

# Hypergraph Modeling for Saliency Detection and Beyond



Yao Li

School of Computer Science

The University of Adelaide

A thesis submitted for the degree of

*Master of Engineering Science*

20/12/2013

# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Figures</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Saliency detection . . . . .	1
1.2 Scene text detection . . . . .	2
1.3 Overview of contributions . . . . .	4
1.4 Outline . . . . .	6
<b>2 Background</b>	<b>8</b>
2.1 Saliency detection . . . . .	8
2.1.1 Local approaches . . . . .	9
2.1.2 Global approaches . . . . .	9
2.2 Scene text detection . . . . .	10
2.2.1 Texture-based approaches . . . . .	10
2.2.2 Region-based approaches . . . . .	12
2.2.3 Scene text detection aided by saliency . . . . .	14
<b>3 Contextual hypergraph modeling for salient object detection</b>	<b>15</b>
3.1 Cost-sensitive SVM saliency detection . . . . .	15
3.2 Hypergraph modeling for saliency detection . . . . .	17
3.2.1 Hypergraph modeling . . . . .	18
3.2.2 Adaptive hyperedge construction . . . . .	20
3.2.3 Hyperedge saliency evaluation . . . . .	21

3.3	Saliency fusion . . . . .	23
3.4	Experiments . . . . .	23
3.4.1	Experimental setup . . . . .	23
3.4.2	Evaluation of individual approaches . . . . .	25
3.4.3	Evaluation of different parameter settings . . . . .	26
3.4.4	Evaluation of saliency fusion . . . . .	27
3.4.5	Comparison with other approaches . . . . .	29
3.4.6	Application to image retargeting . . . . .	29
<b>4</b>	<b>Characterness: an indicator of text in the wild</b>	<b>37</b>
4.1	Characterness model . . . . .	37
4.1.1	Candidate region extraction . . . . .	37
4.1.2	Characterness evaluation . . . . .	39
4.1.2.1	Characterness cues . . . . .	39
4.1.2.2	Bayesian multi-cue integration . . . . .	42
4.2	Labeling and grouping . . . . .	43
4.2.1	Character labeling . . . . .	43
4.2.1.1	Labeling model overview . . . . .	43
4.2.1.2	The design of unary potential . . . . .	44
4.2.1.3	The design of pairwise potential . . . . .	44
4.2.2	Text line formulation . . . . .	45
4.3	Evaluation of the characterness model . . . . .	46
4.3.1	Experimental setup . . . . .	47
4.3.2	Comparison with other approaches . . . . .	48
4.4	Evaluation of the proposed scene text detection approach . . . . .	49
4.4.1	Experimental setup . . . . .	49
4.4.2	eMSER versus MSER . . . . .	51
4.4.3	Evaluation of Bayesian multi-cue integration . . . . .	51
4.4.4	Evaluation of characterness cues . . . . .	52
4.4.5	Comparison with other approaches . . . . .	53
<b>5</b>	<b>Conclusions</b>	<b>61</b>
	<b>References</b>	<b>63</b>

## Abstract

Salient object detection aims to locate objects that capture human attention within images. Previous approaches often pose this as a problem of image contrast analysis. In this work, we model an image as a hypergraph that utilizes a set of hyperedges to capture the contextual properties of image pixels or regions. As a result, the problem of salient object detection becomes one of finding salient vertices and hyperedges in the hypergraph. The main advantage of hypergraph modeling is that it takes into account each pixel's (or region's) affinity with its neighborhood as well as its separation from image background. Furthermore, we propose an alternative approach based on center-versus-surround contextual contrast analysis, which performs salient object detection by optimizing a cost-sensitive support vector machine (SVM) objective function. Experimental results on four challenging datasets demonstrate the effectiveness of the proposed approaches against the state-of-the-art approaches to salient object detection.

