EDITORIAL

HENRIK I. CHRISTENSEN
Laboratory of Image Analysis, Aalborg University
DK-9220 Aalborg, Denmark

KEVIN W. BOWYER
College of Engineering, University of South Florida
Tampa, FL 33620-5399, USA

and

HORST BUNKE
Institut für Informatik und angewandte Mathematik
Universität Bern, CH-3012 Bern, Switzerland

1. INTRODUCTION

This book is a reprint of a special issue of the *International Journal of Pattern Recognition and Artificial Intelligence.* It consists of papers that are extensions and revisions of work that has been presented at the SPIE Conference on *Applications of Artificial Intelligence X: Machine Vision and Robotics*, held in Orlando, Florida, 22–24 April 1992. All papers have been separately reviewed for this special issue. We thank the reviewers and authors for keeping to a very tight schedule.

Machine vision has now been an active area of research for more than 35 years and it is gradually being introduced for real world application. Most of the applications which were developed and built in the seventies and the eighties were based on dedicated methods combined with use of specific application knowledge. I.e. in a typical application special sensory equipment such as laser range cameras was used often in combination with well controlled lighting (i.e. artificial or structured lighting). For description of geometric information or motion it was often assumed that the environment was constrained to a limited number of well defined objects which were modeled *a priori* and that most of their characteristics were utilized in the image processing and analysis.

Little insight into the general problem of image based scene description and interpretation was gained from these applications, as the applications to a large extent were based directly on image derived features. The approaches generally lacked robustness and often became ill-posed for even a slight variation in the conditions of the original application. Very little in the way of general "high level" algorithms came out of this. The reconstruction approach set forward by Marr in his now famous book *Vision*[1] thus received little attention in terms of use in industry.

Recent research has, however, indicated that some of the robustness and the ill-posed problem may be eliminated if the algorithms are applied in the context of a controllable sensor system. The explicit control of the sensory system to improve robustness and eliminate ill-posed conditions is often termed "active vision".[2] In addition it has been suggested that the general machine vision problem of providing a full 3-D symbolic description of the environment without any prior knowledge is much too hard to be solved at this time and that robust solutions may be found provided that task specific knowledge is utilized in the design and processing for a specific application. Such an approach to machine vision is termed "purposive vision". The aim of purposive vision is not to default to the strategy for construction of application, which was adopted in the seventies and the eighties, but rather to complement "general" machine vision techniques with domain specific information to facilitate control of the entire system so as to provide the needed robustness. Control is thus a significant issue in purposive vision.

A significant application area for machine vision is in robotics. Much of the work in robotics has been based on use of sensory modalities such as ultrasonic sonars, as it has been difficult to obtain sufficiently good depth data using "shape from X" techniques. The introduction of *a priori* information may, however, change this situation. For use in well-known scenarios, it is possible to construct a model of the environment and subsequently compare sensor readings with predictions obtained from the model environment. Progress in areas such as CAD modeling has implied that today it is possible to integrate CAD systems into the control of robots. The introduction of such models implies at the same time that it is possible to exploit machine vision methods, as the needed *a priori* information may be extracted from the CAD model.

To ensure that the systems constructed may be used not only for one specific application but rather for a variety of applications the trend is towards use of layered control. In layered control the hardware of the robot is interfaced to the rest of the system through device level software. This software transforms robot-specific commands and feedback into a standard representation which may be shared by several different platforms. It thus becomes simple to change robots without a need for a complete redesign of all the software. Above the device level is a set of control layers which handle path planning, control and the associated perception. In robot control, at least for mobile robots, there is a trend towards use of a set of different layers. Each layer is responsible for a specific task for the robot. For example, one layer may be responsible for "survival" and be responsible for making sure that the robot does not bump into objects in the environment and that it moves away if it is on a collision course with another object in the environment. Another layer might be responsible for construction of a map of the environment to facilitate navigation or localization of target objects. The use of different layers for different tasks is different from the approach which traditionally has been used in robotics, where control is integrated in a "perceive-plan-control" cycle. The two approaches to control of a robot are illustrated in Fig. 1.

