Enhancement of visual sensing and servoing for micromanipulation

Sun, Wen Ting

2008

Sun, W. T. (2008). Enhancement of visual sensing and servoing for micromanipulation.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/41440>

<https://doi.org/10.32657/10356/41440>

Enhancement of Visual Sensing and Servoing for Micromanipulation

Wenting Sun

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University

in fulfilment of the requirement for the degree of

Doctor of Philosophy

2008

Acknowledgement

At the outmost I would like to express my deepest gratitude to my supervisor, Associate Professor Chin Teck Chai, for his support, guidance, understanding, inspiration, enthusiastic help and invaluable advice, for allowing me considerable freedom in conducting this research. I also wish to express my sincere gratitude to Associate Professor Rajagopalan Devanathan, for his encouragement and invaluable suggestions during the first year of my candidature.

I am grateful to Dr Andrew Shacklock, who has contributed his useful guidance and helpful directions during the research stage. I benefited a lot from his invaluable advice and academic spirit. I am also thankful to Dr Luo Hong, Mr Wong Lye Seng in Mechatronics group, Singapore Institute of Manufacturing Technology for their beneficial suggestions and warm help in my experiments.

I would like to express my appreciation to School of Electrical and Electronic Engineering of Nanyang Technological University, Singapore for its financial support and the necessary research facilities.

Last but not least I acknowledge the continual encouragement and support provided by my parents and friends.

Summary

Micromanipulation is the technology to handle objects with micro scale and/or meso scale features under micro scale tolerance, where $10^{-6}m$ to $10^{-4}m$ is referred as "micro scale" and $10^{-4}m$ to $10^{-3}m$ is referred as "meso scale" [1]. Micromanipulation is critical to many growing technologies, such as microsurgery, microelectronics, MEMS, life sciences etc. However it is very difficult for the operator to carry out micromanipulation tasks, he does not have direct access to the micro world, either can he feel or hear a tool interacting with the target, so that sensory input such as vision and force sensing has to be relied on. On the other hand, the scale difference between the micro world and the macro world, the reduced depth of field of micro imaging system make perception a particular problem for micromanipulation. Moreover, the need for large travel ranges and reconfigurable systems introduces kinematics uncertainties, which make the automated operations for micromanipulation such as positioning, pick and place difficult especially when large motion range and small positioning accuracy are both required. Hence, there is a necessity to improve the perception and automating many of the low level tasks to assist the human.

In the literature, several image processing systems have been established for micromanipulation tasks. Most existing methods addressed matching between several frames of images. However, in real situations, long sequence of images plays very important role in monitoring the whole process of manipulation. There are also a

few coarse-fine image based visual servoing systems developed for micromanipulation. However, the image based systems are not well established, few approaches have been proposed to study formally the 3-D parameter estimation [2], especially when image based visual servoing is proposed for micromanipulation, the loss of 3-D information makes the depth loss in coarse phase a very difficult problem. Furthermore, most previous work in image based visual servoing shows the feasibility of visual servo control for stationary targets, a control law for moving target scenarios has not been rigorously developed [3], in such case, target motion in fine phase faces challenges in image based paradigm. The potentials of vision based techniques in assisting and improving micromanipulation need to be explored. In this thesis, we investigated possible vision based techniques for assisting and enhancing micromanipulation process and developed approaches aiming to improve visual tracking and servoing in micromanipulation.

There is evidence that features with redundancy have the advantages in solving correspondence problems (identification of corresponding features over sensor displacements). However, geometric constraints on end points of redundant features are not available. Integrating points and redundant features will be a key factor for achieving improvement in recognition performance that rely on single type of features. In this thesis, two novel line-point integration micro image matching approaches are proposed. Making use of the redundancy provided by the line segments, the primitive similarity comparison is carried out, the distance measure that integrates both line segment and point features is proposed to recover the underlying transformation between every two consecutive micro images. Robust estimation and guided matching are applied to increase the number of matched features as well as to refine the transformation. Simulation studies have been conducted in this thesis to investigate the requirements for successfully implementation of the

proposed method. Based on the matching results, the kinematics parameters are tracked through a combined framework of homography decomposition and Bayesian estimation. Experiments with MEMS wafer images are produced to validate the proposed method. The proposed visual matching and tracking strategy in this thesis is able to track up to 1000 micro images and enlarge the field of view to 22 times of the original single microscopic image. Comparing to previous matching method for micromanipulation which matches pairs of micro images, the results obtained by the proposed approach in this thesis lead to a much larger recognizable area under the microscope, and constitutes much more prior information that is important for further localization or manipulation in the micro environment. The study has also shown close relation between matching performance with the feature quality, correct putative matching rate and feature density.

Grounded on the studies in the traditional image based visual servoing, the extent of improvements in coarse-fine image based visual servoing micromanipulation systems is explored. A novel Kalman filter based depth tracking strategy (KFRS) is proposed where both the depth of the observed features and the depth changing rate are formulated as the state vector, and estimated online. The proposed strategy is applicable to coarse-fine image based visual servoing micromanipulation so that it is able to drive the manipulation stage to the focus plane of the microscope during coarse phase. To compensate target motion in fine phase, the strategy of image based visual servoing with motion compensation is proposed, where a novel formulation of the image jacobian is developed incorporating the motion parameters, so that the unknown target motion can be compensated during the servoing. Simulation studies are presented for both the depth adaptive approach and the motion compensation approach as well as their implementation into the coarse-fine micromanipulation. Without exact knowledge of the depth during coarse phase servoing,

the proposed strategy in this thesis with avoidance procedures of near singular situations is able to achieve accurate positioning results while avoiding large retreat motion of conventional methods. The positioning of field of view onto the interested area can be achieved regardless of the target motion during fine phase servoing for micromanipulation applications with the proposed servoing strategy for moving target.

This thesis constitutes initial effort toward the development of improved visual sensing and servoing system for micromanipulation in a coarse-fine paradigm. The performance of the proposed system is partially validated with experimental test rig as well as through simulation studies. The results show promising direction in assisting human operator and improving operation process for micromanipulation with visual sensing and servoing enhancement.

Table of Contents

Acknowledgement	i
Summary	ii
List of Tables	xii
List of Figures	xxiv
1 Introduction	1
1.1 Motivation	1
1.2 Significance of This Research	6
1.3 Scope and Objectives	9
1.4 Major Contributions of This Thesis	11
1.5 Thesis Outline	12
2 Literature Review	14
2.1 State of The Art	15

TABLE OF CONTENTS	vii
2.1.1 Functionality of Micromanipulation Systems	15
2.1.2 Existing Micromanipulation Systems	17
2.1.3 Summery	24
2.2 A Review of Visual Tracking	24
2.2.1 Image Matching	24
2.2.2 Image Tracking	26
2.2.3 Summery	28
2.3 A Review of Visual Servoing	28
2.3.1 Image Based Visual Servoing (IBVS)	29
2.3.2 Depth Adaptive Image Based Visual Servoing	30
2.3.3 Dynamic Image Based Visual Servoing	33
2.3.4 Summary	33
3 Micro Image Matching	35
3.1 Introduction	35
3.2 Traditional Approaches	36
3.3 Line-Point Integrated Micro Image Matching	42
3.3.1 Line Segment Based Line-Point Matching (LSG)	43
3.3.2 L-shape Based Line-Point Matching (LSH)	68
3.3.3 Comparison Between LSG and LSH	81
3.4 Evaluation of Line-Point Integrated Matching Approaches	85

TABLE OF CONTENTS **viii**

3.4.1	Feature Quality Requirements	85
3.4.2	Putative Matching Rate Requirements	91
3.4.3	Feature Density Requirements	96
3.5	Summary	102
4	Micro Image Tracking	104
4.1	Introduction	104
4.2	Bayesian Tracking	108
4.3	Homography Based Micro Image Tracking (HBT)	114
4.3.1	Tracking of Translational Motions	115
4.3.2	Tracking of Rotational Motions	119
4.4	Implementation and Experiment Results	126
4.5	Summary	137
5	Depth Adaptive Coarse-Fine Image Based Visual Servoing With Motion Compensation (DAMC) For Micromanipulation	140
5.1	Introduction	140
5.2	Proposed Depth Adaptive Approach	143
5.2.1	Notation	143
5.2.2	Proposed Strategy	145
5.2.3	Simulations	151
5.2.4	Discussion	164

TABLE OF CONTENTS	ix
5.3 Proposed Motion Compensation Approach	164
5.3.1 A Brief Review of Rigid Body Velocity	164
5.3.2 Proposed Strategy	166
5.3.3 Simulations	170
5.3.4 Discussion	182
5.4 DAMC For Micromanipulation	182
5.4.1 Implement Depth Adaptive Approach into Coarse Phase . . .	182
5.4.2 Implement Motion Compensation Approach into Fine Phase .	186
5.4.3 Simulation and Results	191
5.5 Summary	198
6 Conclusions	199
6.1 Summary of Research	199
6.2 Summary of Achievements	200
6.3 Recommendation for Future Research	201
Author's Publications	204
References	207
Appendices	229
A Estimating the homography	229

Contents x

A.1 Planar Homography	229
A.2 Inter-image Homography	230
A.3 Homography Estimation	232
A.3.1 Pseudo Inverse	232
A.3.2 Singular Value Decomposition	233
A.3.3 Non-linear Geometric Solution	234

List of Tables

3.1	Processing Steps For LSG	44
3.2	Processing Steps For LSH	69
3.3	Comparison of LSG and LSH. SR: sample required. IT: inlier threshold. NL: number of line matches. NP: number of point matches. RE: RMS error. TC: time cost.	84
3.4	Monte Carlo simulation steps for evaluating line-point integration based matching approaches with different feature quality	86
4.1	Generic particle filter algorithm [4]	112
4.2	Resampling algorithm [4]	113
4.3	Drawing samples for translational motion	117
4.4	Drawing samples for rotational motion	125
5.1	The <i>reset</i> algorithm	150
5.2	Initial Camera Posture	152
5.3	Desired camera posture ($\theta^* = [x^*, y^*, z^*]^T$, $\mathbf{c}^* = [\omega_x^*, \omega_y^*, \omega_z^*]^T$) and final reached camera posture ($\theta^f = [x^f, y^f, z^f]^T$, $\mathbf{c}^f = [\omega_x^f, \omega_y^f, \omega_z^f]^T$) . .	163

LIST OF TABLES **xii**

5.4 Initial Camera Posture 171

5.5 Coarse phase image based visual sevoing with depth adaption algrithm185

5.6 Fine phase image based visual servoing with motion compensation . . 190

List of Figures

1.1	(a) the whole wafer (b) the microscopic view of the wafer	3
1.2	The process flow of coarse-fine image based visual servoing system for micromanipulation	5
1.3	Concept of navigation with microscope	7
1.4	Illustration of coarse-fine micromanipulation	8
2.1	Functionality of micromanipulation systems	16
2.2	Set-up of the nanorobot system [5]	18
2.3	Autonomous embryo injection system [6]	19
2.4	Microscope and micromanipulator [7]	20
2.5	Sequencing of the different operating microassembly tasks [8]	21
2.6	Laboratory arrangement of a fine-course feed system on the milling center: 1, CNC machine tool; 2, 6-DOF platform; 3, hydraulic control arrangement; 4, computer for the platform control; 5, display of the measuring device. [9]	21
2.7	Positioning system configuration	22

LIST OF FIGURES

xiv

2.8	Position based system	29
2.9	Image based system	29
3.1	(a)(b) image 1 and image 2 from a sequence of the MEMS wafer. The images are 484×648 pixels.	37
3.2	(a) gradient in x ($\frac{\partial I}{\partial x}$), (b) gradient in y ($\frac{\partial I}{\partial y}$), (c) gradient in t ($\frac{\partial I}{\partial t}$), (d) flow field	37
3.3	Corners detected. There are 300 corners in each image	38
3.4	matches found by point cross correlation	38
3.5	(a)The grayscale image of the template (b) The binary image of the template	39
3.6	Binary images of the original grayscale images shown in Figure 3.1	40
3.7	Correlation shown in frequency domain	40
3.8	The correlation peaks found in the images. The color channels are allocated to show the matching results. Red Channel: the original image; Green, Blue Channels: correlation	41
3.9	Matched template shown in each of the images	41
3.10	Illustration of correlation for I and I'_i	45
3.11	Illustration of neighboring support for a candidate match	47
3.12	Illustration of different distance measure computation strategy	53
3.13	Truncation error in line segment extraction	55
3.14	(a) image 1 (b) image 2	55

LIST OF FIGURES

xv

3.15	Simulated end points on the line segments in image 1 and image 2 . . .	56
3.16	Distance computation with end points (a) distance (b) histogram of distance	56
3.17	Distance computation with line parameters (a) distance (b) histogram of distance	57
3.18	Distance computation with line parameters and center points of line segments (a) distance (b) histogram of distance	57
3.19	Distance computation with Simplified Segment Hausdorff distance (a) distance (b) histogram of distance	58
3.20	Difference between symmetric transfer error and reprojection error. (a) symmetric transfer error (b) reprojection error. $\mathbf{x}$ and $\mathbf{y}$ are mea- sured points, under the estimated homography, the points $\mathbf{y}$ and $\hat{\mathbf{y}} = \mathbf{H}\mathbf{x}$ do not correspond perfectly. The estimated points $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, do correspond perfectly by the homography $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$. Using the notation $d(\mathbf{x}, \mathbf{y})$ for the Euclidean image distance between $\mathbf{x}$ and $\mathbf{y}$, the symmetric transfer error is $d(\mathbf{x}, \mathbf{H}^{-1}\mathbf{y})^2 + d(\mathbf{y}, \mathbf{H}\mathbf{x})^2$, the repro- jection error is $d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{y}, \hat{\mathbf{y}})^2$	60
3.21	(a)(b) detected line segments superimposed on the images. There are 120 line segments on image 1 and 131 line segments on image 2. (c)(d) 64 putative matches shown by numbers. (e)(f) inliers - 8 correspondences consistent with the estimated homography.	62
3.22	(a)spot detection in image 1. (b)spot detection in image 2.	63

LIST OF FIGURES**xvi**

3.23 (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 17 point matches and 19 line correspondences.	63
3.24 (a) point matches shown by the arrow linking point features. (b) yellow segments: line segments in image 1, red segments: line segments in image 2 projected back into image 1 according to the estimated homography.	64
3.25 (a)(b) detected line segments superimposed on the images. There are 116 line segments on image 1 and 106 line segments on image 2. (c)(d) 51 putative matches shown by numbers. (e)(f) inliers - 5 correspondences consistent with the estimated homography.	65
3.26 (a) spot detection in image 1. (b) spot detection in image 2.	66
3.27 (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 32 point matches and 16 line correspondences.	66
3.28 (a) point matches shown by the arrow linking point features. (b) yellow segments: line segments in image 1, red segments: line segments in image 2 projected back into image 1 according to the estimated homography.	67
3.29 L-shape structure grouped with elongated line segments and point . .	70
3.30 The relation of intersection between two coplanar lines in two images	72

LIST OF FIGURES

xvii

- 3.31 (a)(b) grouped L-shapes (red) based on the detected line segments (yellow) superimposed on the images. There are 258 line segments and 71 L-shapes on image 1, 255 line segments and 78 L-shapes on image 2. (c)(d) 35 putative matched L-shapes shown by numbers. (e)(f) inlier - 7 corresponding L-shapes consistent with the estimated homography. 76
- 3.32 (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 17 point matches and 45 line correspondences. 77
- 3.33 (a) point matches shown by the arrow linking point features. (b) yellow segments: matched line segments in image 1, red segments: matched line segments in image 2 projected back into image 1 according to the estimated homography. 77
- 3.34 (a)(b) grouped L-shapes (red) based on the detected line segments (yellow) superimposed on the images. There are 170 line segments and 44 L-shapes on image 1, 158 line segments and 41 L-shapes on image 2. (c)(d) 8 putative matched L-shapes shown by numbers. (e)(f) inlier - 5 corresponding L-shapes consistent with the estimated homography. 79
- 3.35 (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 28 point matches and 19 line correspondences. 80

LIST OF FIGURES

xviii

3.36 (a) point matches shown by the arrow linking point features. (b) yellow segments: matched line segments in image 1, red segments: matched line segments in image 2 projected back into image 1 according to the estimated homography.	80
3.37 Sets of micro images for testing	83
3.38 (a) Distribution of RMS error wrt. line segment noise (b) Distribution of RMS error wrt. point noise	87
3.39 (a) RMS error wrt. changing of feature noise with initial correct putative matching rate 70%. (b) RMS error wrt. changing of feature noise with initial correct putative matching rate 50%. (c) RMS error wrt. changing of feature noise with initial correct putative matching rate 20%.	88
3.40 (a) Distribution of inlier rate wrt. the noise on end point of line segments (b) Distribution of inlier rate wrt. the noise on points . . .	89
3.41 (a) Distribution of time cost wrt. the noise on end point of line segments (b) Distribution of time cost wrt. the noise on points . . .	90
3.42 Distribution of RMS error under feature quality changes	91
3.43 rate of final matched points vs. rate of correct putative point matches	93
3.44 rate of final matched line segments vs. rate of correct putative line matches	93
3.45 RMS error with respect to rate of putative matches of line segments and points	94

LIST OF FIGURES

xix

3.46 (a) Distribution of time cost wrt. rate of correct putative line matches (b) Distribution of time cost wrt. rate of correct putative point matches	95
3.47 Distribution of RMS error under changing of putative matching rate .	95
3.48 (a) Distribution of RMS error wrt. No of line features (b) Distribution of RMS error wrt. No of point features	97
3.49 (a) RMS error wrt. changing of feature density with initial correct putative matching rate 70% (b) RMS error wrt. changing of feature density with initial correct putative matching rate 50%	99
3.50 Distribution of No of line segment inliers wrt. No of line features . .	100
3.51 (a) Distribution of time cost wrt. No of line features (b) Distribution of time cost wrt. No of point features	100
3.52 Distribution of RMS error with feature density change	101
4.1 Illustration of visual tracking for micro manipulation	106
4.2 Translational scenario	107
4.3 Rotational scenario	108
4.4 Illustration of rotation estimation from small motion vectors. $\vec{a}$, $\vec{b}$: true motion vectors, $\vec{a}'$, $\vec{b}'$: detected motion vectors, O : true rotation axis, O' : estimated rotation axis.	121
4.5 The comparison between large and small motion vector in estimating rotation axis. The red ellipse shows the 95% confidence interval. . . .	121
4.6 Illustration of rotation parameter reconstruction from homography . .	123

LIST OF FIGURES**xx**

4.7	Micro motion workstation with multiple visual sensors	127
4.8	Micro image positions, dots: particles, ellipse: 95% confidence interval for each iteration	129
4.9	HBT estimation of state vectors for translational motion	129
4.10	(a)Histogram of inter image translational motion in x (b)Histogram of inter image translational motion in y	130
4.11	Mosaic made up of 40 frames micro images with translational motion. (a) With HBT. (b) Without HBT.	130
4.12	Enlarged field of view shown in perspective view for translational motion with HBT.	131
4.13	Particle cloud spread of rotation axis through all the iterations, dots: particles, ellipse: 95% confidence interval for each iteration. Darker points: the particles from the last iteration	132
4.14	Particle cloud spread of rotation angle through all the iterations. Darker points: the particles from the last iteration.	132
4.15	HBT estimation of state vectors a_1 vs. a_2 for rotational motion . . .	133
4.16	HBT estimation of state vectors a_1 vs. θ for rotational motion	133
4.17	HBT estimation of state vectors a_2 vs. θ for rotational motion	134
4.18	Histogram of inter image rotational axis estimation (a) a_1 (b) a_2 . . .	134
4.19	Histogram of inter image rotational motion in θ	134
4.20	Mosaic made up of 40 micro image frames with rotational motion (a) With HBT (b) Without HBT.	135

LIST OF FIGURES

xxi

4.21	Enlarged field of view shown in perspective view for rotational motion with HBT.	136
4.22	Mosaic made up of 1000 frames micro images with translational motion with HBT compared to single micro image field of view at the same scale. The covered area is 22 times of the single micro image. (a) Mosaic of 1000 frames (b) One of the single image	137
5.1	Illustration of problem formulation	143
5.2	Camera view comparison.(example 1)	153
5.3	Image points tracking errors. solid line: GD, dashed line: KFRS (example 1)	154
5.4	Camera posture evolution in the absolute frame. left: GD, right:KFRS (example 1)	155
5.5	Comparison of evolution of camera posture, left: translational posture, right: rotational posture. solid line: GD, dotted line: KFRS. (example 1)	156
5.6	Adapted depth (example 1)	156
5.7	Camera view comparison. (example 2)	157
5.8	Image points tracking errors. solid line: GD, dashed line: KFRS (example 2)	157
5.9	Camera posture evolution in the absolute frame. left: GD, right:KFRS (example 2)	158

LIST OF FIGURES

xxii

5.10 Comparison of evolution of camera posture, left: translational posture, right: rotational posture. solid line: GD, dotted line: KFRS. (example 2)	159
5.11 Adapted depth (example 2)	159
5.12 Camera view of the image points evolution (example 3)	160
5.13 Image points tracking error (example 3)	160
5.14 Evolution of camera posture, left: translational posture, right: rotational posture. (example 3)	161
5.15 Camera motion w.r.t the object. (example 3)	162
5.16 Depth adaption comparison between with and w/o reset process. (example 3)	162
5.17 Illustration of problem formulation	166
5.18 2D trajectory in the image plane. (example 1) 'o' initial features, '+'target features.	172
5.19 Image feature tracking error. (example 1)	172
5.20 Estimated object moving velocity w.r.t $\mathcal{F}^k$. (example 1)	173
5.21 Estimated depth (example 1)	173
5.22 Camera trajectory (example 1)	174
5.23 Evolution of camera motion w.r.t the moving object.(example 1) . . .	175
5.24 Evolution of the camera translational posture difference w.r.t the moving target.(example 1)	176

LIST OF FIGURES

xxiii

5.25 Evolution of the camera rotational posture difference w.r.t the moving target. (example 1)	176
5.26 2D trajectory in the image plane. (example 2) 'o' initial features, '+'target features	177
5.27 Image feature tacking error. (example 2)	177
5.28 Estimated object moving velocity w.r.t $\mathcal{F}^k$. (example 2)	178
5.29 Estimated depth (example 2)	178
5.30 Camera trajectory (example 2)	179
5.31 Evolution of camera motion w.r.t the moving object. (example 2) . .	180
5.32 Evolution of the camera translational posture difference w.r.t the moving target. (example 2)	181
5.33 Evolution of the camera rotational posture difference w.r.t the moving target. (example 2)	181
5.34 Simplified Ray Diagram for Typical Optical Microscope	186
5.35 (a) Simulated environment for coarse fine micromanipulation (b) trajectory of the grid pad during coarse servoing	192
5.36 (a) initial macro view image, red rectangle: macro image features (b) target macro view image, red cross: macro image features	193
5.37 The adapted depth during macro view image based visual servoing . .	194
5.38 Macro image feature trajectory. 'o' initial features, '+'target features	195
5.39 Macro image feature error	195
5.40 (a) initial micro view image (b) target micro view image	196

LIST OF FIGURES

xxiv

5.41 (a) final micro view image (b) target micro view image	197
5.42 Average micro image feature error	197
A.1 Plane to plane camera model	229
A.2 Inter-image homography	230
A.3 The interaction of coplanar line and point pairs	232

Chapter 1

Introduction

1.1 Motivation

With the development of modern science and technology, miniaturization has been a trend for recent research and study. More functionalities have been encapsulated into smaller and smaller volume. Micromanipulation has found its potential in many industrial [10] and bioengineering fields [11–13]. The need for hybrid MEMS devices, such as actuators, sensors, optical devices etc is apparent [14–16]. Delicate fabrication and operation techniques are greatly needed. However, due to the scale difference between micro and macro world, the uncertainties in micro environment such as vibration, illumination changes, drifting with unknown motion parameters, whose effects are magnified through the micro imaging system, become the main difficulties for micromanipulation [6, 17–19].

Comparing to conventional objects, micro objects are delicate and easy to be broken, hence safe and reliable sensing system is required to prevent damage and provide high resolution information. In this sense, the non-contact property of visual sensors

offer an attractive option for micromanipulation. The ability to obtain high magnification information of micro object and micro environment makes visual sensors such as optical microscope necessary in most micromanipulation applications [12, 20, 21]. With the rapid growth of computer vision and image processing technologies, some image based methods have gradually been implemented into the domain of micromanipulation [6, 20]. Visual sensing and servoing have begun to play more and more important roles in assisting and enhancing the micromanipulation operations.

Micro images obtained from various visual sensors, such as optical microscope, AFM (atomic force microscope), and SEM (scanning electron microscope) contain high resolution information, but the robustness of these images is limited by small field of view and uncertainties caused by illumination changes, vibration etc. It is therefore difficult to either register the images taken from microscope with the global view or understand the underlying behavior of micro objects through the micro view images. Furthermore, low depth of field restrains the focus to be within very shallow region, thus the features detected from the images captured from a microscope are mostly on the same plane, this prevent the commonly used geometric constraints such as epipolar geometry to be valid. Effective and robust visual enhancement and interpretation method is therefore a significant factor to improve the micromanipulation in perception and assist the operator in proper decision making.

In micromanipulation process, difficulties also lie in the different mechanics of micro world from conventional world as well as the positional accuracy of the mechanisms. In micro world, the adhesive forces dominate instead of the gravitational forces [22, 23], in such a way that the micro parts will often stick to the tools. Dropping them off hence becomes a big problem. To assemble the micro parts, the micromanipulator often has to have a moving mechanism of more than $10^{-2}m$ order

Figure 1.1: (a) the whole wafer (b) the microscopic view of the wafer

and position accuracy of less than $10^{-4}m$ order. For instance, in the Figure 1.1(a), a wafer with diameter $13cm$ is shown, where the red rectangular is the microscopic field of view. The microscopic view of this wafer is shown in Figure 1.1(b), for which the dimension of the field of view is less than $1cm$, and the accuracy requirement of alignment within the image is often in the range of 10^{-4} to $10^{-5}m$. The mechanism of the manipulator in micromanipulation has to produce certain speed to achieve efficiency in long range of motion as well as very slow and precise movement to

achieve accuracy in small range of motion. Thus there is always a trade-off between efficiency and accuracy. Compared to macro domain, in which accuracies in the range of $10^{-3}m$ can be achieved with sensorless manipulators, for micro part whose dimension is within 10^{-3} to $10^{-6}m$, the requirement for accuracy is often micro or submicro (10^{-5} to $10^{-7}m$). This requirement is far beyond the calibration range of conventional open loop precision assembly devices used in industry [24]. Hence close loop sensor based system is necessary in most of the micromanipulation systems [12, 13, 17, 20, 25]. In terms of the above mentioned requirements, the recent developed coarse-fine image based visual servoing framework [10, 26] shows many advantages, such as the ability to travel along large distance as well as make accurate operations, and the ability to compensate for uncertainty not only in the calibration of camera-lens systems, but also in the manipulators and workspace [20].

In the coarse-fine image based visual servoing, there are usually two cameras, one is to provide the global view, the other is mounted on the microscope. The servoing is divided into two phases, coarse phase and fine phase. The process flow is illustrated in Figure 1.2. The coarse phase is based on the images acquired by the global view camera and the fine phase is based on the images acquired by the microscope. Since the high magnification optical system often has a high numerical aperture and thus very small depth of field, the depth of the object with respect to the microscope has significant effect on the image quality. However the depth parameter from the measurement system are not ensured with sufficient confidence, consequently manual adjustment of the lens of the microscope has to be carried out in between of the coarse phase and the fine phase. The need for the depth loss to be estimated and compensated online in coarse phase is present for efficient switch from coarse phase to fine phase, as well as effective information retrieval from unreliable measurements and relevant lessening in time consuming manual adjustments.

Figure 1.2: The process flow of coarse-fine image based visual servoing system for micromanipulation

Furthermore, to achieve precise manipulation on micro objects, the field of view of the microscope has to be kept in the interested area during the whole process. It is thus necessary to compensate the possible object motion during manipulation. For image based micromanipulation, enhancing the visual servoing system to be able to accomplish positioning regardless of target motion is significant for improvement in robustness.

1.2 Significance of This Research

In this research, we proposed novel vision based strategies for micromanipulation system. There are two main functions in the proposed system. One is assisting navigation in the micro environment through enhanced visual appearance. The other is facilitating positioning tasks for micromanipulation through enhanced visual servoing.

When conducting micromanipulation tasks, it is highly demanding and very tedious for the operator to find a certain area over the space that needs operation due to the limited field of view. It is difficult to register the micro field of view with the global view, hence the local information cannot be recognized in a global point of view. To solve this problem, the core is establishing mapping between various view points, that is to model the microscope coordinate system as it changes relative to the object in the scene. However, single micro image presents very small field of view, the correspondences between the micro view image and the global view image is difficult to be found. Hence the map-type reconstruction using geometric primitives recovered from image data becomes very significant. With the out-of-view reconstruction on the microscope views and the motion inferred from the recovered kinematics, reference can be found to relate the current microscope coordinate system with the global view and the relative motion from the current position. This concept is illustrated in Figure 1.3. With the help of this recognition process, the burden of finding interested area under the microscope can be reduced by automated matching and tracking process.

1.2. Significance of This Research

7

Figure 1.3: Concept of navigation with microscope

To improve visual servoing, the proposed strategy is to adapt depth during the coarse phase and compensate for target motion in fine phase. The general coarse-fine micro manipulation system is shown in Figure 1.4. In the coarse phase, visual servoing is conducted based on the image acquired by the macro view camera. However the depth of the features relative to the macro view camera z is not available from the image. This restriction is self-imposed but realistic because even if the depth parameters are given, they will not be known with sufficient confidence to execute the task reliably. That is to say that the depth parameter must be updated in situ. Furthermore, since the micro imaging system has very shallow depth of field, the quality of images obtained from micro view camera is largely depended on the distance d between the micro view camera and the object, if the distance changes, certain lens adjustments have to be carried out. If the depth adaption of course phase servoing is able to drive the stage so that the focus of micro view is achieved after coarse phase servoing, the burden of manual adjustment of lens can be reduced with lens adjustment fixed at a certain focused plane. Moreover, when switched to fine phase servoing, the micro images provide very important information to conduct positioning. However, there are many factors to causes micro object

1.2. Significance of This Research

8

motion in micro environment such as temperature, humidity changes etc. Common techniques for compensating these motion errors include either the use of expensive cooling systems, or waiting hours for the thermal equilibrium of the devices to stabilize [26]. If the micro object motion can be compensated in the fine servoing phase, the efficiency and reliability is expected to be improved with the desired position achieved regardless of the micro object motion.

Figure 1.4: Illustration of coarse-fine micromanipulation

In this thesis, these two very important problems in practical micromanipulation operations, assisting navigation in the micro environment through enhanced visual appearance and facilitating positioning tasks through enhanced visual servoing, will be the main concern. Various visual sensing and servoing techniques are studied, strategies aiming at solving these problems are developed and evaluated.

1.3 Scope and Objectives

The main aim of this research is to design and develop vision based techniques to enhance visual sensing and servoing for micromanipulation. In this study, we have two objectives.

Enhance micro image matching and tracking

To develop and demonstrate a micro image matching and tracking scheme that is not only able to match micro images in spite of repetitive patterns and the effect of image noise but also able to recover the underlying motion parameters, reconstruct the original micro scene to assist human operator.

Although some research has been devoted to model the micro environment and locate the micro object [12, 17, 20, 25, 27], the number of frames involved is relatively small, and it is mainly based on template matching which is not suitable when repetitive patterns are present. On the other hand, the homogeneous transformations that relate consecutive micro images in a sequence encode much motion information of the micro object itself. If the micro images can be matched, the transformation can be estimated, and the underlying motion model can be interpreted properly, applications such as registering each micro image in the perspective view, reconstruct the micro environment and enlarge the limited field of view is attainable. From the enlarged field of view, improved image features can be utilized for coarse phase processing.

Enhance visual servoing for micromanipulation

To enhance the coarse-fine image based visual servoing so that it is able to estimate

and compensate depth online and complete servoing tasks regardless of target motion which is advantageous in solving problems caused by uncertainty of depth in image based approaches and object drifting in micromanipulation.

Image based visual servoing eliminates the need to perform explicit inverse perspective projection mapping, it simplifies the observer dynamics and is much easier to implement. These advantages make image based visual servoing a popular visual feedback approach for micromanipulation [20, 26]. However, since the image based visual servoing uses image features without knowing the actual pose, it suffers from problems caused by uncertainties in depth which constitutes an important part in image Jacobian formulation. The evolution of Jacobian during servoing affected by depth, determines the control law and consequently the posture of the manipulator. When the Jacobian is beyond the range to be in good condition, singularity problems occur. If the depth can be estimated and adapted online to the desired depth which is set to be the depth on focus of the microscope during coarse servoing, then the uncertainty in the lost depth can be compensated while at the same time focus of the microscope can be achieved from the coarse servoing process. In the existing application of image based visual servoing for micromanipulation [6, 20, 28, 29], the depth information during the coarse servoing phase is not considered. For the cases when object tends to drift and vibrate in micromanipulation, there has not been mature technique to compensate target motion in the current image based visual servoing approaches.

In summary, the primary objective of this research is to improve low level image processing and servoing to assist human interventions and enhance robustness of the micromanipulation system.

