



Videre: Journal of Computer Vision Research

Quarterly Journal

Winter 2000, Volume 1, Number 4

The MIT Press

Introduction

Computer Vision Systems

Henrik I. Christensen
James L. Crowley

Videre: Journal of Computer Vision Research (ISSN 1089-2788) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals \$30.00, Institutions \$125.00. Canadians add additional 7% GST. Prices subject to change without notice.

Subscribers are licensed to use journal articles in a variety of ways, limited only as required to insure fair attribution to authors and the Journal, and to prohibit use in a competing commercial product. See the Journals World Wide Web site for further details. Address inquiries to the Subsidiary Rights Manager, MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2864; fax: (617) 258-5028; e-mail: journals-rights@mit.edu.

Computer Vision Systems

Henrik I. Christensen,¹
James L. Crowley²

1 Introduction

Computer vision has been an active area of research for more than three decades. A quick review of the field reveals that image processing and pattern recognition has been tremendously successful in terms of delivering operational systems. Everyday barcode scanners are used in supermarkets, and pattern recognition techniques are used for such purposes as identification, bill recognition, and address recognition. This progress has been driven by significant increases in computer power and the widespread use of multimedia technology. The use of imaging technology for everyday tasks and the wide use of images as part of the Web has resulted in a tremendous reduction in price and delivery of images to virtually every desktop. In terms of full-scale computer vision applications that involve motion estimation, depth recovery, and scene interpretation, fairly limited progress has been achieved. Few systems have been deployed for regular use. A few notable exceptions include the Highway driving system by Prof. Dickmanns, the NAVLAB system from Carnegie Mellon University, and a few systems for medical diagnostics and computer-aided surgery. Overall, it has, however, not been possible to deliver the expected set of applications due to a lack of robustness and methods for integration of methods into full systems.

At the verge of a new millenium, there is now a number of promising efforts to construct fully operational systems. The construction of systems gradually provides the insight to formulate basic methods for the design, analysis, implementation, and evaluation of operational and robust systems that have a wide domain of application. Most of the established conferences in computer vision such as ICCV, ECCV, CVPR, ACCV, and ICPR focus on component methods for computer vision with little emphasis on integration. To provide a forum for discussion of system issues, the first International Conference on Computer Vision Systems (ICVS) was held January, 1999, at the beautiful island of Gran Canaria, Spain. The program included more than sixty papers and three invited lectures by Prof. Dickmanns (automated highway driving), Prof. Kanade (autonomous helicopter flying), and Prof. Brady (image-based medical diagnostics). A total of ninety people attended the conference. A set of outstanding papers from the program were selected for a *VIDERE* special issue. A total of four papers are presented in this issue. All of the papers have been extended, revised, and subjected to the normal review process.

2 Scanning the Issue

The first paper “Active Knowledge-Based Scene Analysis” by Paulus et al. describes a methodology for the design of knowledge-based interpretation of scenes. Many earlier systems have relied on extensive (passive) knowledge bases that have been coupled to standard reasoning engines. In contrast, the presented approach combined declarative knowledge

1. Royal Institute of Technology
Stockholm, Sweden
hic@nada.kth.se
2. INRIA Alpes
Grenoble, France
jim.crowley@inrialpes.fr

and associated reasoning methods into active knowledge-based units. The system uses a modern object-oriented design approach that enables the design of a toolkit for the easy integration of components into operational systems. To illustrate the potential of the developed toolkit, two different applications are presented. One application is used for indexing and recognition of a set of objects in a traditional office setting. The second application is a system for landmark-based navigation in an indoor environment. The paper nicely illustrates how modern software engineering techniques, when combined with standard methods for signal analysis and pattern recognition, can be used to design a scalable toolkit for recognizing objects in the presence of significant clutter.

Automatic recognition in an indoor environment for robot navigation has a significant potential for a range of applications. The second paper, "Integrating Vision-Based Behaviors with an Autonomous Robot" by Schlegel et al., presents a hybrid deliberative architecture for the integration of visual behaviors for robot navigation. Through tight coupling of perception and action in combination with the use of dedicated models, it is possible to achieve real-time performance. This paper presents the overall architecture and a range of visual behaviors for recognition of landmarks and objects and the simple tracking of people. Through integration of such visual behaviors, it is possible to provide basic navigation functionalities and methods for intelligent interaction with humans in the environment. The presented architecture and visual methods have been implemented on a B21 robot that can perform navigation in an office environment.

The integration of multiple sensory modalities allow the construction of efficient interfaces for intelligent systems. The third paper (by Wachsmuth et al.) entitled "Integration of Vision and Speech Understanding using Bayesian Networks," presents a methodology for multimodel integration. Many earlier approaches to scene interpretation have ignored uncertainty. In this paper, an integrated approach is presented. A spoken dialogue is used to generate a partial model for interpretation. The partial model provides a context for interpretation of visual input. To enable the handling of ambiguity and the flexible handling of the inherent uncertainty in processing visual data, the processing is controlled by a Bayesian Network. The network is generated dynamically in response to spoken input. Through adaptation of a dynamic approach, it is possible to generate computationally efficient strategies for interpretation. The presented approach is evaluated in the context of a system that allows automatic assembly of a toy-like object based on spoken commands. The paper illustrates how uncertainty and dynamic modeling can be used for efficient scene interpretation.

A notorious problem in computer vision is object recognition. Typically, recognition strategies are engineered to handle specific objects or classes of objects. Such an approach is typically labor intensive and iterative. Our fourth paper, "ADORE: Adaptive Object Recognition" by Draper et al., presents a methodology for the design of recognition strategies using learning techniques. Based on training using a wide range of feature detection and extraction methods, it is possible to automatically generate recognition strategies that can be deployed for the recognition of complex objects. To illustrate the potential of learning for generation of recognition strategies, the method is evaluated in the context of remote imaging. A set of houses are recognized in serial images. This task is typically carried out manually, and it is thus possible to compare performance of the automatic strategy to "ground truth." Performance

characterization is gradually becoming mainstream, and this is an excellent illustration of this trend. In addition, the paper illustrates how machine learning can be used with success in computer vision.

This issue of *VIDERE* also includes a review of the book “An Introduction to Natural Computation,” by Dana H. Ballard. This book brings together a wide range of disciplines to demonstrate how different disciplines contribute to competent systems. The book is thus an excellent illustration of how an interdisciplinary approach must be adopted to enable the design, analysis, and deployment of full systems.

3 The Future

Based on the success of the first ICVS, follow-up activities are being discussed. Although it is not yet obvious how and when a second conference will be organized, it is felt that a follow-up is needed as systems gradually are becoming available. It is considered essential that the basic methodology for integration of systems is defined in a formal framework to enable more widespread use.

Acknowledgments

This special issue would never have been made if it were not for Chris Brown and Giulio Sandini. They organized most of the work and handled all the practical matters. We are most grateful for this opportunity to present computer vision systems to a wide audience.

ICVS received financial support in particular from the Spanish government and a number of agencies in Spain; we are most grateful for this support. Finally, the conference would never have grown into what it was without the enormous effort of Jorge Cabrera. Jorge went beyond all expectation in his local organization of the conference. Finally, we want to thank MIT Press and the reviewers for the assistance in making this an excellent issue.