To appear: First World Congress on Intelligent Manufacturing"
Mayaguez, Puerto Rico, February 13-17, 1995.

Constraint-Based Visual Robotic Systems

James J. Little, David G. Lowe, Alan K. Mackworth, Dinesh K. Pai, Robert J. Woodham
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada V6T 1Z4

Abstract

The project "Constraint-Based Visual Robotic Systems" is one of six in the theme "Integrated Systems for Dynamic Environments" within the Institute for Robotics and Intelligent Systems, a Canadian Network of Centres of Excellence. The project focuses on building integrated robotic systems with visual sensing. Four main issues have been identified as key scientific bottlenecks: learning, integration of vision modules, model acquisition and integrating perception with action. The central problem of our theme, building integrated systems for dynamic environments, is addressed by delivering formal and practical tools for using vision and constraint-based control in mobile robotic platforms. Our hypothesis is that a constraint-based approach will provide a simple, effective and unitary architectural framework for designing and building such systems.

One goal is to provide vision services to robots across a wide variety of platforms, including eye-heads and mobile robots. A critical aspect of these systems is integrating multiple views, from varying modules, over time; common processes and methods will form an "Integral" interface between robot and vision services. Within the project, a 3-D object recognition system is being developed that learns an efficient indexing function from image features to object models, using a sample of images of each object. This approach is being demonstrated for robot manipulation tasks. The Calibrated Imaging Facility (CIF) continues to develop as a testbed facility for IRIS researchers and partner organizations. In particular, near real-time implementations of photometric stereo and multiple light source optical flow support experiments to test the accuracy of shape and motion determination in a variety of geometric and radiometric configurations.

The Dynamite system facilitates experiments in competitive and cooperative visual robotic behaviour. The Constraint Net framework has been used to describe path planning and tracking for that application. In addition, we have developed a novel spherically-symmetric mobile platform that can operate in cluttered environments and rugged terrain because of its high degree-of-freedom design and ability to recover from toppling.

1 Introduction

This project is designed to attack the research problem of building constraint-based visual robotic systems. A recent NSF Workshop [15] addressed, in part, the current state of the art in computational vision. The summary report from that workshop targeted four main scientific issues underlying seven proposed "Grand Challenge Applications". The issues are i) learning, ii) integration of vision modules, iii) model acquisition and iv) perception and action. We are targeting each of these since they are the key scientific bottlenecks. Learning and vision have hitherto developed as unrelated research areas. There is a large gap between theoretical and practical use of integration of vision modules, although it is known that the issue is critical; particular attention must be paid to cooperative integration of physics-based vision. Model acquisition must be automated for successful industrial applications. Finally, perception and action in dynamic environments is now seen as central. A recent paper on the Toronto IRIS Stereo Head [11] gives an excellent account of work on active vision in terms of a survey of some recent experimental eye-head systems, such as we use. Finally, a recent overview of a research program for situated perception [10] emphasizes the need for a single computational model for building embedded perceptual robotic systems. The formal theory of a situated agent models the robot and its environment as coupled dynamical systems. This is an approach to the central problem of our theme: building integrated systems for dynamic environments.

The major technological challenge is to provide a uniform architectural framework for building integrated constraint-based visual robotic systems. Our goals are: to provide a specific framework for system integration for visual robotic systems; to advance the use of mobile visual platforms in changing environments so that they can be tailored to specific environments, including both smooth and rough terrain; to develop systems for learning visual models and for extracting dynamic surface shape; and to elaborate the constraint-based design paradigm and develop tools for specifying, simulating, implementing and verifying robotic systems.

2 Research Program

The overall objective is to develop new techniques for responsive vision on mobile platforms in dynamic environments. We are developing practical tools for building constraint-based visual robotic systems. Vision systems for monitoring and controlling the robot plant and the environment are being built for various testbed platforms, including visual telerobotic control on machines currently used in the resource and construction industries. The Constraint Net (CN) language is being used as a tool for building visual robotic systems. 3-D vision is being enhanced with methods for learning new models. Real-time systems integration software for all platforms is being developed. New techniques for deformable surface visual interpretation are under development. A symmetric visual platform is being specialized for visual inspection and manipulation tasks in cluttered environments and rugged terrain. We are integrating our new responsive vision systems with the legged mobile platform, and demonstrating the interaction of these integrated systems with real, dynamic environments. Legged platforms are well suited for robust locomotion on rough terrain [1]. We are developing complete integrated constraint-based visual robotic systems in several realistic dynamic environments.

