SUMMARY OF IMAGE UNDERSTANDING RESEARCH AT THE UNIVERSITY OF MASSACHUSETTS

Edward M. Riseman
Allen R. Hanson

COINS Technical Report 88-32

May 1988

Summary of Image Understanding Research
at the University of Massachusetts[1]

Edward M. Riseman and Allen R. Hanson

Computer and Information Science Department
University of Massachusetts at Amherst

ABSTRACT

This paper summarizes several areas of research at the University of Massachusetts that are partially or entirely supported under the DARPA Image Understanding Program. Many of the individual efforts discussed below are further developed in other papers in these proceedings. The summary is divided into several areas:

1. Knowledge-Based Vision
2. Database Support for Symbolic Vision Processing
3. Motion Processing
4. Perceptual Organization (Grouping)
5. Image Understanding Architecture
6. Integrated Vision Benchmark for Parallel Architectures
7. Mobile Vehicle Navigation.

Although we discuss each area separately, a fundamental goal of the computer vision research environment at UMass is the integration of a diverse set of research efforts into a system that is ultimately intended to achieve real-time image interpretation. Two of our major system integration

[1]This work is supported by the Air Force Office of Scientific Research under grant number AFOSR-86-0021, by the Defense Advanced Research Projects Agency under contract numbers DACA76-85-C-0008, DACA76-86-C-0015, F30602-87-C-0140, and N00014-82-K-0464, by the Defense Mapping Agency under contract DMA800-85-C-0012, and by the National Science Foundation under grant number DCR-8500332.

efforts are the VISIONS static interpretation system, which is a knowledge-based computer vision system utilizing parallel modular processes that communicate via a blackboard [DRA87b,HAN87a]. The second system integration effort involves an autonomous mobile vehicle for navigating through a partially known environment [ARK87a,ARK87b,ARK87c].

1 Knowledge-Based Vision

1.1 VISIONS Scene Interpretation

Research in knowledge-based vision has continued via the development of the VISIONS Schema System and its application to a variety of domains, including road scenes, house scenes, and aerial images [DRA87a,DRA87b,HAN87a,REY87]. Integrating concepts from artificial intellignce and computer vision, the Schema System provides a frame work for building prototype application systems employing a knowledge-based approach.

A primary characteristic of this work is the utilization of coarse-grained parallelism at the semantic level, where modular processes called schemas provide object-specific control strategies for recognition of object instances in expected contexts. Schemas partition both knowledge and computation in terms of natural object classes for a given domain; (note that the term "object" is used loosely to include object parts, objects, and scene contexts comprised of sets of objects). Each object class has a corresponding schema which contains declarative and control knowledge specific to that class. To identify an instance of the object class in an image, a schema instance is created, which is an executable copy of the schema associated with the object class. One schema instance is invoked for each object instance hypothesized to be in the scene.

Schemas run as independent concurrent processes. Communication between schema instances is carried out asynchronously by means of a global blackboard through which schemas cooperate and compete to identify and locate the significant objects present in the scene. Consistency is enforced by the requirement that each spatial area of the image is to have only one interpretation, and that all such local interpretations are to be mutually compatible.

There is now a serious effort getting underway to bring the Schema System up on a shared-memory multiprocessor (a Sequent) to begin to explore these ideas on a real machine. The issues here significantly overlap much of what appears in the remaining areas of this summary since those systems will be the components controlled by the interpretation system. In particular this research effort complements that of the Image Understanding Architecture (IUA) [WEE88a,WEE88b], since only a vertical slice of the IUA is being constructed with a single symbolic processor at the the top of the three-level architecture. Thus, the multiprocessor investigation of schemas and the experimental experience of applying it should give insight into the top-level design of the IUA before any attempt is made to scale it up (see section 5. of this summary).

1.2 Object Recognition Using 3D Models

Our primary concern with the schema approach has been on control issues in the parallel recognition of scenes and objects. The experimental knowledge bases constructed to date have utilized primarily two-dimensional processes, with only limited use of three-dimensional information. However, the importance of three-dimensional geometric models is quite obvious. Burns and Kitchen [BUR87a,BUR87b,BUR88] are developing an object recognition system that is designed to handle the computational complexity posed by a large model base of geometric object models, an unconstrained viewpoint, and the two-dimensional structural detail that will appear in the various views of an object.

The design is based on two ideas. The first is that before recognition, in an off-line process, the three-dimensional model base can be precompiled into descriptions, called predictions, of the potential appearance of all objects from all viewpoints. This reduces the recognition process during interpretation to a 2D matching process and avoids the more complex 3D-to-2D transformations between 3D models and 2D images. The second is to represent all the predictions in a prediction hierarchy. The nodes in this hierarchy are partial 2D descriptions that are common to multiple object views and hence constitute shared processing subgoals during matching. Thus, at any stage of extracting 2D geometric structure, the prediction hierarchy provides an indexing path to the 3D models that might have produced them.

A current focus of this work is on the automatic compilation of a prediction hierarchy from a 3D model base. A prototype system using a set of polyhedral objects and projections from an unconstrained range of viewpoints is under development [BUR88].

1.3 Goal-Directed Control of Low-Level Processes for Image Interpretation

In a recently completed thesis, Kohl [KOH88,KOH87] adopted the view that the task of image interpretation should be viewed as a coordinated process in which high-level interpretation strategies and low-level image segmentation processes interact through the intermediate-level of processing. He developed a system called GOLDIE (Goal-Directed Intermediate Level Executive) which mediates this interaction and which provides top-down control over low- and intermediate-level processes. Requests for low-level processing are represented in the form of a goal, which contains in its structure constraints on the form of the resulting data. Posting a goal to GOLDIE results in the creation of an intermediate level process which is an instantiation of a schema control strategy for achieving the goal. Thus, the set of schemas represented at the intermediate level define the types of goals which may be processed.

The active schemas utilize knowledge of the image domain as well as measurements on the image itself to translate the goals into appropriate low-level process specifications, which include the identification of the image features, algorithms, and algorithm parameters to be used by the system to accomplish the task. The low-level process specifications are then executed by a process controller, which returns the results to the intermediate-level control process. The constraints on the results are then checked and, if satisfied, the results are returned to the requesting high-level schema. If the results are unsatisfactory, a new process specification is generated and the cycle repeats until all possible avenues are exhausted. Execution of a process may result in the generation of subgoals, causing a recursive invocation of GOLDIE.

Although GOLDIE was developed primarily to control the process of region segmentation, the development of the segmentation incorporates both edge and line information. In particular the initialization schema of GOLDIE, which is executed at system startup in the absence of high level goals, uses both representations and has proven to be a very effective region segmentation algorithm. Future versions of GOLDIE will extend the schema-based control paradigm to include all of the low and intermediate level algorithms currently in use.

