Knowledge Directed Reconstruction from Multiple Aerial Images*

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

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`

Abstract

Image understanding (IU) techniques for automatic site reconstruction have demonstrated success within restricted domains and for small numbers of model classes. However, these techniques often fail when applied out of context and do not "scale-up" into a more general solution. Under the APGD program, we are constructing a knowledge-based site reconstruction system that automatically selects the correct algorithm according to the current context, applies it to a focused subset of the data, and constrains the interpretation of the result through the explicit use of knowledge.

1 Introduction

The extraction and reconstruction of building models from aerial images has become an important area of research in recent years. Significant progress has been made and several systems perform reasonably well within their appropriate domains [Collins'95, Herman'94, Lin et al.'94, Chellapa et al.'94]. For example, recent testing of the Ascender I system has shown it capable of automatically extracting a large percentage of the buildings within a subregion of the Fort Hood dataset [Collins, et al'96]. Although these results are significant, the system was designed to perform well under particular contexts and is only capable of detecting the single class of buildings whose rooftops are flat rectilinear polygons.

The modest successes attained by Ascender I and similar systems can, we believe, be traced to their narrow scope and application to highly constrained data. The class of flat roofed rectilinear buildings is very clearly defined by a set of geometric and spatial properties which are useful for recognition if the incidence of distracting classes is small (that is, when the data is suitably constrained). This idea of a (set of) local expert(s) for recognizing instances of an object class played a prominent role in early work on the Schema system [Draper'89], as well as other systems in the aerial image domain [Chellapa et al.'94, Huertas and Nevatia'80, Gifford and McKeown'94, Jaynes'96a, Matsuyama'85]. Under this model, robustness is achieved by providing multiple reconstruction/recognition strategies which are applicable under well defined conditions and generality is achieved by increasing the number of object classes to describe a larger fraction of the world.

Work has begun on Ascender II, a geometric site modelling system based on this general framework. The design of Ascender II is founded on three basic principles: 1) Specific image understanding strategies are clearly successful under particular contexts for a particular class of objects but may break down when applied in contexts that exceed the design constraints. 2) Domain knowledge, knowledge acquired from partial processing of the data, and knowledge about available image understanding strategies are all valuable in constraining the reconstruction problem. 3) A successful system will contain many specific strategies but will selectively apply them in the correct context, with the correct set of parameters, and will fuse the results of individual strategies into a complete reconstruction.

2 Ascender II

The Ascender II system explicitly represents both knowledge and context to support a purposeful reconstruction of the site using geometric and spatial reasoning and intelligent control of sophisticated IU algorithms. The system is divided into a vi-

*Funded by the RADIUS project under DARPA/Army TEC contract number DACA76-92-C-0041, by NSF grant number CDA-8922572, and in part by the National Council for Scientific Research-Brazil grant number 260185/92.2.

sual subsystem and a knowledge base. The *visual subsystem* resides within the Radius Common Development Environment (RCDE) [Mundy et al.'92] and 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. The *knowledge base* is based on belief networks and is constructed using HUGIN [Andersen'89], a system for designing belief networks and influence diagrams. The knowledge base consists of reasoning mechanisms, a control system, and the belief network that represents the current set of knowledge about the site. The two systems communicate through Unix socket IP mechanisms. Figure 1 shows an overview of the system.

Reasoning takes place over *regions of discourse* that represent a subset of the available data. Regions of discourse may be image regions, a particular building model, or other sets of data that may have been produced by the system.

Processing of the data proceeds in a straightforward way. First, the knowledge base is consulted and an appropriate IU algorithm and subset of the current region are selected. For example, a search for line evidence along the center of a building region may be invoked to gather evidence for the presence of a peaked roof building. The choice of algorithm and subset of the data is sent as a request to the visual subsystem for processing. The IU algorithm is applied to the data and the database is updated with the result (for example, a set of line features may be produced). The visual subsystem then converts the result into a single value that represents the belief that the requested evidence was present. This belief is passed to the knowledge base where it is used to update the belief network. The next appropriate action is then selected based on the control policy.