The following examples demonstrate the ongoing research in our group and highlight the research directions as well as the facilities we are using.

2.1 3-D Model Acquisition and Use

Our system for 3-D object recognition is based on recent developments in the area of computer learning. This system is more efficient and robust at performing object recognition than current approaches, because the mapping from image features to interpretations is learned from a sample of images of the object. The learning approach offers many advantages, including a more accurate representation of appearance, ability to use a wide range of features, and a large reduction in search. This vision system will be demonstrated in conjunction with a robot for grasping and manipulating objects. Another research objective is real-time identification and tracking of known target shapes. This is a task with near-term industrial applications, as it will allow for low-cost systems that can perform in real time.

The research on 3-D object recognition is aimed at overcoming a key shortcoming of current methods, which is their difficulty in indexing from image features to matching object models. In recent years, considerable progress has been made on the problems of fitting models to images and verifying an interpretation once an initial correspondence has been proposed [9, 8]. However, current methods for model-based recognition are limited to using small numbers of well-defined objects due to the computational costs of the required search process. This proposed research will develop techniques to learn indexing functions from image features to model correspondences using a set of training examples derived from previous instances in which each object has been recognized. An additional form of learning will build a 3-D model representation directly from the features extracted from a set of images [13].

This approach builds on that proposed by [4], in which they propose the use of a Radial Basis Function (RBF) learning method to perform 3-D object recognition. Their implementation uses feature vectors derived from a simple wire-frame synthetic model that assumes pre-segmented and labelled features. In order to extend this method to work with a wide range of real images, solutions must be found to the issues of segmentation, labelling and feature vector construction. In addition, this form of learning is unlikely to solve the full object recognition problem, as it does not attempt to form a consistent integration of object features that are not in the original feature vector. Therefore, we are proposing this technique as a method for indexing to greatly reduce search, while relying on model fitting to provide final reliability of interpretation. The choice of a learning method is somewhat incidental, and we expect to use our own weighted nearest-neighbour kernel algorithms rather than the RBF method in later experiments.

2.2 Vision Architectures

We also are investigating the architecture of responsive vision systems by assembling hardware and software that operate robots in dynamic environments. We are identifying suitable arrangements under a variety of operating constraints, from simple autonomous systems to complex situated agents requiring substantial computational power. This involves more than just system configuration; there are many open issues in determining the proper interactions between the vision component of a robot and the reasoning and action components. World modeling requirements and computing system facilities are combined into a vision-based operating system. These concepts are demonstrated in tracking systems, navigating mobile robots and other robotic systems.

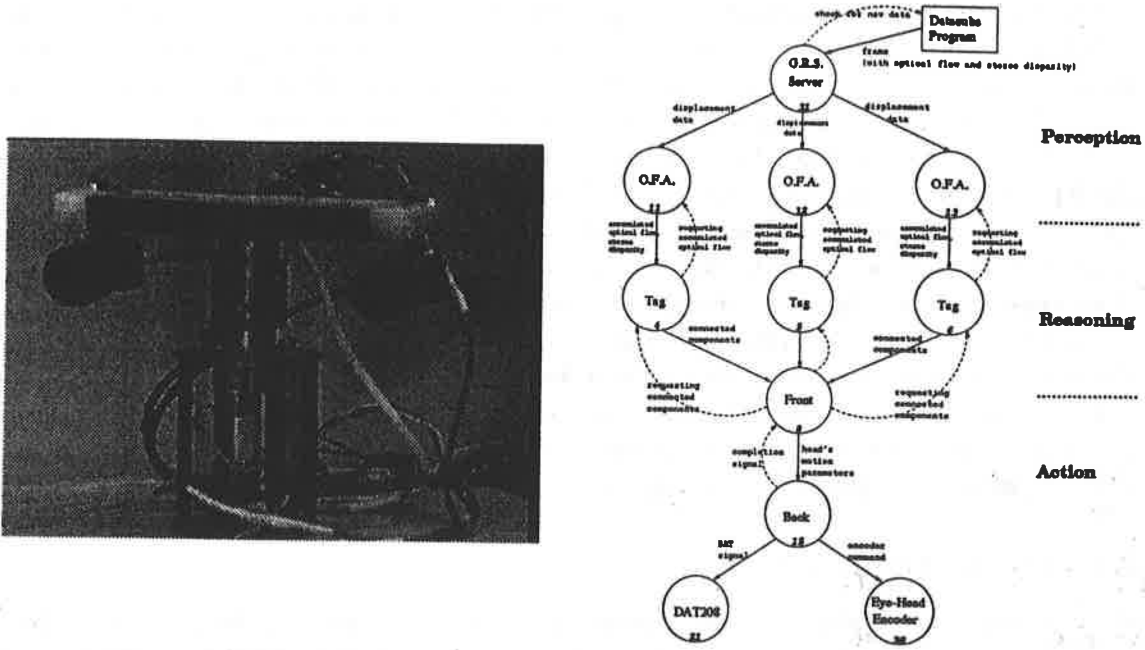
The "Integral" operating system for perception offers familiar interfaces: buffering, caching, hierarchy and interrupts, that permit the reasoning and action components to proceed as if perception were seamless and continuous, masking the discrete and bursty nature of its output. The motion tracker [7] demonstrates how to build a tracker that has several processes running at different rates in near real-time. The tracker follows a moving object, with no knowledge of its target, based on dense optical flow input.