1.4 Information Fusion

A constraint based approach to uniformly combining information from multiple representations and sources of sensory data has been developed [BEL86,RIS87]. The approach is important to research in intermediate level grouping, knowledge-based model matching, and information fusion. The techniques extend the capabilities of an earlier system [HAN87b] that applied constraints to attributes of single types of extracted image events, called tokens. Relational measures are defined between symbolic tokens so that subsets of tokens across representations can be selected and grouped on the basis of constraint functions applied to these relational measures.

Since typical low-level representations involve hundreds or thousands of tokens in each representation, even binary relational measures can involve very large numbers of token pairs. Control strategies for ordering and filtering tokens, based upon constraints on token attributes and token relationships, can be formed to reduce the computation involved in producing token aggregations. The system was demonstrated using region and line data and a simple set of relational measures. The approach can be naturally extended to include tokens extracted from motion, stereo, and range data.

1.5 Evidential-Based Control in Knowledge-Based System

Wesley [WES82,WES83,WES86,WES88] takes a somewhat different view of control in knowledge-based systems. Arguing that such systems which operate in real-world environments must necessarily reason about their actions from information that is inherently uncertain, impecise, and occasionally incorrect, he develops a system called OCULUS in which control related information is viewed as evidence for and against hypothesized control decisions. Using the Shafer-Dempster theory of evidence [DEM68,SHA76] as the mathematical foundation for evidential reasoning, OCULUS derives control related evidence from control knowledge sources which make measurements on the state of a knowledge-based interpretation system. Evidential decision measures were developed to help choose an action based upon the results of the inference process.

A simulated vision environment was constructed using results of the VISIONS system interpretation. Simulated KSs were parameterized via characteristics such as accuracy and ambiguity. Wesley's results demonstate that an evidential approach to control can improve OCULUS's ability to correctly interpret images by as much as 30% in most cases, with less than a 10% increase in effort. Furthermore, the system is shown to degrade gracefully as the quality of the information supplied by the control knowledge sources decreases. Wesley suggests that the domain independent control strategies developed in OCULUS might be effective in knowledge-based systems that operate in other real-world task domains.

2 Database Support for Symbolic Vision Processing

It is becoming increasingly evident that intermediate-level vision, and the perceptual grouping processes encompassed, are an extremely important component of any knowledge-based interpretation system. Our current view is that a major goal of the perceptual organization processes is to reduce the substantial gap which exists between the extracted image descriptors and the high level knowledge representations of the objects. The more abstract the intermediate level tokens are, the more computationally efficent the matching is between high level descriptions and the intermediate level tokens, where general world knowledge is used to constrain the set of possible interpretations.

The intermediate level may be viewed as simply a symbolic representation of primitive image 'events' as points, regions, lines, contours, areas, surfaces, etc. and their features, created by an iconic to symbolic transformation of the image data. However, recent work in vision has shown it to be much more than a passive level of data representation. Many of the recently developed grouping operators, for example, function at the intermediate level by building more abstract structures from the primitive descriptions [BOL87,DOL88,DOL86,FIS86,LOW85,REY87,WIL88b,WIT83].

Consequently, we view the intermediate level as hosting active processes which construct more abstract tokens from less abstract ones. Universally applicable similarity operators and geometric constraints are employed on the evolving spatial structures. In order to facilitate research on image interpretation systems, where data and control are closely coupled throughout all three stages, mechanisms must be provided for efficient structuring of the data and processes.

In addition, the complexity of many vision systems requires the cooperation and interaction of many researchers and the integration of their subsystems. The applications are far too large for an individual to solve on his own. Thus, the intermediate level representations and software environment must support, at a minimum, the following:

- a single uniform data interface to both high and low levels;
- sharing of data between levels, and between researchers at all levels;
- integration of research results into a monolithic system;
- standard handling of common relational and geometric queries, to reduce the programming overhead of coding them from scratch;
- distribution of data and processes over several machines and in several computer languages (C, LISP, FORTRAN)
- an efficient programming environment for intermediate level algorithm development.

Unfortunately, current understanding of this level of vision makes it impossible to predict the kind of structures which must be represented, the types of access to these structures, the kinds of relationships which might exist between them, or the range and type of descriptive features attached to them. At this point it appears that quite a diverse set of representations and mechanisms are employed in various vision system components. We can minimally assume that the intermediate level must support known methods of information fusion and perceptual organization, and provide the flexibility to support the representation and manipulation of geometric and structural relations. For example, the types of data which should be representable at this level include:

- points: endpoints, points of high curvature, vertices, virtual points, etc;
- lines and curves: edges, straight lines, curve segments;
- areas: regions, surface patches, focus of attention areas, etc.;
- relations: adjacency, containment, intersection, etc;
- structures: grouped lines and edges, edge-vertex tuples (e.g.corners), line chains, geometric structures, and generally subsets of tokens defined by a relation.

Each has an associated set of features, or descriptors, whose definition may vary as research progresses. Consequently, there are two fundamental types of data access that must be supported: access to tokens by name and by feature value (associative access); note that we also treat relations as features. It is rarely the case that a token definition stays constant over the course of an interpretation. Tokens may be split or merged with other tokens, features recomputed, and tokens may take part in many set relationships with other tokens.

2.1 ISR1

Research into intermediate level grouping mechanisms [BOL87,DOL88,FIS86,LOW85,REY87] [WIL88b,WIT83] and the development of the VISIONS schema system [DRA87a,DRA87b] [DRA88,HAN78a,HAN78b,HAN78c,HAN86] have led us towards the development of a flexible and efficient intermediate level of representation called the Intermediate Symbolic Representation (ISR) [BRO88,DRA87a,HAN87a].

ISR1 was implemented in 1985 primarily as a data interface between the output of the low-level image segmentation and feature extraction processes running in C on a DEC VAX and the high-level symbolic interpretation system running in Lisp on a TI Explorer. The unit of representation in ISR1 is the token, composed of a name and a list of features. The features are described through a lexicon and tokens sharing a common lexicon are organized into a tokenset. Each feature entry in the lexicon consists of a datatype and an optional on-demand function for computing the feature value. Standard feature datatypes include type real, integer, pointer, extents, and bitplane. Extents is simply the coordinates of the bounding rectangle of the token in the image plane. Features of type bitplane are binary masks defining the spatial coverage of the token in the image.

Since a tokenset may be viewed as a two-dimensional array, access to elements in the array are by token name (the rows) and constraints on feature values (the columns). Associative access of elements are returned as a list or as an array. One of the major design deficiencies of ISR1 was that there were no convenient mechanisms for representing and storing these elements.

Standard I/O and file handling utilities are provided for creating libraries of tokensets and lexicons. File transfer is a common method of communication among the different processors in our environment.

