AI Magazine Volume 25 Number 2 (2004) (© AAAI)

Cognitive Vision

Henrik I. Christensen

■ The integration of AI and vision has been a long-term goal of both disciplines for more than three decades. This special issue illustrates some recent work on bridging the gap.

Forty years ago, the integration of computational vision and AI was considered in the thesis of L. G. Roberts (1963). Back then, the issue was anticipated relatively easily, and a full integration of the two fields was expected within a decade. Processing of images is known to result in noisy segmentation, and in general, data might not be consistent in space or time, making AI methods appealing. In early AI, the handling of noisy, partially inconsistent data was, at best, a major challenge. Nevertheless, there have been major efforts in knowledge-based interpretation and vision, for example, in Hanson and Riseman's (1978) seminal book and as part of the Defense Advanced Research Projects Agency's Image Understanding Program. One landmark was the ACRONYM system (Brooks 1983, 1981), which performed knowledge-based interpretation of three-dimensional objects. Despite these efforts, the emergence of fully integrated systems for complex tasks has been relatively limited. One of the limitations has been access to adequate computer facilities; another has been the lack of efficient methods for reasoning under uncertainty.

Because of the early efforts in integration of AI and computer vision, the available computer power has increased by more than 5×10^7; at the same time, there has been tremendous progress in reasoning under uncertainty (now an area with its own major conference). In addition, computer vision has also matured in terms of modeling, and today, there are well-developed models for description of features and assemblies of features, motion, and objects that use a sound Bayesian framework. Consequently, vision is gradually becoming a methodology that fits well with the corresponding methods in AI. Thus, there has recently been a major upsurge in research that integrates modern methods from AI and vision into systems. One of these research efforts is a major program run by the European Commission called the Cognitive Vision Program. This program, begun in 2000, currently involves 10 major research projects and a network for community integration called ECVISION. In addition, there is an effort in the United States to examine cognitive systems.

Scanning the Issue

In this special issue of *AI Magazine,* three different approaches to cognitive vision are presented.

In the first article, Ernst Dickmanns presents the results of a long-term effort to use vision for the autonomous control of cars. More than 20 years ago, Dickmanns embarked on an effort to build a vehicle that could drive autonomously on the road. The initial system could drive autonomously on a highway. The sheer need for computer power implied that the car had to be a truck; otherwise, it would have been too small to hold all the computers. Gradually, the methodology has been refined and extended into an impressive system for autonomous driving, automatic lane changing, interpretation of traffic patterns, and so on. The advances in computer power have at the same time resulted in a reduction in system size to a system that today can easily fit into a small rack in the back of a small car. The article illustrates how careful modeling of the imaging process, the target vehicle, and the domain of traffic allows design of highly competent systems that can be used not only on the highway but also in heavy traffic in urban environments. This study illustrates a holistic approach to the design of an advanced AI system for a target domain.

The second article, by Hans-Helmutt Nagel, is another example of the long-term effort to construct vision systems for a known hard problem. More than 20 years ago, Nagel initiated work on interpretation of traffic and dynamics in a scene. In vision, his Hamburg taxi sequence is still considered a landmark in motion interpretation. Gradually, the work on interpretation of motion has grown into systems that allow for interpretation of advanced interaction. Today, there are methods for interpreting human activities and traffic patterns. Through a long-term effort, the work at Karlsruhe has grown from computer vision– based motion analysis into an AI effort that allows interpretation of motion patterns and dynamic interaction. Today, the effort is at a level where dynamic scene interaction can be described in terms of natural language, which again illustrates how a number of different methods in terms of modeling, signal-symbol processes, and interpretation come together to allow for the design of a highly competent system. At the same time, the work addresses a number of the fundamental problems involved in the design of truly cognitive systems. Although Nagel's article uses a particular domain to exemplify the problem of cognitive systems, at the same time, it points to many of the basic problems involved in a systems approach to the design of intelligent systems.

The third article, by Gösta Granlund and Anders Moe, addresses another of the fundamental problems involved in the design of cognitive systems—the issue of object recognition. Early research in vision relied on a reconstruction approach to recognition (Marr 1982) in which features were assembled into geometric structures as a basis for identification. One problem here is the lack of a task that can be used to make the problem tractable. In addition, much of the early research relied on edges or regions as a single cue for recognition. No single cue is, in general, robust across environmental variations. At the same time, advances in statistical methods have allowed a change in paradigm toward appearance-based methods that are all biologically plausible (Edelman 1999). The article by Granlund and Moe addresses the issue of integrating visual cues with descriptors to allow efficient indexing that will lead to recognition. At Linköping, there has been another long-term effort to provide robust statistically significant image features that have a sound basis in signal processing. Over time, the effort has been refined and extended from feature processing into recognition and categorization. The article illustrates how these methods integrate into a system to be used for autonomous systems such as robots or flying vehicles.

These three articles have different application domains, but all of them are based on long-term efforts to address the integration of vision and interpretation. It is characteristic that at the outset, they all take a holistic approach to the overall problem. At the same time, they illustrate that today, it is possible to construct highly competent systems. However, at the same time, it is evident that the efforts involve a number of different disciplines, and as such, the sheer scope of the problem calls for relatively large teams to construct operational systems.

Future Directions

The progress reported in the three articles in this special *AI Magazine* section is highly encouraging. At the same time, it is gradually recognized that vision is part of a larger puzzle. If we are to construct intelligent artifacts such as humanoids and service robots, the vision system must be considered as a component of a more complex system. For tasks such as manipulation, we need to integrate vision with haptics to handle objects. For locomotion through the environment, other modalities such as inertial sensors provide important information for the task. At the same time, execution of tasks such as "pick up the milk from the refrigerator" requires task planning, localization (recognition), manipulation, servoing, and so on. To address these problems, the vision subsystems must be integrated into cognitive systems; so, the National Science Foundation, the European Union Commission of the European Community, the Korea Science and Engineering Foundation, and the Japanese Ministry of International Trade and Industry have recently launched efforts to examine cognitive systems. These efforts will, without doubt, offer interesting challenges not only in vision but at all levels of intelligent system construction, that is, in terms of architectures, language (and language acquisition), perception, planning, learning, and integration.

References

Brooks, R. A. 1983. Model-Based Three-Dimensional Interpretation of Two-Dimensional Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5(2): 140–149.

Brooks, R. A. 1981. Symbolic Reasoning among 3D Models and 2D Images. *Artificial Intelligence* 17(1–3): 285–348.

Edelman, S. 1999. *Representation and Recognition in Vision.* Cambridge, Mass: The MIT Press.

Hanson, A., and Riseman, E., eds. 1978. *Computer Vision Systems*. San Diego, Calif.: Academic.

Marr, D. C. 1982. *Vision.* San Francisco: W. H. Freeman.

Roberts, L. G. 1963. Machine Perception of 3D Solids. Ph.D. dissertation, Dept. of Electrical Engineering, Massachusetts Institute of Technology.

Henrik I. Christensen is a full professor of computer science at the Royal Institute of Technology, Stockholm Sweden. He is also the director of the Centre for Autonomous Systems and the coordinator of the EURON network. He has published more than 180 papers on vision, robotics, and AI, and serves on the editorial board of a number of prominent journals. His e-mail address is hic@nada.kth.se.