Progress in Computer Vision at the University of Massachusetts

Allen R. Hanson
Edward M. Riseman

COINS TR 89-44

May 1989

Progress in Computer Vision at the University of Massachusetts[1]

Allen R. Hanson and Edward M. Riseman
Computer Vision Research Laboratory
Dept. of Computer and Information Science
University of Massachusetts
Amherst, MA 01003

ABSTRACT

This report summarize progress in image understanding research at the University of Massachusetts over the past year. Many of the individual efforts discussed in the paper are further developed in other papers in this proceedings. The summary is organized into several areas:

I. Autonomous Vehicle Navigation
II. Motion Processing
III. Knowledge-Based Interpretation
IV. Image Understanding Architecture

The research program in computer vision at UMass has as one of its goals the integration of a diverse set of research efforts into a system that is ultimately intended to achieve real-time image interpretation in a variety of vision applications. A highlight of our recent research effort is the initial integration of a perceptually-based navigation system for our local mobile robot; this work is represented in four other papers in this proceedings.

I. AUTONOMOUS VEHICLE NAVIGATION

I.1. Mobile Robot Project

The UMass Mobile Robot project is investigating the problem of enabling a mobile automaton to navigate intelligently through indoor and outdoor environments. At the foundation of our work is the premise that higher-level vision beyond the first stages of sensory processing is needed for perceptual control of the robot. In particular the system will greatly benefit from, and in many cases require, the use of knowledge and models of objects in the environment.

This project is discussed in several papers in this proceedings: our approach to world modeling, planning, and primitive task execution are presented in Fennema et al (Fennema, Hanson et al. 1989); the world model is developed in a solid modelling package, Geometer, described in (Connolly and Weiss 1989); mechanisms for optimal 2D model matching, used to locate landmarks derived from the world model and an estimate of the robot's current position, are discussed in (Beveridge, Weiss et al. 1989b); and methods for determining the pose of the robot from the matches are developed in (Kumar 1989).

In the early phases of this research, we wish to balance generality with setting sufficient constraints on the initial research goals to be achievable. Therefore, the experiments focus on robust goal-oriented navigation through a partially-modeled, unchanging environment which does not include any unmodeled obstacles. Later experiments will soften these

---

[1] This work was supported in part by the Defense Advanced Research Projects Agency under contract numbers F30602-87-C-0140, DACA76-85-C-0008, and DACA76-86-C-0015, and by the National Science Foundation under grant number DCR-8500332.

constraints to deal with unmodeled or moving objects, as well as learning in partially modeled or new environments.

I.2. Implementation on a Multiprocessor

Mobile robot navigation is a computationally intensive activity. One approach to achieving real-time navigation that we are investigating involves the use of a shared-memory multiprocessor (Sequent Symmetry). A C-based language called RS (Pocock) has been developed for control of a set of real-time cooperative processes, and it has been ported to the Sequent. Similarly, a group of about ten high level modules involved in the navigation system are also being ported.

The near-term goal of this effort is to reproduce the experimental capability of navigating in a modeled environment. Here, the systems issues are paramount and issues of scheduling limited resources with hard real-time constraints is our research focus. We hope to demonstrate the multiprocessor version of mobile robot navigation by early Fall,1989.

II. MOTION PROCESSING

A major area of research in our laboratory is the analysis of sequences of images derived from a moving sensor. The environments that we are examining include indoor hallway and room scenes, as well as outdoor scenes of the UMass campus. The goals of this work include the recovery of sensor motion decomposed into its translational and rotational components, the recovery of environmental depth of key surfaces and objects, and the detection of independently moving objects and, if possible, their motion parameters.

II.1. Stereoscopic Motion Analysis and the Detection of Discontinuities

One of the most important problems in stereo and motion processing is the recovery of depth and motion boundaries. A number of algorithms for computing optic flow make a global smoothness assumption that tends to unnaturally smooth across depth and motion discontinuities, and this makes later detection of these boundaries very difficult. On the other hand, knowledge of these discontinuities is very important for the flow and disparity computations to be correct, especially at occlusion boundaries.