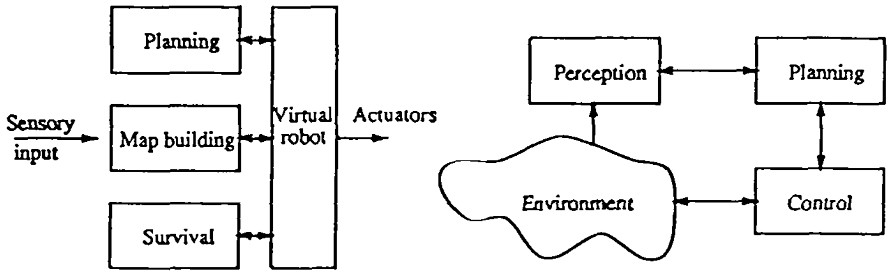

Fig. 1. Two different architectures for control of robots. The figure to the left is a layered control architecture where different layers are responsible for different tasks, while all the tasks are integrated in a single control loop in the architecture to the right.

2. ACTIVE VISION

The research topic *Active Vision* is related to controlled change of the parameters of the sensory system to facilitate improved robustness. Active vision research was initially proposed by R. Bajcsy at the University of Pennsylvania[3] based on efficiency considerations. Initial results using the approach were later presented by E. Krotkow in the context of stereo reconstruction, where images recorded at two different locations are used for calculation of distance to scene objects.[4] In this work a set of stereo images are corresponded and whenever ambiguous matches are obtained a focusing procedure is used for verification of correct depth. In the system presented by Krotkow, facilities for change of the angle between the cameras and computer control of the camera lenses are utilized to determine depth. Later research has demonstrated that several other algorithms also may benefit from use of controlled change of sensory parameters.[2] During the late eighties mainly groups in the USA were working in the area.[5–7] Recently the research has, however, gained interest also in Europe and Japan. Presently a large number of groups are doing either theoretical or experimental research in the area.

To study active vision at the experimental level it is necessary to have actuators which facilitate change of both intrinsic and extrinsic parameters for the sensory system. The intrinsic parameters (focus, focal length, and aperture) may be controlled through use of motorized lenses, which are well-known from the pocket cameras you can buy in the local photoshop. For the control of extrinsic parameters several options are available. The control may be divided according to the degrees of freedom available in the human visual system. Humans may control the orientation of the eyes both horizontally (eye vergence) and vertically (eye panning). In addition humans may also rotate each eye around the optical axis (cyclotorsion). These three degrees of freedom define configurations for the eye, but in addition both neck motion (pan and tilt) and body motion (change of position in 3-D space) may be used in the selection of view point. To facilitate experimental research with emulation of some of the flexibility available in biological vision it is necessary to build robot heads which have similar degrees of freedom.

Today only a single camera head is available commercially[8] and this head was constructed explicitly for outdoor surveillance tasks and is thus not very good as a laboratory tool to be used for experiments. The construction of a robot camera head requires skills in both mechanical and electronic engineering. The interfacing between the two areas must also be covered to ensure that a reasonable design is provided. This intermediate area, often termed mechatronics, is traditionally not well covered by the groups doing computer vision research. There is thus a need for communication of the experience gained by the groups which have undertaken the task of building their own robot camera head. It may be expected that lessons may be learned from this experience and that ideas may be adopted for those interested in building their own robot camera heads. The set of papers presented in this issue provides results for a set of both *old* and *new* camera heads.

3. MODEL BASED ROBOT NAVIGATION

Most of the robot systems used in the industry today are programmed specifically for the task they perform and much of the control of the robot is performed in an open loop, which implies that the robot does not utilize feedback information to adapt its actions to changes in the environment.