The UBC Vision Engine [6] exemplifies distributed, heterogeneous systems that have multi-rate processes. The Engine is general-purpose and supports all levels of vision. The system consists of multiple architectures: pipelined (a Datacube MaxVideo200) image processor and a MIMD multicomputer (20 T800 2MB Transputers, connected via a crossbar), connected by a bidirectional video-rate interface. The Transputer system controls an eye/head platform for vergence, pan and tilt, as shown in Figure 1(a).

Our tracker continuously monitors its environment, overlapping interpretation and sensing. The tracker has multiple stages: correlation motion on the Datacube, optical flow accumulation, target selection and eye-head control. The first two stages comprise the Perception component, distributed over two different machines. The next stage is the Reasoning component, which analyses the Perception output to determine the target. The final stage translates the target location in image coordinates into controls for the eye-head. Figure 1 (b) depicts the data flow in the system.

A *smart buffer* process [7] is responsible solely for receiving the motion data stream. The smart buffer contains an *active monitor* that waits for data to arrive, and keeps track of the data until it has been properly stored for later retrieval by the reasoning system. The monitor is "active" since it asynchronously accepts motion data whenever it is available.

The Reasoning stage finds connected components in 300 to 500ms; this is the principal delay in the system, necessitating the multi-rate interface. The Reasoning stage sees the smart buffer as a data object



(a) (b)

Figure 1: (a) The UBC Eye/Head. (b) Software Components and Data Flow.

that contains the optical flow. The active monitor sees the smart buffer as its client. The full cycle, optical flow computation and accumulation, flow cancellation and target finding requires 800ms, mostly due to component labeling.

The system has two clients: the target process needs accumulated motion data, and a filtering process needs current depth data from stereo. Filtering selects a target in a specific range of depths. To connect a third client, for example, one to use stereo data cues or motion cues to detect looming, the system need only arrange that the stereo and motion buffers provide read-only access to a daemon that detects looming.

2.3 Real-time Tracking and Control

The Dynamo testbed is a collection of independently controlled mobile robot vehicles that play soccer [2]. Figure 2 shows the soccer field and robot soccer players (remotely controlled 1/24th scale cars). The system demonstrates offboard vision processing and distributed processing. The vision component was originally prototyped on the Datacube MaxVideo200. Currently the system is realized as simple custom hardware to process RGB signals, followed by run-length encoding and centroid calculation on Transputers. A single off-board RGB camera sensor communicates its signals to the centralized sensor processor. The sensor processor provides positional information to the control processes for each competing soccer player at 60Hz (once per image field) with a lag of at most 5 ms after the end of field. The structure of the full system is shown in Figure 3.

The Dynamo system has been used to explore novel reactive strategies for control [14] as well as to provide the testbed for ideas on control, specification, and reasoning about real-time systems [18]. The Constraint Net formalism describes a robotic system as a coupled pair of dynamical systems: the robot and its environment. We plan to use CN to model the perceptual, coordination and control

