

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

# Video Tracking of People under Severe Occlusions

**Zui Zhang**

A thesis submitted for the Degree of Doctor of Philosophy of the  
University of Technology, Sydney

Feb, 2010

## 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

---

# ACKNOWLEDGEMENTS

On the completion of this research, I wish to thank those people and organizations for their support and help kindly given to me during the past three years. My principal supervisor, Professor Massimo Piccardi of University of Technology, Sydney, has been a significant presence in my life. His careful supervision is a true gift for the conduct of this study. His critical comments and suggestions on each piece of writing I made have strengthened this study significantly. I will always be thankful for his persistent guidance and encouragement.

I am grateful to Dr Hatice Gunes and Dr Oscar Perez Concha, for keeping their hearts and office doors open to me whenever I want a talk. Their wisdom, knowledge and professional experience has been beneficial to this study in many aspects: research methodology, experiments, and thesis writing. It has been an honour to work with both of them during my candidacy.

I am indebted to all colleagues in my Lab for their participation in every presentation I made, and for their valuable comments and critical questions during this study. Special thanks are due to Ahmed Otoom for his useful help and suggestions for my research. My special thanks are also extended to the industrial partner, iOmniscient, for their financial and technical support of this research. I am also grateful to many anonymous reviewers across the world, for their fantastic comments on each of the manuscripts submitted for publication. Last but not least, I am deeply indebted to my family. Without their patient and considerable support, this research could not be completed.

# ABSTRACT

Video surveillance in dynamic scenes, especially for humans and vehicles, is currently one of the most active research topics in computer vision and pattern recognition. The goal of this research is to develop a real-time automatic tracking system which is both reliable and efficient by utilizing computational approaches. The literature has presented many valuable methods on object tracking; however, most of those algorithms can only perform effectively under simple scenarios. There are a few algorithms which attempt to accomplish object tracking in a complex dynamic scene and have successfully achieved their goals when the dynamic scene is not too complex. However no system yet is capable of accurately handling object tracking, especially human tracking, in a crowded environment with frequent and continuous occlusions. Therefore, the goal of my research is to develop an effective human tracking algorithm which takes into account and overcomes the various factors involved in a complex dynamic scene. The founding idea is that of dividing the human figure into five main parts, and track each individually under a constraint of integrity. Data association in new frames is performed on each part, and is inferred for the whole human figure through a fusion rule. This approach has proved a good trade off between model complexity and actual computability. Experimental results have confirmed the effectiveness of the methodology.

# TABLE OF CONTENTS

<i>List of Figures</i>	<i>IX</i>
<i>List of Tables</i>	<i>XV</i>
<i>Publications List</i>	<i>XVI</i>
<b>CHAPTER 1</b> <i>Introduction</i>	<i>1</i>
<b>1.1 Overview and Research Issues</b>	<b>1</b>
<b>1.2 Objectives</b>	<b>2</b>
<b>1.3 Significance</b>	<b>3</b>
<b>1.4 Thesis Structure</b>	<b>4</b>
<b>CHAPTER 2</b> <i>Related Work</i>	<i>7</i>
<b>2.1 Background Modelling</b>	<b>8</b>
2.1.1 Image Averaging	9
2.1.2 Single Gaussian Model	9
2.1.3 Gaussian Mixture Models	10
2.1.4 Kernel Density Estimation Methods	11
<b>2.2 Foreground Segmentation</b>	<b>12</b>
2.2.1 Background Subtraction	14
2.2.2 Temporal Differencing	15
2.2.3 Optical Flow	15
<b>2.3 Shadow Removal</b>	<b>16</b>
2.3.1 Statistical Approach	17
2.3.2 Deterministic Approach	17
<b>2.4 General Object Tracking</b>	<b>18</b>
2.4.1 Kalman Filter	20
2.4.2 Particle Filters	22
2.4.3 Multiple Hypothesis Tracking	24
2.4.4 Active Contour-based Tracking	24
2.4.5 Kernel-based Tracking	26
2.4.6 Blob-based Tracking	27

