MACHINE LEARNING IN COMPUTER VISION

Computational Imaging and Vision

Volume 29

Machine Learning in Computer Vision

by

N. SEBE

University of Amsterdam,
The Netherlands

IRA COHEN

HP Research Labs, U.S.A.

ASHUTOSH GARG

Google Inc., U.S.A.

and

THOMAS S. HUANG

University of Illinois at Urbana-Champaign,
Urbana, IL, U.S.A.

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3274-9 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-10 1-4020-3275-7 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-1-4020-3274-5 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-1-4020-3275-2 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands.

To my parents
Nicu

To Merav and Yonatan
Ira

To my parents
Asutosh

To my students:
Past, present, and future
Tom

Contents

Foreword

It started with *image processing* in the sixties. Back then, it took ages to digitize a Landsat image and then process it with a mainframe computer. Processing was inspired on the achievements of signal processing and was still very much oriented towards programming.

In the seventies, *image analysis* spun off combining image measurement with statistical pattern recognition. Slowly, computational methods detached themselves from the sensor and the goal to become more generally applicable.

In the eighties, model-driven *computer vision* originated when artificial intelligence and geometric modelling came together with image analysis components. The emphasis was on precise analysis with little or no interaction, still very much an art evaluated by visual appeal. The main bottleneck was in the amount of data using an average of 5 to 50 pictures to illustrate the point.

At the beginning of the nineties, vision became available to many with the advent of sufficiently fast PCs. The Internet revealed the interest of the general public im images, eventually introducing *content-based image retrieval*. Combining independent (informal) archives, as the web is, urges for interactive evaluation of approximate results and hence weak algorithms and their combination in weak classifiers.

In the new century, the last analog bastion was taken. In a few years, sensors have become all digital. Archives will soon follow. As a consequence of this change in the basic conditions datasets will overflow. Computer vision will spin off a new branch to be called something like *archive-based* or *semantic vision* including a role for formal knowledge description in an ontology equipped with detectors. An alternative view is *experience-based* or *cognitive vision*. This is mostly a data-driven view on vision and includes the elementary laws of image formation.

This book comes right on time. The general trend is easy to see. The methods of computation went from dedicated to one specific task to more generally applicable building blocks, from detailed attention to one aspect like filtering to a broad variety of topics, from a detailed model design evaluated against a few data to abstract rules tuned to a robust application.

From the source to consumption, images are now all digital. Very soon, archives will be overflowing. This is slightly worrying as it will raise the level of expectations about the accessibility of the pictorial content to a level compatible with what humans can achieve.

There is only one realistic chance to respond. From the trend displayed above, it is best to identify basic laws and then to learn the specifics of the model from a larger dataset. Rather than excluding interaction in the evaluation of the result, it is better to perceive interaction as a valuable source of instant learning for the algorithm.

This book builds on that insight: that the key element in the current revolution is the use of machine learning to capture the variations in visual appearance, rather than having the designer of the model accomplish this. As a bonus, models learned from large datasets are likely to be more robust and more realistic than the brittle all-design models.

This book recognizes that machine learning for computer vision is distinctively different from plain machine learning. Loads of data, spatial coherence, and the large variety of appearances, make computer vision a special challenge for the machine learning algorithms. Hence, the book does not waste itself on the complete spectrum of machine learning algorithms. Rather, this book is focussed on machine learning for pictures.

It is amazing so early in a new field that a book appears which connects theory to algorithms and through them to convincing applications.

The authors met one another at Urbana-Champaign and then dispersed over the world, apart from Thomas Huang who has been there forever. This book will surely be with us for quite some time to come.

Arnold Smeulders
University of Amsterdam
The Netherlands
October, 2004

Preface

The goal of computer vision research is to provide computers with human-like perception capabilities so that they can sense the environment, understand the sensed data, take appropriate actions, and learn from this experience in order to enhance future performance. The field has evolved from the application of classical pattern recognition and image processing methods to advanced techniques in image understanding like model-based and knowledge-based vision.

In recent years, there has been an increased demand for computer vision systems to address "real-world" problems. However, much of our current models and methodologies do not seem to scale out of limited "toy" domains. Therefore, the current state-of-the-art in computer vision needs significant advancements to deal with real-world applications, such as navigation, target recognition, manufacturing, photo interpretation, remote sensing, etc. It is widely understood that many of these applications require vision algorithms and systems to work under partial occlusion, possibly under high clutter, low contrast, and changing environmental conditions. This requires that the vision techniques should be robust and flexible to optimize performance in a given scenario.

