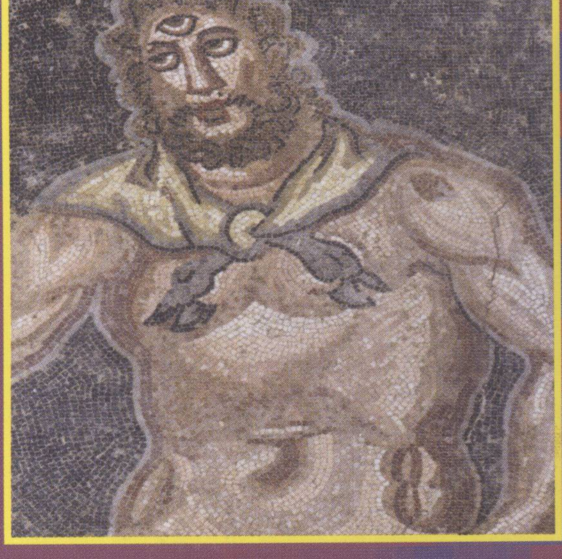
State-of-the-Art Survey

LNCS 4170

Jean Ponce
Martial Hebert
Cordelia Schmid
Andrew Zisserman (Eds.)

Toward Category-Level Object Recognition



Springer

Jean Ponce Martial Hebert
Cordelia Schmid Andrew Zisserman (Eds.)

Toward Category-Level Object Recognition

Springer

Volume Editors

Jean Ponce
Ecole Normale Supérieure
Département d'Informatique
45 rue d'Ulm, 75230 Paris Cedex 05, France
E-mail: jean.ponce@ens.fr

Martial Hebert
Carnegie Mellon University
The Robotics Institute
Pittsburgh PA 15213, USA
E-mail: hebert@ri.cmu.edu

Cordelia Schmid
INRIA Rhône-Alpes
665, avenue de l'Europe, 38330 Montbonnot, France
E-mail: cordelia.schmid@inrialpes.fr

Andrew Zisserman
University of Oxford
Department of Engineering Science
Parks Road, Oxford OX1 3PJ, UK
E-mail: az@robots.ox.ac.uk

The cover illustration is a detail from the mosaic of Ulysses offering wine to Polyphemus, Villa Romana del Casale, Sicily. Permission to reproduce this detail was kindly granted by the J. Paul Getty Trust. © The J. Paul Getty Trust, 2006. All rights reserved.

Library of Congress Control Number: 2006938342

CR Subject Classification (1998): I.4, I.2.10, I.2.6, I.5.4, F.2.2

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-68794-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-68794-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11957959 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 4170

Preface

Object recognition —or, in a broader sense, scene understanding— is the ultimate scientific challenge of computer vision: After 40 years of research, robustly identifying the familiar objects (chair, person, pet), scene categories (beach, forest, office), and activity patterns (conversation, dance, picnic) depicted in family pictures, news segments, or feature films is still far beyond the capabilities of today's vision systems. On the other hand, truly successful object recognition and scene understanding technology will have a broad impact in application domains as varied as defense, entertainment, health care, human–computer interaction, image retrieval and data mining, industrial and personal robotics, manufacturing, scientific image analysis, surveillance and security, and transportation.

Although research in computer vision for recognizing 3D objects in photographs dates back to the 1960s, progress has been relatively slow and only now do we see the emergence of effective techniques for recognizing object categories with different appearances under large variations in the observation conditions. While much of the early work relied almost exclusively on geometric methods, modern recognition techniques are appearance-based, in which methods from standard statistical pattern recognition are applied to image descriptors. Tremendous progress has been achieved in the past five years, thanks in large part to the integration of new data representations, such as invariant semilocal features, developed in the computer vision community with the effective models of data distribution and classification procedures developed in the statistical machine-learning community.