2.4.7 Other approaches	28
<b>2.5 Human Tracking</b>	<b>29</b>
2.5.1 Appearance models	30
2.5.2 3-Dimension Human Model	31
2.5.3 Body-part Model	32
2.5.4 Other Approaches	35
<b>2.6 Head/Face Detection</b>	<b>36</b>
2.6.1 Feature-invariant Detection	37
2.6.2 Template-based Detection	38
2.6.3 Appearance-based Detection	39
<b>2.7 Summary</b>	<b>39</b>
<b>CHAPTER 3 Methodology</b>	<b>41</b>
<b>3.1 Overview</b>	<b>42</b>
<b>3.2 Foreground Segmentation</b>	<b>44</b>
<b>3.3 Initialization</b>	<b>45</b>
<b>3.4 Prediction</b>	<b>46</b>
3.4.1 Gating	47
3.4.2 Target Dimension Prediction	49
<b>3.5 Occlusion Event Detection</b>	<b>51</b>
<b>3.6 Head Detection</b>	<b>52</b>
3.6.1 Appearance Model	53
3.6.2 Feature Model	54
<b>3.7 Construct The Human Model</b>	<b>55</b>
<b>3.8 Model Comparison</b>	<b>56</b>
3.8.1 Best Candidate Model Selection	56
3.8.2 Candidate Model Correction	58
<b>3.9 Model Updates</b>	<b>58</b>
<b>3.10 Summary</b>	<b>60</b>
<b>CHAPTER 4 Head Detection</b>	<b>62</b>
<b>4.1 Hair Colour Distribution Model</b>	<b>63</b>
<b>4.2 Skin Colour Distribution Model</b>	<b>70</b>
<b>4.3 Head Shape Modelling and Fitting</b>	<b>72</b>
4.3.1 Tilted-ellipse Fitting	72
4.3.2 Best-fitting Ellipse	73
<b>4.4 Head Detection Experiments</b>	<b>75</b>
4.4.1 General Head Images	76
4.4.2 Video Surveillance Datasets	79
<b>4.5 Further improvement</b>	<b>85</b>

4.5.1 Colour Modelling Based on MGM	86
4.5.2 Head Detection Based on SURF Features	86
<b>4.6 Summary</b>	<b>88</b>
<b>CHAPTER 5 <i>Human Model</i></b>	<b>90</b>
<b>5.1 Model Selection</b>	<b>91</b>
<b>5.2 Geometrical Construction</b>	<b>93</b>
<b>5.3 Feature Set</b>	<b>96</b>
5.3.1 Possible Features	96
5.3.2 Feature Selection	99
<b>5.4 Summary</b>	<b>101</b>
<b>CHAPTER 6 <i>Data Association</i></b>	<b>102</b>
<b>6.1 Local Match</b>	<b>103</b>
6.1.1 Comparison Between Features	104
6.1.2 Construction of Feature Difference PDF	106
<b>6.2 Global Match</b>	<b>109</b>
<b>6.3 Model Correction</b>	<b>109</b>
<b>6.4 Model Update</b>	<b>111</b>
<b>6.5 Summary</b>	<b>112</b>
<b>CHAPTER 7 <i>Experiments and Evaluation</i></b>	<b>114</b>
<b>7.1 Video Sets</b>	<b>115</b>
<b>7.2 Comparative Algorithms</b>	<b>116</b>
<b>7.3 Experiment Evaluation Methods</b>	<b>118</b>
7.3.1 Trajectory Accuracy Analysis	119
7.3.2 Segmentation Accuracy Analysis	120
7.3.3 Occlusions Events Analysis	121
7.3.4 Trajectory-based Analysis	122
<b>7.4 Experiment Results</b>	<b>123</b>
7.4.1 PETS2009 Sparse Crowd People Tracking Dataset	124
7.4.2 AVSS2007 I-LIDS Abandoned Baggage Detection Medium Density	158
7.4.3 ETISEO Metro Station subset	168
<b>7.5 Parameter Settings</b>	<b>173</b>
<b>7.6 Summary</b>	<b>176</b>
<b>CHAPTER 8 <i>Conclusions and Further Research</i></b>	<b>177</b>
<b>Appendix A</b>	<b>179</b>
<b>A.1 Software Interface</b>	<b>179</b>
<b>A.2 Running the Tracking System</b>	<b>185</b>

