



UNIVERSITÄT  
DES  
SAARLANDES

Universität des Saarlandes  
Max-Planck-Institut für Informatik



MAX-PLANCK-GESELLSCHAFT

# Keyframe-based Visual-Inertial Odometry for Small Workspace

Masterarbeit im Fach Informatik  
Master's Thesis in Computer Science  
von / by  
**Xi Li**

angefertigt unter der Leitung von / supervised by  
DR. ROLAND ANGST

betreut von / advised by  
DR. ROLAND ANGST

begutachtet von / reviewers  
DR. ROLAND ANGST  
PROF. DR. JOACHIM WEICKERT

Saarbrücken, June 2016

**Eidesstattliche Erklärung**

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

**Statement in Lieu of an Oath**

I hereby confirm under oath that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

**Eidesstattliche Erklärung**

Ich erkläre hiermit an Eides Statt, dass die vorliegende Arbeit mit der elektronischen Version übereinstimmt.

**Statement in Lieu of an Oath**

I hereby confirm the congruence of the contents of the printed data and the electronic version of the thesis.

**Einverständniserklärung**

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

**Declaration of Consent**

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

---

(Datum / Date)

---

(Unterschrift / Signature)

## Acknowledgments

# Keyframe-based Visual-Inertial Odometry for Small Workspace

by  
Xi Li

Submitted to the Department of Computer Science  
on June 1, 2016, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Computer Science

## **Abstract**

Thesis Supervisor: Roland Angst  
Title: Dr.



# Contents

<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	Motivation and Contribution . . . . .	9
1.2	Outline of the Thesis . . . . .	11
<b>2</b>	<b>Overview of Visual-inertial Odometry</b>	<b>12</b>
2.1	World Representations and Notations . . . . .	12
2.2	Filter Versus Keyframe . . . . .	12
<b>3</b>	<b>Background on Quaternion Algebra</b>	<b>14</b>
3.1	Definition of Quaternion . . . . .	14
3.2	Properties of Quaternion . . . . .	14
3.3	Quaternions and Rotation operations . . . . .	14
3.4	Time-derivatives and Time-integration on Quaternion . . . . .	14
<b>4</b>	<b>Modular Sensor Fusing</b>	<b>15</b>
4.1	Error-state Kalman Filter for IMU Integration . . . . .	15
4.1.1	Motivation . . . . .	15
4.1.2	System Kinematics . . . . .	15
4.1.3	State Propagations . . . . .	15
4.1.4	State Reset . . . . .	15
4.2	Camera as Complementary Sensory Data . . . . .	15
4.2.1	Motivation . . . . .	15
4.2.2	Self-adapt Map Scale . . . . .	15
4.2.3	Keyframe-based Bundle Adjustment . . . . .	15
4.3	Visual-inertial Odometry Pipeline Overview . . . . .	15
<b>5</b>	<b>Experiments</b>	<b>16</b>
5.1	Synthetic Dataset . . . . .	16
5.2	Some other experiments . . . . .	16
<b>6</b>	<b>Summary, Discussion and Future Works</b>	<b>17</b>
<b>A</b>	<b>Integration Methods</b>	<b>18</b>
A.1	Runge-Kutta Numerical Integration Methods . . . . .	18
A.2	Closed-form Integration Methods . . . . .	18



# Chapter 1

## Introduction

In past few years, the development of *Robotics* has surpassed people's expectation. The word *Robotics* has been first appeared in science fiction "Liar!" by Issac Asimov [21], it referred to science and technology of robots. By definition, *Robotics* is a research branch that related to design, control and application of robots, as well as processing feedback from robots.

Modern robots have been classified into several categories(e.g., Mobile robot, industrial robot, service robot, education robot etc.) with their usages. Among those categories, full-autonomous or semi-autonomous mobile robots attracts more and more researches. Such robots have abilities to move around in their moving space, with or without humans' control. The aim of research in mobile robots is to help us accomplish various hard tasks, whether domestically, commercially or militarily. These tasks, such as assisting disabled people, defusing bombs, or repair equipment in dangerous place is either risky or high expense for human beings.

