

# Toward Full Autonomy for Vision-Guided Robots

## From Self-Calibration to Self-Directed Exploration

**Doctoral Thesis**

**Author(s):**

Heng, Lionel

**Publication date:**

2014

**Permanent link:**

<https://doi.org/10.3929/ethz-a-010410672>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

DISS. ETH NO. 22400

# **Toward Full Autonomy for Vision-Guided Robots: From Self-Calibration to Self-Directed Exploration**

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by

**LIONEL HENG**

M.S. in Computer Science  
Stanford University

born on 12.08.1983

citizen of Singapore

accepted on the recommendation of

Prof. Marc Pollefeys  
Prof. Roland Siegwart  
Prof. Wolfram Burgard

2014

# Abstract

Many building blocks have to be put in place for full robot autonomy to happen: self-calibration, control, perception, path planning, and high-level behaviors. Hence, a considerable number of algorithms has to run on-board a robot, which however, has a finite amount of computational resources. Vision-based perception algorithms are inherently computationally expensive, and this fact makes full autonomy for a vision-guided robot even more challenging to achieve. Given only 2D image data, computationally efficient algorithms are needed to compute accurately and in real-time the robot's current pose and the 3D geometric representation of the environment.

An accurate 3D map of the environment with a high update rate is a primary requisite for high-level behaviors on the robot. We propose a 3D occupancy mapping algorithm that runs in real-time, has constant space complexity, and scales well to large environments. Through extensive real-world experiments with a micro aerial vehicle (MAV) equipped with a forward-looking stereo camera, we demonstrate our 3D occupancy mapping algorithm to work well in both indoor and outdoor environments. In addition, we build on the 3D mapping results by showing wall-following and frontier-based exploration behaviors. With automated 3D reconstruction in mind, we continue our work on exploration by proposing an algorithm for simultaneous exploration and coverage. Simulation experiments show the feasibility of this algorithm.

The use of multiple cameras increases a robot's perceptual awareness of its environment. However, transformations between cameras must be known with high accuracy. In this case, an unsupervised and markerless calibration algorithm is ideal as expert supervision and modification of the environment are not required. We propose SLAM-based self-calibration algorithms for two different configurations of multi-sensor systems: one with multiple monocular cameras and a calibrated odometry system, and one with multiple calibrated stereo cameras. Metric scale is inferred from odometry and calibrated stereo

---

respectively. These self-calibration algorithms do not assume an overlapping field of view between any two cameras. We also come up with an infrastructure-based calibration algorithm that leverages image-based localization and the map built by SLAM-based self-calibration to calibrate a multi-camera system in near real-time.

Motion estimation algorithms for single cameras cannot easily be extended to multi-camera systems. Using the generalized camera model, we propose a 3-point motion estimation algorithm for a calibrated multi-camera system. This algorithm uses inertial information from an IMU and estimates the relative motion with metric scale. In addition, we use this multi-sensor system on a MAV to demonstrate real-time and on-board SLAM with loop closures.

All these works enable vision-guided robots to take a big step towards full autonomy. The work on self-calibration and infrastructure-based calibration enables a robot to perform self-maintenance, facilitating long-term operation in the field without expert supervision. The mapping algorithm allows robots to efficiently map out large environments without running out of memory. The SLAM implementation allows robots to use multiple cameras at the same time for pose estimation. The successful demonstration of high-level behaviors in the form of wall-following, frontier-based exploration, and simultaneous exploration and coverage serves to show that full autonomy for vision-guided robots can be a realistic goal.

# Zusammenfassung

Um die vollständige Autonomie von Robotern zu ermöglichen, werden viele Baublöcke benötigt: Selbstkalibrierung, Lenkung, Perzeption, das Planen von Pfaden und intelligentes Verhalten. Daher muss eine erhebliche Anzahl von Algorithmen auf dem Roboter selbst ausgeführt werden. Allerdings sind die Ressourcen des Roboters für Berechnungen limitiert. Algorithmen zur bildbasierten Perzeption sind inherent rechenintensiv. Dieser Umstand erschwert das Erreichen vollständiger Autonomie eines durch Bilder gesteuerten Roboters zusätzlich. Um die aktuelle Positionierung des Roboters und die dreidimensionale Repräsentation der Umgebung akkurat und in Echtzeit alleine anhand von 2D-Bilddaten zu berechnen, werden effiziente Algorithmen benötigt.

