Intelligent Control for Automatic Model Acquisition from Aerial Images[1]

Christopher O. Jaynes, Mauricio Marengoni, Allen Hanson, and Edward Riseman

Computer Vision Laboratory
Dept. of Computer Science
Box 34610, University of Massachusetts
Amherst, MA. 01003
E-mail: jaynes@cs.umass.edu, URL: http://vis-www.cs.umass.edu

Keywords: Bayesian control, object recognition, 3D reconstruction, model acquisition

Abstract

A system is presented that automatically acquires 3D geometric site models from multiple aerial images. Although there have been a number of different site reconstruction algorithms developed, they address a small number of object classes and typically fail when applied outside of the domain for which they were designed. Information about context sensitive performance is encoded along with each of the available vision algorithms in the form of a Schema. Schemas gather information about the scene that is then used by the system for reasoning, control, and production of a geometric model.

Evidence produced by the execution of a Schema is stored in a geometric database that may be used for a final scene reconstruction, and is fused with other evidence using a Bayesian network. Use of Bayesian networks allow the explicit representation of domain knowledge, while use of schemas allow algorithms to be selected and executed within contexts in which they are likely to succeed. Results show how the framework extends the capability of the Ascender I system, a building model acquisition system, by automatic classification image regions prior to geometric reconstruction.

1 Introduction

The extraction and reconstruction of geometric models from images is an important practical focus of the computer vision community. Significant progress has been made in several constrained subareas and systems perform reasonably well within the domains for which they were designed. These (sub)efforts can be characterized by the contextual constraints embedded into the algorithms at the time they were designed. These implicit and explicit restrictions vary from the type and characteristics of the data required for processing to the classes of objects addressed by the algorithm. Although these algorithms perform well within the particular contexts for which they were designed, they often degrade significantly within different domains.

An alternative to the monolithic IU system is one composed of a set of smaller systems which are experts at a particular visual task, such as recognizing a specific class of objects. Our working hypothesis

[1] Funded by the APGD project under DARPA contract number DACA76-97-K-0005 and in part by the National Council for Scientific Research-Brazil grant number CNP 260185/92.2.

is that both generality and robustness can be achieved by integrating these sets of experts into a larger system which provides the appropriate infrastructure and communication channels. The key is then selecting the right strategy, at the right time, and applying it to the right data. It is necessary to fuse the results from individual experts into a coherent site model. The framework described here is similar in some respects to the Schema system [Draper'89], and several knowledge directed vision systems [Rimey'92, Musman'93, Sarkar'95], as well as other reconstruction systems from the aerial image domain [Chellapa et al.'94, Huertas and Nevatia'80, Gifford and McKeown'94, Jaynes'96].

2 System Overview

The Ascender II system is divided into visual and reasoning subsystems (see Figure 1). The *visual subsystem* contains a library of IU algorithms, a geometric database that contains available data (images, line segments, functional classifications, etc.), as well as models that may have been acquired through processing. Display of the acquired models and a user interface is supplied by the Radius Common Development Environment (RCDE) [Mundy et al.'92], a geometric modeling package. The *Reasoning subsystem* is used for classification of polygons identified in aerial images, it is divided into two parts, knowledge base and controller. The *knowledge base* is composed of a set of belief networks which are constructed using HUGIN [Andersen'89], a system for designing belief networks and influence diagrams. The *controller* uses the knowledge base to decide on the algorithm to be applied in the image.

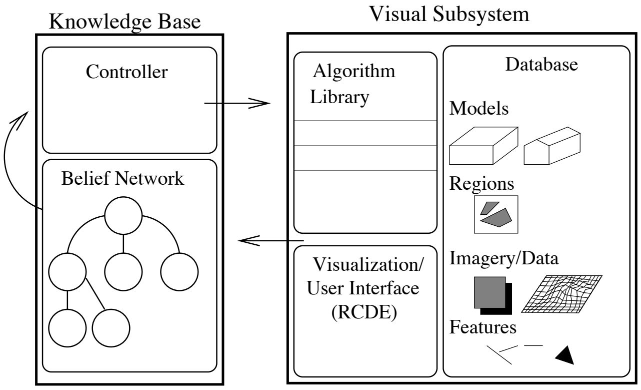

Figure 1: System overview. Control decisions are based on the current knowledge about the site. Vision algorithms, stored in the visual subsystem, gather evidence about the site, update the knowledge base and produce geometric models.

