

# Contents

|          |                                                                                 |           |
|----------|---------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Throwing Down the Visual Intelligence Gauntlet</b>                           | <b>1</b>  |
|          | <i>Cheston Tan, Joel Z. Leibo, Tomaso Poggio</i>                                |           |
| 1        | Artificial Intelligence: Are We There Yet?                                      | 1         |
| 1.1      | A Compass for the Uncharted Journey towards Intelligence                        | 2         |
| 2        | The Neuromorphic Approach to Visual Intelligence                                | 3         |
| 2.1      | Anatomy and Physiology of the Primary Visual Cortex (V1)                        | 3         |
| 2.2      | Hubel-Wiesel Models: Successive Tuning and Pooling                              | 5         |
| 2.3      | Consistency with Experimental Results at Multiple Levels                        | 6         |
| 2.4      | From Neuroscience Models to Engineering Applications                            | 10        |
| 3        | What's Next in the Quest for Visual Intelligence?                               | 10        |
| 3.1      | Going beyond "What Is Where"                                                    | 10        |
| 3.2      | From Perception to Abstraction                                                  | 11        |
| 3.3      | A Loose Hierarchy of Visual Tasks                                               | 11        |
| 3.4      | It's Time to Try again – The MIT Intelligence Initiative                        | 12        |
|          | References                                                                      | 13        |
| <b>2</b> | <b>Actionable Information in Vision</b>                                         | <b>17</b> |
|          | <i>Stefano Soatto</i>                                                           |           |
| 1        | Preamble                                                                        | 17        |
| 2        | Introduction                                                                    | 20        |
| 3        | Preliminaries                                                                   | 22        |
| 3.1      | Notation and Conventions                                                        | 24        |
| 3.2      | Visibility and Quantization                                                     | 25        |
| 3.3      | Invariant and Sufficient Statistics                                             | 26        |
| 4        | Placing the Ecological Approach to Visual Perception onto Computational Grounds | 27        |

|          |                                                                                                        |           |
|----------|--------------------------------------------------------------------------------------------------------|-----------|
| 4.1      | Actionable Information . . . . .                                                                       | 27        |
| 4.2      | Invertible and Non-invertible Nuisances . . . . .                                                      | 27        |
| 4.3      | The Actionable Information Gap . . . . .                                                               | 29        |
| 4.4      | Information Pickup . . . . .                                                                           | 31        |
| 5        | Representational Structures . . . . .                                                                  | 33        |
| 5.1      | Computing Actionable Information . . . . .                                                             | 33        |
| 5.2      | Closing the Actionable Information Gap . . . . .                                                       | 34        |
| 6        | Empirical Consequences of the Definitions . . . . .                                                    | 34        |
| 6.1      | Exploration via Information Pickup . . . . .                                                           | 35        |
| 7        | Summary and Discussion . . . . .                                                                       | 42        |
|          | References . . . . .                                                                                   | 45        |
| <b>3</b> | <b>Learning Binary Hash Codes for Large-Scale Image Search . . . . .</b>                               | <b>49</b> |
|          | <i>Kristen Grauman, Rob Fergus</i>                                                                     |           |
| 1        | Introduction . . . . .                                                                                 | 50        |
| 2        | Search Algorithms for Binary Codes . . . . .                                                           | 52        |
| 2.1      | Linear Scan in Hamming Space . . . . .                                                                 | 53        |
| 2.2      | Semantic Hashing . . . . .                                                                             | 54        |
| 2.3      | Locality Sensitive Hashing . . . . .                                                                   | 56        |
| 2.4      | Recap of Search Strategy Tradeoffs . . . . .                                                           | 58        |
| 3        | Supervised Methods for Learning Binary Projections . . . . .                                           | 59        |
| 3.1      | Forms of Supervision to Define Semantic Similarity . . . . .                                           | 60        |
| 3.2      | Hash Functions for Learned Mahalanobis Kernels . . . . .                                               | 61        |
| 3.3      | Learned Binary Embeddings for Semantic Similarity . . . . .                                            | 68        |
| 3.4      | Other Supervised Methods . . . . .                                                                     | 72        |
| 4        | Unsupervised Methods for Defining Binary Projections . . . . .                                         | 73        |
| 4.1      | Binary Codes for Specific Similarity Functions . . . . .                                               | 73        |
| 4.2      | Spectral Hashing . . . . .                                                                             | 74        |
| 4.3      | Kernelized Locality Sensitive Hashing . . . . .                                                        | 79        |
| 4.4      | Other Unsupervised Methods . . . . .                                                                   | 83        |
| 5        | Conclusions . . . . .                                                                                  | 83        |
|          | References . . . . .                                                                                   | 83        |
| <b>4</b> | <b>Bayesian Painting by Numbers: Flexible Priors for Colour-Invariant Object Recognition . . . . .</b> | <b>89</b> |
|          | <i>Jeroen C. Chua, Inmar E. Givoni, Ryan P. Adams, Brendan J. Frey</i>                                 |           |
| 1        | Introduction . . . . .                                                                                 | 89        |
| 2        | The Colour-Invariant Admixture Model . . . . .                                                         | 94        |
| 2.1      | The Standard Stel Model . . . . .                                                                      | 95        |
| 2.2      | The Stel Model as a Generative Admixture . . . . .                                                     | 96        |
| 2.3      | Inference via Gibbs Sampling . . . . .                                                                 | 98        |
| 2.4      | Estimating the Posterior Distribution over Stels . . . . .                                             | 101       |
| 3        | Using Stels for Supervised Learning Tasks . . . . .                                                    | 102       |
| 4        | Experimental Evaluation . . . . .                                                                      | 103       |
| 4.1      | Experimental Setup . . . . .                                                                           | 104       |
| 4.2      | Learning a Flexible Number of Stels . . . . .                                                          | 104       |