This book exemplifies this progress. It is the outcome of two workshops that were held in Taormina in 2003 and 2004, and brought together about 40 prominent vision and machine-learning researchers interested in the fundamental and applicative aspects of object recognition, as well as representatives of industry. The main goals of these two workshops were (1) to promote the creation of an international object recognition community, with common datasets and evaluation procedures, (2) to map the state of the art and identify the main open problems and opportunities for synergistic research, and (3) to articulate the industrial and societal needs and opportunities for object recognition research worldwide.

These concerns are reflected in this book. Collecting all the workshops' contributions into a single book would have been impossible. We chose instead to select a relatively small number of papers that illustrate the breadth of today's object recognition research and the arsenal of techniques at its disposal and that discuss current achievements and outstanding challenges.

The book is divided into five parts. Each part includes a series of chapters written by contributors to the workshops. Most of the chapters are descriptions of technical approaches, intended to capture the current state of the art. Some of the chapters are of a tutorial nature. They cover fundamental building blocks for object recognition techniques.

Part I of the book introduces general background material on the state of object recognition research. We begin with a review of the history of the field, which sets the stage for the more recent developments reported later in the book. We then discuss the need for consistent evaluation procedures and common, challenging, datasets. This is a crucial aspect since, as the field matures, systematic evaluation of the different approaches becomes increasingly important. We conclude Part I with a discussion of the industrial needs and opportunities. As we shall see, the technology has matured to a point at which exciting applications are becoming possible.

Part II focuses on recognizing *specific* objects, an area where significant progress has occurred over the past five years. This is in part due to the advent of effective techniques for detecting and describing image patches with a controlled degree of invariance, together with efficient matching and indexing algorithms that exploit both local appearance models and powerful global geometric constraints arising from perspective imaging. As demonstrated by the five chapters making up this part of the book, reliable methods for localizing specific objects in photographs and video clips despite occlusion, clutter, and changes in viewpoint are now available.

Part III of the book attacks the difficult problem of category-level object recognition. In the methods described in these chapters, object categories are represented by collections of image patches (fixed image windows or invariant patches such as those used in Part II), potentially augmented with weak spatial layout constraints. The emphasis is on the generative or discriminative techniques used to learn the distribution of these features and their relationships, and subsequently used to classify the image instances.

Part IV investigates part-based object models that incorporate stronger structural components in the form of explicit geometric constraints, or tree-structured part assemblies, for example. The emphasis there is on the definition and identification of parts as well as on efficient algorithms for detecting object instances as part assemblies in images.

Finally, Part V of the book is concerned with classifying the image pixels into object foreground vs background (as opposed to simply detecting an object instance). As shown in the chapters making up this part, this process leads to a new, well-posed view of image segmentation incorporating both bottom-up and top-down interpretation processes.

This book is a testimony to the amazing progress achieved in object recognition research in the past five years. But much remains to be done: We can now recognize a limited number of categories in constrained settings (e.g., from particular viewpoints). However, *understanding* an image or video still remains an open problem. We must also improve current datasets and evaluation criteria to avoid toy problems and to allow meaningful comparisons (see the chapter on "Datasets" in Part I, for more on this issue). Further, category-level object recognition is today essentially viewed as a statistical pattern matching problem. The emphasis is in general on the features defining the patterns and the machine-learning techniques used to learn and recognize them, rather than on the representation of object, scene, and activity categories or the integrated interpretation of the various scene elements. Future progress will require explicitly addressing the representational issues involved in object recognition and, more generally, scene understanding. Contextual issues and hierarchical, incremental learning of a large number of categories must also be addressed. Exciting times lie ahead.

Acknowledgments. The two workshops were supported in part by the National Science Foundation under grant IIS-0335780, DARPA, the Institut National de la Recherche en Informatique et Automatique, the PASCAL European Network of Excellence IST-2002-506778, France Telecom, General Electric, Intel, Lockheed Martin, Microsoft Research, Toyota, and Xerox.

