlap with any desired feature (of the detected class). Again, the rate may be measured in terms of number of objects or by some weighting.
- 3) **Accuracy of Models:** This is more difficult to measure. We can measure errors in 2-D or 3-D. Typically we want to know the accuracy in terms of size, shape and location. Size error can be computed in terms of volumes or by other parameters such as area and height. Shape differences may be harder to characterize, except perhaps by the amount of overlap (in 2-D or 3-D). The error metric can be made specific to a shape, for example, for a rectangular structure, measurements of the three sides and the center may suffice.
- 4) **Confidence Factor:** We expect that our systems will be able to assign confidence factors to the detected features (and even to components of these features if necessary). These could be included as modifications to the above measures. For example, a false alarm indicated with lower confidence could be counted as being less severe than one with a high confidence.

3.2 Testing

For evaluations to be meaningful, the system must be tested on a wide variety of images that contain a variety of desired objects in a variety of environments, and imaged under a variety of conditions (possibly with a variety of sensor types). The results



Figure 4 Buildings detected by multi-view hierarchical system.

grams under orthography (generally applicable when the sensor is relatively far compared to the size of the object). Next, we show an example of building detection using multiple images. We have developed a system that combines matching and grouping operations in a hierarchical fashion [Noronha & Nevatia, 1997]. Figure 4 shows the results of this procedure on a portion of the Ft. Hood dataset using three separate views. This is one of the most difficult parts of the Ft. Hood scene; trees obscure some of the sides and buildings have roofs at multiple heights. Note that there are no false alarms and most of the buildings are detected and delineated correctly.

In past work, it has been common to use geometrical properties. With the availability of multiple

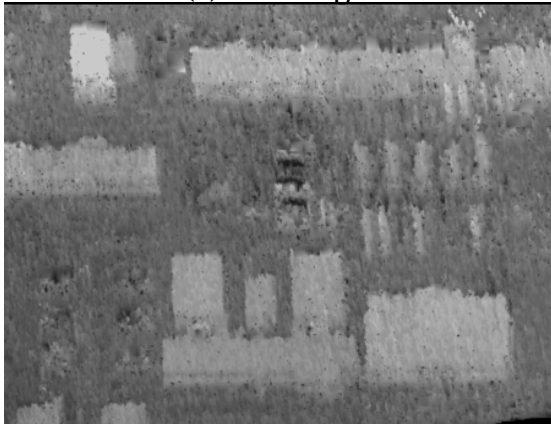
sources, grouping will need to use features from multiple images and combine them depending on the sensor characteristics. Combinations using geometric properties will be easier than combining sensor level data. We will also need a method to accumulate evidence of objects from a variety of uncertain sources.

d) Context and Domain Knowledge

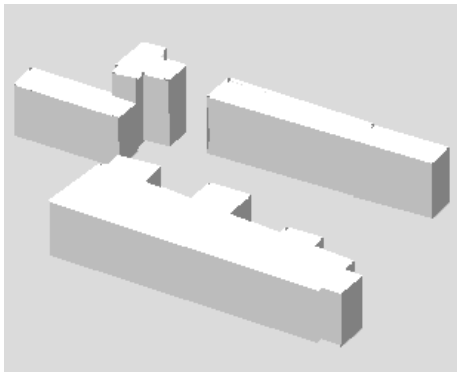
Context and domain knowledge are important sources of information for object extraction. Presence of one set of objects can help reinforce or suggest the presence of others. For example, in an airport complex, the hypotheses for an airplane and a terminal reinforce each other if proper relations exist between them. Extraction of a certain road, or a transportation network in general, helps in identifi-



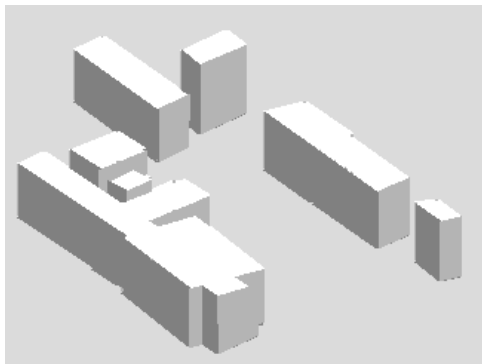
(a) EO Image



(b) Smoothed IFSAR image



(c) Buildings using IFSAR alone



(d) Buildings using EO cued by IFSAR

Figure 3 Building extraction using EO and IFSAR

can aid in the processing in many important ways. They can be useful in detecting vegetation and surface material and may help in distinguishing between natural and cultural features. They can be useful for distinguishing shadow regions (shadow regions have different spectral properties, in a normal color image they typically have a higher blue component from scattered light). Hyperspectral images should also aid in segmentation of images as different kinds of materials can be clustered by spectral properties more easily than by intensity alone. IR images may also be useful for detecting vegetation and some components of a 3-D structure.

c) Perceptual Grouping

Much of the information about an object resides in its geometric properties. However, the geometry of the object needs to be inferred from the image. Early segmentation processes typically give low-level features that are fragmented and the desired object features are not separated from those in the background or those due to surface markings, shadows or noise. The goal of perceptual grouping is to organize these features into meaningful groups that give the desired object geometry. Use of certain kinds of sensors may provide better connected features with fewer distracting ones, but the process of perceptual organization is still required.

