Camera Motion Estimation for Multi-Camera Systems

Jae-Hak Kim

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

August 2008
This thesis is submitted to the Department of Information Engineering, Research School of Information Sciences and Engineering, The Australian National University, in fulfilment of the requirements for the degree of Doctor of Philosophy.

This thesis is entirely my own work, except where otherwise stated, describes my own research. It contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning.

Jae-Hak Kim
31 July 2008

Supervisory Panel:

Prof. Richard Hartley
Dr. Hongdong Li
Prof. Marc Pollefeys
Dr. Shyjan Mahamud

The Australian National University
The Australian National University
ETH, Zürich
National ICT Australia

In summary, this thesis is based on materials from the following papers, and my perceived contributions to the relevant chapters of my thesis are stated:


Jae-Hak Kim, Hongdong Li and Richard Hartley, “Motion Estimation for Non-overlapping Multi-Camera Rigs: Linear Algebraic and \( L_\infty \) Geometric Solutions,” submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, (70 per cent of my contribution and related to chapters 9 and 10)
‘Peace be with you.’ John 20:21
I would like to sincerely thank Prof. Richard Hartley for giving me the opportunity to study at the Australian National University and at NICTA and for supervising me during my Ph.D. course. His guidance gave me a deep understanding of multiple view geometry and truly let me know how beautiful geometry is. I would like to thank Dr. Hongdong Li, who gave me a lot of advice on my research work and encouraged me to conceive new ideas. I would also like to thank Prof. Marc Pollefeys, who gave me a chance to visit his vision group at the University of North Carolina, Chapel Hill, as a visiting student and inspired me to find a research topic for my Ph.D. thesis. I also thank Dr. Shyjan Mahamud, who directed my approach to solving problems. I would like to thank National ICT Australia (NICTA) for providing Ph.D. scholarships during the last four years.

I would like to thank Dr. Jan-Michael Frahm for many discussions about this research as well as other members of the vision group at the UNC-Chapel Hill – Dr. Jean-Sebastien Franco, Dr. Philipppos Mordohai, Brian Clipp, David Gallup, Sudipta Sinha, Paul Merrel, Changchang Wu, Li Guan and Seon-Ju Kim. They gave me a warm welcome and provided good research atmosphere during my visit to UNC-Chapel Hill.

I would like to thank Prof. Berthold K. P. Horn and Prof. Steve Maybank who answered my questions and discussed the history and terminologies of epipolar geometry via emails. I would also like to thank Dr. Jun-Sik Kim who pointed out an error in a figure that has been used in this thesis.

I also thank Prof. Joon H. Han, Prof. Daijin Kim, Prof. Seungyong Lee, Prof. Yongduek Seo and Prof. Jong-Seung Park. They have been my strongest support from Korea, and they have shown me how exciting and interesting computer vision is.

I would also like to thank people and organizations for providing me pictures and illustrations that are used in my thesis – AAAS, AIST, Breezesystems Inc., Dr. Carsten Rother, Google, The M.C. Escher Company, NASA/JPL-Caltech, Point Grey Research Inc., Prof. Richard Hartley, Sanghee Seo, Dr. Simon Baker, Timothy Crabtree, UNC-Chapel Hill and Wikipedia.

I am grateful to the RSISE graduate students and academic staff – Andreas M. Maniotis, Dr. Antonio Robles-Kelly, Dr. Arvin Dehghani, Dr. Brad Yu, Dr. Chanop Silpa-Anan, Dr. Chunhua Shen, Desmond Chick, Fangfang Lu, Dr. Guohua Zhang, Dr. Hendra Nurdin, Jae Yong Chung, Dr. Jochen Trumpf, Junae Kim, Dr. Kaiyang Yang, Dr. Kristy Sim, Dr. Lei Wang, Luping Zhou, Manfed Doudar, Dr. Nick Barnes, Dr. Paulette Lieby, Dr. Pei Yean Lee,
Pengdong Xiao, Peter Carr, Ramtin Shams, Dr. Robby Tan, Dr. Roland Goecke, Sung Han Cha, Surya Prakash, Tamir Yedidya, Teddy Rusmin, Vish Swaminathan, Dr. Wynita Griggs, Yifan Lu, Yuhang Zhang and Zhouyu Fu. They welcomed me to RSISE and helped me survive as a Ph.D. student in Australia.

