Lecture Notes in Computer Science 8408

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Jia Li Wen Gao

Visual Saliency Computation

A Machine Learning Perspective

Springer

Authors

Jia Li
Peking University, Haidian District, Beijing, 100871, China
E-mail: jia.li@pku.edu.cn

Wen Gao
Peking University, Haidian District, Beijing, 100871, China
E-mail: wgao@pku.edu.cn

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-319-05641-8 e-ISBN 978-3-319-05642-5
DOI 10.1007/978-3-319-05642-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014934972

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

There are two sides to every story. The same might be said about visual attention. On one hand, due to our limited perceptual and cognitive resources, we have to focus our attention to the most pertinent subset of the available sensory data. On the other hand, we are active seekers of visual information rather than passive recipients, and thus our human vision system often operationalizes attention as a selection mechanism responsible for filtering out unwanted information in a visual scene.

Closely related to attention is the concept of saliency. The saliency of an item - be it an object, a person, a pixel, etc. - arises from its contrast relative to its neighbors. In addition, saliency can be influenced by training: for example, particular letters can become salient for human subjects by training. Saliency detection is considered to be a key attentional mechanism, which guides visual attention. Although attention and saliency have been intensely studied, what we know is still far less than what we do not know. I am, however, optimistic that active investigation in psychology, cognitive neuroscience, and computer science will help reveal more and more about the inner workings of human vision systems.

Visual attention has played an essential role in our biological survival through the development of humanity. It will still play an essential role in this coming Big Data age. As is reported, it will take an individual over half a million years to watch all the videos that currently cross the network each month and the number will increase to 5 million years in 2017. Over 300 million new photos are added to Facebook every day. To "survive" in this era, we have to overcome a number of challenges. How can we efficiently process the massive amount of visual data with limited computational resources? What algorithms do we develop to extract pertinent information to succinctly represent the visual data? How can we model each individual's interest and provide information only relevant to that individual? Visual saliency computation, which measures the importance of various visual subsets in an image or video, is key to addressing these challenges.

This monograph, written by Dr. Jia Li and Prof. Wen Gao, is timely. It uniquely approaches visual saliency computation from a machine learning perspective. Instead of directly simulating the "known" mechanisms of the human brain, they propose to incorporate the modern machine learning algorithms to automatically mine the probable saliency mechanisms from user data. In this process, the prior knowledge, which is believed to be stored in the higher regions of the human brain, can be effectively and efficiently modeled to guide the computation of visual saliency. A multi-task learning technique is developed to infer what a human subject may attend to in an incoming scene by analyzing the users' activities when watching similar scenes in the past. Moreover, they present a statistical learning approach to infer such priors from millions of images in an unsupervised manner. In particular, they have discovered that acquiring the ordering of various visual subsets is more crucial than obtaining their real saliency values. Accordingly, they propose a learning-to-rank approach and formulate the problem of saliency computation within a ranking framework. Subsequently, they examine the properties of salient targets and describe how to extract a salient object from an image as a whole. Extensive experiments are conducted, and the results demonstrate that learning-based approaches have superior performance compared to traditional saliency models.

The book is a delight to read. The authors have put tremendous effort into making it clear and precise. A reader can start from the very basic question of "What is visual saliency?" and progressively explore the problems in modeling saliency, extracting salient objects, mining prior knowledge, evaluating performance, and using saliency in real-world applications. This timely book will be very valuable to researchers and practitioners who are interested in visual information processing and multimedia.

December 2013 Zhengyou Zhang

Acknowledgements

This monograph is based on Dr. Jia Li's doctoral dissertation, under the supervision of Prof. Wen Gao. This book is translated from the dissertation and extended with new content during the two authors' attachment to Peking University.

Many of our colleagues provided us with valueable assistance during the writing of doctor dissertation as well as this edition, which includes valuable material, dataset, suggestions, feedbacks, and comments. We want to thank Dr. Yuanning Li, Dr. Jingjing Yang, Prof. Shuqiang Jiang, Ms. Shu Fang, Ms. Guanghui Cai, Ms. Dan Li and Mr. Xiaobiao Wang for their kind help. A special thank to Prof. Yonghong Tian at Peking University, the leader of the multimedia learning research group at the National Engineering Laboratory on Video Technology at Peking University, who was in the advisory committee on Dr. Li's Ph.D. dissertation, and made many contributions on research and publication. We want to give our thanks to Prof. Fang Fang, Prof. Tiejun Huang and Prof. Yizhou Wang at Peking University, Prof. Qingming Huang at University of Chinese Academy of Sciences, Prof. Xilin Chen at Institute of Computing Technology, Chinese Academy of Sciences, they read the manuscript and gave many comments and suggestions which helped us to improve this book. We are also grateful to Prof. Alex Kot who provided a prosperous environment for writing the book during Dr. Li's visiting to the ROSE lab at Nanyang Technological University, Singapore.

Finally, we appreciate the financial support from the research funds. This work is supported in part by grants from the Chinese National Natural Science Foundation under contract No. 61370113 and No. 61035001, National Basic Research Program of China (No. 2009CB320900) and the Supervisor Award Funding for Excellent Doctoral Dissertation of Beijing (No. 20128000103). This research is also partially supported by the Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by the Interactive & Digital Media Programme Office, Media Development Authority.

Contents