October 2006

Jean Ponce
Martial Hebert
Cordelia Schmid
Andrew Zisserman

Lecture Notes in Computer Science

For information about Vols. 1–4247

please contact your bookseller or Springer

Vol. 4355: J. Julliand, O. Kouchnarenko (Eds.), B 2007: Formal Specification and Development in B. XIII, 293 pages. 2006.

Vol. 4345: N. Maglaveras, I. Chouvarda, V. Koutkias, R. Brause (Eds.), Biological and Medical Data Analysis. XIII, 496 pages. 2006. (Sublibrary LNBI).

Vol. 4338: P. Kalra, S. Peleg (Eds.), Computer Vision, Graphics and Image Processing. XV, 965 pages. 2006.

Vol. 4337: S. Arun-Kumar, N. Garg (Eds.), FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science. XIII, 430 pages. 2006.

Vol. 4333: U. Reimer, D. Karagiannis (Eds.), Practical Aspects of Knowledge Management. XII, 338 pages. 2006. (Sublibrary LNAI).

Vol. 4331: G. Min, B. Di Martino, L.T. Yang, M. Guo, G. Ruenger (Eds.), Frontiers of High Performance Computing and Networking – ISPA 2006 Workshops. XXXVII, 1141 pages. 2006.

Vol. 4330: M. Guo, L.T. Yang, B. Di Martino, H.P. Zima, J. Dongarra, F. Tang (Eds.), Parallel and Distributed Processing and Applications. XVIII, 953 pages. 2006.

Vol. 4329: R. Barua, T. Lange (Eds.), Progress in Cryptology - INDOCRYPT 2006. X, 454 pages. 2006.

Vol. 4326: S. Göbel, R. Malkewitz, I. Iurgel (Eds.), Technologies for Interactive Digital Storytelling and Entertainment. X, 384 pages. 2006.

Vol. 4325: J. Cao, I. Stojmenovic, X. Jia, S.K. Das (Eds.), Mobile Ad-hoc and Sensor Networks. XIX, 887 pages. 2006.

Vol. 4320: R. Gotzhein, R. Reed (Eds.), System Analysis and Modeling: Language Profiles. X, 229 pages. 2006.

Vol. 4319: L.-W. Chang, W.-N. Lie (Eds.), Advances in Image and Video Technology. XXVI, 1347 pages. 2006.

Vol. 4318: H. Lipmaa, M. Yung, D. Lin (Eds.), Information Security and Cryptology. XI, 305 pages. 2006.

Vol. 4313: T. Margaria, B. Steffen (Eds.), Leveraging Applications of Formal Methods. IX, 197 pages. 2006.

Vol. 4312: S. Sugimoto, J. Hunter, A. Rauber, A. Morishima (Eds.), Digital Libraries: Achievements, Challenges and Opportunities. XVIII, 571 pages. 2006.

Vol. 4311: K. Cho, P. Jacquet (Eds.), Technologies for Advanced Heterogeneous Networks II. XI, 253 pages. 2006.

Vol. 4309: P. Inverardi, M. Jazayeri (Eds.), Software Engineering Education in the Modern Age. VIII, 207 pages. 2006.

Vol. 4308: S. Chaudhuri, S.R. Das, H.S. Paul, S. Tirthapura (Eds.), Distributed Computing and Networking. XIX, 608 pages. 2006.

Vol. 4307: P. Ning, S. Qing, N. Li (Eds.), Information and Communications Security. XIV, 558 pages. 2006.

Vol. 4306: Y. Avrithis, Y. Kompatsiaris, S. Staab, N.E. O'Connor (Eds.), Semantic Multimedia. XII, 241 pages. 2006.

Vol. 4305: A.A. Shvartsman (Ed.), Principles of Distributed Systems. XIII, 441 pages. 2006.