One approach to this problem is to integrate motion and stereo data. (Balasubramanyam 1989) uses information in both the stereo and motion sequences at two time instances to define a measure of confidence in the presence of motion and depth discontinuities. This measure can be applied early, prior to the full computation of flow and disparity fields. The general idea is to use coarse disparity and flow estimates from hierarchical correlation processes (Anandan 1989) to locate and label depth and motion discontinuities; smoothing is then inhibited across these boundaries. Discontinuities that are continuous (i.e. unbroken) in the other dimension are favored. Initial results are presented on both synthetic and real stereo-motion imagery.

II.2. Smoothness Constraints for Optic Flow and Surface Reconstruction

Snyder (Snyder 1989) has developed a theoretical analysis of smoothness constraints that is used in the computation of optic flow and surface reconstruction. It is derived from a mathematical foundation that lends insight to the heuristic justification of other smoothness constraints. Under several simple assumptions he derives the most general possible smoothness constraint, which turns out to be quadratic in the first derivatives of the flow field, and quadratic in the first and second derivatives of the grey-level image intensity function. All the best-known smoothness constraints are special cases of this general form, and the relationship to a few are examined.

II.3. 3D Interpretation of Rotational Motion from Image Trajectories

The research of Sawhney and Oliensis (Sawhney and Oliensis 1989) addresses the problem of discovering the motion parameters of independently moving objects in their natural coordinate system. This paper focuses on analyzing an extended time sequence of images of an object rotating uniformly around an axis of arbitrary location and orientation. It demonstrates how the abstraction of continuous descriptions of multi-frame data can lead to the recovery of scene motion and structure. Image traces of 3D feature points are generated from image point correspondences over a sequence of frames. These traces are described by continuous curves that are obtained by fitting conic arcs to the set of points. The goal is motion-based grouping of image traces to provide constraints (that are unavailable in only a few frames) sufficient to extract the motion parameters of independently moving objects in their natural coordinate system.

II.4. A Motion Data Set from the ALV

Motion analysis has remained an extremely difficult research area. One of the difficulties has been the lack of motion data with ground truth of known accuracy. In particular, this sort of data has not been collected for robot vehicles operating under realistic conditions in outdoor environments. Thus, the proper scientific evaluation of motion algorithms intended for practical application has been impossible.

In response to this general problem, our group decided to collect a reasonably large data set from the ALV (Dutta, Manmatha et al. 1989a; Dutta, Manmatha et al. 1989b) Motion sequences of about 30 frames each were collected at five different outdoor sites with different road surfaces, including on-road, dirt-road, and off-road scenarios, Data from the video camera, laser range finder, and land navigation system were recorded simultaneously under stop-and-shoot and move-and-shoot scenarios. Ground truth data was obtained using traditional surveying methods. The data is being made available to the general community and can be obtained by communicating with Ms. Valerie Cohen at UMass (E-mail address is VCohen@CS.UMass.EDU).

III. KNOWLEDGE-BASED INTERPRETATION

III.1. New Schema system

Research in knowledge-directed vision has begun to focus on two new goals. The first is to build a control system for vision that is provably near-optimal under certain assumptions. The second is to use machine learning techniques to assume part of the role of the knowledge engineer. Work is underway on the Schema System II which is designed to achieve both of these goals.

Previous research with the Schema System (Draper, Brolio et al. 1989) centered around the realization that different interpretation techniques are needed to recognize different objects. The process of finding an automobile in an image is not the same as finding a road or a tree. The problem has always been how to combine multiple recognition techniques into a single, coherent system. The Schema System II approaches this problem by viewing object recognition as search through the space of knowledge states. The general idea is to use a compile-time analysis to trace all possible paths through knowledge space in order to find the most efficient routes.