<b>A.3 System Architecture</b>	<b>189</b>
<b>A.4 System Flowchart</b>	<b>190</b>
<b><i>Bibliography</i></b>	<b><i>191</i></b>



# LIST OF FIGURES

<i>Figure 1-1 Thesis structure .....</i>	<i>6</i>
<i>Figure 3-1 Tracking procedure .....</i>	<i>44</i>
<i>Figure 3-2 Foreground segmentation with shadow removal. (a) the original image, (b) the foreground image .....</i>	<i>45</i>
<i>Figure 3-3 Track initialization. (a) manually draw the contour of the target, (b) extract the region of the selected target ( the blob of the targe) .....</i>	<i>46</i>
<i>Figure 3-4 Gating (red, squre bounding box). (a) normal event, (b) occlusion event ...</i>	<i>49</i>
<i>Figure 3-5 Peroidic motion of human movement.....</i>	<i>50</i>
<i>Figure 3-6 Human Model.....</i>	<i>55</i>
<i>Figure 4-1 Example of hair colour training samples manually extracted from a range of head images : (A) brown, (b) red, (c) blond and (D) black.....</i>	<i>65</i>
<i>Figure 4-2 PDF model of the black hair colour cateogry.....</i>	<i>66</i>
<i>Figure 4-3 PDF model of the brown hair colour category.....</i>	<i>67</i>
<i>Figure 4-4 PDF model of the brown hair colour category.....</i>	<i>68</i>
<i>Figure 4-5 PDF model of Red hair colour category.....</i>	<i>69</i>
<i>Figure 4-6 Example of skin colour training samples manually extracted from a range of head images under different illumination conditions: (a) bright, (b) standard, (c) dark .....</i>	<i>70</i>
<i>Figure 4-7 Illustration of the tilted-ellipse fitting procedure. (a) original iamge and (b) result of the head detection procedure .....</i>	<i>73</i>
<i>Figure 4-8 Illustration of the head shape modelling and fitting procedure. (a) result of head detection; (b) fitting procedure.....</i>	<i>75</i>

<i>Figure 4-9 Results of the head detector applied on images from the WWW.....</i>	<i>77</i>
<i>Figure 4-10 Results of the head detector applied on iamges from the FERET dataset .</i>	<i>78</i>
<i>Figure 4-11 Lab environment: results on manually segmented AVSS 2007 sequences.</i>	<i>81</i>
<i>Figure 4-12 Surveillance environment: results on surveillance video sequences taken from avss 2007 dataset .....</i>	<i>82</i>
<i>Figure 4-13 surveillance environment: results on surveillance video sequences taken from etiseo dataset.....</i>	<i>84</i>
<i>Figure 5-1 Various human model representations. (a) centroid, (b) elliptical bounding box, (c) rectangular bounding box, (d) articulated model, (e) feature points, (f) body contour.....</i>	<i>93</i>
<i>Figure 5-2 Human model .....</i>	<i>94</i>
<i>Figure 5-3 colour histogram comparison difference between observation with state.</i>	<i>100</i>
<i>Figure 5-4 Area comparison difference between observation with state.....</i>	<i>100</i>
<i>Figure 5-5 Perimeter comparison difference between observation with state.....</i>	<i>100</i>
<i>Figure 5-6 Displacement comparison difference between observation with state.....</i>	<i>101</i>
<i>Figure 5-7 Overlapping regions comparison difference between observation with state .....</i>	<i>101</i>
<i>Figure 6-1 Histogram of area difference .....</i>	<i>107</i>
<i>Figure 6-2 Histogram of perimeter difference .....</i>	<i>107</i>
<i>Figure 6-3 Histogram of color histogram difference .....</i>	<i>108</i>
<i>Figure 6-4 Histogram of displacement difference.....</i>	<i>108</i>
<i>Figure 6-5 Histogram of overlapping region difference.....</i>	<i>108</i>
<i>Figure 7-1 Opencv video surveillance system pipeline.....</i>	<i>116</i>
<i>Figure 7-2 Switching of algorithms in CCMSPF tracker .....</i>	<i>118</i>
<i>Figure 7-3 Tracking evaluation criteria (blue trajectory represents the tracking result, red trajectory represents the ground-truth) .....</i>	<i>120</i>
<i>Figure 7-4 Tracking Results from the proposed methodology.....</i>	<i>129</i>
<i>Figure 7-5 Comparison of target centre positions (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>130</i>
<i>Figure 7-6 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>130</i>