Vol. 4304: A. Sattar, B.-H. Kang (Eds.), AI 2006: Advances in Artificial Intelligence. XXVII, 1303 pages. 2006. (Sublibrary LNAI).

Vol. 4302: J. Domingo-Ferrer, L. Franconi (Eds.), Privacy in Statistical Databases. XI, 383 pages. 2006.

Vol. 4301: D. Pointcheval, Y. Mu, K. Chen (Eds.), Cryptology and Network Security. XIII, 381 pages. 2006.

Vol. 4300: Y.Q. Shi (Ed.), Transactions on Data Hiding and Multimedia Security I. IX, 139 pages. 2006.

Vol. 4297: Y. Robert, M. Parashar, R. Badrinath, V.K. Prasanna (Eds.), High Performance Computing - HiPC 2006. XXIV, 642 pages. 2006.

Vol. 4296: M.S. Rhee, B. Lee (Eds.), Information Security and Cryptology – ICISC 2006. XIII, 358 pages. 2006.

Vol. 4295: J.D. Carswell, T. Tezuka (Eds.), Web and Wireless Geographical Information Systems. XI, 269 pages. 2006.

Vol. 4294: A. Dan, W. Lamersdorf (Eds.), Service-Oriented Computing – ICSOC 2006. XIX, 653 pages. 2006.

Vol. 4293: A. Gelbukh, C.A. Reyes-Garcia (Eds.), MICAI 2006: Advances in Artificial Intelligence. XXVIII, 1232 pages. 2006. (Sublibrary LNAI).

Vol. 4292: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXII, 906 pages. 2006.

Vol. 4291: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXI, 916 pages. 2006.

Vol. 4290: M. van Steen, M. Henning (Eds.), Middleware 2006. XIII, 425 pages. 2006.

Vol. 4289: M. Ackermann, B. Berendt, M. Grobelnik, A. Hotho, D. Mladenič, G. Semeraro, M. Spiliopoulou, G. Stumme, V. Svatek, M. van Someren (Eds.), Semantics, Web and Mining. X, 197 pages. 2006. (Sublibrary LNAI).

Vol. 4288: T. Asano (Ed.), Algorithms and Computation. XX, 766 pages. 2006.

Vol. 4287: C. Mao, T. Yokomori (Eds.), DNA Computing. XII, 440 pages. 2006.

Vol. 4286: P. Spirakis, M. Mavronicolas, S. Kontogiannis (Eds.), Internet and Network Economics. XI, 401 pages. 2006.

Vol. 4285: Y. Matsumoto, R. Sproat, K.-F. Wong, M. Zhang (Eds.), Computer Processing of Oriental Languages. XVII, 544 pages. 2006. (Sublibrary LNAI).

Vol. 4284: X. Lai, K. Chen (Eds.), Advances in Cryptology – ASIACRYPT 2006. XIV, 468 pages. 2006.

Vol. 4283: Y.Q. Shi, B. Jeon (Eds.), Digital Watermarking. XII, 474 pages. 2006.

Vol. 4282: Z. Pan, A.D. Cheok, M. Haller, R.W.H. Lau, H. Saito, R. Liang (Eds.), Advances in Artificial Reality and Tele-Existence. XXIII, 1347 pages. 2006.

Vol. 4281: K. Barkaoui, A. Cavalcanti, A. Cerone (Eds.), Theoretical Aspects of Computing - ICTAC 2006. XV, 371 pages. 2006.

Vol. 4280: A.K. Datta, M. Gradinariu (Eds.), Stabilization, Safety, and Security of Distributed Systems. XVII, 590 pages. 2006.

Vol. 4279: N. Kobayashi (Ed.), Programming Languages and Systems. XI, 423 pages. 2006.

Vol. 4278: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Part II. XLV, 1004 pages. 2006.

Vol. 4277: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Part I. XLV, 1009 pages. 2006.

Vol. 4276: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, Part II. XXXII, 752 pages. 2006.