In addition to a novel method for salient object detection, we tackle scene text detection, a challenging research problem in the both vision and document analysis community, from the saliency detection perspective. Motivated by the need to consider the widely varying forms of natural text, we propose a bottom-up approach to the problem which reflects the 'characterness' of an image region. In this sense our approach mirrors the move from saliency detection methods to measures of 'objectness'. In order to measure the characterness we develop three novel cues that are tailored for character detection, and a Bayesian method for their integration. Because text is made up

of sets of characters, we then design a Markov random field (MRF) model so as to exploit the inherent dependencies between characters. We experimentally demonstrate the effectiveness of our characteriness cues as well as the advantage of Bayesian multi-cue integration. The proposed text detector outperforms state-of-the-art methods on a few benchmark scene text detection datasets. We also show that our measurement of ‘characteriness’ is superior than state-of-the-art saliency detection models when applied to the same task.

## Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

**Signature**

**Date**

## Acknowledgements

First I would like to thank Dr. Chunhua Shen, my principal supervisor during my master study years. His dedication to research inspired me a lot, and will definitely influence me throughout my research career in the coming years. Under his supervision, not only have I learnt specific knowledge in the field, but I also know how to think as a mature researcher, including the ability to find unsolved challenges and address them in novel ways.

I would also like to thank staffs and visitors in the Australian Center for Visual Technologies (ACVT). As a non-native English speaker, I made grammar errors in academic manuscripts now and then. I would like to thank my co-supervisor, Prof. Anton van den Hengel, for his time and effort in revising those manuscripts. I also want to thank him for providing living cost to me for a year. In addition, I would like to thank Dr. Xi Li, a research fellow in our research center. A paper accepted by the International Conference on Computer Vision this year indicates that our cooperation is very successful. The paper is also a cornerstone of this thesis. Dr. Wenjing Jia is a lecturer from the University of Technology Sydney, who visited our lab for several months. She helped me a lot when I arrived in Adelaide. I would like to thank her for her selfless help and our cooperation on research topic of scene text detection. Other staffs in the ACVT, such as Prof. Ian Reid and Dr. Anthony Dick, have given me thoughtful suggestions on research. In a word, I am very lucky to be a member of the ACVT.

As a master by research student, I spent most time with PhD students. Guosheng Lin, Zhen Zhang, Josh Boys, Fayao Liu, Zhenhua Wang, Quoc Huy Tran, Trung Thanh Pham, Julio Zaragoza, Anh Tuan Ngo

all have helped me a lot on both research and life. Having friends with different culture backgrounds really enriches my life experience.

Finally, thanks go to the excellent support from staffs in the School of Computer Science at the University of Adelaide. Pru Carter, Julie Mayo, Sharyn Liersch and Jo Rogers are administrative staffs of our school. They have offered me enormous support within the school such as booking flight and accommodation for conferences, which enables me focusing my main effort on research. Thanks also go to our Head of School, David Suter, and our Postgraduate Coordinator Frank Neumann. I am really grateful for what they have done.



# List of Figures

1.1	Illustration of our approaches to salient object detection. . . . .	2
3.1	Illustration of cost-sensitive SVM for saliency detection. The saliency score is computed using based on the SVM classification results. . . . .	16
3.2	Illustration of hypergraph modeling for saliency detection using nonparametric clustering. . . . .	18
3.3	Illustration of salient object detection using two different types of graphs (i.e., hypergraph and standard pairwise graph). Clearly, our hypergraph saliency measure is able to accurately capture the intrinsic structural properties of the salient object. . . . .	20
3.4	Illustration of the gradient magnitude information for hyperedge saliency evaluation. The left subfigure shows the original image, and the middle subfigure displays the gradient magnitude map $I_g^*$ obtained by binarizing $I_g$ using the adaptive threshold $\mathcal{T}$ , as illustrated in the right subfigure. . . . .	20
3.5	Illustration of $M_g$ and $I_g^* \circ M_g$ for hyperedge saliency evaluation. The top row shows the multi-scale hyperedges; the middle row displays the scale-specific $M_g$ that indicates the pixels (within a narrow band) along the boundary of the scale-specific hyperedge; and the bottom row exhibits the filtered gradient magnitude map $I_g^* \circ M_g$ . . . . .	22