|          |                                                                                                                         |            |
|----------|-------------------------------------------------------------------------------------------------------------------------|------------|
| 4.3      | Object Recognition Using CIA                                                                                            | 108        |
| 4.4      | Effect of Hyperparameter Choices on Learnt Stels                                                                        | 111        |
| 5        | Summary                                                                                                                 | 115        |
|          | References                                                                                                              | 116        |
| <b>5</b> | <b>Real-Time Human Pose Recognition in Parts from Single Depth Images</b>                                               | <b>119</b> |
|          | <i>Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake</i> |            |
| 1        | Introduction                                                                                                            | 119        |
| 1.1      | Related Work                                                                                                            | 121        |
| 2        | Data                                                                                                                    | 122        |
| 2.1      | Depth Imaging                                                                                                           | 123        |
| 2.2      | Motion Capture Data                                                                                                     | 123        |
| 2.3      | Generating Synthetic Data                                                                                               | 124        |
| 3        | Body Part Inference and Joint Proposals                                                                                 | 124        |
| 3.1      | Body Part Labeling                                                                                                      | 124        |
| 3.2      | Depth Image Features                                                                                                    | 125        |
| 3.3      | Randomized Decision Forests                                                                                             | 126        |
| 3.4      | Joint Position Proposals                                                                                                | 127        |
| 4        | Experiments                                                                                                             | 128        |
| 4.1      | Qualitative Results                                                                                                     | 130        |
| 4.2      | Classification Accuracy                                                                                                 | 130        |
| 4.3      | Joint Prediction Accuracy                                                                                               | 131        |
| 5        | Discussion                                                                                                              | 133        |
|          | References                                                                                                              | 134        |
| <b>6</b> | <b>Scale-Invariant Vote-Based 3D Recognition and Registration from Point Clouds</b>                                     | <b>137</b> |
|          | <i>Minh-Tri Pham, Oliver J. Woodford, Frank Perbet, Atsuto Maki, Riccardo Gherardi, Björn Stenger, Roberto Cipolla</i>  |            |
| 1        | Introduction                                                                                                            | 137        |
| 2        | Background                                                                                                              | 139        |
| 2.1      | Global Approaches vs. Local Approaches                                                                                  | 139        |
| 2.2      | Learning a Local Appearance Model                                                                                       | 141        |
| 2.3      | Finding Local Modes in the Vote Space                                                                                   | 143        |
| 2.4      | Mean Shift                                                                                                              | 145        |
| 3        | The SRT Distance and Its Mean                                                                                           | 148        |
| 3.1      | Distance Definition                                                                                                     | 148        |
| 3.2      | Mean Computation                                                                                                        | 150        |
| 3.3      | SRT Mean Shift                                                                                                          | 153        |
| 4        | Experiments                                                                                                             | 154        |
| 4.1      | Experimental Setup                                                                                                      | 154        |
| 4.2      | Inference                                                                                                               | 154        |
| 4.3      | Evaluation                                                                                                              | 155        |
| 4.4      | Results                                                                                                                 | 156        |