Eine genaue dreidimensionale Karte der Umgebung, welche mit hoher Frequenz aktualisiert wird, ist die Hauptvoraussetzung für das intelligente Verhalten des Roboters. Wir schlagen einen Algorithmus basierend auf einer dreidimensionalen Karte der Raumebelegung vor. Der Algorithmus läuft in Echtzeit, verfügt über konstante Speicherkomplexität und funktioniert auch für grosse Umgebungen gut. Durch ausgiebige Experimente in der realen Welt mit einem kleinen Fluggefährt ('Micro Aerial Vehicle', kurz MAV), welches mit einer nahe vorne ausgerichteten Kamera ausgestattet wurde, zeigen wir, dass unser Algorithmus zur Erstellung einer dreidimensionalen Karte der Raumebelegung sowohl inner- als auch ausserhalb von Gebäuden funktioniert. Zusätzlich bauen wir auf den Resultaten zur Erstellung dreidimensionaler Karten auf, indem wir das Folgen entlang von Wänden und grenzbasierte Erkundung zeigen. Mit Blick auf automatisierte dreidimensionale Rekonstruktion setzen wir unsere Arbeit an der Erkundung fort, indem wir einen Algorithmus zur simultanen Erkundung und Abdeckung vorschlagen. Simulationen zeigen die Umsetzbarkeit dieses Algorithmus.

Die Verwendung mehrerer Kameras erhöht die Wahrnehmung der Umgebung des Roboters. Allerdings müssen die Transformationen zwischen den

---

Kameras mit hoher Genauigkeit bekannt sein. In diesem Fall ist ein unbeaufsichtigter Algorithmus, welcher keine Marker verwendet, ideal, da er keine Beaufsichtigung durch Experten und keine Anpassung der Umgebung benötigt. Wir schlagen Algorithmen zur SLAM-basierten Selbstkalibrierung für zwei verschiedene Konfigurationen von Systemen mit mehreren Sensoren vor: einen mit mehreren monokularen Kameras und einem kalibrierten System für Odometrie und einen mit mehreren kalibrierten Stereokameras. Die metrische Skala wird von der Odometrie, beziehungsweise von kalibrierten Stereoaufnahmen abgeleitet. Diese Algorithmen zur Selbstkalibrierung gehen nicht von überlappenden Gesichtsfeldern zwischen irgendwelchen zwei Kameras aus. Zusätzlich schlagen wir einen Algorithmus zur auf Infrastruktur basierenden Kalibrierung vor, welcher die bildbasierte Lokalisierung und die durch SLAM-basierte Selbstkalibrierung erstellte Karte ausnutzt, um ein Multi-Kamera-System in Nahe-Echtzeit zu kalibrieren.

Algorithmen zur Schätzung von Bewegungen für einzelne Kameras können nicht einfach für Multi-Kamera-Systeme erweitert werden. Unter Verwendung des generalisierten Kameramodells schlagen wir einen 3-Punkte-Algorithmus zur Schätzung der Bewegung eines kalibrierten Multi-Kamera-Systems vor. Dieser Algorithmus verwendet die Informationen eines IMU-Systems und schätzt die relative Bewegung mit einer metrischen Skala. Zusätzlich verwenden wir dieses Multi-Sensor-System auf einem MAV um SLAM inklusive dem Schliessen von Schleifen in Echtzeit zu demonstrieren.

All diese Arbeiten ermöglichen den bildbasiert gesteuerten Robotern einen grossen Schritt in Richtung vollständiger Autonomie. Die Arbeit an Selbstkalibrierung und Kalibrierung basierend auf Infrastruktur ermöglicht es einem Roboter, sich selbst instand zu halten, was wiederum lange andauernde praktische Einsätze ohne Beaufsichtigung von Experten ermöglicht. Der Algorithmus zur Erstellung von Karten erlaubt es Robotern grosse Umgebungen aufzuzeichnen, ohne dass der Speicherplatz ausgeht. Die Implementierung von SLAM erlaubt es Robotern mehrere Kameras gleichzeitig für das Schätzen von Posen zu verwenden. Die erfolgreiche Demonstration von intelligentem Verhalten in Form vom Folgen entlang von Wänden, der grenzbasierten Erkundung und der simultanen Erkundung und Abdeckung dient dem Zweck zu zeigen, dass vollständige Autonomie für bildbasiert gesteuerte Roboter ein realistisches Ziel sein kann.