1.4 Major Contributions of This Thesis

The major contributions of this thesis are summarized as below:

- (a) A novel scheme of micro image matching has been introduced based on the integration of line segment and point features. To compensate the effect of noise on micro image matching, the principles of robust estimation and nonlinear optimization are implemented. In this thesis, the computational cost of this scheme is further reduced 37% by the structural formulation of representative features. Comparing to conventional matching methods which have difficulties in matching images with repetitive patterns or under large displacements, the proposed schemes are able to match multiple frames of micro images with repetitive patterns, scale difference, large displacement and image noise caused by the uncertainties of motion model.
- (b) A homography based motion tracking scheme is proposed to recover motion parameters from the matched micro images. The motion model is deduced from the homography transformation which is estimated during the matching process. To recover the behavior of the moving object regardless of the noise effect, a sequential Monte Carlo simulation is conducted to represent the state distribution, the motion parameters are estimated iteratively based on Bayesian theory. Both translational and rotational motion behavior are taken into consideration. The tracking results are further applied to reconstruct the micro environment and enlarge the field of view of microscope. Compared with previous work which can only match pairs of micro images, the proposed homography based motion tracking scheme is able to match large sequence of micro images as well as recover the tracking map, which can be used to provide much more information of the underlying micro environment for human

intervention and navigation.

- (c) A novel depth adaptive image based visual servoing approach is proposed to estimate depth and compensate uncertainties of the lost depth in image based visual servoing which is able to achieve reduced trajectory and avoid singularity. This framework is further developed to improve the coarse phase of coarse-fine image based visual servoing system for micro manipulation, which is able to adapt depth in the coarse phase. Compared with previous depth adaptive approaches, the proposed method achieves accurate positioning, while singularity and large retreat of motion that is often occurred in conventional methods is avoided.
- (d) A dynamic image based visual servoing framework for compensating unknown translational motion of the object is developed. The motion parameters are designed to be included in the Jacobian relation so that its estimate can be adapted during the servoing. This framework is further developed to improve the fine phase of coarse-fine image based visual servoing system for micro-manipulation, which is able to compensate for micro object drifting. This strategy resolves the problem of planar target motion scenario that has not been formally studied in the previous work.

1.5 Thesis Outline

The remainder of the thesis is organized as follows:

In Chapter 2, the state of the art of micromanipulation systems is reviewed, the literature survey of the relevant research areas in visual tracking and visual servoing is presented.

Chapter 3 presents the scheme of micro image matching based on integration of line segment and point features. The experiment results with MEMS wafer images are presented and compared. Analytical studies are carried out with simulations to investigate the applicable criteria for the proposed approaches.

In Chapter 4, the homography based micro image tracking scheme is introduced. The motion model for both translation and rotation is deduced from the homography parameterizations. The motion estimation of sequence of micro images are derived based on a particle filter paradigm. Examples are presented to demonstrate the ability of the proposed method in reconstruction and field of view enlargement.

In Chapter 5, the strategy of coarse-fine image based visual servoing with both depth adaption and motion compensation (DAMC) for micromanipulation is proposed. The Kalman filter depth adaptive image based visual servoing scheme is developed for the coarse phase servoing. Novel jacobian relation is proposed for motion compensation. The role of projective motion vector in image-motion mapping is studied, the jacobian relation with image feature velocity of the moving target included is constructed. Implementation of this framework into the fine phase of image base visual servoing for micromanipulation is demonstrated. Simulations are presented to validate the proposed approaches.

Chapter 6 concludes the work completed in this thesis and recommends future directions.

Chapter 2

Literature Review

In this chapter, the current status of micromanipulation systems and review of related areas is presented. Manipulation and assembly at the micro scale is a critical issue in diverse industries as the trend for miniaturization continues. The biomedical applications that require precise manipulation of delicate living material is also increasing. However, there are many problems and uncertainties encountered when working at the micro scale. Micromanipulation is also a multi disciplinary area, it relates many other research areas such as robotics, computer vision and physics etc. To cope with the difficulties induced by uncertainties, robust visual tracking is important and inevitable. How to locate the narrow field of view and recover the spacial information from 2D micro image data is largely related to the conventional image tracking and reconstruction techniques such as image matching, transformation estimation etc. There is also an important role for visual servoing in the positioning task. Some of the techniques of visual tracking and visual servoing have already been successfully put into micro domain [6, 20]. In this chapter, state of the art of micromanipulation systems in both biological and mechanical applications are reviewed. Related techniques in visual tracking and visual servoing

are also surveyed, potential relations between these areas and their application in micromanipulation are identified.

2.1 State of The Art

In recent years, micromanipulation has been involved in many bioengineering area and mechanical applications. Several micromanipulation systems have been developed to integrate multiple functionalities and different manufacturing processes. However the difference of scale between micro and macro domains becomes a very important issue. In this section, various micromanipulation systems are reviewed, the necessary requirements for micromanipulation system design are identified.

2.1.1 Functionality of Micromanipulation Systems

With the development of micromanipulation, more and more functionalities have been integrated. Figure 2.1 shows the diverse application areas for micromanipulation systems. In various operations that the micromanipulation systems conduct, precise and reliable alignment and positioning of the micro object is the fundamental functionality for successful manipulations.

In bioengineering area, micromanipulation is needed for transferring material into cells or other biological objects. The objects are kept in liquid and there is a time limit of manipulation for the objects to keep living. Conventionally, the injection is conducted manually. Only with the developments of new manipulating tools, can biomanipulation be automated. The positioning of bio objects is often a difficult

2.1. State of The Art

16

Figure 2.1: Functionality of micromanipulation systems

task, there are non-contact and contact type of handling techniques for bio objects. For non-contact type there are: laser trapping [30], where the laser trap manages to levitate the cell and hold it in position, and electromanipulation [31], by which the electric field produces torque to place the cell. The most commonly used contact type technique is mechanical manipulation according to the classification in Figure 2.1, by which MEMS (microelectromechanical system) based cell holder is applied to aid the manipulation [6], while the injection is conducted with the help of visual feedback [6, 12, 20].

In mechanical applications, micromanipulation is often used for micro assembly. Micro assembly tasks diverged according to different process techniques, but the main applications for micro assembly systems are microfluidic systems and MEMS systems. To assemble microfluidic devices, such as valves, nozzles, pumps, mixers, and flow sensors, a 3-D insertion operation is often needed to connect these devices into system. Precise and reliable alignment has to be made to avoid damage, [32] is a typical example of the interconnection techniques used in microfluidic system assembly. There are two types of MEMS system, optical MEMS system and hybrid MEMS system. Optical MEMS devices including microlenses, micromirrors,

switches, and beam-splitters, need to be aligned in desired positions and connected by optical fibers to provide inlet and outlet with microscale accuracy [33]. The hybrid MEMS system integrates sensors, information processing circuits and actuators [1, 34], besides the physical and mechanical processing difference, the basic requirements for both optical MEMS and hybrid MEMS are precise and reliable alignment operations.

2.1.2 Existing Micromanipulation Systems

In the literature, several micromanipulation systems which are compact in design and ready to integrate more functionalities have come into being in the recent decade.

One class is the single use manufacturing systems which is evolved from IC fabrication industry and concerns the mass production on the same wafer of several scores of microscale MEMS using silicon-based techniques. Products such as pressure micro sensors [35], semiconductors [36–38] are equipped with this class of assembly. However, these assemblies are usually bulky and costly, besides, they can only be excellent for very specific and large volume of products. But for small-scale production, which needs to produce a wide range of expensive and high-quality products, such as biomedical endoscopes and micro optical products, such systems become not well-adapted. The need to assemble three-dimensional hybrid components (silicon, glass, metal) using meso and micro components with more reconfigurability and flexibility drives new classes of micro assembly stations to come out.

Micro robots based micro assembly stations (MMS) become popular due to the

2.1. State of The Art

18

reason that micro robot is flexible, multi functional, and it's capable of moving over long distances and manipulating in the range of a few nanometers [5, 39–41]. The associated problems of MMS regarding to the assembly planning [42, 43], control issues [44], design and typical operating problems [45] are also gradually studied.

Figure 2.2 shows the 'Nanorobotics' developed in the Swiss Federal Institute of Technology in Zurich (ETHZ) [5]. It is a 3 dof micromanipulation workcell. The right arm is the object carrier and has 4 dof. It is composed of Abalone [46], a 3 dof planar mechanism and a z micrometric stage. The left arm, or the tool carrier also has 4 dof. An xyz-stage driven by 3 DC-motors provides the 3 translations of the tool which can be positioned at the center of the field of view of the light microscope. The repeatability of the device is $1\text{ }\mu\text{m}$. A piezoelectric rotating actuator [46], allows the rotation of the gripper about the y axis of the stage with a resolution of $0.1\mu\text{rad}$.

Figure 2.2: Set-up of the nanorobot system [5]

The system designed in [6] used for autonomous injection is a hybrid micromanipulation unit (see Figure 2.3). This unit includes a holding pipette, an injection pipette, two standard pipette holders, a high precision 3 DOF micro robot, and a

coarse manipulator. The injection and holding pipettes are both processed using a micropipette puller. The dimensions of the pipette tips are 1mm in inner diameter for the injection pipettes and 20mm in outside diameter for the holding pipettes. Both the holding pipettes and injection pipettes are held by pipette holders. A 3 DOF microrobot is used in XYZ axes each of which has a travel of 2.54 cm with a step resolution of 40nm.

Figure 2.3: Autonomous embryo injection system [6]

Another micromanipulation system is proposed in [7]. The microscope in this system is a handstand type with an oil immersion lens driven by a piezoelectric actuator. The manipulator side consists of a coarse moving part (driving with stepping motors, with range $25\text{mm} \times 25\text{mm} \times 25\text{mm}$) and a fine moving stage (driving with piezoelectric actuators, with range: $100\mu\text{m} \times 100\mu\text{m} \times 100\mu\text{m}$).

A 6 dof micro assembly workcell is presented in [8] (see Figure 2.5). The configura-

Figure 2.4: Microscope and micromanipulator [7]

tion of the micro handling system consists of two concentrated micro motion manipulators (main and sub) equipped with a micro tool holder having 6 degrees-of-freedom (DOF) of mobility. It is composed of 2 DOF coarse motion worktable (driving with piezoelectric actuators) with a motion range of $30mm \times 30mm \times 30mm$ and an accuracy of $50\mu m$ along x - and y - axes. It supports a 2-DOF fine positioning system (driving with piezoelectric actuators, the movable ranges: $100\mu m \times 100\mu m \times 100\mu m$) with an absolute position accuracy of $40nm$ controlled via closed loop position feedback controllers. Finally, the orientation of the end-effectors can also be changed by two accurate ultrasonic micro-motors (accuracy: 0.5°) along θ_x - and θ_z - axis.

In [9], the utilization of the rigid platform as a secondary micro metric coordinate table for precision machining is discussed (see Figure 2.6). In this system, welded joints connecting the changeable links with the platform and base provide the high stiffness. Hydraulically controlled micro actuators use longitudinal elastic deformations for elongation of the legs. As shown in Figure 2.6, the XYZ moving units of the machine tool represent a coarse feed subsystem, while the platform fastened on the machine table represents a fine subsystem.

Figure 2.5: Sequencing of the different operating microassembly tasks [8]

Figure 2.6: Laboratory arrangement of a fine-course feed system on the milling center: 1, CNC machine tool; 2, 6-DOF platform; 3, hydraulic control arrangement; 4, computer for the platform control; 5, display of the measuring device. [9]

In [1], a hybrid supervisory wafer-level 3D microassembly system is developed (see Figure 2.7). In this system, large range coarse motions are provided by the 4 DOF coarse positioning unit. Small range fine motions are provided by the 4 DOF micromanipulator. Planar motion in the horizontal direction is provided by an open frame high precision XY table with a travel of 32 cm (12 inch) and a repeatability of $1\mu m$ in both directions. Position feedback with a resolution of $0.1\mu m$ is provided

Figure 2.7: Positioning system configuration

by two linear encoders on the table. The three Cartesian DOFs are provided by the micromanipulator, which also provides yaw motions with its rotational DOF.

From the design of these systems, the following observations can be made. Firstly, optical systems are inevitable in MMS systems. [5] uses light microscope to observe the stage. The inverted microscope is used to obtain visual information in [6]. Optical microscope is implemented and a metal halide lamp is used over the sample holder to increase visual luminance of the sample in [7]. A reflecting light-type optical microscope (OM) is used as a top-view vision sensor for part pickup and a lateral microscopic view is used for fine adjustment of the position and orientation during micro assembly in [8]. These visual sensors not only play important role for operator to access micro world but are also used for visual feedback control in many operations.

Secondly, most of the MMS systems have multiple degrees of freedom (DOF) and coarse-fine configuration, such as in [5, 7, 8]. In the MMS system proposed in [7], a coarse moving part and a fine moving stage are integrated to provide 3 degrees

of freedom. The 'Nanorobotics' proposed in [5] have an xyz-stage, z micrometric stage and the 3-dof micro robot "Abalone", which provides the fine mechanism. The micro assembly workcell presented in [8] has 6 dof composed of a coarse motion worktable and a fine positioning system. In these multiple DOFs systems, DOFs are often separated into different devices so that the structural complexity of the manipulator can be reduced.

Another observation is that to meet the requirements of accomplishing multiple micromanipulation tasks in one unit, hybrid micro manipulation system evolves with more functionalities [1,6]. In [6], the micromanipulation unit consists of several different types of components, each has its own functionality, such as holding pipette, injection pipette, pipette holders, micro robot, and coarse manipulator. The whole system can be used not only for holding the embryos but also depositing DNA. The system developed in [1] is also a hybrid system, where 4 DOF positioning unit and 4 DOF micromanipulator are integrated to handle complex micromanipulation operations in multiple degrees of freedom.

In summary, visual sensing data is very important information to ensure good performance for the MMS systems. Coarse-fine positioning system is necessary in most of the systems, which helps to realize fast and accurate localization of the micro objects. Multi degrees of freedom are often separated into different devices such as object carrier, metric stage, gripper etc, in such way that the structural complexity of the manipulator can be reduced.

2.1.3 Summery

To summarize the state of art of micromanipulation, vision is one of the most important sensing information for micromanipulation. To enhance and make good use of visual sensing is essential for reducing burden and promoting automation in micromanipulation tasks. In micromanipulation systems, different applications often require different system configuration and customization. There is no optimal configuration to support complex micromanipulation operations, although coarse-fine mechanism is prevalent in many systems. There is also a lack of general-purpose automated micromanipulation machine. Micromanipulation is far from standard due to the highly customization needed for different tasks. Many supported techniques, such as 3D visual servo, computer vision and micro force control still remain to be developed to handle the problems encountered in terms of robustness and efficiency for micromanipulation.

2.2 A Review of Visual Tracking

In computer vision, visual tracking is the act of consistently locating a desired feature in each image of an input sequence. It can be used to estimate the motion, recover the structure of the scene etc. In visual tracking tasks, the matching problem has to be solved between every two images of the sequence.

2.2.1 Image Matching

Image matching is one of the most difficult basic problems in computer vision. In the general image matching tasks, given the image point m_1 in one image, the image

point m_2 in the other image, the process of finding corresponding points m_1 and m_2 is referred as image matching.

Although there are different types of image features, such as colors, textures, and shapes, the grey level primitives, corners [47,48], edges [49–51], and contours [52–54] are most commonly used since they are easy to be extracted and the geometric constraints are available to be imposed to improve the matching performance. The methods for solving the image matching problem can be categorized into 3 classes:

Correlation based In this case, the similarity is measured by correlation computation around certain image features [55–59]. Thus given a feature in the first image, the most similar region is searched in the second image. To reduce the computational cost, geometric constraints, such as epipolar geometry can be utilized to limit the size of search region. The limitation in this method is the computational cost is relatively high.

Optical flow Optical flow method usually takes advantage of the intensity difference over time, by comparing the difference of intensity of an area, the velocity field which warps one image into another is computed. The optical flow can give important information about the spatial arrangement of the objects viewed and the rate of change of this arrangement [60]. It has been proved that the motion parameters and the shape of the target can be estimated from two or more successive optical flow images [61]. A lot of work has been done to recover shape and object motion information from optical flow estimation [62–64]. The limitation of using optical flow estimation is that the computed optical flow could be inaccurate due to the image noise and quantization error. Thus obtaining the information of shapes or motion parameters from optical flow may turn out to be unreliable. Hence, optical

2.2. A Review of Visual Tracking

26

flow is more suitable to be used when the images are similar to each other or the movement is slight.

Direct method Direct method, such as dynamic programming [65–67], genetic algorithm [68] and neural networks [69], avoid both correlation and full image optical flow computation, the matching problem is often formulated as the minimization of a function of many discrete variables. However, the ambiguity is difficult to be determined by direct method.

Besides the above methods, when higher level features are available, the geometric characteristics associated with them can be used to aid matching. It could be more reliable than using grey level values [51]. For instance, lines as higher level features are determined by a large number of pixels, the redundancy makes it possible to locate those features more accurately in the image plane [70].

2.2.2 Image Tracking

In image tracking, the image transformation parameters recovered from image frames are used to track the pose of an object. There are two types of tracking method, feature based [71, 72], in which the relation between features in different images is estimated, and model based [73, 74], where the relation of features in different image frames with the same model is estimated. Image tracking is important for many applications such as image registration, recovery of the structure from motion etc.

A lot of work has been done in this area. By representing the displacement of image feature as a 2-D random Gaussian variable, Kalman filter can be applied. In [75], the Kalman filter is used to predict the motion. In [76], a Kalman filter based active

contour model is proposed to track nonrigid objects. In [77], Kalman filter is used to predict target pixel. However, Kalman filter based approaches have limitation such as it can only handle linear model with noise of Gaussian distribution. To improve the tracking to be able to deal with uncertainty in a more coherent manner and cover all the process of the tracking, Bayesian based approaches have been gradually developed for image tracking. Probabilistic tracking using Monte Carlo simulations to estimate the posterior density via a Bayesian frame work is presented in [78]. A particle filter frame work integrating a special indirect measurement model is proposed for visual tracking in [79]. A fully probabilistic Relevance Vector Machine (RVM) to generate observations with Gaussian distributions is used for long-term region tracking in [80]. Different Bayesian algorithms for direct target tracking in image sequences in a situation of random target aspect and unknown clutter parameters are discussed in [81], a generic simultaneous tracking and verification algorithm based on sequential Monte Carlo sampling methods is presented in [82]. For tracking nonlinear motion model with non-Gaussian noise, particle filter appears advantageous [83]. Furthermore, particle filter is able to represent the state distribution completely by using sequential importance sampling and resampling and has many features such as flexibility and ease of implement, which make it very popular recently in solving tracking problems [84–87]. However it is very challenging to design the model and apply particle filtering approaches in different tracking problems, especially micromanipulation.

2.2.3 Summery

In this section, we reviewed the existing image matching and tracking techniques. Although they are well developed areas in the literature, to identify suitable methods and adapt these methods to solve problems in micromanipulation would be very challenging, especially when uncertainties are prominent in the micro images and nonlinearities are sometimes present in the motion dynamics. For image matching tasks, if features with redundancy can be extracted and matched correctly, the noise on single point can be compensated by many more points that represented by the same redundant feature. So looking into those redundant features would be very interesting in micro image matching. For the tracking tasks, since the dynamics could be nonlinear and the image noise, which is not necessarily linear, could be induced from various sources, the particle filter appears attractive in the ability to represent the complete distribution through Monte Carlo method.

2.3 A Review of Visual Servoing

The term "visual servoing" which means the closed-loop position control for a robot end-effector by visual feedback, is first introduced by Hill and Park [88]. Visual servoing system is often classified in two groups [89]: position-based (see Figure2.8) and image-based systems (see Figure2.9). In position based system, the input is computed in 3D Cartesian space, while in image based system, the input is computed in 2D image space. In general, image-based visual servoing is known to be more robust not only with respect to camera but also to robot calibration error [90].

Comparing to position based control, in image based visual servoing, the step of conjunction with a geometric model to determine the pose of the target is omitted, and servoing is done on the basis of image features directly. Thus the difficulties

Figure 2.8: Position based system

Figure 2.9: Image based system

in calibration and geometric model identification are circumvented. The problem is also simplified by eliminating the necessity to compute 3D parameters. For micro-manipulation tasks, the above mentioned features are extremely useful since instead of 3D information, 2D image is usually all that can be obtained from the microscope.

2.3.1 Image Based Visual Servoing (IBVS)

In image based visual servoing system, the feature vector $f(t)$ is extracted from the image and compared with the reference feature $f_{ref}(t)$ in the feature space. The feature $f(t)$ would be some function of the relative pose of the camera $X_c(t)$ ($(x_1(t), \dots, x_m(t))$). In general this function is non-linear and coupled, such that motion of the camera along one degree of freedom results in a complicated motion of all features. This relationship can be linearized about the operating point as:

$$\delta f(t) = J_v(x) \delta x \quad (2.3.1)$$

where J_v is defined as image Jacobian, which is first introduced by Weiss et al. [89], named as the feature sensitivity matrix:

$$J_v(x) = \left[\frac{\partial f}{\partial x} \right] = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k(x)}{\partial x_1} & \cdots & \frac{\partial f_k(x)}{\partial x_m} \end{pmatrix} \quad (2.3.2)$$

where m is the dimension of the task space, k is the number of image features.

Since the image jacobian plays a crucial part in IBVS, different image Jacobian estimation methods affect the performance of IBVS systems greatly [91]. Furthermore, the selection of image features used in IBVS is very important to ensure the completion of specified tasks [91, 92]. And to implement IBVS successfully, several problems that are often encountered in IBVS such as singularity and ambiguity etc. have to be considered carefully. (Singularity is the situation when the image Jacobian goes bad-conditioned and ambiguity is the situation when multiple solutions are available for the same image). Besides, although depth does not have much effect on the stability of IBVS control scheme, depth parameters are prerequisite information and affects the convergence speed for IBVS [93]. Hence to achieve good performance of IBVS, depth should be estimated and updated online during the servoing process.

2.3.2 Depth Adaptive Image Based Visual Servoing

The early work for IBVS are depended on some known 3D model of the scene. For example, in IBVS systems [94, 95] and [96], the depth information is supposed to be known. In [97], Grosso et al. presented PBVS and IBVS approaches, in both the actual position of the target in the Cartesian space is assumed. Not until recently,

has depth estimation been investigated for IBVS. The published work along this line can be divided up into three categories: (1) Depth estimation with structure from motion. (2) Depth estimation with least square approaches. (3) indirect estimation by adapting full coupled Jacobian.

The most commonly used depth estimation techniques are derived from computer vision theories, such as structure from motion [98–100]. In this sense, the features of an object from a desired image is compared to features of the object in the current image, geometric relations are exploited to reconstruct the scene model. The information obtained from this reconstruction, especially depth, is used to develop controllers. In [101], Martinet et al. use stereo relations to recover depth, a two phase estimation technique is developed in [102, 103] to estimate camera displacement, hence the depth information to construct image Jacobian. However in these two cases, the camera has to be calibrated. They are model dependent, so that they are not suitable for unknown environment. Another method for partially reconstructing the 3D structure is by using homography decomposition [104, 105]. By decomposing the inter image homography into a rotation matrix and a rank 1 matrix, the depth from the camera to the object plane can be estimated. This approach is utilized in many applications. For instance, in [106], the rotation matrix obtained from homography decomposition is used to form a Lyapunov-based adaptive control strategy to compensate for the lack of unknown depth measurements. Similar decomposition is used in [107] for wheeled mobile robots. Several hybrid control strategies are also designed based on this decomposition, such as [90, 108, 109]. Although this method provides solutions to partially recover the 3D structure [110], the decomposition is rather complicated and disambiguation is difficult from the multiple solutions.

2.3. A Review of Visual Servoing

32

Some least square approaches have also been implemented into depth estimation, recursive-least-square algorithm is used in [111] for depth identification, a 3D depth estimation procedure is introduced in [2], based on prediction errors. However, the convergence rate cannot be ensured.

An alternative for estimating depth directly is to estimate and adapt the full Jacobian. The early development of Jacobian estimation is similar to an exponentially weighted recursive least square (RLS) update equation for vectors [112], where the vector format is applied to the Jacobian matrix estimation. Broyden's updating method is demonstrated by Jagersand in [113]. Based on the validity of Broyden's method, a nonlinear least square optimization approach (Newton's method) is proposed by J.A. Piepmeyer [114], the moving target scenario is confronted in [115,116] by taking quasi-Newton steps. In [117], Zhenyuan Deng shows that adaptive control is more advantageous for complex visual servo tasks. Similar to depth estimation error, convergence in visual servoing is relatively insensitive to error in the Jacobian estimation [91], but the estimation error affects the trajectory and time taken towards the goal. The benefit of this Jacobian estimation technique is the easiness to be incorporated with robust and adaptive control methods. However, for fine manipulation tasks in multiple degrees of freedom, certain model information has to be provided to give a good initial estimate of Jacobian to avoid erroneous movements and reduce processing time. Furthermore, possible singularity is difficult to be avoided during adaptation of the full Jacobian matrix.

2.3.3 Dynamic Image Based Visual Servoing

In the literature, there have not been much work done regarding to tracking moving target. One of the important research regarding to this issue is reported in [116,118], which is based on [113,119], where the complete Jacobian is iteratively updated by Dynamic Quasi-Newton method for model free IBVS systems [114,115,117]. However, it's difficult to get good initial estimate and to ensure the convergence rate for full Jacobian adaption. Another dynamic image based visual servoing task is considered in [120], which is a classical task of mobile target tracking using a pan-and-tilt camera. The image servoing technique regards to retrieving the target position in the image from its estimate motion. However, since only two rotational degrees of freedom are considered, it's not suitable for manipulation applications where many translational motions are involved.

2.3.4 Summary

In this section, we reviewed the visual servoing technique, and especially the existing image based visual servoing methods for depth adaption and moving target. In the two trends of image based visual servoing, namely solving the servoing problem by 3D reconstruction from images and solving the servoing problem by optimization based method with pure image data, the latter appears to be more suitable for micromanipulation cases, since the 3D structure reconstruction is extraordinary difficult when the projective properties are very weak due to low depth of field of microscopic imaging. It also shows that the existing techniques are not adequate to meet the requirements for micromanipulation. The depth recovering and motion compensation have to be dealt with altered models. The singularity and 3D ambiguity need to be avoided, at least locally. Since in micromanipulation, the travelling

2.3. A Review of Visual Servoing

34

range is limited in the order of cm even for coarse phase movement, the out of view concern is not a critical issue, since it's relatively easy to keep all the features in the field of view for both coarse phase and fine phase servoing as long as the switch from coarse phase to fine phase has been successfully accomplished.

Chapter 3

Micro Image Matching

3.1 Introduction

In micromanipulation, the image has to be recognized so that appropriate decisions can be made for either localization or navigation. However, the features in micro images are often sparse, repetitive patterns occur regularly, and the image is easily affected by changing of environment, such as lighting, temperature, vibration etc. Although many standard image matching approaches are available, they are not suitable for the cases of micro images. It is difficult to recognize microscopic images without certain prior information.

In this chapter, we developed novel methods for image matching across micro view images. Two methods integrating both line segment and point features for matching micro images are proposed to obtain reliable matching results. One considers line segment and point as separate features, the other uses specific structure that consists of both line segment and point features. Both methods are tested and compared, the performance is evaluated through series of simulations.

3.2 Traditional Approaches

Traditional image matching approaches are mainly based on optical flow [62–64] or correlation of single points [58, 59, 121, 122], where geometric constraints such as epipolar geometry can be applied to constrain the solutions. However, they are not suitable for solving matching problems of micro images.

In the following example, we illustrated the difficulties of applying conventional matching techniques into micro images. Optical flow is the apparent motion of brightness patterns in the image. Generally, optical flow corresponds to the motion field. Figure 3.2 shows the optical flow computation [123] of a pair of micro images taken from an image sequence of a MEMs wafer located on a manipulation stage shown in Figure 3.1. The stage is undergoing an unknown motion, seen from the microscope the motion field should be consistent. However, the flow field computed in Figure 3.2(d) shows that the optical flow does not consistent with each other. Figure 3.2(a), Figure 3.2(b) show the partial derivatives of the image (I) with respect to x and y direction in the image respectively. Figure 3.2(c) shows the partial derivative of image (I) with respect to time (t). In this case, the apparent motion of brightness patterns in the image computed by optical flow cannot present the motion field. The dominance of repetitive patterns makes the direction of the intensity gradient differ from the direction of motion.

3.2. Traditional Approaches

37

Figure 3.1: (a)(b) image 1 and image 2 from a sequence of the MEMS wafer. The images are 484×648 pixels.

Figure 3.2: (a) gradient in x ($\frac{\partial I}{\partial x}$), (b) gradient in y ($\frac{\partial I}{\partial y}$), (c) gradient in t ($\frac{\partial I}{\partial t}$), (d) flow field

3.2. Traditional Approaches

38

This set of images are also tested by single point correlation based method. The corners are first detected shown in Figure 3.3 based on [124]. The matches found by point cross correlation [122] is shown in Figure 3.4.

Figure 3.3: Corners detected. There are 300 corners in each image

Figure 3.4: matches found by point cross correlation

3.2. Traditional Approaches

39

From Figure 3.4, it can be seen that the matching result is far from acceptable. Correct matches are difficult to be found. The repetitive patterns lead to wrong matches very often.

Another method that has very important role in matching image patterns is template matching [125–127]. Among extensively studied template matching applications, template matching by correlation is the most commonly used [128]. In the next example, we applied a template matching method to images shown in Figure 3.1, which used matched spatial filter based on correlation [128]. Figure 3.5 shows the template used for matching. Figure 3.6 shows the binarized images of Figure 3.1. The correlation in frequency domain between each of the images and the template is shown in Figure 3.7. The peak of correlation is found in Figure 3.8. And the template is finally matched into the original images in Figure 3.9.

Figure 3.5: (a)The grayscale image of the template (b) The binary image of the template

3.2. Traditional Approaches

40

Figure 3.6: Binary images of the original grayscale images shown in Figure 3.1

Figure 3.7: Correlation shown in frequency domain

The results show that template matching provides acceptable results in matching the template image into the original images. However, for long sequence of images, such as hundreds of micro images, the template has to be updated very often, which requires additional human interaction. And to optimize the matching results based on the features in spacial domain is difficult to achieve, when the template is computed in frequency domain.

3.2. Traditional Approaches

41

Figure 3.8: The correlation peaks found in the images. The color channels are allocated to show the matching results. Red Channel: the original image; Green, Blue Channels: correlation

Figure 3.9: Matched template shown in each of the images

In summary, traditional matching techniques are not sufficient to solve the problems in matching micro images. The difficulty of finding good matches caused by repetitive patterns, uncertainties need to be considered and well compensated during matching process. Local methods such as point cross correlation is not suitable in micro images, to robustly match sequence of micro images correctly, more global information such as the underlying structure of the features has to be included. Fur-

thermore, since the geometric constraints prevalent in perspective view images such as epipolar geometry are not valid for micro images any more due to the reason that the features are mostly on the same plane, other constraints have to be established to achieve satisfactory matching results with micro images.

3.3 Line-Point Integrated Micro Image Matching

In this section, two integrated image matching methods with both line segments and point features across multiple view micro images are presented, specifically Line Segment Based Line-Point Matching(LSG) and L-shape Based Line-Point Matching(LSH). LSG uses line segments and points as separate features, while LSH uses L-shape structures which are formed by line segments and point.

The difference of pixel coordinates of the same feature between two images is called apparent motion. Generally, the apparent motion is not a 2D geometrical transformation. However, since the features on the wafer can be viewed as coplanar ensembles, the apparent motion can be approximated by a 2D transformation. As similarity or affine transformation may not approximate the apparent motion well over large image regions [51], to improve matching, the homography, which represents the general planner mappings and forms the projective general linear group $PGL_3(h)$, is used to approximate the apparent motion between two images in the proposed approach. Robust estimation, optimization and guided match are also implemented in the proposed approach to reduce the effect of outliers, optimize the estimated homography, and increase the number of matches.

3.3.1 Line Segment Based Line-Point Matching (LSG)

Introduction

In this section, we introduce the Line Segment based line-point matching method: Line Segment Grouping (LSG) method.

Line segments are high level image features, which have several advantages in solving image matching problems. They are very common in man-made environment such as Micro-Electro-Mechanical structures. Line segment captures more information contained in the images than that captured by point features [129], furthermore, they incorporate the continuity constraints [130], which requires the gradient of disparity to be continuous. However, line segment matching has been addressed to be a difficult problem due to several reasons. First, there is deficiency caused by line extracting algorithms. Second, the end points of a line segments are not well defined, geometric constraints of end point is not available [50]. In the literature, various approaches have been adopted for line matching problem. In [131], matching of individual line segments is dealt based on geometric attributes. Groups of line segments are matched by graph-matching [132]. In [133], line matching was proposed by dynamic programming in Hough space. Spacial domain dynamic programming is used in [134]. Trifocal geometry based approach and homography based approach are introduced for both short and long range of motions in [50]. Multi-stage primitive election procedure is proposed in [135]. However, the problem caused by microscopic imaging, such as repetitive patterns, sparseness of features etc. have not been considered in line segment matching, neither do these approaches show ability in handling micro images.