|          |                                                                                           |            |
|----------|-------------------------------------------------------------------------------------------|------------|
| 5        | Conclusion .....                                                                          | 159        |
|          | References .....                                                                          | 160        |
| <b>7</b> | <b>Multiple Classifier Boosting and Tree-Structured Classifiers .....</b>                 | <b>163</b> |
|          | <i>Tae-Kyun Kim, Roberto Cipolla</i>                                                      |            |
| 1        | Introduction .....                                                                        | 163        |
| 2        | Multiple Classifier Boosting .....                                                        | 166        |
| 2.1      | MCBoost: Multiple Classifier Boosting .....                                               | 168        |
| 2.2      | Experiments .....                                                                         | 172        |
| 3        | Online Multiple Classifier Boosting for Object Tracking .....                             | 176        |
| 3.1      | Joint Boosting and Clustering .....                                                       | 178        |
| 3.2      | Online MCBQ for Object Tracking .....                                                     | 179        |
| 3.3      | Results .....                                                                             | 181        |
| 4        | Conversion of a Boosting Classifier into a Decision Tree by<br>Boolean Optimisation ..... | 184        |
| 4.1      | Conversion of a Boosting Classifier into a Tree .....                                     | 186        |
| 4.2      | Experiments and Discussions .....                                                         | 191        |
| 5        | Summary and Conclusion .....                                                              | 194        |
|          | References .....                                                                          | 195        |
| <b>8</b> | <b>Simultaneous Detection and Tracking with Multiple Cameras .....</b>                    | <b>197</b> |
|          | <i>Murtaza Taj, Andrea Cavallaro</i>                                                      |            |
| 1        | Introduction .....                                                                        | 197        |
| 2        | Multi-camera Tracking: A Brief Overview .....                                             | 198        |
| 3        | Multi-camera Detection Volume .....                                                       | 199        |
| 4        | Track-Before-Detect on the Detection Volume .....                                         | 201        |
| 4.1      | Single Target Track-Before-Detect .....                                                   | 201        |
| 4.2      | Track-Before-Detect Particle Filter .....                                                 | 203        |
| 4.3      | Multi-camera, Multi-target Track-Before-Detect .....                                      | 206        |
| 4.4      | Discussion .....                                                                          | 210        |
| 5        | Conclusions .....                                                                         | 212        |
|          | References .....                                                                          | 213        |
| <b>9</b> | <b>Applications of Computer Vision to Vehicles: An Extreme Test .....</b>                 | <b>215</b> |
|          | <i>Alberto Broggi, Stefano Cattani, Paolo Medici, Paolo Zani</i>                          |            |
| 1        | Motivation .....                                                                          | 215        |
| 2        | Sensors .....                                                                             | 216        |
| 2.1      | Design Considerations .....                                                               | 216        |
| 2.2      | Perception Systems .....                                                                  | 218        |
| 3        | Processing Systems .....                                                                  | 221        |
| 3.1      | Lane Detection .....                                                                      | 221        |
| 3.2      | Stereo Obstacle Detection .....                                                           | 227        |
| 3.3      | Laser Obstacle Detection .....                                                            | 231        |
| 3.4      | Vehicle Detection .....                                                                   | 236        |
| 4        | Statistics .....                                                                          | 242        |
| 4.1      | Preliminary Test .....                                                                    | 242        |

- 4.2 The Challenge Statistics ..... 243
- 5 Conclusions ..... 246
  - 5.1 Laser Obstacle Detection ..... 247
  - 5.2 Lane Detection ..... 247
  - 5.3 Stereo Obstacle Detection ..... 247
  - 5.4 Vehicle Detection ..... 247
  - 5.5 The Final Word ..... 248
- References ..... 248