I would also like to thank Hossein Fallahi and Dr. Andy Choi, who were residents with me at Toad Hall.

I would like to thank my close friends from Korea – Jongseung Kim, Jaenam Kim, Hui-jeong Kim, Yechan Ju, Hochan Lim, Kyungho Kim and Jaewon Jang – who supported me.

I also thank my friends in POSTECH – Hyukmin Kwon, Semoon Kil, Byunghwa Lee, Jiwoon Hwang, Dr. Kyoo Kim, Yumi Kim, Dr. Hyeran Kim, Hyosin Kim, Dr. Sunghyun Go, Dr. Changhoon Back, Minseok Song, Dr. Hongjoon Yeo, Dr. Jinmung Won, Dr. Gilje Lee, Sookjin Lee, Hyekyung Lim, Chanjin Jeong and Chunkyu Hwang.

I would like to thank the Korean students in Canberra – Anna Jo, Christina Yoon, Eunhye Park, Eunkyung Park, Haksoo Kim, Inho Shin, Jane Hyo Jin Lee, Kyungmin Lee, Minkyung Moon, Miseon Moon, Sanghoon Lee, Sangwoo Ha, Se-Heon Oh, Sung-Hun Lee, Taehyun Kim, Thomas Han and Wonkeun Chang. They have been so kind to me as the same international student studying in Australia. I will never forget the good times that we had together.

In particular, I would like to thank Fr. Albert Se-jin Kwon, Fr. Michael Young-Hoon Kim, Br. Damaso Young-Keun Chun and Fr. Laurie Foote for providing spiritual guidance.

Last, but not the least, I would like to thank my relatives and family – Clare Kang, Gloria Kim, Natalie Kim, Yunmi Kim, Hyuncheol Kim, Hocheol Kim, Yeonmi Kim, my father and my mother. I would especially like to thank my wife, Eun Young Kim, who supported me with her love and sincere belief. Also, thank my little child to be borned soon.

Thanks to all people I do not remember now, and as always, thanks be to God.
Abstract

The estimation of motion of multi-camera systems is one of the most important tasks in computer vision research. Recently, some issues have been raised about general camera models and multi-camera systems. Using many cameras as a single camera is studied [60], and the epipolar geometry constraints of general camera models is theoretically derived. Methods for calibration, including a self-calibration method for general camera models, are studied [78, 62]. Multi-camera systems are an example of practically implementable general camera models and they are widely used in many applications nowadays because of both the low cost of digital charge-coupled device (CCD) cameras and the high resolution of multiple images from the wide field of views. To our knowledge, no research has been conducted on the relative motion of multi-camera systems with non-overlapping views to obtain a geometrically optimal solution.

In this thesis, we solve the camera motion problem for multi-camera systems by using linear methods and convex optimization techniques, and we make five substantial and original contributions to the field of computer vision. First, we focus on the problem of translational motion of omnidirectional cameras, which are multi-camera systems, and present a constrained minimization method to obtain robust estimation results. Given known rotation, we show that bilinear and trilinear relations can be used to build a system of linear equations, and singular value decomposition (SVD) is used to solve the equations. Second, we present a linear method that estimates the relative motion of generalized cameras, in particular, in the case of non-overlapping views. We also present four types of generalized cameras, which can be solvable using our proposed, modified SVD method. This is the first study finding linear relations for certain types of generalized cameras and performing experiments using our proposed linear method. Third, we present a linear 6-point method (5 points from the same camera and 1 point from another camera) that estimates the relative motion of multi-camera systems, where cam-
eras have no overlapping views. In addition, we discuss the theoretical and geometric analyses of multi-camera systems as well as certain critical configurations where the scale of translation cannot be determined. Fourth, we develop a global solution under an $L_\infty$ norm error for the relative motion problem of multi-camera systems using second-order cone programming. Finally, we present a fast searching method to obtain a global solution under an $L_\infty$ norm error for the relative motion problem of multi-camera systems, with non-overlapping views, using a branch-and-bound algorithm and linear programming (LP). By testing the feasibility of LP at the earlier stage, we reduced the time of computation of solving LP.