2.2 ISR2

ISR1 was used heavily over a period of years by researchers whose individual research focus was distributed reasonably uniformly over all three levels of abstraction. During this period of time, a number of design deficiencies were noted in ISR1; two of the major problems which necessitated the redesign were:

- The separation of the lexicon from a tokenset created problems. When the lexicon had to be modified, old tokensets no longer had valid descriptions; (for example, when a feature was added to a set of region tokens). Short-term solutions resulted in a proliferation of stored tokensets and a great deal of confusion at the application level.
- Sets of associatively accessed tokens could not be conveniently manipulated, made into tokensets, nor stored as features of other tokens. In particular, it was difficult to relate tokens across token types (such as regions and lines).

In response to these problems a decision was made to design a new version of the symbolic database [BRO88]. ISR2 retains the basic flavor of ISR1, including tokensets, the basic token access functions, and features and feature datatypes. The lexicon concept was eliminated in favor of associating the feature descriptions with the tokenset itself. Recognizing that there were other pieces of information which apply to the tokenset as a whole (such as generation dates, image information, and processing history), each tokenset is now organized as a simple frame, with slots for the various features of the tokenset. Frame features include the simple types integer, real, string, frame, and tokenset and the complex types composite, sort, slice, and virtual. The frame feature allows frame hierarchies to be constructed. The tokenset feature points to the tokenset or tokensubset associated with the frame. The composite feature is a generalization of feature types like bitplane and extents from ISR1 (i.e. they are multi-valued features). Virtual features are features whose values can be calculated but not stored, hence they are always calculated on demand. They serve much the same purpose as methods in an object-oriented programming language. Sort and slice datatypes provide facilities for defining and maintaining partitions based on feature values; for example, a typical application for a slice feature might be to create and maintain a grid for fast 2D spatial access to tokens from the image coordinates. Other modifications to ISR1 include a more comprehensive file management system to deal with the frame hierarchies, the addition of several types of demons (on-demand functions), and extensions to the command language to support the new capabilities.

Like ISR1, much of ISR2 will be implemented in C with a LISP user interface. Implementation is expected to begin in the near future. Since vision is such a dynamic research environment, it would not be unreasonable to expect ISR3 after experience with ISR2 is obtained.

3 Motion

3.1 Background

In the area of motion analysis, research has continued down several avenues: a subsystem for extracting depth from approximate translational motion for the CMU NAVLAB, an algorithm for binocular motion analysis, a closed form solution for recovery of general motion parameters under assumptions of constant motion, and an algorithm for spatio-temporal grouping and tracking of linear structures and recovery of their associated depth.

Over the last 7 years, our research group has developed a variety of motion algorithms, and in most cases applied them to real-world image sequences, including domains of robot arm workspaces, indoor hallways, and outdoor sidewalk/road scenes. In particular, experimental investigations of translational motion sequences demonstrated some degree of robustness. Anandan [ANA85b] [ANA87a,ANA87b,ANA88] developed an algorithm for determining feature point correspondences between frames that allowed the computation of dense displacement fields with associated confidences. This capability could be used to effectively track points across frames. Lawton [LAW83,LAW84] showed that the focus-of-expansion (FOE) often could be extracted from a sensor undergoing pure translational motion (i.e two degrees of freedom) to within a few degrees of accuracy. Glazer [GLA87a,GLA87b], in his recently completed Ph.D. thesis, developed two algorithms for the efficient computation of image motions using hierarchical multiresolution methods operating over image data pyramids.

Bharwani [BHA85,BHA86] developed a multi-frame algorithm for depth extraction under known translational motion which iteratively predicts the image motion of a feature point in future frames, determines correspondence by a search over the limited predicted area, and then refines the depth estimate using the new match. Snyder [SNY86] analyzed the effects of uncertainty in the location of the FOE and feature points in the image on the computation of depth, and showed how this analysis could be used to quantitatively provide predictions for constraining the search window used for matching these points in future frames.

Adiv [ADI85a,ADI85b,ADI85c] developed an algorithm for general sensor motion (five degrees of freedom) in an environment with objects undergoing independent general motion, the goal being to recover the motion parameters of both the sensor and any visible moving objects. This latter problem is much harder, and although there was some empirical demonstration of capabilities, there was an assumption that this algorithm would be computationally more complex, and perhaps less robust, than the algorithms for translational motion.

In this volume research is presented on extracting depth from approximate translational motion, intended for practical use in obstacle avoidance on the ALV [DUT88]; extracting depth from stereoscopic motion [BAL88]; and motion analysis that is simplified by assumptions of constant motion [PAV88]; and depth computation from grouped geometric structures [WIL88a,WIL88b]. We will discuss each of these briefly below.

3.2 Processing Approximate Translational Motion for the ALV

Previous research in motion analysis led us to attempt to deal with a real application subsystem for the CMU NAVLAB [THOR87]. The goal was to detect obstacles in the path of the vehicle at distances beyond the limits of the ERIM range sensor, i.e. at distances beyond 40 feet. Initial results from Bharwani's algorithm implied the possibility of extracting usable depth of obstacles at distances between 40 and 80 feet. By applying an FOE extraction algorithm prior to the depth extraction algorithm, there was an expectation that an effective subsystem could be developed. To accomplish this in actual imaging situations on a moving vehicle has turned out to be far more difficult.

In dynamic imaging situations where the sensor is undergoing primarily translational motion with a relatively small rotational component, it might seem likely that "approximate" translational motion algorithms can be effective in determining depth. Although translational motion is the dominant form of motion and is approximately constant over a long sequence of frames, there usually are local variations due to irregularities in the road surface (bumps, holes, and undulations), as well as minor rotation of the vehicle as it translates. This is often manifested in changes in the location of the FOE (i.e. effectively it produces a different translational motion), and in rotational motions that must be removed if correct values of depth are to be extracted from the feature displacements. An attempt to correct for these effects via a relatively simple preprocessing algorithm, without utilizing full analysis of the general motion problem, also led to difficulties. The issues and our experimental efforts to deal with what we considered to be the relatively simple problem of approximate translational motion are discussed in [DUT88] in this volume.

The problems led us recently to apply the Anandan and Adiv algorithms for general motion to the sequences of approximate translational motion with significantly improved results; this approach is also reported in [DUT88]. The conclusion is that while the FOE might be approximately extracted, in many real situations general motion analysis must be applied in order to determine depth of points, even when sensor motion is primarily translational with only small amounts of rotation. One obvious hardware solution (at significantly increased cost) is the use of a gyro-stabilized platform so that sensor motion typically will be much closer to the case of pure translational motion. Alternatives to these approaches to motion processing to extract motion parameters and depth are outlined in the next two subsections, and in the perceptual organization section where spatio-temporal grouping is used to derive depth from geometric structures.

