

Visual SLAM for Autonomous Navigation of MAVs

Visual SLAM for Autonomous Navigation of MAVs

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard-Karls-Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Shaowu Yang

aus Chongqing, China

Tübingen

2014

Tag der mündlichen Qualifikation: 28.07.2014
Dekan: Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter: Prof. Dr. Andreas Zell
2. Berichterstatter: Prof. Dr. Andreas Schilling

Berichte aus der Robotik

Shaowu Yang

Visual SLAM for Autonomous Navigation of MAVs

D 21 (Diss. Universität Tübingen)

Shaker Verlag
Aachen 2016

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Tübingen, Univ., Diss., 2014

Copyright Shaker Verlag 2016

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-4228-3

ISSN 1434-8098

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

This thesis focuses on developing onboard visual simultaneous localization and mapping (SLAM) systems to enable autonomous navigation of micro aerial vehicles (MAVs), which is still a challenging topic considering the limited payload and computational capability that an MAV normally has. In MAV applications, the visual SLAM systems are required to be very efficient, especially when other visual tasks have to be done in parallel. Furthermore, robustness in pose tracking is highly desired in order to enable safe autonomous navigation of an MAV in three-dimensional (3D) space. These challenges motivate the work in this thesis in the following aspects.

Firstly, the problem of visual pose estimation for MAVs using an artificial landmark is addressed. An artificial neural network (ANN) is used to robustly recognize this visual marker in cluttered environments. Then a computational projective-geometry method is implemented for relative pose computation based on the retrieved geometry information of the visual marker. The presented vision system can be used not only for pose control of MAVs, but also for providing accurate pose estimates to a monocular visual SLAM system serving as an automatic initialization module for both indoor and outdoor environments.

Secondly, autonomous landing on an arbitrarily textured landing site during autonomous navigation of an MAV is achieved. By integrating an efficient local-feature-based object detection algorithm within a monocular visual SLAM system, the MAV is able to search for the landing site autonomously along a predefined path, and land on it once it has been found. Thus, the proposed monocular visual solution enables autonomous navigation of an MAV in parallel with landing site detection. This solution relaxes the assumption made in conventional vision-guided landing systems, which is that the landing site should be located inside the field of view (FOV) of the vision system before initiating the landing task.

The third problem that is addressed in this thesis is multi-camera visual SLAM for robust pose tracking of MAVs. Due to the limited FOV of a single camera, pose tracking using monocular visual SLAM may easily fail when the MAV navigates in unknown environments. Previous work addresses this problem mainly by fusing information from other sensors, like an inertial measurement unit (IMU), to achieve robustness of the whole system, which does not improve the robustness of visual SLAM itself. This thesis investigates solutions for improving the pose tracking robustness of a visual SLAM system by utilizing multiple cameras. A mathematical analysis of how measurements from multiple cameras should be integrated in the optimization of visual SLAM is provided. The resulting theory allows those measurements to be used for both robust pose tracking

and map updating of the visual SLAM system. Furthermore, such a multi-camera visual SLAM system is modified to be a robust constant-time visual odometry. By integrating this visual odometry with an efficient back-end which consists of loop-closure detection and pose-graph optimization processes, a near-constant time multi-camera visual SLAM system is achieved for autonomous navigation of MAVs in large-scale environments.

Kurzfassung

Diese Arbeit konzentriert sich auf die Entwicklung von integrierten Systemen zur gleichzeitigen Lokalisierung und Kartierung (Simultaneous Localization and Mapping, SLAM) mit Hilfe visueller Sensoren, um die autonome Navigation von kleinen Luftfahrzeugen (Micro Aerial Vehicles, MAVs) zu ermöglichen. Dies ist noch immer ein anspruchsvolles Thema angesichts der meist begrenzten Nutzlast und Rechenleistung eines MAVs. Die dafür eingesetzten visuellen SLAM Systeme müssen sehr effizient zu sein, vor allem wenn parallel noch andere visuelle Aufgaben durchgeführt werden sollen. Darüber hinaus ist eine robuste Positionsschätzung sehr wichtig, um die sichere autonome Navigation des MAVs im dreidimensionalen (3D) Raum zu ermöglichen. Diese Herausforderungen motivieren die vorliegende Arbeit gemäß den folgenden Gesichtspunkten:

