Faculty of Engineering and Information Technology
University of Technology Sydney

Visual Saliency Prediction For Stereoscopic Image

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Hao Cheng

July 2018

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Acknowledgments

Foremost, I would like to express my gratitude to my chief supervisor, Assoc. Prof. Jian Zhang, my co-supervisor Assoc. Prof. Qiang Wu at the University of Technology Sydney, and my supervisor Prof. Ping An at Shanghai University. I am extremely grateful for all the advice and guidance so unselfishly given to me over the last three and half years by these three distinguished academics. This research would not have been possible without their encouragement, continuous support and insight, especially during my leave of absence from the University.

Meanwhile, I am grateful to the doctors and nurses of Royal Prince Alfred Hospital. Without their help and treatment, I could not have recovered from myelitis. I would also like to thank the physiotherapists at MetroRehab Hospital for helping me to stand up again during physical therapy. My sincere appreciation and gratitude go to the people who helped me in my leave of absence for all their freely-given encouragement and care.

The wonderful support and assistance provided by many people during this research are very much appreciated by me and my family. I am very grateful for the help that I received from all of these people, but there are some special individuals who deserve thanks by name.

A very sincere thank you is definitely owed to my loving wife, Yaping Wang. Thank you for taking care of me when I was receiving treatment for six months. A very special thanks to my little son Luojia Cheng for bringing me a lot of happiness. Special thanks are also owed to my mother Hua Shao for all her tremendous assistance and unfailing support.

I thank my fellow labmates in the Big Data Technologies Centre (GBDTC) and the Advanced Analytics Institute: Shangrong Huang and Yucheng Wang for their invaluable advice and insightful discussions throughout my research, and for all the help during my leave of absence due to my illness. I would also like to thank my colleagues: Renhua Song, Jing Ren, Wenbo Wang, Dongyan Guo, Ying Cui, Jinsong Xu, Xiaoshui Huang and many more, for their selfless support over the course of my PhD candidature and for all the fun that we have shared over the last three and half years. Finally, to anyone whom I have not mentioned, please forgive me. I can most definitely assure you that you have occupied a unique and special place in my thoughts. Thank you!

Hao Cheng
January 2018 @ UTS

Contents

Certificate . . . . . **i**
Acknowledgment . . . . . **iii**
List of Figures . . . . . **ix**
List of Tables . . . . . **xi**
List of Publications . . . . . **xiii**
Abstract . . . . . **xv**

Chapter 1 Introduction **1**
1.1 Background 1
1.1.1 Visual saliency 2
1.1.2 Application of visual saliency 3
1.1.3 Classification of visual saliency 3
1.1.4 Development of 3D saliency detection 5
1.2 Research issues 6
1.2.1 Depth factor 6
1.2.2 Mechanisms of stereoscopic vision 7
1.2.3 Relationship among these mechanisms 7
1.3 Research contributions 7
1.4 Thesis structure 9

Chapter 2 Literature Review **12**
2.1 The 3D saliency model 12
2.1.1 Development of 3D saliency detection 12
2.1.2 Classification of 3D saliency detection 15

2.1.3 Recent research about 3D saliency detection 17
2.2 The fundamental method of saliency analysis 18
2.2.1 SLIC and multi-scale integration 18
2.2.2 Bayesian integration 21
2.2.3 Center bias 23
2.3 Experimental datasets and measurements 24
2.4 Summary 26

Chapter 3 A Preliminary Saliency Model for Stereoscopic Images 27
3.1 Introduction 27
3.2 Proposed stereo saliency detection 29
3.2.1 Local-global saliency 30
3.2.2 Surrounded enhancement 32
3.2.3 Stereo center prior enhancement 35
3.3 Experiments 37
3.3.1 Experimental setup 38
3.3.2 Experimental results and comparisons 39
3.4 Conclusion and discussion 47

Chapter 4 Stereoscopic Visual Saliency Prediction Based on Stereo Contrast and Stereo Focus 48
4.1 Introduction 48
4.2 Methodology 51
4.3 Proposed stereoscopic visual saliency prediction model 55
4.3.1 Pre-processing 55
4.3.2 Stereo contrast model 57
4.3.3 Stereo focus model 60
4.3.4 Enhancement 63
4.3.5 Bayesian integration scheme 65
4.4 Results and discussion 66
4.4.1 Experimental setup 67