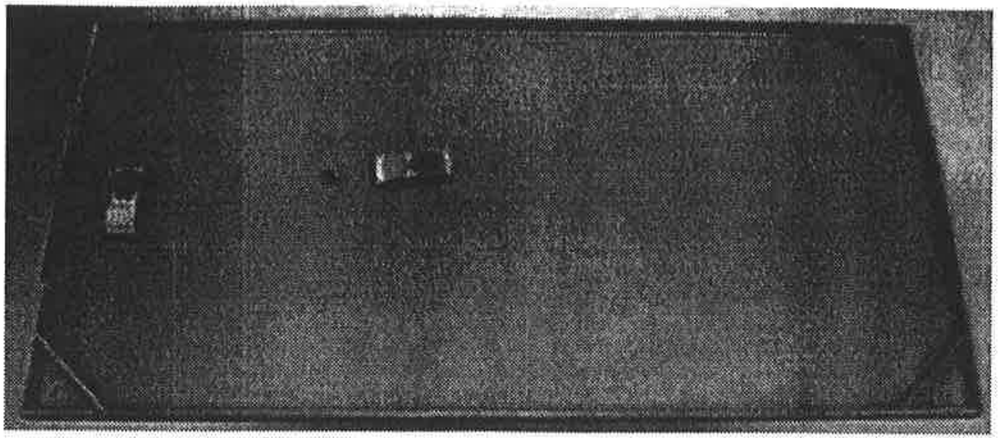


Figure 2: Dynamites Playing Soccer

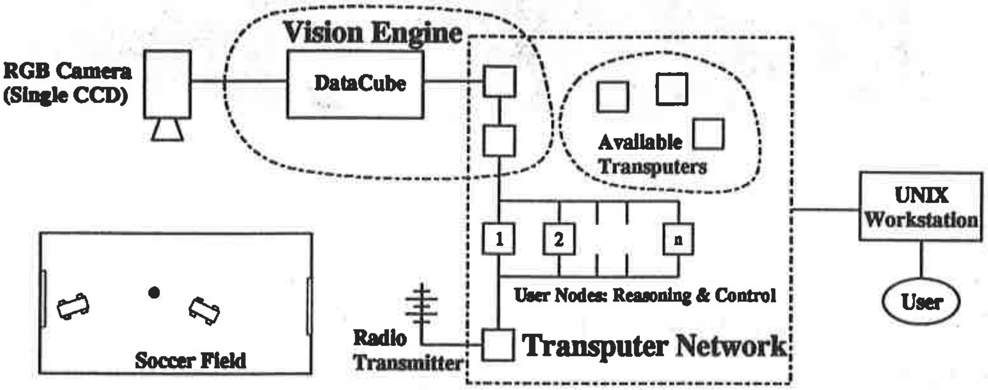


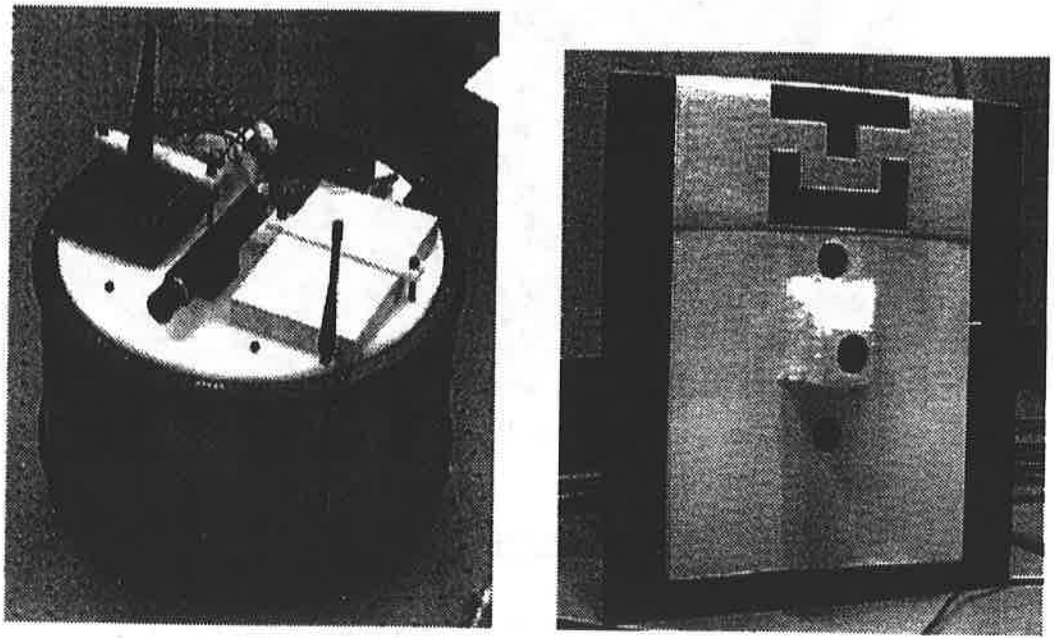
Figure 3: Dynamo Architecture

software required for our Dynamo multiple mobile robot testbed using both offboard [2] and onboard visual systems. The tasks include navigation in rugged terrains and multi-player real-time games like soccer [10] and hockey that serve not as ultimate applications but crucial experimental development environments. Work on verifiable control for robots using embedded constraint solvers is also being pursued.

