AGENT–BASED 3D VISUAL TRACKING

Tak Keung CHENG

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

November, 1998
Revised on December, 1999
Final Revised on July, 2000

Department of Computer Science and Software Engineering
The University of Melbourne
This is to certify that

1. the thesis comprises only my original work,

2. due acknowledgement has been made in the text to all other material used,

3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies, appendices and footnotes.

Signed by Tak Keung CHENG
Abstract

We describe our overall approach to building robot vision systems, and the conceptual systems architecture as a network of agents, which run in parallel, and cooperate to achieve the system’s goals. We present the current state of the 3D Feature-Based Tracker, a robot vision system for tracking and segmenting the 3D motion of objects using image input from a calibrated stereo pair of video cameras.

The system runs in a multi-level cycle of prediction and verification or correction. The currently modelled 3D positions and velocities of the feature points are extrapolated a short time into the future to yield predictions of 3D position. These 3D predictions are projected into the two stereo views, and are used to guide a fast and highly focused visual search for the feature points. The image positions at which the features are re-acquired are back-projected in 3D space in order to update the 3D positions and velocities. At a higher level, features are dynamically grouped into clusters with common 3D motion. Predictions from the cluster level can be fed to the lower level to correct errors in the point-wise tracking.
Acknowledgments

I would like to thank my principal supervisor Dr Les Kitchen for providing his knowledgeable advice and guidance to my research during the course of the thesis. Without his help and guidance, this thesis would have not been possible. I would also like to thank Professor Zhi-Qiang Liu for his expert knowledge and advice through out my research work. A sincere thanks you to Dr James Cooper, who provided much assistance in the earlier stages of my candidature, Professors Peter Thorne and Professors Leon Sterling — Heads of the Department of Computer Science at the University of Melbourne — for their support and care during my candidature.

The great support by my colleagues (Andrew Howard, Nick Barnes, Soodamani Ramalingam, Sandy Dance and others) at the Computer Vision and Machine Intelligence Laboratory and Department of Computer Science, for providing a stimulating environment for my research.

I am extremely greatful for the financial support of both The ARC and The University of Melbourne. A big thank you to my dearest Venus and parents for their patience and unlimited support, without them I would not have achieved this.

Finally, thanks to a cast of thousands, who contributed bits and pieces to help me to finish this thesis.
Contents

Abstract i

Acknowledgments ii

Preface xiv

1 Introduction 1

1.1 Robot Vision ......................................................... 1

1.2 Visual Tracking ..................................................... 2

1.3 Problems and Philosophy ........................................... 5

1.4 Thesis’s Contribution ............................................... 9

1.5 Plan of the thesis ................................................... 10

2 Literature Review 12

2.1 Machine Visual Tracking System ................................. 12

2.2 Agents-Based Architecture ........................................ 15

2.3 Camera Calibration .................................................. 16

2.4 Motion Analysis ....................................................... 19

2.5 Convex Hull Algorithms ............................................ 20
# CONTENTS

## 3 System Architecture

3.1 “Task-Directed” Robot Vision Architecture ........................................... 22
3.2 Agent-Based Architecture ................................................................. 23
   3.2.1 Design Issues of Agent-based Architecture ................................. 24
3.3 System Design and Overview .............................................................. 26
3.4 Hardware and Software Setup ............................................................ 32

## 4 Camera Calibration and Converting Agent

4.1 Introduction ......................................................................................... 34
   4.1.1 Camera Calibration Agent ........................................................... 35
      4.1.1.1 Camera Model .................................................................. 35
      4.1.1.2 Nonlinear Camera Calibration Method ................................. 40
      4.1.1.3 Tsai’s Two Stage Calibration Method ................................. 42
   4.1.2 Converting Agent .......................................................................... 44
4.2 Design and Implementation ................................................................. 45
4.3 Experimental Results ........................................................................... 46
4.4 Discussion ........................................................................................... 52

## 5 Feature and Replenishment Agents

5.1 Introduction ......................................................................................... 54
   5.1.1 Stereo Feature Agent .................................................................... 54
      5.1.1.1 Models for Feature Tracking .............................................. 58
   5.1.2 Feature Replenishment Agent ....................................................... 61
   5.1.3 Threshold Calibration .................................................................... 63
CONTENTS

5.2 Experimental Results .................................................. 67
5.3 Data Communications .................................................. 70
5.4 Discussion ................................................................. 70

6 Motion Estimation Agent ................................................. 73
6.1 Introduction ............................................................... 73
   6.1.1 General Motion Estimation Technique ...................... 74
   6.1.2 Kalman Filtering .................................................. 76
6.2 Data Communications .................................................. 78
6.3 Experimental Results and Discussion .............................. 79