We tested our proposed methods by performing experiments with synthetic and real data. The Ladybug2 camera, for example, was used in the experiment on estimation of the translation of omnidirectional cameras and in the estimation of the relative motion of non-overlapping multi-camera systems. These experiments showed that a global solution using $L_\infty$ to estimate the relative motion of multi-camera systems could be achieved.
# Contents

Acknowledgements iv

Abstract vi

1 Introduction 1
   1.1 Problem definition ........................................ 4
   1.2 Contributions ............................................. 5
   1.3 Overview .................................................. 5

2 Single-Camera Systems 7
   2.1 Geometry of cameras ....................................... 7
      2.1.1 Projection of points by a camera .................... 9
      2.1.2 Rigid transformation of points ..................... 10
      2.1.3 Rigid transformation of cameras .................. 11
   2.2 Epipolar geometry of two views .......................... 13
      2.2.1 Definitions of views and cameras .................. 13
      2.2.2 History of epipolar geometry ..................... 14
      2.2.3 Interpretation of epipolar geometry ............... 17
      2.2.4 Mathematical notation of epipolar geometry ...... 18
         2.2.4.1 Pure translation (no rotation) case .......... 18
         2.2.4.2 Pure rotation (no translation) case .......... 21
         2.2.4.3 Euclidean motion (rotation and translation) case . 21
         2.2.4.4 Essential matrix from two camera matrices ... 22
         2.2.4.5 Fundamental matrix ............................ 23
   2.3 Estimation of essential matrix ............................ 26
## Contents

- **2.3.1** 8-point algorithm ........................................ 26
- **2.3.2** Horn’s nonlinear 5-point method ......................... 30
- **2.3.3** Normalized 8-point method .............................. 33
- **2.3.4** 5-point method using a Gröbner basis .................. 34
- **2.3.5** The $L_\infty$ method using a branch-and-bound algorithm ........ 34

### 3 Two- and Three-camera Systems

- **3.1** Two-camera systems (stereo or binocular) .................. 36
- **3.2** Motion estimation using stereo cameras .................... 37
- **3.3** Three-camera systems (trinocular) .......................... 39
- **3.4** Trifocal tensor .............................................. 39
- **3.5** Motion estimation using three cameras ..................... 44

### 4 Multi-camera Systems

- **4.1** What are multi-camera systems? .............................. 45
  - **4.1.1** Advantages of multi-camera systems .................. 46
- **4.2** Geometry of multi-camera systems .......................... 46
  - **4.2.1** Rigid transformation of multi-camera systems ....... 47
- **4.3** Essential matrices in multi-camera systems ............... 48
- **4.4** Non-perspective camera systems ............................ 49

### 5 Previous Related Work

- **5.1** Motion estimation using a large number of images .......... 55
  - **5.1.1** Plane-based projective reconstruction ................. 55
  - **5.1.2** Linear multi-view reconstruction and camera recovery .... 58
- **5.2** Recovering camera motion using $L_\infty$ minimization .......... 60
- **5.3** Estimation of rotation ....................................... 60
  - **5.3.1** Averaging rotations .................................... 60
  - **5.3.2** Lie-algebraic averaging of motions .................... 61
- **5.4** General imaging model ..................................... 61
## Contents