2.1 Hierarchical Bayesian Controller

A Bayesian network is a probabilistic inference system, represented by a directed acyclic graph, that denotes causal dependencies within the domain under consideration. Each node in the network represents a random variable and each arc represents a relationship between the variables. More details about Bayesian networks can be found in Pearl [Pearl'88] and a summary of Bayesian networks for computer vision can be found in [Rimey'92].

The knowledge base is a hierarchical system composed of a set of networks divided by different levels of detail. Reasoning takes place over *regions of discourse* that represent a subset of the available data. Regions of discourse may be image regions, volumes in the world, a particular building model, or other sets of data that may have been produced by the system. As opposed to systems that partition the space under consideration *a priori*, regions of discourse are formed, merged, and destroyed during processing of the data. Each level of the network hierarchy provides relevant information about a region at a particular scale of detail. Processing within the network is restricted to a single level until recognition at the root node occurs. This recognition is then used to start processing within a new network that will provide a more detailed view of the object being recognized. Given enough processing resources and time, the process continues until the root nodes of at least one of the networks in each level has provided recognition.

Specific site knowledge can influence processing of the scene and is stored in the prior probability distributions at each node in the network. For example, if 80% of the regions detected in a certain area are buildings and the only two possible classifications for a region are either building or open field, then the expected frequency of a particular region being an open field is 20%, and the expectation of finding region heights greater than zero will be proportionally higher than regions of small height.

All nodes in the network searched by the controller have an associated schema that encodes how relevant evidence is fused and propagated through the network. The schemas are selected based on the uncertainty of the node. A node is considered completely uncertain if the distribution of beliefs among the states is uniform. The node with highest uncertainty is selected using the expression below.

$$node \;=\; arg\; min_n(max(Belief(n)) - \frac{1}{S_n})$$

where S_n represents the number of states of node n. The schema related to the node selected is then invoked in order to reduce uncertainty. The findings of this action are then returned to the controller, entered as evidence and propagated through the network.

After each new piece of evidence is propagated through the network the maximum belief and the second maximum are computed and compared. If the maximum is at least twice the value of the second maximum the controller stops and gives to the region the label defined by the state with the maximum belief.

2.2 Visual Subsystem

The visual subsystem is composed of two parts: a function library that stores the set of IU algorithms available to the system, and a geometric database that contains available data in the form of imagery, partial models, and other collateral information about the scene (such as classification of functional areas).

At the request of the controller, an algorithm is selected from the library and run on a region that currently resides within the geometric database. New regions may be produced as a result of processing and these are stored in the database for future reference. In addition, the controller may request that regions be merged, split or eliminated.

The algorithm library contains information about each of the algorithms available to the system for selection as well as a definition of the contexts in which each algorithm can be applied. This information is stored in the form of a schema, encoding the preconditions that are required for the algorithm to be executed, the expected type of data that the algorithm will produce, and the algorithm itself. If preconditions for a particular algorithm are not met, then an alternative algorithm may be executed if it is available within the schema. If there are no algorithms that can be run in the current context, then the corresponding belief value cannot be extracted by the visual subsystem and must be inferred from the Bayesian network. The set of algorithms used for the results in the paper are shown in table 1.

The library of algorithms presented here were developed to address aspects of the site reconstruction problem from aerial images. For example, finding regions that may contain buildings, classifying building rooftop shapes, and determining the position of other cultural features, are all important tasks for the model acquisition system. The IU algorithms may be very "lightweight", be expected to perform only in a constrained top-down manner, and usable in more than one context. Other algorithms may also be very complex and themselves contain multiple strategies and associated control; several of the algorithms presented here are sophisticated procedures.

If the framework is to be truly general useful, the cost of engineering a new schema must not be prohibitive, something that proved to be a problem in earlier knowledge-based vision systems [Draper'89]. Only two components are necessary to convert an IU algorithm into an evidence policy that is usable by the system. First, the context in which the algorithm is intended to be run must be defined. Currently, the definition of allowable contexts is straightforward and only disallows algorithms to be run in invalid contexts (on the wrong type of data, for example). This is similar to the Context Sets introduced in the Condor system [Strat'93]. This definition of context is expected to be too simple for our needs and eventually the framework will be extended to allow the definition of a performance profile for each algorithm that defines the expected performance of the algorithm under a variety of different contexts. Secondly, a method for deriving a certainty value from the output of the algorithm must be defined. This certainty value is used by the system to update the knowledge base using Bayesian inference.