7 Clustering Agents .......................................................... 83
7.1 Introduction ............................................................. 83
   7.1.1 Motion-Based Clustering .................................... 84
   7.1.2 Proximity Position-Based Clustering ..................... 85
   7.1.3 High-Level Clustering (HLC) ............................... 90
7.2 Feature Recovery Mechanism and Other Functions .......... 92
7.3 Data Communications ................................................ 93
7.4 Experimental Results and Discussion ......................... 95
   7.4.1 Data Simulation Experiment ............................... 95
   7.4.2 Real Image Experiment ...................................... 98
   7.4.3 Fuzzy and Motion Clustering Experiments ............ 104
7.5 Conclusion .............................................................. 107

8 Convex Hulls for Moving Sets of 2D Points ...................... 114
8.1 Introduction .................................................. 114
8.2 Mergehull algorithm ............................................. 116
8.3 Iterative convex hull algorithm .............................. 117
8.4 Experimental results and discussion .......................... 120
8.5 Enhancing Feature Replenishment by Convex Hull ............. 125
8.6 Conclusions .................................................. 131

9 Experimental Results for 3DFBT .......................... 132
  9.1 Multiple Moving Objects in Off-Line Mode ................. 132
  9.2 Multiple Objects in Real-Time Mode ........................ 134
  9.3 Mobile Robot with 3DFBT .................................. 138
  9.4 3DFBT using Different Clustering Agent ................... 142

10 Conclusion and Future Work .............................. 155
  10.1 Conclusion ............................................... 155
  10.2 Future Work ............................................. 157

Bibliography .................................................. 159

A Published Papers ........................................... 177
List of Figures

1.1 “Traditional” Computer Vision. ................................. 5
1.2 Interaction in Time. ........................................... 7
1.3 Prediction/Verification Cycle. ............................... 8

3.1 Communication path between three people ..................... 25
3.2 Languages model used for the translation ...................... 26
3.3 System Design for Object Recognition ........................ 27
3.4 System Architecture for 3D Feature Based Tracker ............ 28
3.5 Interface of Display agent ..................................... 32

4.1 Pinhole Camera Model .......................................... 36
4.2 2D image coordinate system .................................... 38
4.3 Data communication path between three front-end agents .... 46
4.4 Set up for calibration and the world coordinates used for the experiments. 48
4.5 Test object for the camera calibration process in experiment 1 (Baseline is 30cm.) ............................................. 48
4.6 Test object for the camera calibration process in experiment 2 (Baseline is 23cm.) ............................................. 49
4.7 Test object for the camera calibration process in experiment 3 (Baseline is 13cm.) 

4.8 Test object for the camera calibration process in experiment 4 (Baseline is 5.3cm.) 

4.9 The 3D space RMS error for four baseline experiments. 

4.10 The 2D space RMS error for four baseline experiments 

5.1 Epi-polar and depth constraints. 

5.2 Matching window which is set up by epi-polar constraintst 

5.3 Prediction/Verification Cycle for vector-based feature tracking. 

5.4 16 Blocks with labels. 

5.5 Left Image. 

5.6 Right Image. 

5.7 Normalise error value vs Feature. 

5.8 Percentage of accepted features vs threshold value. 

5.9 Image sequence: From the top to bottom — frame numbers 1, 15, 30, 50 

5.10 Data Communication Path for Stereo Feature agent. 

6.1 Rigid Body Motion in 3D space. 

6.2 Data Communication Path for Motion Estimation and Clustering agent. 

6.3 Summary of Kalman Filter Experiment. 

6.4 Summary of Newton Raphson Experiment. 

6.5 Summary of Error Analysis on noise data. 

7.1 Data Communication Path for internal Clustering agents group.
LIST OF FIGURES

7.2 Test object for the camera calibration process. ........................ 99
7.3 Set up for calibration and the world coordinates used for this experiment. 99
7.4 Stereo pair shot when the process started. ............................... 101
7.5 Stereo pair shot when the process finished. ............................... 102
7.6 Experiments for Single Object Simulation (Number of Objects). ...... 106
7.7 Experiments for Single Object Simulation (Object’s membership). .... 107
7.8 Experiments for Two Object Simulation (Number of Objects). ......... 108
7.9 Experiments for Two Object Simulation (Object A’s membership). .... 109
7.10 Experiments for Two Object Simulation (Object B’s membership). ... 109
7.11 Experiments for Four Object Simulation (Number of Objects). ......... 110
7.12 Experiments for Four Object Simulation (Object A’s membership). ... 111
7.13 Experiments for Four Object Simulation (Object B’s membership). ... 111
7.14 Experiments for Four Object Simulation (Object C’s membership). ... 112
7.15 Experiments for Four Object Simulation (Object D’s membership). ... 112

8.1 Variation 1: 50000 feature points (Gaussian Distribution) .............. 120
8.2 Variation 2: 50000 feature points (Gaussian Distribution) .............. 121
8.3 Variation 1: Summary for Experiment Results ............................. 122
8.4 Variation 2: Summary for Experiment Results ............................. 122
8.5 Variation 1: 50000 feature points (Uniform Distribution) ............... 123
8.6 Variation 2: 50000 feature points (Uniform Distribution) ............... 123
8.7 Variation 1: Experiment results for uniform distribution ................ 124
8.8 Variation 2: Experiment results for uniform distribution ................ 124
8.9 Situation when our method cannot detect any new features. ............. 126
LIST OF FIGURES