3.3 Steroscopic Motion Analysis

By carrying out motion analysis with a pair of cameras - stereoscopic motion - the additional constraints can significantly reduce the complexity of the analysis on a theoretical level. Balasubramanyan and Snyder [BAL87a,BAL87b,BAL88] have developed an algorithm to extract the parameters of motion in depth: the single component of translation in depth (i.e. parallel to the line of sight) and the two components of rotation in depth (i.e. rotations that are not around the line of sight). This is achieved by building upon the work of Adiv for segmenting the flow field into rigid independently moving objects [ADI85a], and the formulation of Waxman and Duncan [WAX86] of the ratio of the relative optic flow between a stereo pair of images to the disparity between them as a linear function of the image coordinates. Experimental results are presented for simulated data of general motion of both the sensor and independently moving objects. Work is currently underway to test the effectiveness on real scenes.

3.4 Analysis of Constant General Motion

Another way to introduce additional constraints to the problem of general motion analysis in an effort to achieve practical, robust algorithms is via Shariat's formulation: constant but arbitrary general motion of a rigid object [SHA86]. This leads to a set of difference equations across a sequence of images, relating the positions of a feature in the image plane to the motion parameters of the projected point. The solution obtained is a set of 5th order non-linear polynomial equations in the unknown motion parameters, whose solution requires a Gauss-Newton non-linear least-squares method with carefully designed initial guess schemes. Pavlin [PAV88] has derived a closed-form solution for the rigid object trajectory by integrating the differential equations describing the motion of a point on the tracked object. The integrated equations are non-linear only in angular velocity, and are linear in all other motion parameters. These equations allow the use of a simple least-square error minimization criterion in an iterative search for the motion parameters.

4 Perceptual Organization (Grouping)

4.1 Token-Based Approaches to Motion and Perceptual Organization

The problems cited in Section 3 with respect to the extraction of motion and depth information using traditional optical flow techniques have led us toward the exploration of methods for combining the local flow/displacement fields with larger token-like structures. It is our position that the inherently local measurement of visual motion provided by optical flow is insufficient to meet the varied requirements of dynamic image understanding. The approach we are developing involves computing the correspondence between tokens of arbitrary spatial scale produced by perceptual organization processes. Such tokens often map directly to environmental structure, and descriptions of their movement often correlate more closely with the motion of physical objects than does the local motion information contained in the displacement field. A token match represents more than just a spatial displacement; also explicit in this representation are the time-varying values of those parameters which define the token or which can be extracted from the structure of the token.

In two papers in this volume, Williams and Hanson [WIL88a,WIL88b] describe work in progress toward this goal. The premise of this work is that the structure obtained from perceptual organization processes can be combined with the local motion information contained in the flow field to provide a more robust estimate of motion and depth parameters. The approach can be viewed as augmenting the rather limited use of spatial structure in tradional approaches with the richer descriptive vocabulary of spatial structure provided by the perceptual organizational processes over both space and time. In this sense, the spatially organized structures (such as lines, regions, curves, vertices, intersections, rectangular groups, etc.), which are actively constructed from the image, can be considered to be interest operators of large spatial extent.

In the first paper [WIL88a], a method for computing the temporal correspondence between straight line segments is presented. We consider the two frame case here, but the method is extensible and has been extended to multiple frames. A straight line perceptual organization process, developed by Boldt and Weiss [BOL87,WEI86], is applied to both frames independently to provide straight lines in each frame. A displacement field is also computed from the two frames using the algorithm developed by Anandan [ANA87b,ANA88]. After filtering the straight lines on length and constrast to reduce the line set in both images, the displacment field is used to construct a search area in frame 2 for each line in frame 1. Since a one-to-one correspondence between lines is unlikely, a minimal mapping approach [ULL79] is used to compute the correspondence between the frame 1 and frame 2 line sets; such a mapping is called a minimal bipartite cover. The similarity measure used to compute the cover involves the similarity and spatial separation of the candidate token matches. By computing the connected components of the bipartite graph, the global matching problem is conveniently divided into smaller, individually tractable pieces which reflect the scope of potential interactions; a simple blind search of the subgraphs is used to extract the bipartite cover minimizing the positional and similarity discrepancy metric.

The matching results obtained are quite good. The system has been run repeatedly on successive frames of several multi-frame sequences. In the multi-frame case, a directed acyclic graph is constructed which represents the splitting and merging patterns of line segments over time. Work is in progress to analyze the trajectories of the tokens over time.

In the second paper [WIL88b], a method for computing depth from the line correspondences is described using the temporal change in the length of virtual lines constructed from the intersections of the Boldt lines [BOL87]. We use virtual lines because the length of the original lines is not reliable, although their orientation and lateral displacement is quite precise. This "looming" method is also generalized to areas. The method is generally applicable to structures whose total extent in depth is small compared to the depth of its centroid (that is, for those cases in which perspective projection can be approximated by scaled orthographic projection [THOM87]) and which do not exhibit any independent motion. The technique does not depend on the complete determination of egomotion parameters of the sensor, but it does require the computation of the translation component of the sensor in the direction of motion. An analysis of the sensitivity of the algorithm to errors in the measured variables is planned for the near future; experimental results on real image sequences have shown that the algorithm may be quite robust.

4.2 The Perceptual Organization of Image Curves

Most of our work in perceptual organization [BEL86,BOL87,BUR86,DOL86,DOL88,REY84] [REY86b,REY87,RIS87,WEI85,WEI86,WIL88a,WIL88b] has been focussed on rectilinear structures (e.g. straight lines, corners, parallel line pairs, and the like). Unfortunately, not all of the world can be described by straight lines. Consequently, Dolan [DOL88] has been exploring methods for extending the general technique developed by Boldt [BOL87,WEI86] to the simultaneous extraction of curves, straight lines, and corners (including cusps); these are the primitive descriptive elements. The basic operation cycle consists of linking, grouping, and replacement, which takes place at increasing perceptual scales, resulting in a hierarchical scale-space description of these important image events.

The linking stage finds subsets within the set of initial local edge tokens that satisfy the binary constraints of the particular grouping principles employed. The grouping mechanisms perform a detailed geometric analysis on sets of linked tokens whose extent is within the current scale; in Dolan's system, this also entails classification and ranking of the token sequences as one of the basic primitive elements. Replacement mechanisms encode the geometry of a surviving group by substituting a single token for the group. The process then repeats at the next scale.

4.3 Extracting Geometric Structure

Reynolds and Beveridge [REY87] have been developing a perceptual grouping system for the extraction of rectilinear structures from an initial set of line primitives obtained using the straight line extraction algorithm developed by Burns, Hanson, and Riseman [BUR86]. The lines are represented as nodes in a graph. The grouping criteria are the geometric relations of spatially proximate collinear, spatially proximate parallel, spatially proximate orthogonal, or any subset of these relations; the relations form the links in the graph.