For mobile robot, finding physical location of itself in unknown environment is normally crucial, such an ability(e.g., *Robot Navigation*) allows mobile robots avoid risky obstacles and finally arrive the goal position. Roughly speaking, *Robot Navigation* is a computing system which processes the information from external sources(e.g., sensors) and apply an algorithm to navigate robot, and sometimes build a map of environment. In *Robot Navigation*, robots sense environmental information by their sensors. These sensors, either locally (e.g., camera, inertial measurement unit (IMU)), or globally (e.g., Global Positioning System) detect events or changes in environment, and transfer their data to robots. Generally, sensors equipped in mobile robot (Local Sensor) are designed light, small, and inexpensive considering convenient movement and low expense. Camera and IMU sensor are considered most common local sensor in small mobile robot system.

A camera is a optical instrument for capturing images. Modern camera has several advantages for *Robot Navigation*. First, the core of camera chip set is cheap and easily installed in any mobile robot system; Second, camera often brings very rich information as it simulates the functioning of human eye. By recognizing key-points [19, 13, 18] in certain images, system observes the *landmarks* in environment, and those *landmarks* will localize robots by *Triangulation* [3, 9, 5].

A IMU sensor (Figure 1-1) often combines *gyroscope* and *accelerometer*, sometimes



*magnetometer*, to measure specific force, angular rate and magnetic field regarding to its local frame. IMU is one of main component in *Inertial Navigation System*, which firstly used in air plane, spacecraft, guided missiles, and now also in mobile robot [2, 10, 15, 7]. IMU sensor utilize *Dead Reckoning* to track device’s position, such technology tries to integrate IMU data over time by assuming the movement model of devices fixed(i.e., acceleration and rotation rate is constant over small period of time).

## 1.1 Motivation and Contribution

The main motivation of this thesis is to improve the navigation accuracy by fusing camera data and IMU sensor data.

Single sensor-based navigation system may not satisfy the requirements of high-quality localization by mobile robot. GPS-based navigation system has been used for outdoor devices for long time. However, such a system suffered from localizing in indoor environment, and also the accuracy of general GPS is not high, i.e., 3 to 5 meters error [20]. For mobile robot, which may move from in-door to out-door, and requires high-accuracy navigation, GPS is mostly not used, or as baseline of navigation [8].

Vision-based navigation system [3, 9], or simultaneous localization and mapping (SLAM) [4, 6, 16] gives an acceptable navigation result. [3] recognizes corner feature by single camera, and it uses a extended kalman filter framework to track the uncertainty and propagates the system state, however it can not handle large-scale scene. [9] utilizes key-frame based bundle adjustment to update the map, improves both accuracy and efficiency, it still met some problem in large-scale scene, becasue vision-based method often is a trade-off between computational complexity and localization accuracy due to the rich information and low output frequency by camera.

Single *Inertial Navigation System* recognize its pose by *Dead Reckoning* [14, 11]. However, the problem of *Dead Reckoning* error accumulation; Only few directions are observable [8] by IMU during whole navigation process. When object corrupts movement assumption, a correction step by external data will be needed.

Fusing camera and IMU data has many advantages. On the one hand, it can decrease computational time by making good use of high-frequency IMU data; On the other hand, it can reduce the error accumulation of IMU integration by the correction of vision-based navigation result within several turn. Fuse the camera and IMU data for robot navigation is not novel. [15] applies multi-state kalman filter (MSCKF) on state update, decrease the computational complexity by only keeping few last information of keyframe. [8] analysis the consistency of MSCKF, correct the ways of IMU integration to obtain consistency of sytem, therefore increase the accuracy. [7] exploits a way to optimize manifold information, therefore solving data fusing problem in a non-linear optimization scheme. Unfortunately, codes and data of above methods are not accessible, that motivates to explore possible methods to increase accuracy of *vision-inertial navigation system* both theoretically and experimentally.