4.4.2 Performance comparison with different combinations of components. 69
4.4.3 Comparison of our proposed method with other methods. 72
4.5 Conclusion and discussion 77

Chapter 5 A Computational Model for Stereoscopic Visual Saliency Prediction 79
5.1 Introduction 79
5.2 Related work 81
5.3 The proposed computational model for stereoscopic visual saliency 83
5.3.1 PE-CZ-BE mechanisms 84
5.3.2 The three modules based on the PE-CZ-BE mechanisms 86
5.3.3 Control function 87
5.3.4 The selection strategy 88
5.3.5 Framework based on a multi-feature saliency model . . 91
5.4 Experiments 99
5.4.1 Experimental setup 99
5.4.2 Performance of the features and components 101
5.4.3 Comparison with the state-of-the-art methods 101
5.5 Conclusions 106

Chapter 6 Conclusions and Future Work 108
6.1 Conclusions 108
6.2 Future work 109

Bibliography 111

List of Figures

1.1 Thesis Structure. Ch. 1 introduced the research background. Ch.2 provides the literature review. Ch. 3 presents a preliminary saliency model for stereoscopic images. Ch. 4 propose stereoscopic saliency detection based on stereo contrast and stereo focus. Ch. 5 presents a computational model for stereoscopic 3D visual saliency based on three mechanisms: pop-out effect, comfort zone, and background effect.Ch. 6 provides a final summary of this research and also suggests some future directions. 11

2.1 (a) a typical depth weight model and (b) a typical depth saliency model. 16
2.2 Examples of SLIC. 20

3.1 The framework of the proposed stereo saliency detection method 30
3.2 Global and local range. 32
3.3 Visual comparison of various saliency detection models. 43
3.4 Visual comparison of various saliency detection models. . . . 44

4.1 Stereo perception based on the different parallax 52
4.2 Stereo comfort zone based on human stereo vision 54
4.3 The framework of the proposed stereo saliency model 56
4.4 Global and local range. 58
4.5 A example of stereo contrast map. 60

4.6 The examples of the stereo focus maps. 63
4.7 An example of the proposed visual saliency prediction. (a) is the original left image and depth map. (b) shows the maps computed by the stereo contrast and stereo focus models. (c) shows the maps after clustering. (d) Final saliency map and ground truth. 64
4.8 An example of the proposed visual saliency prediction 73
4.9 Stereo comfort zone based on human stereo vision. DSM represents the depth saliency map in (Wang, DaSilva, LeCallet & Ricordel 2013) 77

5.1 Examples of the different combinations of the mechanisms. The first row depicts the left version of several stereoscopic images. The second row shows the corresponding depth maps. The last row shows the ground truths. 85
5.2 Flow chart of the stereoscopic saliency model 89
5.3 The main steps of the framework based on the multi-feature analysis. 92
5.4 Attention maps at pixel-level and at superpixel-level 93
5.5 Global and local range. 94
5.6 Global and local saliency maps. 95
5.7 A saliency map based on IP, and a ground-truth saliency map based on the GP 95
5.8 Samples of four features and combined saliency maps. 97
5.9 Samples of the proposed saliency model. 98
5.10 Examples of the components of the proposed model 103

List of Tables

3.1 The Pseudo-code 37

3.2 Comparison between each component in database (Wang et al. 2013) 40

3.3 Comparison between each component in database (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012) 40

3.4 Comparison between the proposed framework with others. DSM represents the depth saliency map in (Wang et al. 2013) . . . 42

3.5 Comparison between different 3D saliency detection models. "+" means the combination by simple summation by study (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). "×" means the combination by point-wise multiplication (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). DSM represents the depth saliency map in (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). 45

4.1 The Pseudo-code 66

4.2 Comparison between different component orders in database (Wang et al. 2013) 70

4.3 Comparison between different component orders in database (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012) . . . 70

4.4 Comparison between the proposed framework with others. DSM represents the depth saliency map in (Wang et al. 2013) . . . 75

4.5 Comparison between different 3D viusal saliency prediction models. "+" means the combination by simple summation by study (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). "×" means the combination by point-wise multiplication (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). DSM represents the depth saliency map in (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012). 76