The field of machine learning is driven by the idea that computer algorithms and systems can improve their own performance with time. Machine learning has evolved from the relatively "knowledge-free" general purpose learning system, the "perceptron" [Rosenblatt, 1958], and decision-theoretic approaches for learning [Blockeel and De Raedt, 1998], to symbolic learning of high-level knowledge [Michalski et al., 1986], artificial neural networks [Rowley et al., 1998a], and genetic algorithms [DeJong, 1988]. With the recent advances in hardware and software, a variety of practical applications of the machine learning research is emerging [Segre, 1992].

Vision provides interesting and challenging problems and a rich environment to advance the state-of-the art in machine learning. Machine learning technology has a strong potential to contribute to the development of flexible and robust vision algorithms, thus improving the performance of practical vision systems. Learning-based vision systems are expected to provide a higher level of competence and greater generality. Learning may allow us to use the experience gained in creating a vision system for one application domain to a vision system for another domain by developing systems that acquire and maintain knowledge. We claim that learning represents the next challenging frontier for computer vision research.

More specifically, machine learning offers effective methods for computer vision for automating the model/concept acquisition and updating processes, adapting task parameters and representations, and using experience for generating, verifying, and modifying hypotheses. Expanding this list of computer vision problems, we find that some of the applications of machine learning in computer vision are: segmentation and feature extraction; learning rules, relations, features, discriminant functions, and evaluation strategies; learning and refining visual models; indexing and recognition strategies; integration of vision modules and task-level learning; learning shape representation and surface reconstruction strategies; self-organizing algorithms for pattern learning; biologically motivated modeling of vision systems that learn; and parameter adaptation, and self-calibration of vision systems. As an eventual goal, machine learning may provide the necessary tools for synthesizing vision algorithms starting from adaptation of control parameters of vision algorithms and systems.

The goal of this book is to address the use of several important machine learning techniques into computer vision applications. An innovative combination of computer vision and machine learning techniques has the promise of advancing the field of computer vision, which will contribute to better understanding of complex real-world applications. There is another benefit of incorporating a learning paradigm in the computational vision framework. To mature the laboratory-grown vision systems into real-world working systems, it is necessary to evaluate the performance characteristics of these systems using a variety of real, calibrated data. Learning offers this evaluation tool, since no learning can take place without appropriate evaluation of the results.

Generally, learning requires large amounts of data and fast computational resources for its practical use. However, all learning does not have to be on-line. Some of the learning can be done off-line, e.g., optimizing parameters, features, and sensors during training to improve performance. Depending upon the domain of application, the large number of training samples needed for inductive learning techniques may not be available. Thus, learning techniques should be able to work with varying amounts of a priori knowledge and data.

The effective usage of machine learning technology in real-world computer vision problems requires understanding the domain of application, abstraction of a learning problem from a given computer vision task, and the selection of appropriate representations for the learnable (input) and learned (internal) entities of the system. To succeed in selecting the most appropriate machine learning technique(s) for the given computer vision task, an adequate understanding of the different machine learning paradigms is necessary.

A learning system has to clearly demonstrate and answer the questions like what is being learned, how it is learned, what data is used to learn, how to represent what has been learned, how well and how efficient is the learning taking place and what are the evaluation criteria for the task at hand. Experimental details are essential for demonstrating the learning behavior of algorithms and systems. These experiments need to include scientific experimental design methodology for training/testing, parametric studies, and measures of performance improvement with experience. Experiments that exihibit scalability of learning-based vision systems are also very important.

In this book, we address all these important aspects. In each of the chapters, we show how the literature has introduced the techniques into the particular topic area, we present the background theory, discuss comparative experiments made by us, and conclude with comments and recommendations.

Acknowledgments

This book would not have existed without the assistance of Marcelo Cirelo, Larry Chen, Fabio Cozman, Michael Lew, and Dan Roth whose technical contributions are directly reflected within the chapters. We would like to thank Theo Gevers, Nuria Oliver, Arnold Smeulders, and our colleagues from the Intelligent Sensory Information Systems group at University of Amsterdam and the IFP group at University of Illinois at Urbana-Champaign who gave us valuable suggestions and critical comments. Beyond technical contributions, we would like to thank our families for years of patience, support, and encouragement. Furthermore, we are grateful to our departments for providing an excellent scientific environment.