The contributions of this thesis are as follows,

- We exploit a highly flexible, realistic software to generate synthetic IMU sensor

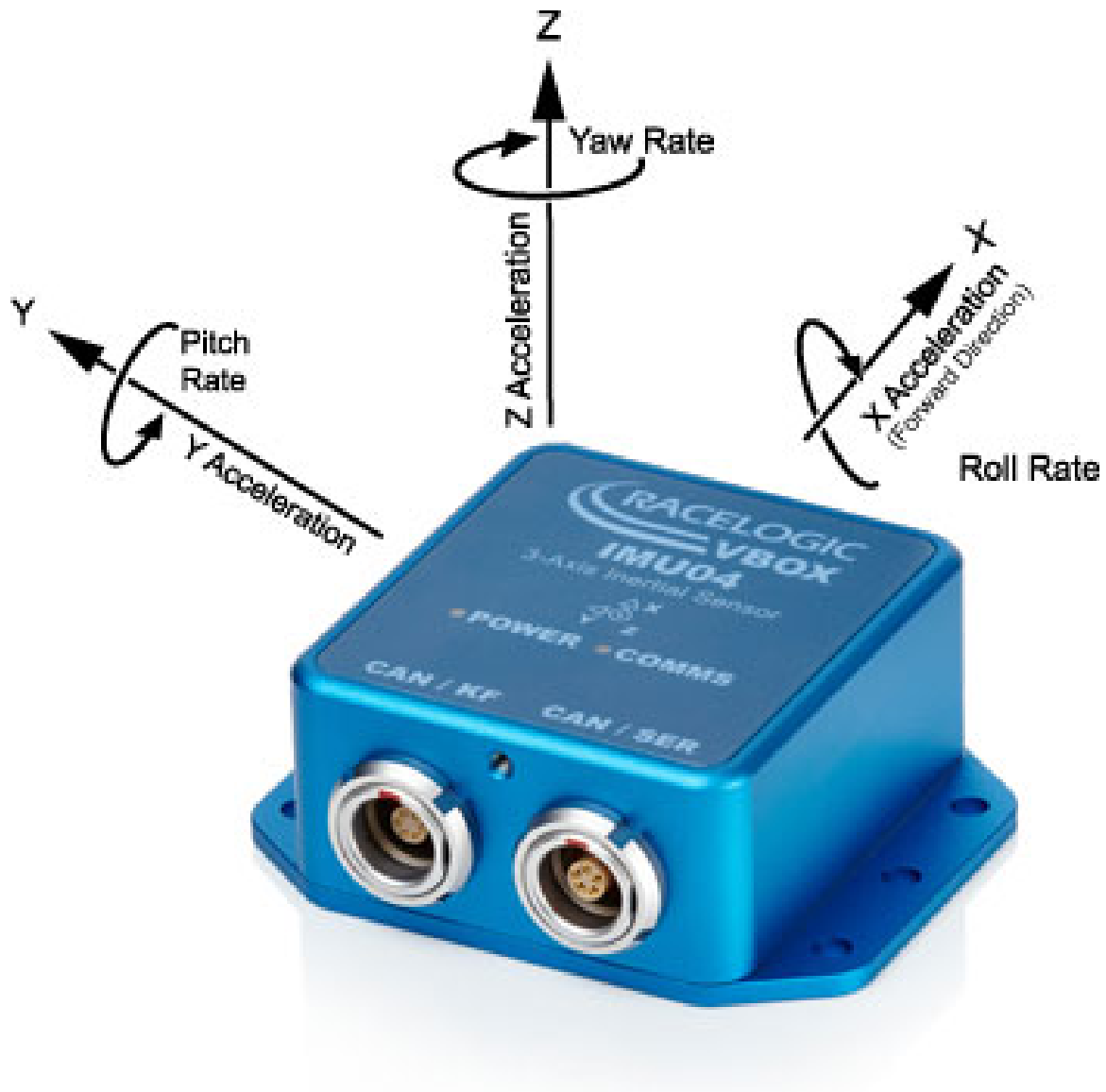


Figure 1-1: IMU sensor with gyroscope and accelerometer measures rotational rate and acceleration of X, Y, Z axis regarding its local frame. Note that IMU sensor normally has bias and irreducible noises, it is necessary to apply a calibration like cameras. The frequency of IMU output is normally larger than 100 Hz. Source: [1]

data and corresponding vision data, that is the main source to provide experimental data in this thesis.