2.4 Real-time Localization

ROLL (Real-time Onboard Localization with Landmarks) identifies its position in real-time, using passive visual localization of a single landmark [3]. ROLL is implemented on a Real-World Interface platform (see Figure 4(a)), named Spinoza, with offboard video processing sent via FM narrowcast (antenna at left). Real-time processing occurs on a TI C40 system. The landmark (Figure 4(b)) contains a visual bar code as well as a structure that contains disks whose centroids can be rapidly and accurately computed. Their relative positions plus the known dimensions of the target allow ROLL to compute its own position and orientation at 10Hz. ROLL controls the RWI base via a spread-spectrum modem (at right in Figure 4(a)).

The system is distributed over the C40s and a Sparcstation, which communicates with the RWI. Robust localization can use the platform's odometry, but there is significant latency between the odometry and the vision server on the C40s. Prediction by Kalman filtering must compensate for the varying temporal shift between the measurements.



(a) (b)

Figure 4: (a) Spinoza. (b) Landmark.

2.5 Spherically Symmetric, High Degree of Freedom Robots

We have developed a new class of spherically symmetric, high degree of freedom robots called "platonic beasts." A robot in this family is kinematically equivalent to a symmetric polyhedron, one of the five Platonic solids, with identical multi-purpose limbs attached to its vertices. The symmetry and regularity of the design have several advantages including robustness to toppling, novel gaits such as the *rolling* gait (shown in Figure 5), and fault tolerance. For instance, the robot has no preferred orientation and hence can never be incapacitated by falling on its 'back' after losing a foothold.

We have designed, simulated, and constructed a prototype platonic beast robot in our lab [12]. The robot has four limbs, each with three degrees of freedom, and is controlled by a network of four embedded 32-bit microcontrollers.

We are currently designing a low-cost spherically symmetric, six-limbed robot which will serve as a crucial testbed for our integrated vision systems, combining visual terrain perception, proximity detection, and locomotion. The spherically symmetric visual platform is based on our existing prototype and modular design [12].

The platform can be equipped with a variable number of low-resolution CCD camera modules which are integrated into the limbs of the robot. A camera module will be used multi-modally, either simply as a remote, active, targetable camera relaying images to a host over a wireless communication link, or as part of an onboard vision system for fast proximity detection and slower terrain mapping. The robot will integrate terrain perception by combining visual proximity detection and terrain mapping with contact and orientation sensors. The platform can also use its limbs in many modes because of its symmetry and redundancy; it can use its limbs both for locomotion and manipulation (Figure 6). The system allows us an opportunity to develop new locomotion algorithms that explicitly exploit perceived terrain constraints.