5.1 Comparison between four features and two components in the dataset (Wang et al. 2013) 102
5.2 Comparison between each component in the dataset (Lang, Nguyen, Katti, Yadati, Kankanhalli & Yan 2012) 102
5.3 Comparison between the proposed framework and others. DSM represents the depth saliency map in (Wang et al. 2013) . . . 105
5.4 Comparison between 3D saliency detection models. 107

List of Publications

Papers Published

- **Hao Cheng**, Jian Zhang, Qiang Wu, and Ping An (2017), Stereoscopic Saliency Detection Based on Stereo Contrast And Stereo Focus. The EURASIP Journal on Image and Video Processing. Volumn 2017, pp.32-61.

- **Hao Cheng**, Jian Zhang, Ping An, Zhi Liu (2015), A Novel Saliency Model for Stereoscopic Images. *in* 'Proceedings of the Internetional Conference on Digital Image Computing: Techniques and Applications (**DICTA2015**)', pp.1-7.

Papers to be Submitted/Under Review

- **Hao Cheng**, Jian Zhang, Qiang Wu, and Ping An (2016) A Computational Model for Stereoscopic 3D Visual Saliency Based on Surrounding Region, Contrast and Interest Point, to be submit the IEEE Transactions on Multimedia.

Abstract

Saliency prediction is considered to be key to attentional processing. Attention improves learning and survival by compelling creatures to focus their limited cognitive resources and perceptive abilities on the most interesting region of the available sensory data. Computational models for saliency prediction are widely used in various fields of computer vision, such as object detection, scene recognition, and robot vision. In recent years, several comprehensive and well-performing models have been developed. However, these models are only suitable for 2D content. With the rapid development of 3D imaging technology, an increasing number of applications are emerging that rely on 3D images and video. In turn, demand for computational saliency models that can handle 3D content is growing. Compared to the significant progress in 2D saliency research, studies that consider depth factor as part of stereoscopic saliency analysis are rather limited. Thus, the role depth factor in stereoscopic saliency analysis is still relatively unexplored.

The aim of this thesis is to fill this gap in the literature by exploring the role of depth factors in three aspects of stereoscopic saliency: how depth factors might be used to leverage stereoscopic saliency detection; how to build a stereoscopic saliency model based on the mechanisms of human stereoscopic vision; and how to implement a stereoscopic saliency model that can adjust to the particular aspect of human stereoscopic vision reflected in specific 3D content. To meet these three aims, this thesis includes three distinct computation models for stereoscopic saliency prediction based on the past and present outcomes of my research. The contributions of the thesis are as

follows:

Chapter 3 presents a preliminary saliency model for stereoscopic images. This model exploits depth information and treats the depth factor of an image as a weight to leverage saliency analysis. First, low-level features from the color and depth maps are extracted. Then, to extract the structural information from the depth map, the surrounding Boolean-based map is computed as a weight to enhance the low-level features. Lastly, a stereoscopic center prior enhancement based on the saliency probability distribution in the depth map is used to determine the final saliency.

The model presented in Chapter 4 predicts stereoscopic visual saliency using stereo contrast and stereo focus. The stereo contrast submodel measures stereo saliency based on color, depth contrast, and the pop-out effect. The stereo focus submodel measures the degree of focus based on monocular vision and comfort zones. Multi-scale fusion is then used to generate a map for each of the submodels, and a Bayesian integration scheme combines both maps into a stereo saliency map.

However, the stereoscopic saliency model presented in Chapter 4 does not explain all the phenomena in stereoscopic content. So, to improve the models robustness, Chapter 5 includes a computational model for stereoscopic 3D visual saliency with three submodels based on the three mechanisms of the human vision system: the pop-out effect, comfort zones, and the background effect. Each mechanism provides useful cues for stereoscopic saliency analysis depending on the nature of the stereoscopic content. Hence, the model in Chapter 5 incorporates a selection strategy to accurately determine which submodel should be used to process an image. The approach is implemented within a purpose-built, multi-feature analysis framework that assesses three features: surrounding region, color and depth contrast, and points of interest.

All three models were verified through experiments with two eye-tracking databases. Each outperforms the state-of-the-art saliency models.