A Neural Framework for Visual Scene Analysis with Selective Attention

by

Eric Wai-Shing Chong, B.E.(Hons)

A thesis submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy

The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
Department of Electrical and Electronic Engineering
June, 2001
Contents

Abstract ix
Declaration xi
Acknowledgments xiii
List of Publications xv
List of Principal Symbols xvii
List of Abbreviations xix
List of Figures xxviii

1 Introduction 1
1.1 Background and Motivation 1
1.2 Research Objectives 2
1.3 Research Methodology and Approach 3
1.3.1 Shape-Based Representation 4
1.3.2 Self-Organising Neural Architectures 5
1.4 Major Contributions of the Thesis 5
1.5 Outline of the Thesis 6
2 Vision
2.1 Visual Perception ........................................... 9
2.2 The Visual System ........................................... 10
  2.2.1 The Neuron ............................................ 11
2.3 Object Recognition ......................................... 12
  2.3.1 Object-Centred versus Viewer-Centred Representations ............................................ 13
2.4 Selective Attention ......................................... 13
  2.4.1 Psychology of Attention ............................... 15
  2.4.2 Neurophysiology of Attention ....................... 16
2.5 Summary .................................................... 17

3 Object Recognition Approaches and Models ................. 19
3.1 Computational Approaches ................................ 19
  3.1.1 Model-Based Methods ................................. 20
  3.1.2 Appearance-Based Methods ........................... 23
3.2 Neuro-Vision Systems ..................................... 25
  3.2.1 Artificial Neural Networks ............................ 26
  3.2.2 Learning in Neural Nets ............................... 27
  3.2.3 Network Architectures and Learning Paradigms ......................... 28
  3.2.4 Related Neural Architectures for Object and Pattern Recognition ......................... 30
3.3 Adaptive Resonance Theory ............................... 32
  3.3.1 ART2 and ART3 ........................................ 34
  3.3.2 Parameter Estimation - A Case Study of ART ......................... 40
3.4 Selective Attention Adaptive Resonance Theory ........... 48
  3.4.1 The SAART Chemical Synapse and Neural Layers ......................... 50
  3.4.2 The SAART Architecture ................................ 54
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Summary</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Models of Visual Object Recognition</td>
<td>57</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction and Overview</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Translation Invariance</td>
<td>58</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Stages of Operation</td>
<td>58</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Partitioning of the Input Field</td>
<td>60</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Bottom-Up Activation of Stored Memory</td>
<td>60</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Selective Transfer of Bottom-Up and Top-Down Patterns</td>
<td>63</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Matching of Bottom-Up and Top-Down Patterns</td>
<td>65</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Mismatch Reset</td>
<td>65</td>
</tr>
<tr>
<td>4.3</td>
<td>Recognition in Cluttered Images</td>
<td>65</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Stages of Operation</td>
<td>68</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Incorporation of Adaptive Resonance Theory</td>
<td>69</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Implementation of Selective Attention</td>
<td>71</td>
</tr>
<tr>
<td>4.4</td>
<td>Preattentive Processing: Deploying Automatic Attention</td>
<td>72</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Stages of Operation</td>
<td>73</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Attentional Capture: High Activity Region Selection</td>
<td>76</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Selective Transfer of High Activity Region</td>
<td>78</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Attentional Shift Considerations</td>
<td>79</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Attentional Shift Implementation</td>
<td>81</td>
</tr>
<tr>
<td>4.5</td>
<td>Rotation Invariance</td>
<td>83</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Rotation Invariant Model Propositions and Assumptions</td>
<td>86</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Stages of Operation</td>
<td>87</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Implementation of Mental Rotation</td>
<td>90</td>
</tr>
<tr>
<td>4.6</td>
<td>Distortion Invariance</td>
<td>94</td>
</tr>
</tbody>
</table>
CONTENTS

4.6.1 Stages of Operation ............................................. 97
4.6.2 Band Transformation ........................................... 101
4.6.3 Shape Attraction ............................................... 101
4.7 Integrated Model of Architectural Framework ................. 102
4.8 Conclusions ...................................................... 107