Vol. 4275: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, Part I. XXXI, 1115 pages. 2006.

Vol. 4274: Q. Huo, B. Ma, E.-S. Chng, H. Li (Eds.), Chinese Spoken Language Processing. XXIV, 805 pages. 2006. (Sublibrary LNAI).

Vol. 4273: I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, L. Aroyo (Eds.), The Semantic Web - ISWC 2006. XXIV, 1001 pages. 2006.

Vol. 4272: P. Havinga, M. Lijding, N. Meratnia, M. Wegdam (Eds.), Smart Sensing and Context. XI, 267 pages. 2006.

Vol. 4271: F.V. Fomin (Ed.), Graph-Theoretic Concepts in Computer Science. XIII, 358 pages. 2006.

Vol. 4270: H. Zha, Z. Pan, H. Thwaites, A.C. Addison, M. Forte (Eds.), Interactive Technologies and Sociotechnical Systems. XVI, 547 pages. 2006.

Vol. 4269: R. State, S. van der Meer, D. O'Sullivan, T. Pfeifer (Eds.), Large Scale Management of Distributed Systems. XIII, 282 pages. 2006.

Vol. 4268: G. Parr, D. Malone, M. Ó Foghlú (Eds.), Autonomic Principles of IP Operations and Management. XIII, 237 pages. 2006.

Vol. 4267: A. Helmy, B. Jennings, L. Murphy, T. Pfeifer (Eds.), Autonomic Management of Mobile Multimedia Services. XIII, 257 pages. 2006.

Vol. 4266: H. Yoshiura, K. Sakurai, K. Rannenberg, Y. Murayama, S. Kawamura (Eds.), Advances in Information and Computer Security. XIII, 438 pages. 2006.

Vol. 4265: L. Todorovski, N. Lavrač, K.P. Jantke (Eds.), Discovery Science. XIV, 384 pages. 2006. (Sublibrary LNAI).

Vol. 4264: J.L. Balcázar, P.M. Long, F. Stephan (Eds.), Algorithmic Learning Theory. XIII, 393 pages. 2006. (Sublibrary LNAI).

Vol. 4263: A. Levi, E. Savaş, H. Yenigün, S. Balcısoy, Y. Saygın (Eds.), Computer and Information Sciences – ISCIS 2006. XXIII, 1084 pages. 2006.

Vol. 4262: K. Havelund, M. Núñez, G. Roşu, B. Wolff (Eds.), Formal Approaches to Software Testing and Runtime Verification. VIII, 255 pages. 2006.

Vol. 4261: Y. Zhuang, S. Yang, Y. Rui, Q. He (Eds.), Advances in Multimedia Information Processing - PCM 2006. XXII, 1040 pages. 2006.

Vol. 4260: Z. Liu, J. He (Eds.), Formal Methods and Software Engineering. XII, 778 pages. 2006.

Vol. 4259: S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H.S. Nguyen, R. Słowiński (Eds.), Rough Sets and Current Trends in Computing. XXII, 951 pages. 2006. (Sublibrary LNAI).

Vol. 4257: I. Richardson, P. Runeson, R. Messnarz (Eds.), Software Process Improvement. XI, 219 pages. 2006.

Vol. 4256: L. Feng, G. Wang, C. Zeng, R. Huang (Eds.), Web Information Systems – WISE 2006 Workshops. XIV, 320 pages. 2006.

Vol. 4255: K. Aberer, Z. Peng, E.A. Rundensteiner, Y. Zhang, X. Li (Eds.), Web Information Systems – WISE 2006. XIV, 563 pages. 2006.

Vol. 4254: T. Grust, H. Höpfner, A. Illarramendi, S. Jablonski, M. Mesiti, S. Müller, P.-L. Patranjan, K.-U. Sattler, M. Spiliopoulou, J. Wijsen (Eds.), Current Trends in Database Technology – EDBT 2006. XXXI, 932 pages. 2006.

