Object Categorization

This edited volume presents a unique multidisciplinary perspective on the problem of visual object categorization. The result of a series of four highly successful workshops on the topic, the book gathers many of the most distinguished researchers from both computer and human vision to reflect on their experience, identify open problems, and foster a cross-disciplinary discussion with the idea that parallel problems and solutions have arisen in both domains.

Twenty-seven of these workshop speakers have contributed chapters, including fourteen from computer vision and thirteen from human vision. Their contributions range from broad perspectives on the problem to more specific approaches, collectively providing important historical context, identifying the major challenges, and presenting recent research results. This multidisciplinary collection is the first of its kind on the topic of object categorization, providing an outstanding context for graduate students and researchers in both computer and human vision.

Sven J. Dickinson is Professor of Computer Science at the University of Toronto. From 1994 until 2000, he was Assistant Professor at Rutgers University, where he held joint appointments in the Department of Computer Science and the Rutgers Center for Cognitive Science. He was cochair of the 1997, 1999, 2004, and 2007 IEEE International Workshops on Generic Object Recognition and Categorization and cochair of the First International Workshop on Shape Perception in Human and Computer Vision in 2008.

Aleš Leonardis is a Full Professor and Head of the Visual Cognitive Systems Laboratory at the University of Ljubljana and Adjunct Professor at the Faculty of Computer Science, Graz University of Technology. He was a researcher and visiting professor at the University of Pennsylvania, Vienna University of Technology, Swiss Federal Institute of Technology, and University of Erlangen.

Bernt Schiele is Full Professor of Computer Science at TU Darmstadt, Germany. He obtained his Ph.D. from INPG in Grenoble, France. He was a researcher at Carnegie Mellon University and visiting Assistant Professor at Massachusetts Institute of Technology, as well as Assistant Professor at the Swiss Federal Institute of Technology in Zurich (ETH Zurich).

Michael J. Tarr is the Co-Director of the Center for the Neural Basis of Cognition and a Professor of Psychology at Carnegie Mellon University. From 1989 to 1995, he was an Assistant Professor of Psychology and Computer Science at Yale University. From 1995 to 2009, he was a Professor of Cognitive and Linguistic Sciences at Brown University. While at Brown he also served as the Co-Director of the Center for Vision Research.

Object Categorization

Computer and Human Vision Perspectives

Edited by

Sven J. Dickinson

University of Toronto

Aleš Leonardis

University of Ljubljana

Bernt Schiele

Darmstadt University of Technology

Michael J. Tarr

Carnegie Mellon University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521887380

First published 2009

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Object categorization : computer and human vision perspectives / edited
by Sven J. Dickinson . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-88738-0
1. Computer vision – Congresses. 2. Pattern recognition systems – Congresses.
I. Dickinson, Sven J. II. Title.
TA1634.O25 2009
006.3′7–dc22 2009026893

ISBN 978-0-521-88738-0 Hardback

Contents

Preface

The recognition of object categories has a rich history in computer vision. In the 1970s, generic object recognition systems sought to model and recognize objects based on their coarse, prototypical shape. These early systems employed complex 3-D models, which offered invariance to viewpoint (including image translation, rotation, and scale), articulation, occlusion, and minor within-class shape deformation. Despite powerful modeling paradigms, however, these early systems lacked the low- and intermediate-level segmentation, grouping, and abstraction machinery needed to recover prototypical shapes from real images of real objects. Over the next two decades, the recognition community began to back away from this "holy grail" of recognition, bringing new models closer to the image in an effort to reduce the representational gap between extractable image features and model features. During this time, the community migrated from the CAD-based vision era, in which exact 3-D geometry was specified, to the appearance-based vision era, in which exact 2-D photometry was specified (either globally, or locally at interest points). Almost in parallel, approaches to biological vision have followed a roughly similar path; that is, there has been a migration from CAD-inspired structural models comprised of 3-D parts, to image-based models preserving much of an object's input appearance, to, most recently, hybrid fragment-based models that rely on hierarchies of more localized image features.