Zuerst wird das Problem bearbeitet, die Pose eines MAVs mit Hilfe einer künstlichen Markierung visuell zu schätzen. Ein künstliches neuronales Netz wird verwendet, um diese visuelle Markierung auch in anspruchsvollen Umgebungen zuverlässig zu erkennen. Anschließend wird ein Verfahren aus der projektiven Geometrie eingesetzt, um die relative Pose basierend auf der gemessenen Geometrie der visuellen Markierung zu ermitteln. Das vorgestellte Bildverarbeitungssystem kann nicht nur zur Regelung der Pose des MAVs verwendet werden, sondern auch genaue Posenschätzungen zur automatischen Initialisierung eines monokularen visuellen SLAM-Systems im Innen- und Außenbereich liefern.

Anschließend wird die autonome Landung eines MAVs auf einem beliebig texturierten Landeplatz während autonomer Navigation erreicht. Durch die Integration eines effizienten Objekterkennungsalgorithmus, basierend auf lokalen Bildmerkmalen in einem monokularen visuellen SLAM-System, ist das MAV in der Lage den Landeplatz autonom entlang einer vorgegebenen Strecke zu suchen, und auf ihm zu landen sobald er gefunden wurde. Die vorgestellte Lösung ermöglicht somit die autonome Navigation eines MAVs bei paralleler Landeplatzerkennung. Diese Lösung lockert die gängige Annahme in herkömmlichen Systemen zum kamerageführten Landen, dass der Landeplatz vor Beginn der Landung innerhalb des Sichtfelds des Bildverarbeitungssystems liegen muss.

Das dritte in dieser Arbeit bearbeitete Problem ist visuelles SLAM mit mehreren Kameras zur robusten Posenschätzung für MAVs. Aufgrund des begrenzten Sichtfelds von einer einzigen Kamera kann die Posenschätzung von monokularem visuellem SLAM leicht fehlschlagen, wenn sich das MAV in einer unbekannten Umgebung bewegt. Frühere Arbeiten versuchten dieses Problem hauptsächlich durch die Fusionierung von Informationen anderer Sensoren, z.B. eines Inertialsensors (Inertial Measurement Unit, IMU)

zu lösen um eine höhere Robustheit des Gesamtsystems zu erreichen, was die Robustheit des visuellen SLAM-Systems selbst nicht verbessert. Die vorliegende Arbeit untersucht Lösungen zur Verbesserung der Robustheit der Posenschätzung eines visuellen SLAM-Systems durch die Verwendung mehrerer Kameras. Wie Messungen von mehreren Kameras in die Optimierung für visuelles SLAM integriert werden können wird mathematisch analysiert. Die daraus resultierende Theorie erlaubt die Nutzung dieser Messungen sowohl zur robusten Posenschätzung als auch zur Aktualisierung der visuellen Karte. Ferner wird ein solches visuelles SLAM-System mit mehreren Kameras modifiziert, um in konstanter Laufzeit robuste visuelle Odometrie zu berechnen. Die Integration dieser visuellen Odometrie mit einem effizienten Back-End zur Erkennung von geschlossener Schleifen und der Optimierung des Posengraphen ermöglicht ein visuelles SLAM-System mit mehreren Kameras und fast konstanter Laufzeit zur autonomen Navigation von MAVs in großen Umgebungen.

Acknowledgements

First of all, I am most grateful to my supervisor, Prof. Andreas Zell, not only for his guidance and support throughout my study towards this thesis, but also for this care about my daily life, especially when I just arrived in Tübingen. He has provided me enough freedom in doing research, which helps me to focus on my research topics, and furthermore, to be able to appreciate the beautiful landscape outside my office in Sand.

I am very grateful to other co-authors of my previous publications: Sebastian Scherer, Konstantin Schauwecker, and Andreas Masselli. I have benefited greatly from collaborations with them. Furthermore, I would like to thank the colleagues who have done proofreading of individual chapters of this thesis: Yuyi Liu, Konstantin Schauwecker, Jacobo Jimenez, Nedim Šrndić and Sebastian Scherer.

I would like to thank Prof. Andreas Schilling for providing us with the access to the external pose tracking system, which plays a major role in evaluating the proposed vision systems in this thesis. I am also very grateful to him for his valuable comments to this thesis.