3.6	PR curves based on three different configurations: 1) using the SVM saliency approach only; and 2) using the hypergraph saliency approach only; 3) combining the SVM and hypergraph saliency approaches. Clearly, the saliency detection performance of using the third configuration outperform that of using the first and second configurations. From left to right: MSRA-1000, SOD, SED-100, and Imgsal-50. . . . .	25
3.7	Illustration of our saliency detection approach based on different parameter settings. (a) shows the PR curves of using different settings of $\lambda$ ; (b) displays the PR curves with different configurations of the scale space (determined by $\gamma$ ); and (c) exhibits the PR curves in different cases of scale numbers. . . . .	26
3.8	Evaluation of two different saliency fusion configurations on the MSRA-1000 dataset: 1) varying the hypergraph saliency approach while keeping the SVM saliency approach fixed; and 2) changing the SVM saliency approach while fixing the hypergraph saliency approach. (a) shows the PR curves of the saliency detection approaches associated with the first configuration while (b) displays the PR curves of the saliency detection approaches corresponding to the second configuration. . . . .	28
3.9	Saliency detection examples of our different approaches on the MSRA-1000 dataset. Clearly, the SVM saliency approach is able to locate the salient objects while the hypergraph saliency approach is capable of capturing the intrinsic structural information on the salient objects. . . . .	31
3.10	Quantitative PR curves of all the thirteen approaches on the four datasets. The rows from top to bottom correspond to MSRA-1000, SOD, SED-100, and Imgsal-50, respectively. Clearly, our approach achieve a better PR performance than the other competing approaches in most cases. . . . .	32

## LIST OF FIGURES

---

3.11	Quantitative ROC curves of all the thirteen approaches on the four datasets. The rows from top to bottom correspond to MSRA-1000, SOD, SED-100, and Imgsal-50, respectively. Clearly, our approach achieve a better ROC performance than the other competing approaches in most cases. . . . .	33
3.12	Quantitative F-measure performance of all the thirteen approaches on the four datasets. The columns from left to right correspond to MSRA-1000, SOD, SED-100, and Imgsal-50, respectively. Here, GS is a shorthand form of GS_SP. It is clear that our approach achieve a good F-measure performance on the four datasets. . . .	34
3.13	Salient object detection and segmentation examples on the MSRA-1000 dataset. For each example, the top row shows the input image and its corresponding saliency maps obtained by different approaches, and the bottom row displays the ground truth and the salient object segmentation results associated with the saliency maps. It is clear that our approach obtains the visually more consistent saliency detection and segmentation results than the other competing approaches. . . . .	35
3.14	Qualitative image retargeting performance comparison between [1] and ours. From left to right: images, our results, results of [1]. Clearly, the performance of our approach is better than that of [1].	36
4.1	Overview of our scene text detection approach. The charactereness model consists of the first two phases. . . . .	38
4.2	Cases that the original MSER fails to extract the characters while the modified eMSER succeeds. . . . .	39
4.3	Efficient stroke width computation [2] (best viewed in color). Note the color variation of non-characters and characters on (c). Larger color variation indicates larger stroke width variance. . . . .	40
4.4	Sample text (left) and four types of edge points represented in four different colors (right). Note that the number of edge points in blue is roughly equal to that in orange, and so for green and crimson. . . . .	42

4.5	Observation likelihood of characters (blue) and non-characters (red) on three characteriness cues <i>i.e.</i> , SW (top row), PD (middle row), and eHOG (bottom row). Clearly, for all three cues, observation likelihoods of characters are quite different from those of non-characters, indicating that the proposed cues are effective in distinguishing them. Notice that 50 bins are adopted. . . . .	56
4.6	Quantitative precision-recall curves performance of all the eleven approaches. Clearly, our approach achieves significant improvement compared with state-of-the-art saliency detection models for the measurement of ‘characteriness’. . . . .	57
4.7	Quantitative F-measure performance of all the eleven approaches. Clearly, our approach achieves significant improvement compared with state-of-the-art saliency detection models for the measurement of ‘characteriness’. . . . .	57
4.8	Visual comparison of saliency maps. Clearly, the proposed method highlights characters as salient regions whereas state-of-the-art saliency detection algorithms may be attracted by other stuff in the scene.	58
4.9	Sample outputs of our method on the ICDAR datasets and OSTD dataset. Detected text are in yellow rectangles. . . . .	59
4.10	False negatives of our approach. Clearly, there are two kinds of characters that our approach cannot handle, (i) characters in extremely blur and low resolution (top row), (ii) characters in uncommon fonts (bottom row). . . . .	60