3 Experimental Results: Extending the Ascender System

An experiment was conducted to demonstrate how the introduction of the knowledge directed framework into the site reconstruction process can improve the completeness and accuracy of the final site model. The dataset used contains seven overlapping aerial views of the site. The area includes several cultural features typically found at an urban site such as buildings, parking lots, road networks. A ground truth site model was constructed by hand through alignment of building models in all seven images. A corresponding elevation map, generated from the groundtruth model, was also included in the dataset

Algorithm Name	**Preconditions**	**Description**
`Line Count`	(Image)	Compute number of lines in image region [Weiss'86]
`Line Frequency`	(Image)	Compute line frequency as the number of lines versus region size
`T Junctions`	(Image,Camera)	Detect image lines, intersect to compute junctions, back-project to a nominal world plane to determine if there are junctions of type 'T'
`L Junctions`	(Image, Camera)	Search region for junctions, type 'L'
`# T Junctions`	(Image, Camera)	Count 'T' junctions in region
`# L Junctions`	(Image, Camera)	Count 'L' junctions in region
`T Junction Contrast`	(Image, Camera)	Median contrast of 'T' junctions. Junction contrast is computed as the average contrast of the two lines that gave rise to the junction.
`L Junction Contrast`	(Image, Camera)	Median contrast of 'L' junctions
`L Junction Boundary`	(Image, Camera)	Search region boundary for 'L' junctions
`Image Variance`	(Image)	Compute greyscale variance in region
`Shadow`	(Image, Camera, Sun Position)	Search image in areas neighboring region. Detect shadows as dark areas with high contrast edges. Determine if shadow is flat, non-flat, or does not exist.
`Region Height`	(Image, Multi-Views) (Elevation Map)	In the case of multiple overlapping views, match boundary segments across views to compute height [Collins'96] If a DEM is available, extract corresponding elevation information, return median height.
`Model Index`	(Elevation Map)	Fit mesh to elevation map, correlate with library of parametric surfaces return model in rank-order. [Jaynes'97] Parametric Model types are: Flat, Peak, FlatPeak, Cylinder
`Model Index X`	(Elevation Map)	Correlate with all models of type **X** only. Return best correlation score and model parameters.
`Model Fit`	(Elevation Map, Model)	Fit the parametric model to region within the elevation map using robust fitting technique [Jaynes'96]
`Width`	()	Width of Region
`Length`	()	Length of Region
`Ratio`	()	Width to Height ration of Region

Table 1: Set of algorithms currently available to the Ascender II system. The preconditions required and a short description of each algorithm are also shown.

for the experiment.

The experiment was performed in two stages. First, Ascender I, a hand-crafted building reconstruction system, was executed on the area under consideration. [Collins'96] The system was developed over several years and tuned to extract flat rectilinear buildings. The system detects rectilinear structure in a single image of the dataset and uses the known relative camera pose between other views to compute a polygon height. This final site model is a set of these boundaries extruded to the ground. Although the system has been shown to be effective in detecting a large percentage of buildings at the site, production of false positives in the site model can be a problem. Although the Ascender system attempts to eliminate these false positives by searching other images for sufficient edge evidence, often this is not sufficient to discriminate between true buildings and false positives.

Ascender II uses Bayesian networks to recognize regions within the site model prior to reconstruction. The process will eliminate many of the false positives produced by Ascender I and allow for a larger class of cultural features to be recognized and added to the final site model. Because the primary interest in the recognition process is concerned with identifying buildings, sublevels for the preliminary knowledge base have been developed only for the building branch. The first level classifies a region into one of the classes (`Building, Parking Lot, Open Field, Complex, Other`).

A second network attempts a finer classification of `building` regions into either (`Multi-level Building, Single Level Building`). The first level and building-class specific network are shown in in figure 2. If a building is classified as a single building then the network presented at left of figure 3 is called. This network is used to classify the roof top in the building. If the controller finds a “good” classification the `Model Fit` schema is called to confirm the classification. If the controller can not uniquely determine the rooftop type then an “expensive” schema is invoked to fit a set of models in the data and the best model is returned. [Jaynes'97] On the other hand if a multi level building was found, the network shown at right of figure 3 is called. Using this network the controller checks if the building is really a multilevel or perhaps a single building with a peak roof. If the controller gets a confirmation for multilevel building then it asks the vision subsystem to split the region and loads the single building network in order to recognized the rooftop type for each part of the multilevel building.

