

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

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

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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 valuable 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

1	Introduction	1
1.1	What Is Visual Saliency	1
1.2	Visual Saliency in Human Vision System	3
1.2.1	Basic Concepts	4
1.2.2	Neurobiological Mechanisms of Human Vision	6
1.2.3	Psychological Explanations of Visual Saliency	13
1.2.4	Summary	17
1.3	How to Compute Visual Saliency	18
1.4	Structure of the Book	20
	References	20
2	Benchmark and Evaluation Metrics	23
2.1	Overview	23
2.2	Benchmarks	24
2.2.1	Image Benchmark	24
2.2.2	Video Benchmark	29
2.3	Evaluation Metrics	31
2.3.1	Receiver Operating Characteristic	31
2.3.2	Normalized Scanpath Saliency	34
2.3.3	Kullback-Leibler Divergence	35
2.3.4	Recall and Precision	36
2.3.5	Other Metrics	37
2.3.6	Summary	39
2.4	Notes	39
	References	40
3	Location-Based Visual Saliency Computation	45
3.1	Overview	45
3.2	Saliency Models in Spatial Domain	46
3.2.1	Saliency from Local Irregularity	47
3.2.2	Saliency from Global Irregularity	52

3.3	Saliency Models in Transform Domain	56
3.3.1	Saliency from Frequency Analysis	57
3.3.2	Saliency from Subspace Analysis	59
3.4	Saliency Models in Spatiotemporal Domain	64
3.4.1	Saliency from Surprise	64
3.4.2	Saliency from Spatiotemporal Irregularity	67
3.5	Notes	68
	References	69
4	Object-Based Visual Saliency Computation	73
4.1	Overview	73
4.2	Salient Object Extraction from Location-Based Saliency Map	74
4.2.1	Frequency-Tuned Saliency	75
4.2.2	Complementary Saliency	77
4.3	Object-Based Salient Region Detection	83
4.3.1	Regional Contrasts	84
4.3.2	Context and Shape Priors	86
4.3.3	Saliency Filters	89
4.3.4	Single Image Optimization	92
4.4	Model Comparison	96
4.5	Notes	98
	References	99
5	Learning-Based Visual Saliency Computation	101
5.1	Overview	101
5.2	Task-Driven Saliency Model	103
5.2.1	The Bottom-Up Component	105
5.2.2	The Probabilistic Multi-task Learning Framework	109
5.2.3	Experiments	116
5.3	Knowledge-Based Saliency Model	127
5.3.1	The Bayesian Formulation	129
5.3.2	Saliency with Statistical Priors	130
5.3.3	Experiments	139
5.4	Notes	146
	References	147
6	Mining Cluster-Specific Knowledge for Saliency Ranking	151
6.1	Overview	151
6.2	Multi-task Learning to Rank for Visual Saliency Computation	152
6.2.1	Introduction	152
6.2.2	Problem Formulation	154
6.2.3	The Multi-task Learning to Rank Approach	155

6.2.4	Experiments	163
6.3	Notes	176
	References	177
7	Removing Label Ambiguity in Training Saliency	
	Model	179
7.1	Overview	179
7.2	A Cost-Sensitive Learning to Rank Approach	181
7.2.1	The Approach	181
7.2.2	Experiments	184
7.3	A Multi-instance Learning to Rank Approach	187
7.3.1	Problem Formulation	188
7.3.2	The Learning Algorithm	190
7.3.3	Experiments	198
7.4	Notes	211
	References	211
8	Saliency-Based Applications	215
8.1	Overview	215
8.2	Retargeting	216
8.2.1	Image Retargeting	216
8.2.2	Video Retargeting	217
8.3	Advertising	221
8.3.1	Linear Advertising	223
8.3.2	Non-linear Advertising	224
8.4	Retrieval	225
8.5	Summarization	227
8.6	Compression	229
8.7	Notes	230
	References	231
9	Conclusions and Future Work	233
9.1	Contribution Summarization	233
9.2	Future Work	235
	References	237
Index		239