Neighborhood of features can make great effort in finding correct matches based

3.3. Line-Point Integrated Micro Image Matching

44

on the structure near the interested features, which has been demonstrated in [136] for point features. However, since points are not sufficient to encode the underlying structure, they perform badly in micro images as demonstrated in Figure 3.4. In this part, we extend this neighborhood voting approach into matching line segments in micro view images where the matching score is computed based on both the correlation computation of the line segments and the neighborhood line segments vote, then the estimated homography transformation is used as the geometric constrain in matching micro images. Particularly, the primitive matches are established by neighborhood voting of grey level correlation. Then the relaxation process is adopted to eliminate ambiguity. After that, robust estimation is employed to estimate the homography from the putatively matched line segments features. Guided by this homography, point matches and more line segment matches are found. Finally, non-linear optimization is applied iteratively to refine the homography estimation and search for more matches of lines and points that are consistent with the estimated homography, until the number of matches converges. The processing steps for Line Segment Based Line-Point Matching (LSG) is given by Table 3.1:

Table 3.1: Processing Steps For LSG

1. Compute grey level correlation
- establish goodness score for candidate pairs
2. Compute the measure of support for a match candidate
- find putative matches
3. Relaxation process
- disambiguate the putative matches
4. Robust estimation
- estimate the transformation between two images
5. Guided match
- find more matches and refine the transformation

Figure 3.10: Illustration of correlation for l and l'_i

Compute grey level correlation

Since the constraint between images is not yet known, the search should be performed in the whole image. Given center point c_1 of a line segment l in one image, a rectangular search box around the same coordinate point as c_1 in the other image can be selected. The computation of correlation is performed on the correlation window between line segment l in the first image and each of the line segments lying within the search window in the second image l'_i (see Figure 3.10). The search window reflects some a priori knowledge about the disparities between the matched line segments, and also restrain the search region to reduce computational cost. Based on the general point correlation computation [137], the correlation score for line segments l and each l'_i is defined as:

$$Score(l, l'_i) = \frac{\sum \sum_{\Omega} [I_1(x_1 + i, y_1 + j) - \bar{I}_1][I_2(x_2 + i, y_2 + j) - \bar{I}_2]}{mn\sqrt{\sigma_1^2\sigma_2^2}} \quad (3.3.1)$$

where Ω is the correlation window which is a rectangular box around l and l'_i rotating

3.3. Line-Point Integrated Micro Image Matching

46

θ_1, θ_2 respectively, and has the size $m \times n$. θ_1, θ_2 are the acute angle between the line segment and the horizontal direction of the images. The length of the correlation box m is defined as the average length of the rectangular boxes around $\mathbf{l}$ and $\mathbf{l}'_i$ respectively. n is the width of the rectangular box. $\bar{I}_1, \bar{I}_2$ are the average intensity of the correlation box in image 1 and image 2 respectively. σ_1 and σ_2 are the standard deviation of intensity of the two correlation boxes, defined as:

$$\sigma_k = \sqrt{\frac{\sum \sum_{\Omega} I_k^2(x_k + i, y_k + j)}{mn} - \bar{I}_k^2} \quad (3.3.2)$$

where $k = \{1, 2\}$, refers to image 1 and image 2 respectively.

This correlation score indexes the similarity of the neighborhood region of each line segment in image 1 and it's matching candidate in image 2. It ranges from -1 to 1. The closer it is to 1, the more similar the pair of line segments is, while the closer it is to -1, the more different the pair of line segments is. Since we define the correlation window to be located at the center of the line segments detected and taking the average length of the candidate pair as the length of the correlation window, the ambiguity caused by different length of line segments is eliminated. This correlation score gives us a rough estimate of similarities, and can be presented as a measure of the matching candidate's goodness .

Compute measure of support for a match candidate

To cope with the problem caused by repetitive patterns in microscopic images, the neighborhood information of the detected features are taken into account by adopting the concept of matching strength [122]. In this thesis, the original match-

3.3. Line-Point Integrated Micro Image Matching

47

ing strength of point is extended into line segment cases by defining the matching strength based on the properties of line segments where the support from neighboring line segments are accumulated to vote for the candidate line segment.

Consider a pair of match candidate $\mathbf{l}_{1i}, \mathbf{l}_{2j}$ (see Figure 3.11). Let $N(\mathbf{l}_{1i}) = \{N_1(\mathbf{l}_{1i}), N_2(\mathbf{l}_{1i})\}$ be the neighboring area for $\mathbf{l}_{1i}$, and $N(\mathbf{l}_{2j}) = \{N_1(\mathbf{l}_{2j}), N_2(\mathbf{l}_{2j})\}$ be the neighboring area for $\mathbf{l}_{2j}$ centered at the ends of the average segments $\bar{\mathbf{l}}_{1i}$ and $\bar{\mathbf{l}}_{2j}$ respectively, within a disc of radius R . The average segments are obtained by taking the direction of the original line segments in each image and taking the average length of the candidate line segment pair. When the center (c_{1k}) of line ($\mathbf{l}_{1k}$) enters the neighboring area ($N_1(\mathbf{l}_{1i})$), $\mathbf{l}_{1k}$ is considered as the neighboring line of $\mathbf{l}_{1i}$. If $(\mathbf{l}_{1i}, \mathbf{l}_{2j})$ is a good match, many matches $(\mathbf{l}_{1k}, \mathbf{l}_{2l})$ will be expected, where $\mathbf{l}_{1k} \in N(\mathbf{l}_{1i})$ and $\mathbf{l}_{2l} \in N(\mathbf{l}_{2j})$, such that the geometric position of $\mathbf{l}_{1k}$ relative to $\mathbf{l}_{1i}$ is similar to that of $\mathbf{l}_{2l}$ relative to $\mathbf{l}_{2j}$.

Figure 3.11: Illustration of neighboring support for a candidate match

More formally, the new formulation of matching strength for two matching candidate line segments is given by:

3.3. Line-Point Integrated Micro Image Matching

48

$$\begin{aligned}
 M_s(l_{1i}, l_{2j}) & \\
 = c_{ij} [1 + \eta (\sum_{l_{1k} \in N_1(l_{1i})} c_{kl} \max_{l_{2l} \in N_1(l_{2j})} \delta_1 + \sum_{l_{1k} \in N_2(l_{1i})} c_{kl} \max_{l_{2l} \in N_2(l_{2j})} \delta_2)]
 \end{aligned} \tag{3.3.3}$$

where c_{ij} and c_{kl} represent the goodness of match candidate (l_{1i}, l_{2j}) and (l_{1k}, l_{2l}) respectively, they can take the correlation score defined in (3.3.1). η is a weighting parameter, by changing which, the weight put on the correlation score of candidate line segment pair itself or the neighboring voting to decide a match can be tuned, thus it can be chosen to trust the similarity of the candidate line segment pair more or trust their neighboring area more. When $\eta \leq 0.5$, the correlation score between candidate pair itself is put on more weight. This is suitable for the cases when the candidate line segment pair can be extracted with better quality or the features in the image are sparse. When $\eta > 0.5$, more weight is put on the neighboring voting, this is applicable to the situations, when the candidate segment pair itself is likely to be deteriorated by truncation errors or line extraction error, i.e. when the line segments are very short or the edges have been broken into several segments. In this case, the similarity of the neighboring region can help to compare the similarity of the candidate segment pair despite of the effect of these noises.

In the above equation:

$$\delta_i = \begin{cases} \exp^{-(\frac{\Delta d}{d} + \frac{\Delta \theta}{\theta})}, & \Delta d < \epsilon_d; \Delta \theta < \epsilon_\theta; \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.4}$$

$i = 1, 2$, $\epsilon_d, \epsilon_\theta$ are thresholds on the relative difference. The relative difference of distance $\frac{\Delta d}{d}$ and angle $\frac{\Delta \theta}{\theta}$ are used to score the goodness of neighboring similarity

3.3. Line-Point Integrated Micro Image Matching

49

(see Figure 3.11), where:

$$\Delta d = |d_{ik} - d_{jl}| \quad (3.3.5)$$

$$\bar{d} = (d_{ik} + d_{jl})/2 \quad (3.3.6)$$

$$\Delta \theta = |\theta_{ik} - \theta_{jl}| \quad (3.3.7)$$

$$\bar{\theta} = (\theta_{ik} + \theta_{jl})/2 \quad (3.3.8)$$

When the relative position between $\mathbf{l}_{1i}$ and $\mathbf{l}_{1k}$ is similar to the relative position between $\mathbf{l}_{2j}$ and $\mathbf{l}_{2l}$, $\frac{\Delta d}{d} \rightarrow 0$, $\frac{\Delta \theta}{\bar{\theta}} \rightarrow 0$ so that $\delta_i \rightarrow 1$. The correlation score for a neighboring pair of line segments will be accumulated to the matching strength for the candidate pair $(\mathbf{l}_{1i}, \mathbf{l}_{2j})$. In this way the neighboring line segments contributes to the matching strength of the candidate pair. The more neighboring matching pairs with similar relative position exists, the larger the matching strength.

To find the matched candidate, the ambiguity caused by repetitive patterns needs to be eliminated. Although the line segment itself may be similar to other candidates in the other image, the accumulation of their neighboring areas can be differed considerably as the accumulation area grows. This matching strength which encapsules both the grey level similarity of the candidate matching pair and the similarity of neighboring structure becomes a better measure to encode similarity for line segments. However, the matches are not yet decided at this stage, in the following, the relaxation process is applied to find the matches and disambiguate them to be one to one matching.

Relaxation Process

To disambiguate the candidate pairs, relaxation process for point [138] is extended to line segment to iteratively minimize the total energy for line segment matches.

The following energy function for line segment matches is used:

$$J = \sum_{(l_{1i}, l_{2j})} M_s(l_{1i}, l_{2j}) \quad (3.3.9)$$

where $M_s(l_{1i}, l_{2j})$ is the matching strength of the candidate line segment pair computed by (3.3.3). The matches are then updated at each iteration until the energy converges. The proposed disambiguation strategy for line segment takes account of two parts of the neighborhood area for each line segments and eliminate the bad matches. All the matching candidate above a threshold are selected to be potential matches. Then the matching strength table $T_{ms} = \{T_{ms1}, T_{ms2}\}$ for line segment is formed, which consists of all the matching strength computed from image 1 to image 2 (T_{ms1}) and matching strength computed from image 2 to image 1 (T_{ms2}). The ambiguity score for each line segment from image 1 to image 2 (U_a) and from image 2 to image 1 (U_b) is developed from the point ambiguity score [122] by replacing the matching strength of point with the matching strength of line segment:

$$U_i = 1 - M_s^{(2)}/M_s^{(1)} \quad (3.3.10)$$

where $i = \{a, b\}$, $M_s^{(1)}$ and $M_s^{(2)}$ are the best and second best matching strength for the line segment computed from (3.3.3). U_i is ranging from 1 to 0. The larger U_i is, the less ambiguous the matching candidate is.

3.3. Line-Point Integrated Micro Image Matching

51

With the two tables, one consists of the elements from T_{ms} with all the unique candidate pairs, the other consists of $T_{ua} = \{\frac{U_{ak}+U_{bl}}{2}\}$, where U_{ak} and U_{bl} are the ambiguity scores correspond to all the unique candidate pairs from image 1 to image2 and from image2 to image 1 respectively, the last m percent of matches in T_{ms} and the last m percent of matches in T_{ua} are removed as wrong matches from the sorted table of descending order. The left matches both with high similarity and low ambiguity are then taken as matched segments. The iteration is repeated until the total energy J of the left matches is converged. Since the matches obtained in this way are not symmetric, which means there may be several line segments in one image match to a single line in the other image, a score incorporating both similarity and ambiguity defined by $\gamma = M_s(\mathbf{l}_{1i}, \mathbf{l}_{2j}) + U_a(\mathbf{l}_{1i}, \mathbf{l}_{2j})$ is developed, the line segment with the largest γ in image 2 is then selected to be the match of the single line segment in image 1.

Robust Estimation

As stated in [98], any four point matches in general positions of the image (no three of them are collinear) can determine a homography matrix $\mathbf{H}$ and any four line matches in general positions of the image (no three of them are concurrent) can determine the homography matrix $\mathbf{H}^{-T}$. $\mathbf{H}$ and $\mathbf{H}^{-T}$ are both 3×3 , invertible matrices, they represent the 2D transformation between the two images to be matched, which constrains all the features in the image. Since the epipolar geometry is unable to be recovered from features on the same plane, the homography becomes very important information for finding more matches and understanding the underlying motion behavior in micromanipulation. However, it is difficult to estimate this homography matrix with noisy image data and possible outliers from micro images.

3.3. Line-Point Integrated Micro Image Matching

52

To obtain the homography reliably, robust estimation is necessary.

In the literature, several robust model fitting techniques have been implemented into computer vision, such as the M-estimator [139,140], LMS (least median of squares) [122] and RANSAC (random sample consensus) [141], where LMS and RANSAC are both random sampling techniques. It has been shown in [142] that in the range of the well-used robust estimators, random sampling techniques provide the best solution. However, when there are more than 50% outliers, LMS cannot provide a good solution. And M-estimator suffers when the initial estimate was poor. It has also been addressed [98,142] that if the M-estimator method was initialized using random sampling, the combination provides better results than random sampling alone. Hence in the homography estimation problem with line segment matches in micro images dealt with in this thesis, the RANSAC (Random Sample Consensus) algorithm [143] followed by optimization with M-estimator is adapted to recover the homography robustly.

The main idea of RANSAC is using Monte Carlo technique to draw N random minimum size subsamples s (for homography estimation $s = 4$) that is needed to determine the model. N is determined by the assumed probability of outliers ϵ , in such a way that the probability that at least one of the N subsamples is free of outliers is:

$$p = 1 - [1 - (1 - \epsilon)^s]^N \quad (3.3.11)$$

Thus N is given by:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \quad (3.3.12)$$

The design of distance measure for determining whether a match is inlier or not is crucial for RANSAC to work. For point cases, the symmetric error or reprojective

Figure 3.12: Illustration of different distance measure computation strategy

error is used, and the threshold according to the variance of the image noise level is set to judge the consistency of the putative matches with the homography estimated from each subsample [98]. However, for line segment matching case, the problem is more complicated.

For line segment matches, there are several ways to estimate the homography and define the distance measure, which are illustrated in Figure 3.12: 1.using the end points

3.3. Line-Point Integrated Micro Image Matching

54

of line segments (Figure 3.12(a)), 2. using the line parameters (Figure 3.12(b)), 3. using the combination of both line parameters and center point of the line segments (Figure 3.12(c)) and 4. using Hausdoff distance between two sets of geometric points (Figure 3.12(d)). Since end points of line segments are not well-defined in the image, the end points of segments in image 1 often differ from the end points of the corresponding segments in image 2, an example is shown in Figure 3.13, where the detected end points of segment a, b, c, d are different points on the line in the two images. Hence line parameters are preferred compared to end points of the line segments. On the other hand, if only the line parameters are used, it's difficult to disambiguate when multiple parallel line segment candidates are present. The integration of line parameters and points is advantageous in solving the difficulty for line based method when parallel lines are prevalent such as in micro image matching cases (Figure 3.13). Considering the above situations, the distance measure incorporating both line parameters and the center point of the line segment is introduced in Equation 3.3.13.

$$d_i = \sqrt{\gamma[d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i)^2 + d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2] + d_{geo}(\mathbf{c}_i, \mathbf{H}^{-1} \mathbf{c}'_i)^2 + d_{geo}(\mathbf{c}'_i, \mathbf{H} \mathbf{c}_i)^2} \quad (3.3.13)$$

Where $\mathbf{l}_i = [a, b, c]^T$ and $\mathbf{l}'_i = [a', b', c']^T$ are the vectors to parameterize the line where matched line segment candidates lie on, so that $ax + by + c = 0$, $a'x + b'y + c' = 0$. $\mathbf{c}_i = [x_1, x_2, x_3]^T$ and $\mathbf{c}'_i = [x'_1, x'_2, x'_3]^T$ are the homogeneous coordinates of the center of the line segment of a candidate pair in image 1 and image 2 respectively. γ is defined as the weighting factor to restrain the data from line segment matches and data from point matches to have the same magnitude, it can take the average ratio between point distance and line distance of a large amount of data. d_{geo} is the geometric error [98] defined in Equation 3.3.14.

$$d_{geo}(\mathbf{x}, \mathbf{y}) = ((x_1/x_3 - y_1/y_3)^2 + (x_2/x_3 - y_2/y_3)^2)^{1/2} \quad (3.3.14)$$

where $\mathbf{x} = [x_1, x_2, x_3]^T$, $\mathbf{y} = [y_1, y_2, y_3]^T$ are the homogeneous coordinates of points (or the three element vector of lines).

Figure 3.14: (a) image 1 (b) image 2

The conjecture of the distance measure is also validated by the following simulation. In this simulation, four line segments in general positions are generated (see Figure 3.14). Then the detected end points are simulated according to Gaussian distribution along the direction of the line so that the distance between the simulated end

3.3. Line-Point Integrated Micro Image Matching

56

Figure 3.15: Simulated end points on the line segments in image 1 and image 2

Figure 3.16: Distance computation with end points (a) distance (b) histogram of distance

points and the original end points are distributed with standard deviation $0.25l$, l is the length of the line segment. Then a Gaussian distributed random noise with standard deviation 0.5 is added to the generated end points (see Figure 3.15). The homographies are then computed in four ways illustrated in Figure 3.12. For case 1 (Figure 3.12(a)), the end points are directly used to compute the homography and the distance. For case 2 (Figure 3.12(b)), the line parameters are computed from the line determined by the end points simulated, then the homography and distance are computed with the line parameters. For case 3 (Figure 3.12(c)), the homography is computed with the line parameters, but the distance is computed based on both line parameters and the center points of the simulated line segments. For case 4 (Figure 3.12(d)), the homography and the simplified segment Hausdorff

3.3. Line-Point Integrated Micro Image Matching

57

Figure 3.17: Distance computation with line parameters (a) distance (b) histogram of distance

Figure 3.18: Distance computation with line parameters and center points of line segments (a) distance (b) histogram of distance

distance are computed based on the end points [144]. The distance computation results of 100 experiments are illustrated in Figure 3.16, Figure 3.17, Figure 3.18 and Figure 3.19.

Since the end points are simulated on the line segments, a good definition of distance measure should represent the Gaussian distributed random noise added on the end points but not sensitive to the position difference of the end points along the line. From the results, it can be observed that for case 1 and case 4, large distance appears often, which indicates that the computation of homography has been deteriorated by the uncertain end point positions. For case 2, although the distance is distributed more consistently than case 1 and case 4, the magnitude of the distance

Figure 3.19: Distance computation with Simplified Segment Hausdorff distance (a) distance (b) histogram of distance

is too small, and constraints on the magnitude is not available without knowing the point correspondence distance. For case 3, the distance measure shows consistent distribution, and the magnitude has been constrained.

From this simulation, it can be seen that the integration with both line segments and points is more advantageous in defining the distance measure for choosing inliers.

Guided Match The homography estimated by robust estimation described above can be used to find more matched features, and further refinement of the estimation can be made by M-estimator, which is a nonlinear optimization with certain cost function, for instance, using Lavenberg-Marquardt method [145, 146].

To include more point features in the optimization, the spots on the wafer image are detected by series of image processing procedures, such as binarization, close operation, and thresholding etc. With plenty of features detected in the image, both line segments and point features, two thresholds t_1, t_2 can be set for selecting new line matches and new point matches respectively. New line segment pairs and new point pairs satisfying (3.3.23) and (3.3.16) are taken into account to perform nonlinear optimization.

3.3. Line-Point Integrated Micro Image Matching

59

$$d_{li} = d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i)^2 + d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2 < t_1 \quad (3.3.15)$$

$$d_{pi} = d_{geo}(\mathbf{x}_i, \mathbf{H}^{-1} \mathbf{x}'_i)^2 + d_{geo}(\mathbf{x}'_i, \mathbf{H} \mathbf{x}_i)^2 < t_2 \quad (3.3.16)$$

In the above equations, $\mathbf{l}_i$ and $\mathbf{x}_i$ are the parameterizations of line and points respectively.

Nonlinear optimization is performed to minimize the reprojection error over the whole image so that not only the homography transformation but also the correspondence itself can be refined, the following cost function is defined:

$$\begin{aligned} f(\mathbf{H}) = & \gamma \sum_{i=1}^{n_1} (d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2 + d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i)^2) \\ & + \sum_{i=1}^{n_2} (d_{geo}(\mathbf{x}'_i, \mathbf{H} \mathbf{x}_i)^2 + d_{geo}(\mathbf{x}_i, \mathbf{H}^{-1} \mathbf{x}'_i)^2) \end{aligned} \quad (3.3.17)$$

In the right hand side of Equation 3.3.17, the first term is the summed geometric error of the line matches in the two images, the last term is the summed geometric error of the matched points in the two images. γ is to adjust the magnitude between point and line features. n_1, n_2 are the numbers of line segment matches and point matches respectively. The residual error computed in both images can thus be given as:

$$\begin{aligned} \epsilon_{res} = & \frac{1}{2} \left\{ \sqrt{\frac{\gamma}{4n_1} \sum_{i=1}^{n_1} (d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2 + d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i)^2)} \right. \\ & \left. + \sqrt{\frac{1}{4n_2} \sum_{i=1}^{n_2} (d_{geo}(\mathbf{x}'_i, \mathbf{H} \mathbf{x}_i)^2 + d_{geo}(\mathbf{x}_i, \mathbf{H}^{-1} \mathbf{x}'_i)^2)} \right\} \end{aligned} \quad (3.3.18)$$

Figure 3.20: Difference between symmetric transfer error and reprojection error. (a) symmetric transfer error (b) reprojection error. $\mathbf{x}$ and $\mathbf{y}$ are measured points, under the estimate $\mathbf{H}$ homography, the points $\mathbf{y}$ and $\hat{\mathbf{y}} = \mathbf{H}\mathbf{x}$ do not correspond perfectly. The estimated points $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, do correspond perfectly by the homography $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$. Using the notation $d(\mathbf{x}, \mathbf{y})$ for the Euclidean image distance between $\mathbf{x}$ and $\mathbf{y}$, the symmetric transfer error is $d(\mathbf{x}, \mathbf{H}^{-1}\mathbf{y})^2 + d(\mathbf{y}, \mathbf{H}\mathbf{x})^2$, the reprojection error is $d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{y}, \hat{\mathbf{y}})^2$

Optimized homography and location of new matches can be obtained through iterative Levenberg-Marquardt optimization of the cost function until the number of inliers converges. In this way, the reprojection errors of all the matched features are minimized (Figure 3.20 shows the difference between symmetric transfer error and reprojection error). It can be seen that rather than minimize the symmetric error, the perfectly matched points $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are sought out so that the correspondences are corrected at the same time as the homography is refined.

Implementation and Illustrative Examples

In this section, the experimental results on MEMS wafer micro images with LSG are presented. To illustrate the effectiveness of this method, we tested the algorithm on different wafer images. The micro images are taken with Pulnix TM-6702 camera that is mounted on the microscope, while the manipulator stage with the wafer positioned on undergoes unknown planner motions. The vision system consists of a digitizer and a frame grabber(Matrox).

Example 1 Line matching results for a pair of micro images are shown in Figure 3.21. Figure 3.22 shows the feature increasing process, the lower threshold is 0.7, the upper threshold is 0.9. The final matching results are presented in Figure 3.23 and Figure 3.24. In this example, the search window is defined within ± 150 pixels, the minimum length of line segments is 30 pixels, the inlier thresholds are $t_1 = 0.0001$, $t_2 = 1.5$. A total 800 samples are required. The guided matching needs 2 iterations. The root-mean-squared (RMS) error is reduced from 0.8574 to 0.6176 after the Levenberg-Marquardt optimization. The operation time is 124.9800s based on Pentium 4 CPU 3.00GHz, RAM 1.0GB, MatLab6.5.1.

3.3. Line-Point Integrated Micro Image Matching

62

Figure 3.21: (a)(b) detected line segments superimposed on the images. There are 120 line segments on image 1 and 131 line segments on image 2. (c)(d) 64 putative matches shown by numbers. (e)(f) inliers - 8 correspondences consistent with the estimated homography.

3.3. Line-Point Integrated Micro Image Matching

63

Figure 3.22: (a)spot detection in image 1. (b)spot detection in image 2.

Figure 3.23: (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 17 point matches and 19 line correspondences.

Figure 3.24: (a) point matches shown by the arrow linking point features. (b) yellow segments: line segments in image 1, red segments: line segments in image 2 projected back into image 1 according to the estimated homography.

Example 2 In this example, the rotational motion is present. Figure 3.25 shows the line segment matching results. Features in the image are increased by spot detection as shown in Figure 3.26, the lower threshold is 0.7, the upper threshold is 0.9. The final matched point features and line features are shown in Figure 3.27 and Figure 3.28. In this example, the search window is ± 150 pixels, the minimum length of line segments is 30 pixels. The inlier thresholds are $t_1 = 0.0001$, $t_2 = 1.5$. A total 2000 samples are required. The guided match required 4 iterations. The root-mean-squared (RMS) residual error is reduced from 0.9824 to 0.7477 after Levenberg-Marquardt optimization. The operation time is 138.3730s based on Pentium 4 CPU 3.00GHz, RAM 1.0GB, MatLab6.5.1.

3.3. Line-Point Integrated Micro Image Matching

65

Figure 3.25: (a)(b) detected line segments superimposed on the images. There are 116 line segments on image 1 and 106 line segments on image 2. (c)(d) 51 putative matches shown by numbers. (e)(f) inliers - 5 correspondences consistent with the estimated homography.

3.3. Line-Point Integrated Micro Image Matching

66

Figure 3.26: (a)spot detection in image 1. (b)spot detection in image 2.

Figure 3.27: (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 32 point matches and 16 line correspondences.

Figure 3.28: (a) point matches shown by the arrow linking point features. (b) yellow segments: line segments in image 1, red segments: line segments in image 2 projected back into image 1 according to the estimated homography.

Discussion

In this section, we presented Line Segment Based Line-Point Matching method(LSG) for micro image matching. The difficulties of matching repetitive patterns and inaccuracy caused by noises which is magnified by the microscopic imaging system are overcome by our proposed method with integration of grey level correlation based neighborhood voting, robust estimation and guided nonlinear optimization with both line segment features and point features. The matching and estimation results are shown to be satisfactory. The disadvantage of this method is that the grey level correlation computation is computationally expensive. For micro images with line segments of certain structure, the geometric information can be used to measure the similarity instead of computing the grey level correlation to reduce computational complexity. The next section is contributed to this approach.

3.3.2 L-shape Based Line-Point Matching (LSH)

Introduction

To reduce the burden of grey level correlation computation, in this section, a method to find primitive matched features making use of the geometric structure induced by grouped features is introduced, the matching results are constrained with homography transformation estimation. This method is a combination of direct matching and correlation computation of points.

On the surface of micro devices, there often exist junctions with certain angle (such as shown in Figure 3.1). These junctions are well located strong geometric features which also help to reduce ambiguity. If we can generate rough cues of the structure of the observed scene from these features, the search space can be reduced. Although such features are less numerous but they are rich in information, and their geometric characteristics can be used to reduce computational cost of grey level correlation.

In the literature, several methods utilizing such features described above have been proposed in the macro domain. In [51], the angle and length ratio of line segments are used to compute similarity invariants. Junctions of lines, such as T-junction is used for hypothesis testing to detect building structures in [147]. The orientation, length of line segments and intensity difference are applied to compute matching score [129]. The corner structure which consists of four line segments, with certain geometric structure specified, is used to calibrate the camera [148]. However, to cope with the problems of microscopic images, complicated corner structure is not suitable since they may not be found in micro images, on the other hand, the structure cannot be too simple or else it could be repetitive itself and disambiguating

3.3. Line-Point Integrated Micro Image Matching

69

will be difficult. Hence, the approach should make use of relatively simple structures and has to be able to disambiguate similar patterns.

Based on these requirements, we propose to use L-shape structure that is relatively simple and can be disambiguated by the angle and length of both sides. In the following, a method to group line segments and point into L-shape structure is presented, a novel similarity matching algorithm to generate matching scores is also developed for micro images. We expect that using the L-shape structure to compute goodness of similarity is able to reduce the computational complexity of grey level line correlation used in LSG method (Section 3.3.1). In LSH, the robust estimation and optimization of homography is also conducted. Once the homography is estimated, guided match can be employed to find new line matches and point matches. The processing steps for L-Shape Based Line-Point Matching (LSH) is given by Table 3.2.

Table 3.2: Processing Steps For LSH

1. Group features into L-shape and compute the similarity score - establish primitive matchings for candidate L-shapes
2. Robust estimation - estimate and optimize the transformation between two images
3. Guided match - find more matches and refine the transformation

Feature Grouping and Similarity Computation

The L-shape structure is defined with two line segments intersecting at one point. For line segments that do not intersect, the elongated ones are used to form L-shape (see Figure 3.29).

3.3. Line-Point Integrated Micro Image Matching

70

Figure 3.29: L-shape structure grouped with elongated line segments and point

After groups of L-shape features are found in both of the images to be matched, the similarity comparison can be performed. First, L-shapes that accord with the following conditions are selected to be the candidate matches:

$$\theta_1 < \theta \quad (3.3.19)$$

$$\theta_2 < \theta \quad (3.3.20)$$

$$|PP'| < d \quad (3.3.21)$$

where θ_1 is the angle between PP_1 and $P'P'_1$, θ_2 is the angle between PP_2 and $P'P'_2$, $|PP'|$ is the distance between P and P' , θ , d are the thresholds selected according to the a priori knowledge of the motion.

Then the similarity score can be defined as follows:

$$Score(L, L') = -\gamma \log(\alpha\beta) \quad (3.3.22)$$

where

3.3. Line-Point Integrated Micro Image Matching

71

$$\alpha = \sqrt{\frac{|||PP_1|| - ||P'P'_1|||}{\min(||PP_1||, ||P'P'_1||)} \cdot \frac{|||PP_2|| - ||P'P'_2|||}{\min(||PP_2||, ||P'P'_2||)}} \quad (3.3.23)$$

$$\beta = \frac{1}{\max(||PP_1||, ||P'P'_1||, ||PP_2||, ||P'P'_2||)} \quad (3.3.24)$$

$$\gamma = \text{cor}(P, P') + 0.5(\text{cor}(P_1, P'_1) + \text{cor}(P_2, P'_2)) \quad (3.3.25)$$

The larger $\text{Score}(L, L')$ is, the similar the two candidate shapes L and L' are. In the formulation, α denotes the relative length difference of both sides of the L-shape. It takes the normalized length difference. Since the line segment extraction could break the original line segments that are similar to each other, the product is to select the candidate L-shape with either side similar. For good candidate matches, $\frac{|||PP_1|| - ||P'P'_1|||}{\min(||PP_1||, ||P'P'_1||)}$, or $\frac{|||PP_2|| - ||P'P'_2|||}{\min(||PP_2||, ||P'P'_2||)}$ should be less than 1, if both line segments are similar in the two images, the product gives an even smaller value, hence the score defined in 3.3.22 is much larger, which indicates very similar patterns in the two images. β is used to weight the importance of the segments. Longer segments are relied on more than shorter ones in this design. The $\log$ function is to distinguish the very good matches. $\text{cor}(a, b)$ is the intensity difference in the nearby area of a and b in image 1 and image 2 respectively. The joint point is more important in determining the location of the L-shape, and is less easy to be affected by breaking of line segments due to feature detection errors, hence it takes more weight. Notice that, in this method, only point correlation is computed, which is much less computationally expensive than line correlation computation.

The L-shape structure pairs with high score are selected as potential matches. Since the L-shapes detected in the image is not massive, the disambiguation is done by taking the pair with the largest score. The initial matched L-shapes are processed with robust estimation to obtain the homography between two images.

Robust Estimation

Since the L-shape structure is used, the distance measure should be developed to include information from each element of the structure, namely, two segments and one point.

Consider the following geometric constraints: for two pairs of matched line segments $l_1 \leftrightarrow l'_1$ and $l_2 \leftrightarrow l'_2$, if they correspond to two coplanar line segments in the space, as shown in Figure 3.30, and satisfy:

$$l'_1 \simeq \mathbf{H}^{-T} l_1 \quad (3.3.26)$$

$$l'_2 \simeq \mathbf{H}^{-T} l_2 \quad (3.3.27)$$

then their intersections P and P' (in homogeneous form) must satisfy:

$$P' \simeq \mathbf{H} P \quad (3.3.28)$$

Figure 3.30: The relation of intersection between two coplanar lines in two images

3.3. Line-Point Integrated Micro Image Matching

73

With this relation, the integration of line and point parameters is more easier to be applied for L-shape. The distance measure to select inliers for L-shape is thus defined as follows:

$$d_i = \sqrt{\gamma[d_{geo}(\mathbf{l}_{1i}, \mathbf{H}^T \mathbf{l}'_{1i})^2 + d_{geo}(\mathbf{l}'_{1i}, \mathbf{H}^{-T} \mathbf{l}_{1i})^2 + d_{geo}(\mathbf{l}_{2i}, \mathbf{H}^T \mathbf{l}'_{2i})^2 + d_{geo}(\mathbf{l}'_{2i}, \mathbf{H}^{-T} \mathbf{l}_{2i})^2] + d_{geo}(\mathbf{P}_i, \mathbf{H}^{-1} \mathbf{P}'_i)^2 + d_{geo}(\mathbf{P}'_i, \mathbf{H} \mathbf{P}_i)^2} \quad (3.3.29)$$

L-shapes with $d_i < t$, where t is the threshold, are selected as inliers. In the above equation, γ is the weighting factor to restrain same magnitude of data from line segments and points, it can take the ratio of norm between the point and line vectors.

RANSAC robust estimation can thus be performed by repeating N times to randomly select 2 pairs of matched L-shapes, and compute $\mathbf{H}^{-T}$. Finally, $\mathbf{H}^{-T}$ with the largest inliers is chosen to be the correct one. N is determined by adaptive algorithm [98] similar to what was presented in Section 3.3.1.

Guided Match Once the homography is estimated through robust estimation, more matched features can be found efficiently with guided match. Similar to what was presented in section 3.3.1, new point features such as spots on the pattern and spots on the background of the wafer image can be found, then two thresholds are set to select new matches of line segments and spots in the image. Then the cost function (Equation 3.3.30) is optimized by Levenberg Marquart method iteratively, until no new matches are found.

$$\begin{aligned}
f(\mathbf{H}) = & \gamma \sum_{i=1}^n (d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2 + d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i))^2 \\
& + \sum_{i=1}^m (d_{geo}(\mathbf{x}'_i, \mathbf{H} \mathbf{x}_i)^2 + d_{geo}(\mathbf{x}_i, \mathbf{H}^{-1} \mathbf{x}'_i))^2 \\
& + \sum_{i=1}^p (d_{geo}(\mathbf{P}'_i, \mathbf{H} \mathbf{P}_i)^2 + d_{geo}(\mathbf{P}_i, \mathbf{H}^{-1} \mathbf{P}'_i))^2
\end{aligned} \tag{3.3.30}$$

In the above equation, n is the number of matched line segments which consists of two pairs of matched line segments for each matched L-shape and all the new matched line segments found by guided match, m is the number of matched L-shapes, p is the number of matched spots found by guided match. γ is to restrain the magnitude of data from line segments and points. In the right side of Equation 3.3.30, the first term is to compute the geometric error of all the line segment matches including both the ones from L-shape and new matched line segments obtained from guided match, the second term is to compute the geometric error of the matched L-shape junction points in both images, the last term is the geometric error of all the other matched spots found by guided match.