Introduction of robots may be expected at a larger number of places in industry, but to allow this it is necessary that the adaptation of the robot for a specific task become simpler. Today each new task requires extensive re-programming of the robot, a task which is often performed manually. It may be expected that future factories to a large extent will produce items which are customized to the needs of individual customers. To allow this, more general control tools are needed. There is thus a tremendous potential for application of robotics but the associated adaptation to new tasks must be solved in order for this potential to be realized.

To increase the flexibility of robots it is desirable to be able to build systems which have facilities for exploiting sensory feedback to allow adaptive interaction with the environment. Preferably robot systems should be built with a "generic" set of facilities for planning and control which utilizes a model of the environment. The model of the environment may either be acquired off-line or it may be built on-line based on sensory information. In a compromise, a coarse model may be provided *a priori* and the model may then be elaborated on-line at the locations where interaction is requested.

Recent research has focussed on how robots may exploit information available in CAD models. The CAD models are often available *a priori*, as they are used in the design process. Through use of such models the robot control system may perform planning of missions. This planning can be based on general techniques (such as graph-based search) and they will thus be applicable even if the CAD model is later changed for use in another environment. In addition the CAD model may also be used for prediction of what data the sensors will produce and the interpretation of sensory information may consequently be simplified or made more robust.

In combination with model based robot navigation, several groups have also implemented a device level controller which hides the hardware details of the robot. The result is a virtual robot that may be commanded by "generic" commands. The use of such virtual robots allows a simple change of robot, but in addition a robot may also be simulated through a simple software module. Such simulated robots may be used for operator training and visualisation of how a process will be carried out before it is performed with the real robot. Use of virtual and simulated robots will thus allow operators to program a robot and test it before it is introduced in the production line. This may facilitate detecting failures ahead of time. The use of both models and virtual robots appears to have a large potential for replacing the off-line programming used today and for providing the flexibility which may allow more widespread use of robots in production lines.

Several groups have worked on the use of models for robot navigation and sensor interpretation. An example is the HARVEY project at the University of Massachusetts[9] where a model of the campus is used in outdoor navigation with a mobile robot. Two other systems for model based mobile robot navigation are presented in this issue.

4. REACTIVE ROBOT CONTROL

Robot systems have traditionally been built using the "perceive-plan-control" cycle shown on the right in Fig. 1. In this cycle all the actions needed are handled in a coordinated manner.

Brooks[10] has suggested that robot systems should rather be built as layered systems where different competences are distributed over several different layers. The lowest layers in the control are related to survival; i.e. avoid obstacles. Higher level layers are related to more complex goal directed activities. Brooks claims that such robot systems may be built using a minimum of representations. This claim remains to be demonstrated for large scale systems, but the idea is intriguing and a number of groups are thus pursuing robotic research using such an architectural model.

One of the problems which has received the most attention is the control in the "survival" layer. This is probably due to the fact that it is manageable to build and demonstrate methods at this layer while higher level layers require much more information and they are correspondingly much harder to demonstrate. There are, however, a few efforts directed at construction of fully operational systems based to a large extent on this idea. Examples are the robotics work at the MIT AI laboratory,[10] and the CEC-funded Basic Research project "FIRST".[11]

At the "survival" layer typical functionalities are *dodge* (avoid a moving obstacle), *escape* (move away from another object), and *track* (follow an identified object; i.e. the cordless "dog" leash). These functionalities are often termed behaviors in the sense that they to a certain extent emulate basic animal activities, where the action is driven by activities in the external environment.

5. HIGHLIGHTS

This issue contains nine papers which are highlighted below. The papers are divided into three groups: robot camera heads, model based robot navigation, and reactive control.