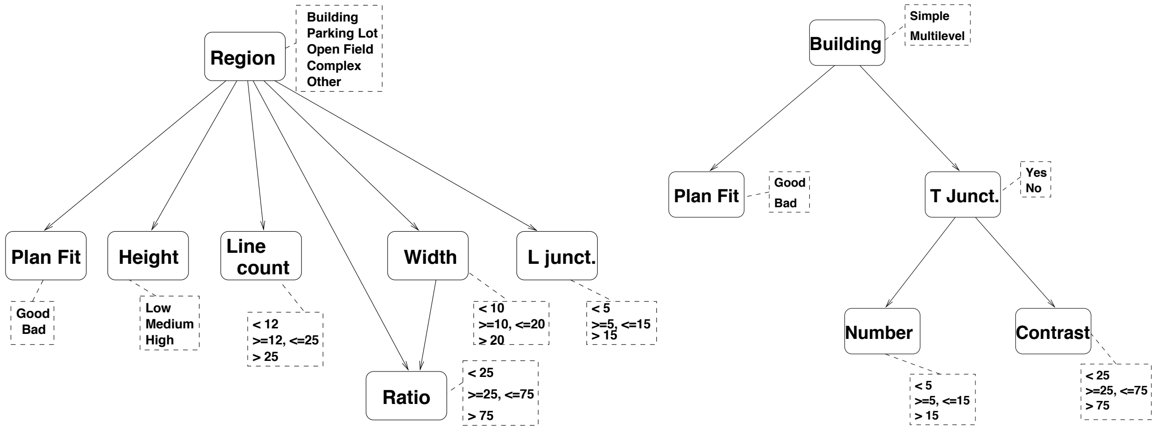

Figure 2: The network at left works in the coarsest level. It tries to classify a region into one the possible outcomes: Building, Parking Lot, Open Field, Complex, Other. The network at right works in level 1. For each building found in level 0 this network is invoked and tries to classify the building as either a simple building or a multilevel building.

3.1 Results

For each polygon produced by the Ascender I system, a region was produced by the visual subsystem and a request for classification was issued to the knowledge base. After the selection of an appropriate evidence policy, the action selected is passed to the visual system where the actual processing is accomplished. Evidence values are returned to the knowledge base where they are used to update the network. The system was run on the 42 regions extracted by Ascender I; processing was stopped when a belief value for one of the states reached the limit condition or the controller was unable to select a new action. The results of classification using the controller is presented in figure 4.

Region Classification	
Region Type	Total
Simple Buildings	22
Multilevel Buildings	1
Parking Lots	4
Open Fields	13
Unknown	1

Building Classification	
Building Type	Total
Flat	24
Peaked	0
Curved	0
Flat Peak	0

For all regions classified as buildings, the network was able to further classify the rooftop shape as flat. Recognition of a flat roof is dependent on the planar fit error and the results of the `Index Models` function that found the planar model to match the region's elevation data better than the other available models. We have begun experiments on datasets that include several different rooftop shapes in order to demonstrate the level of detail the system is able to achieve.

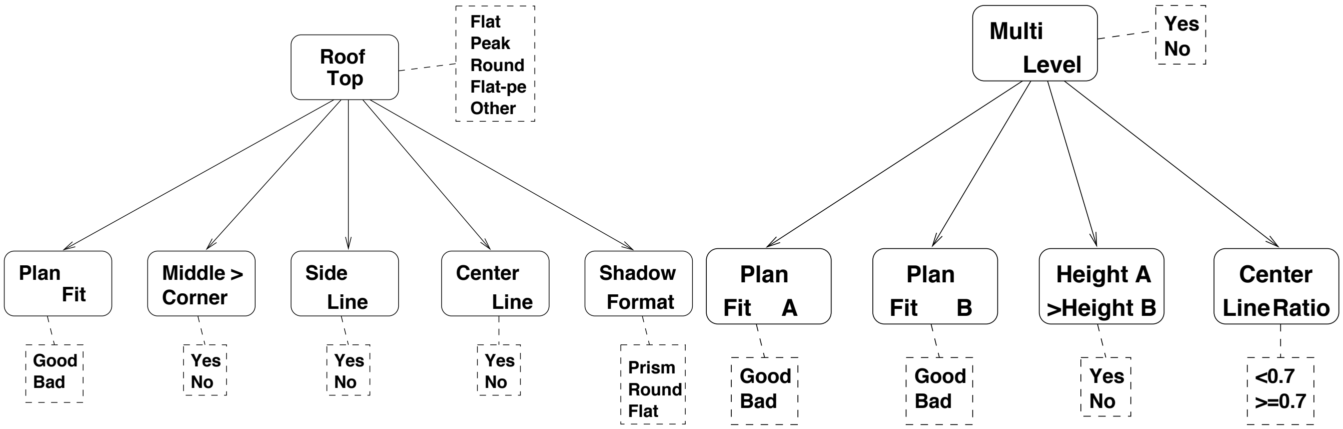

Figure 3: Both networks are invoked in level 2. The network at left is called after a single building is detected and it is used to make roof top classification. The network at right is called when a multilevel building is detected. The reasoning system calls this network to confirm the classification and calls the network at left for each level of the building.