To measure the residual error in both images, the following Equation is adopted:

$$\begin{aligned}
\epsilon_{res} = & \frac{1}{3} \left\{ \sqrt{\frac{\gamma}{4n} \sum_{i=1}^n (d_{geo}(\mathbf{l}'_i, \mathbf{H}^{-T} \mathbf{l}_i)^2 + d_{geo}(\mathbf{l}_i, \mathbf{H}^T \mathbf{l}'_i))^2} \right. \\
& + \sqrt{\frac{1}{4m} \sum_{i=1}^m (d_{geo}(\mathbf{x}'_i, \mathbf{H} \mathbf{x}_i)^2 + d_{geo}(\mathbf{x}_i, \mathbf{H}^{-1} \mathbf{x}'_i))^2} \\
& \left. + \sqrt{\frac{1}{4p} \sum_{i=1}^p (d_{geo}(\mathbf{P}'_i, \mathbf{H} \mathbf{P}_i)^2 + d_{geo}(\mathbf{P}_i, \mathbf{H}^{-1} \mathbf{P}'_i))^2} \right\}
\end{aligned} \tag{3.3.31}$$

Implementation and Experiment Results

In this section, the experimental results on MEMS wafer micro images with LSH are presented. We use the same two set images used for LSG.

Example 1 The L-shape matching results are shown in Figure 3.31. As the same thresholds are utilized to detect spots on the image, the same results are obtained for feature increasing process as presented in Figure 3.22. The final matched point features and line features are presented in Figure 3.32 and Figure 3.33. In this example, the distance threshold is $\pm 150pixels$, the angle threshold is $\pm 40deg$, the point correlation window size is $5 \times 5pixels$. The minimum length of line segments to form L-shape is $20pixels$. The inlier threshold is 2. A total 3 samples are required for RANSAC robust estimation. 3 iterations are needed for guided matching. The RMS (root-mean-squared) residual error computed according to (3.3.31) after Levenberg-Marquardt optimization is 0.5148 (before 0.8713). The operation time is 54.6160 s based on Pentium 4 CPU 3.00GHz, RAM 1.0GB, MatLab6.5.1.

3.3. Line-Point Integrated Micro Image Matching

76

Figure 3.31: (a)(b) grouped L-shapes (red) based on the detected line segments (yellow) superimposed on the images. There are 258 line segments and 71 L-shapes on image 1, 255 line segments and 78 L-shapes on image 2. (c)(d) 35 putative matched L-shapes shown by numbers. (e)(f) inlier - 7 corresponding L-shapes consistent with the estimated homography.

3.3. Line-Point Integrated Micro Image Matching

77

Figure 3.32: (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 17 point matches and 45 line correspondences.

Figure 3.33: (a) point matches shown by the arrow linking point features. (b) yellow segments: matched line segments in image 1, red segments: matched line segments in image 2 projected back into image 1 according to the estimated homography.

3.3. Line-Point Integrated Micro Image Matching

78

Example 2 Figure 3.34 shows the L-shape matching results. Since the same threshold is used to increase the features, the same feature increasing results are obtained as shown in Figure 3.26. The final matched point features and line features are shown in Figure 3.35 and Figure 3.36. In this example, the distance threshold is $\pm 150pixels$, the angle threshold is $\pm 40deg$, the point correlation window size is $5 \times 5pixels$. The minimum length of line segments to form L-shape is $20pixels$. The inlier threshold is 2. A total 3 samples are required for RANSAC robust estimation. 2 iterations are needed for guided match. The RMS residual error computed according to (3.3.31) after Levenberg-Marquardt optimization is 0.6329 (before 0.9021). The operation time is 34.9990 s based on Pentium 4 CPU 3.00GHz, RAM 1.0GB, MatLab6.5.1.

3.3. Line-Point Integrated Micro Image Matching

79

Figure 3.34: (a)(b) grouped L-shapes (red) based on the detected line segments (yellow) superimposed on the images. There are 170 line segments and 44 L-shapes on image 1, 158 line segments and 41 L-shapes on image 2. (c)(d) 8 putative matched L-shapes shown by numbers. (e)(f) inlier - 5 corresponding L-shapes consistent with the estimated homography.

3.3. Line-Point Integrated Micro Image Matching

80

Figure 3.35: (a)(b) image 1 and image 2 with final set of correspondences after guided match and maximum likelihood estimation superimposed, there are 28 point matches and 19 line correspondences.

Figure 3.36: (a) point matches shown by the arrow linking point features. (b) yellow segments: matched line segments in image 1, red segments: matched line segments in image 2 projected back into image 1 according to the estimated homography.

Discussion

In this section, an image matching technique with grouped features, L-shape Based Line-Point Matching (LSH), is presented. Instead of the grey level correlation computation of line segments, the invariance from grouped structure as initial similarity measurements is used, based on the assumption that the apparent motion between two images can be approximated by a planar geometric transformation, specifically, similarity. Under this assumption, the relative length difference is used as invariance to compare goodness of similarity, further, the constrain of angle of L-shape is used to reduce the number of potential matching candidates. To encode the grey level similarity, the correlation score of the end and junction points of L-shape is incorporated into the similarity comparison process. The junction point is put on high weight, since it's more significant in representing the location of L-shape and more robust to line segmentation errors. After initial computation of matches, robust estimation and guided match are performed to increase the number of matched features and optimize the homography estimate. The experiment with MEMS wafer images show satisfactory results with reduced computational cost. However, there are also some drawbacks of this method, since L-shapes may not be able to be acquired from microscopic images sometimes, for instance, only parallel lines are in the image etc, this method could fail on less structured microscopic images.

3.3.3 Comparison Between LSG and LSH

The performance of LSG and LSH for sets of micro images (Figure 3.37) are compared in Table 3.3. The running time is calculated based on Pentium 4 CPU 3.00GHz, RAM 1.0GB, MatLab6.5.1. Since in the feature grouping process, many line segments have been removed from consideration except those who can form

L-shapes, those line segments forming putative matched L-shapes detected by LSH is much less than the number of putative matched line segments detected by LSG, hence line segments detected by LSG often present more information in the image, so that LSG is more robust to illumination changes. This is validated from the experiment results, unless in the cases when the number of features is extremely small, such as in set 9, which needs the inlier threshold to be increased, normally inlier threshold $1.5 - 2$ is sufficient for LSG to achieve good results regardless of the illumination changes. However, the inlier threshold for LSH has to be tuned according to the illumination situation, for instance, threshold for image set 10 is 4, while the inlier threshold for image set 1 is 5. We can also see that, the time cost of LSH is much smaller than LSG, as stated already this is due to the correlation computation has been replaced by more efficient invariance computation. The computational time is reduced by 37.12% in average. From this analysis, it can be seen that efficiency and accuracy are two factors that have to be trade off. If more efficient process is needed, the parameters have to be modified according to the changing environment, if the accuracy is needed, efficiency has to be sacrificed to some extent.

Figure 3.37: Sets of micro images for testing

3.3. Line-Point Integrated Micro Image Matching

84

Table 3.3: Comparison of LSG and LSH. SR: sample required. IT: inlier threshold. NL: number of line matches. NP: number of point matches. RE: RMS error. TC: time cost.

Category	SR	IT(pixel)	NL	NP	RE	TC(s)
LSG(set1)	2000	2	24	26	0.6610	118.6490
LSH(set1)	60	5	65	27	0.6737	57.9750
LSG(set2)	1000	2	23	21	1.0958	47.1210
LSH(set2)	50	3	17	22	1.4094	30.3210
LSG(set3)	500	2	21	29	0.8299	41.0320
LSH(set3)	10	5	29	24	0.7458	35.2520
LSG(set4)	2000	2	38	39	0.8861	41.0000
LSH(set4)	30	5	42	39	0.6329	33.0370
LSG(set5)	800	2	25	20	1.0422	47.8690
LSH(set5)	30	5	6	13	1.0476	27.5930
LSG(set6)	800	2	53	71	0.7913	111.4490
LSH(set6)	30	2	52	43	0.5679	43.2610
LSG(set7)	2000	2	38	22	0.4503	80.9400
LSH(set7)	60	4	34	20	0.9482	57.2390
LSG(set8)	1000	2	34	20	0.7065	52.9710
LSH(set8)	100	4	17	32	0.8525	45.4520
LSG(set9)	500	3	41	29	0.8262	66.6230
LSH(set9)	80	2	39	15	0.3519	53.8930
LSG(set10)	2000	2	48	25	0.4983	96.8433
LSH(set10)	50	4	34	31	0.8921	72.4715
LSG(set11)	500	2	44	28	0.7649	87.0318
LSH(set11)	70	2	39	17	0.6487	49.7480
LSG(set12)	800	2	31	33	0.7184	108.5746
LSH(set12)	50	2	29	25	0.4105	59.6640

3.4 Evaluation of Line-Point Integrated Matching Approaches

In this section, the line-point integrated matching approach is evaluated through simulations. The aim is to test the performance of the proposed methods under different situations and investigate the factors that affect the results, especially, the necessary requirements that make the proposed methods work are studied and the possible pitfalls that make them fail are discovered. The main idea lying in the proposed two methods, MSG and MSH, is an integration of line segment and point features, the difference is MSG uses gray level correlation, while MSH uses similarity invariance to obtain the initial putative matches. In the following, the factors that are common to both methods are investigated, so that the analysis will be applicable for both MSG and MSH.

3.4.1 Feature Quality Requirements

To apply the proposed approaches, line detection methods have to be adopted first, however the truncation error and segmentation error often accompany with the detection results. Bad feature quality will probably lead to large percentage of wrong matches. In the following study, the underlying problems of feature quality in micro image matching are investigated. The noise associated with each feature is used to represent the quality of features, the line parameters are computed from the two end points of the line segments, the point parameters are extracted from the points simulated directly. Monte Carlo simulations are conducted on the proposed line-point integration based algorithms with simulated noise. The simulation steps are illustrated in Table 3.4. The distribution of noise is simulated as random number that is evenly distributed in the interval $(0, 3)$. In this simulation, n is taken as 100,

3.4. Evaluation of Line-Point Integrated Matching Approaches 86

thus there are 100 point features and 100 line segment features in each image, and N is taken as 10000. The initial correct putative matching rate represents the percentage of features that are assumed to be correctly matched by putative matching through similarity comparison.

Table 3.4: Monte Carlo simulation steps for evaluating line-point integration based matching approaches with different feature quality

<p>Repeat N times</p> <ul style="list-style-type: none"> • Generate uniformly distributed $n \geq 4$ image points in the image plane with noise. • Generate uniformly distributed $2n$, ($n \geq 4$) end points in the image with noise to construct line segments. • Use predefined homography to transfer the points and line segments into the other image, add random noise to the transferred features • Obtain putative matches according to the rate of correct putative matches • Use RANSAC to robustly compute the homography and find line match inliers • Use guided match to find more consistent point and line matches. <p>Compute the statistical properties of the results</p>

3.4. Evaluation of Line-Point Integrated Matching Approaches

87

The distribution of the RMS error computed according to Equation 3.3.18 obtained from the simulation is shown in Figure 3.38. It shows that the RMS error appears to be more sensitive to point noise than line segment noise on the end points. Figure 3.38(a) is nearly evenly distributed regardless of the noise level on the end points of the line segments, however, when the noise on the points increases, the RMS error increases accordingly. This can be explained by the reason that the computation on line segments is basically on the direction of lines, the noise on the end point has not as much ability as the noise on point features to influence the matching. However, when the line segments are very short, we can expect this effect of noise on end points of line segments to increase. This result also indicates that during guided matching, the threshold to select more features have to be carefully selected, especially for the point features, if points with large image noise is selected, it could deteriorate the matching results in terms of high RMS error. Figure 3.39 shows the result of the RMS error with respect to feature noise changing with different initial correct putative matching rate, from which, we can observe that the higher the initial putative matching rate the lower the RMS error, while the lower the initial putative matching rate the higher the final RMS error. This also shows the importance of getting more correct initial matches before robust estimation and guided match.

Figure 3.38: (a) Distribution of RMS error wrt. line segment noise (b) Distribution of RMS error wrt. point noise

3.4. Evaluation of Line-Point Integrated Matching Approaches

88

(a)

(b)

(c)

Figure 3.39: (a) RMS error wrt. changing of feature noise with initial correct putative matching rate 70%. (b) RMS error wrt. changing of feature noise with initial correct putative matching rate 50%. (c) RMS error wrt. changing of feature noise with initial correct putative matching rate 20%.

3.4. Evaluation of Line-Point Integrated Matching Approaches

89

The distribution of inlier rate is shown in Figure 3.40. It can be seen in this result that the rate of inliers is greatly related to the noise of line segments but is rarely correlated to the point noise. As the noise on end point of line segments increases, the rate of inliers found decreases. And if the noise is too large, the number of inliers could reduce to be very small so that it prevents the robust estimation of homography to be successfully executed, for instance the number of inliers is less than 4. This could probably lead to failure of matching. There are two strategies to prevent this situation, one is to increase the total number of line segments, which can be achieved by tuning the threshold in detecting line segments, the other is to increase the rate of inliers by reducing the noise in the image. As we can see from previous description, the homography estimation plays crucial part in the whole matching process, so to obtain good matching results, the effective number of inliers (≥ 4) have to be assured.

Figure 3.40: (a) Distribution of inlier rate wrt. the noise on end point of line segments (b) Distribution of inlier rate wrt. the noise on points

In Figure 3.41, the time consumption related to different noise level is shown. Since this process does not include the time taken by the putative matching process, it

3.4. Evaluation of Line-Point Integrated Matching Approaches

90

only indicates the underlying relations but cannot show the absolute time cost. The results show close correlation between the noise level on the line segments to the time cost during matching. Intuitive conclusion can be made that the larger the noise on line segments, the slower the process. However, this relation does not have the same effect on point features.

Figure 3.41: (a) Distribution of time cost wrt. the noise on end point of line segments
(b) Distribution of time cost wrt. the noise on points

Finally, the histogram of RMS error of all the simulated instances is shown in Figure 3.42. This figure shows the overall distribution of RMS error within the range of noise that is simulated.

In summary, the noise on point features has more effect on the RMS error than the noise on the end point of line segments, but the number of inliers and the time cost are more depended on noise on the end point of line segments. To ensure good matching results, the effect of noise on both points and line segments has to be considered. Noise on line segments has crucial part in determining whether the matching will succeed or not, if the number of inliers is too small, the process will

Figure 3.42: Distribution of RMS error under feature quality changes

probably fail. If the noise on point features is too large, the results could be deteriorated in terms of matching with large RMS error. Thus, we can see that the feature quality is very important for the whole matching process, to obtain good matching results, the feature has to be well detected, and the noise associated with it has to be relatively small.

3.4.2 Putative Matching Rate Requirements

From observation of the experiments, it shows that the initial matching rate of putative matches often influence the final matching results. For instance, when only a few features are matched during putative matching process, good final matching results cannot be guaranteed. This problem will be analyzed in the following.

To investigate the effect of this factor, Monte Carlo simulation is conducted to simulate the situations of different putative matching rate. The rate of putative point matching rate can be considered as the rate of initial recovered point matches

3.4. Evaluation of Line-Point Integrated Matching Approaches 92

with respect to total matches. The rate of putative line matching rate can be considered as the rate of initial recovered line matches through similarity comparison with respect to the total matches. The simulation steps are stated in Table 3.4. The noise on both point features and end point of line segments is fixed to a uniformly distributed random noise on the interval $(0, 1)$. The putative matching rate is simulated as uniformly distributed random number on the interval $(0, 1)$ for line segment features and point features respectively. $n = 100$ points and line segments are used. N is taken as 10000.

From the simulation, it can be found out that the rate of correct putative point matches is largely related to the number of point matches finally recovered, Figure 3.43 shows the close relation between them. However this is not necessary the case for line segments. As can be seen from Figure 3.44, low rate of putative line segment matches does not necessarily related to low rate of final matching rate of line segments although the simulation shows some tendency. Figure 3.45 shows the relation between RMS error and the rate of correct putative matches of line segments and points. In this figure, the brown surface indicates unsuccessful cases of matching. From this figure, we can see that when the rate is very low (below 0.2), the matching cannot succeed, which means very low correct putative matching rate will fail the whole matching process. To ensure the matching can be processed successfully, the rate has to be kept above 0.2.

3.4. Evaluation of Line-Point Integrated Matching Approaches

93

Figure 3.43: rate of final matched points vs. rate of correct putative point matches

Figure 3.44: rate of final matched line segments vs. rate of correct putative line matches

Figure 3.45: RMS error with respect to rate of putative matches of line segments and points

Figure 3.46 shows the time cost according to the rate of correct putative matches. The time is more sensitive to the rate of putative point matches than to the line segment matches. The increase of time cost with increase of the rate of putative point matches is obvious from the figure, it also shows some tendency of increasing time cost with higher rate of putative line matches. This is due to the reason that as the correctness rate increases, more features will be included into the optimization process which will take more time to process with a larger scale optimization problem. Finally, the distribution of RMS error is shown in Figure 3.47, which implies the distribution of matching error in 10000 experiments of random selected putative matching rate of line segments and points.

3.4. Evaluation of Line-Point Integrated Matching Approaches

95

Figure 3.46: (a) Distribution of time cost wrt. rate of correct putative line matches
(b) Distribution of time cost wrt. rate of correct putative point matches

Figure 3.47: Distribution of RMS error under changing of putative matching rate

3.4. Evaluation of Line-Point Integrated Matching Approaches 96

To summarize, the correct putative matching rate is not very essential as long as the rate is above the necessary value. Although it does have certain effect on the number of final matched features and the time cost, its effect on the matching process is not as critical as the feature quality. However, the feature quality may influence the putative matching rate in an indirect way, for instance, when salient features with relatively small noise are prominent, it will be easier to carry on the putative matching process, and correct putative matching rate could probably increase at these circumstances.

3.4.3 Feature Density Requirements

It is intuitive to notice that, the density of feature has certain effect on the matching process. The investigation of this relation will be covered in the following.

Monte Carlo simulation is conducted according to Table 3.4. In this testing, the number of features n is simulated as a random number evenly distributed in the interval (10, 510) for line segments and points respectively. Correct putative matching rate is fixed as 70%, and the feature noise is fixed to be following uniform distribution in the interval (0,1). N is taken as 10000. Since in this simulation the length of line segments is not restrained, the results represent the general comparison between point features and line segment features.

Figure 3.48 shows the relation between the RMS error and the feature density. It indicates that the RMS error will decrease with the increase of line features but will increase with the raise of point features. This result also validates the fact that line features are more robust than point features. Similar observations can be obtained in Figure 3.49, where the surface is correlated to the RMS error under different

3.4. Evaluation of Line-Point Integrated Matching Approaches

97

Figure 3.48: (a) Distribution of RMS error wrt. No of line features (b) Distribution of RMS error wrt. No of point features

density of line segment features and point features. This result also shows that improvement of matching can be achieved by increasing the number of line segment features in the image.

Close relationship between the number of line segment inliers after RANSAC and the density of line segment features can be shown in Figure 3.50. The number of inliers increases with the number of line features in the image. This result is rather intuitive, since the more line segment features are in the image, the more matched inliers consistent with the homography are likely to be found. We can also notice that the cloud disperses with increase of number of line segments. This is probably due to the increased uncertainties in the number of inliers when there are more line segment features involved in the image.

Figure 3.51 shows the relation between time cost and the feature density. It shows that the time cost increases as the density of feature increases regardless of type

3.4. Evaluation of Line-Point Integrated Matching Approaches **98**

of feature. Finally, the distribution of RMS error in 10000 experiments with the changing of feature density is illustrated in Figure 3.52.

3.4. Evaluation of Line-Point Integrated Matching Approaches

99

(a)

(b)

Figure 3.49: (a) RMS error wrt. changing of feature density with initial correct putative matching rate 70% (b) RMS error wrt. changing of feature density with initial correct putative matching rate 50%

3.4. Evaluation of Line-Point Integrated Matching Approaches 100

Figure 3.50: Distribution of No of line segment inliers wrt. No of line features

Figure 3.51: (a) Distribution of time cost wrt. No of line features (b) Distribution of time cost wrt. No of point features

Figure 3.52: Distribution of RMS error with feature density change

To summarize, feature density affects both the goodness of matching and the time cost during matching. The effect of point feature density and line feature density to RMS error is in reverse. If too many points are in the image, the effect of noise associated with them will probably reduce the accuracy of matching, but if the number of line segments increases the result will be improved. However, line segments are not reliable in some particular circumstances such as when the most of line segments are parallel, the direction perpendicular to the parallel line direction will be much more difficult to determine with pure line segment features. Hence, proper number of point features are necessary to guarantee the proposed methods. As to the time cost, large density of features will lead to more processing time. From the above analysis, we can see that it requires careful consideration to make trade-off between efficiency and accuracy when matching with micro images.

3.5 Summary

In this chapter, line-point integration based micro image matching approaches are proposed, LSG and LSH. LSG uses the matching strength to find putative matches through comparison of gray level correlation and neighborhood voting. LSH takes advantage of the similarity invariants such as ratio of length, angle, which lead to less costly computation in obtaining putative matches. In both methods, robust estimation, guided matching and nonlinear optimization are implemented to acquire an improved estimate of the homography and increased number of matches. The aim of the proposed methods is to reduce the effect of uncertainties in processing with micro images and obtain good matching results regardless of repetitive patterns and image noises. The experiment results validate the proposed methods in matching MEMS wafer images under microscope.

Simulation of quantitative data is conducted to evaluate the proposed methods and investigate the underlying factors that influence the matching results. From the simulation, it is found out that feature quality is very important to ensure success of the proposed methods. The final matching error is more sensitive to point noise than line segment noise, but quality of line segment feature have great effect on the number of inliers and the time cost. The part of feature density is depended on the type of feature. Large density of point features will probably lead to increase of matching error but large density of line segment features often reduces the error. And dense features often lead to long computational time. Although the correct putative matching rate does not have as much effect as the previously mentioned two factors, the correctness rate of putative point matches determines the number of final matched points and large correctness rate is related to long processing time. Hence in applying the proposed methods, to guarantee good matching results, the

noise associated with the features cannot be too big so that the RMS error won't be too high and the number of inliers should be ensured so that the matching process can be carried out successfully. Furthermore, to complete the process efficiently, feature cannot be too dense, especially the point features. The putative matching rate has to be at least above 0.2 to avoid failure.

Chapter 4

Micro Image Tracking

4.1 Introduction

Visual tracking and reconstruction play important roles in many macro view tasks [149–151]. Since structural ambiguity normally exists with two-frame motion in macro domain [152], many research activities have been established to infer the object pose with sequence of frames [153, 154]. In micro domain, feature identification, tracking and localization are very time consuming processes. This is because the features have a high degree of symmetry, the view is very narrow compared to the size of the scene, and the kinematics of the manipulator stage are not intuitive to be known. From the transformations between images, it is possible to recover the object motion. Especially for rotations, as However, robust and physically consistent solutions are difficult to be obtained. To enforce the reconstruction to be consistent with the kinematics constraints and reduce the dimension of computation so that the results are less sensitive to noise, it is necessary to make considerable use of prior information or regularization techniques.

In this chapter, a homography based tracking (HBT) technique of micro images for the coarse-fine micromanipulation is proposed, in which two views of the scene are provided, the micro view and the perspective view. The main idea is to estimate the kinematic parameters from the homography obtained during the image matching, accumulate sequence of micro images to build a mosaic. This mosaic represents a map of the visited areas from the microscope, and can be used to register micro images with the perspective view or present an enlarged field of view of microscope. In other way, this mosaic provides information in different views, both macro view and micro view. Once the kinematics of the manipulator stage is recovered, the relation between features from different views can be established, hence cheaper tracking system without extensive calibration and anti-vibration configuration that is easy to setup is possible to be achieved. This idea is illustrated in Figure 4.1, where H_u is the homography between every two consecutive micro images. The motion of manipulator stage is not always necessarily planar, however, to simplify the problem, the motion can be locally approximated by general planar motions, where physical constraints and prior knowledge can be used to refine the solutions. Therefore, in this chapter the planar motion is the main concern.

Figure 4.1: Illustration of visual tracking for micro manipulation

From another point of view, in micro images the limited field of view makes it very difficult to tell the difference between rotation and translation, furthermore, the decomposition methods such as eigen value decomposition are usually unstable in the presence of small disturbances which occur very often in micro environment. But any position obtained through planar motion can be realized by carrying out translation and rotation separately, the problem will be much more straightforward and the physical constraints are more easily to be used knowing the motion is translation or rotation. Hence for the ease of analysis, the two cases of general planar motion, translation and rotation are investigated separately. The following scenarios are designed in this chapter (see Figure 4.2 and Figure 4.3) to realize the proposed ideas of HBT. In Figure 4.2 and Figure 4.3, a MEMS wafer is placed on the manipulator stage which is undergoing translational motion or rotational motion with constant speed respectively. Sequence of images are taken from the microscope and the CCD camera which provides the global view. Then with only the type of motion (translation or rotation) provided, the tracking is conducted without any knowledge of the camera calibration or exact motion parameters. The detailed process and results will be illustrated in the following sections.

Figure 4.2: Translational scenario

This chapter is organized as follows. Section 4.2 introduces the Bayesian tracking. In Section 4.3, the proposed method is introduced. The implementation and ex-

Figure 4.3: Rotational scenario

periment results are presented in Section 4.4. Summary of this chapter is made in Section 4.5.

4.2 Bayesian Tracking

In the problem of tracking [155–157], the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$ of a target is given by:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}) \quad (4.2.1)$$

where $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_x}$ is possibly nonlinear function of the state $\mathbf{x}_{k-1}$, $\{\mathbf{w}_{k-1}, k \in \mathbb{N}\}$ is an i.i.d. (independent and identically distributed) process noise sequence, n_x, n_w are dimensions of the state and process noise vectors, respectively, and $\mathbb{N}$ is the set of natural numbers. The objective of tracking is to recursively estimate $\mathbf{x}_k$ from measurements:

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{v}_k) \quad (4.2.2)$$

where $h_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_z}$ is possibly nonlinear function, $\{\mathbf{v}_k, k \in \mathbb{N}\}$ is an i.i.d

measurement noise sequence, and n_z, n_v are dimensions of the measurement and measurement noise vectors, respectively. The tracking problem is to seek the estimates of $\mathbf{x}_k$ based on the set of all available measurements $\mathbf{z}_{1:k} = \{z_i, i = 1, \dots, k\}$ up to time k .

From a Bayesian perspective [4, 158, 159], the tracking problem is to recursively calculate some degree of belief in the state $\mathbf{x}_k$ at time k , taking different values, given the data $\mathbf{z}_{1:k}$ up to time k . Thus it is required to construct the pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. It is assumed that the initial pdf $p(\mathbf{x}_0|\mathbf{z}_0) \equiv p(\mathbf{x}_0)$ of the state vector, which is also known as the prior, is available. Then, in principle, the pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ may be obtained, recursively, in two stages: prediction and update.

Suppose that the required pdf $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ at time $k-1$ is available. The prediction stage involves using the system model (4.2.1) to obtain the prior pdf of the state at time k via the Chapman-Kolmogorov equation:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \quad (4.2.3)$$

In (4.2.3), the fact that $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ described as a Markov process of order one has been used. The probabilistic model of the state evolution $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is defined by the system equation (4.2.1) and the known statistics of $\mathbf{w}_{k-1}$. At time step k , a measurement $\mathbf{z}_k$ becomes available and this may be used to update the prior via Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (4.2.4)$$

where the normalizing constant:

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k \quad (4.2.5)$$

depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$ defined by the measurement model (4.2.2) and the known statistics of $\mathbf{v}_k$. In the update stage (4.2.4), the measurement $\mathbf{z}_k$ is used to modify the prior density to obtain the required posterior density of the current state. The recurrence relations (4.2.3) and (4.2.4) form the basis for the optimal Bayesian solutions. Although in some strict set of cases, solutions exist using Kalman filter or grid-based filters, the recursive propagation of the posterior density cannot be determined analytically due to the difficulty in solving the integration [4, 155].

An important method for integration especially for high dimension is Monte Carlo method [160, 161]. The Monte Carlo method approximates a definite integral by uniformly sampling from the domain of integration, and averaging the function values at the samples. However, there are two major limitations to the basic Monte Carlo approach: 1) the accuracy improves only linearly with the number of samples and 2) more samples are needed if the integral has peaks in some small regions and is very small in other area [155].

To handle these limitations, there exists several methods such as importance sampling, rejection sampling, Gibbs sampling and various combined techniques. Sequential Monte Carlo methods [157, 158], otherwise known as particle filter [83, 162], or condensation [163], are very popular in recent years as a numerical approximation approach to compute the tracking recursion for complex models. Particle filter, a sequential Monte Carlo based method, allows for a complete representation of the state distribution using sequential importance sampling, resampling and has some of the good properties of both importance sampling and Markov chain Monte Carlo

(MCMC). On the one hand, the samples are properly weighted, and the significance of a sample is represented by its weight, on the other hand, by reusing the samples, particle filter can keep track of a slowly varying density [155]. Furthermore, particle filter has features such as simplicity, flexibility, ease of implementation, and has been successfully used in modelling a wide range of challenging applications [157]. Due to these advantages, in the micro image tracking problem discussed in this thesis, a tracking system integrating particle filter estimation method is proposed.

To resolve the common problem with the sequential importance sampling particle filter, the degeneracy phenomenon (where after a few iterations, all but one particle will have negligible weight), the resampling procedure can be used [4, 158], where some suitable measure of degeneracy such as the effective sample size need to be implemented [158, 164]. In the following, the particle filter is briefly introduced.

Let $\{\mathbf{x}_{0:k}^i, w_k^i\}$ denote a random measure that characterizes the posterior pdf $p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})$, where $\{\mathbf{x}_{0:k}^i, i = 0, \dots, N\}$ is a set of support points with associated weights $\{w_k^i, i = 1, \dots, N\}$ and $\mathbf{x}_{0:k} = \{\mathbf{x}_j, j = 0, \dots, k\}$ is the set of all states up to time k . $q(\cdot)$ is the proposal distribution called the importance density. Then if the samples $\mathbf{x}_{0:k}^i$ were drawn from an importance density $q(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})$, then the weights are given by:

$$w_k^i \propto \frac{p(\mathbf{x}_{0:k}^i|\mathbf{z}_{1:k})}{q(\mathbf{x}_{0:k}^i|\mathbf{z}_{1:k})} \quad (4.2.6)$$

By expanding the proposal distribution as $q(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k})q(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1})$, where the assumption is made that the current state is not dependent on future observations. Furthermore, under the assumption that the states correspond to a Markov process and that the observations are conditionally independent given the

4.2. Bayesian Tracking

112

states, the recursive modified weight is then:

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{z}_k^i | \mathbf{x}_k^i) p(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i, \mathbf{z}_k)} \quad (4.2.7)$$

With this derivation, the Generic Particle Filter algorithm can be summarized in Table 4.1, where N_{eff} is the effective sample size to determine whether a significant degeneracy phenomena has occurred. When the degeneracy happens the resampling process is conducted to reduce the effect.

Table 4.1: Generic particle filter algorithm [4]

<ul style="list-style-type: none"> • For $i = 1:N$ <ul style="list-style-type: none"> – Draw $\mathbf{x}_k^i \sim q(\mathbf{x}_k \mathbf{x}_{k-1}^i, \mathbf{z}_k)$ – Assign the particle a weight, w_k^i, according to $w_k^i \propto w_{k-1}^i \frac{p(\mathbf{z}_k^i \mathbf{x}_k^i) p(\mathbf{x}_k^i \mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i \mathbf{x}_{k-1}^i, \mathbf{z}_k)} \quad (4.2.8)$ end For • Normalize the importance weights: <ul style="list-style-type: none"> For $i = 1:N$ <ul style="list-style-type: none"> – Normalize: $w_k^i = \frac{w_k^i}{\sum_{j=1}^N w_k^j}$ end For • Calculate N_{eff}, according to $N_{eff} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}$ • If $N_{eff} < N$ <ul style="list-style-type: none"> – resample, according to resampling algorithm (Table 4.2) end If

The resampling algorithm is stated in Table 4.2, where the basic idea is to eliminate particles that have small weights and to concentrate on particles with large weights. The resampling step involves generating a new set $\{\mathbf{x}_k^j\}_{j=1}^N$ by resampling N times

from an approximate discrete representation of $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. The resulting sample is in fact an i.i.d. sample; therefore, the weights are reset to $w_k^j = 1/N$.

Table 4.2: Resampling algorithm [4]

• Initialize the CDF: $c_1 = 0$ • For $i = 2 : N$ – Construct CDF: $c_k^i = c_k^{i-1} + w_k^i$ end For • Start at bottom of the CDF: $i = 1$ • Draw a starting point: $u_1 = \mathbb{U}[0, N^{-1}]$ • For $j = 1:N$ – Move along the CDF: $u_j = u_1 + \frac{j-1}{N}$ – While $u_j > c_k^i$ $i = i + 1$ end While – Assign sample: $\mathbf{x}_k^j = \mathbf{x}_k^i$ – Assign weight: $w_k^j = 1/N$ end For

In the following discussion, particle filter paradigm will be used to model the micro image tracking problem. The tracking problem will be considered provided knowing the type of motion. The translational motion and rotational motion will be investigated independently. In translational motion, the essential variables to be recovered is the inter image translational displacement between two consecutive images with respect to the camera, presented in pixel coordinates. The kinematic parameters need to be obtained from rotational motion is the rotation axis and rotation angel with respect to the camera in pixel coordinates. Suitable models need to be designed to present these parameters in the Bayesian process of particle filter estimation.