In construction of camera heads the set of available heads may be divided into two generations. The first generation of heads were built in the late eighties by a set of universities from USA. The first paper in the *robot camera head* section is a paper by Ferrier and Clark from Harvard University. The head described is one of the initial set of heads built in the first generation. The binocular head has facilities for change of the angle between the cameras (the vergence angle) and it is also equipped with motorized lenses. At the time of construction it was rather difficult to obtain off-the-shelf motorized lenses for a machine vision camera and they have thus chosen to use lenses for a standard 24 × 36 mm camera (and have reverse engineered the interfacing to such lenses). The constructed head is considered typical for the set of initial heads which was constructed in the late eighties. For control of the camera head the Harvard group based their system design on the human control system and they tried to implement a closed loop control cycle which has a performance similar to human vision. Very impressive results were obtained with this approach as described in the paper.

Lately a considerable interest in active vision has resulted in construction of a variety of new robot camera heads, all of which build on the experiences obtained with the first generation heads. In this issue five recent heads are described. In the second paper by Kourosh and Eklundh the KTH binocular head is described. In the design of this head special attention has been paid to accuracy and high speed to provide a research tool which to the extent possible may emulate human eye and neck motion. This robot camera head is considered one of the best heads available in the world today. It is often specified as the reference head due to its high accuracy. In addition the group at the Royal Institute of Technology in Stockholm has implemented an extensive set of reflex level behaviors on the head and there is thus an extensive experience with this head.

In the third paper by Milios, Jenkin and Tsotsos the IRIS head from The University of Toronto is described. This head is also a binocular platform with independent control of pan and tilt for each camera, but in addition the researchers have chosen to implement motor control of cyclotorsion (rotation around the optical axis). In stereo matching where the optical axes of the cameras are not parallel the epi-polar lines (lines where corresponding features are located as a function of depth) are no longer horizontal, but they have a slant. This complicates the stereo matching and it is thus desirable to have facilities for change of the camera geometry so that the epi-polar lines again become horizontal. One mechanism for achieving horizontal epi-polar lines is the use of cyclotorsion. The IRIS head has just been completed and there is thus only limited experience with use of cyclotorsion in stereo matching. It will, however, be of interest to monitor this effort, as the results obtained may indicate how future camera heads should be constructed to accomodate as simple as possible image analysis. In the fourth paper by Christensen a low-cost camera head which has been constructed at Aalborg University is described. This paper illustrates how it is possible to construct a fairly simple platform which will allow initial experiments in active vision. The head presented is the first in a series of prototypes being built at Aalborg University and a focus of this paper is thus communication of the lessons learned from construction of an initial prototype.

Most of the heads which have been constructed have been designed by researchers from a computer science or electrical engineering background. The mechanical design may have suffered from this bias. At least this is the statement from Pretlove and Parker in the fifth paper. They have built the Surrey robot camera head based on an extensive mechanical analysis and they have exploited the latest in CIM based technology for the construction. The result is a small robot head which may be fitted at the end of a traditional robot arm while still leaving space for the end effector (i.e. a gripper). The design involves use of miniature cameras and very small lenses and the entire design has been optimized to ensure that it occupies a minimum of the available payload on a robot arm.

Most of the available camera heads have a large number of degrees of freedom. To ensure a maximum of flexibility the second generation heads have been designed with independent degrees of freedom, but this also implies that control becomes more difficult. It is thus desirable to have standard device level software interfaces to the robot camera heads to allow flexible modifications to the camera heads without a need for recoding of all the software. To ensure simple interfacing to complicated devices such as a robot camera head, Crowley, Bobet and Mesrabi have proposed a standard software interface protocol for interaction with the heads. This standard protocol and its use in connection with a robot head built at LIFIA in France is described in the last paper in the robot camera head section.