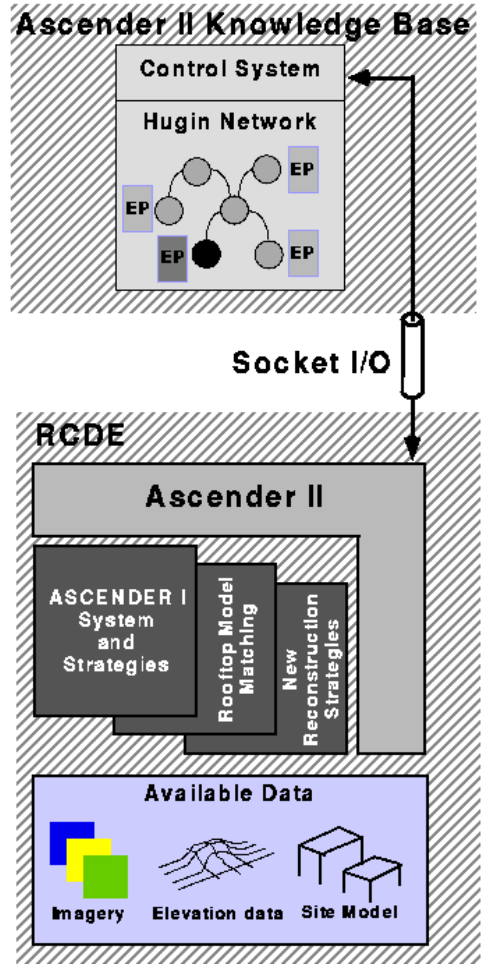


Figure 1: Ascender II system overview. Control decisions are based on the current knowledge about the site. Vision algorithms, stored in the RCDE, gather evidence about the site and produce a site model. Ascender I provides one set of IU strategies relevant to site reconstruction.

2.1 Knowledge Base

The knowledge base is capable of representing the current context, specific site knowledge (either engineered or acquired as part of processing), and general domain knowledge relevant to site modelling. Knowledge is stored in a *Schema Network*. The representation is a combination of two important ideas drawn from the field of Artificial Intelligence; Schemas [Draper'89] and Belief Networks [Jensen'96]. The network encodes information about how and when algorithms can be applied in the current context and explicitly represents the causal dependencies found in a particular domain.

Nodes within the network represent discrete variables that are associated with the domain. For example, the node `Building-Roof-Shape` may have the discrete states {`flat, peaked, curved, composite`}. At each node, an *evidence policy* contains information about how evidence for a peaked roof building may be acquired. Contextual rules, part of the node's evidence policy, assist in the selection of the correct algorithm, data, and parameters, for the given context.

Edges within the network represent a conditional dependence between a parent and child node. Associated with each node is a conditional probability table that contains a probability for each state of the node given the values of the parents. The *Belief* for a parent node, then, can be computed from the values of the children using causal influence [Russel'95]. As evidence is added to the network (through the execution of an evidence policy), the effect is propagated throughout the network and a new set of belief values are computed.

2.1.1 Action Selection

The problem of action selection within the schematic network is a significant one. Currently we take a greedy approach. In order to gather evidence about a node n, the corresponding evidence policy is invoked. If the evidence policy at n is empty (there are no IU algorithms directly applicable to computing belief(n)) or there are no available algorithms for the current context, then the children of n are visited. The node whose belief value contains the highest uncertainty is selected and either its evidence policy is invoked or its children are visited. Once evidence has been computed, the belief values are propagated back through the network and a new action is selected. This implies that there must be at least one evidence policy available at each of the leaf nodes within the network.

Certainty for node n is defined as difference of the maximum belief and the belief value if all states at n are equally likely.

$$max(Belief(n) - \frac{1}{(Num\ states)_n})$$

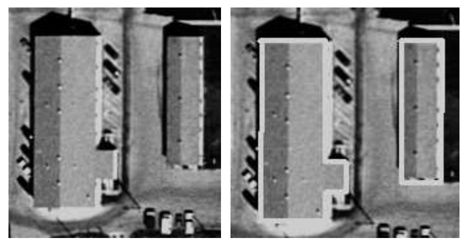
2.2 Visual Subsystem

The visual subsystem is comprised 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 site (such as classification of functional areas).

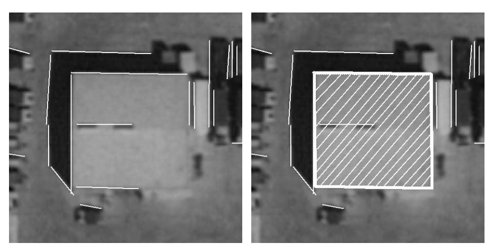
The library of Ascender II algorithms must address aspects of the site reconstruction problem. For example, finding regions that may contain buildings, classifying building rooftop shapes, and determining the position of other cultural features, are all important tasks for the Ascender II system. Many of the IU algorithms may be very "lightweight" and are expected to perform only in a constrained top-down manner. This is due to the fact that the IU algorithms are responsible for gathering evidence for a particular hypothesis put forward by the knowledge base. For example, an algorithm that detects the presence of local maximum in a region of the elevation data can be viewed as a car detector when invoked on a parking lot area. The same algorithm may detect the presence of a rooftop structure when applied to a known building area.

