

## DOCTORAL THESIS

### Visual attention: saliency detection and gaze estimation

Peng, Qinmu

*Date of Award:*  
2015

[Link to publication](#)

#### **General rights**

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

Visual attention is an important characteristic in the human vision system, which is capable of allocating the cognitive resources to the selected information. Many researchers are attracted to the study of this mechanism in the human vision system and have achieved a wide range of successful applications. Generally, there are two tasks encountered in the visual attention research including visual saliency detection and gaze estimation. The former is normally described as distinctiveness or prominence as a result of a visual stimulus. Given images or videos as input, saliency detection methods try to simulate the mechanism of human vision system, predicting and locating the salient parts in them. While the later involves physical device to track the eye movement and estimate the gaze points.

As for saliency detection, it is an effective technique for studying and mimicking the mechanism of the human vision system. Most of saliency models can predict the visual saliency with the boundary or the rough location of the true salient object, but miss the appearance or shape information. Besides, they pay little attention to the image quality problem such as low-resolution or noises. To handle these problems, in this thesis, we propose to model the visual saliency from local and global perspectives for better detection of the visual saliency. The combination of the local and global saliency scheme employing different visual cues can make full use of their respective advantages to compute the saliency. Compared with existing models, the proposed method can provide better saliency with more appearance and shape information, and can work well even in the low-resolution or noisy images. The experimental results demonstrate the superiority of the proposed algorithm.

Next, video saliency detection is another issue for the visual saliency computation. Numerous works have been proposed to extract the video saliency for the tasks of object detection. However, one might not be able to obtain desirable saliency for inferring the region of foreground objects when the video presents low contrast or complicated background. Thus, this thesis develops an salient object detection approach with less demanding assumption which gives higher detection performance. The method computes the visual saliency in each frame using a weighted multiple manifold ranking algorithm. It then computes motion cues to estimate the motion saliency and localization prior. By adopting a new energy function, the data term depends on the visual saliency and localization prior; and the smoothness term depends on the constraint in time and space. Compared to existing methods, our approach automatically segments the persistent foreground object while preserving the potential shape. We apply our method to challenging benchmark videos, and show competitive or better results than the existing counterparts.

Additionally, to address the problem of gaze estimation, we present a low cost and efficient approach to obtain the gaze point. As opposed to eye gaze estimation techniques requiring specific hardware, e.g. infrared high-resolution camera and infrared light sources, as well as a cumbersome calibration process, we concentrate on visible-imaging and present an approach for gaze estimation using a web camera in a desktop environment. We combine intensity energy and edge strength to locate the iris center and utilize the piecewise eye corner detector to detect the eye corner. To compensate for head movement causing gaze error, we adopt a sinusoidal head model (SHM) to simulate the 3D head shape, and propose an adaptive weighted facial features embedded in the pose from the orthography and scaling with iterations algorithm (AWPOSIT), whereby the head pose can be estimated. Consequently, the gaze estimation is obtained by the integration of the eye vector and head movement information. The proposed method is insensitive to the light conditions, and the experimental results show the efficacy of the proposed approach.

# Table of Contents

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Motivations of this Project . . . . .	3
1.3 Main Contributions of this Thesis . . . . .	5
1.4 Organization of this Thesis . . . . .	6
<b>Chapter 2 Literature Review of Related Works</b>	<b>7</b>
2.1 Image Saliency Detection . . . . .	7
2.2 Video Saliency Detection . . . . .	10
2.3 Gaze Estimation . . . . .	12
<b>Chapter 3 Image Saliency Detection Based on Local and Global Visual Cues</b>	<b>15</b>
3.1 Introduction . . . . .	15

3.2	Proposed Method . . . . .	17
3.2.1	An Image Pre-processing: A Smoothing Scheme with Automatic Selection of Smoothing Parameter . . . . .	18
3.2.2	Detection of Local Saliency . . . . .	20
3.2.3	Detection of Global Saliency . . . . .	24
3.2.4	Combination of Different Saliencies . . . . .	28
3.3	Experimental Results . . . . .	31
3.3.1	Experimental Settings . . . . .	31
3.3.2	Results and Analysis . . . . .	33
3.4	Concluding Remarks . . . . .	40
 <b>Chapter 4 Video Saliency Detection for Salient Object Segmentation</b>		<b>42</b>
4.1	Introduction . . . . .	42
4.2	Preliminary . . . . .	44
4.3	Proposed Method . . . . .	45
4.3.1	Visual Saliency Computing . . . . .	45
4.3.2	Motion Cues Computing . . . . .	47
4.3.3	Video Object Segmentation . . . . .	50
4.4	Experimental Evaluation . . . . .	52
4.4.1	Comparison with Saliency Detection Approaches . . . . .	52
4.4.2	Salient Object Segmentation . . . . .	53
4.5	Concluding Remarks . . . . .	55
 <b>Chapter 5 Gaze Estimation</b>		<b>58</b>
5.1	Introduction . . . . .	58
5.2	The Proposed Method . . . . .	59
5.2.1	Eye Region Detection . . . . .	60
5.2.2	Eye Features Detection . . . . .	62
5.2.3	Eye Vector and Calibration . . . . .	66
5.2.4	Head Pose Estimation . . . . .	67

5.3	Experimental Evaluation . . . . .	74
5.3.1	Eye Center Detection . . . . .	74
5.3.2	Head Pose Estimation . . . . .	75
5.3.3	Gaze Estimation . . . . .	77
5.4	Concluding Remarks . . . . .	82
<b>Chapter 6 Conclusions and Future Work</b>		<b>84</b>
6.1	Conclusions . . . . .	84
6.2	Future Work . . . . .	86
<b>Bibliography</b>		<b>88</b>
<b>Curriculum Vitae</b>		<b>102</b>