4.3 Homography Based Micro Image Tracking (HBT)

In this section, the matching results and the homography transformation computed from Chapter 3 are used to estimate the kinematics of the manipulator stage. Since on the one hand, the motion such as rotation follows nonlinear model and the noise which is not necessarily Gaussian in micro images is a significant factor for tracking, on the other hand, particle filters provide robust tracking frameworks as they are neither limited to linear systems nor require the noise to be Gaussian, in this thesis, it is proposed to use particle filter for tracking the kinematics variables as they evolve over time.

The main idea of this method is to design the model representation of kinematics parameters from the homography relation, construct a sample-based representation of the entire pdf from the matching results. Multiple copies (particles) of the kinematics variable are used, each one associated with a weight that signifies the quality of that specific particle. An estimate of the kinematics variables is obtained by the weighted sum of all the particles. The algorithm is recursive with prediction and update phases. The prediction of variables is based on the system model, and the re-evaluation with the sensory information which evolves over time is validated by data taken from every two consecutive images each time. At times the particle with small weights will be eliminated by *resampling* process.

To implement the particle filter, the choice of the proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$ is the most critical design issue. The optimal proposal distribution is often given by [165]:

$$q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k) \quad (4.3.9)$$

4.3. Homography Based Micro Image Tracking (HBT)

115

which is the true conditional state density given the previous state history and all observations. However, sampling from this is impractical for arbitrary densities. Consequently the transition prior is the most popular choice of proposal distribution [166, 167]:

$$q(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{z}_k) \doteq p(\mathbf{x}_k | \mathbf{x}_{k-1}) \quad (4.3.10)$$

In the following, this sampling routine will be implemented for both translational and rotational motions in micromanipulation.

4.3.1 Tracking of Translational Motions

Dynamic model

When the translational motion of the manipulation stage is given by a linear motor with prescribed speed such as the scenario described in Section 4.1, a model with velocity close to constant value can be given by:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w} \quad (4.3.11)$$

The observation model is:

$$\mathbf{z}_k = \mathbf{x}_k + \mathbf{v} \quad (4.3.12)$$

where $\mathbf{x}_k = [x_k, y_k]^T$ is the state vector representing the inter image translation, $\mathbf{z}_k = [\Delta s_x^k, \Delta s_y^k]^T$ is the position difference of the corresponding features in image coordinates between two images. $\mathbf{w}$ is the process noise, and $\mathbf{v}$ is the measurement noise.

$\mathbf{z}_k = [\Delta s_x^k, \Delta s_y^k]^T$ can be computed from:

$$\mathbf{z}_k = \begin{pmatrix} \Delta s_x^k \\ \Delta s_y^k \end{pmatrix} = \mathbf{s}_k - \mathbf{s}_{k-1} \quad (4.3.13)$$

where $\mathbf{s}_{k-1}$, $\mathbf{s}_k$ are corresponded feature positions of the $k-1$ th and k th frame in image coordinates respectively.

Let $\tilde{\mathbf{s}}_{k-1}$ and $\tilde{\mathbf{s}}_k$ be the homogeneous form of $\mathbf{s}_{k-1}$, $\mathbf{s}_k$ in two consecutive frames, the following homography relation holds for a planar translational motion:

$$\tilde{\mathbf{s}}_k \simeq \mathbf{H}_k \tilde{\mathbf{s}}_{k-1} \quad (4.3.14)$$

where

$$\mathbf{H}_k = \begin{pmatrix} 1 + \epsilon & \epsilon & \hat{x}_k \\ \epsilon & 1 + \epsilon & \hat{y}_k \\ \epsilon & \epsilon & 1 \end{pmatrix} \quad (4.3.15)$$

$\simeq$ defines the equality up to a scale. ϵ denotes the small perturbations occurred during the homography estimation. $\hat{x}_k, \hat{y}_k$ give the estimate of the inter image translation. Since the homography can be recovered for every two consecutive images (from Chapter 3), the statistical characteristics of the set $\{\hat{x}_1, \hat{y}_1, \dots, \hat{x}_n, \hat{y}_n\}$ computed from a number of consecutive micro images can be used to sample the states.

Prediction

To initialize the process, the set of parameters $\{\hat{x}_1, \hat{y}_1, \dots, \hat{x}_n, \hat{y}_n\}$ computed from numbers of homographies (as in Equation (4.3.15)) of consecutive images of the same

4.3. Homography Based Micro Image Tracking (HBT)

117

translational motion need to be collected first as the prior information. As there are many sources in micromanipulation that could cause noise in the estimation of the inter image translation and it's difficult to partition these noises, the noise is modelled by adding an additive Gaussian process noise.

Then the prediction can be formally presented by Table 4.3:

Table 4.3: Drawing samples for translational motion

<p>From the old sample set $\{\mathbf{x}_{k-1}^i, i = 1, \dots, N\}$ at time step $k - 1$, construct a new sample set $\{\mathbf{x}_k^i, i = 1, \dots, N\}$ at time step k:</p> <p>for $i = 1: N$</p> <ul style="list-style-type: none"> • if $k = 1$ $\mathbf{x}_k^i = \mathcal{N}(\hat{\mathbf{x}}_i, \sigma);$ • else $\mathbf{x}_k^i = \mathcal{N}(\mathbf{x}_{k-1}^i, \sigma).$ <p>end for</p> <p>(N is the number of samples, $\hat{\mathbf{x}}_i = [\hat{x}_i, \hat{y}_i]^T$, where $\hat{x}_i$ and $\hat{y}_i$ are defined in Equation (4.3.15), σ depends on the assumption of the process noise)</p>

The process noise is difficult to be estimated, hence, in practice σ is taken as the standard deviation of $\{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n\}$.

Update

Based on (4.2.8)(4.3.10), the weight for the i th particle can be presented as proportional to the probability of $\mathbf{z}_k$ given $\mathbf{x}_k$:

$$w_k^i \propto p(\mathbf{z}_k | \mathbf{x}_k) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta s_x}} e^{-\frac{(x_k^i - \rho_{\Delta s_x})^2}{2\sigma_{\Delta s_x}^2}} \frac{1}{\sqrt{2\pi}\sigma_{\Delta s_y}} e^{-\frac{(y_k^i - \rho_{\Delta s_y})^2}{2\sigma_{\Delta s_y}^2}} \quad (4.3.16)$$

4.3. Homography Based Micro Image Tracking (HBT)

118

where

$$[\Delta s_x^k, \Delta s_y^k]^T = \mathbf{s}_k - \mathbf{s}_{k-1} \quad (4.3.17)$$

$$\rho_x = \frac{1}{m-1} \sum_{k=2}^m \Delta s_x^k \quad (4.3.18)$$

$$\rho_y = \frac{1}{m-1} \sum_{k=2}^m \Delta s_y^k \quad (4.3.19)$$

$$\sigma_x = \sqrt{\frac{1}{m-1} \sum_{k=2}^m (\Delta s_x^k - \rho_{\Delta s_x})^2} \quad (4.3.20)$$

$$\sigma_y = \sqrt{\frac{1}{m-1} \sum_{k=2}^m (\Delta s_y^k - \rho_{\Delta s_y})^2} \quad (4.3.21)$$

x_k^i and y_k^i present the elements of the i th particle $\mathbf{x}_k^i$, $\mathbf{s}_k$ and $\mathbf{s}_{k-1}$ are all the matched image features in the k th and the $(k-1)$ th frame respectively, m is the number of matches between k th and $(k-1)$ th frame.

Then the particles can be updated and processed by particle filter (Table 4.1) and *resampling* process (Table 4.2). Each iteration corresponds to one evolution from one frame to the next. Finally, the inter image translation between the $(k-1)$ th and the k th frame can be estimated from:

$$\hat{\mathbf{x}}_k = \mathbb{E}(\mathbf{x}_k | \mathbf{z}_k) \approx \sum_{i=1}^N w_k^i \mathbf{x}_k^i \quad (4.3.22)$$

In this modelling of translational motion, the translational parameters recovered from a set of estimated homographies are used as prior information to initialize the process and the distribution of all the matches found in every two frames contributes in modelling the weight assigned to each particle. With more and more information

collected during the iteration, the translational parameters over sequence of image observations can be approximated through this Monte Carlo simulation.

4.3.2 Tracking of Rotational Motions

Dynamic model

In this practice, the rotation with predefined angular velocity generated by a linear motor such as the scenario described in Section 4.1 can be modelled with fixed rotation axis and the angular velocity close to constant values with respect to the absolute frame when the manipulator stage undergoes planar rotational motion.

The system dynamic model is thus modelled by:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{w} \quad (4.3.23)$$

The observation model is given by:

$$\mathbf{z}_k = f(\mathbf{x}_k) + \mathbf{v} \quad (4.3.24)$$

The state vector consists of three elements: $\mathbf{x}_k = [a_1^k, a_2^k, \theta^k]^T$, where $[a_1^k, a_2^k]^T$ is the rotation axis, θ^k is the rotation angle between the k th and the $(k-1)$ th frame with respect to the absolute frame. f is some function of $\mathbf{x}_k$, $\mathbf{w}$ is the process noise, $\mathbf{v}$ is the measurement noise.

To formulate the observation vector, the correspondences found by the matching

4.3. Homography Based Micro Image Tracking (HBT)

120

algorithms in Chapter 3 is utilized:

$$\mathbf{z}_k = \|\mathbf{s}_k - \mathbf{s}_{k-1}\| \quad (4.3.25)$$

where $\mathbf{s}_k, \mathbf{s}_{k-1}$ denote the corresponded image features in the k th and $(k-1)$ th frame respectively.

Similar to the translational cases, the transition prior is taken as the proposal distribution:

$$q(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{z}_k) \doteq p(\mathbf{x}_k | \mathbf{x}_{k-1}) \quad (4.3.26)$$

In rotational motion of micro view images, it is often difficult to distinguish rotation from translation (see Figure 4.4). Estimation of rotational parameters from micro images is difficult and is often sensitive to noise. In general cases, the rotation can be estimated from two motion vectors, however, under microscope the motion vector is usually very small relative to the distance between the motion vector and the rotation axis, such as depicted in the example in Figure 4.4. $\vec{a}, \vec{b}$ are the true motion vectors, the detected ones maybe locate at $\vec{a}', \vec{b}'$, then the estimation noise could be very large. The noises come not only from CCD of the camera but also the image processing operations involved, such as edge operators, corner detector etc, which induce the uncertainty into the estimation of rotation axis (O' from O). And smaller motion vector are often more liable to introduce large uncertainties in the estimation of rotation axis.

Figure 4.4: Illustration of rotation estimation from small motion vectors. $\vec{a}$, $\vec{b}$: true motion vectors, $\vec{a}'$, $\vec{b}'$: detected motion vectors, O : true rotation axis, O' : estimated rotation axis.

Figure 4.5: The comparison between large and small motion vector in estimating rotation axis. The red ellipse shows the 95% confidence interval.

4.3. Homography Based Micro Image Tracking (HBT)

122

Figure 4.5 shows the difficulty of estimating rotation axis when the motion vector is very small. In Figure 4.5(a) and Figure 4.5(b), the same angle is simulated between the two motion vectors, Figure 4.5(a) shows large motion vector, and Figure 4.5(b) presents small motion vector. The noise with standard deviation 1 pixel is simulated at ends of the motion vectors for both Figure 4.5(a) and Figure 4.5(b). It can be observed that the uncertainties increased greatly when the motion vectors become small.

To obtain the distribution of the parameters to be estimated, a set of homographies computed from the same motion need to be used to provide the prior information. Assume $\tilde{\mathbf{s}}_{k-1}$ and $\tilde{\mathbf{s}}_k$ are the homogeneous form of $\mathbf{s}_{k-1}, \mathbf{s}_k$, which are the corresponding image features in two consecutive frames, the following homography relation holds:

$$\tilde{\mathbf{s}}_k \simeq \mathbf{H}_k \tilde{\mathbf{s}}_{k-1} \quad (4.3.27)$$

where

$$\mathbf{H}_k = \begin{pmatrix} \cos \hat{\theta}^k & \sin \hat{\theta}^k & (1 - \cos \hat{\theta}^k) \hat{a}_1^k - \sin \hat{\theta}^k \hat{a}_2^k \\ -\sin \hat{\theta}^k & \cos \hat{\theta}^k & (1 - \cos \hat{\theta}^k) \hat{a}_2^k + \sin \hat{\theta}^k \hat{a}_1^k \\ 0 & 0 & 1 \end{pmatrix} \quad (4.3.28)$$

$\simeq$ denotes the equality up to a scale. The derivation of this relation is presented below, in which anti clockwise direction is defined to be the angle direction.

Figure 4.6 shows the micro view image, where $(x_1, y_1, 1)^T$, and $(x_2, y_2, 1)^T$ are the homogeneous image coordinates before and after the rotation. The image of the rotation axis is denoted by $\mathbf{a} = (a_1, a_2, 1)^T$, which is also in its homogeneous form. θ in the figure is in clockwise direction (the cross product of x axis to y axis), ρ is the radius. Notice that, the y axis of image coordinates is reversed with respect to the traditional convention of coordinate system, this is due to the CCD imaging process.

Figure 4.6: Illustration of rotation parameter reconstruction from homography

From Figure 4.6, the following relations can be derived:

$$x_2 = x_1 - \rho \cos \alpha + \rho \cos(\alpha - \theta) \quad (4.3.29)$$

$$y_2 = y_1 + \rho \sin(\alpha - \theta) - \rho \sin \alpha \quad (4.3.30)$$

and:

$$\rho \cos \alpha = -a_1 + x_1 \quad (4.3.31)$$

$$\rho \sin \alpha = -a_1 + y_1 \quad (4.3.32)$$

substituting Equations (4.3.31) and (4.3.32) into Equations (4.3.29) and (4.3.30), after some manipulation, the relation of the rotation parameters can be shown:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} (1 - \cos \theta)a_1 - a_2 \sin \theta \\ (1 - \cos \theta)a_2 + a_1 \sin \theta \end{pmatrix} \quad (4.3.33)$$

4.3. Homography Based Micro Image Tracking (HBT)

124

Hence the homography is reconstructed as (4.3.28).

From the relation between the homography and the rotation parameters described in Equation (4.3.28), the distribution of the rotation parameters can be estimated from the set of homographies under the same rotational motion according to:

$$\hat{\theta}^k = \arccos(H_{11}^k) \quad (4.3.34)$$

$$\begin{pmatrix} \hat{a}_1^k \\ \hat{a}_2^k \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{-\sin(\hat{\theta}^k)}{2(\cos \hat{\theta}^k - 1)} \\ \frac{-\sin \hat{\theta}^k}{2(1 - \cos \hat{\theta}^k)} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} H_{13}^k \\ H_{23}^k \end{pmatrix} \quad (4.3.35)$$

where $k = 1, \dots, n$, n is the number of homography samples.

Prediction

The prediction can be initialized by computing the set of parameters $\{\hat{\theta}^1, \hat{a}_1^1, \hat{a}_2^1, \dots, \hat{\theta}^n, \hat{a}_1^n, \hat{a}_2^n\}$ from the homographies acquired from consecutive images of the same rotational motion in prior. Similar to the translation case, additive Gaussian noise is applied to model various noise caused by vibration, drifting, feature extraction etc. Then the prediction of particles can be generated according to Table 4.4:

4.3. Homography Based Micro Image Tracking (HBT)

125

Table 4.4: Drawing samples for rotational motion

<p>From the old sample set $\{\mathbf{x}_{k-1}^i, i = 1, \dots, N\}$ at time step $k - 1$, construct a new sample set $\{\mathbf{x}_k^i, i = 1, \dots, N\}$ at time step k:</p> <p>for $i = 1: N$</p> <ul style="list-style-type: none"> • if $k = 1$ $\mathbf{x}_k^i = \mathcal{N}(\hat{\mathbf{x}}_i, \sigma)$; • else $\mathbf{x}_k^i = \mathcal{N}(\mathbf{x}_{k-1}^i, \sigma)$. <p>end for</p> <p>(N is the number of samples, $\hat{\mathbf{x}}^i = [\hat{\theta}^i, \hat{a}_1^i, \hat{a}_2^i]^T$, where $\hat{\theta}^i$, $\hat{a}_1^i$ and $\hat{a}_2^i$ are defined in Equation (4.3.34) and Equation (4.3.35), σ depends on the assumption of the process noise, σ is estimated by the standard deviation of $\{\hat{\mathbf{x}}^1, \dots, \hat{\mathbf{x}}^n\}$ in practice.)</p>

Update

In the update stage, the weight is calculated for each particle according to the probability of $\mathbf{z}_k$ given $\mathbf{x}_k$. Since the rotation axis and angle is not measurable from the image directly, to obtain the probability of $p(\mathbf{z}_k|\mathbf{x}_k)$, the measurable pixel positions in horizontal and vertical direction are used to present the goodness of each particle in an indirect way:

$$w_k^i \propto p(\mathbf{z}_k|\mathbf{x}_k) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta\mathbf{s}_x}} e^{-\frac{(\Delta\mathbf{s}_x^i - \rho_{\Delta\mathbf{s}_x})^2}{2\sigma_{\Delta\mathbf{s}_x}^2}} \frac{1}{\sqrt{2\pi}\sigma_{\Delta\mathbf{s}_y}} e^{-\frac{(\Delta\mathbf{s}_y^i - \rho_{\Delta\mathbf{s}_y})^2}{2\sigma_{\Delta\mathbf{s}_y}^2}} \quad (4.3.36)$$

In Equation (4.3.36), $[\Delta\mathbf{s}_x^i, \Delta\mathbf{s}_y^i]^T$ is the position difference $([x_2, y_2]^T - [x_1, y_1]^T)$ computed from each particle according to the rule presented by Equation (4.3.33), and $[\rho_{\Delta\mathbf{s}_x}, \rho_{\Delta\mathbf{s}_y}]^T$ is the mean of the position difference computed directly from the corresponded features between two images. $\sigma_{\Delta\mathbf{s}_x}$, $\sigma_{\Delta\mathbf{s}_y}$ are the standard deviation of $[\Delta\mathbf{s}_x, \Delta\mathbf{s}_y]^T$ to signify the confidence for each measurement.

With the simulation of the particles' weight from Equation 4.3.36. The particles can be updated through particle filter (Table 4.1) and resampling (Table 4.2), each iteration corresponds to one evolution from one frame to the next. Hence the state vector $\mathbf{x}_k = [a_1^k, a_2^k, \theta^k]^T$ representing the inter image rotation axis and rotation angle can be estimated at each stage by:

$$\hat{\mathbf{x}}_k = \mathbb{E}(\mathbf{x}_k | \mathbf{z}_k) \approx \sum_{i=1}^N w_k^i \mathbf{x}_k^i \quad (4.3.37)$$

In this section, the evolution of rotational motion is modelled in a particle filter manner, where precalculated homographies provide the information to start the iteration of state prediction and the position difference of features are used to indirectly reflect the goodness of the particles for updating. Through this process, the rotation parameters are expected to be estimated at increasing confidence with more data available.

4.4 Implementation and Experiment Results

In this section, the proposed HBT method is evaluated based on the scenario described in Section 4.1. The micro motion workcell setup in the experiments is shown in Figure 4.7. The workcell consists of independent X, Y and θ motion powered by Normag linear motors. The multiple visual sensory array consists of a Pulnix TM-6702 camera mounted on the microscope and a Pulnix TM-6CN CCD camera providing the global view. The vision system includes a digitizer and a frame grabber (Matrox). The system is able to take images from the micro view and global view at the same rate (2 frames/sec in the following experiments).

At the beginning of the experiment, once the manipulation stage is moving, a se-

Figure 4.7: Micro motion workstation with multiple visual sensors

quence of micro images need to be taken and the set of homographies need to be calculated first (using the algorithms presented in Chapter 3) as prior information. Then as the stage keeps moving, tracking will be solved by HBT with the matching between each pair of frames recovered (using the algorithms presented in Chapter 3). Thus the kinematics parameters from the sequence of micro images are able to be estimated. This estimate can further be used to stitch the micro images together, construct an Euclidean map and enlarge the field of view, so that navigation can be performed within this map using the recovered kinematics.

In the following demonstration for translation, the motion parameters of a 40-frame sequence is tracked, where 80 parameters need to be estimated. For each of the 40 frames, 2 parameters are specified: translational motion in x direction and translational motion in y direction. 1000 particles are sampled for estimation.

Figure 4.8 shows the tracked translation parameters of all the frames tracked. With reference to the first frame in the sequence, the particles representing the motion parameters of the consecutive frames are plotted along the motion direction, so that the tendency of motion can be observed clearly. Figure 4.9 shows the reduced uncertainty with more iterations involved, where the initial states derived from the initial set of homographies as prior information representing the same motion to be estimated and the final estimated states refined by particle filter is shown. The confidence of states is increased through the HBT process as can be seen from Figure 4.9. The histogram of the translational motion, which presents the posterior density of the states, is shown to be non-Gaussian in Figure 4.10. Finally, the stitched micro images are shown in Figure 4.11 and the enlarged field of view is re-projected into the perspective view, shown in Figure 4.12. The difference between the mosaic result with and without HBT shown in Figure 4.11 demonstrates the ability of the proposed HBT method in reducing the misalignment due to uncertainty in tracking for translational motion.

4.4. Implementation and Experiment Results

129

Figure 4.8: Micro image positions, dots: particles, ellipse: 95% confidence interval for each iteration

Figure 4.9: HBT estimation of state vectors for translational motion

Figure 4.10: (a) Histogram of inter image translational motion in x (b) Histogram of inter image translational motion in y

Figure 4.11: Mosaic made up of 40 frames micro images with translational motion. (a) With HBT. (b) Without HBT.

Figure 4.12: Enlarged field of view shown in perspective view for translational motion with HBT.

For rotational motion, 120 parameters of the model are estimated in the following example. Between each consecutive pair micro images of the 40 frames, 3 parameters are specified: rotation axis presented by two element vector and the rotation angle. Similar to the translational case, 1000 particles are generated for simulation.

Figure 4.13 and Figure 4.14 present the spread of particles evolved in 39 iterations, darker points show the result after the final iteration. It can be observed that the uncertainty decreases comparing to the initial samples with more iterations performed. Similar results can also be illustrated by Figure 4.15, Figure 4.16 and Figure 4.17, where the initial states estimated from the set of homographies representing the same motion and the final estimated states refined by particle filter are shown. The histogram of the motion parameters are shown in Figure 4.18 and

4.4. Implementation and Experiment Results

132

Figure 4.19. Figure 4.20 presents the stitched micro images. And the difference with and without HBT validate the performance of the proposed method. The enlarged field of view is finally re-projected into the perspective view, shown in Figure 4.21.

Figure 4.13: Particle cloud spread of rotation axis through all the iterations, dots: particles, ellipse: 95% confidence interval for each iteration. Darker points: the particles from the last iteration

Figure 4.14: Particle cloud spread of rotation angle through all the iterations. Darker points: the particles from the last iteration.

Figure 4.15: HBT estimation of state vectors a_1 vs. a_2 for rotational motion Figure 4.16: HBT estimation of state vectors a_1 vs. θ for rotational motion

Figure 4.17: HBT estimation of state vectors a_2 vs. θ for rotational motion Figure 4.18: Histogram of inter image rotational axis estimation (a) a_1 (b) a_2 Figure 4.19: Histogram of inter image rotational motion in θ

(a)

(b)

Figure 4.20: Mosaic made up of 40 micro image frames with rotational motion (a) With HBT (b) Without HBT.

Figure 4.21: Enlarged field of view shown in perspective view for rotational motion with HBT.

Finally, a mosaic consists of 1000 images is presented in Figure 4.22 to validate the proposed method in tracking large scale number of images. Since the number of images is very large, in the fringe area there is slight misalignment due to the accumulation of noise, however, the overall reconstruction recovers much larger field of view than single micro images.

From the reconstruction results shown in Figure 4.12 and Figure 4.21, it can be observed that the enlarged field of view not only has high resolution of micro images, but also have large field of view with respect to the single micro image field of view. This helps to relocate the interested area that is observed under the microscope in a larger view, with the recovered kinematics, every single image that is within the mosaic can be revisited by issuing a reverse command.

Figure 4.22: Mosaic made up of 1000 frames micro images with translational motion with HBT compared to single micro image field of view at the same scale. The covered area is 22 times of the single micro image. (a) Mosaic of 1000 frames (b) One of the single image

4.5 Summary

In this chapter, HBT method for tracking both translational and rotational motion of the manipulator stage is proposed. Since there is significant effect of noise in micro environment that influences the tracking results greatly, a Bayesian based approach with combination of homography decomposition is resorted to reduce the uncertainties as well as improve the state estimation result in an iterative manner.

In HBT, the relation between the inter image homography and the kinematic parameters in translational and rotational motion behaviors is investigated. The particle

filter paradigm in HBT, updates the particles according to the weight computed from more and more data involved in the process. By using HBT, the uncertainty in the state estimation is reduced so that a smoothly fitted mosaic can be built based on the estimation result.

Further, the mosaic can be registered with the global view image to represent the enlarged micro field of view with respect to the global view. The implementation in a MEMS wafer motion workstation demonstrates the validity of the proposed HBT strategy.

In this chapter, translation is treated separately from rotation, while it is common to have translation and rotation occur simultaneously. However, it is very difficult to decouple and resolve the translation and rotation parameters from the homography relation between two micro images. The homography relation involves both translation and rotation is shown in Equation 4.5.38. Assume $\tilde{\mathbf{s}}_{k-1}$ and $\tilde{\mathbf{s}}_k$ are the homogeneous form of $\mathbf{s}_{k-1}, \mathbf{s}_k$, which are the corresponding image features in two consecutive frames, the following homography relation holds:

$$\tilde{\mathbf{s}}_k \simeq \begin{pmatrix} \cos \hat{\theta}^k & \sin \hat{\theta}^k & (1 - \cos \hat{\theta}^k)\hat{a}_1^k - \sin \hat{\theta}^k \hat{a}_2^k + \hat{x}^k \\ -\sin \hat{\theta}^k & \cos \hat{\theta}^k & (1 - \cos \hat{\theta}^k)\hat{a}_2^k + \sin \hat{\theta}^k \hat{a}_1^k + \hat{y}^k \\ 0 & 0 & 1 \end{pmatrix} \tilde{\mathbf{s}}_{k-1} \quad (4.5.38)$$

$\simeq$ denotes the equality up to a scale, $\hat{x}^k$ and $\hat{y}^k$ are the translational parameters, $\hat{a}_1^k$ and $\hat{a}_2^k$ are the rotation axis in x and y direction respectively and $\hat{\theta}^k$ is the rotation angle. From this relation shown in Equation 4.5.38, it can be seen that there are not enough independent variables available to resolve five parameters from homographies in prior to initialize the tracking process. Besides that, when the rotation angle is small, and the rotation axis is far outside of the image, it's very difficult to

have accurate estimation of rotation axis or translation because of the difficulty to differentiate rotation from translation. Instead of computing the parameters from the homographies acquired from consecutive images, some learning algorithms such as supervised learning or neural network might be used to obtain the initial state vectors from the homographies. Then the tracking can be conducted based on particle filter similar to the proposed tracking method. However, considering the effect of uncertainty in micromanipulation, such as what is shown in Figure 4.5, the cost function has to be designed carefully to obtain initial state estimation.

In summary, micro image tracking is achieved with reduced uncertainty by the proposed HBT method which combines particle filter paradigm and homography decomposition method in this chapter. The proposed method also provides a way to enlarge the limited field of view and helps localization and navigation under microscope.

Chapter 5

Depth Adaptive Coarse-Fine Image Based Visual Servoing With Motion Compensation (DAMC) For Micromanipulation

5.1 Introduction

The coarse-fine mechanism is prevalent in many design of micromanipulation systems for the sake of trading off between efficiency and accuracy [7, 8]. As a visual feedback approach, image based visual servoing is popular and has been successfully applied into many micromanipulation tasks [1, 20]. However, the depth information is difficult to be recovered in image based visual servoing. Furthermore, there are many spacial uncertainties caused by object motion in micromanipulation [168], such as thermal drift etc., and if the motion is ignored the manipulation would fail very often. To enhance visual servoing in micromanipulation, it is important

to adapt depth online during servoing, so that the performance of visual servoing can be improved. In compensating the motion during manipulation, including the object motion in the servoing design is the crucial part in order to balance for the object motion during servoing.

In the literature, there are several methods for solving the losing depth in image based visual servoing, such as homography decomposition approaches [90, 108, 109], least square based approaches [2, 111], and full jacobian adaption method [112–114]. However, these approaches have their own limitations to be implemented into coarse-fine image based visual servoing for micromanipulation. For the homography decomposition, disambiguation is difficult; singularity cannot be avoided for full jacobian adaption, while least square approaches cannot guarantee the convergence rate. In visual servoing of a moving target, the existing methods only consider rotational degrees of freedom [120], or convergence to the desired position in the image space, the relative posture is not guaranteed in the workspace [116, 118].

In this chapter, the depth adaptive image based visual servoing with motion compensation (DAMC) is introduced. A novel depth adaptive image based visual servoing approach is designed for coarse phase of image based visual servoing and a motion compensation frame work is developed for the fine phase image based visual servoing.

We propose to use multiple view multiple scale image based visual servoing to achieve both efficiency and accuracy. Since high magnification of microscope often leads to high numerical aperture and thus very small depth of field, we propose to capture the perspective view image with the interested object in focus under the microscope as desired macro view image. The servoing process will consist of two phases. The first phase is to use macro view images for coarse positioning, so that

not only the macro part can be efficiently moved along long distance, but also the focus of microscopic image can be achieved to some extent at the same time. To adjust the macro view image to the desired one with which the interested object is focused under the microscope, the depth has to be adapted. The second phase is conducted by micro view image based visual servo with motion compensation, which is a 3 degrees of freedom planer motion control task.

In the following discussion, we assume the image feature points in macro view image can be obtained, and the correspondences of the image feature points are solved by the approaches proposed in Chapter 3. For coarse positioning, the depth is changing with time of each servoing step, so the depth adaptive algorithm has to be developed to meet this requirement. Since the fine positioning is actually a planner motion control, the depth during this phase is assumed to be fixed, but target drift has to be compensated during this process.

The organization of this chapter is as follows: Section 5.2 introduces the proposed depth adaptive approach for image based visual seroving which is able to improve the convergence rate and have the mechanism to avoid singularity. Simulation results are also presented to validate this method. In Section 5.3, we introduce the new frame work for motion compensation of image based visual servoing and validated the proposed method by simulations. In Section 5.4, DAMC is realized by implementing the depth adaptive approach into coarse phase image based visual servoing and motion compensation approach into fine phase image based visual seroving for micromanipulation. The simulations of DAMC for micromanipulation are presented to evaluate the performance of the proposed system. Finally, this chapter is summarized in Section 5.5.

5.2 Proposed Depth Adaptive Approach

In this section, we propose a depth adaptive approach in the eye-in-hand image based visual servoing paradigm. The problem can be formulated as: given a set of points observed from the desired camera $\mathcal{F}^*$, then from a generic camera $\mathcal{F}$, as shown in Figure 5.1, drive the camera to the desired posture from any initial one posture with image information. We start with the introduction of the notations, followed by the derivation of proposed algorithms and simulation results.

Figure 5.1: Illustration of problem formulation

5.2.1 Notation

Let $\mathbf{I}_n$ denote the $n \times n$ identity matrix, $\mathbf{e}_i$ denote the i th column of $\mathbf{I}_n$, $\mathbf{0}_n$ be the $n \times 1$ null vector and $[\mathbf{a}]_\times$ be the skew symmetric matrix of $\mathbf{a} \in \mathbb{R}^3$, defined as

5.2. Proposed Depth Adaptive Approach

144

$$[\mathbf{a}]_{\times} = \begin{pmatrix} 0 & -\mathbf{a}(3) & \mathbf{a}(2) \\ \mathbf{a}(3) & 0 & -\mathbf{a}(1) \\ -\mathbf{a}(2) & \mathbf{a}(1) & 0 \end{pmatrix}$$

Moreover we define:

- $\mathcal{F}^0, \mathcal{F}, \mathcal{F}^*$: absolute frame, current camera, and desired camera frames.
- $\{\theta, \mathbf{c}\}$: current camera posture, where $\theta \in \mathbb{R}^3$ is the orientation expressed in exponential coordinates and $\mathbf{c}$ is the camera center with respect to $\mathcal{F}^0$.
- $\mathbf{R} \in SO_3$, $\mathbf{t} \in \mathbb{R}^3$ rotational and translational components of transformation changing coordinates from $\mathcal{F}^0$ to $\mathcal{F}$ according to $\mathbf{R} = e^{[-\theta]_{\times}}$, $\mathbf{t} = -e^{[-\theta]_{\times}} \mathbf{c}$.
- $\phi \in SO_3$, $\eta \in \mathbb{R}^3$: rotation and translation of $\mathcal{F}^*$ with respect to $\mathcal{F}$ expressed in $\mathcal{F}$.
- $\mathbf{x}_i \in \mathbb{R}^3$: i th point expressed in $\mathcal{F}^0$.
- $\tilde{\mathbf{s}}_i = [\mathbf{s}_i^T, 1]^T = [u_i, v_i, 1]^T \in \mathbb{R}^3$: image projections of the i th point in $\mathcal{F}$ expressed in normalized homogeneous coordinates according to

$$\tilde{\mathbf{s}}_i = \frac{\mathbf{R}\mathbf{x}_i + \mathbf{t}}{\mathbf{e}_3^T(\mathbf{R}\mathbf{x}_i + \mathbf{t})}$$
 where $\mathbf{s}_i = [u_i, v_i]^T$ is called the image plane coordinates of the i th point.
- $\mathbf{p}_i \in \mathbb{R}^3$: image projections of the i th point in $\mathcal{F}$ expressed in pixel homogeneous coordinates according to $\mathbf{p}_i = \mathbf{M}\tilde{\mathbf{s}}_i$, where

$$\mathbf{M} = \begin{pmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the intrinsic camera calibration matrix.