In figure 4 region "C", which is a parking lot, was classified as a building, partly due to the number of lines detected which was uncharacteristically low for most parking lot regions. Region "D", which is composed of a parking lot and building with some area of grass should have been classified as complex, but the system classified it as parking lot. This mistake is understandable in light of the the fact that the `complex` classification includes a mixture of features from all other models. Parking lot features, such as many short lines and a rough elevation map, are not only present in region "D" but prevalent. The other two regions that were misclassified are the two small regions in the parking lots, a car to the right of region C, that was classified as open field instead of unknown and a truck to the left of region "A" which was also classified as open field. Region "B" was classified correctly as a building in level 1 but in level 2 the system exhausted all actions and was not able to decide between simple building and multilevel building. However, the maximum belief obtained for that region was 66% for single building versus 34% for multilevel. Region "A", which is a parking lot, had the same problem in level 1, the system exhausted all actions and at the end the highest belief presented was for parking lot (59%) and the second highest value was for open field (31%). The only multilevel building in the scene, region "E" was properly classified.

In the overall classification process the system used only about 41% of the actions available. An interesting result in this process is that an area of the same type, say "open field", was classified using a

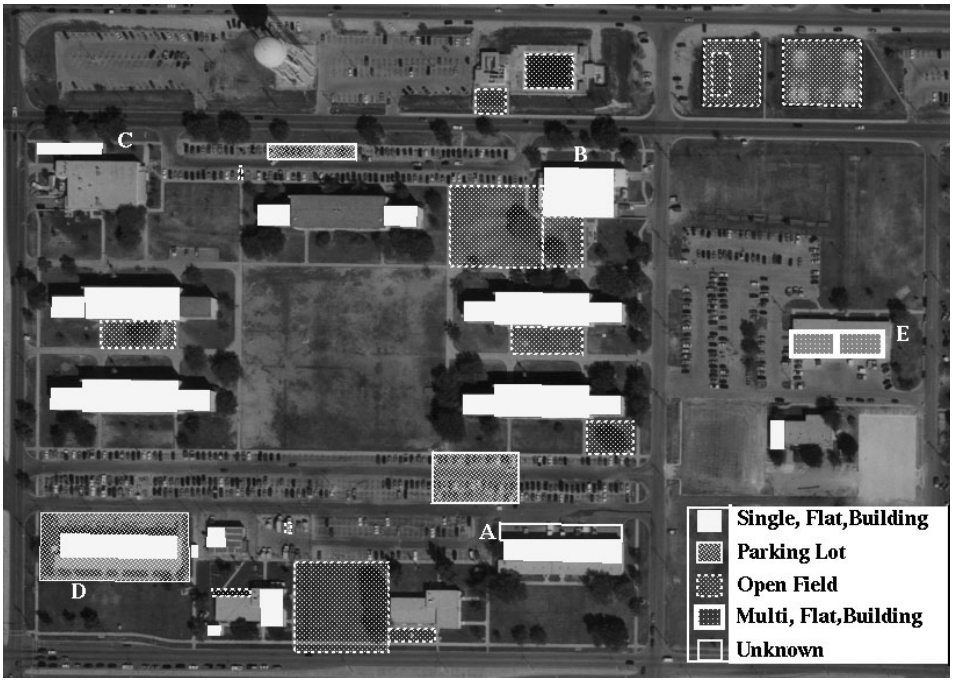

Figure 4: Classification results on regions produced by the Ascender I system. Four regions were misclassified and in one region the system was not able to decide on the classification. Letters referred to in the text.

different set of actions. The final models achieved with the use of the knowledge framework and without are compared in the figures below.