<i>Figure 7-7 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>130</i>
<i>Figure 7-8 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>131</i>
<i>Figure 7-9 Tracking Results from the proposed methodology.....</i>	<i>132</i>
<i>Figure 7-10 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>133</i>
<i>Figure 7-11 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>133</i>
<i>Figure 7-12 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>134</i>
<i>Figure 7-13 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>134</i>
<i>Figure 7-14 Tracking Results from the proposed methodology.....</i>	<i>135</i>
<i>Figure 7-15 Comparison of the target centre position.....</i>	<i>136</i>
<i>Figure 7-16 Comparison of x-coordinate of the target centre .....</i>	<i>137</i>
<i>Figure 7-17 Comparison of y-coordinate of the target centre .....</i>	<i>137</i>
<i>Figure 7-18 Comparison of target dimension differences .....</i>	<i>137</i>
<i>Figure 7-19 Tracking Results from the proposed methodology.....</i>	<i>139</i>
<i>Figure 7-20 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>140</i>
<i>Figure 7-21 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>140</i>
<i>Figure 7-22 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>140</i>
<i>Figure 7-23 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>141</i>
<i>Figure 7-24 Tracking Results from the proposed methodology.....</i>	<i>143</i>
<i>Figure 7-25 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>143</i>
<i>Figure 7-26 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>144</i>

<i>Figure 7-27 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>144</i>
<i>Figure 7-28 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>144</i>
<i>Figure 7-29 Tracking Results from the proposed methodology.....</i>	<i>146</i>
<i>Figure 7-30 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>147</i>
<i>Figure 7-31 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>147</i>
<i>Figure 7-32 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>147</i>
<i>Figure 7-33 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>148</i>
<i>Figure 7-34 Tracking Results from the proposed methodology.....</i>	<i>149</i>
<i>Figure 7-35 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>150</i>
<i>Figure 7-36 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>150</i>
<i>Figure 7-37 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>151</i>
<i>Figure 7-38 Comparison of target dimension differences (all dotted lines labelled as MS are multiple responses from the MS trackers) .....</i>	<i>151</i>
<i>Figure 7-39 Tracking Results from the proposed methodology.....</i>	<i>153</i>
<i>Figure 7-40 Comparison of the target centre position (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>153</i>
<i>Figure 7-41 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>154</i>
<i>Figure 7-42 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers).....</i>	<i>154</i>
<i>Figure 7-43 Comparison of target dimension differences .....</i>	<i>154</i>
<i>Figure 7-44 Tracking Results from the proposed methodology.....</i>	<i>156</i>
<i>Figure 7-45 Comparison of the target centre position.....</i>	<i>157</i>