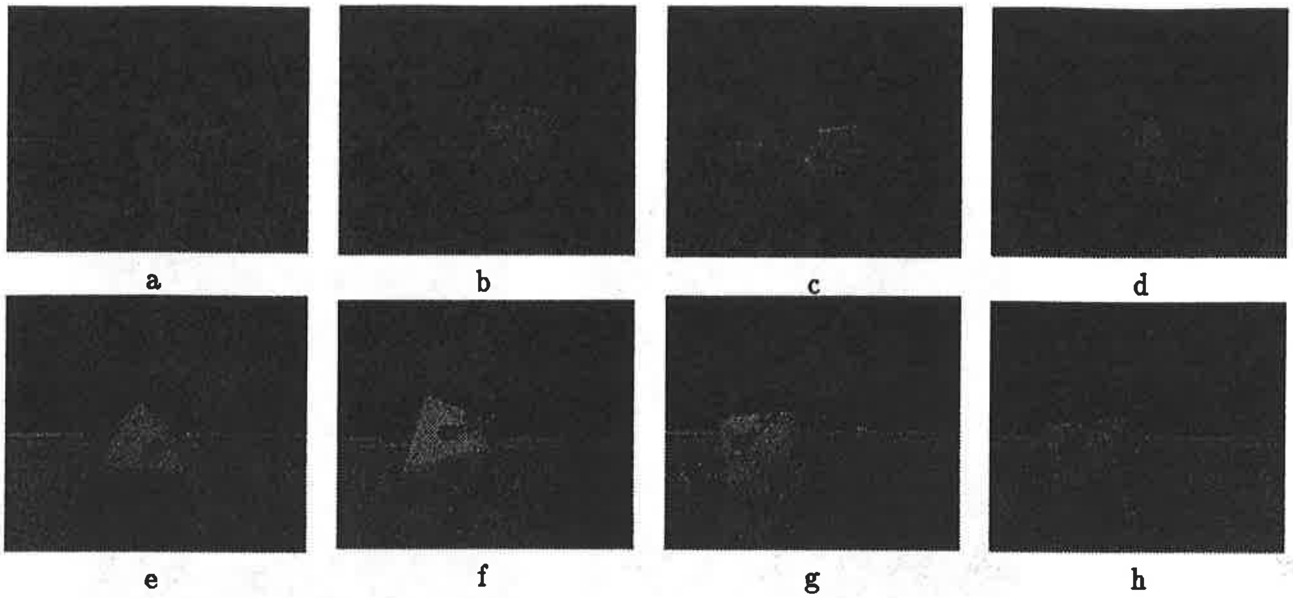


Figure 5: The symmetric platform undergoing a canonical tumble

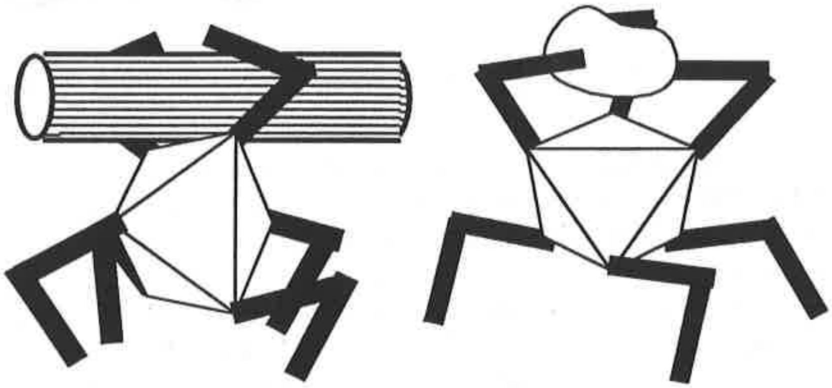
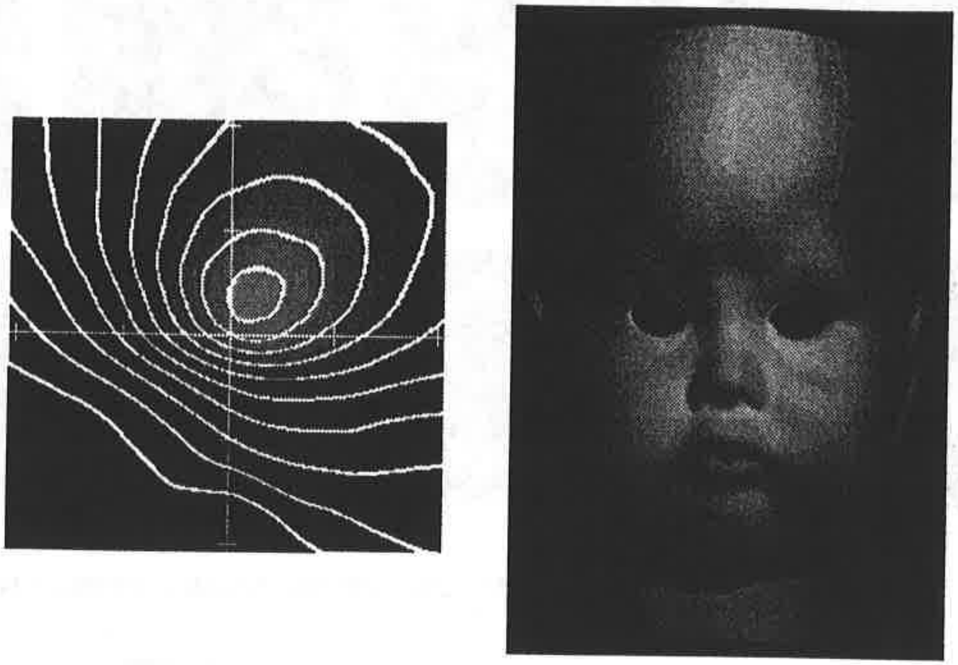


Figure 6: A 6-beast can be used as a 4-limbed crawling machine with a 2 fingered hand or as a 3-fingered mobile hand.