5.5 Convex optimization in multiple view geometry ................................... 62

6 Translation Estimation from Omnidirectional Images .......................... 64
  6.1 Omnidirectional camera geometry ............................................. 65
  6.2 A translation estimation method .............................................. 67
    6.2.1 Bilinear relations in omnidirectional images ..................... 67
    6.2.2 Trilinear relations ....................................................... 70
    6.2.3 Constructing an equation ............................................. 71
    6.2.4 A simple SVD-based least-square minimization ............... 73
  6.3 A constrained minimization .................................................. 73
  6.4 Algorithm ................................................................. 75
  6.5 Experiments ............................................................... 75
    6.5.1 Synthetic experiments ................................................. 75
    6.5.2 Real experiments ...................................................... 80
  6.6 Conclusion ........................................................................ 83

7 Robust 6 DOF Motion Estimation for Non-Overlapping Multi-Camera Rigs ........ 84
  7.1 Related work ............................................................... 85
  7.2 6 DOF multi-camera motion ................................................ 87
  7.3 Two camera system – Theory ................................................. 88
    7.3.1 Geometric interpretation ............................................... 90
    7.3.2 Critical configurations ................................................ 91
  7.4 Algorithm ................................................................. 92
  7.5 Experiments ............................................................... 94
    7.5.1 Synthetic data .......................................................... 94
    7.5.2 Real data .............................................................. 95
  7.6 Conclusion ............................................................... 101

8 A Linear Estimation of Relative Motion for Generalized Cameras ............ 104
  8.1 Previous work ............................................................. 105
# Contents

8.2 Generalized essential matrix for multi-camera systems .......................... 106
  8.2.1 Plücker coordinates ....................................................... 107
  8.2.2 Pless equation ............................................................ 108
  8.2.3 Stewénius’s method ...................................................... 110
8.3 Four types of generalized cameras .................................................... 111
  8.3.1 The most-general case .................................................. 114
  8.3.2 The locally-central case ................................................. 114
  8.3.3 The axial case ............................................................. 116
  8.3.4 The locally-central-and-axial case ................................... 117
8.4 Algorithms .............................................................................. 118
  8.4.1 Linear algorithm for generalized cameras ............................... 118
  8.4.2 Minimizing $||Ax||$ subject to $||Cx|| = 1$ .......................... 120
  8.4.3 Alternate method improving the result of the linear algorithm .... 121
8.5 Experiments .............................................................................. 121
  8.5.1 Synthetic experiments ..................................................... 121
  8.5.2 Real experiments ............................................................ 122
8.6 Conclusion ................................................................................. 130

9 Visual Odometry in Non-Overlapping View Using Second-order cone program-
ming .......................................................................................... 132
  9.1 Problem formulation ............................................................ 132
    9.1.1 Geometric concept ....................................................... 133
    9.1.2 Algebraic derivations .................................................... 135
    9.1.3 Triangulation problem .................................................. 135
  9.2 Second-order cone programming ................................................. 136
  9.3 Summarized mathematical derivation ........................................... 136
  9.4 Algorithm ............................................................................. 137
  9.5 Experiments ........................................................................... 138
    9.5.1 Real data ................................................................. 138
### Contents

9.6 Discussion .......................................................... 142

10 Motion Estimation for Multi-Camera Systems using Global Optimization 143

10.1 The $L_\infty$ method for a single camera ...................... 144

10.2 Branch-and-bound algorithm .................................. 148

10.3 Theory ............................................................. 149

10.4 Algorithm ......................................................... 155

10.5 Experiments ....................................................... 156

10.5.1 Synthetic data experiments ................................. 156

10.5.2 Real data experiments ...................................... 158

10.5.2.1 First real data set ....................................... 158

10.5.2.2 Second real data set ..................................... 162

10.6 Conclusion ....................................................... 167

11 Conclusions and discussions ................................. 168

Appendix ................................................................. 171

Bibliography .............................................................. 174

Index ................................................................. 183
Chapter 1

Introduction

In this thesis, we investigate the relative motion estimation problem of multi-camera systems to develop linear methods and a global solution. Multi-camera systems have many benefits such as rigid motion for all six degrees of freedom without 3D reconstruction of the scene points. Implementations of multi-camera systems can be found in many applications but few studies have been done on the motion of multi-camera systems so far.