– $\mathbf{v}, \omega \in \mathbb{R}^3$: translational and rotational camera velocities expressed in $\mathcal{F}$.

5.2.2 Proposed Strategy

Main Idea

For image based visual servoing systems, the velocity of image point $\dot{\mathbf{s}}_i = [\dot{u}_i, \dot{v}_i]^T$ is related to the camera velocity $\mathbf{V} = [\mathbf{v}^T, \omega^T]^T$ by the image Jacobian relation [169]:

$$\dot{\mathbf{s}}_i = \mathbf{J}_i(\mathbf{r})\mathbf{V} \quad (5.2.1)$$

where

$$\mathbf{J}_i = \begin{pmatrix} -\frac{1}{z_i} & 0 & \frac{u_i}{z_i} & u_i v_i & -(1 + u_i^2) & v_i \\ 0 & -\frac{1}{z_i} & \frac{v_i}{z_i} & 1 + v_i^2 & -u_i v_i & -u_i \end{pmatrix} \quad (5.2.2)$$

i denotes the i th point. Since the depth z_i is normally unknown, it is usually set to the desired depth z_i^* [91] or estimated by least square approach [2, 111].

The main idea of our strategy consists of realizing the depth estimation by constructing the system model with both depth and the depth changing rate, and use Kalman filter method to estimate the unknown parameters with measurable image features obtained at each iteration. To form a linear state updating rule, we chose

5.2. Proposed Depth Adaptive Approach

146

a constant acceleration state model:

$$\begin{cases} \hat{\alpha}(k+1) = \hat{\alpha}(k) + \Delta t \dot{\hat{\alpha}}(k) + \nu_1 \\ \dot{\hat{\alpha}}(k+1) = \gamma \dot{\hat{\alpha}}(k) + \nu_2 \end{cases} \quad (5.2.3)$$

where α is the depth of the specific feature, $\hat{\alpha}$ is the estimation of α , Δt is the sampling period, γ is the degree of correlation between successive accelerations and ranging from 0 to 1 (0.8 in the simulation). ν_1, ν_2 are the zero-mean Gaussian white noise values on the chosen model.

As the uncertain parameter $\alpha(k)_i = 1/z(k)_i, k = 1, \dots, n$ appears linearly in the image jacobian relation, the observation model can be obtained in the following way:

$$\mathbf{s}(k)_i - \mathbf{s}(k-1)_i - \mathbf{B}(k-1)_i \omega(k-1) = \alpha(k-1)_i \mathbf{A}(k-1)_i \mathbf{v}(k-1) + \eta \quad (5.2.4)$$

where $\mathbf{A}(k-1)_i$ and $\mathbf{B}(k-1)_i$ are defined in (5.2.5), (5.2.6):

$$\mathbf{A}(k-1)_i = \begin{pmatrix} -1 & 0 & u(k-1)_i \\ 0 & -1 & v(k-1)_i \end{pmatrix} \quad (5.2.5)$$

$$\mathbf{B}(k-1)_i = \begin{pmatrix} u(k-1)_i v(k-1)_i & -(1 + u(k-1)_i^2) & v(k-1)_i \\ 1 + v(k-1)_i^2 & -u(k-1)_i v(k-1)_i & -u(k-1)_i \end{pmatrix} \quad (5.2.6)$$

$\mathbf{v}(k-1)$ and $\omega(k-1)$ are the translational and rotational camera velocity generated from the previous iteration. η is the zero-mean Gaussian white noise value on the chosen model.

Once the parameters are estimated, control law (5.2.7) can be provided to move the

5.2. Proposed Depth Adaptive Approach

147

camera at each iteration:

$$\mathbf{V} = -\lambda \hat{\mathbf{J}}^+(\hat{\alpha})(\mathbf{s}(k) - \mathbf{s}^*) \quad (5.2.7)$$

where $\hat{\mathbf{J}} = [\hat{\mathbf{J}}_1^T, \hat{\mathbf{J}}_2^T, \dots, \hat{\mathbf{J}}_n^T]^T$ is stacked by jacobian matrix for each feature, with substitution of $\hat{\alpha}_i$ into (5.2.2).

It is well known that the sufficient condition to ensure the global asymptotic stability of image based visual servoing system is [91, 170]:

$$\hat{\mathbf{J}}^+ \mathbf{J} > 0 \quad (5.2.8)$$

However, when image Jacobian is reaching or nearing a task singularity, the condition (5.2.8) can no longer be ensured.

It has been discovered and well identified that the jacobian is singular if image feature is composed of three points such that they are collinear, or belong to a cylinder containing the camera optical center [171–173]. It was also demonstrated in [91] that whatever the number of points and their configuration, the image jacobian may become singular during the visual servoing, if image points are chosen as visual features. Furthermore, if multiple point (≥ 4) features are used to construct image jacobian, some points in specific configuration are more likely to reach or near a task singularity, such as shown in [91]. In other words, the bad condition occurs in several *subjacobiens* related to some of the feature points. To reduce the chances where the goodness of the image jacobian is demolished because elements of the subjacobian turn to infinity or very small numbers, we use the reset scheme to identify the points reaching singularity during prediction step and reassign the estimated state associated with them to good ones, so that the bad condition and

5.2. Proposed Depth Adaptive Approach

148

possible singularity is avoided before the control signal is generated.

Algorithm

Given the initial state $\begin{pmatrix} \hat{\alpha}(0) \\ \dot{\hat{\alpha}}(0) \end{pmatrix} \in \mathbb{R}^{2 \times 1}$ and the initial covariance matrix $\mathbf{P}(0) \in \mathbb{R}^{2 \times 2}$, do the following at each iteration:

Predict the state:

$$\begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \hat{\alpha}(k-1) \\ \dot{\hat{\alpha}}(k-1) \end{pmatrix} \quad (5.2.9)$$

Predict the error covariance:

$$\mathbf{P}(k-) = \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix} \mathbf{P}(k-1) \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix}^T + \mathbf{Q}(k-1) \quad (5.2.10)$$

Update Kalman gain:

$$\mathbf{K}(k) = \mathbf{P}(k-)\mathbf{H}^T[\mathbf{H}\mathbf{P}(k-)\mathbf{H}^T + \mathbf{R}(k)]^{-1} \quad (5.2.11)$$

$$\text{where } \mathbf{H} = \begin{pmatrix} \mathbf{A}(k-1)\mathbf{v}(k-1) & \mathbf{0}_2 \end{pmatrix}$$

Update state estimate:

$$\begin{pmatrix} \hat{\alpha}(k) \\ \dot{\hat{\alpha}}(k) \end{pmatrix} = \begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix} + \mathbf{K}(k)[\mathbf{s}(k) - \mathbf{s}(k-1) - \mathbf{B}(k-1)\omega(k-1) - \mathbf{H}(k) \begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix}] \quad (5.2.12)$$

5.2. Proposed Depth Adaptive Approach

149

Update the error covariance matrix:

$$\mathbf{P}(k) = [\mathbf{I}_2 + \mathbf{K}(k)\mathbf{H}(k)]\mathbf{P}(k-) \quad (5.2.13)$$

Update the estimation of image jacobian matrix according to:

$$\hat{\mathbf{J}}(k) = \begin{pmatrix} \hat{\alpha}(k)\mathbf{A}(k) & \mathbf{B}(k) \end{pmatrix} \quad (5.2.14)$$

$$\begin{cases} \text{reset (Table 5.1),} & \text{if singular or bad conditioned points are detected;} \\ V = -\lambda \hat{\mathbf{J}}^+(\hat{\alpha})(\mathbf{s}(k) - \mathbf{s}^*), & \text{else.} \end{cases}$$

where $\hat{\mathbf{J}}^+$ is the pseudo inverse of $\hat{\mathbf{J}}$. *reset* operation in the above algorithm is to reset the state associated with bad conditioned subjacobians which cause singular and bad conditioned jacobian at the prediction step before the control signal is generated.

Suppose the estimation of α at the k th iteration is given by $\Upsilon(k) = [\hat{\alpha}_1, \hat{\alpha}_i, \dots, \hat{\alpha}_n]$, $\hat{\alpha}_i$ denotes the estimate according to the i th point. $\Upsilon(k)$ can be used to estimate the jacobian status after substitution. Particularly, if the estimation is good enough, every element of $\Upsilon(k)$ should be within a range, and close to each other. If several elements of $\Upsilon(k)$ become too big or too small, they may probably cause the Jacobian singularity or bad condition. The states associated with these elements can be identified and reset by the *reset* algorithm (Table 5.1).

Let $\mathbf{x}(k) = [\hat{\alpha}(k)_i, \dot{\hat{\alpha}}(k)_i]^T$ and $\mathbf{P}(k)$ be the state and covariance matrix associated with the i th feature point at the k th iteration, and let $\hat{\alpha}_i \in [\delta_1, \delta_2]$ be the estimate within the range, in which we call $[\delta_1, \delta_2]$ *the range of goodness*. Let p be the num-

5.2. Proposed Depth Adaptive Approach

150

ber of $\hat{\alpha}_i$ that is within the range, let $\hat{\alpha}_o$ be the estimate out of the range, q be the number of $\hat{\alpha}_o$ that is out of the range.

Let $\kappa = \|[\mathbf{J}(\hat{\alpha}_i^1)^T, \mathbf{J}(\hat{\alpha}_i^2)^T, \dots, \mathbf{J}(\hat{\alpha}_i^p)^T]^T\| \|[\mathbf{J}(\hat{\alpha}_i^1)^T, \mathbf{J}(\hat{\alpha}_i^2)^T, \dots, \mathbf{J}(\hat{\alpha}_i^p)^T]^T\|^{-1}$ be the condition number, the following process is conducted:

Table 5.1: The *reset* algorithm

$n_1 = 0$
while $q > 0$, do
$\begin{cases} \hat{\alpha}_o^l = \frac{1}{p} \sum_{j=1}^p \hat{\alpha}_i^j, l = 1, 2, \dots, q, & \text{if } \kappa \leq \theta; \\ \mathbf{x}(k) = \mathbf{x}(k - n_1 - 1), \mathbf{P}(k) = \mathbf{P}(k - n_1 - 1), & \text{if } \kappa > \theta, n_1 < k - 1; \\ \mathbf{x}(k) = \mathbf{x}(1) + \varsigma, \mathbf{P}(k) = \mathbf{P}(1) + \zeta, & \text{if } \kappa > \theta, n_1 \geq k - 1. \end{cases}$
$n_1 = n_1 + 1$

Where θ is the threshold to determine whether the subjacobian matrix is bad conditioned. n_1 is the number of the *reset* process iterations that have been processed at the current step. ς and ζ are the matrix of random entries according to $\varsigma_i \sim \mathcal{N}(0, k_1 \|\mathbf{x}_i\|)$, $\zeta_i \sim \mathcal{N}(0, k_2 \|\mathbf{P}_i\|)$, where ς_i , ζ_i are the elements for each feature point, $\mathbf{x}_i$, $\mathbf{P}_i$ are the element of $\mathbf{x}$ and $\mathbf{P}$ for each feature point respectively, k_1, k_2 are the scales to generate a small perturbation around the initial values of the elements of $\mathbf{x}$ and $\mathbf{P}$.

The above process indicates that when the subjacobian associated with the estimated states within the range is well-conditioned and not singular, the states are assumed to be good estimates, so the mean value is taken to replace the ones out of the range. When the subjacobian is bad-conditioned, and the reset process iteration number is less than the overall visual servo iteration number, the previous state is set to be the current one. In other words, we detect the instance when bad data appears and reset the state associated with it with the previous one. At

5.2. Proposed Depth Adaptive Approach

151

each iteration of the process, the state will be reset one step backward until the subjacobian is well-conditioned or all the estimates enter the range of goodness. In case that the process iteration number is larger than the overall iteration, which means all the previous states have been tried out, we set the current state to be the initial state with a random perturbation. k_1, k_2 denote the scale of the perturbation.

Therefore, the proposed strategy is able to detect the bad data which causes singularity and bad-conditioned jacobian, and by resetting their states, new estimate can be obtained in the neighborhood of the previous state.

Remarks

$\dot{\hat{\alpha}}$ is not observable from the image data, only very abrupt changes of $\dot{\hat{\alpha}}$ can be observed through the function of $\hat{\alpha}$, and since we only need the estimate of α to estimate the image jacobian, $\dot{\hat{\alpha}}$ does not need to be very accurate. As long as $\dot{\hat{\alpha}}$ can detect and compensate the abrupt accelerations, the servoing steps can be proceeded successfully. In this sense, the observation model is sufficient for depth adaption task.

5.2.3 Simulations

In this section, we present a series of simulations to evaluate the proposed strategy.

In the simulation, a configuration of $n = 4$ points is used to estimate the image Jacobian. The screen size is 600×600 pixels, $\mathbf{M} = [1000, 0, 300; 0, 1000, 300; 0, 0, 1]$. $\rho = 0.8$ and the sampling period is $\Delta t = 10ms$. Specifically, the object considered

5.2. Proposed Depth Adaptive Approach

152

are cubes, and the extracted features are the center of the red cross on the cube. The initial camera posture is reported in Table 5.2. We compared the performance of proposed method denoted by KFRS (kalman filter based method with resetting process) and the traditional least square method denoted by GD (gradient decent).

Table 5.2: Initial Camera Posture

Example	$c(\text{cm})$	$\theta(\text{deg})$
1	[48.0000, -87.0000, 188.0000]	[-1.3403, 2.4829, 0.9660]
2	[69.0000, 37.0000, 149.0000]	[12.7267, -19.5759, -1.4706]
3	[139.0000, 155.0000, 115.0000]	[34.5192, -5.5187, -12.2484]

In example 1 and 2, we can see that the traditional least square method gives biased estimation. To reach the desired camera posture, large camera retreat has to be performed (example 1), which is very dangerous since it may lead to jacobian singularity. If the adaption gain for least square method is tuned to damp the retreat, ambiguity occurs (example 2), which prevent the camera reach the actual desired posture, but converged to a local minima. The results for example 1 are shown in Figure 5.2, Figure 5.3, Figure 5.4, Figure 5.5 and Figure 5.6. The results for example 2 are shown in Figure 5.7, Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11. Large camera retreat during servoing is avoided in the proposed method (see Figure 5.5). This is due to the reason that Kalman filter is unbiased estimator and it has minimum variance. It can also be shown that a better depth adaption convergence rate is obtained by KFRS (see Figure 5.11). Example 3 tested the proposed method for near Jacobian singularity case. The results are shown in Figure 5.12, Figure 5.13, Figure 5.14, Figure 5.15, and Figure 5.16. Observe that, with the reset process, the jacobian singularity with very large estimated depth is avoided (see Figure 5.16).

The desired and final reached camera posture for both GD and KFRS of these

5.2. Proposed Depth Adaptive Approach

153

examples are listed in Table 5.3 for comparison.

5.2. Proposed Depth Adaptive Approach

154

Figure 5.3: Image points tracking errors. solid line: GD, dashed line: KFRS (example 1)

5.2. Proposed Depth Adaptive Approach

155

Figure 5.4: Camera posture evolution in the absolute frame. left: GD, right:KFRS (example 1)

5.2. Proposed Depth Adaptive Approach

156

Figure 5.5: Comparison of evolution of camera posture, left: translational posture, right: rotational posture. solid line: GD, dotted line: KFRS. (example 1)

Figure 5.6: Adapted depth (example 1)

5.2. Proposed Depth Adaptive Approach

157

Figure 5.7: Camera view comparison. (example 2)

Figure 5.8: Image points tracking errors. solid line: GD, dashed line: KFRS (example 2)

5.2. Proposed Depth Adaptive Approach

158

Figure 5.9: Camera posture evolution in the absolute frame. left: GD, right:KFRS (example 2)

5.2. Proposed Depth Adaptive Approach

159

Figure 5.10: Comparison of evolution of camera posture, left: translational posture, right: rotational posture. solid line: GD, dotted line: KFRRS. (example 2)

Figure 5.11: Adapted depth (example 2)

5.2. Proposed Depth Adaptive Approach

160

Figure 5.12: Camera view of the image points evolution (example 3)

Figure 5.13: Image points tracking error (example 3)

Figure 5.14: Evolution of camera posture, left: translational posture, right: rotational posture. (example 3)

5.2. Proposed Depth Adaptive Approach

162

Figure 5.15: Camera motion w.r.t the object. (example 3)

Figure 5.16: Depth adaption comparison between with and w/o reset process. (example 3)

5.2. Proposed Depth Adaptive Approach

163

Table 5.3: Desired camera posture ($\theta^* = [x^*, y^*, z^*]^T, \mathbf{c}^* = [\omega_x^*, \omega_y^*, \omega_z^*]^T$) and final reached camera posture ($\theta^f = [x^f, y^f, z^f]^T, \mathbf{c}^f = [\omega_x^f, \omega_y^f, \omega_z^f]^T$)

example 1	$x(cm)$	$y(cm)$	$z(cm)$	$\omega_x(deg)$	$\omega_y(deg)$	$\omega_z(deg)$
desired	111.0000	-26.0000	195.0000	-25.5175	-11.8567	3.4882
final(GD)	111.5647	-25.8356	194.4583	-25.5473	-11.8672	3.5383
final(KFRS)	110.6253	-25.7039	195.6674	-25.3829	-11.7885	3.4451
example 2	$x(cm)$	$y(cm)$	$z(cm)$	$\omega_x(deg)$	$\omega_y(deg)$	$\omega_z(deg)$
desired	79.0000	16.0000	214.0000	-4.6881	0.9793	0.5381
final(GD)	47.0590	-82.0864	174.2824	-32.7478	6.4708	3.1313
final(KFRS)	78.8839	15.5256	213.8809	-4.8148	1.0064	0.5529
example 3	$x(cm)$	$y(cm)$	$z(cm)$	$\omega_x(deg)$	$\omega_y(deg)$	$\omega_z(deg)$
desired	17.0000	124.0000	167.0000	9.9185	12.3996	-2.5144
final(GD)	NA	NA	NA	NA	NA	NA
final(KFRS)	17.0191	124.0045	166.9945	9.9159	12.3966	-2.5146

5.2.4 Discussion

A visual servoing strategy for dealing with depth adaption problem has been proposed. The approach consists of a kalman filter based depth estimation strategy and a state reset process. The extrinsic camera parameters are not required. All image points are tracked to the desired ones. Moreover, the camera posture is converged to the desired. Near singular cases are dealt with the state reset process. The jacobian singularity and large camera retreat is avoided. Simulation results have validated the performance.

5.3 Proposed Motion Compensation Approach

In this section we propose a novel motion compensation approach in the eye-in-hand image based visual servoing paradigm. The problem can be formulated as: given a set of points observed from the generic camera $\mathcal{F}$, the visual system provides the evolution of the image projections as the camera and the object move in real time, shown in Figure 5.17. The point correspondences are assumed to be known. The goal is to drive the camera to the desired posture relative to the moving target so that the image points reach the desired ones for any initial one with pure image information.

5.3.1 A Brief Review of Rigid Body Velocity

Let's consider the relation between a differential motion vector, for instance, $[\mathbf{v}_e^T, \omega_e^T]^T = [v_1, v_2, v_3, \omega_1, \omega_2, \omega_3]^T$ (different from $[\mathbf{v}^T, \omega^T]^T$) and the homogeneous transformation $\mathbf{T}$.

5.3. Proposed Motion Compensation Approach

165

From mechanics we know that:

$$\dot{\mathbf{R}} = [\omega_e]_{\times} \mathbf{R} \quad (5.3.15)$$

where $\mathbf{R} \in SO_3$ is the orthogonal rotational matrix. This can be generalized to:

$$\dot{\mathbf{T}} = \begin{pmatrix} [\omega_e]_{\times} & \mathbf{v}_e \\ \mathbf{0}_3^T & 1 \end{pmatrix} \mathbf{T} = \mathbf{g} \mathbf{T} \quad (5.3.16)$$

where $\mathbf{g} = \begin{pmatrix} [\omega_e]_{\times} & \mathbf{v}_e \\ \mathbf{0}_3^T & 1 \end{pmatrix}$, Equation (5.3.16) can be rewritten into discrete time format:

$$\mathbf{T}(k) = [\mathbf{I} + \mathbf{g} \Delta t] \mathbf{T}(k-1) \quad (5.3.17)$$

Let $\mathbf{q} = [q_1, q_2, q_3, 1]^T$ be the coordinate of a point in $\mathcal{F}^0$, $\mathbf{p} = \mathbf{T}(k-1)\mathbf{q}$ be its coordinate in $\mathcal{F}^{k-1}$, the point velocity with respect to $\mathcal{F}^k$ is :

$$\dot{\mathbf{p}} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{T}(k)\mathbf{q} - \mathbf{T}(k-1)\mathbf{q}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{g} \Delta t \mathbf{T}(k-1)\mathbf{q}}{\Delta t} = \mathbf{g} \mathbf{p} \quad (5.3.18)$$

Let $\nu = [\nu_1, \nu_2, \nu_3, 0]^T$ be a vector in its homogeneous coordinate with respect to $\mathcal{F}^0$, $\varsigma = \mathbf{T}(k-1)\nu = [\varsigma_1, \varsigma_2, \varsigma_3, 0]^T$ be its coordinate in $\mathcal{F}^{k-1}$, thus the vector velocity with respect to $\mathcal{F}^k$ is given by:

$$\begin{aligned} \dot{\varsigma} &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{T}(k)\nu - \mathbf{T}(k-1)\nu}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{g} \Delta t \mathbf{T}(k-1)\nu}{\Delta t} \\ &= \mathbf{g} \varsigma = \begin{pmatrix} [\omega_e]_{\times} [\varsigma_1, \varsigma_2, \varsigma_3]^T \\ 0 \end{pmatrix} \end{aligned} \quad (5.3.19)$$

Figure 5.17: Illustration of problem formulation

5.3.2 Proposed Strategy

Our main idea is to include the projective presentation of the target motion into the image jacobian development, which is more natural to estimate the unknown target feature changes due to target motion from pure image information. By constraining the target feature changes due to target motion to be at a certain ratio in the whole image motion, the control law to move the camera can be generated.

Formulation of Image Jacobian With Integration of Motion Parameters

Let $\mathbf{P}$ be a point on the object, expressed in $\mathcal{F}^k$. Suppose the target motion velocity can be presented by a vector $\xi \in \mathbb{R}^3$ with respect to $\mathcal{F}^k$. According to the rigid body velocity relation with the differential motion vector (Equations (5.3.18), (5.3.19)), the velocity of the point $\mathbf{P}$, expressed relative to $\mathcal{F}^k$ is given by:

$$\dot{\mathbf{P}} = -[\omega]_{\times} \mathbf{P} - \mathbf{v} + \Omega \xi \quad (5.3.20)$$

5.3. Proposed Motion Compensation Approach

167

where

$$\begin{aligned}\boldsymbol{\Omega} &= e^{[\boldsymbol{\omega}(k-1)]_{\times} \Delta t} = \mathbf{I} - \Delta t [\boldsymbol{\omega}(k-1)]_{\times} \\ &\quad - \frac{\Delta t^2}{2!} [\boldsymbol{\omega}(k-1)]_{\times}^2 - \frac{\Delta t^3}{3!} [\boldsymbol{\omega}(k-1)]_{\times}^3 - \dots\end{aligned}\quad (5.3.21)$$

$\mathbf{v}, \boldsymbol{\omega} \in \mathbb{R}^3$ are translational and rotational velocity screw of the camera. $\boldsymbol{\Omega}$ is the rotation between the component of $\mathcal{F}^k$ and the component of $\mathcal{F}^{k-1}$, with respect to $\mathcal{F}^{k-1}$. Here we take '–', because the differential motion vector here refers to the camera motion, rather than the motion of the object itself.

We omit the high order terms to give the following estimate:

$$\boldsymbol{\Omega} = e^{[\boldsymbol{\omega}(k-1)]_{\times} \Delta t} \approx \mathbf{I} - \Delta t [\boldsymbol{\omega}(k-1)]_{\times} \quad (5.3.22)$$

Equation (5.3.20) is rewritten as:

$$\dot{\mathbf{P}} = -[\boldsymbol{\omega}]_{\times} \mathbf{P} - \mathbf{v} + (\mathbf{I} - \Delta t [\boldsymbol{\omega}(k-1)]_{\times}) \boldsymbol{\xi} \quad (5.3.23)$$

The first two terms $-[\boldsymbol{\omega}]_{\times} \mathbf{P} - \mathbf{v}$ corresponds to the image motion due to the camera motion, the last term $(\mathbf{I}_{3 \times 3} - [\boldsymbol{\omega}]_{\times}) \boldsymbol{\xi}$ corresponds to the image motion due to the target motion.

To simplify the notation, let $\mathbf{P} = [x, y, z]^T$, $\boldsymbol{\xi} = [\xi_x, \xi_y, \xi_z]^T$. The perspective projection equations are given by:

$$\mathbf{s} = \begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{z} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.3.24)$$

5.3. Proposed Motion Compensation Approach

168

Substituting the perspective projection equation (5.3.24) into equation (5.3.20), we can write the derivatives of the coordinates of $\mathbf{P}$ in terms of the image feature velocity u, v and the object motion velocity ξ as:

$$\dot{x} = zv\omega_z - z\omega_y - v_x + \xi_y\omega_z - \xi_z\omega_y + \xi_x \quad (5.3.25)$$

$$\dot{y} = z\omega_x - zu\omega_z - v_y - \xi_x\omega_z + \xi_z\omega_x + \xi_y \quad (5.3.26)$$

$$\dot{z} = zu\omega_y - zv\omega_x - v_z + \xi_x\omega_y - \xi_y\omega_x + \xi_z \quad (5.3.27)$$

Using the quotient rule on Equation (5.3.24), the following can be derived:

$$\begin{aligned} \dot{u} &= \frac{z\dot{x} - x\dot{z}}{z^2} \quad (5.3.28) \\ &= \frac{1}{z^2} \{ z(zv\omega_z - z\omega_y - v_x + \xi_y\omega_z - \xi_z\omega_y) - x(zu\omega_y - zv\omega_x - v_z \\ &\quad + \xi_x\omega_y - \xi_y\omega_x + \xi_z) \} \\ &= -\frac{1}{z}v_x - \frac{u}{z}v_z + (uv + \frac{u\xi_y}{z})\omega_x + (-1 - u^2 - \frac{\xi_z}{z} - \frac{u\xi_x}{z})\omega_y \\ &\quad + (v + \frac{\xi_y}{z})\omega_z + \frac{\xi_x}{z} - \frac{u\xi_z}{z} \end{aligned}$$

Similarly

$$\begin{aligned} \dot{v} &= -\frac{1}{z}v_y + \frac{v}{z}v_z + (1 + v^2 + \frac{\xi_z}{z} + \frac{v\xi_y}{z})\omega_x + (-uv - \frac{v\xi_x}{z})\omega_y \\ &\quad + (-u - \frac{\xi_x}{z})\omega_z + \frac{\xi_y}{z} - \frac{v\xi_z}{z} \end{aligned} \quad (5.3.29)$$

Finally, we may rewrite these two equations in matrix form to obtain:

$$\begin{aligned} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} &= \begin{pmatrix} -\frac{1}{z} & 0 & \frac{u}{z} & uv + \frac{u\xi_y}{z} & -1 - u^2 - \frac{\xi_z}{z} - \frac{u\xi_x}{z} & v + \frac{\xi_y}{z} \\ 0 & -\frac{1}{z} & \frac{v}{z} & 1 + v^2 + \frac{\xi_z}{z} + \frac{v\xi_y}{z} & -uv - \frac{v\xi_x}{z} & -u - \frac{\xi_x}{z} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \omega \end{pmatrix} \\ &\quad + \begin{pmatrix} \frac{\xi_x}{z} - \frac{u\xi_z}{z} \\ \frac{\xi_y}{z} - \frac{v\xi_z}{z} \end{pmatrix} \end{aligned} \quad (5.3.30)$$

5.3. Proposed Motion Compensation Approach

169

Equation (5.3.30) can be written as:

$$\dot{\mathbf{s}} = \mathbf{J}(\mathbf{s}, \xi, z)[\mathbf{v}^T, \omega^T]^T + \mathbf{f}(\mathbf{s}, \xi, z) \quad (5.3.31)$$

where $\mathbf{f} = [\frac{\xi_x}{z} - \frac{u\xi_z}{z}, \frac{\xi_y}{z} - \frac{v\xi_z}{z}]^T$, both $\mathbf{J}(\mathbf{s}, \xi, z)$ and $\mathbf{f}(\mathbf{s}, \xi, z)$ are functions of $\mathbf{s}, \xi, z$.

Control Law

We define the error vector between the current and desired feature points as:

$$\mathbf{e}_k = \mathbf{s}_k - \mathbf{s}^* \quad (5.3.32)$$

Let $\mathbf{V} = [\mathbf{v}^T, \omega^T]^T = [v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^T$ be the velocity screw of the camera motion and $\hat{\mathbf{J}} = [\hat{\mathbf{J}}_1^T, \hat{\mathbf{J}}_2^T, \dots, \hat{\mathbf{J}}_n^T]^T$, $\hat{\mathbf{f}} = [\hat{\mathbf{f}}_1^T, \hat{\mathbf{f}}_2^T, \dots, \hat{\mathbf{f}}_n^T]^T$, then a control law that minimizes a norm of the error $\|\mathbf{e}_k\|$ is given by:

$$\mathbf{V} = \hat{\mathbf{J}}^+(-\lambda \mathbf{e}_k - \hat{\mathbf{f}}) \quad (5.3.33)$$

where $\hat{\mathbf{J}}^+$ is the pseudo inverse of $\hat{\mathbf{J}}$.

Parameter Setting and Adaption

To make the servoing process efficient, we constrain the norm ratio of the two image motion vectors, $\dot{\mathbf{e}}_k$ and $\mathbf{f}$ to be constant. Suppose c to be this constant (taking 2 to 2.5 in the simulations), the control gain λ is given by:

$$\lambda = \frac{c\|\mathbf{f}\|}{\|\mathbf{e}_k\|} \quad (5.3.34)$$

5.3. Proposed Motion Compensation Approach

170

To obtain $\hat{\mathbf{J}}$ and $\hat{\mathbf{f}}$, we can estimate the unknown parameters ξ and z based on the change in the affine model by least square method. Given the initial estimate $\hat{\xi}(0)$ and $\hat{z}(0)$ and the initial control input $\hat{\mathbf{V}}(0)$. The cost function $\mathbf{G}(k)$ at each iteration k is minimized, given by:

$$\begin{aligned} \mathbf{G}(k) = & \|\mathbf{s}(k) - \mathbf{s}(k-1) - \mathbf{J}(\mathbf{s}(k-1), \hat{\xi}(k-1), \hat{z}(k-1))\mathbf{V}(k-1)\Delta t \\ & - \mathbf{f}(\mathbf{s}(k-1), \hat{\xi}(k-1), \hat{z}(k-1))\Delta t\| \end{aligned} \quad (5.3.35)$$

where Δt is the sampling period.

Based on the estimate $\hat{\xi}(k-1)$ and $\hat{z}(k-1)$ which are obtained by minimizing $\mathbf{G}_k$, $\hat{\xi}(k)$ and $\hat{z}(k)$, can be approximated:

$$\hat{\xi}(k) = \mathbf{\Omega}\hat{\xi}(k-1) \quad (5.3.36)$$

$$\hat{z}(k) = \hat{z}(k-1) - v_z(k-1)\Delta t \quad (5.3.37)$$

where $\mathbf{\Omega}$ is given by Equation (5.3.22).

5.3.3 Simulations

In this section, we present a series of simulations to evaluate the proposed strategy.

In the simulation, a configuration of $n = 4$ points is used as image feature. The screen size is 600×600 pixels, $\mathbf{M} = [1000, 0, 300; 0, 1000, 300; 0, 0, 1]$. The sampling period is $\Delta t = 50ms$. Specifically, the objects considered are cubes, and the extracted features are the centers of the red crosses on the cube. The initial camera posture $\{\mathbf{c}, \theta\}$ and object translational velocity $\mathbf{v}_o$ with respect to $\mathcal{F}^0$ is reported in Table 5.4.

5.3. Proposed Motion Compensation Approach

171

The obtained results are presented in the following. The 2d image trajectory is shown in Figure 5.18, Figure 5.26, image feature tracking error is shown in Figure 5.19 and Figure 5.27. Object motion estimation is shown in Figure 5.20, Figure 5.28. Depth estimation is shown Figure 5.21, Figure 5.29. The 3d posture evolution is presented by Figure 5.22, Figure 5.23 and Figure 5.30, Figure 5.31. From the results we can observe that, the image feature approaches to the desired one in a trajectory that is almost a straight line which can avoid out of view problems. The camera posture converged to the desired relative posture with respect to the current object. In Figure 5.24, Figure 5.25 and Figure 5.32, Figure 5.33, the error of the posture difference between the camera and the object are shown, we can see that the relative posture difference error between current camera and current moving target is converged to zero, which means the desired relative posture between the camera and the object is achieved. From the results, it is also shown that the estimation noise appeared in the object motion and the depth has low effect on the 3D trajectory.

Table 5.4: Initial Camera Posture

Example	$\mathbf{c}(\text{cm})$	$\theta(\text{rad})$	$\mathbf{v}_o(\text{cm/s})$
1	[110.0000, 57.0000, 225.0000]	[-0.0314, -0.0356, 0.0013]	[42, 32, 4]
2	[47.0000, 131.0000, 168.0000]	[-0.0070, 0.0522, -0.0108]	[3.4, 4.4, 6]

5.3. Proposed Motion Compensation Approach

172

Figure 5.18: 2D trajectory in the image plane. (example 1) 'o' initial features, '+' target features.