- We present error-state kalman filter system kinematic based on quaternion, explore the multiple ways to integrate IMU data due to different movement models.
- We propose a novel method to update fusing result based on key-framed bundle adjustment.
- We propose a real-time visual-inertial framework implemented in C++, which is stable, scalable, high-accuracy and low-latency.

## 1.2 Outline of the Thesis

In Chapter 2 we first overview our visual-inertial odometry, including world representation, important notations, we also compare filter method and key-frame based method in SLAM problem, and in the end we explain our choice in this thesis.

Then we enter the Chapter 3, which introduces *Quaternion Algebra*. In this chapter, we introduces the basic operations on quaternion, the relationship among quaternion, rotation matrix and rotation vector. In this chapter, we also explain how to integrate or derivative quaternion over time.

Chapter 4 is main part of this thesis. In Section 4.1, we study the Error-State Kalman Filter (ESKF), and apply it to IMU integration; Section 4.2 we summarize how to fuse camera into ESKF, and how to optimize it by key-frame based bundle adjustment. In Section 4.3, we overview the pipeline of our visual-inertial odometry system.

We show our experiment results in Chapter 5. First, the process of generating synthetic IMU and camera data is presented. Then we run several experiments to show our proposed visual-inertial odometry has higher accuracy than single IMU integration, visual SLAM, as our system is still running in real-time.

In the end, we summarize and discuss our work and analysis the potential future work in Chapter 6.

## Chapter 2

# Overview of Visual-inertial Odometry

In this chapter, we will overview visual-inertial odometry system. In section 2.1, we will introduce world representation(e.g., world frame, camera frame and IMU frame) together with basic notations in our odometry system. Then in section 2.2, we will discuss two important scheme, filter method and keyframe Bundle Adjustment(keyframe BA) in SLAM algorithm, and explain why we finally choose keyframe-based method.

### 2.1 World Representations and Notations

Visual-inertial odometry [12], literally, is an odometry system, which received environment information by visual (camera) and inertial (IMU) sensor. VIO is similar with well-known visual odometry (VO) problem [17], with an additional IMU sensor, it tries to estimate agent's pose as agent keep moving in the environment. One big difference between VIO and SLAM algorithm is that VIO does not or only build a simple map, whereas SLAM normally maintains and continuously updating a map.

To setup a VIO system, we need to first define the ways to represent the world. Globally, we have a world frame  $\mathcal{W}$ ; World frame  $\mathcal{W}$  is set to a right-handed Cartesian coordinate system that every objects has an absolute pose (translation and rotation) in it. Then we have local frame for each sensor, i.e., IMU frame  $\mathcal{I}$  and camera frame  $\mathcal{C}$ ; Every time camera and IMU sensor obtain observations within their own local frame, we need to integrate those data and estimate the pose of those sensors in world frame  $\mathcal{W}$ . Figure 2-1 shows the overall world representations.

In this master thesis, we denote measurement  $M$  in particular frame  $\mathcal{F}$  as  $M_{\mathcal{F}}$ . To further simplify, any measurement that is **not** in world frame shall be denoted particularly. For example, the translation  $P$  in camera frame will be denoted as  $P_{\mathcal{C}}$ , and the translation  $P$  in world frame will be denoted as  $P$ .

### 2.2 Filter Versus Keyframe

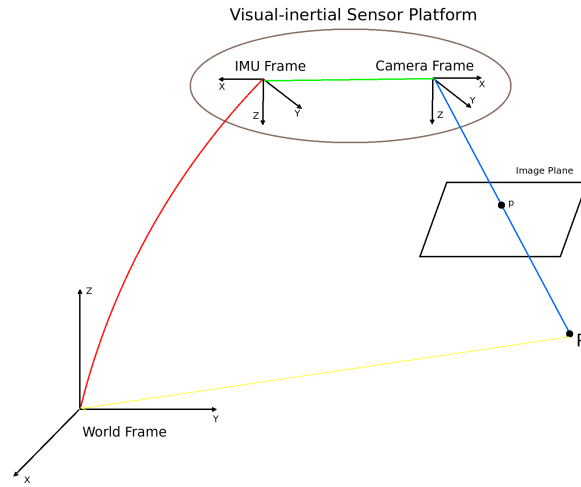


Figure 2-1: This figure shows the connection among world frame  $\mathcal{W}$ , IMU frame  $\mathcal{I}$  and camera frame  $\mathcal{C}$ . Green line shows the transformation between camera and IMU, which can be pre-calibrated. Red line is the pose of IMU in world frame. Camera frame observes point  $p$  of object  $P$  in image plane, and connects a object  $P$  by blue line, and the coordinate of object  $P$  in world frame is presented as yellow line.