Our approach to perceptual grouping is a hierarchical one. Lower level features, such as lines, are grouped into successively higher levels, such as parallel or symmetric lines, which in turn are grouped into features that may correspond to surfaces of desired objects. The surfaces may then be further grouped to give volumetric objects. Multiple hypotheses are possible at each level. Our approach is to select among them only at levels where sufficient information is available to do so. This results in many hypotheses being generated at each level. Also, the grouping process may use either 2-D or 3-D features, but the final *verification* step for the objects should use 3-D reasoning wherever possible.

The properties that are used for grouping and for selecting among possible groups are of key importance. We believe that these properties should derive from an analysis of expected invariances for classes of objects under various imaging conditions. For example, we know that surfaces of a rectangular parallelepiped will project to parallelo-

entation discontinuities and separate them from image discontinuities caused by sources such as markings, shadows and highlights.

Some sensors, such as IFSAR, provide direct measurement of height, however, this information may be incomplete and may contain significant isolated errors. Height can also be extracted from multiple 2-D images by making correspondences between features visible in two or more images. The process of finding correspondences is, however, a difficult task in itself. One key issue is the level of features at which this matching should be performed. Low level features are easier to compute but more ambiguous. Higher level features are easier to match but difficult to infer reliably from image data. We use a hierarchical approach where features are matched at various levels and the matching at one level helps compute features at higher levels. We have constructed an early version of a hierarchical multi-image system [Noronha & Nevatia, 1997].

b) Use of Multiple Sources

Multiple sources of data can help in many ways. One is in estimating heights even if each sensor only gives a 2-D image. We can also take advantage

of other characteristics of the sensors as the different sensors may be better at extracting different kinds of information. For example, IFSAR images have height information. In an ideal case, certain kinds of features, such as a flat horizontal roof, may be found just by thresholding on height. Surface boundaries may be found by first derivative discontinuities in the image. However, IFSAR data is likely to contain *holes* and some of the values may have gross errors. Thus, results obtained by simple processing are not likely to be of sufficient quality for accurate building models. We believe that IFSAR images will be highly useful for detection, but detailed description and delineation may require use of supplementary sensor data.

This is illustrated in Figure 3. The results using IFSAR alone are obtained by thresholding height to get roof locations and the EO results are derived by our monocular building detection system with the approximate location given by the IFSAR roof areas; these results are shown only to indicate the potential of combining sensor modalities and not to imply these problems are solved.

Multispectral and hyperspectral images, if available

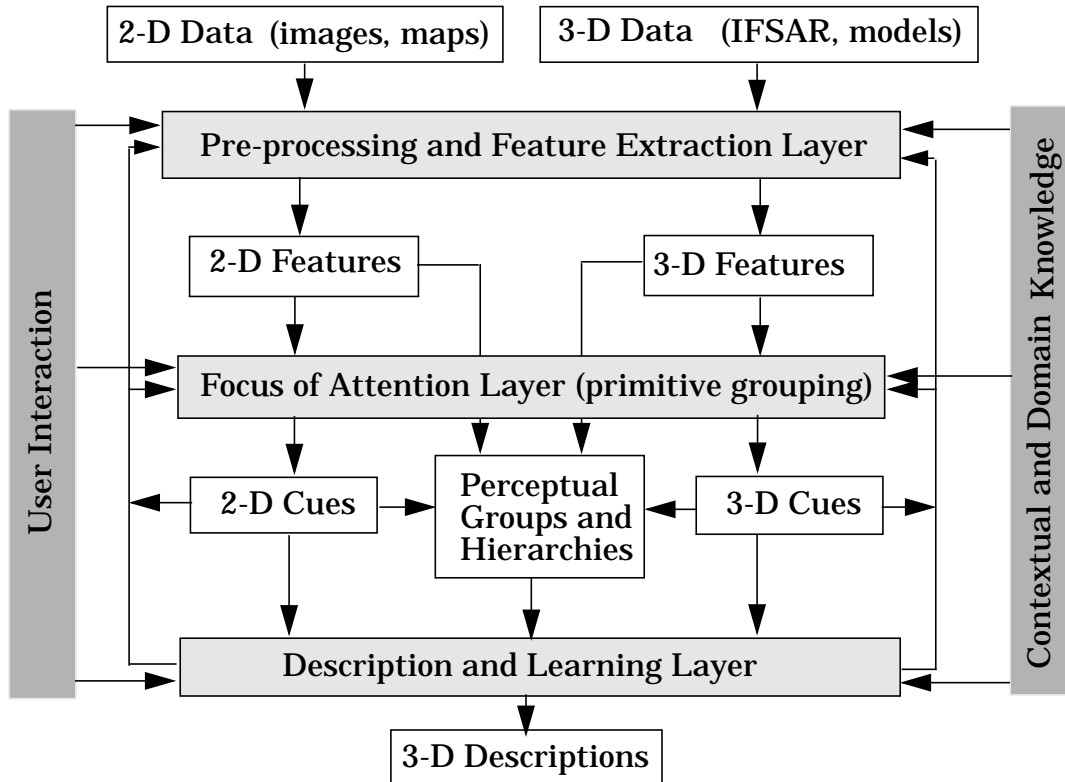


Figure 2 Schematic Diagram of the 3-D Grouping Approach

shadows, noise and other features, may be present as can be seen in the example of Figure 1. The system must complete the desired object boundaries and discard the extraneous ones. It must also infer the 3-D structure of the objects, which is not explicit in an intensity image and needs to be inferred.



