

# MACHINE LEARNING IN COMPUTER VISION

# Computational Imaging and Vision

---

Managing Editor

MAX A. VIERGEVER

*Utrecht University, Utrecht, The Netherlands*

Editorial Board

RUZENA BAJCSY, *University of Pennsylvania, Philadelphia, USA*

MIKE BRADY, *Oxford University, Oxford, UK*

OLIVIER D. FAUGERAS, *INRIA, Sophia-Antipolis, France*

JAN J. KOENDERINK, *Utrecht University, Utrecht, The Netherlands*

STEPHEN M. PIZER, *University of North Carolina, Chapel Hill, USA*

SABURO TSUJI, *Wakayama University, Wakayama, Japan*

STEVEN W. ZUCKER, *McGill University, Montreal, Canada*

# 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*

All Rights Reserved

© 2005 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

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	xi
Preface	xiii
1. INTRODUCTION	1
1    Research Issues on Learning in Computer Vision	2
2    Overview of the Book	6
3    Contributions	12
2. THEORY: PROBABILISTIC CLASSIFIERS	15
1    Introduction	15
2    Preliminaries and Notations	18
2.1    Maximum Likelihood Classification	18
2.2    Information Theory	19
2.3    Inequalities	20
3    Bayes Optimal Error and Entropy	20
4    Analysis of Classification Error of Estimated ( <i>Mismatched</i> ) Distribution	27
4.1    Hypothesis Testing Framework	28
4.2    Classification Framework	30
5    Density of Distributions	31
5.1    Distributional Density	33
5.2    Relating to Classification Error	37
6    Complex Probabilistic Models and Small Sample Effects	40
7    Summary	41

3.	THEORY:	
	GENERALIZATION BOUNDS	45
1	Introduction	45
2	Preliminaries	47
3	A Margin Distribution Based Bound	49
	3.1 Proving the Margin Distribution Bound	49
4	Analysis	57
	4.1 Comparison with Existing Bounds	59
5	Summary	64
4.	THEORY:	
	SEMI-SUPERVISED LEARNING	65
1	Introduction	65
2	Properties of Classification	67
3	Existing Literature	68
4	Semi-supervised Learning Using Maximum Likelihood Estimation	70
5	Asymptotic Properties of Maximum Likelihood Estimation with Labeled and Unlabeled Data	73
	5.1 Model Is Correct	76
	5.2 Model Is Incorrect	77
	5.3 Examples: Unlabeled Data Degrading Performance with Discrete and Continuous Variables	80
	5.4 Generating Examples: Performance Degradation with Univariate Distributions	83
	5.5 Distribution of Asymptotic Classification Error Bias	86
	5.6 Short Summary	88
6	Learning with Finite Data	90
	6.1 Experiments with Artificial Data	91
	6.2 Can Unlabeled Data Help with Incorrect Models? Bias vs. Variance Effects and the Labeled-unlabeled Graphs	92
	6.3 Detecting When Unlabeled Data Do Not Change the Estimates	97
	6.4 Using Unlabeled Data to Detect Incorrect Modeling Assumptions	99
7	Concluding Remarks	100

5. ALGORITHM: MAXIMUM LIKELIHOOD MINIMUM ENTROPY HMM	103
1 Previous Work	103
2 Mutual Information, Bayes Optimal Error, Entropy, and Conditional Probability	105
3 Maximum Mutual Information HMMs	107
3.1 Discrete Maximum Mutual Information HMMs	108
3.2 Continuous Maximum Mutual Information HMMs	110
3.3 Unsupervised Case	111
4 Discussion	111
4.1 Convexity	111
4.2 Convergence	112
4.3 Maximum A-posteriori View of Maximum Mutual Information HMMs	112
5 Experimental Results	115
5.1 Synthetic Discrete Supervised Data	115
5.2 Speaker Detection	115
5.3 Protein Data	117
5.4 Real-time Emotion Data	117
6 Summary	117
6. ALGORITHM: MARGIN DISTRIBUTION OPTIMIZATION	119
1 Introduction	119
2 A Margin Distribution Based Bound	120
3 Existing Learning Algorithms	121
4 The Margin Distribution Optimization (MDO) Algorithm	125
4.1 Comparison with SVM and Boosting	126
4.2 Computational Issues	126
5 Experimental Evaluation	127
6 Conclusions	128
7. ALGORITHM: LEARNING THE STRUCTURE OF BAYESIAN NETWORK CLASSIFIERS	129
1 Introduction	129
2 Bayesian Network Classifiers	130
2.1 Naive Bayes Classifiers	132
2.2 Tree-Augmented Naive Bayes Classifiers	133



3	Switching between Models: Naive Bayes and TAN Classifiers	138
4	Learning the Structure of Bayesian Network Classifiers: Existing Approaches	140
4.1	Independence-based Methods	140
4.2	Likelihood and Bayesian Score-based Methods	142
5	Classification Driven Stochastic Structure Search	143
5.1	Stochastic Structure Search Algorithm	143
5.2	Adding VC Bound Factor to the Empirical Error Measure	145
6	Experiments	146
6.1	Results with Labeled Data	146
6.2	Results with Labeled and Unlabeled Data	147
7	Should Unlabeled Data Be Weighed Differently?	150
8	Active Learning	151
9	Concluding Remarks	153
8.	APPLICATION: OFFICE ACTIVITY RECOGNITION	157
1	Context-Sensitive Systems	157
2	Towards Tractable and Robust Context Sensing	159
3	Layered Hidden Markov Models (LHMMs)	160
3.1	Approaches	161
3.2	Decomposition per Temporal Granularity	162
4	Implementation of SEER	164
4.1	Feature Extraction and Selection in SEER	164
4.2	Architecture of SEER	165
4.3	Learning in SEER	166
4.4	Classification in SEER	166
5	Experiments	166
5.1	Discussion	169
6	Related Representations	170
7	Summary	172
9.	APPLICATION: MULTIMODAL EVENT DETECTION	175
1	Fusion Models: A Review	176
2	A Hierarchical Fusion Model	177
2.1	Working of the Model	178
2.2	The Duration Dependent Input Output Markov Model	179

<i>Contents</i>	ix
3 Experimental Setup, Features, and Results	182
4 Summary	183
10. APPLICATION: FACIAL EXPRESSION RECOGNITION	187
1 Introduction	187
2 Human Emotion Research	189
2.1 Affective Human-computer Interaction	189
2.2 Theories of Emotion	190
2.3 Facial Expression Recognition Studies	192
3 Facial Expression Recognition System	197
3.1 Face Tracking and Feature Extraction	197
3.2 Bayesian Network Classifiers: Learning the “Structure” of the Facial Features	200
4 Experimental Analysis	201
4.1 Experimental Results with Labeled Data	204
4.1.1 Person-dependent Tests	205
4.1.2 Person-independent Tests	206
4.2 Experiments with Labeled and Unlabeled Data	207
5 Discussion	208
11. APPLICATION: BAYESIAN NETWORK CLASSIFIERS FOR FACE DETECTION	211
1 Introduction	211
2 Related Work	213
3 Applying Bayesian Network Classifiers to Face Detection	217
4 Experiments	218
5 Discussion	222
References	225
Index	237

## 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 in 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 exhibit 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.