In this chapter, we give a general introduction to multi-camera systems and their applications, followed by our contributions and an overview of this thesis.

Recently, the popularity of digital cameras such as digital SLR (single-lens reflex) cameras, compact cameras and mobile phones with built in camera has increased due to their decreased cost. Barry Hendy from Kodak Australia [29] plotted the “pixels per dollar” as a basic measure of the value of a digital camera and used the information to recommend a retail price for Kodak digital cameras. This law is referred to as “Hendy’s Law”. On the basis of this law, it can be concluded that the resolution of a digital camera is becoming higher and the price per pixel of the camera sensor is becoming lower every year. It is no longer difficult or expensive to set up an application that uses several cameras.

It is considered that multicamera systems (a cluster of cameras or a network of cameras) have many benefits in real applications such as visual effects and scientific research. The first study on virtualized reality projects that use virtual views captured by a network of cameras was conducted by Kanade et al. in 1995 [54]. Their system was used to capture touchdowns in the Super Bowl, which is the championship game of professional American football, and it was used to look around the event from other point of virtual views. In 1999, a similar visual
A software controlling 120 cameras using 5 laptops. [www.breezesys.com](http://www.breezesys.com) (Courtesy of Breezesystems, Inc)

Effect known as “bullet time” was implemented in the film “The Matrix”, where the camera appears to orbit around the subject of the scene. This was done by placing a large number of cameras around the subject of the scene. Digital Air is a well-known company that produces Matrix-like visual effects for commercial advertisements [9]. Another company, Breezesys, Inc. [6], sells consumer-level software that allows the simultaneous capture of multiple images by multiple cameras controlled by a single laptop, as shown in Figure 1.1. Thus, the use of multi-camera systems in various applications is becoming popular and their use is expected to increase in the near future.

In the last two decades, many studies have been conducted on the theory and geometry of single-camera systems which are used to capture images from two views, three views and multiple views [11, 10, 27]. However, the theory and geometry of multi-camera systems have not been fully studied or clarified yet. This is because in addition to recording multiple views of a scene using a network of cameras or an array of cameras, there are more challenging tasks such as obtaining spatial and temporal information as the multi-camera system moves around the environment.

This process of obtaining the orientation and position information is known as the “visual odometry” problem or “the problem of estimation of relative motion of multi-camera systems”. A good example of this is as follows: The Mars Exploration Rovers, Spirit and Opportunity,
Figure 1.2: The Mars Exploration Rovers in motion. The rovers are equipped with 9 cameras: four Hazcams are mounted on the front and rear ends for hazard avoidance, two Navcams are mounted on the head of the rovers for navigation, two Pancams are mounted on the head to capture panoramas, and one microscopic camera is mounted on the robotic arm. (Courtesy NASA/JPL-Caltech)

landed on Mars in January 2004. As shown in Figure 1.2, these rovers were equipped with nine cameras distributed between their heads, legs and arms. Although the rovers were equipped with navigation sensors such as IMU (inertial measurement unit) and odometry sensors on their wheels, the estimated distance travelled by the rovers on Mars was not very accurate. This could have been due to several reasons, for example, the rover wheels could not obtain a proper grip on the ground on Mars, which caused the wheels to spin without moving. This resulted in the recording of false measurements by the odometry unit. Another reason could have been the accidental failure of the IMU and odometry equipment. In such a case, visual sensors such as the nine cameras might be used to determine the location of the rovers on Mars. To our knowledge, no research has been conducted on getting an optimal solution to predict the
motion of multi-camera systems. Hence, if we develop an optimal solution, it can be applied to control the motion of planetary rovers, UAVs (unmanned aerial vehicles), AUVs (autonomous underwater vehicles) and domestic robots such as Spirit and Opportunity on Mars, Aerosonde, REMUS and iRobot’s Roomba.