Figure 5.19: Image feature tracking error. (example 1)

Figure 5.20: Estimated object moving velocity w.r.t $\mathcal{F}^k$. (example 1)

Figure 5.21: Estimated depth (example 1)

Figure 5.22: Camera trajectory (example 1)

5.3. Proposed Motion Compensation Approach

175

Figure 5.23: Evolution of camera motion w.r.t the moving object.(example 1)

5.3. Proposed Motion Compensation Approach

176

Figure 5.24: Evolution of the camera translational posture difference w.r.t the moving target.(example 1)

Figure 5.25: Evolution of the camera rotational posture difference w.r.t the moving target. (example 1)

5.3. Proposed Motion Compensation Approach

177

Figure 5.26: 2D trajectory in the image plane. (example 2) 'o' initial features, '+' target features

Figure 5.27: Image feature tracking error. (example 2)

5.3. Proposed Motion Compensation Approach

178

Figure 5.28: Estimated object moving velocity w.r.t $\mathcal{F}^k$. (example 2)

Figure 5.29: Estimated depth (example 2)

Figure 5.30: Camera trajectory (example 2)

Figure 5.31: Evolution of camera motion w.r.t the moving object. (example 2)

5.3. Proposed Motion Compensation Approach

181

Figure 5.32: Evolution of the camera translational posture difference w.r.t the moving target. (example 2)

Figure 5.33: Evolution of the camera rotational posture difference w.r.t the moving target. (example 2)

5.3.4 Discussion

we have investigated the use of eye-in-hand image based visual servoing for tracking moving target. The methods presented are able to track the moving target under translational motions, and keep the relative posture between the camera and the moving target so that not only the image feature reaches the desired ones, but also the camera keeps the desired relative posture difference from the moving target. During the servoing, only image data is utilized, no further interpretation of the 3D structure is required. Satisfactory results are obtained from the simulations.

5.4 DAMC For Micromanipulation

In this section we implemented the depth adaptive approach into coarse phase of image based visual servoing for micromanipulation, the motion compensation approach is implemented into the fine phase of image based visual servoing for micromanipulation. The simulation is presented to validate the DAMC method for micromanipulation.

5.4.1 Implement Depth Adaptive Approach into Coarse Phase

In previous discussion, the eye-in-hand problem is considered, which means the robot end-effector mounted with a camera is moved with respect to the camera coordinate system. However, for micromanipulation tasks, the camera is usually fixed, it is necessary to convert the eye-in-hand relation into an eye-to-hand relation, which means the robot end-effector is moved with respect to the instant frame attached to the end-effector itself rather than the camera frame.

5.4. DAMC For Micromanipulation

183

Assume $\mathbf{v}_e = [v_{ex}, v_{ey}, v_{ez}]^T \in \mathbb{R}^3$, $\omega_e = [\omega_{ex}, \omega_{ey}, \omega_{ez}]^T \in \mathbb{R}^3$ represent the translational and rotational velocities with respect to the instant end-effector frame $\mathcal{F}^e$ respectively. To convert the velocity $\mathbf{V} = [\mathbf{v}^T, \omega^T]^T$ with respect to camera frame $\mathcal{F}^k$ into the velocity $\mathbf{V}_e = [\mathbf{v}_e^T, \omega_e^T]^T$ with respect to the end-effector frame $\mathcal{F}^e$, we only need the relative posture between the initial end-effector and the camera.

Suppose $\mathbf{T} = \begin{pmatrix} \mathbf{R}^{ec} & \mathbf{t}^{ec} \\ \mathbf{0}_3^T & 1 \end{pmatrix}$ to be the transformation from coordinates with respect to the camera frame to the coordinate with respect to the initial end-effector frame, the velocity $\mathbf{V}$ with respect to the instant camera frame $\mathcal{F}^k$ and the velocity $\mathbf{V}_e$ with respect to the instant end-effector frame $\mathcal{F}^e$ is related by:

$$\mathbf{V} = \mathbf{W}\mathbf{V}_e \quad (5.4.38)$$

where

$$\mathbf{W} = \begin{pmatrix} \mathbf{R}^{ec} & \mathbf{R}^{ec}[\mathbf{t}_e]_{\times} \\ \mathbf{0} & \mathbf{R}^{ec} \end{pmatrix} \quad (5.4.39)$$

So the image jacobian for eye-to-hand configuration ($\mathbf{J}_e$) is given by:

$$\mathbf{J}_e = \mathbf{J}\mathbf{W} \quad (5.4.40)$$

where $\mathbf{J}$ is the image jacobian in eye-in-hand configuration.

Thus the depth adaptive approach can be directly applied into an eye-to-hand con-

figuration by rewriting Equation (5.2.7) into:

$$\begin{aligned}\mathbf{V}_e &= -\lambda \hat{\mathbf{J}}_e^+ (\mathbf{s}(k) - \mathbf{s}^*) \\ &= -\lambda (\widehat{\mathbf{J}(\hat{\alpha})\mathbf{W}})^+ (\mathbf{s}(k) - \mathbf{s}^*)\end{aligned}\tag{5.4.41}$$

The algorithm of image based visual servoing with depth adaption for coarse phase can be stated in Table 5.5:

Table 5.5: Coarse phase image based visual servoing with depth adaption algorithm

Given the initial state of depth and the derivative of depth $\mathbf{x}(0) = [\hat{\alpha}(0), \dot{\hat{\alpha}}(0)]^T$, and the initial covariance matrix $\mathbf{P}(0)$, at each iteration do:

- predict the state according to

$$\begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \hat{\alpha}(k-1) \\ \dot{\hat{\alpha}}(k-1) \end{pmatrix} \quad (5.2.9)$$

- predict the error covariance according to

$$\mathbf{P}(k-) = \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix} \mathbf{P}(k-1) \begin{pmatrix} 1 & \Delta t \\ 0 & \gamma \end{pmatrix}^T + \mathbf{Q}(k-1) \quad (5.2.10)$$

- update the Kalman gain according to

$$\mathbf{K}(k) = \mathbf{P}(k-) \mathbf{H}^T [\mathbf{H} \mathbf{P}(k-) \mathbf{H}^T + \mathbf{R}(k)]^{-1} \quad (5.2.11)$$

- update the state estimate according to

$$\begin{pmatrix} \hat{\alpha}(k) \\ \dot{\hat{\alpha}}(k) \end{pmatrix} = \begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix} + \mathbf{K}(k) [\mathbf{s}(k) - \mathbf{s}(k-1) - \mathbf{B}(k-1)\omega(k-1) - \mathbf{H}(k) \begin{pmatrix} \hat{\alpha}(k-) \\ \dot{\hat{\alpha}}(k-) \end{pmatrix}] \quad (5.2.12)$$

- update error covariance matrix according to

$$\mathbf{P}(k) = [\mathbf{I}_2 + \mathbf{K}(k) \mathbf{H}(k)] \mathbf{P}(k-) \quad (5.2.13)$$

- update the estimation of eye-in-hand image jacobian $\hat{\mathbf{J}}$ according to

$$\hat{\mathbf{J}}(k) = \begin{pmatrix} \hat{\alpha}(k) \mathbf{A}(k) & \mathbf{B}(k) \end{pmatrix} \quad (5.2.14)$$

- convert the eye-in-hand image jacobian into eye-to-hand image jacobian according to

$$\mathbf{J}_e = \mathbf{J} \mathbf{W} \quad (5.4.40)$$

- compute the control signal according to

$$\begin{aligned} \mathbf{V}_e &= -\lambda \mathbf{J}_e^+ (\mathbf{s}(k) - \mathbf{s}^*) \\ &= -\lambda (\widehat{\mathbf{J}(\hat{\alpha}) \mathbf{W}})^+ (\mathbf{s}(k) - \mathbf{s}^*) \end{aligned} \quad (5.4.41)$$

5.4.2 Implement Motion Compensation Approach into Fine Phase

To implement the motion compensation approach into fine phase, we have to consider the imaging system of microscope. The simplified optical microscope ray model is shown in Figure 5.34.

Figure 5.34: Simplified Ray Diagram for Typical Optical Microscope

The intermediate image is projected at a distance g behind the posterior principal focus of the objective. By similar triangles:

$$h' = \frac{Mg}{f_0} \quad (5.4.42)$$

where f_0 is the posterior objective focal length and M is the object size in task frame. The total linear magnification is given by:

$$h = \frac{m}{M} = \frac{gc}{f_0 f_e} \quad (5.4.43)$$

where g is the optical tube length.

5.4. DAMC For Micromanipulation

187

Based on the formation of micro image, the following interior relation holds:

$$\begin{bmatrix} p_x \\ p_y \end{bmatrix} = \begin{bmatrix} h \cdot a_x & 0 & x_0 \\ 0 & h \cdot a_y & y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (5.4.44)$$

where $\mathbf{p}_i = (p_x, p_y)$ is pixel coordinates of object features in the microscopic image, $\mathbf{x} = (x, y, z)$ is object coordinates with respect to the micro camera frame. (x_0, y_0) is the principle point of microscopic image, and (a_x, a_y) is pixel resolution of the microscopic image.

The image coordinates $\mathbf{s} = [u, v]^T$ is given by:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} p_x \\ p_y \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} h \cdot a_x & 0 \\ 0 & h \cdot a_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.4.45)$$

To map the motion in task space into sensor space, differentiating (5.4.45), we have:

$$\dot{u} = h a_x \dot{x} \quad (5.4.46)$$

$$\dot{v} = h a_y \dot{y}$$

Recall the derivative relation of a point $\mathbf{P} = [x, y, z]^T$ in motion, developed in Equations (5.3.20), (5.3.25), (5.3.26), (5.3.27), we have the following relation for the micro view:

$$\dot{x} = y\omega_z - z\omega_y - v_x + \xi_y\omega_z - \xi_z\omega_y + \xi_x \quad (5.4.47)$$

$$\dot{y} = z\omega_x - x\omega_z - v_y - \xi_x\omega_z + \xi_z\omega_x + \xi_y \quad (5.4.48)$$

5.4. DAMC For Micromanipulation

188

where $\xi = [\xi_x, \xi_y]^T$ is the target motion velocity with respect to the microscope.

Thus the following can be derived:

$$\dot{u} = ha_x(y\omega_z - z\omega_y - v_x + \xi_y\omega_z - \xi_z\omega_y + \xi_x) \quad (5.4.49)$$

$$\dot{v} = ha_y(z\omega_x - x\omega_z - v_y - \xi_x\omega_z + \xi_z\omega_x + \xi_y) \quad (5.4.50)$$

written in the form of image jacobian, we have:

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} -ha_x & 0 & 0 & 0 & -ha_xz - ha_x\xi_z & ha_xy + ha_x\xi_y \\ 0 & -ha_y & 0 & ha_yz + ha_y\xi_z & 0 & -ha_yx - ha_y\xi_x \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \omega \end{pmatrix} + \begin{pmatrix} ha_x\xi_x \\ ha_y\xi_y \end{pmatrix} \quad (5.4.51)$$

Since in general manipulations under microscope, only three degrees of freedoms are considered, namely, two translational DOFs and one rotational DOF on the plane of focus, the above equation can be simplified into:

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} -ha_x & 0 & ha_xy + ha_x\xi_y \\ 0 & -ha_y & -ha_yx - ha_y\xi_x \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ \omega_z \end{pmatrix} + \begin{pmatrix} ha_x\xi_x \\ ha_y\xi_y \end{pmatrix} \quad (5.4.52)$$

rewritten (5.4.52), we have:

$$\dot{\mathbf{s}} = \mathbf{J}(\mathbf{s}, \xi)(v_x, v_y, \omega_z)^T + \mathbf{f}(\mathbf{s}, \xi) \quad (5.4.53)$$

where

$$\mathbf{J} = \begin{pmatrix} -ha_x & 0 & ha_xy + ha_x\xi_y \\ 0 & -ha_y & -ha_yx - ha_y\xi_x \end{pmatrix} \quad (5.4.54)$$

$$\mathbf{f} = \begin{pmatrix} ha_x\xi_x \\ ha_y\xi_y \end{pmatrix} \quad (5.4.55)$$

$\mathbf{f}$ is the function of $\mathbf{s}, \xi$. Hence the micro image based visual servoing with moving target can be solved with the motion compensation method proposed in Section 5.3. Consider the relative posture of the microscope and the robot end-effector, in our design the end-effector is the manipulation stage with three planner degrees of freedoms. In the following, we assume the optical axis of the microscope is aligned with the the z axis of the stage, x, y directions of the stage are aligned with x and y direction of the image through preprocessings, in which case, the control signal computed in eye-in-hand paradigm coincides with the eye-to-hand one.

The fine phase image based visual servoing with motion compensation is stated in Table 5.6:

Table 5.6: Fine phase image based visual servoing with motion compensation

Given the initial estimate $\xi(0)$ and the initial control input $[\hat{v}_x(0), \hat{v}_y(0), \hat{\omega}_z(0)]^T$, do the following at each iteration:

- minimizing the cost function

$$\mathbf{G}(k) = \|\mathbf{s}(k) - \mathbf{s}(k-1) - \mathbf{J}(\mathbf{s}(k-1), \hat{\xi}(k-1), \hat{z}(k-1))\mathbf{V}(k-1)\Delta t - \mathbf{f}(\mathbf{s}(k-1), \hat{\xi}(k-1), \hat{z}(k-1))\Delta t\| \quad (5.3.35)$$

at each iteration to obtain $\hat{\xi}(k-1)$

- approximate $\hat{\xi}(k)$ by

$$\hat{\xi}(k) = \mathbf{\Omega}\hat{\xi}(k-1) \quad (5.3.36),$$

taking λ according to

$$\lambda = \frac{c\|\mathbf{f}\|}{\|\mathbf{e}_k\|} \quad (5.3.34)$$

- compute Jacobian estimate according to

$$\mathbf{J} = \begin{pmatrix} -ha_x & 0 & ha_xy + ha_x\xi_y \\ 0 & -ha_y & -ha_yx - ha_y\xi_x \end{pmatrix} \quad (5.4.54),$$

compute feature motion due to the target motion according to

$$\mathbf{f} = \begin{pmatrix} ha_x\xi_x \\ ha_y\xi_y \end{pmatrix} \quad (5.4.55)$$

- compute the control signal according to

$$\mathbf{V} = \hat{\mathbf{J}}^+(-\lambda\mathbf{e}_k - \hat{\mathbf{f}}) \quad (5.3.33)$$

5.4.3 Simulation and Results

In this section, we validate the proposed DAMC method with simulations. One macro view camera and one microscope are involved to constitute the visual system, the manipulation stage with micro object on is simulated such as shown in Figure 5.35(a). During the macro view image based visual servoing with depth adaption phase (DA), the stage is driven to be aligned with the focused plane of the microscope using the macro view visual features. Due to the resolution of macro view visual system, the micro view features usually won't reach the target micro view features after the coarse phase servoing. Then under the microscope, our proposed micro view image based visual servoing with motion compensation (MC) is applied to move the stage so that the micro view features are reaching the target positions as well as compensate the target motion so that the micro view features are kept at the desired ones.

In the simulation, the pin-hole camera model is used for macro view camera, and parallel projection model is used for micro view camera. The object is presented by the grid pad. The distance between every two micro features is $20\mu m$. The micro object motion is simulated as $3\mu m/iter$ in x direction and $2\mu m/iter$ in y direction. The intrinsic camera calibration matrix for macro view is $\mathbf{T}_{int} = \begin{pmatrix} 1500 & 0 & 384 & 0 \\ 0 & 1500 & 298 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, the microscope is placed $2.5cm$ above the desired plane of the stage, we assume that it is only able to focus on the plane of the target stage. The product of the magnification and the pixel resolution in x direction ha_x and y direction ha_y are set to be the same as 10000, the principle point of the microscopic image is $x_0 = 324$ and $y_0 = 242$ respectively. The simulated environment is shown in Figure 5.35(a), the initial and target stage postures are also presented, the obvious

Figure 5.35: (a) Simulated environment for coarse fine micromanipulation (b) trajectory of the grid pad during coarse servoing

depth difference between them can be noted. Notice that to perform the servoing, only the intrinsic parameters of the macro and micro view cameras, and the relative difference between the macro view camera and the initial stage posture are needed, the exact posture of either the cameras or the current and desired manipulation stage are not required. Using the proposed image based visual servoing with depth adaption (DA), the image features from the initial macro view image and the image features from the desired macro view image are compared. The servoing is conducted based on the algorithm presented in Table 5.5. In the simulation, uniformly distributed random noise with standard deviation 1 pixel is simulated in each image acquired.

Figure 5.36: (a) initial macro view image, red rectangle: macro image features (b) target macro view image, red cross: macro image features

The servoing results are presented in the following. The trajectory of the grid pad during coarse servoing are shown in Figure 5.35(b), where the four vertex posture of the grid pad in each iteration are presented. Figure 5.36 shows the initial and target images for the coarse phase image based visual servoing. The macro view features in the initial image are presented by the red boxes, and the macro view image features in the target image are presented by red crosses. By the proposed method, the depth of the manipulation stage is adapted (Figure 5.37), the macro image features approach to the desired ones (Figure 5.38). The initial true depth of the four features are $z_0 = \begin{pmatrix} 22.7827 & 22.4999 & 22.3938 & 22.6766 \end{pmatrix} (cm)$, the initial guess depth set to start the servoing is $\hat{z}_0 = \begin{pmatrix} 10.0000 & 12.5000 & 13.5000 & 9.8000 \end{pmatrix} (cm)$. The final depth is adapted to $z_f = \begin{pmatrix} 17.2075 & 16.9711 & 16.7444 & 16.9808 \end{pmatrix} (cm)$, while the desired depth is $z_d = \begin{pmatrix} 17.2373 & 17.0006 & 16.7800 & 17.0167 \end{pmatrix} (cm)$. The feature errors for coarse phase is shown in Figure 5.39.

Figure 5.37: The adapted depth during macro view image based visual servoing

Figure 5.38: Macro image feature trajectory. 'o' initial features, '+' target features

Figure 5.39: Macro image feature error

5.4. DAMC For Micromanipulation

196

When the macro image features reach the desired ones, the manipulation stage reaches the focus plane of the microscope, fine tuning can also be performed manually on the microscope to make sure that the micro features can be seen from the micro view camera. Figure 5.40 (a)(b) shows the initial micro image obtained after coarse phase servoing and the desired micro image. It can be seen that they are not aligned. The initial guess of motion is $0.1\mu\text{m}/\text{iter}$ for both directions while the actual motion is $3\mu\text{m}/\text{iter}$ in x direction and $2\mu\text{m}/\text{iter}$ in y direction. In the simulation, uniformly distributed random noise with standard deviation 0.5 pixel is simulated in each micro image obtained. The image based visual servoing with motion compensation (MC) is performed according to Table 5.6. The results are presented as follows. The final micro image and the desired micro image are shown in figure 5.41 (a)(b) for comparison. The average feature error during micro view image based visual servoing is shown in Figure 5.42. It can be seen that the the micro view image features approach the desired micro view image features and are kept at the desired positions when the target motion of $3\mu\text{m}/\text{iter}$ in x direction and $2\mu\text{m}/\text{iter}$ in y direction and image noise are present.

Figure 5.40: (a) initial micro view image (b) target micro view image

From the above results, the depth of the manipulation stage has been adapted

Figure 5.41: (a) final micro view image (b) target micro view image

Figure 5.42: Average micro image feature error

to the desired so that the microscope is directly focused on the micro part, only slightly manual adjustment is needed. The motion compensation method helps to compensate the target motion during micro image based visual servoing, so that the micro view features reach the desired ones regardless of the target motion.

5.5 Summary

In this chapter, we presented the depth adaptive coarse-fine image based visual servoing with motion compensation (DAMC) for micromanipulation. This approach consists of two main parts, depth adaptive image based visual servoing for coarse phase micromanipulation and the fine phase image based visual servoing with motion compensation. They are aiming to solve two problems in general coarse-fine micromanipulation. One is to efficiently switch from coarse phase to fine phase with auto focus of the interested object. The other is to compensate the target motion of the micro object during fine phase servoing. By introducing the Kalman filtering based depth adaptive approach with state resetting process, we realized the depth adaption for coarse phase visual servoing where the depth of the interested manipulation stage is adapted so that it evolves to the plane that is on focus of the microscope without any particular knowledge of the 3D model of the scene. New formulation of image jacobian including the target motion model is developed to compensate for object motion during fine phase image based visual servoing. The proposed methods are first validated with generic eye-in-hand model and then implemented into the coarse-fine image based visual servoing for micromanipulation. Simulation results show satisfactory results of the proposed DAMC approach.

Chapter 6

Conclusions

6.1 Summery of Research

The primary objective of this thesis is to investigate the possible visual enhancement approaches for micromanipulation. The first part of this thesis focuses on the enhancement of visual tracking and recognition of the microscopic images. The second part of this thesis concentrates on the improvement of traditional visual servoing technique in micromanipulation.

Review of recent literature has provided supporting evidence that indicates considerable interest in coarse-fine image based visual servoing framework for both efficiency and accuracy of micromanipulation tasks. However, problems associated with the limited depth of field, field of view and target motion are very difficult to be avoided. In this thesis the visual improvements for these problems have been the main concern. Experiments and simulation studies give general insight into the performance of the proposed methods.

6.2 Summary of Achievements

In this thesis, we have developed specific tools and techniques in enhancing visual sensing and servoing which represent advances upon existing state of art.

In particular, we proposed and developed matching and tracking methods for long sequence of micro images. Although matching and tracking in macro view have been extensively studied, the problem in micro view is hardly mature. Based on the redundancy property of the image features, we designed two novel line-point integration micro image matching approaches, which are able to robustly match micro images. We also demonstrate the ability of the proposed methods in enhancing matching pairs of micro images in traditional approaches to matching 1000 micro images.

Through the use of underlying homography relations between every two matched micro images, which consist of rich information of the object motion, we developed a tracking algorithm to recover the kinematics parameters as well as reconstruct the micro environment. By decomposing the homography, the kinematics parameters are estimated through a Bayesian framework. Hence the proposed tracking strategy achieved recovery of the underlying structure of the micro world without any specific calibration information. The proposed approach allows us to provide the human operator with advanced information of the micro scene to assist micromanipulation and navigation under the microscope.

In the literature, although several methods with coarse-fine image based visual servoing for micromanipulation have been reported, the limitations of image based approaches on the loss of depth during macro phase and the uncertainties in the mi-

cro object motion have not been fully considered. In this thesis, based on the study of the existing image based visual seroving methods in macro view, we developed a novel coarse fine image based visual seroving approach with both the depth adaption in the coarse phase and motion compensation in the fine phase, so that the depth in coarse phase is adapted and the micro object motion is compensated during fine seroving. To adapt depth in coarse phase, the depth and its derivative are taken as the state vector, a kalman filter model is designed to estimate the state vector online. The estimated depth is then provided into the image based visual seroving to generate control signal. In conventional image based visual seroving, the jacobian only presents the relation between the feature velocity and the control signal, the lack of motion modelling in the jacobian makes the motion compensation very difficult. In this thesis, we proposed a new formulation of image Jacobian, which includes the part of the object motion. This formulation helps to estimate image feature motion due to the object motion during the visual seroving thus makes it possible to compensate the effect of object motion as well as drive the image features to the desired ones.

6.3 Recommendation for Future Research

In terms of further methodological developments, we believe that our work raises many avenues for future research. In particular, we highlighted the important role played by the matching and tracking in recovering micro structure, especially in the context of high noise and uncertainty of the micro environment. We believe that further analysis in image matching and tracking for larger and complicated structures will lead to further developments concerning the recognition of the complete

scene. As the limited field of view and low depth of field could bring more difficulties, 3D segmentation techniques can be applied so that different faces of the 3D micro structure can be distinguished, with the help of CAD model, virtual reality, more information of the micro structure is possible to be reconstructed, hence the navigation will not only be restricted in thousands of images, but extended to even larger and more complicated space.

In this thesis, the importance of homography has been shown in recovering the kinematics of translational and rotational object motion respectively. We believe that there is significant potential to investigate coupled motions with both translation and rotation, in which case, the homography has to be decomposed to recover both translational parameters and rotational parameters. Furthermore, the investigation of the tolerance of the proposed homography based tracking approach can be extended to homography based tracking cases in the mixture of translational and rotational motions.

Although a prototype of the proposed DAMC system has been simulated and studied, there is still a long way to implement the proposed strategy into the real micromanipulation tasks. For specific micromanipulation tasks, different hardware specifications, requirements in physics and mechanics have to be considered. Whole system design could be very specific for particular micromanipulation tasks. Further, user graphical interface should be provided, two views, the micro view and macro view, should be presented to the operator for proper operation. However, in the direction of enhancing visual sensing and servoing to facilitate micromanipulation, we believe that we have made important first steps in this direction, in particular through our recognition of the ability of the visual approaches in both reconstructing the micro scene and controlling the manipulation stage.

6.3. Recommendation for Future Research

203

Moreover, perhaps a promising route towards future advances may be to adopt a similar philosophy which underlies in this thesis into the effort of cooperating movements between the manipulation stage and the manipulator through proper path planning strategy. If the vision based approaches can be thought of as an effective sensing and servoing way, then we believe that image processing, machine vision, visual servoing, which have achieved many amazing results in the engineering world, have an enormous potential in micromanipulation which we are only just beginning to realize.

Author's Publications

1. **W T Sun**, R Devanathan, T C Chin, A Shacklock, Multi View and Multi Scale Image Fusion For Micromanipulation, *The 2nd International Conference on Mechatronics and Information Technology*, Jecheon, Korea, Dec 4-6, 2003 (accepted).
2. **W T Sun** and T C Chin, Image-based Visual Servo for Micromanipulation: A Multiple View and Multiple Scale Approach, *Proceedings of 2004 IEEE International Symposium on Micro-Nano Mechatronics and Human Science*, Nagoya, Japan, pp.341-346, Oct 31-Nov3, 2004.
3. **W T Sun** and T C Chin, Vision Based Micromanipulation Using Planar Features: A Multiple View Approach, *Proceedings of 2004 IEEE International Conference on Robotics, Automation and Mechatronics (RAM-2004)*, Singapore, pp.77-82, Dec1-3, 2004.
4. R Devanathan, **W T Sun**, T C Chin, A Shacklock, Multi View and Multi Scale Image Based Visual Servo For Micromanipulation, *Studies in Computational Intelligence (Editors: Srikanta Patnaik, Lakhmi C. Jain, Spyros G. Tzafestas, et al)*, Vol 8, pp.105-134, 2005.

5. A Shacklock, **W T Sun**, Integrating Microscope and Perspective Views, *Proceedings of 2005 IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 18-22, pp.456-461, 2005.
6. A Shacklock, **W T Sun**, Ephemera: Tracking of Motion in Microscope Views, *Proceedings of IEEE International Conference on Robotics and Automation*, Barcelona, Spain, pp.1847-1852, April 18-22, 2005.
7. **W T Sun**, T C Chin, A Shacklock, Dynamic Image Based Visual Servoing: A Hybrid Approach, *4th Asian Conference on Industrial Automation and Robotics*, Bangkok, Thailand, pp.124-129, May11-13, 2005.
8. **W T Sun**, T C Chin, A Shacklock, Line Segments Matching Across Micro Images, *4th Asian Conference on Industrial Automation and Robotics*, Bangkok, Thailand, pp.329-334, May11-13, 2005.
9. **W T Sun**, T C Chin, A Shacklock, Micro Image Matching With Grouped Features, *Proceedings of 2005 IEEE International Symposium on Micro-Nano Mechatronics and Human Science*, Nagoya, Japan, pp.297-302, Nov 7-9, 2005.
10. **W T Sun**, T C Chin, A Shacklock, Line Segments Matching Across Micro Images, *International Journal for Manufacturing Science and Technology*, Vol 8, No. 2, pp.23-28, 2006.
11. **W T Sun**, T C Chin, A Shacklock, Image Fusion and Tracking for Micro-manipulation, *Journal of Micromechatronics*, Vol 3, No. 2, pp.159-185, 2006.

Author's Publications

206

12. **W T Sun**, T C Chin, A Shacklock, A Depth Adaptive Approach for Image Based Visual Servoing, *Journal of Robotics and Autonomous Systems*(revised).

References

- [1] G. Yang, J. A. Gaines, and B. J. Nelson, "A supervisory wafer-level 3d microassembly system for hybrid mems fabrication," *Journal of Intelligent and Robotic Systems*, vol. 37, pp. 43–68, May 2003.
- [2] F. Conticelli and B. Allotta, "Discrete-time robot visual feedback in 3d positioning tasks with depth adaptation," *IEEE/ASME Transactions on Mechatronics*, vol. 6, no. 3, pp. 356 – 363, 2001.
- [3] J.A.Piepmeyer, G.V.McMurray, and H.Lipkin, "Uncalibrated dynamic visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 1, pp. 143 – 147, 2004.
- [4] M. Arulampalam, S. Maskell, N.Gordon, and T.Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [5] I. A.Codourey, "Visual control of a microrobot operating under a microscope," *Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 993–1000, Nov 1996.
- [6] Y. Sun and B. Nelson, "Autonomous injection of biological cells using visual servoing," *Experimental Robotics VII*, pp. 169–178, 2001.

- [7] A.Kawaji, F.Arai, and T.Fukuda, "Calibration for contact type of micro-manipulation," *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems*, vol. 2, pp. 715–720, 1999.
- [8] A. Ferreira, C. Cassier, and S. Hirai, "Automatic microassembly system assisted by vision servoing and virtual reality," *IEEE/ASME Transactions on Mechatronics*, vol. 9, pp. 321 – 333, June 2004.
- [9] V. T. Portman, B.-Z. Sandler, and E. Zahavi, "Rigid 6-dof parallel platform for precision 3-d micromanipulation," *International Journal of Machine Tools and Manufacture*, vol. 41, pp. 1229–1250, July 2001.
- [10] Y.Sun, M.A.Greminger, D.P.Potasek, and B.J.Nelson, "A visually servoed mems manipulator," *Proc. 8th Int. Symp. on Experimental Robotics (ISER02)*, July 2002.
- [11] Z.Wang, B.G.Hu, L.C.Liang, and Q.Ji, "Cell detection and tracking for micromanipulation vision system of cell-operation," *Proc. IEEE Conference on System, Man, and Cybernetics, Nashville, Tennessee*, October 2000.
- [12] H.Yamamoto and J.Sakiyama, "Stereoscopic visual servo system for microinjection," *Proc. 19th IEEE Instrumentation and Measurement Technology Conference, IMTC/2002.*, vol. 2, pp. 1109 –1112, 2002.
- [13] F.Arai, A.Kawaji, P.Luangjarmekom, T.Fukuda, and K.Itoigawa, "Three-dimensional bio-micromanipulation under the microscope," *Proc. IEEE International Conference on Robotics and Automation*, vol. 1, pp. 604 –609, 2001.
- [14] P.Kallio and M.Lind, "A 3 dof piezohydraulic parallel micromanipulator," *Proc. IEEE international Conference on Robotics and Automation*, pp. 1823–1828, 1998.

- [15] M.Goldfarb and J.H.Lipsey, "Design of a minimum surface-effect tendon-based microactuator for micromanipulation," *Proc. IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1460 –1465, 1997.
- [16] T. Morita, K. Shimizu, M. Hasegawa, K. Oka, and T.Higuchi, "A miniaturized levitation system with motion control using a piezoelectric actuator," *Proc. IEEE Transactions on Control Systems Technology*, vol. 10, pp. 666–670, 2002.
- [17] T.Sano and H.Yamamoto, "Study of micromanipulation using stereoscopic microscope," *Proc. 17th IEEE Technology Conference on Instrumentation and Measurement*, vol. 3, pp. 1227 –1231, 2000.
- [18] L. Xudong, Z. Guanghua, and B. Shusheng, "Development of global vision system for biological automatic micro manipulation system," *Proc. IEEE International Conference on Robotics and Automation*, pp. 127–132, 2001.
- [19] K. S. Hata, S. Ichi, and I. Ishimaru, "3-d vision systems and its computer model for micro-operation," *Proc. 32nd International Symposium on Robotics*, pp. 698–703, 2001.
- [20] Y.Mezouar and P. Allen, "Visual servoed micropositioning for protein manipulation tasks," *Proc. IEEE/RSJ International Conference on Intelligent Robots and System*, vol. 2, pp.1766 –1771, 2002.
- [21] H. B. F. J. A.Burkle, F.Schmoeckel, "A versatile vision system for micromanipulation tasks"," in *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 271 –276, 2001.
- [22] R. Eibl and M. Benoit, "Molecular resolution of cell adhesion forces," *IEE Proceedings on Nanobiotechnology*, vol. 151, pp. 128–132, June 2004.

- [23] F. Arai and T. Andou, D.and Fukuda, "Adhesion forces reduction for micro manipulation based on micro physics," *IEEE The Ninth Annual International Workshop on Micro Electro Mechanical Systems*, pp. 354–359, Feb 1996.
- [24] B. Vikramaditya and B. Nelson, "Visually guided microassembly using optical microscopes and active vision techniques," *Proc. IEEE International Conference on Robotics and Automation*, pp. 3172–3177, April 1997.
- [25] H.Yamamoto and T.Sano, "Study of micromanipulation using stereoscopic microscope," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, pp. 182 –187, Apr 2002.
- [26] S.J.Ralis, B.Vikramaditya, and B.J.Nelson, "Micropositioning of a weakly calibrated microassembly system using coarse-to-fine visual servoing strategies," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 23, no. 2, pp. 123–131, 2000.
- [27] S. Lee, K.Kim, D. Kim, J.O.Park, and G.T.Park, "Recognizing and tracking 3d-shaped micro parts using multiple visions for micromanipulation," *Proc. IEEE International Symposium on Micromechatronics and Human Science*, pp. 203–210, 2001.
- [28] F.Arai, T.Sugiyama, P.Luangjarmekorn, A.Kawaji, T.Fukuda, K.Itoigawa, and A.Maeda, "3d viewpoint selection and bilateral control for biomicromanipulation," *Proc. IEEE International Conference on Robotics and Automation*, vol. 1, pp. 947 –952, 2000.
- [29] A.Kawaji, F.Arai, and T.Fukuda, "3d attitude control system for biomicromanipulation," in *Proceedings of 2001 International Symposium on Micromechatronics and Human Science*, pp. 197 –202, 2001.