2.6 Physics-based Shape and Motion Estimation

Another specific objective is to integrate "physics-based" shape and motion estimation for robot vision tasks including 3-D model acquisition, object recognition, localization and inspection. The more general objective is to develop our Calibrated Imaging Facility (CIF) as a widely used and generally useful scientific testbed to study the role that 3-D scene parameters, including object shape, object material, scene illumination, viewpoint and motion play in the success (and failure) of computer vision techniques. The CIF is based on an optical bench with mounting hardware for controlled positioning and motion of cameras, light sources and test objects; it is integrated with other ISDE-6 vision facilities, specifically with the DataCube image processing hardware, transputers and C40s.

Geometric constraints determine where in an image a point in the scene will appear and radiometric constraints determine its brightness and colour. "Physics-based vision" takes seriously radiometric, as well as geometric, constraints; it has led to a variety of new methods for determining 3-D scene parameters from 2-D images including shape-from-shading, photometric stereo and optical flow. Using spectral



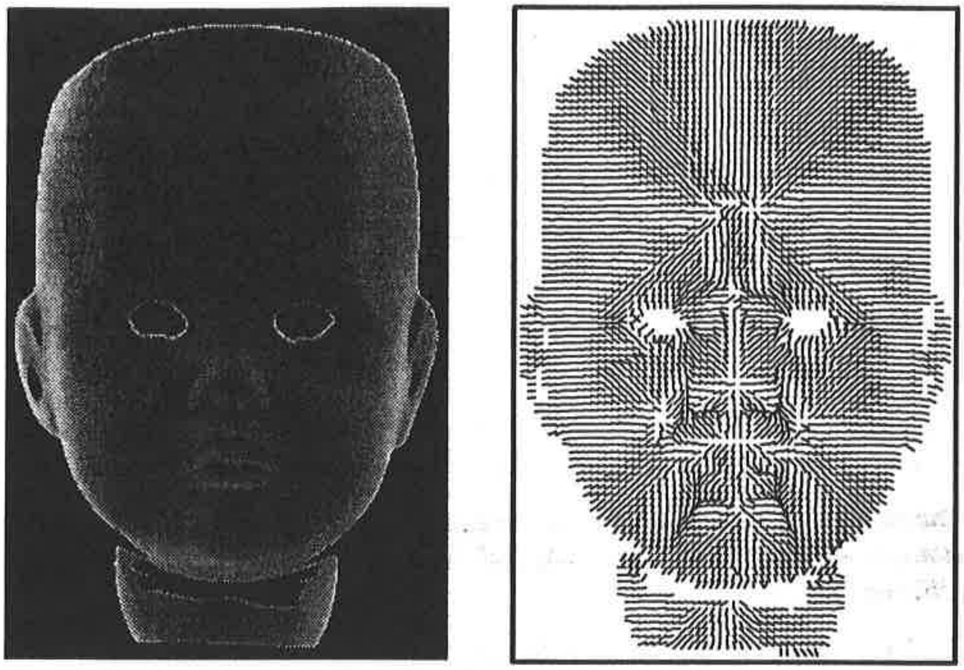
(a) (b)

Figure 7: (a) Empirical reflectance map. (b) Image of pottery head.

multiplexing, the CIF allows one to realize the additional constraint provided by viewing a scene simultaneously under different conditions of illumination. Recently, near real-time performance on gradient estimation from photometric stereo has been demonstrated using 512x480 24-bit RGB images (15 Hz on DataCube hardware and 10 Hz on a 6-node C40 system). Figures 7(a) and (b) show an example reflectance map (relation between surface orientation and brightness) and input image for gradient estimation. Figures 8 (a) and (b) show the gradient magnitude and direction for this image. This will be extended, in the C40 environment, to include near real-time estimation of curvature [17] and optical flow [16]. Curvature data is especially valuable because it is viewpoint invariant. Comparison involves shape estimation for tasks including recognition, localization and inspection and the ability to distinguish optical flow (a radiometric concept) from motion (a geometric concept). A prototype 3-D (rigid body) model acquisition system using photometric stereo and range data as input uses orientation-based representations of shape [5] including the radial function, the support function and the first and second curvature functions. Recently developed viewpoint invariant surface reconstruction techniques will be used to piece together visible surface data into full 3-D models. The system will have the unique capability of handling surfaces whose shape deforms as well as moves.