<i>Figure 7-46 Comparison of x-coordinate of the target centre</i> .....	157
<i>Figure 7-47 Comparison of y-coordinate of the target centre</i> .....	157
<i>Figure 7-48 Comparison of target dimension differences</i> .....	158
<i>Figure 7-49 Tracking Results from the proposed methodology</i> .....	160
<i>Figure 7-50 Comparison of the target centre position</i> .....	161
<i>Figure 7-51 Comparison of x-coordinate of the target centre</i> .....	161
<i>Figure 7-52 Comparison of y-coordinate of the target centre</i> .....	161
<i>Figure 7-53 Comparison of target dimension differences</i> .....	162
<i>Figure 7-54 Tracking Results from the proposed methodology</i> .....	163
<i>Figure 7-55 Comparison of the target centre position</i> .....	164
<i>Figure 7-56 Comparison of x-coordinate of the target centre</i> .....	164
<i>Figure 7-57 Comparison of y-coordinate of the target centre</i> .....	164
<i>Figure 7-58 Comparison of target dimension differences</i> .....	165
<i>Figure 7-59 Tracking Results from the proposed methodology</i> .....	166
<i>Figure 7-60 Comparison of the target centre position</i> .....	166
<i>Figure 7-61 Comparison of x-coordinate of the target centre</i> .....	167
<i>Figure 7-62 Comparison of y-coordinate of the target centre</i> .....	167
<i>Figure 7-63 Comparison of target dimension differences</i> .....	167
<i>Figure 7-64 Tracking Results from the proposed methodology</i> .....	171
<i>Figure 7-65 Comparison of x-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers)</i> .....	172
<i>Figure 7-66 Comparison of y-coordinate of the target centre (all dotted lines labelled as MS are multiple responses from the MS trackers)</i> .....	172
<i>Figure 7-67 Comparison of target dimension differences</i> .....	173
<i>Figure A-1 Initial interface of the human tracker</i> .....	179
<i>Figure A-2 Video/sequence selection interface</i> .....	180
<i>Figure A-3 Object segmentation property interface</i> .....	181
<i>Figure A-4 Head detection property interface</i> .....	182
<i>Figure A-5 Object tracking property interface</i> .....	183
<i>Figure A-6 Display output selection interface</i> .....	184
<i>Figure A-7 Play the video/sequence in the human tracker interface</i> .....	185
<i>Figure A-8 Manually select the target</i> .....	186

*Figure A-9 Initialize the human model for the target..... 187*  
*Figure A-10 Track the target..... 188*  
*Figure A-11 system architecture ..... 189*  
*Figure A-12 System flowchart..... 190*

# LIST OF TABLES

<i>Table 3-1 Comparison with the particle filter.....</i>	<i>61</i>
<i>Table 4-1 Accuracy and false alarm rate on pixel-level averaged over 15 images: manual extraction vs the proposed algorithm.....</i>	<i>79</i>
<i>Table 4-2 Hit rate (HR) and false alarm rate (FAR) for both experiments .....</i>	<i>83</i>
<i>Table 4-3 Analysis of false alarm rate (FAR) and miss detection under different reasons .....</i>	<i>83</i>
<i>Table 4-4 Comparison between our algorithm (our) and the standard face detection algorithm (std).....</i>	<i>85</i>
<i>Table 7-1 Comparison between different trackers .....</i>	<i>125</i>
<i>Table 7-2 Comparison between different trackers (for severed occluded trajectories) .....</i>	<i>125</i>
<i>Table 7-3 Segmentation accuracy of different trackers .....</i>	<i>126</i>
<i>Table 7-4 Frequencies and performance on occlusion events for a total of 26 targets .....</i>	<i>127</i>
<i>Table 7-5 Precision measurements of different trackers.....</i>	<i>159</i>
<i>Table 7-6 segmentation accuracy of different tracker .....</i>	<i>169</i>
<i>Table 7-7 Frequencies and performance on occlusion events .....</i>	<i>169</i>
<i>Table 7-8 The value of important parameters tuned for each video set.....</i>	<i>173</i>

# PUBLICATIONS LIST

- **Z. Zhang**, O. P. Concha, and M. Piccardi (2010), Video tracking of people under heavy occlusions, to be submitted to *IEEE Transactions on Systems, Man and Cybernetics*.
- **Z. Zhang**, H. Gunes and M. Piccardi (2009), Head detection for video surveillance based on categorical hair and skin colour models, *The 2009 IEEE International Conference on Image Processing (ICIP 2009)*, 7-10 November 2009, Cairo, Egypt, 1137-1140.
- **Z. Zhang**, H. Gunes and M. Piccardi (2008), Tracking people in crowds by a part matching approach, *The Fifth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2008)*, 1-3 September 2008, Santa Fe, USA, 88-95.
- **Z. Zhang**, H. Gunes and M. Piccardi (2008), An accurate algorithm for head detection based on XYZ and HSV hair and skin color models, *The 2008 IEEE International Conference on Image Processing (ICIP 2008)*, 12-15 October 2008, San Diego, USA, 1644-1647.
- **Z. Zhang** and M. Piccardi (2007), A review of tracking methods under occlusions, *Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2007)*, 16-18 May 2007, Tokyo, Japan, 146-149.