Extreme Value Theory-Based Methods for Visual Recognition

Synthesis Lectures on Computer Vision

Editors
Gérard Medioni, ***University of Southern California***
Sven Dickinson, ***University of Toronto***

Synthesis Lectures on Computer Vision is edited by Gérard Medioni of the University of Southern California and Sven Dickinson of the University of Toronto. The series publishes 50- to 150-page publications on topics pertaining to computer vision and pattern recognition. The scope will largely follow the purview of premier computer science conferences, such as ICCV, CVPR, and ECCV. Potential topics include, but not are limited to:

- Applications and Case Studies for Computer Vision
- Color, Illumination, and Texture
- Computational Photography and Video
- Early and Biologically-inspired Vision
- Face and Gesture Analysis
- Illumination and Reflectance Modeling
- Image-Based Modeling
- Image and Video Retrieval
- Medical Image Analysis
- Motion and Tracking
- Object Detection, Recognition, and Categorization
- Segmentation and Grouping
- Sensors
- Shape-from-X
- Stereo and Structure from Motion
- Shape Representation and Matching

- Statistical Methods and Learning
- Performance Evaluation
- Video Analysis and Event Recognition

Extreme Value Theory-Based Methods for Visual Recognition
Walter J. Scheirer
2017

Data Association for Multi-Object Visual Tracking
Margrit Betke and Zheng Wu
2016

Ellipse Fitting for Computer Vision: Implementation and Applications
Kenichi Kanatani, Yasuyuki Sugaya, and Yasushi Kanazawa
2016

Computational Methods for Integrating Vision and Language
Kobus Barnard
2016

Background Subtraction: Theory and Practice
Ahmed Elgammal
2014

Vision-Based Interaction
Matthew Turk and Gang Hua
2013

Camera Networks: The Acquisition and Analysis of Videos over Wide Areas
Amit K. Roy-Chowdhury and Bi Song
2012

Deformable Surface 3D Reconstruction from Monocular Images
Mathieu Salzmann and Pascal Fua
2010

Boosting-Based Face Detection and Adaptation
Cha Zhang and Zhengyou Zhang
2010

Image-Based Modeling of Plants and Trees
Sing Bing Kang and Long Quan
2009

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Extreme Value Theory-Based Methods for Visual Recognition

Walter J. Scheirer

ISBN: 978-3-031-00689-0 paperback
ISBN: 978-3-031-01817-6 ebook

DOI 10.1007/978-3-031-01817-6

A Publication in the Springer series

SYNTHESIS LECTURES ON COMPUTER VISION

Lecture #10
Series Editors: Gérard Medioni, *University of Southern California*
Sven Dickinson, *University of Toronto*
Series ISSN
Print 2153-1056 Electronic 2153-1064

Extreme Value Theory-Based Methods for Visual Recognition

Walter J. Scheirer
University of Notre Dame

SYNTHESIS LECTURES ON COMPUTER VISION #10

ABSTRACT

A common feature of many approaches to modeling sensory statistics is an emphasis on capturing the "average." From early representations in the brain, to highly abstracted class categories in machine learning for classification tasks, central-tendency models based on the Gaussian distribution are a seemingly natural and obvious choice for modeling sensory data. However, insights from neuroscience, psychology, and computer vision suggest an alternate strategy: preferentially focusing representational resources on the extremes of the distribution of sensory inputs. The notion of treating extrema near a decision boundary as features is not necessarily new, but a comprehensive statistical theory of recognition based on extrema is only now just emerging in the computer vision literature. This book begins by introducing the statistical Extreme Value Theory (EVT) for visual recognition. In contrast to central-tendency modeling, it is hypothesized that distributions near decision boundaries form a more powerful model for recognition tasks by focusing coding resources on data that are arguably the most diagnostic features. EVT has several important properties: strong statistical grounding, better modeling accuracy near decision boundaries than Gaussian modeling, the ability to model asymmetric decision boundaries, and accurate prediction of the probability of an event beyond our experience. The second part of the book uses the theory to describe a new class of machine learning algorithms for decision making that are a measurable advance beyond the state-of-the-art. This includes methods for post-recognition score analysis, information fusion, multi-attribute spaces, and calibration of supervised machine learning algorithms.

KEYWORDS

visual recognition, extreme value theory, machine learning, statistical methods, decision making, failure prediction, information fusion, score normalization, open set recognition, object recognition, information retrieval, biometrics, deep learning

Contents

Preface

The choice of a probability distribution can have a profound effect on the results coming from a model for an underlying analysis task—and not always in a good way. Consider the case of the 2008 financial crisis. Many different quantitative methods are deployed by economists and financial analysts to gauge the health of the economy. Some of these methods attempt to model the overall composition of a market sector, while others look at specific circumstances that may impact the market at large. A central tendency model would lead to an understanding of average market forces, and would have good predictive power when significant market forces move collectively in a certain direction (perhaps toward a bull or bear market).

An analyst using such a model back in 2007 would have had a rosy picture of the economy—a correction as large as the looming financial crisis was a seemingly improbable event. What eventually brought the financial system to the brink was a series of *extreme* market movements as firms holding mortgage-backed securities collapsed. An alternative modeling strategy would have focused not on average movement in the market, but on the tails of distributions representing specific market returns. Under such a model, where the extrema contribute to the model in a meaningful way, the financial crisis was well within the realm of possibility.