3 Summary

Our project focuses on building integrated robotic systems with visual sensing. We address the central problem of our theme, integrated systems for dynamic environments, by delivering formal and practical tools for using vision and constraint-based control in mobile robots. Our hypothesis is that a constraint-based approach will provide a simple, effective and unitary architectural framework for designing and building such systems. Our robotic systems will establish new standards for speed of interaction and density of data flow in real-time robotics. They will act as prototypes for the design of commercial systems by demonstrating how complex perceptual computing serves real-time robotic activities.



(a) (b)

Figure 8: (a) Gradient magnitude. (b) Gradient direction.

4 Acknowledgments

We would like to thank Rod Barman, Stewart Kingdon, Jeff Brewster, Jiping Lu, Johnny Kam, Art Pope, Scott Ralph, Michael Sahota, and Ying Zhang for all their work on the systems described here. This research is funded by the Natural Sciences and Engineering Research Council, the Institute of Robotics and Intelligent Systems, a Network of Centres of Excellence managed by Precarn, Inc.

Dinesh Pai is a Fellow of the B.C. Advanced Systems Institute. David Lowe is a Scholar and Alan Mackworth is a Fellow of the Canadian Institute for Advanced Research.

References

[1] J. Bares, M. Hebert, T. Kanade, E. Krotkov, T. Mitchell, R. Simmons, and W. Whittaker. Ambler — an autonomous rover for planetary exploration. *Computer*, 22(6):18–26, 1989.

[2] R. Barman, S. Kingdon, J.J. Little, A.K. Mackworth, D.K. Pai, M. Sahota, H. Wilkinson, and Y. Zhang. DYNAMO: real-time experiments with multiple mobile robots. In *Intelligent Vehicles Symposium*, Tokyo, July 1993.

[3] J. Brewster. Real-time onboard localization with landmarks. Master's thesis, The University of British Columbia, Vancouver, BC, 1994.

[4] S. Edelman and T. Poggio. Bringing the grandmother back into the picture: a memory-based view of object recognition. AI-Memo-1181, MIT-AI, April 1990.

[5] Y. Li and R. J. Woodham. Orientation-based representations of 3-D shape. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994*, pages 182–187, Seattle, WA, 1994.

[6] J. J. Little, R. A. Barman, S. J. Kingdon, and J. Lu. Computational architectures for responsive vision: the vision engine. In *Proceedings of CAMP-91, Computer Architectures for Machine Perception*, pages 233–240, December 1991.

[7] James J. Little and Johnny Kam. A smart buffer for tracking using motion data. In *Proc. Workshop on Computer Architectures for Machine Perception*, pages 257–266, December 1993.

[8] D. G. Lowe. Fitting Parameterized Three-Dimensional Models to Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[9] David G. Lowe. The viewpoint consistency constraint. *International Journal of Computer Vision*, 1(1):57–72, 1987.

[10] A. K. Mackworth. On seeing robots. In A. Basu and X. Li, editors, *Computer Vision: Systems, Theory, and Applications*, pages 1–13. World Scientific Press, Singapore, 1993.

[11] E. Milios, H. Jenkin, and J. Tsotsos. Design and performance of TRISH, a binocular robot head with torsional eye movements. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):51–68, 1993.

[12] D. K. Pai, R.B. Barman, and S. Ralph. Platonic beasts: a new family of multilimbed robots. In *Proc. IEEE Conf. on Robotics and Automation, 1994*, pages 1019–1025, 1994.

[13] Arthur Pope and David G. Lowe. Learning object recognition models from images. In *Proc. 4th International Conference on Computer Vision*, pages 296–301, May 1993.

[14] Michael K. Sahota. Reactive deliberation: An architecture for real-time intelligent control in dynamic environments. In *Proc. 12th National Conference on Artificial Intelligence*, pages 1303–1308, 1994.

[15] B. W. Wah et al. Report on workshop on high performance computing and communications for grand challenge applications: Computer vision, speech and natural language processing, and artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering*, 5(1):138–154, feb 1993.

[16] R. J. Woodham. Multiple light source optical flow. In *Proc. 3rd International Conference on Computer Vision*, pages 42–46, Osaka, Japan, 1990.

[17] R. J. Woodham. Gradient and curvature from the photometric stereo method, including local confidence estimation. *Journal of the Optical Society of America, A*, 11:3050–3068, 1994.

[18] Y. Zhang and A. K. Mackworth. Modeling behavioral dynamics in discrete robotic systems with logical concurrent objects. In S. G. Tzafestas and J. C. Gentina, editors, *Robotics and Flexible Manufacturing Systems*, pages 187–196. Elsevier Science Publishers B.V., 1992.