- [30] J. Conia, B. Edwards, and S. Voelkel, "The micro-robotic laboratory: Optical trapping and scissoring for the biologist," *Journal of Clinical Laboratory Analysis*, vol. 11, no. 1, pp. 28–38, 1997.
- [31] F. A. Kawaji and T. Fukuda, "3d attitude control system for bio-micromanipulation," *Proc. Intl Symposium on Micromechatronics and Human Science (MHS 2001)*, pp. 197–202, 2001.
- [32] B. Gray, D. Jaeggi, N. Mourlas, B. van Driehuisen, K. Williams, N. Maluf, and G. Kovacs, "Novel interconnection technologies for integrated microfluidic systems," *Sensors and Actuators*, vol. 77, no. 1, pp. 57–65, 1999.
- [33] R. Hauffe, U. Siebel, K. Petermann, R. Moosburger, J. Kropp, and F. Arndt, "Methods for passive fiber chip coupling of integrated optical devices," *IEEE Trans. Advanced Packaging*, vol. 24, no. 4, pp. 450–455, 2001.
- [34] H. Last, M. Deeds, D. Garvick, R. Kavetsky, P. Sandborn, E. Magrab, and S. K. Gupta, "Nano-to-millimeter scale integrated systems," *IEEE Trans. Components and Packaging Technology*, vol. 22, no. 2, pp. 338–343, 1999.
- [35] C. Pramanik, T. Islam, H. Saha, J. Bhattacharya, S. Banerjee, and S. Dey., "Design, fabrication, testing and simulation of porous silicon based smart mems pressure sensor," *18th International Conference on VLSI Design held jointly with 4th International Conference on Embedded Systems Design (VL-SID'05)*, pp. 235–240, 2005.
- [36] L. Wein, "Scheduling semiconductor wafer fabrication," *IEEE Transactions on Semiconductor Manufacturing*, vol. 1, pp. 115–130, Aug 1988.
- [37] M. Moslehi, R. Chapman, M. Wong, A. Paranjpe, H. Najm, J. Kuehne, R. Yeakley, and C. J. Davis, "Single-wafer integrated semiconductor device

REFERENCES

212

- processing," *IEEE Transactions on Electron Devices*, vol. 39, pp. 4–32, Jan 1992.
- [38] Y. Hamakawa, "Fabrication of a new type of synthetic semiconductor in space," *IEEE First World Conference on Photovoltaic Energy Conversion*, vol. 1, pp. 1–8, Dec 1994.
- [39] B. Magnussen, S. Fatikow, and U. Rembold, "Actuation in microsystems: Problem field overview and practical example of the piezoelectric robot for handling of microobjects," *Proc. of the INRIA/IEEE Conference on Emerging Technologies and Factory Automation*, vol. 3, pp. 21–27, 1995.
- [40] S. Fatikow, B. Magnussen, and U. Rembold, "A piezoelectric mobile robot for handling of microobjects," *Proc. of the Int. Symposium on Microsystems, Intelligent Materials and Robots*, pp. 189–192, 1995.
- [41] S. Fatikow and U. Rembold, "An automated microrobot-based desktop station for micro assembly and handling of microobjects," *Proc. Of the IEEE Conference on Emerging Technologie and Factory Automation*, pp. 586–592, 1996.
- [42] J. S. A. Mardanov and S. Fatikow, "An automated assembly system for a microassembly station," *Computers in Industry*, vol. 38, no. 2, pp. 93–102, 1999.
- [43] K. Santa, S. Fatikow, and G. Felso, "Control of microassembly-robots by using fuzzy-logic and neural networks," *Computers in Industry*, vol. 39, pp. 219–227, July 1999.
- [44] S. FATIKOW, "Assembly planning and plan decomposition in an automated microrobot-based microassembly desktop station," *Journal of Intelligent Manufacturing*, vol. 9, pp. 73–92, Jan 1998.

-
- [45] U. Rembold, "Autonomous microrobots," *Journal of Intelligent and Robotic Systems*, vol. 19, pp. 375–391, Aug 1997.
 - [46] Z. Wolfgang, R. Bchi, A. Codourey, and R. Siegwart, "Inertial drives for micro- and nanorobots: Two novel mechanisms," *SPIE Conf. on Microrobots and Micromechanical Systems*, pp. 80–88, Oct 1995.
 - [47] Y. Kudo, T. Kajiyama, and W. Mitsuhashi, "Extraction of feature points suitable for matching image correspondence," *SICE 2003 Annual Conference*, vol. 1, pp. 26–29, Aug 2003.
 - [48] X. Zhang, Y. Liu, and T. Huang, "Determining 3d structure and motion of man-made objects from image corners," *2002. Proceedings. Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 26–30, 2002.
 - [49] C.F.Olson, "Maximum-likelihood image matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 853 – 857, June 2002.
 - [50] C. Schmid and A.Zisserman, "Automatic line matching across views," *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 666 – 671, June 1997.
 - [51] E. B. Patrick Gros, Olivier Bournez, "Using local planar geometric invariants to match and model images of line segments matching line segments using local planar invariants," *Computer Vision and Image Understanding*, pp. 1–38, 1998.
 - [52] C. Zhuang, "A new approach for partial matching of contours," *1993 IEEE Region 10 Conference on TENCN '93. Proceedings. Computer, Communication, Control and Power Engineering.*, vol. 1, pp. 426–429, Oct 1993.

- [53] M.-Y. Shih and D.-C. Tseng, "Versatile surface model reconstruction from serial planar contours," *IEEE Proceedings of the 22nd Annual International Conference on Engineering in Medicine and Biology Society*, vol. 4, pp. 2910–2913, 2000.
- [54] S. Furuie and A.V.Dias, "Shape-based approach for pointwise tracking of contours," *Computers in Cardiology*, pp. 437–440, 2001.
- [55] M. Yasuda, K.Yamamoto, and H.Yamada, "Effect of the perturbed correlation method for optical character recognition," *Proceedings of the Second International Conference on Document Analysis and Recognition*, pp. 830–833, Oct 1993.
- [56] J.P.Didon and F.Langevin, "Fast matching of mr images: validation of a projection based cross-correlation method," *IEEE Proceedings of the 16th Annual International Conference Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers.*, vol. 1, pp. 578–579, Nov 1994.
- [57] K. Kaluzynski, C. Xunchang, S.Y.Emelianov, A.R.Skovoroda, and M.O'Donnell, "Strain rate imaging using two-dimensional speckle tracking," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 48, pp. 1111–1123, Jul 2001.
- [58] R.Bansal, L. Staib, C. Zhe, A.Rangarajan, J.Knisely, R.Nath, and J.S.Duncan, "Entropy-based dual-portal-to-3-dct registration incorporating pixel correlation," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 29–49, Jan 2003.
- [59] J. Kim and J. Fessler, "Intensity-based image registration using robust correlation coefficients," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 1430–1444, Nov 2004.

-
- [60] B. Horn and B. Schunck, "Determining optical flow," *AI Memo 572, Massachusetts Institute of Technology*, 1980.
 - [61] J. Weber and J. Malik, "Rigid body segmentation and shape description from dense optical flow under weak perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 139–143, Feb 1997.
 - [62] S. Morita, "Eye movement computation model based on short-term memory optical flow image," *Proceedings of IEEE International Conference on Multi-sensor Fusion and Integration for Intelligent Systems*, pp. 3–8, Aug 2003.
 - [63] X. Zhuang, R. M. Haralick, and Y. Zhao, "From depth and optical flow to rigid body motion," *Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 393–397, Jun 1988.
 - [64] T. Yamaguchi and H. Nagano, "Shape from optical flow under self- and object-motion," *SICE Annual Conference*, vol. 1, pp. 34–37, Aug 2003.
 - [65] R. Yip and W. Ho, "Multi-level based stereo line matching with structural information using dynamic programming," *Proceedings of International Conference on Image Processing*, vol. 1, pp. 341–344, Sept 1996.
 - [66] J. A. Yang, "An algorithm for parallel stereo matching based on dynamic programming," *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 1, pp. 636–641, Oct 1996.
 - [67] E. Milios and E. G. M. Petrakis, "Shape retrieval based on dynamic programming," *IEEE Transactions on Image Processing*, vol. 9, pp. 141–147, Jan 2000.
 - [68] C. Pan and S. Ma, "3d motion estimation of human by genetic algorithm," *Proceedings. 15th International Conference on Pattern Recognition*, vol. 1, no. 1, pp. 159–163, 2000.
-

- [69] D.Tzovaras, N.Ploskas, and M.G.Strintizis, "Rigid 3-d motion estimation using neural networks and initially estimated 2-d motion data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 158–165, Feb 2000.
- [70] J.Weng, T. Huang, and N. Ahuja, "Estimating motion and structure from line matches: performance obtained and beyond," *Proceedings of the 10th International Conference on Pattern Recognition*, vol. 1, pp. 168 – 172, June 1990.
- [71] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Proc. Intl Conf. Computer Vision*, pp. 259–268, 1987.
- [72] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *Proc. Conf. Computer Vision and Pattern Recognition*, p. 142C149, 2000.
- [73] D. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *Intl J. Computer Vision*, vol. 8, no. 2, pp. 113–122, 1992.
- [74] M. Black and A. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," *Proc. European Conf. Computer Vision*, vol. 1, pp. 329–342, 1996.
- [75] D. Jang and H.-I. Choi, "Moving object tracking using active models," *1998 International Conference on Image Processing*, vol. 3, pp. 648–652, Oct 1998.
- [76] N. Peterfreund, "Robust tracking of position and velocity with kalman snakes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 564–569, June 1999.

- [77] S. Hong, "Steady-state analysis of computational load in correlation-based image tracking," *IEE Proceedings Vision, Image and Signal Processing*, vol. 149, pp. 168–172, June 2002.
- [78] B. Li, R. Chellappa, and H. Moon, "Monte carlo simulation techniques for probabilistic tracking," *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 75–82, Nov 2001.
- [79] H. Zhang, W. Huang, Z. Huang, and B. Zhang, "A particle filtering framework with indirect measurements for visual tracking," *Control, Automation, Robotics and Vision Conference*, vol. 1, pp. 723–728, Dec 2004.
- [80] O. Williams, A. Blake, and R. Cipolla, "Sparse bayesian learning for efficient visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1292–1304, 2005.
- [81] M. Bruno, "Bayesian methods for multiaspect target tracking in image sequences," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1848–1861, July 2004.
- [82] L. Baixin and R. Chellappa, "A generic approach to simultaneous tracking and verification in video," *IEEE Transactions on Image Processing*, vol. 11, pp. 530–544, May 2002.
- [83] N. Gordon, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *IEE Proceedings on Radar and Signal Processing*, vol. 140, pp. 107–113, 1993.
- [84] T. Bando, T. Shibata, K. Doya, and S. Ishii, "Switching particle filters for efficient real-time visual tracking," *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, pp. 720–723, Aug 2004.

- [85] C. Chang and R. Ansari, "Kernel particle filter for visual tracking," *IEEE Signal Processing Letters*, vol. 12, pp. 242–245, March 2005.
- [86] J. Wang, D. Zhao, W. Gao, and S. Shan, "Interacting multiple model particle filter to adaptive visual tracking," *Third International Conference on Image and Graphics*, pp. 568–571, Dec 2004.
- [87] S. K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Transactions on Image Processing*, vol. 13, pp. 1491–1506, Nov 2004.
- [88] J. Hill and W. I. Park, "Real time control of a robot with a mobile camera," *Proc. 9th ISIR*, pp. 233–246, Mar 1979.
- [89] A. C. Sanderson, L. E. Weiss, and C. P. Neuman, "Dynamic sensor based features for resolved rate visual feedback control," *IEEE Trans. Robot. Automat.*, vol. 7, pp. 404–417, Oct 1987.
- [90] E. Malis, F. Chaumette, and S. Boudet, "2 1/2d visual servoing," *IEEE Int. Conf. on Robotics and Automation*, April 2000.
- [91] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control* (D. Kriegman, G. S. Hager, and A. Morse, eds.), pp. 66–78, LNCIS Series, No 237, Springer-Verlag, 1998.
- [92] R. Mahony, P. Corke, and F. Chaumette, "Choice of image features for depth-axis control in image based visual servo control," *IEEE/RSJ International Conference on: Intelligent Robots and Systems*, vol. 1, pp. 390 – 395, 2002.

- [93] A.A.Rizzi and D.E.Koditschek, "An active visual estimator for dexterous manipulation," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 697–713, 1996.
- [94] K.Hashimoto, T.Ebine, and H.Kimura, "Visual servoing with hand-eye manipulator-optimal control approach," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 766 – 774, 1996.
- [95] R.Kelly, "Robust asymptotically stable visual servoing of planar robots," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 759 – 766, 1996.
- [96] B. Nelson, N.P.Papanikolopoulos, and P.K.Khosla, "Robotic visual servoing and robotic assembly tasks," *IEEE Robotics and Automation Magazine*, vol. 2, no. 3, pp. 23 – 31, 1996.
- [97] E. Grosso, G.Metta, A.Oddera, and G.Sandini, "Robust visual servoing in 3-d reaching tasks," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 732 – 742, 1996.
- [98] R.Hartley and A.Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [99] J.Oliensis and Y.Genc, "Fast and accurate algorithms for projective multi-image structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 546 – 559, 2001.
- [100] M.I.A.Lourakis, S.V.Tzurbakis, A.A.Argyros, and S.C.Orphanoudakis, "Feature transfer and matching in disparate stereo views through the use of plane homographies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 271 – 276, 2003.

- [101] P.Martinet and E.Cervera, "Stacking jacobians properly in stereo visual servoing," *IEEE International Conference on Robotics and Automation*, vol. 1, pp. 717 – 722, 2001.
- [102] G.Chesi and K.Hashimoto, "A simple technique for improving camera displacement estimation in eye-in-hand visual servoing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1239 – 1242, 2004.
- [103] G. Chesi and K. Hashimoto, "Improving camera displacement estimation in eye-in-hand visual servoing: a simple strategy," *IEEE International Conference on Robotics and Automation*, vol. 3, pp. 3911 – 3916, 2003.
- [104] O. D. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment.," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, no. 3, pp. 485–508, 1988.
- [105] Z. Zhang and A. Hanson, "3d reconstruction based on homography mapping," *Proc. Image Understanding Workshop*, pp. 1007 – 1012, 1996.
- [106] J.Chen, D. Dawson, W. Dixon, and A. Behal, "Adaptive homography - based visual servo tracking," *Proceedings. 2003 IEEE/RSJ International Conference on: Intelligent Robots and Systems*, vol. 1, pp. 230–235, 2003.
- [107] Y.Fang, D. D. ad W.E. Dixon, and M. de Queiroz, "Homogrpahy-based visual sevoing of wheeled mobile robots," *Proc of the 41st IEEE Conference on Desiion and Control*, vol. 3, pp. 2866–2871, Dec 2002.
- [108] K.Deguchi, "Optimal motion control for image-based visual servoing by decoupling translation and rotation," *1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 705 – 711, 1998.

- [109] E. Malis and F. Chaumette, "Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods," *IEEE Transactions on Robotics and Automation*, vol. 18, pp. 176–186, April 2002.
- [110] E. Vincent and R. Laganier, "Detecting planar homographies in an image pair," *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis*, pp. 182 – 187, 2001.
- [111] H. Fujimoto, "Visual servoing of 6 dof manipulator by multirate control with depth identification," *Proceedings. 42nd IEEE Conference on Decision and Control*, vol. 5, pp. 5408 – 5413, 2003.
- [112] K. Hosoda and M. Asada, "Versatile visual servoing without knowledge of true jacobian," *IEEE RSJ/GI International Conference on Intelligent Robots and Systems*, pp. 186–193, 1994.
- [113] M. Jagersand, O. Fuentes, and R. Nelson, "Experimental evaluation of uncalibrated visual servoing for precision manipulation," *Proceeding International Conference on Robotics and Automation*, pp. 2874–2870, 1997.
- [114] J. A. Piepmeyer, G. V. McMurray, and H. Lipkin, "A dynamic jacobian estimation method for uncalibrated visual servoing," *Proceedings. 1999 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 944 – 949, 1999.
- [115] J. A. Piepmeyer, G. V. McMurray, and H. Lipkin, "A dynamic quasi-newton method for uncalibrated visual servoing," *ICRA*, pp. 1595–1600, 1999.
- [116] J. A. Piepmeyer, G. V. McMurray, and H. Lipkin, "Uncalibrated dynamic visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 1, pp. 143 – 147, 2004.

REFERENCES

222

- [117] D. Zhenyuan and M. Jagersand, "Evaluation of model independent image-based visual servoing," *Proceedings First Canadian Conference on Computer and Robot Vision*, pp. 138 – 144, 2004.
- [118] J. A. Piepmeyer and H. Lipkin, "Uncalibrated eye-in-hand visual servoing," *The International Journal of Robotics Research*, vol. 22, pp. 805–819, October–November 2003.
- [119] K. Hosoda and M. Asada, "Versatile visual servoing without knowledge of true jacobian," *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robotics and Systems*, vol. 1, pp. 186–193, Sept 1994.
- [120] A. Crtual and F. Chaumette, "Application of motion-based visual servoing to target tracking," *The International Journal of Robotics Research*, vol. 20, no. 11, pp. 878–890, 2001.
- [121] G. Wang, H.-T. Tsui, Z. Hu, and F. Wu, "Reconstruction of structured scenes from two uncalibrated images," *Pattern Recognition Letters*, vol. 26, pp. 207–220, 2005.
- [122] Z. Zhang, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," tech. rep., INRIA, 1994.
- [123] A. Bab-Hadiashar and D. Suter, "Robust optic flow computation," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 59 – 77, 1998.
- [124] C. Achard, E. Bigorgne, and J. Devars, "A sub-pixel and multispectral corner detector," *Proceedings. 15th International Conference on Pattern Recognition, 2000.*, vol. 3, pp. 959–962, 2000.
- [125] X. Mu, M. Artiklar, M. Hassoun, and P. Watta, "Training algorithms for robust face recognition using a template-matching approach," *Proceed-*

REFERENCES

223

- ings. IJCNN '01. International Joint Conference on Neural Networks*, vol. 4, pp. 2877–2882, Jul 2001.
- [126] E. Y. Kim, J. S. Chang, and H. J. Kim, “Automatic text location using cluster-based template matching,” *16th International Conference on Pattern Recognition*, vol. 3, p. 30423, 2002.
- [127] K. Fredriksson, V. Makinen, and G. Navarro, “Rotation and lighting invariant template matching,” *Information and Computation*, vol. 205, pp. 1096–1113, Jul 2007.
- [128] R. Brunelli and T. Poggio, “Template matching: Matched spatial filters and beyond,” *Pattern Recognition*, pp. 30(5):751–768, 1997.
- [129] M. I. Lourakis, “Establishing straight line correspondences,” *Technical Report FORTH-ICS / TR-208*, August 1997.
- [130] S.D.Sharghi and F.A.Kamangar, “Geometric feature-based matching in stereo images,” *IEEE Proceedings of Information, Decision and Control*, pp. 65–70, Feb 1999.
- [131] R. Deriche and O.Faugeras, “Tracking line segments,” *Proceedings of Europe Conference on Computer Vision*, pp. 259–268, 1990.
- [132] P. Gros, “Matching and clustering: two steps toward automatic object modeling in computer vision,” *International Journal of Robotics Research*, vol. 14, no. 6, pp. 633–642, 1995.
- [133] Z.N.Li, “Stereo correspondence based on line matching in hough space using dynamic programming,” *IEEE Transactions on System Man and Cybernetics*, vol. 24, no. 1, pp. 144–152, 1994.

REFERENCES

224

- [134] S.H.Lee and J. Leou, "A dynamic programming approach to line segment matching in stereo vision," *Pattern Recognition*, vol. 27, no. 8, pp. 961–986, 1994.
- [135] S. Bouchafa and B.Zavidovique, "Cumulative level-line matching for image registration," *Proceedings of the 12th International Conference Image Analysis and Processing*, pp. 176–181, Sept 2003.
- [136] G. Xu and Z. Zhang, eds., *Epipolar Geometry in Stereo, Motion and Object Recognition - A Unified Approach*. Kluwer Academic Publishers, 1996.
- [137] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Massachusetts: Addison-Wesley, 1992.
- [138] X.Gang and Z.Zhengyou, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Kluwer Academic, 1996.
- [139] S.I.Olsen, "Epipolar line estimation," *Proc. Second European Conf. Comput. Vision*, pp. 307–311, May 1992.
- [140] S. Z. Li, H. Wang, and W. Y. C. Soh, "Robust estimation of rotation angles from image sequences using the annealing m-estimator," *Journal of Mathematical Imaging and Vision*, vol. 8, pp. 181–192, 1998.
- [141] T. OKABE and Y. SATO, "Object recognition based on photometric alignment using ransac," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 221–228, 2003.
- [142] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *International Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

- [143] M. Fishchler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–395, June 1981.
- [144] C. Guerra and V. Pascucci, "Line-based object recognition using hausdorff distance: from range images to molecular secondary structure," *Image and Vision Computing Journal*, vol. 23, no. 4, pp. 405–415, 2005.
- [145] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.
- [146] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.
- [147] Z. W. Kim, A. Huertas, and R. Nevatia, "Automatic description of complex buildings with multiple images," *Fifth IEEE Workshop on Applications of Computer Vision*, pp. 155–162, Dec 2000.
- [148] P. Liu, J. Shi, J. Zhou, and L. Jiang, "Camera self-calibration using the geometric structure in real scenes," *Proceedings of Computer Graphics International*, pp. 262–265, July 2003.
- [149] T. Miyake, T. Murakami, and K. Ohnishi, "An approach to a visual tracking algorithm for a fast motion of an object," *Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society*, vol. 3, pp. 1326 – 1331, 1998.
- [150] W. Kim and J.-J. Lee, "Object tracking based on the modular active shape model," *Mechatronics*, vol. 15, pp. 371–402, April 2005.

- [151] B.-G. Kim and D.-J. Park, "Unsupervised video object segmentation and tracking based on new edge features," *Pattern Recognition Letters*, vol. 25, pp. 1731–1742, November 2004.
- [152] R. Eagle and A. Blake, "Two-dimensional constraints on three-dimensional structure from motion tasks," *Vision Research*, vol. 35, pp. 2927–2941, Oct 1995.
- [153] J. M. Winn and A. Blake, "Generative affine localisation and tracking," *Advances in Neural Information Processing Systems*, vol. 17, pp. 1505–1512, 2004.
- [154] B. Frey and N. J. Fast, "large-scale transformation-invariant clustering," *Advances in Neural Information Processing Systems*, pp. 721–727, 2001.
- [155] L. Baoxin and R. Chellappa, "A generic approach to simultaneous tracking and verification in video," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 530–544, 2002.
- [156] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forsell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.
- [157] J. Vermaak, N. Ikoma, and S. J. Godsill, "Sequential monte carlo framework for extended object tracking," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 353–363, 2005.
- [158] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

-
- [159] A. Doucet and X. Wang, "Monte carlo methods for signal processing: a review in the statistical signal processing context," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, 2005.
 - [160] J. E. Gentle, *Random Number Generation and Monte Carlo Methods*. Springer-Verlag New York, Inc, 2nd ed., 2003.
 - [161] G. Casella and C. P. Robert, *Monte Carlo Statistical Methods*. Springer-Verlag, Inc, 2nd ed., 2004.
 - [162] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian nonlinear state space models," *J. Comput. Graph. Statist.*, vol. 5, pp. 1–25, 1996.
 - [163] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *Int. J. Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.
 - [164] N. Bergman, *Recursive Bayesian estimation: Navigation and tracking applications*. PhD thesis, Kinkoping University, 1999.
 - [165] V. Zaritskii, V. Svetnik, and L. Shimelevich, "Monte-carlo techniques in problems of optimal information processing," *Automation and Remote Control*, vol. 36, no. 3, pp. 2015–2022, 1975.
 - [166] D. Avitzour, "A stochastic simulation bayesian approach to multitarget tracking," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 142, no. 2, pp. 41–44, 1995.
 - [167] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 1–25, 1996.
 - [168] A. A. Requicha, S. Meltzer, F. T. Arce, J. Makaliwe, H. Siken, S. Hsieh, D. Lewis, B. Koel, and M. Thompson, "Manipulation of nanoscale components

REFERENCES

228

- with the afm: Principals and applications,” *Proceedings of IEEE International COnference on Nanotechnology*, pp. 81–86, Oct 2001.
- [169] S.Hutchinson, D.Gregory, Hager, and P.I.Corke, “A tutorial on visual servo control,” *IEEE Transactions on Robotics and Automation*, vol. 12, pp. 651–670, October 1996.
- [170] C. Samson, M. L. Borgne, and B. Espiau, *Robot Control: The Task Function Approach*. Oxford, United Kingdom: Clarendon Press, 1991.
- [171] F.Chaumette, P.Rives, and B.Espiau, “Classification and realization of the different vision-based tasks,” *Visual Servoing*, pp. 199–228, 1993.
- [172] H.Michel and P.Rives, “Singularities in the determination of the situation of a robot effector from the perspective view of three points,” *Technical Report 1850, INRIA*, Feb 1993.
- [173] N.Papanikolopoulos, “Selection of features and evaluation of visual measurements during robotic visual servoing tasks,” *Journal of Intelligent and Robotics Systems*, no. 13, pp. 279–304, 1995.
- [174] W.Press, B.Flannery, S.Teukolsky, and Vetterling, *Numerical Recipes in C*. Cambridge University Press, 1988.

Appendix A

Estimating the homography

A.1 Planar Homography

When the points on a world plane are mapped to points on the image plane, the projection is a specialization of the general central projection, homography. It is described as a plane to plane projection, or a 2D-2D projective mapping, see Figure A.1.

Figure A.1: Plane to plane camera model

A homography is a 3×3 non-singular matrix. The world point $\mathbf{X} = (X, Y, Z)^T$ and

its corresponding image point in its homogeneous form $\mathbf{x} = (x, y, w)^T$ is related by:

$$\mathbf{x} \simeq \mathbf{H}\mathbf{X} \quad (\text{A.1.1})$$

where H is 3×3 homography matrix. $\simeq$ is equality up to scale. Thus it has only eight degrees of freedom, for the scale of the matrix does not affect the equation.

A.2 Inter-image Homography

When a planar surface is viewed from two different viewpoints, an inter-image homography is induced between the two images. Points on the world plane can be transferred from one image to the other by means of homography mapping, see Figure A.2.

Figure A.2: Inter-image homography

In the above figure, π is a plane in the world frame $\mathcal{F}_0$, being viewed from different viewpoints, defined by the frames $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. $\mathbf{H}$ is the inter-image homography between image points from two images. The apparent motion is completely specified once the matrix $\mathbf{H}$ is determined.

To estimate this homography, either point correspondences or line correspondences can be used. For the case of point, given a 3D point $\mathbf{X} = [X, Y, Z]^T$ in the world space, the corresponding image points in homogeneous coordinate from two image frames are denoted by $\mathbf{x} = [x, y, w]^T$ and $\mathbf{x}' = [x', y', w']^T$. If the point $\mathbf{X}$ lies on a plane in space (see Fig. A.3), then $\mathbf{x}$ and $\mathbf{x}'$ are related by a homography $\mathbf{H}$ induced by the plane.

$$\mathbf{x}' \simeq \mathbf{H}\mathbf{x} \quad (\text{A.2.2})$$

where $\mathbf{H}$ is in general a full rank 3×3 matrix defined up to scale. It can be computed from at least four point correspondences on the same plane π in general position (no three of them are collinear).

In the case of line correspondences, $\mathbf{l} \leftrightarrow \mathbf{l}'$, similar relation holds:

$$\mathbf{l} \simeq \mathbf{H}^{-T}\mathbf{l}' \quad (\text{A.2.3})$$

then $\mathbf{H}$ can be linearly computed from four line pairs corresponding to lines on the same plane in space (i.e. no three of them are concurrent). If point $\mathbf{x}$ lie on a line $\mathbf{l}$, then the transformed point $\mathbf{x}'$ lies on the line $\mathbf{l}'$. In this way, incidence of points on lines is preserved (see Fig. A.3).

Figure A.3: The interaction of coplanar line and point pairs

A.3 Homography Estimation

From Equation A.2.2, each point correspondence between two images provides two equations which are linear in the $\mathbf{H}$ matrix elements.

$$H_{31}x' + H_{32}x'y + H_{33}x' = H_{11}x + H_{12}y + H_{13} \quad (\text{A.3.4})$$

$$H_{31}xy' + H_{32}yy' + H_{33}y' = H_{21}x + H_{22}y + H_{23} \quad (\text{A.3.5})$$

Suppose n is the number of correspondences obtained from the image, if $n \geq 4$, the homography can be solved by the following three standard methods.

A.3.1 Pseudo Inverse

Fixing H_{33} to be one, the other 8 elements can be solved using pseudo-inverse with $2n$ equations. This is a non-homogeneous linear solution. It suffers from the inaccuracy of numerical computation of the matrix pseudo inverse and poor estimates when the elements are close to zero.

A.3.2 Singular Value Decomposition

By taking cross product, Equation A.2.2 can be explicitly derived to:

$$\mathbf{x}' \times \mathbf{H}\mathbf{x} = \begin{pmatrix} 0 & -w' & y' \\ w' & 0 & -x' \\ -y' & x' & 0 \end{pmatrix} \begin{pmatrix} \mathbf{H}_{11}x + \mathbf{H}_{12}y + \mathbf{H}_{13}w \\ \mathbf{H}_{21}x + \mathbf{H}_{22}y + \mathbf{H}_{23}w \\ \mathbf{H}_{31}x + \mathbf{H}_{32}y + \mathbf{H}_{33}w \end{pmatrix} = \mathbf{0} \quad (\text{A.3.6})$$

Only two of the three rows are independent. Let $\mathbf{h} = (\mathbf{H}_{11}, \mathbf{H}_{12}, \mathbf{H}_{13}, \mathbf{H}_{21}, \mathbf{H}_{22}, \mathbf{H}_{23}, \mathbf{H}_{31}, \mathbf{H}_{32}, \mathbf{H}_{33})^T$, the homogeneous equation A.3.6 for n point correspondences is given by:

$$\begin{pmatrix} x_1 & y_1 & w_1 & 0 & 0 & 0 & -x_1x'_1 & -y_1x'_1 & -x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & w_1 & -x_1y'_1 & -y_1y'_1 & -y'_1 \\ x_2 & y_2 & w_2 & 0 & 0 & 0 & -x_2x'_2 & -y_2x'_2 & -x'_2 \\ 0 & 0 & 0 & x_2 & y_2 & w_2 & -x_2y'_2 & -y_2y'_2 & -y'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & w_n & 0 & 0 & 0 & -x_nx'_n & -y_nx'_n & -x'_n \\ 0 & 0 & 0 & x_n & y_n & w_n & -x_ny'_n & -y_ny'_n & -y'_n \end{pmatrix} \mathbf{h} = \mathbf{A}\mathbf{h} = \mathbf{0} \quad (\text{A.3.7})$$

The problem of computing $\mathbf{h}$ can be solved by singular value decomposition of matrix $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, and taking the singular vector corresponding to the minimum singular value (the last column of $\mathbf{V}$).

The homogeneous method eliminates the problems caused by non-homogeneous method, but it minimize the algebraic error d_{alg} (see Equation A.3.8) , and the

A.3. Homography Estimation

234

algebraic error does not have a geometric meaning.

$$d_{alg}(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 = (y'_i \hat{w}'_i - w'_i \hat{y}'_i)^2 + (w'_i \hat{x}'_i - x'_i \hat{w}'_i)^2 \quad (\text{A.3.8})$$

where

$$\hat{\mathbf{x}}_i = \mathbf{H}^{-1} \mathbf{x}'_i = [\hat{x}_i, \hat{y}_i, \hat{w}_i]^T \quad (\text{A.3.9})$$

$$\hat{\mathbf{x}}'_i = \mathbf{H} \mathbf{x}_i = [\hat{x}'_i, \hat{y}'_i, \hat{w}'_i]^T \quad (\text{A.3.10})$$

are the estimate of $\mathbf{x}_i = [x_i, y_i, w_i]^T$ and $\mathbf{x}'_i = [x'_i, y'_i, w'_i]^T$ respectively.

A.3.3 Non-linear Geometric Solution

If we use the geometric error as the element of cost function:

$$\begin{aligned} f(\mathbf{H}) &= \sum_{i=1}^n \left\{ \left(\frac{x_i}{w_i} - \frac{\hat{x}_i}{\hat{w}_i} \right)^2 + \left(\frac{y_i}{w_i} - \frac{\hat{y}_i}{\hat{w}_i} \right)^2 + \left(\frac{x'_i}{w'_i} - \frac{\hat{x}'_i}{\hat{w}'_i} \right)^2 + \left(\frac{y'_i}{w'_i} - \frac{\hat{y}'_i}{\hat{w}'_i} \right)^2 \right\} \\ &= \sum_{i=1}^n \{ d_{geo}(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d_{geo}(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \} \end{aligned} \quad (\text{A.3.11})$$

Where n is the number of matches, d_{geo} is the geometric error. Then we can obtain the Maximum Likelihood Estimation of the $\mathbf{H}$ matrix.

Compared with the other two algebraic methods, the quantity minimized in Equa-

tion A.3.11 has its geometric meaning, which is corresponded to the error involved in the measurement. However, there is no closed form of solution to this nonlinear minimization problem, in which Levenberg-Marquardt [174] is most often employed. Usually, the linear method is used to provide a solution that is close to the global minima, for instance, the SVD method, then the nonlinear iterative method is employed to refine the solution to reach the global minima.