

Contents

1 Introduction	1
1.1 Contributions	4
1.2 Outline	6
References	7
2 Background	9
2.1 Attention Models	9
2.1.1 Visual Attention	10
2.1.2 Auditory Attention	16
2.1.3 Multimodal Attention	21
2.2 Applications of Attention Models	22
2.2.1 Image Processing and Computer Vision	23
2.2.2 Audio Processing	24
2.2.3 Robotics	24
2.2.4 Computer Graphics	25
2.2.5 Design, Marketing, and Advertisement	26
References	27
3 Bottom-Up Audio-Visual Attention for Scene Exploration	35
3.1 Related Work and Contributions	37
3.1.1 Spectral Visual Saliency	37
3.1.2 Visual Saliency and Color Spaces	39
3.1.3 Visual Saliency and Faces	40
3.1.4 Auditory Saliency	41
3.1.5 Audio-Visual Saliency-Based Exploration	41
3.1.6 Scene Analysis	43
3.2 Visual Attention	44
3.2.1 Spectral Visual Saliency	46
3.2.2 Color Space Decorrelation	63
3.2.3 Modeling the Influence of Faces	74
3.3 Auditory Attention	81
3.3.1 Auditory Novelty Detection	81

3.3.2	Evaluation	85
3.4	Saliency-Based Audio-Visual Exploration	87
3.4.1	Gaussian Proto-Object Model	88
3.4.2	Auditory Proto-Objects	88
3.4.3	Visual Proto-Objects	89
3.4.4	Audio-Visual Fusion and Inhibition	92
3.4.5	Evaluation	94
3.5	Multiobjective Exploration Path	98
3.5.1	Exploration Path	99
3.5.2	Exploration Strategies	99
3.5.3	Evaluation	101
3.6	Summary and Future Directions	105
	References	107
4	Multimodal Attention with Top-Down Guidance	115
4.1	Related Work and Contributions	117
4.1.1	Joint Attention	117
4.1.2	Visual Attention	120
4.2	Debiased Salient Object Detection	124
4.2.1	The MSRA Dataset	125
4.2.2	MSRA’s Photographer Bias	126
4.2.3	Salient Object Detection	129
4.2.4	Debiased Salient Object Detection and Pointing	133
4.3	Focusing Computational Attention in Human-Robot Interaction	134
4.3.1	Pointing Gestures	136
4.3.2	Language	145
4.4	Gaze Following in Web Images	159
4.4.1	Approach	160
4.4.2	The Gaze@Flickr Dataset	161
4.4.3	Evaluation	163
4.5	Summary and Future Directions	168
	References	170
5	Conclusion	177
5.1	Summary	177
5.2	Future Work	179
	Appendix A: Applications	181
	Appendix B: Dataset Overview	187
	Appendix C: Color Space Decorrelation: Full Evaluation	191
	Appendix D: Center Bias Integration Methods	199



<http://www.springer.com/978-3-319-33794-4>

Multimodal Computational Attention for Scene
Understanding and Robotics

Schauerte, B.

2016, XXIV, 203 p. 55 illus., 51 illus. in color., Hardcover

ISBN: 978-3-319-33794-4