This framework is being used for learning information that would otherwise have to be provided by a "knowledge engineer." The system may be able to learn the expected cost of each knowledge source (KS) and the likelihood of each possible KS result to eliminate the need for user-defined control strategies. The system may also be able to learn which object-specific combinations of evidence allow the presence of an object to be inferred, eliminating the need for user supplied confidence mapping functions. The final phase will address acquiring new object descriptions by learning new knowledge sources.

III.2. Perceptual Organization

III.2.1. Perceptual Organization of Curved Lines

Image curves often correspond to the bounding contours of objects as they appear in the image and provide important structural 3D information. In most cases, however, curves do not appear as coherent events in the image and must be reconstructed from fragments obtained from low level processes. Dolan (Dolan and Weiss 1989) is developing a system which exploits principles of perceptual organization, such as proximity and good continuation, to build multi-scale symbolic descriptions of co-curving or curvilinear image structure.

Two primitive geometric descriptors of image structure are employed (straight lines and conic splines) to describe the image structures of interest: collinearities, smooth curves, inflections, corners, and cusps. In order to manage the computational complexity inherent in the organization process, the system follows the iterative linking, grouping, and replacement paradigm developed by Boldt et al (Boldt and Weiss 1987; Weiss and Boldt 1986; Weiss, Hanson et al. 1985) The image primitives from which larger structures are built are unit tangents obtained by finding zero-crossings of the Laplacian and computing the local orientation of each edge point from the local gradient. Preliminary experimental results from this system may be found in (Dolan and Weiss 1989).

III.2.2. Organizing Surface Boundaries

The ability to find sets of points or lines which belong to a single object or define a single surface is extremely important in computer vision. For example, Williams and Hanson (Williams and Hanson 1988) show that when two or more image points are approximately equidistant in depth (which is often the case when they belong to a single object whose extent in depth is small relative to its distance to the camera) then the distance to those points (and therefore the object) can be accurately and reliably recovered. Similarly, the formidable combinatorics inherent in matching 3-D models to image data can only be controlled when there is prior evidence that several points or lines belong to a single object (Beveridge, Weiss et al. 1989b; Burns 1987; Burns and Kitchen 1988). Williams is developing a system for the perceptual organization of surface boundaries which exploits Gestalt grouping principles such as proximity, similarity, good continuity, convexity and symmetry.

Our view is that the goal of perceptual organization is not to assert surface boundaries in the presence of "noise", but rather to assert surface boundaries in the presence of other potentially occluding surfaces. The process of explaining an input set of line segments as the projections of boundaries of opaque surfaces at different depths begins by the insertion of virtual lines and possible vertices corresponding to potential organizations. For each member of this augmented line set, a set of constraints (derived from the physical properties of surfaces) on the role of that line in the final interpretation are asserted.

We envision this constructive or problem posing stage as being followed by an optimization or "problem answering" stage. The optimization problem consists of a linear or quadratic objective function subject to linear constraints. Each virtual line and vertex will either be promoted to a visible surface boundary, a hidden surface boundary or be deleted. As a natural side effect, wherever possible, the sign of occlusion will be determined. This aspect of the work, if successful, is tantamount to figure-ground segregation, and has important implications for obstacle avoidance in robotics and for the enforcement of smoothness constraints for creation of dense depth maps and optical flow fields.

III.3. 2D Model Matching

An important problem in model-driven 3D interpretation is how to use approximate knowledge of the location and orientation of the camera, models of objects in the environment, and the results of low-level vision to determine the image-to-model correspondence. The approach we have taken is to separate 2D model-to-image matching from the determination of the 3D pose parameters (see Section III.4)

Beveridge (Beveridge, Weiss et al. 1989a; Beveridge, Weiss et al. 1989b) assumes that a 2D model has been supplied with rough constraints on its image position (e.g. via an approximate 3D location in a modeled environment). This substantially reduces the search space of possible model-image line correspondences. The goal here is to determine correspondences between model and data lines such that an optimized spatial fit will produce the lowest match error. The search must be carried out across the space of possible line correspondences and this involves dealing with the complexities of grouping fragmented data, and missing or erroneous lines. The rotation and translation of the model that minimizes the error in spatial fit for a given set of line correspondences is computed via a closed-form solution. Interesting experimental results are achieved on images from our mobile robot domain.