Line groups are generated using a connected components analysis of the chosen geometric links. Finally, individual geometric structures (e.g. rectangles, collinear lines, parallel line pairs, corners, etc.) may be identified as subgraphs of the connected components. These techniques have been applied to extraction of objects such as road networks in aerial images.

Object recognition strategies can be represented as relational graphs to be matched to extracted data. The problems associated with fragmentation, as well as merged and missing tokens, makes this a difficult problem. However, multiple representations (such as lines and regions) can be brought together to provide partial redundancy [RIS87]. Thus, current work overlaps issues of constrained graph matching, perceptual organization, and information fusion.

5 Image Understanding Architecture

The Image Understanding Architecture being jointly designed and constructed by the University of Massachusetts and Hughes Research Laboratories [WEE84,WEE87,WEE88a,WEE88b] is a multi-level, heterogeneous, associative, parallel machine to support real-time knowledge-based computer vision. The machine consists of three computational levels (the CAAPP, ICAP, and SPA, respectively), roughly corresponding to the low, intermediate, and high-levels of abstraction currently believed necessary for image understanding [HAN87a]. At each level of the IUA, the processing elements are tuned to the computational granularity of the algorithms required at that particular level of abstraction.

The CAAPP (Content Addressable Array Parallel Processor) is a 512x512 square grid array of custom processors designed to perform low-level image processing tasks. The CAAPP is also specifically designed to interact with the ICAP (through a shared memory) in a tightly coupled fashion for both bottom-up and top-down processing. The ICAP (the Intermediate Communications Associative Processor) is designed to manipulate tokens either extracted from the image data at the CAAPP level or constructed from other tokens at the ICAP level. The ICAP is also a square grid (64x64), built from Texas Instruments TMS320-C25 signal processing chips, each with 256K bytes of local memory, and 384K bytes of dual-ported memory for CAAPP/ICAP and ICAP/SPA communication and data storage. The ICAP can operate in either synchronous MIMD or pure MIMD mode.

The SPA (Symbolic Processing Array) is constructed of processors capable of running a LISP-based blackboard system at each node. The SPA processors operate in MIMD mode with intra-level communication through the blackboard and inter-level communication through the dual-ported memory shared with the ICAP processors. Since only a 1/64th prototype is currently being built, the SPA level will consist of a single processor in the Motorola 68020 class; in the full machine the SPA level will consist of 64 or more processors, each capable of running LISP. The detailed architecture of the SPA level of the full machine has not yet been fully defined, but research in progress oriented towards porting the schema system to a commercial shared-memory multiprocessor is expected to provide insight into its structure.

The CAAPP and ICAP levels are controlled by a dedicated Array Control Unit which contains two separate control processors. The Macro-controller is a 68020-based system which supports the software development environment and provides an interface to the programmer. The Micro-controller is a custom processor driven by horizontal microcode and is capable of issuing an instruction to the CAAPP every 100 nanoseconds. Access to the Micro-controller is through a library of CAAPP control subroutines. In this way, the advantages of a high-level program development environment are combined with the speed advantages of the Micro-controller. The ACU is also accessible from the SPA level, providing knowledge-driven control of both the low-level and the intermediate-level processes. The programmer's model for this environment is described in [WEE88b].

5.1 The Coteric Network on the CAAPP

The requirements of high-speed, fine-grained bi-directional inter- and intra-level communication and control have led to the development of very general associative processing techniques to support the communication requirements [WEE88a,WEE88b]. Currently, there are four mechanisms for communication between CAAPP cells. One method used the hardware implementation of the associative processing capabilities to accomplish global feedback and rebroadcast. Communication can also take place through the ICAP level via the backing store. The third mechanism uses the traditional S,E,W,N neighborhood network between adjacent CAAPP processors.

The fourth mechanism involves a new and powerful variation on the nearest neighbor mesh called the Coterie Network. The coterie technique allows the CAAPP mesh to be partitioned into independent groups of processors that share a local associative Some/None circuit. The independent groups of processors can then respond to globally broadcast instructions in a locally data-dependent way, which permits the parallelism in the mesh to be more flexibly exploited. For example, each coterie might correspond to a single region in a region segmentation; each region could then be processed independently and in parallel.

The coterie mechanism is implemented through a network of software controlled switches, one switch between each adjacent processor. Opening the switch between two processors effectively isolates them from communicating with each other over the open line. Creating a closed path of open switches creates an island of processors isolated from the remainder of the mesh. Leaving the switches inside the 'island boundary' closed creates an internal communication network for the processors participating in the coterie. Each processor may write or read a bit from the network; when more than one processor writes to the network, the result is the logical OR of the output bits of the processors. The shared mesh is thus functionally equivalent to the global Some/None network, but local to the coterie. Several image processing algorithms which utilize the Coterie Network are discussed in [WEE88b].

5.2 Status

The IUA programming environment currently exists in software simulation on an Explorer LISP workstation, augmented with a Texas Instruments Odyssey parallel signal co-processor. Portions of the simulation are also available on VAX and SUN systems.

At the hardware level, final versions of the custom CAAPP chips are currently being fabricated through MOSIS. A CAAPP-CISM-ICAP-ISSM test structure has been successfully breadboarded by Hughes Research Laboratories. The prototype IUA system is scheduled for completion in the

Fall of 1988.

6 Integrated Vision Benchmark for Parallel Architectures (DARPA IU Benchmark: Round 2)

The most recent attempt at constructing a vision benchmark for parallel architectures emerged from the DARPA IU community in 1986. This benchmark consists of ten prototypical vision tasks: Gaussian convolution, zero-crossing detection and output of border lists, connected components labeling, Hough transform, determining the convex hull, constructing a minimal spanning tree, computing the visibility of vertices in a 3-D model, finding a minimal cost path, and subgraph matching. Each of these tasks were benchmarked individually and the results reported in [ROS87].

While useful information was gained from this exercise, there were significant shortcomings. In particular, it was recognized that the individual benchmarks did not adequately represent the performance of a machine on an integrated vision task, such as knowledge-based image interpretation. In response a group of researchers from the University of Maryland and the University of Massachusetts accepted the responsibility of formulating an integrated benchmark representing a more realistic interpretation scenario that transcends several different representations and forms of processing that are typical of complex vision applications. The goal was to generate a benchmark which would lead to a better understanding of vision architecture requirements and the performance bottlenecks in different classes of machines. A secondary goal was to utilize as many of the algorithms as possible from the first benchmark in order to minimize the coding impact on participants of the second benchmarking task. The benchmark was developed with some input from other members of the DARPA IU community [WEE88c].