It is not an enormous leap of faith to believe that the same phenomenon occurs in computer vision. Most often we find that the extrema (e.g., edges, attributes, parts, salient objects) in a scene contained within a digital image define visual appearance, and not the average pixel content (*e.g.*, background). This calls for statistical modeling that does not deemphasize or ignore the rare facets of images or the features extracted from them. However in practice, this is not what we find. Remarkably, the field of computer vision has maintained a steady fixation with central tendency modeling, in spite of the complex nature of the underlying statistics of natural scenes. Extrema may be rare, but their influence is more often than not considerable.

These observations lead us to the topic at hand: the statistical extreme value theory. From predicting floods to downturns in the market, extreme value theory is a workhorse of predictive modeling in many fields outside of computer science. However, we are just starting to see its emergence within computer vision—an exciting and much welcomed development. Admittedly, there is some safety in central tendency modeling, as it allows one to invoke the central limit theorem, and assume that the normal distribution applies in approximation. But as we shall see, the first extreme value theorem functions in much the same way, and gives us access to a number of limiting distributions that apply in the tails of overall distributions, regardless of form. Given such flexibility, researchers within computer vision may find the extreme value theory becoming an indispensable part of their statistics toolkit once they get a feel for it.

This book is a summary of close to a decade of research into the application of the statistical extreme value theory to visual recognition. Unlike various references found in the statistics literature, it is intended to be a practical introductory guide to extreme value theory-based algorithms, and thus eschews proofs and other non-essential formalities. The interested reader is encouraged to dig deeper into the cited papers for that material as necessary. Further, this book can be read as a companion volume to the "Statistical Methods for Open Set Recognition" tutorial that was presented at CVPR 2016 in Las Vegas. The material from that tutorial, including slides and code, is available at the following URL: `http://www.wjscheirer.com/misc/openset/`. While this book represents a milestone of sorts for a budding research area within computer vision, we are sure to see even more intriguing work in this direction in the coming years.

Walter J. Scheirer
January 2017

Acknowledgments

I would like to extend my thanks to Gerard Medioni and Sven Dickinson, the editors of this Synthesis Lectures on Computer Vision series, for inviting me to contribute. I would also like to thank all of the researchers who granted me permission to use their figures in this book. Without their contribution, this volume would not have turned out as well as it did. I am also indebted to my editor, Diane Cerra, for her guidance and patience during the preparation of this manuscript. A group of expert reviewers from the field of machine learning provided helpful feedback that led to the polished version of the book that you have in your hands today. I am very grateful for their input. Finally, the idea for applying extreme value theory to recognition problems in computer vision came out of work undertaken at the University of Colorado, Colorado Springs and Securics, Inc. during the summer of 2008. The core group from that era of Terry Boult, Anderson Rocha, and Abhi Bendale remains instrumental to this research area—thanks for the great work and fun over the years, guys!

Walter J. Scheirer
January 2017

Figure Credits

Figure 1.2 (far left) Based on: J. W. Tanaka and O. Corneille. Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & Psychophysics*, 69(4):619–627, May 2007.

Figure 1.2 (second from left) Based on: D. Leopold, I. Bondar, and M. Giese. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442:572–575, August 2006.

Figure 1.2 (middle) Based on: J. W. Tanaka and M. J. Farah. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 46:225–245, 1993.

Figure 1.2 (second from right) From: L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, February 2001. Copyright © 2001 Nature Publishing Group. Used with permission.

Figure 1.2 (far right) From: E. Barenholtz and M. Tarr. Visual judgment of similarity across shape transformations: Evidence for a compositional model of articulated objects. *Acta Psychologica*, 128:331–338. Copyright © 2008 Elsevier. Used with permission.

Figure 3.1 From: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689–1695. Copyright © 2011 IEEE. Used with permission.

Figure 3.2 Based on: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689-1695, August 2011.

Figure 3.3 Based on: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689-1695, August 2011.

Figure 3.4	Based on: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689-1695, August 2011.
Figure 3.5a	Based on: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1689-1695, August 2011.
Figure 3.6	Based on: V. Fragoso and M. Turk. SWIGS: A swift guided sampling method. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* June 2013.
Figure 4.1	From: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Robust fusion: Extreme value theory for recognition score normalization. In *European Conference on Computer Vision (ECCV)*. Copyright © 2010 Springer-Verlag Berlin Heidelberg. Used with permission.
Figure 4.2	Based on: W. J. Scheirer, A. Rocha, R. Michaels, and T. E. Boult. Robust fusion: Extreme value theory for recognition score normalization. In *European Conference on Computer Vision (ECCV)*, September 2010.
Figure 4.3	From: V. Fragoso, P. Sen, S. Rodriguez, and M. Turk. EVSAC: Accelerating hypotheses generation by modeling matching scores with extreme value theory. In *IEEE International Conference on Computer Vision (ICCV).* Copyright © 2013 IEEE. Used with permission.
Figure 4.4	From: X. Gibert, V. Patel, and R. Chellappa. Sequential score adaptation with extreme value theory for robust railway track inspection. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. Copyright © 2015 IEEE. Used with permission.
Figure 5.1	Based on: W. J. Scheirer, N. Kumar, P. N. Belhumeur, and T. E. Boult. Multi-attribute spaces: Calibration for attribute fusion and similarity search. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* June 2012a.
Figure 5.2	From: W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35(7):1757–1772. Copyright © 2013 IEEE. Used with permission.

Figure 5.3	From: L. P. Jain, W. J. Scheirer, and T. E. Boult. Multi-class open set recognition using probability of inclusion. In *European Conference on Computer Vision (ECCV)*. Copyright © 2014 Springer. Used with permission.
Figure 5.7	From: A. Bendale and T. E. Boult. Towards open set deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Copyright © 2016 IEEE. Used with permission.