In general, the motions of camera systems can be considered to be Euclidean motions that have six degrees of freedom in three-dimensional (3D) space. So, the main aim of this study is to estimate the motion for all six degrees of freedom. However, in single-camera systems that capture two images, the relative motion can be estimated for only five degrees of freedom: three degrees for rotation and two degrees for translation direction. The scale of translation cannot be estimated from the single-camera system unless 3D structure is recovered. However, in the case of non-overlapping multiple rigs, 3D structure recovery problem is not as easy as in the case of systems with overlapping views such as stereo systems and monocular SLAM (Simultaneous Localization and Mapping) systems.

1.1 Problem definition

In this thesis, we investigate the motion of multi-camera systems. We investigate motion estimation problems such as the translational motion of an omni-directional camera, the motion of a non-overlapping 8-camera system on a vehicle using a linear method and the motion of a 6-camera system (Ladybug2 camera) using second-order cone programming (SOCM) or linear programming (LP) under $L_\infty$ norm.

In general, the motion of multi-camera systems is a rigid motion. Therefore, there are 6 degrees of freedom for rotation and translation. Taking advantage of the spatial information (exterior calibration parameters) of cameras in multi-camera systems, we can estimate the relative motion of multi-camera systems for six degrees of freedom.

Given known camera parameters, we capture image sequences using a multi-camera system. Then, pairs of matching points are detected and found using feature trackers. Using these pairs of matching points, we estimate the relative motion of multi-camera systems for all the six degrees of freedom.
1.2 Contributions

In this thesis,

1. We show that if the rotation of the camera across multiple views is known, it is possible to estimate the translation more accurately using a constrained minimization method based on singular value decomposition (SVD).

2. We also show that the motion of non-overlapping images can be estimated from a minimal set of 6 points of which 5 points are from one camera and 1 point is from another camera. Theoretical analysis of the critical configuration that makes it impossible to solve the relative motion of multi-camera systems is also studied.

3. A linear method to estimate the orientation and position of a multi-camera system (or a general camera model) is studied by considering the rank deficiency of equations and experiments. To our knowledge, no experiments using linear methods have been performed by other researchers in the field of computer vision.

4. Using global optimization and the convex optimization techniques, we solved the problem of estimation of motion using SOCP.

5. We solved the problem of estimation of motion using LP with a branch-and-bound algorithm. Approaches 4 and 5 provide a framework to obtain a global solution for the problem of estimation of relative motion in multi-camera systems (even with non-overlapping views) under the $L_\infty$ norm.

We performed experiments with synthetic and real data to verify our algorithms, and they mostly showed robust and good results.

1.3 Overview

In chapter 1, we provide a general overview of the problems in the estimation of multi-camera systems and demonstrate how multi-camera systems can be used in real applications.
In chapters 2 to 4, we provide brief overviews of the single-camera system, two-camera system, three-camera system, multi-camera system and their motion estimation problems. In chapter 5, we discuss previous related works.

The main work of this thesis is presented in chapters 6, 7, 8, 9 and 10. In chapter 6, we show how constrained minimization allows the robust estimation from omnidirectional images. In chapter 7, we show how using six points, we can estimate the relative motion of non-overlapping views, and we also show that there is a degeneracy configuration that makes it impossible to estimate the motion of non-overlapping multi-camera rigs. In chapter 8, we reveal a linear method for estimation of the motion of a general camera model or non-overlapping multi-camera systems along with an intensive analysis of the rank deficiency in generalized epipolar constraint equations. In chapter 9, we study the geometry of multi-camera systems and demonstrate how using their geometry, we can convert the motion problem to a convex optimization problem using SOCP. In chapter 10, we attempt to improve the method proposed in chapter 9 by developing a unified framework to derive a global solution for the problem of estimation of camera motion in multi-camera systems using LP and a branch-and-bound algorithm. Finally, in chapter 11, conclusions and discussions are presented.