Vol. 4253: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. XXXII, 1301 pages. 2006. (Sublibrary LNAI).

Vol. 4252: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. XXXIII, 1335 pages. 2006. (Sublibrary LNAI).

Vol. 4251: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LXVI, 1297 pages. 2006. (Sublibrary LNAI).

Vol. 4250: H.J. van den Herik, S.-C. Hsu, T.-s. Hsu, H.H.L.M. Donkers (Eds.), Advances in Computer Games. XIV, 273 pages. 2006.

Vol. 4249: L. Goubin, M. Matsui (Eds.), Cryptographic Hardware and Embedded Systems - CHES 2006. XII, 462 pages. 2006.

Vol. 4248: S. Staab, V. Svátek (Eds.), Managing Knowledge in a World of Networks. XIV, 400 pages. 2006. (Sublibrary LNAI).

¥701.00元

Table of Contents

I Introduction

II Recognition of Specific Objects

III Recognition of Object Categories

IV Recognition of Object Categories with Geometric Relations

V Joint Recognition and Segmentation

Part I

Introduction

Object Recognition in the Geometric Era: A Retrospective

Joseph L. Mundy

Division of Engineering,
Brown University
Providence, Rhode Island
mundy@lems.brown.edu

Abstract. Recent advances in object recognition have emphasized the integration of intensity-derived features such as affine patches with associated geometric constraints leading to impressive performance in complex scenes. Over the four previous decades, the central paradigm of recognition was based on formal geometric object descriptions with a focus on the properties of such descriptions under perspective image formation. This paper will review the key advances of the geometric era and investigate the underlying causes of the movement away from formal geometry and prior models towards the use of statistical learning methods based on appearance features.

1 Introduction

Object recognition by computer has been an active area of research for nearly five decades. For much of that time, the approach has been dominated by the discovery of analytic representations (models) of objects that can be used to predict the appearance of an object under any viewpoint and under any conditions of illumination and partial occlusion. The expectation is that ultimately a representation will be discovered that can model the appearance of broad object categories and in accordance with the human conceptual framework so that the computer can "tell" what it is seeing.

Advantages of Geometric Description. From the earliest attempts at recognition, geometric representations have dominated the development of the theory and resulting algorithms and systems. There are a number of reasons why geometry has played such a central role.

- Invariance to viewpoint - Geometric object descriptions allow the projected shape of an object to be accurately predicted under perspective projection.
- Invariance to illumination - recognizing geometric descriptions from images can be achieved using edge detection and geometric boundary segmentation. Such descriptions are reasonably invariant to illumination variations.
- Well developed theory - geometry has been under active investigation by mathematicians for thousands of years. The geometric framework has achieved a high degree of maturity and effective algorithms exist for analyzing and manipulating geometric structures.

J. Ponce et al. (Eds.): Toward Category-Level Object Recognition, LNCS 4170, pp. 3–28, 2006.

– Man-made objects - a large fraction of manufactured objects are designed using computer-aided design (CAD) models and therefore are naturally described by primitive geometric elements, such as planes and spheres. More complex shapes are also represented with simple geometric descriptions, such as a triangular mesh or polynomial patches.

There are, of course, deficiencies of the geometric approach to recognition, but the discussion of such limitations will be postponed until after a review of the broad sweep of geometric recognition research over the last four decades.

2 The Beginning

In the 1950s and early 1960s ideas from signal processing and detection theory, such as autocorrelation and template matching, were exploited to form the first object recognition systems. Much of the research focus was on 2-d pattern classification applications such as character recognition, fingerprint analysis and microscopic cell classification. These early decades were dominated by methods of statistical pattern recognition and perception classifiers based on parametric learning. Even so, the features used in these classification schemes were often derived from geometric descriptions. For example, an early approach [34] (1962) to the definition of features for character recognition was based on geometric invariance using moments. Geometric invariance will re-appear as a major research thrust in the early 1990s, three decades later. This example illustrates that recognition ideas are continually re-visited as computational power and feature segmentation methods advance.