4 Conclusions

We have demonstrated how a flexible, knowledge-directed control framework can improve the accuracy of model acquisition systems such as Ascender. Our goal is to demonstrate that this flexibility improves system performance and widens its scope of applicability. To this end, work is underway on the development of additional evidence policies for a wider range of building classes. The general framework being employed supports any type of data as long as there are corresponding evidence policies available for interpreting it. Consequently, the system is being extended to include IFSAR elevation maps (in addition to elevation maps from traditional stereo techniques) and multi-spectral imagery for improved ground classifications. More detailed networks for cultural features other than buildings are being developed. For example, we hope to discriminate between full and empty parking lots as well as different types of fields. We are currently testing the framework on a large number of datasets that include several different rooftop classes such as flat, peaked, and curved roofs.

References

[Andersen'89] S. Andersen, K. Olesen, F. V. Jensen,F. Jensen, "HUGIN - A shell for building Bayesian belief universes for expert systems" In *Proceedings of the 11th international joint conference on artificial intelligence*, pp 1080-1085, 1989.

[Weiss'86] R. Weiss and M. Boldt, "Geometric Grouping Applied to Straight Lines," *IEEE Computer Society on Computer Vision and Pattern Recognition*, 1986.

[Breese'95] J.S. Breese and D. Heckerman. "Decision-theoretic case-based reasoning", Microsoft Research, MSR-TR-95-03, January 1995.

[Chellapa et al.'94] R. Chellapa, L. Davis, C. Lin, T. Moore, C. Rodriguez, A. Rosenfeld, X. Zhang, and Q. Zheng. "Site-Model-Based Monitoring of Aerial Images" *Computer Vision and Pattern Recognition (CVPR)*, pp. 694-699, 1997.

[Collins'96] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, E. Riseman. "The ASCENDER System: Automated Site Modelling from Multiple Aerial Images", To Appear: Special Issue in *Computer Vision and Image Understanding (CVIU)* on Building Detection and Reconstruction from Aerial Images, guest editors R. Nevatia, A. Gruen, 1998.

[Cooper'90] F. Cooper, "The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks", *Artifical Intelligence*, vol. 40 pp. 393-405. 1990.

[Draper'89] B. Draper, R. Collins,J. Brolio, A. Hanson, E. Riseman. "The Schema System", *International Journal of Computer Vision*, vol. 2. pp. 209-250. 1989.

[Gifford and McKeown'94] J. Gifford, D. McKeown. "Automating the Construction of Large-Scale Virtual Worlds *Proc. ARPA Image Understanding Workshop*, 1994.

[Herman'94] M. Herman and T. Kanade. "3D Mosaic Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images". *Proc. ARPA Image Understanding Workshop*, 1994.

[Huertas and Nevatia'80] A. Huertas and R. Nevatia. "Detecting Buildings in Aerial Images" *Computer Vision, Graphics, Image Processing.* vol. 13, 1980.

[Jaynes'94] C. Jaynes, F. Stolle and R. Collins "Task Driven Perceptual Organization for Extraction of Rooftop Polygons," *IEEE Workshop on Applications of Computer Vision,* Sarasota, FL, December 1994, pp. 152–159.

[Jaynes'97] C. Jaynes, E. Riseman, and A. Hanson, "Building Reconstruction from Optical and Range Images" *Computer Vision and Pattern Recognition (CVPR)*, San Juan Puerto Rico, June 1997.

[Jaynes'96] C. Jaynes, F. Stolle, H. Schultz, R. Collins, A. Hanson, and E. Riseman. "Three-Dimensional Grouping and Information Fusion for Site Modeling from Aerial Images" *Proc. ARPA Image Understanding Workshop*, pp. 479-490, 1996.

[Mundy et al.'92] J. Mundy, R. Welty, L. Quam, T. Strat, B. Bremner, M. Horwedel, D. Hackett, and A. Hoods, "The RADIUS Common Development Environment", *DARPA Image Understanding Workshop*, pp. 215-226, 1992.

[Musman'93] S. Musman and L. Chang, "A study of Scaling Issues in Bayesian Belief Networks for Ship Classification", *Proc. of 9th Conference on Uncertainty in Artifical Intelligence*, pp 32-39, 1993.

[Pearl'88] J. Perl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[Rimey'92] R. Rimey and C. Brown, "Task-Oriented Vision with Multiple Bayes Nets", Active Vision, eds. A. Blake and A. Yuille, MIT Press, 1992.

[Sarkar'95] S. Sarkar and K. Boyer, "Using Perceptual Inferece Networks to Manage Vision Processes" *Computer Vision and Image Understanding*, vol. 62(1) pp. 27-46, 1995.

[Strat'93] T. Strat. "Employing Contextual Information in Computer Vision", *Proc. ARPA Image Understanding Workshop*, 1993.