Over this period, the recognition problem was sometimes reformulated from generic object recognition to exemplar recognition. For the first time, real object exemplars, with full texture and complex shape, could be recognized. However, it became apparent that these techniques for exemplar recognition did not scale up to generic objects (alternatively called classes or categories). Moreover, as is abundantly clear from the study of biological vision, the generic recognition of object categories is the predominant mode of how we interact with our visual environments. Thus, over the last decade, the mainstream object recognition pendulum has started to swing back toward object categorization. Armed with new features, new segmentation techniques, new optimization and matching techniques, new machine learning methods, and new a understanding of behavioral and neural phenomena, the community is far better prepared to tackle this important problem. Of course, because categorization was absent from the mainstream

for so long, there is a tendency not to look back at earlier problem formulations, challenges, and solutions. We feel that this historical disconnect has not served the community well, and, in many instances, we are facing today's challenges, including the quests for more categorical features (shape vs. appearance), viewpoint invariance, articulation invariance, occlusion invariance, and invariance to within-class structural change, without the clear hindsight of the community's earlier experience.

In an effort to foster greater communication between researchers from disparate disciplines, and to help bridge this historical disconnect, we organized international workshops on generic object recognition at many venues, including CVPR 97 (Ram Nevatia and Sven J. Dickinson, co-chairs), ICCV 99 (Gerard Medioni and Sven J. Dickinson, co-chairs), CVPR 04 (Aleš Leonardis, Bernt Schiele, and Sven J. Dickinson, co-chairs), and ICCV 2007 (Aleš Leonardis, Bernt Schiele, and Sven J. Dickinson, co-chairs). The workshops all had an identical format: bring together ten to twelve of the community's most prominent researchers, whose research spans the evolution of the field, to share their perspectives on the problem. Importantly, these researchers have been drawn from both the biological and computer vision communities with the idea that parallel problems and solutions have arisen in both domains. Moreover, we adhere to the integrative, multidisciplinary approach perhaps best articulated in the seminal work of David Marr. To stimulate discussion, we often purposely chose researchers with opposing viewpoints. We have found that beyond representing all perspectives of a problem, creating a forum for a diversity of views leads to more meaningful and more productive exchanges of ideas. As an added benefit, many of the workshop attendees were graduate students; thus, a broad treatment of the problem, with broad historical context, is particularly important. Speakers were encouraged not to simply present their latest work, but rather to provide a perspective on their experience working on the problem, and to talk about the challenges, successes, and failures. The workshops have been a great success, and attendance has been very high, in some cases outdrawing all other workshops at the conference!

To mark the tenth anniversary of the first such workshop, we decided to invite all the contributors from the four workshops to submit a chapter to what we hope will become a valuable collection of perspectives, from both human and computer vision, on the problem of object categorization. There are many reasons why we believe the time is right for such a collection. As mentioned, the historical disconnect continues to grow at the same time as more researchers enter the recognition community; an institutional memory refresh is especially important for today's researchers and students if we are to maximally benefit from the community's prior work. Perhaps a more compelling reason for assembling such a collection is a renewed interest from computer vision researchers in results originating from cognitive neuroscience, neuroscience, and psychology. In particular, the advent of functional neuroimaging, as well as new neurophysiological methods, has rejuvenated the study of object categorization in humans. Some of the best new researchers from these disciplines are represented in this volume. Even more promising is that there have been, of late, several successful algorithms that are biologically inspired or motivated. By bringing together researchers from different vision subcommunities, we hope to foster interdisciplinary awareness and collaboration, both of which will ultimately help to shed light on the problem of object categorization.

What you hold in your hands is a collection of twenty-seven chapters from some of the top human and computer vision researchers in the field of categorization. Some have worked on the problem for many years (decades) and have a unique perspective to offer researchers and students alike on what trends and issues have shaped the field, the progress we've made, and the challenges we face. Others have been in the field for a decade or less and offer fresh perspectives on old problems. Like our workshops, this volume is aimed at offering a unique, multidisciplinary view that strives to cover this important problem from all sides rather than promote a particular paradigm. Such a perspective is essential for new researchers attempting to understand the broader landscape of the problem so that they can build on a firm foundation. We hope you find the collection as exciting and as useful as we do.

When we sat down to organize the chapters by topic, we quickly found that many chapters defied categorization in that they addressed many topics. Our attempts to cluster chapters into sections led to uneven clusters of chapters, and decisions to put a chapter in one section versus another seemed rather arbitrary. There was also a tendency to cluster human vision chapters together and computer vision chapters together, which defeated our goal of bridging and integrating the two communities. As a result, we decided on a flat (sectionless) structure, with alternating human and computer vision chapters clustered by theme when appropriate. It is our hope that this lack of structure will avoid the biases associated with particular topics and encourage the reader to explore the unique contributions and perspectives offered by each chapter. Moreover, we hope that the interleaved format will naturally encourage human and computer vision researchers to explore each other's community.

