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

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 prospective. 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 characterness 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 'characterness' 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 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