2.1 The Blocks World

The dependence on statistics and signal methods rapidly gave way to the theme of *artificial intelligence*, coined by Marvin Minsky and John McCarthy around 1956. The new approach focussed on establishing a theoretical framework for cognitive tasks, such as vision, where computers could carry out the necessary reasoning using formal logic and other mathematical tools. The plan was to start with a simplification of the world so that the mathematical models can apply rigorously and to solve the resulting recognition problem completely before proceeding to more difficult situations.

For the computer vision problem, this simplification is called *the blocks world* where objects are restricted to polyhedral shapes on a uniform background. Polyhedra have simple and easily represented geometry and the projection of polyhedra into images under perspective can be straightforwardly modeled with a projective transformation. Under this projection, lines in 3-d map to lines in 2-d and polyhedral faces project to polygons. The goal is to be able to recognize general polyhedral shapes in an arbitrary spatial arrangement including significant occlusion of one object by itself or others.

The blocks world framework dominated the vision research agenda for over a decade before it was abandoned to tackle more realistic scenes. It is not that all the problems of recognizing polyhedral objects and structures made up of polyhedra were definitively and completely solved. Instead it became clear that too many assumptions were being made in recognition strategies that could not be expected to hold in real world scenes. This tension between the desire for a sound theoretical basis for recognition and the ability to confront the complexities of recognizing complex objects such as trees and the human form, will re-immerge repeatedly during the geometric era.

2.2 Roberts and the Blocks World

Perhaps the most complete and powerful recognition system of the blocks world was that of L. G. Roberts [64]. Roberts' recognition algorithm exhibited most of the steps that are still followed today, some four decades later. He carefully considered how polyhedra project into perspective images and established a generic library of polyhedral components that could be assembled into a composite structure. His philosophy towards recognition is defined by the quote, '... we shall assume that the objects seen could be constructed out of parts with which we

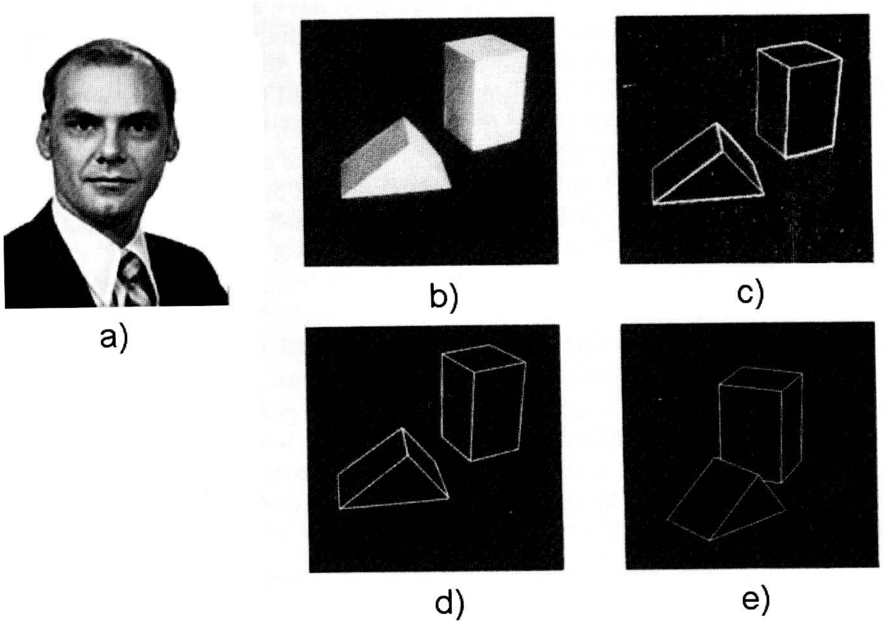


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)