The integrated benchmark involves recognizing an approximately specified "2 1/2 D mobile" sculpture composed of rectangles, given images from intensity and range sensors. The test images are designed so that the data from neither sensor by itself is sufficient to solve the task. The sculpture can be thought of as a semi-rigid mobile consisting of suspended rectangles floating in space with spatial relationships that are fixed only up to some tolerance limit. Each rectangle is oriented normal to the Z axis (the viewing direction) and the image is constructed under orthographic projection. In the image, the rectangles comprising the mobile are interspersed with additional rectangles. The additional rectangles may occlude portions of the mobile object, and some of the adjacent rectangles in the scene may have very similar brightnesses and depths.

A set of symbolic models of several mobile sculptures will be provided when the benchmark is distributed. These models are only approximate in that the sizes, orientations, depths, and spatial relationships of the rectangles are constrained within some tolerance bands. The goal is to determine which of the models are present in the images, the degree to which it is visible (matchable), and to update the model with positional data that has been extracted from the images.

In order to ensure comparable benchmarking results, an algorithm-level solution to the interpretation problem is provided, as well as a set of instrumentation guidelines. It is not intended that this solution be optimal or even that it constitutes a reasonable approach. The goal is to provide a measure of consistency among the solution methods and benchmark results. The solution has been tested on sequential machines by the University of Massachusetts and the University of Maryland in order to remove ambiguity in its statement and to detect as many unexpected problems in its implementation as possible. The instrumentation guidelines are currently being developed and reviewed, and will be distributed at a later date. The benchmark is intended to be widely distributed, and a workshop for comparison of results on a variety of machines will be held at a future date.

7 Mobile Vehicle Navigation

The hardware platform for experimentation in mobile robotics at UMass is a Denning Mobile Robotics vehicle with a B&W television camera and UHF transmitters and receivers for uplink and downlink communication to a Gould IP8500 image processing system connected to a Vax

11/750 computer. Plans are underway to utilize a 12-node Sequent multiprocessor to improve the computational effectiveness of our experimentation environment.

Arkin [ARK87a,ARK87b,ARK87c] used this platform to develop AuRA (Autonomous Robot Architecture), which integrates planning, cartographic, perception, motor, and homeostatic systems into a functional robot navigation system. The system is designed to navigate in the hallways and outdoor environment surrounding our building at UMass.

Aura employs a 'meadow' map as its long-term memory; the meadow map is used for global path planning and contains embedded a priori knowledge to guide sensor expectations used for positional updating. A layered short-term memory based on instantiated meadows represents the currently perceived world. A hierarchical path planner produces a global path free of collisions with all modeled obstacles.

Aura extends the idea of schemas, as currently employed in the VISIONS system, to include the mobile robot domain. The schema-based path execution system handles unexpected and dynamic obstacles not present in the robot's world model. This motor-schema based navigation system produces reactive/reflexive behavior in direct response to sensor events. In addition, new techniques in the treatment of robot uncertainty, which expedite sensory processing, were developed. These include the use of a spatial error map with associated growth and reduction techniques.

Several computer vision sensor strategies have been developed for use within Aura. These include a fast line finding algorithm that is a simplified and more efficient version of the Burns straight line extraction algorithm (at the price of robustness) [BUR86,KAH87], a fast simplified region segmentation algorithm based on the VISIONS region segmentation system [BEV87] , and a depth from motion algorithm [BHA86]. Aura uses both vision and ultrasonic sensing during path traversal.

We are currently rebuilding Aura to make better use of the information available from the visual sensors and to more completely integrate the full spectrum of image understanding techniques developed in the VISIONS project. In particular, we intend to utilize some of the depth from motion algorithms discussed above [BAL88,DUT88,WIL88a,WIL88b]; and some of the simpler object recogniton strategies of the schema system [DRA87a,DRA87b,HAN87b], including strategies for multi-sensor information fusion [BEL86,RIS87].

REFERENCES

[ADI85a] G. Adiv, "Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume PAMI-7, July 1985, pp. 384-401.

[ADI85b] G. Adiv, "Interpreting Optical Flow," Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts at Amherst, September 1985.

[ADI85c] G. Adiv, "Inherent Ambiguities in Recovering 3-D Motion and Structure from a Noisy Flow Field, " *Proc. of the Computer Vision and Pattern Recognition Conference*, San Francisco, CA, June 1985, pp. 7077.

[ANA84] P. Anandan, "Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion," *SPIE Intelligent Robots and Compter Vision Conference,* Volume 521, 1984, pp. 184-194; also *DARPA IU Workshop Proceedings,* 1984; and COINS Technical Report 84-32, University of Massachusetts at Amherst, December 1984.

[ANA85a] P. Anandan and R. Weiss, "Introducing a Smoothness Constraint in a Matching Approach for the Computation of Optical Flow Fields," *Proc. of the Third Workshop on Computer Vision: Representation and Control,* October 1985, pp. 186-196, also in *DARPA IU Workshop Proceedings,* 1985.

[ANA85b] P. Anandan, "Motion and Stereopsis," COINS Technical Report 85-52, University of Massachusetts at Amherst, December 1985; also to appear (in Spanish) in *Vision por Computador,* (Carme Torras, ed.), to be published by Alianza Editorial, Spain.

[ANA87a] P. Anandan, "Measuring Visual Motion From Image Sequences", Ph.D. Dissertation, University of Massachusetts at Amherst, January 1987.

[ANA87b] P. Anandan, "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion", *Proc. of the International Conference on Computer Vision,* London, England, June 1987.

[ANA88] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion", to appear in *International Journal on Computer Vision*, 1988.

[ARK87a] R. Arkin, "Motor Schema-Based Navigation for a Mobile Robot: An Approach to Programming by Behavior", *Proc. of the IEEE International Conference on Robotics and Automation*, Raleigh, NC, pp. 264-271, March 1987.

[ARK87b] R. Arkin, "Towards Cosmopolitan Robots: Intelligent Navigation in Extended Man-Made Environments", Ph.D. Thesis, Computer & Information Science Department, University of Massachusetts at Amherst, September 1987.

[ARK87c] R. Arkin, A. Hanson, and E. Riseman, "Visual Strategies for Mobile Robot Navigation", *Proc. of IEEE Computer Society Workshop on Computer Vision*, Miami, FL November, 1987.

[BAL87a] P. Balasubramanyam, "Extraction of Motion In Depth: A First Step in Stereoscopic Motion Interpretation", presented at the American Institute of Aeronautics and Astronautics Computers in Aerospace IV Conference, Wakefield, MA, October 1987, pp. 277-286.

[BAL87b] P. Balasubramanyam, "Computation of Motion-In-Depth Parameters Using Stereoscopic Motion Constraints", *Proc. of IEEE Computer Society Workshop on Vision*, Miami, FL, November 1987.