Algorithms may also be very sophisticated, such as the reconstruction of flat roof buildings from multiple views (the role of the Ascender I system). Below, several of the more complex algorithms are briefly described in order to demonstrate the type of algorithms made available to the system.

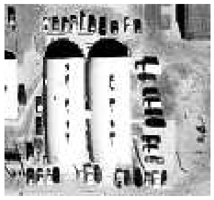
2D Polygon Detection [Jaynes'94]: Search optical image for polygons that represent high confidence rooftop boundaries. Lines and corners are extracted from the image and grouped into perceptually compatible chains. A search of all possible groupings returns the maximal independent set of closed chains.



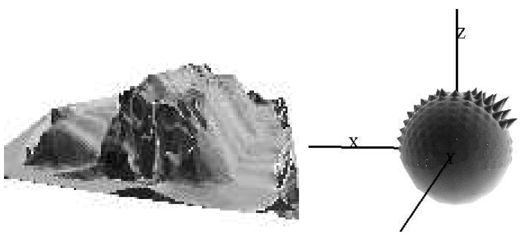
2.5D Feature Grouping: Match image features across multiple images to compute heights and group based on height/shape constraints. For example, compute line heights through a multi-image matching scheme. Group the line segments into sets of two parallel lines at the same height with a third, higher parallel line into regions that may indicate the presence of a peaked roof building.



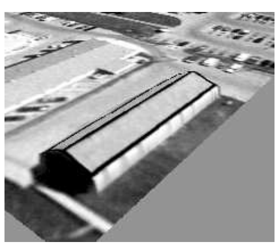
Local Shadow Analysis [Lin et al.'94]: The known sun angle and building model constrains the search for a corresponding shadow in the image. The shape of the shadow can be analyzed to infer the shape of the building rooftop that cast it.



Automatic Model Indexing [Jaynes'97]: Match a region of an elevation image with a surface primitive database. This is accomplished through a construction of the extended Gaussian image for the image region and correlating the surface orientation histogram with the database.



Fitting Parametric Surfaces to DEMs [Jaynes'96b]: Fit a model to a region within the elevation data. The model parameters should have already been determined through another processes (model indexing, for example).



As research into Ascender II continues, more IU algorithms will be added to the system. However, in order for the Ascender II framework to be useful, the cost of adding a new algorithm to the system 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 are 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] and rule packets within the HUB. 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 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. For example, the detection of L-junctions within a region of the image must be converted to a single value that represents the probability that the L-junctions are present.

2.3 Preliminary Tests

An experiment was conducted on a scene from the Fort Hood dataset. The test was both a simple example of the concepts presented here and a demonstration of the communication mechanisms that have been constructed as part of the Ascender II system. A small schematic network (only four nodes) was engineered that attempts to classify rectangular building boundaries (called building footprints) according to the three categories, `Single`, `Multi-level`, and `Multiple` that correspond to the case of a single planar rooftop, several planes or slopes at different heights, or more than one building in the region.

The network used for the test is shown in figure 2. The network encodes the fact that the classification of the footprint is dependent upon the presence of certain junction types along the edges of the region, and the quality of a single planar surface fit to the corresponding elevation data. An evidence policy that defined the plane fit algorithm and its reliance on avalable elevation data was constructed for the `Plan Level` node. Similarly, evidence polices for both L and T junctions were constructed.

Each child node in the network has an associated conditional probability table that encodes object specific knowledge. The conditional probability values are engineered for the specific problem, and, for the test here, were constructed based on our experience with both the evidence policies and the domain.