8.10 Convex Hulls for trackable object ........................................ 127
8.11 Convex Hulls Feature Replenishment for boundary objects .......... 128
8.12 Convex Hulls Feature Replenishment for occlusion objects .......... 129
8.13 Convex Hulls Feature Replenishment for moving occluded objects ... 130

9.1 Experimental set-up .......................................................... 133
9.2 Camera calibration object ...................................................... 133
9.3 Image sequence: (TOP) stereo image taken when tracking started; (BOT-
   TOM) stereo image taken when tracking ended .......................... 135
9.4 Train Set: (TOP) Images taken when system started up; (BOTTOM)
   Images taken at frame number 25 ........................................... 137
9.5 Image sequence: (TOP) stereo image taken when tracking started; (BOT-
   TOM) stereo image taken when tracking ended ......................... 139
9.6 J. Edgar Mobile robot ......................................................... 141
9.7 Robot Set 1: Mobile robot movement direction ........................ 142
9.8 Robot Set 1: (TOP) Images taken when system started up; (BOTTOM)
   Images taken at frame number 10 ......................................... 143
9.9 Robot Set 2: Mobile robot movement direction ........................ 144
9.10 Robot Set 2: (TOP) Images taken when system started up; (BOTTOM)
   Images taken at frame number 15 ....................................... 145
9.11 Images taken when system started up ..................................... 146
9.12 Setting 1: Images taken at frame number 35 .......................... 147
9.13 Setting 2: Images taken at frame number 35 .......................... 148
9.14 Setting 3: Images taken at frame number 35 .......................... 148
9.15 Setting 4: Images taken at frame number 35. ............... 149
9.16 All agents Experiment ........................................... 151
9.17 Fuzzy Clustering Experiment ...................................... 151
9.18 Kalman Engine Experiment ........................................ 151
9.19 Motion Estimation Experiment .................................... 151
9.20 All agents Experiment ............................................. 152
9.21 Fuzzy Clustering Experiment ...................................... 152
9.22 Kalman Engine Experiment ....................................... 152
9.23 Motion Estimation Experiment .................................... 152
## List of Tables

4.1 Two sets of camera parameters are found in experiment 1 (Baseline is 30cm.) 48

4.2 Two sets of camera parameters are found in experiment 2 (Baseline is 23cm.) 49

4.3 Two sets of camera parameters are found in this experiment 3 (Baseline is 13cm.) 50

4.4 Two sets of camera parameters are found in this experiment 4 (Baseline is 5.3cm.) 50

4.5 The RMS error for four baseline experiments. 50

5.1 Normalised error values for three kinds of threshold analysis 65

5.2 Number of trackable and new features in each frame 68

7.1 The motion of the ‘objects’ which are generated by data generator. 97

7.2 The objects and their members which were found by the system during the experiment. 97

7.3 Two sets of camera parameters used in this experiment. 98

7.4 Information for all features when the process is started. 102

7.5 The objects and their members which were found by the system during the experiment. 103
7.6  Comparison of the 3D predicted position and actual position found by the
      system for tracked features in frame number 5 of example sequence. . . . 104

9.1  The objects and their members found by the system. . . . . . . . . . . . . 136
9.2  The objects and their members found by the system. . . . . . . . . . . . . 140
9.3  3DFBT Setting for four experiments . . . . . . . . . . . . . . . . . . . . . . 146
9.4  Setting 1: Clustering and Motion Estimation results in frame 35 . . . . . 149
9.5  Setting 2: Clustering and Motion Estimation results in frame 35 . . . . . 150
9.6  Setting 3: Clustering and Motion Estimation results in frame 35 . . . . . 150
9.7  Setting 4: Clustering and Motion Estimation results in frame 35 . . . . . 150
Preface

Most of the work presented in this thesis has been based on previously published papers. The concept and architecture of the system in Chapter 1 and Chapter 3 are largely based on the work reported in the 2nd Asian Conference in Computer Vision [27], the 1995 National Conference of the Australian Robot Association [28] and DICTA-93 Digital Image Computing: Techniques and Applications [23]. The contents of Chapter 5 has been reported partly in the 3rd International Computer Science Conference [24] and 1995 National Conference of the Australian Robot Association [28]. The work in Chapter 6, part of Chapter 7 and Chapter 9 has been reported partly in the 18th Australasian Computer Science Conference [26], the Australasian Workshop on Parallel and Real-Time System [83] and the 2nd Asian Conference in Computer Vision [25].

The aforementioned publications are all multi-authored. This thesis contains only the work I contributed to all these papers. The rest of the thesis is based on my own work during the candidature except where due acknowledgment has been made in the content of the thesis.
Author/s:
Cheng, Tak Keung

Title:
Agent-based 3d visual tracking

Date:
2000-07

Citation:

Publication Status:
Unpublished

Persistent Link:
http://hdl.handle.net/11343/39431

File Description:
Intro.

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.