In the model based robot navigation section two different but quite similar systems are described. In the USA there is a long tradition for doing mobile research, in particular due to funding from national agencies such as the Department of Energy. The aim of this research has to a large extent been the development of platforms for operation in hostile environments, such as a nuclear power plant. The construction of integrated control systems which allow simple interfacing to the mobile platform, integration of sensor reading and simulations for operator training involves many different disciplines from pattern recognition and CAD modeling to control system theory. Only recently have such systems become available and it may be expected that future generations of such systems may benefit from the lessons learned. In the first paper by Chen and Trivedi a simulation, visualization and control environment for a track-driven robot is described. In this system an operator may perform abstract teleoperation of the vehicle. The system receives overall commands from the operator and the vehicle control system then has facilities for handling the unexpected event to allow successful mission completion. The described system involves thus both autonomous robot control monitoring and facilities for simple interaction with a complex device such as a mobile robot with an on-board manipulator. In the second paper by Roth and Jain a system for monitoring and interaction with a mobile platform performing missions in an indoor environment is described. A major problem associated with model based robot navigation is structuring of information to allow efficient indexing into the model. For handling of this situation the group at the University of Michigan has developed a Context-based-Caching technique that allows on-line use of a model describing a campus or factory environment, even when the model contains a large number of details.

The availability of reasonably priced mobile platforms has resulted in a significant increase in the research related to mobile robot control, and the problems encountered with the traditional "perceive-plan-control" cycle have at the same time paved the way for experiments in control using the purposive architecture proposed by Brooks. In this section Arkin, Carter and MacKenzie present a mobile robot system which has behaviors for *dodge* and *escape*. Such behaviors may appear rather simple but the paper clearly demonstrates the potential of using such behaviors to achieve a powerful performance in an indoor environment. In this work the sensory feedback is provided by the ultrasonic sonars but the behaviors may in principle exploit any kind of sensory information.

Again, we would like to thank the authors and reviewers. We hope that the readers will find this publication timely and informative.

REFERENCES

1. D. Marr, *Vision*, W. H. Freeman and Co., 1982.
2. J. Y. Aloimonos, I. Weiss and A. Bandyopadhyay, "Active vision", in *Proc. 1st Int. Conf. on Computer Vision (ICCV)*, IEEE CS Press, Apr. 1987, pp. 35–54.
3. R. Bajcsy and D. A. Rosenthal, "Visual and conceptual focus of attention", in *Structured Computer Vision*, eds. S. Tanimoto and A. Klinger, Academic Press, 1980.
4. E. Krotkow, Exploratory visual sensing for determining spatial layout with an agile stereo camera system, Ph.D. thesis, University of Pennyslvania, Apr. 1987.
5. D. Ballard, "Animate vision", *Artif. Intell.* 48 (1991) 57–86.
6. L. Abbott and N. Ahuja, "Surface reconstruction by dynamic integration of focus, camera vergence, and stereo", in *Proc. 2nd Int. Conf. on Computer Vision (ICCV)*, eds. R. Bajcsy and S. Ullman, St. Petersburg, FL, IEEE CS Press, Dec. 1989, pp. 532–544.
7. J. J. Clark and N. J. Ferrier, "Model control of an attentive vision system", in *Proc. 2nd Int. Conf. on Computer Vision (ICCV)*, eds. R. Bajcsy and S. Ullman, St. Petersburg, FL, IEEE CS Press, Dec. 1989, pp. 514–523.
8. P. Jeanne, "Real time gaze control (Esprit project p5390)", in *Proc. ESPRIT Day, ECCV'92* (CEC DGXIII/A, ed.), St. Margherita, May 1992.
9. C. Fennema, A. Hanson and E. Riseman, "Towards autonomous mobile robot navigation", in *Proc. Image Understanding Workshop*, DARPA, Morgan Kauffman Publishers Inc., May 1989, pp. 219–231.
10. R. Brooks, "Intelligence without representation", *Artif. Intell.* 47 (1991) 139–160.
11. M. Brady, A. Blake and L. van Gool, "First: Fundamentals of intelligent reliable robot systems (ESPRIT project p. 3274)", Tech. Rep., University of Oxford, UK, 1989.