Figure 1 Image (top) and extracted line segments. (bottom)

The segmentation and 3-D description problems become easier if a 3-D range sensor, such as LADAR or IFSAR is available, as image discontinuities correspond more directly to object or surface bound-

aries. However, some extraneous boundaries due to noise and other sources would remain. Also, sensors such as IFSAR can not provide a complete range map, typically the data has areas of missing elements and may contain some points with grossly erroneous values.

In the absence of an ideal range sensor, we can infer depth from multiple images. However, this requires finding corresponding points or features in two or more images, which is a complex task in itself. Also, such a process gives sparse information about 3-D features which must still be grouped appropriately to extract the desired objects.

We plan to overcome the difficulties of 3-D object detection and description by using a combination of tools: reconstruction and reasoning in 3-D, use of multiple sources of data, perceptual grouping, use of context and domain knowledge, use of previous maps and models, and limited use of human input where applicable. This combination is shown schematically in Figure 2. Reasoning in 3-D has many advantages, as the objects we desire are 3-D objects rather than 2-D surface features. The use of multiple images and multiple sources of data will aid in the problems of 3-D reconstruction and also in resolving ambiguities that may be present in a single image or in images taken from a single sensor. Perceptual grouping is an essential step in selecting and organizing lower level features into meaningful objects. The use of context and domain knowledge will help with the reduction of ambiguities, with help in attribution and in some cases help with choosing the appropriate algorithm parameters. In cases where previous maps or models exist, these can be used to reduce the needed work; instead of building complete models, the system can focus on change detection and updating. Even with use of these multiple tools, some human interaction may be required, either to initiate the automatic processes correctly or to edit the results. We do this when necessary but minimize the effort required from the human operator.

We now describe the elements of our approach in more detail:

a) Reconstructing and Reasoning in 3-D

Explicit 3-D representations are needed for many applications. Knowledge of 3-D also makes the task of segmentation easier as we can find depth and ori-

Knowledge-Based Automatic Feature Extraction

Ram Nevatia and Keith Price

Institute for Robotics and Intelligent Systems
University of Southern California
Powell Hall Room 204, MC-0273
Los Angeles, California 90089-0273
<http://iris.usc.edu/Outlines/apgd-project.html>

Abstract

Constructing geospatial databases is a tedious manual operation. Automatic 3-D feature extraction from 2-D images requires solving a number of problems. We present a plan to attack this task using a combination of tools: reconstruction and reasoning in 3-D, use of multiple sources of data, perceptual grouping, use of context and domain knowledge, use of previous maps and models, and limited use of human input where applicable.

1 Objectives

Geospatial databases are important for a number of battlefield awareness tasks such as mission planning, mission rehearsal, tactical training and damage assessment. Other applications include intelligence analysis for site monitoring and change detection. Geospatial database requirements may vary for the different tasks, but generally knowledge of terrain elevation, surface features and cultural features is needed. Our task focuses on the automatic detection and description of cultural features, particularly buildings.

The importance of cultural features for multiple tasks is quite clear. A mission plan in urban or semi-urban environments must consider buildings and similar structures. These may be the targets of an operation or assets to be utilized in the mission. Road networks and buildings are also key components in a simulated scene for such environments. Damage assessment reports also may be required for the infrastructure. For site monitoring, cultural

features are not only of interest in themselves but also provide context for the detection of other features such as vehicles. The cultural features in a database also help in orienting an analyst to the observation of a new image by registering the image to a site model and by pointing out the major known features of interest. Thus, the construction of geospatial databases containing cultural features is of key importance in all aspects of battlefield environment, from initial site monitoring to mission planning and rehearsal, in the execution of the mission itself, and then in an analysis of the results after an action.

Some commercial *softcopy* systems for constructing geospatial databases are available. These can be helpful in the process of recording the data and carrying out many photogrammetric computations. However, the extraction of the important features, particularly the cultural features, largely remains a manual task. At best, these systems allow a user to choose parameters of a prototype shape; the large number of parameters can make this a tedious and inefficient process. There has been some progress on automated detection in research systems but the capabilities remain limited. We need to significantly expand the classes of objects that can be modeled, the conditions under which they may be imaged and to make the systems more robust and reliable.

2 Research Issues and Technical Approach

The problem of 3-D feature extraction is difficult in many ways. Low level segmentation techniques (such as line detection) give incomplete and imperfect results. Object boundaries may be incomplete and many extraneous boundaries, due to markings,

* This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by Topographic Engineering Center of the U.S. Army.