III.4. 3D Pose refinement

Kumar (Kumar 1989) has developed an optimization technique for finding the 3D camera pose given a set of correspondences between 3D model lines and 2D image lines. The 3D pose is given by the rotation and translation matrices which map the world coordinate system to the camera coordinate system. Using the output of the system described in Section III.3, these algorithms allow updating of the mobile robot position via landmark recognition..

The approach is based on the constraints developed by Liu et al (Liu, Huang et al. 1988), but differs in two significant ways. First, rotation and translation are solved for simultaneously, which makes more effective use of the constraints and is more robust in the presence of noise. Second, the nonlinear least-squares optimization algorithm used to solve for rotation and translation is adapted from Horn (Horn 1987); Horn's technique provides much better convergence properties than does Liu et al's solution method based on Euler angles (Kumar 1989).

IV. IMAGE UNDERSTANDING ARCHITECTURE

IV.1. Image Understanding Benchmark

The second DARPA vision benchmark represents an integrated vision task across a range of typical vision processing (Weems, Riseman et al. 1988) Released in March 1988, the integrated benchmark involves processing from the sensory level, through intermediate processing of symbolic tokens, to matching of data against a set of models, and finally verification of the hypothesized model in a top-down manner. A sequential solution to the benchmark was written at UMass and verified by the University of Maryland.

The benchmark was widely distributed and in October 1988, UMass hosted a workshop in Avon, Connecticut to discuss the results. Representatives were present from the Darpa IU community, as well as from groups who implemented the benchmark on different machines. All of the architectures from the Strategic Computing program that are typically considered for vision applications were involved in the exercise, including the Warp, the Connection Machine, and the IUA (the Image Understanding Architecture). In addition, the benchmark was implemented on the Sun-3 and Sun-4, the Sequent Symmetry 81, Intel iPSC-2, Aspex ASP, and the Alliant FX-80. The results of this workshop are presented in (Weems, Riseman et al. 1989) in this proceedings.

IV.2. Status of IUA Project

The Image Understanding Architecture effort focussed on three main areas: completion of the IUA prototype with Hughes Research Labs; extensions to the IUA software simulator, and implementation of the DARPA Benchmark on the IUA simulator.

At the hardware level, progress towards completion of the IUA prototype included redesign of the feedback concentrator chip (for the associative functions in the CAAPP array), resubmission of the ICAP communication chips after a failed fabrication run, redesign and resubmission of the 64 processor CAAPP test chips (fabrication was delayed over six months due to vendor problems), and the design of an I/O subsystem (by Hughes). After several alternatives were considered, a decision was made to use commercially available components to construct the system controller. The first prototype board is expected to be delivered to UMass from the Hughes Research Labs in August, 1989.

At the software level, major changes and extension were made to the IUA simulator. It was ported to the Sun-3 system, integrated with the Sun windowing environment, and its interactive graphical capabilities were substantially enhanced. A full ICAP simulator was developed and merged with the CAAPP simulator to provide two-level simulation capabilities. This was then ported to the Sequent Symmetry multiprocessor and enlarged to a full sized (512x512) CAAPP and a 4096 processor ICAP. Unfortunately, memory limitations on the Sequent limited useful simulations to a 16 processor ICAP configuration. In addition, numerous improvements were made to the simulator at the basic code level and a number of vision algorithms were coded and tested.

Finally, a major effort to program the benchmark on the IUA simulator was successfully completed and reported as part of the Avon conference discussed earlier. The IUA implementation of the benchmark ran in approximately 80 msecs on the simulator.

V. REFERENCES

Anandan, P. (1989). "A Computational Framework and an Algorithm for the Measurement of Visual Motion," IJCV, Vol. 2(3), pp. 283-310.