[BAL88] P. Balasubramanyam and M. Snyder, "Computation of Motion in Depth Parameters: A First Step in Stereoscopic Motion Interpretation", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[BEL86] R. Belknap, E. Riseman, and A. Hanson, "The Information Fusion Problem and Rule-Based Hypotheses Applied to Complex Aggregations of Image Events", *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 1986, pp. 227-234.

[BEV87] J. Ross Beveridge, J. Griffith, R. Kohler, A. Hanson, and E. Riseman, "Segmenting Images Using Localized Histrograms and Region Merging", COINS Technical Report 87-88, University of Massachusetts at Amherst, September 1987.

[BHA85] S. Bharwani, A. Hanson, and E. Riseman, "Refinement of Environmental Depth Maps over Multiple Frames," *Proc. DARPA IU Workshop,* Miami Beach, FL, December 1985.

[BHA86] S. Bharwani, E. Riseman, and A. Hanson, "Refinement of Environmental Depth Maps Over Multiple Frames", *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, SC, May 1986, pp. 73-80.

[BOL87] M. Boldt and R. Weiss, "Token-Based Extraction of Straight Lines", COINS Technical Report 87-104, University of Massachusetts at Amherst, October 1987.

[BRO88] J. Brolio, J. Ross Beveridge, B. Draper, J. Griffith, G. Reynolds, R. Collins, J. Burrill, L. Williams, A. Hanson, E. Riseman, "Symbolic Data Management for Vision Research", COINS Technical Report, University of Massachusetts at Amherst, 1988.

[BUR88] J.B. Burns and Leslie J. Kitchen, "Rapid Object Recognition From A Large Model-Base Using Prediction Hierarchies", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988. Also COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[BUR87a] J.B. Burns, "Recognition in 2D Imges of 3D Objects from Large Model Bases Using Prediction Hierarchies", *Proc. of the International Joint Conference on Artificial Intelligence*, Milan, Italy, August 1987, pp. 763-766.

[BUR87b] J.B. Burns and L.J. Kitchen, "An Approach to Recognition in 2D Images of 3D Objects From Large Model Bases", COINS Technical Report 87-85, University of Massachusetts at Amherst, August 1987.

[BUR86] J.B. Burns, A.R. Hanson, and E.M. Riseman, "Extracting Straight Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*, No. 4, July 1986, 425-455.

[DEM68] A.P. Dempster, "A Generalizaton of Bayesian Inference", *Journal of the Royal Statistical Society*, Series B., Vol. 30, 1968, pp. 205-247.

[DOL88] John Dolan and Richard Weiss, "The Perceptual Organization of Image Curves", COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[DOL86] J. Dolan, G. Reynolds, and L. Kitchen, "Piecewise Circular Description of Image Curves Using Constancy of Grey-Level Curvature", COINS Technical Report 86-33, University of Massachusetts at Amherst, July 1986.

[DRA88] B. Draper, J. Brolio, R. Collins, A. Hanson, and E. Riseman, "Information Fusion by Distributed Subsystems in a Knowledge-Based Vision Architecture", to appear in *Computer Vision and Pattern Recognition*, 1988.

[DRA87a] B. Draper, R. Collins, J. Brolio, J. Griffith, A. Hanson, and E. Riseman, "Tools and Experiments in the Knowledge-Directed Interpretation of Road Scenes", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 178-193.

[DRA87b] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman, "The Schema System", COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[DUT88] R. Dutta, R. Manmatha, E. Riseman, and M.A. Snyder, "Issues in Extracting Motion Parameters and Depth from Approximate Translational Motion", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[FIS86] M.A. Fischler and R.C. Bolles, "Perceptual Organization and Curve Partitioning", *IEEE PAMI*, 8(1), pp. 100-105, January 1986.

[GLA87a] F. Glazer, "Hierarchical Motion Detection", Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts, Amherst, February 1987.

[GLA87b] F. Glazer, "Hierarchical Gradient-Based Motion Detection", COINS Technical Report 87-77, University of Massachusetts at Amherst, 1987.

[GLA87c] F. Glazer, "Computation of Optic Flow by Multilevel Relaxation", Coins Technical Report 87-64, University of Massachusetts at Amherst, 1987.

[HAN78a] A.R. Hanson and E.M. Riseman (Eds.), *Computer Vision Systems*, New York, Academic Press, 1978.

[HAN78b] A. Hanson and E. Riseman, "Segmentation of Natural Scenes," in *Computer Vision Systems*, (A. Hanson and E. Riseman, Eds.), Academic Press 1978, pp. 129-163.

[HAN78c] A. Hanson, and E. Riseman, "VISIONS: A Computer System for Interpreting Scenes, in *Computer Vision Systems* (A. Hanson and E. Riseman, eds.) pp. 303-333, Academic Press, 1978.

[HAN86] A.R. Hanson and E.M. Riseman, "A Methodology for the Development of General Knowledge-Based Vision Systems," in *Vision, Brain, and Cooperative Computation,* (M. Arbib and A. Hanson, Eds.) 1986, MIT Press, Cambridge, MA.

[HAN87a] A.R. Hanson and E.M. Riseman, "The VISIONS Image Understanding System", in *Advances in Computer Vision,* Chris Brown, (Ed.), Erlbaum Press, 1987. Also COINS Technical Report 86-62, University of Massachusetts at Amherst, December 1986.

[HAN87b] A. Hanson and E. Riseman, "From Image Measurements to Object Hypotheses", COINS Technical Report 87-129, University of Massachusetts at Amherst, December 1987.

[KAH87] P. Kahn, L. Kitchen, and E. Riseman, "Real-Time Feature Extraction: a Fast Line Finder for Vision-Guided-Robot Navigation", COINS Technical Report 87-57, University of Massachusetts at Amherst, July 1987.

[KOH87] C.A. Kohl, A.R. Hanson, and E.M. Riseman, "A Goal-Directed Intermediate Level Executive for Image Interpretation", *Proc. of the International Joint Conference on Artificial Intelligence,* Milan, Italy, August 1987, pp. 811-814.

[KOH88] C.A. Kohl, "Goal-Directed Control for Computer Vision", Ph.D. Dissertation, Computer & Information Science, University of Massachusetts at Amherst, February 1988.

[LAW84] D.T. Lawton, "Processing Dynamic Image Sequences from a Moving Sensor," Ph.D. Dissertation (COINS Technical Report 84-05), Computer and Information Science Department, University of Massachusetts, 1984.

[LAW83] D.T. Lawton, "Processing Translational Motion Sequences", *Computer Graphics and Image Processing,* Vol. 22, pp. 116-144, 1983.

[LEH87] N. Lehrer, G. Reynolds, and J. Griffith, "A Method for Initial Hypothesis Formation in Image Understanding", COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1987.

[LOW85] D.G. Lowe, *Perceptual Organization and Visual Recognition,* Kluwer Academic Publishers, 1985.