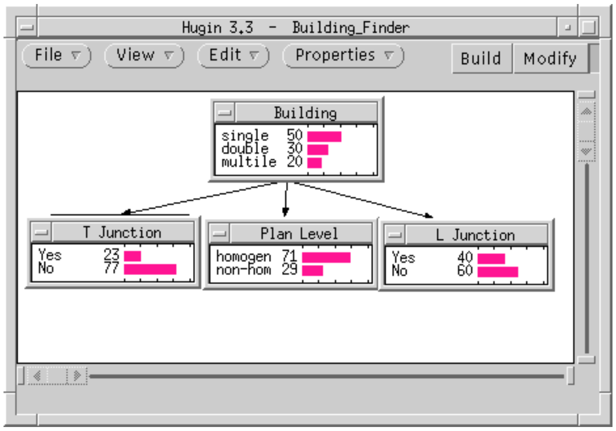


Figure 2: The network used to control the classification of a possible building region into one of the three possibilities: single building, double building or multilevel building.

The 2D polygon detector used in Ascender I was run on a single downlooking view of a subregion of the Fort Hood dataset. Ten of the polygons were selected for classification using the Ascender II system. In all, four polygons contained single level rooftops, four contained multi-level buildings, and two contained more than one building. Image 3 shows a typical polygon for each of the classes.

Figure 3: Three different building footprint cases. Leftmost: Single Rooftop. Center: Multi-level Building. Rightmost: Two distinct buildings.

The system was run on all ten regions and stopped when a belief value for one of the states for `footprint class` exceeded 65% or the controller was unable to select an new action. The region is then classified according to the state of `footprint class` with the maximum belief value. The table below shows the results of the experiment and the number of vision algorithms executed in order to classify the region.

Polygon Type	# Actions	Classification/Belief
Single	1	Single (75%)
Mult-Level	6	Multi-Level (57%)
Multiple	4	Multiple (57%)
Single	4	Single (55%)
Multi-Level	6	Multi-Level (50%)
Multi-Level	2	Single (75%)
Single	2	Single (75%)
Multi-Level	4	Multi-Level (83%)
Single	4	Single (81%)
Multiple	6	Multiple (65%)

3 Future Directions

Ascender II is based on a much more flexible design than was Ascender I. Our goal is to demonstrate that this flexibility improves system performance and widens its scope of applicability. To this end, work is underway on engineering the software architecture of Ascender II and 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. We expect to use the Fort Hood image dataset as well as other datasets as they become available (e.g. Ft. Benning) to demonstrate the Ascender II system.

There are many issues to be addressed during the design and implementation of Ascender II. One issue concerns the ganularity of the IU algorithms employed in the system and how this affects system performance. For example, should Ascender I be dismantled into component parts and reassembled in the knowledge network? Previous attempts to build knowledge-based systems ran into major knowledge engineering problems. The treatment of IU algorithms as black-box evidence gathering mechanisms, regardless of the underlying complexity, may be one way to avoid this. Currently, simple greedy evidence policy is being used to select the next action. What other policies are reasonable and how do the affect the system efficiency? Techniques that compare the expected utility of applying a particular evidence policy to its expected cost will be investigated as one way to answer the question of efficient control.

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.

[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'95] R.Collins, Y.Cheng, C.Jaynes, F.Stolle, X.Wang, A.anson and E.Riseman, "Site Model Acquisition and Extension from Aerial Images," *International Conference on Computer Vision,* Cambridge, MA, June 1995, pp. 888–893.

[Collins, et al'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", Submitted to: Special Issue in *Computer Vision and Image Understanding (CVIU)* on Building Detection and Reconstruction from Aerial Images, guest editors R. Nevatia, A. Gruen, to appear 1998.

[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'96b] 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.

[Jaynes'96a] C. Jaynes, R. Collins, Y. Wang, F. Stolle, H. Schultz, A. Hanson, and E. Riseman. "Automatic Construction of Three-Dimensional Models of Buildings", chapter to appear in ARPA IU RADIUS book. Oscar Firschein and Thomas Strat (eds.), Morgan Kaufmann Publishers, San Francisco, CA, 1997.

[Jensen'96] F. Jensen, *An Introduction to Bayesian Networks* Springer Verlag New York, 1996.

[Lin et al.'94] C. Lin, A. Huertas, R. Nevatia "Detection of Buildings Using Perceptual Grouping and Shadows", *Computer Vision and Pattern Recognition (CVPR)*, pp. 62-69, 1997.

[Matsuyama'85] T. Matsuyama and V. Hwang, "SIGMA: A Framework for Image Understanding: Integration of Bottom-Up and Top-Do wn Processes," *Proceedings of the Ninth IJCAI*, Los Angeles, CA, pp. 908-915, 1985.

[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.

[Russel'95] S. Russel and P. Norvig. Artificial Intelligence, A Modern Approach, Prentice Hall: 1995.

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