Balasubramanyam, P. (1989). "Early Identification of Occlusion in Stereo-Motion Image Sequences," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Beveridge, J. R., R. Weiss and E. Riseman. (1989a). "Matching and Fitting Sets of 2D Line Segments to Broken and Skewed Data," Dept. of Computer and Information Science, University of Massachusetts (Amherst), TR In preparation.

Beveridge, J. R., R. Weiss and E. Riseman. (1989b). "Optimizing Two-Dimensional Model Matching," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Boldt, M. and R. Weiss. (1987). "Token-Based Extraction of Straight Lines," Dept. of Computer and Information Science, University of Massachusetts (Amherst), TR 87-104.

Burns, J. B. (1987). "Recognition in 2D Images of 3D Objects from Large Model Bases Using Prediction Hierarchies," Proc. of International Joint Conference on Artificial Intelligence, Milan, pp. 763-766.

Burns, J. B. and L. J. Kitchen. (1988). "Rapid Object Recognition from a Large Model Base using Prediction Hierarchies," Proc. of DARPA Image Understanding Workshop, Cambridge, MA, pp. 711-719.

Connolly, C. and R. Weiss. (1989). "Geometer: A System for Modelling and Algebraic Manipulation," Proc.

of DARPA Image Understanding Workshop, Palo Alto, CA.

Dolan, J. and R. Weiss. (1989). "Perceptual Grouping of Curved Lines," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Draper, B., J. Brolio, R. Collins, A. Hanson and E. Riseman. (1989). "The Schema System," IJCV, Vol. 2(3), pp. 209-250.

Dutta, R., R. Manmatha, L. Williams and E. Riseman. (1989a). "A Data Set for Quantitative Motion Analysis," Proc. of Conference on Computer Vision and Pattern Recognition, San Diego, CA.

Dutta, R., R. Manmatha, L. Williams and E. Riseman. (1989b). "A Data Set for Quantitative Motion Analysis," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Fennema, C., A. Hanson and E. Riseman. (1989). "Toward Autonomous Vehicle Navigation," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Horn, B. K. P. (1987). "Closed-form solution of absolute orientation using unit quarternions," J. Opt. Soc. A., Vol. 4, pp. 629-642.

Kumar, R. (1989). "Determination of Camera Location and Orientation," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Liu, Y., T. Huang and O. Faugeras. (1988). "Determination of Camera Location From 2D and 3D Line and Point Correspondences," Proc. of CVPR, Ann Arbor, MI, pp. 82-88.

Pocock, G. (1989). *The Design and Analysis of a Real-Time Computer System for Robotics.* forthcoming Ph.D. Thesis, Dept. of Computer and Information Science, University of Massachusetts.

Sawhney, H. and J. Oliensis. (1989). "Description and Interpretation of Rotational Motion from Image Trajectories," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Snyder, M. A. (1989). "On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Weems, C., E. Riseman, A. Hanson and A. Rosenfeld. (1989). "A Report on the Results of the DARPA Integrated Image Understanding Benchmark Exercise," Proc. of DARPA Image Understanding Workshop, Palo Alto, CA.

Weems, C., R. Riseman, A. Hanson and A. Rosenfeld. (1988). "An Integrated Image Understanding Benchmark: Recognition of a 2 1/2 D "Mobile"," Proc. of DARPA Image Understanding Workshop, Cambridge, MA, pp. 111-126.

Weiss, R. and M. Boldt. (1986). "Geometric Grouping Applied to Straight Lines," Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, FL, pp. 489-495.

Weiss, R., A. Hanson and E. Riseman. (1985). "Geometric Grouping of Straight Lines," Proc. of DARPA Image Understanding Workshop, Miami Beach, FL.

Williams, L. R. and A. R. Hanson. (1988). "Translating Optical Flow into Token Matches and Depth from Looming," Proc. of 2nd International Conference on Computer Vision, Tarpon Springs, FL, pp. 441-448.