I am very grateful to the nice study environment that the Chair of Cognitive Systems has provided during the years of my study in Tübingen. I would like to express my gratitude to all the colleagues and former colleagues in our group. Special thanks should be given to Klaus Beyreuther and Vita Serbakova who offer assistance to our work.

I would like to thank all my friends in Tübingen who have shared the great time with me during my study, and kept me away from loneliness when I was in this foreign country being far away from home.

Finally, I would like to thank my family for their great love and support, especially my parents, my parents in law, and my wife Qian Zhang.

The China Scholarship Council (CSC) is gratefully acknowledged for providing me the scholarship for my study in Tübingen. The National University of Defense Technology in China is acknowledged for the support to publishing this work.

Acknowledgements

Contents

Notation	xv
1 Introduction	1
1.1 Motivation	1
1.2 Contributions	5
2 Background	9
2.1 The MAV Platform	9
2.2 MAV Flight Control	11
2.2.1 Quadrotor configuration and dynamic model	12
2.2.2 Attitude control and hover control	13
2.2.3 Trajectory control	15
2.3 Camera Model and Calibration	16
2.3.1 Perspective camera model	16
2.3.2 Camera calibration	20
2.3.3 Extrinsic calibrations	21
2.4 Visual SLAM for MAVs	23
2.4.1 The PTAM system	24
2.4.2 Related work	26
3 Artificial-Landmark-Based Visual Pose Estimation	29
3.1 Introduction	29
3.2 Related Work	30
3.3 Artificial-Landmark Recognition	32
3.3.1 The artificial landmark and coordinate systems	32
3.3.2 Vision algorithm for landmark detection	32
3.3.3 Retrieving the geometry information	37
3.4 6DOF Pose Estimation	38
3.4.1 5DOF camera pose estimation	38
3.4.2 Resolving the remaining DOF	41
3.4.3 6DOF MAV pose	41
3.5 Experiments and Results	42
3.5.1 Landmark detection and ellipse fitting	42
3.5.2 6DOF pose estimation results	43

3.6	Conclusions	50
4	Visual-SLAM-based Autonomous Landing of MAVs	51
4.1	Introduction	51
4.2	Related Work	52
4.3	Monocular Visual SLAM for Autonomous MAVs	54
4.3.1	Using PTAM in near-constant time	54
4.3.2	Automatic initialization of the SLAM system	54
4.4	Landing Site Detection and Pose Estimation	56
4.4.1	Brief overview of the ORB features	56
4.4.2	Applying multi-scale ORB to the SLAM framework	57
4.4.3	Landing site detection by feature matching	57
4.4.4	Locating the landing site within the map	59
4.5	Experiments and Results	60
4.5.1	Experimental setup	60
4.5.2	Landing site position estimation results	62
4.5.3	Autonomous navigation and landing results	62
4.6	Outdoor Experience	67
4.6.1	Outdoor experiment	67
4.6.2	Discussions	69
4.7	Conclusions and Discussions	71
5	Multi-Camera Visual SLAM for Autonomous MAVs	73
5.1	Introduction	73
5.2	Related Work	75
5.3	Multi-Camera Visual SLAM	76
5.3.1	Multi-camera projection and pose tracking	76
5.3.2	Optimizations in the multi-camera SLAM	77
5.3.3	Pose update with multiple cameras	78
5.3.4	Bundle adjustment with multiple cameras	79
5.4	Implementation	80
5.4.1	Organizing the map	80
5.4.2	Pose tracking	81
5.4.3	Mapping	81
5.4.4	Automatic Initialization	82
5.5	Experiments	83
5.5.1	Experimental setup	83
5.5.2	Enabling autonomous navigation	83
5.5.3	Further evaluation through manual flight	86
5.6	Conclusions and Discussions	89
6	Towards Constant-Time Multi-Camera Visual SLAM for MAVs	91

6.1	Introduction	91
6.2	Related Work	93
6.3	Back-End of the SLAM System	94
6.3.1	The global map representation	94
6.3.2	Loop-closure detection	95
6.3.3	Adaptive-window pose-graph optimization	96
6.4	Implementation	98
6.4.1	The visual odometry	98
6.4.2	The back-end	100
6.5	Experiments	101
6.5.1	Enabling autonomous navigation	101
6.5.2	Further evaluations with manual flight data	103
6.6	Conclusions and Discussions	106
7	Conclusions	109
7.1	Summary	109
7.2	Future work	110
	Bibliography	113