Probabilistic and Biologically Inspired Feature Representations

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 50–150 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

Probabilistic and Biologically Inspired Feature Representations
Michael Felsberg
2018

A Guide Convolutional Neural Networks for Computer Vision
Salman Khan, Hossein Rahmani, Syed Afaq, Ali Shah, and Mohammed Bennamoun
2018

Covariances in Computer Vision and Machine Learning
Hà Quang Minh and Vittorio Murino
2017

Elastic Shape Analysis of Three-Dimensional Objects
Ian H. Jermyn, Sebastian Kurtek, Hamid Laga, and Anuj Srivastava
2017

The Maximum Consensus Problem: Recent Algorithmic Advances
Tat-Jun Chin and David Suter
2017

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

Probabilistic and Biologically Inspired Feature Representations

Michael Felsberg

ISBN: 978-3-031-00694-4 paperback
ISBN: 978-3-031-01822-0 ebook
ISBN: 978-3-031-00079-9 hardcover

DOI 10.1007/978-3-031-01822-0

A Publication in the Springer series
SYNTHESIS LECTURES ON COMPUTER VISION

Lecture #16
Series Editors: Gérard Medioni, *University of Southern California*
Sven Dickinson *University of Toronto*
Series ISSN
Print 2153-1056 Electronic 2153-1064

Probabilistic and Biologically Inspired Feature Representations

Michael Felsberg
Linköping University

SYNTHESIS LECTURES ON COMPUTER VISION #16

ABSTRACT

Under the title *"Probabilistic and Biologically Inspired Feature Representations,"* this text collects a substantial amount of work on the topic of channel representations. Channel representations are a biologically motivated, wavelet-like approach to visual feature descriptors: they are local and compact, they form a computational framework, and the represented information can be reconstructed. The first property is shared with many histogram- and signature-based descriptors, the latter property with the related concept of population codes. In their unique combination of properties, channel representations become a visual Swiss army knife—they can be used for image enhancement, visual object tracking, as 2D and 3D descriptors, and for pose estimation. In the chapters of this text, the framework of channel representations will be introduced and its attributes will be elaborated, as well as further insight into its probabilistic modeling and algorithmic implementation will be given. Channel representations are a useful toolbox to represent visual information for machine learning, as they establish a generic way to compute popular descriptors such as HOG, SIFT, and SHOT. Even in an age of deep learning, they provide a good compromise between hand-designed descriptors and a-priori structureless feature spaces as seen in the layers of deep networks.

KEYWORDS

channel representation, channel-coded feature map, feature descriptor, signature, histogram

Contents

Preface

This book comes during a deep learning revolution in computer vision, when performance of, e.g., object classification on ImageNet [Russakovsky et al., 2014] has improved vastly from top-5 error of 26% in 2011 to 16% in 2012. The major paradigm shift has been to move from engineered image features ("pixel f***ing" according to Koenderink and van Doorn [2002]) to learned deep features. So why write this text now? Many recent publications making use of deep learning show a lack of rigor and their way to *throw data at the problem* is unsatisfying from a theoretical perspective. Attempts to put deep networks into established frameworks as done by Mallat [2016] are essential contributions to the field. Deep learning is very important from a practical perspective, but having a well-founded understanding of the underlying features and how they relate to common approaches can only help whether you are using deep learning or engineered features. Indeed, there might be a twist to use channel representations inside of deep networks, but there are further motivations to write this text. One reason is to honor the work by Gösta Granlund, the father of channel representations, who recently finished his academic career. A second motivation is to summarize one of my own branches of research, as I have been working on feature representations since my master's thesis 20 years ago (Felsberg [1998]). Last but not least, this text addresses many mathematical and algorithmic concepts that are useful to know and thus I want to share with students, colleagues, and practitioners. None of those groups of people is addressed exclusively and presumably none will see this as a primary source of information, but I hope that all will find new aspects and try to formulate new research questions as a consequence.

Michael Felsberg
April 2018

Acknowledgments

My work and the topic of this book have been constructively influenced by many colleagues, but primarily I would like to name Gösta Granlund, former head and founder of the Computer Vision Laboratory at Linköping University; Gerald Sommer, my Ph.D. supervisor; my colleagues at CVL who contributed to this book's content in some way or another, Per-Erik Forssén, Reiner Lenz, Klas Nordberg and my Ph.D. students Erik, Fredrik, Johan, Kristoffer, and Martin; and my colleagues outside of CVL who contributed, Remco Duits, Hanno Scharr, Ullrich Köthe, Rudolf Mester, Kai Krajsek, Norbert Krüger, and Richard Bowden.

Besides fruitful scientific exchange with colleagues, research also requires funding, and the collected work here has been financed by a long list of projects. Since we have ongoing projects with all founding agencies that supported our earlier work, I simply name the respective still ongoing projects. This research was partly supported by:

- the Swedish Research Council through a framework grant for the project Energy Minimization for Computational Cameras (2014-6227);
- the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation;
- ELLIIT, the Strategic Area for ICT research, funded by the Swedish Government;
- the Swedish Foundation for Strategic Research (Smart Systems: RIT 15-0097);
- the EC's Horizon 2020 Programme, grant agreement CENTAURO; and
- Vinnova through the grant CYCLA.

Michael Felsberg
April 2018