Finally, there are a number of people we would like to thank for helping to make this volume possible. Ram Nevatia and Gerard Medioni co-chaired the first and second workshops, in 1997 and 1999, respectively. Heather Bergman, from Cambridge University Press, has been incredibly supportive of the volume and very patient with the editors, and her colleague, David Jou, has been extremely helpful on the editorial side. Mario Fritz, from the Darmstadt University of Technology, maintained a wonderful website for the collection and provided valuable technical support. We would also like to thank our sponsors for their generous financial support: ECVision, the European research network for cognitive computer vision systems, sponsored the workshop in 2004, and EuCognition, the European network for the advancement of artificial cognitive systems, and Toyota Europe sponsored the workshop in 2007. Our sincere thanks to you all.

Sven J. Dickinson
University of Toronto

Aleš Leonardis
University of Ljubljana

Bernt Schiele
Darmstadt University of Technology

Michael J. Tarr
Carnegie Mellon University

Contributors

Cecilia Ovesdotter Alm
University of Illinois at Urbana-Champaign

Benjamin Balas
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Moshe Bar
Martinos Center for Biomedical Imaging at MGH
Harvard Medical School

Marlene Behrmann
Department of Pyschology
Carnegie Mellon University

Tamara Berg
SUNY Stony Brook

Marko Boben
University of Ljubljana, Slovenia

Jasmine Boshyan
Martinos Center for Biomedical Imaging at MGH
Harvard Medical School

Kevin Bowyer
University of Notre Dame

Scott L. Brincat
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics

Charles E. Connor
Department of Neuroscience
Johns Hopkins University

James J. DiCarlo
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Sven J. Dickinson
University of Toronto

Shimon Edelman
Department of Psychology
Cornell University

Ali Farhadi
University of Illinois at Urbana-Champaign

Sanja Fidler
University of Ljubljana, Slovenia

D.A. Forsyth
University of Illinois at Urbana-Champaign

Mario Fritz
Darmstadt University of Technology

Kalanit Grill-Spector
Department of Pyschology and Neuroscience
Stanford University

Martial Hebert
Carnegie Mellon University

Julia Hockenmaier
University of Illinois at Urbana-Champaign

Donald D. Hoffman
University of California, Irvine

Kate Humphreys
Institute of Psychiatry

David W. Jacobs
University of Maryland

Benjamin B. Kimia
Division of Engineering
Brown University

Zoe Kourtzi
University of Birmingham

Kestutis Kveraga
Martinos Center for Biomedical Imaging at MGH
Harvard Medical School

Svetlana Lazebnik
University of North Carolina at Chapel Hill

Tai Sing Lee
Carnegie Mellon University

Aleš Leonardis
University of Ljubljana, Slovenia

Marius Leordeanu
Carnegie Mellon University

Nicolas Loeff
University of Illinois at Urbana-Champaign

Gérard Medioni
University of Southern California

Yuri Ostrovsky
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Anitha Pasupathy
Department of Biological Structure
University of Washington

Pietro Perona
California Institute of Technology

Stephen Pizer
University of North Carolina at Chapel Hill

Jean Ponce
LIENS, Ecole Normale Superieure

Jake Porway
University of California, Los Angeles

Brian Potetz
Carnegie Mellon University

Maximilian Riesenhuber
Department of Neuroscience
Georgetown University Medical Center

Edmund T. Rolls
Department of Experimental Psychology
University of Oxford

Jason Samonds
Carnegie Mellon University

K. Suzanne Scherf
Department of Pyschology
Carnegie Mellon University

Bernt Schiele
Darmstadt University of Technology

Cordelia Schmid
INRIA Grenoble

Kaleem Siddiqi
Centre for Intelligent Machines
McGill University

Pawan Sinha
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Louise Stark
University of the Pacific

Tom Stepleton
Carnegie Mellon University

Rahul Sukthankar
Intel Research, Pittsburgh

Melanie Sutton
University of West Florida

Shimon Ullman
Weizmann Institute of Science

Siavash Vaziri
Department of Biomedical Engineering
Johns Hopkins University

Christian Wallraven
MPI for Biological Cybernetics

Gang Wang
University of Illinois at Urbana-Champaign

Jonas Wulff
RWTH Aachen University, Germany

Benjamin Yao
University of California, Los Angeles

Song Chun Zhu
Lotus Hill Research Institute