## Chapter 3

# Background on Quaternion Algebra

3.1 Definition of Quaternion

3.2 Properties of Quaternion

3.3 Quaternions and Rotation operations

3.4 Time-derivatives and Time-integration on Quaternion

# Chapter 4

## Modular Sensor Fusing

### 4.1 Error-state Kalman Filter for IMU Integration

#### 4.1.1 Motivation

#### 4.1.2 System Kinematics

#### 4.1.3 State Propagations

#### 4.1.4 State Reset

### 4.2 Camera as Complementary Sensory Data

#### 4.2.1 Motivation

#### 4.2.2 Self-adapt Map Scale

#### 4.2.3 Keyframe-based Bundle Adjustment

### 4.3 Visual-inertial Odometry Pipeline Overview

# Chapter 5

## Experiments

### 5.1 Synthetic Dataset

### 5.2 Some other experiments



## Chapter 6

# Summary, Discussion and Future Works

# Appendix A

## Integration Methods

### A.1 Runge-Kutta Numerical Integration Methods

### A.2 Closed-form Integration Methods

# Appendix B

## Approximation Methods

[15]

# Bibliography

- [1] Imu sensor <https://www.vboxautomotive.co.uk/index.php/en/products/modules/inertial-measurement-unit>, 2016. [Online; accessed 22-March-2016].
- [2] Maxim A Batalin, Gaurav S Sukhatme, and Myron Hattig. Mobile robot navigation using a sensor network. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 636–641. IEEE, 2004.
- [3] Andrew J Davison. Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1403–1410. IEEE, 2003.
- [4] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):1052–1067, 2007.
- [5] Ethan Eade and Tom Drummond. Monocular slam as a graph of coalesced observations. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.
- [6] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *Computer Vision–ECCV 2014*, pages 834–849. Springer, 2014.
- [7] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. Imu preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. In *Robotics: Science and Systems (RSS)*, 2015.
- [8] Joel A Hesch, Dimitrios G Kottas, Sean L Bowman, and Stergios I Roumeliotis. Consistency analysis and improvement of vision-aided inertial navigation. *Robotics, IEEE Transactions on*, 30(1):158–176, 2014.
- [9] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.

- [10] Taehee Lee, Joongyou Shin, and Dongil Cho. Position estimation for mobile robot using in-plane 3-axis imu and active beacon. In *Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on*, pages 1956–1961. IEEE, 2009.
- [11] Robert W Levi and Thomas Judd. Dead reckoning navigational system using accelerometer to measure foot impacts, December 10 1996. US Patent 5,583,776.
- [12] Mingyang Li and Anastasios I Mourikis. Consistency of ekf-based visual-inertial odometry. *University of California Riverside, Tech. Rep*, 2011.
- [13] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.
- [14] BL McNaughton, Lijiang Chen, and EJ Markus. “Dead reckoning,” landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. *Cognitive Neuroscience, Journal of*, 3(2):190–202, 1991.
- [15] Anastasios Mourikis, Stergios Roumeliotis, et al. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3565–3572. IEEE, 2007.
- [16] Raul Mur-Artal, JMM Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *Robotics, IEEE Transactions on*, 31(5):1147–1163, 2015.
- [17] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–652. IEEE, 2004.
- [18] Edward Rosten, Reid Porter, and Tom Drummond. Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(1):105–119, 2010.
- [19] Jianbo Shi and Carlo Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.
- [20] Wikipedia. Global positioning system — wikipedia, the free encyclopedia, 2016. [Online; accessed 24-March-2016].
- [21] Wikipedia. Robotics — wikipedia, the free encyclopedia, 2016. [Online; accessed 22-March-2016].