5 Model Simulations and Analysis ................................. 109

5.1 Introduction ...................................................... 109
5.2 Learning .......................................................... 110
5.3 Translation Invariance ........................................... 113
  5.3.1 Simulation I .................................................. 114
  5.3.2 Simulation II ................................................ 116
5.4 Recognition in Cluttered Images ............................... 119
  5.4.1 Simulation I .................................................. 119
  5.4.2 Simulation II ................................................ 120
  5.4.3 Simulation III ............................................... 123
5.5 Preattentive Processing: Automatic Attentional Shift and Capture ........................................ 124
  5.5.1 Simulation I .................................................. 126
  5.5.2 Simulation II ................................................ 129
  5.5.3 Simulation III ............................................... 131
5.6 Rotation Invariance ............................................... 133
  5.6.1 Simulation I .................................................. 133
  5.6.2 Simulation II ................................................ 137
5.7 Distortion Invariance .......................................... 137
  5.7.1 Simulation I .................................................. 139
  5.7.2 Simulation II ................................................ 141
5.8 Design of System Parameters ........................................ 143
5.9 Real-World Imagery Simulations .................................... 146
  5.9.1 Learning ...................................................... 146
  5.9.2 Simulation I .................................................. 148
  5.9.3 Simulation II .................................................. 153
  5.9.4 Simulation III .................................................. 154
5.10 Limitations of the Model ............................................ 158
5.11 Conclusions ........................................................ 159

6 Recognition of Moving Objects ....................................... 161
  6.1 Introduction and Overview ....................................... 161
  6.2 The Motion Pathway ............................................. 162
    6.2.1 Apparent Motion .......................................... 163
  6.3 Neural Architecture for Motion-Direction Computation ........ 163
    6.3.1 The Input Layer ........................................... 165
    6.3.2 The Photo-Sensitive Layer ............................... 166
    6.3.3 The Transient Layer ..................................... 166
    6.3.4 The Direction-Selective Layer ........................... 168
    6.3.5 The Selective Attention Layer ........................... 174
  6.4 Simulations and Analysis ....................................... 175
    6.4.1 Motion-Direction Detection ............................... 175
    6.4.2 Effects of Stimulus Contrast and Temporal Frequency ... 179
    6.4.3 Directional Bias ........................................... 186
  6.5 A Visual Motion Cue for Recognition of Moving Objects ........ 190
    6.5.1 A Test Case ................................................ 192
  6.6 Conclusions ....................................................... 194
7 Advanced Framework Features

7.1 Introduction ........................................... 195

7.2 Complementary Selective Attention Adaptive Resonance Theory .................. 195

7.2.1 Complementary Feedforward-Feedback Interactions .............................. 196

7.2.2 Network Implementation ................................ 198

7.2.3 Parts Recognition and Occlusion Simulations .................................... 202

7.2.4 Framework Simulations with CSAART ........................................ 204

7.3 Robust Automatic Attentional Capture ............................................. 207

7.3.1 Automatic Selection Threshold ........................................ 209

7.3.2 Automatic Resizing of Window of Attention .................................... 209

7.3.3 Partially Captured Objects .................................................. 210

7.4 Size Invariance .............................................. 211

7.5 Conclusion .................................................. 212

8 Conclusions and Recommendations ...................................................... 213

8.1 Recapitulation of the Thesis .................................................. 213

8.2 Concluding Statement .................................................. 216

8.3 Recommendations for Future Work ........................................ 217

A Additional Simulations ....................................................... 219
Abstract

Attention is essential to the analysis of visual scenes that consist of multiple objects, especially in cases where the objects are embedded in complex and cluttered backgrounds. Despite its importance, few artificial neural network models of object and pattern recognition have incorporated attentional mechanisms. Recent advances in cognitive neuroscience have provided important information on the neural mechanisms of attention. Significantly, attentional processes involve the modulation of neuronal signals, and are clearly influenced by memory related processes via feedforward-feedback interactions.

This thesis proposes an architectural framework based on neural networks for visual scene analysis with attentional mechanisms. The core of the framework is based on an adaptive resonance theory architecture, which is a self-organising neural network for stable learning of recognition codes. The proposed model exploits the computational role of attention in visual object recognition by modelling the dynamics of attentional processes for perceptual grouping and selective processing. As a result, the proposed model is capable of performing translation, rotation, and distortion invariant 2D object recognition in the presence of background clutter and occlusion. The model is shown to be flexible to extensions by incorporating an elementary motion detection architecture for recognising moving objects. Furthermore, the use of feedforward-feedback modulation has enabled partial or incomplete familiar objects to be recognised in a variety of visual conditions. Biologically, such feedforward-feedback interactions can be used to explain the phenomenon of visual completion.

Simulation studies undertaken demonstrate the effectiveness of the proposed model in recognising 2D objects in many non-ideal visual conditions. The practical feasibility of the neural architecture is demonstrated through its application to real-world images. Despite difficult visual environments, including severe distortion, the simulation results indicate the model can detect, locate and recognise the learned objects from the simulated images.

From the research presented in this thesis, it is concluded that the use of attentional mechanisms can enhance artificial vision systems to cope with difficult visual conditions. It is shown that feedforward-feedback interactions with synaptic modulation are a versatile and powerful mech-
anism for performing many useful functions such as transformations, filtering, gain control, and selective processing in neural network based vision systems.