[PAV85] I. Pavlin, A. Hanson, and E. Riseman, "Analysis of an Algorithm for Detection of Translational Motion," *Proc. DARPA IU Workshop,* Miami Beach, FL, December 1985.

[PAV86] I. Pavlin, E. Riseman, and A. Hanson, "Translational Motion Algorithm With Global Feature Constraints", COINS Technical Report 86-58, University of Massachusetts at Amherst, December 1986.

[PAV88] I. Pavlin, "Motion From a Sequence of Images", *Proc. of the DARPA Image Understanding Workshop,* Cambridge, MA, April 1988.

[REY86a] G. Reynolds, D. Strahman, N. Lehrer, L. Kitchen, "Plausible Reasoning and the Theory of Evidence", COINS Technical Report 86-11, University of Massachusetts at Amherst, April 1986.

[REY86b] G. Reynolds, J. Ross Beveridge, "Geometric Line Organization Using Spatial Relations and a Connected Components Algorithm", COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1986.

[REY84] G. Reynolds, N.Irwin, A. Hanson and E. Riseman, "Hierarchical Knowledge- Directed Object Extraction Using a Combined Region and Line Representation," *Proc. of the Workshop on Computer Vision: Representation and Control,* Annapolis, Maryland, April 30 - May 2, 1984, pp. 238-247.

[REY85] G. Reynolds, D. Strahman, N. Lehrer, "Converting Feature Values to Evidence," *Proc. DARPA IU Workshop,* Miami Beach, FL, 1985.

[REY87] G. Reynolds and J.R. Beveridge, "Searching For Geometric Structure in Images of Natural Scenes", COINS Technical Report 87-03, University of Massachusetts at Amherst, January 1987.

[RIS87] E. Riseman, A. Hanson, and R. Belknap, "The Information Fusion Problem: Forming Token Aggregations Across Multiple Representations", COINS Technical Report 87-48, University of Massachusetts at Amherst, 1987.

[RIS84] E.M. Riseman and A.R. Hanson, "A Methodology for the Development of General Knowledge-Based Vision Systems," *IEEE Proc. of the Workshop on Computer Vision: Representation and Control*, 1984, pp. 159-170.

[ROS87] A. Rosenfeld, "A Report on the DARPA Image Understanding Architectures Workshop", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 298-302.

[SHA76] G. Shafer, "A Mathematical Theory of Evidence", Princeton University Press, 1976.

[SHA86] H. Shariat, "The Motion Problem: How To Use More Than Two Frames", Ph.D. Dissertation, University of Southern California, Los Angeles. Also Technical Report IRIS 202, Institute for Robotics and Intelligent Systems, University of Southern Calfornia, Los Angeles, CA, October 1986.

[SNY86] M. A. Snyder, "The Accuracy of 3D Parameters in Correspondence-Based Techniques," *Proc. of the Workshop on Motion: Representation and Analysis,* Charleston, S.C., May 1986; also COINS Technical Report 86-28, University of Masschusetts at Amherst, July 1986.

[SNY88] M.A. Snyder, "The Precision of 3D Parameters in Correspondence-Based Techniques: The Case of Uniform Translational Motion in a Rigid Environment", to appear in *IEEE PAMI*, 1988.

[THOR87] C. Thorpe, S. Shafer, and T Kanade, "Vision and Navigation for the Carnegie Mellon Navlab", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 143-152.

[THOM87] D. Thompson and J. Mundy, "Three Dimensional Model Matching From an Unconstrained Viewpoint, *Proc. of the IEEE Conference on Robotics and Automation,* Raleigh, NC, 1987.

[ULL79] S. Ullman, "The Interpretation of Visual Motion", MIT Press, Cambridge, MA, 1979.

[WAX86] A.M. Waxman and J.H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion", *IEEE PAMI*, Vol. PAMI-8, November 1986, pp. 715-729.

[WEE84] C. Weems, "Image Processing on a Content Addressable Array Parallel Processsor", Ph.D. Dissertation and COINS Technical Report 84-14, University of Massachusetts at Amherst, September 1984.

[WEE87] C. Weems, S. Levitan, A. Hanson, and E. Riseman, "The Image Understanding Architecture Workshop", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 483-498.

[WEE88a] C. Weems, S. Levitan, A. Hanson, E. Riseman, G. Nash, and D. Shu, "The Image Understanding Architecture", to appear in The International Journal on Computer Vision, 1988. Also Technical Report 87-76, Computer & Information Science, University of Massachusetts at Amherst, August 1987.

[WEE88b] Charles Weems, "Some Sample Algorithms for the Image Understanding Architecture", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WEE88c] C. Weems, E. Riseman, A. Hanson, and A. Rosenfeld, "An Integrated Image Understanding Benchmark: Recognition of a 2 1/2 D 'Mobile'", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WEI86] R. Weiss and M. Boldt, "Geometric Grouping Applied to Straight Lines", *Proceedings on the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, Florida, June, 1986, pp. 489-495.

[WEI85] R. Weiss, A. Hanson, and E. Riseman, "Geometric Grouping of Straight Lines," *Proc. 1985, DARPA IU Workshop*, Miami Beach, FL, 1985

[WES82] L. Wesley and A. Hanson, "The Use of Evidential-Based Model for Representing Knowledge and Reasoning About Images in the VISIONS System", *Proc. Workshop on Computer Vision*, Rindge, NH, August 1982.

[WES83] L. Wesley, "Reasoning About Control: The Investigation of an Evidential Approach", *Proc. 8th IJCAI*, Karlsruhe, West Germany, August 1983, pp. 203-210.

[WES86] L. Wesley, "Evidential Knowledge-Based Computer Vision", *Optical Engineering*, Vol. 25, No. 3, March 1986, pp. 363-379.

[WES88] L. Wesley, "Evidential-Based Control in Knowledge-Based Systems", Ph.D. Dissertation, Computer & Information Science, University of Massachusetts at Amherst, February 1988.

[WEY86] T.E. Weymouth, "Using Object Descriptions in a Schema Network For Machine Vision," Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts at Amherst. Also COINS Technical Report 86-24, University of Massachusetts at Amherst, 1986.

[WIL86] Lance R. Williams and P. Anandan, "A Coarse-to-Fine Control Strategy for Stereo and Motion on a Mesh-Connected Computer", COINS Technical Report 86-19, University of Massachusetts at Amherst, May 1986.

[WIL88a] L. Williams and A. Hanson, "Depth From Looming Structure", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WIL88b] L. Williams and A. Hanson, "Translating Optical Flow Into Token Matches", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WIT83] A.P. Witkin and J.M. Tenenbaum, "What Is Perceptual Organization For?", *IJCAI*, 1983, pp. 1023-1026.