



Doctoral Thesis

Leveraging Geometric Priors and Measurements in 3D Modeling, Calibration and Registration

Author(s):

Zeisl, Bernhard

Publication Date:

2016

Permanent Link:

<https://doi.org/10.3929/ethz-a-010654206> →

Rights / License:

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

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 23157

Leveraging Geometric Priors and Measurements in 3D Modeling, Calibration and Registration

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Bernhard Zeisl

Dipl.-Ing. Univ., Technische Universität München

born 5. August 1982

citizen of Austria

accepted on the recommendation of

Prof. Marc Pollefeys
Prof. Konrad Schindler
Dr. Jürgen Sturm
Dr. Kevin Köser

2016

Abstract

Two of the primary concepts in 3D computer vision are the reconstruction of 3D models and the subsequent estimation of the location and orientation of a camera therein. Due to a remarkable progress in the development of related algorithms in the last decade – for example, already modeling cities from images only – they nowadays form the basis of numerous and diverse applications, like digital map building for later virtual exploration, autonomous navigation, augmented reality, or the digital preservation of cultural heritage. However, algorithms are still error-prone in often occurring settings, *e.g.*, in texture-less scenes, in the presence of strong viewpoint changes between individual images, or they become computationally intractable for large-scale environments.

Therefore, this thesis examines approaches for 3D modeling and registration tasks and puts its focus on the exploitation of structural priors and available geometric measurements to achieve robust algorithms, while simultaneously aiming for computational speedup and improved precision.

In this regard, we consider computational stereo in texture-less environments, and bootstrap the extraction of a meaningful and visually pleasing geometry by enforcing a strong prior that favors vertical wall elements. The resulting parameterization allows us to transform the reconstruction problem into a dynamic programming optimization that computes regularized depth maps at interactive frame rates.

Since, RGB-D cameras can replace computational stereo in indoor settings nowadays, we consider such sensors and present a structure-based auto-calibration method that jointly determines the extrinsic pose between the color camera and depth sensor, as well as the typically present distortions in the depth measurements. The obtained calibration allows for instant, accurate modeling without the need for any artificial calibration targets.

To efficiently compute a consistent model from individual RGB-D scans, automatic model registration is required. Especially terrestrial laser scans of-

ten exhibit strong viewpoint distortions, since the number of scan positions is minimized for reasons of efficiency. We illustrate how to leverage developable surfaces and salient directions, both extracted from the underlying scene geometry, to obtain rectified scene projections that enable discriminative feature matching and consequently fully automatic registration of scans with only limited overlap.

In a final step, we aim for the estimation of the camera pose with respect to a previously built reconstruction. For large-scale 3D models, *e.g.*, on the scale of a city, the 2D-3D correspondence search yields many tentative correspondences with a very low inlier ratio. We illustrate the use of simple, but efficient geometric filters to reject outliers and propose to formulate the camera pose estimation as a voting procedure. This results in a linear run-time, multi-modal pose estimates which are well suited to indicate repetitive structures, and an increased precision compared to state-of-the-art.

Zusammenfassung

Zwei der wichtigsten Konzepte im maschinellen, dreidimensionalen Bildverstehen sind die Rekonstruktion von 3D Modellen und die darauf folgende Schätzung der Position und Orientierung einer Kamera darin. Im letzten Jahrzehnt wurde ein bemerkenswerter Fortschritt in der Entwicklung von entsprechenden Algorithmen erzielt, z.B. erfolgte die 3D Modellierung von Städten ausschliesslich aus Bildmaterial. Deshalb bilden 3D Computer Vision Algorithmen heutzutage die Basis für unterschiedlichste Anwendungen, wie beispielsweise digitale Kartografie (inklusive der späteren virtuelle Erkundung der 3D Karte), autonome Navigation, Augmented Reality, oder auch die digitale Konservierung von Kulturerbe. Allerdings sind die zugrunde liegenden Algorithmen in regelmässig auftretenden Situationen weiterhin fehlerhaft, wie z.B. in wenig strukturierten Szenen, bei stark variierenden Ansichten zwischen einzelnen Bildern, oder werden zu rechenintensiv für grossräumige Umgebungen.

Aus diesem Grund untersucht diese Doktorarbeit neue Ansätze für die 3D Modellierung und Registrierung und setzt den Schwerpunkt auf die Behandlung und den Einbezug von bekannten strukturellen Beschränkungen und geometrischen Messungen mit dem Ziel der Entwicklung von robuster Algorithmen, bei gleichzeitig schnellerer Berechnungen und verbesserter Genauigkeit.

Zu Beginn betrachten wir die Tiefenschätzung aus Stereobildern in wenig strukturierten Umgebungen und ermöglichen die Extraktion einer aussagekräftigen und visuell ansprechenden 3D Geometrie mit Hilfe der Beschränkung auf vertikale Wandelemente. Die daraus resultierende Parameterisierung des Problems erlaubt es die Rekonstruktion als dynamische Programmierung aufzufassen, und ermöglicht die Berechnung von geglättete Tiefenkarten in Echtzeit.

Da heutzutage RGB-D Kameras die passive Stereoberechnung in Innenräumen ersetzen können, widmen wir uns solchen Sensoren und präsentieren eine automatische struktur-basierte Kalibrationsmethode, welche die extrinsische Pose zwischen der Farbkamera und dem Tiefensensor sowie die typischen Verzerrungen in den Tiefenmessungen in einem Schritt ermittelt. Die erzielte

Kalibration ermöglicht eine sofortige und akkurate 3D Modellierung ohne auf künstliche Kalibrationsschablonen angewiesen zu sein.

Um ein konsistentes 3D Modell aus individuellen RGB-D Aufnahmen effizient zu berechnen, benötigt es einer automatischen Registrierung. Besonders bei der Arbeit mit terrestrischen Laser Scannern variieren die Standpunkten oft stark voneinander um die Anzahl der Messpositionen aus Effizienzgründen zu minimieren. In dieser Arbeit zeigen wir, wie abwickelbare Oberflächen und typisch auftretende Richtungen in der zugrunde liegenden Geometrie genutzt werden können um entzerrte, d.h. Standpunkt unabhängige, Szenenprojektionen zu erhalten, welche eine differenzierte Punktkorrespondenzbestimmung und infolgedessen eine vollkommen automatische Registrierung von Messungen mit nur geringer Überlappung ermöglichen.

In einem letzten Schritt zielen wir auf die Schätzung der Kamera Pose in Bezug auf eine zuvor berechnete Rekonstruktion. Für grosse 3D Modelle (z.B. in der Grösse einer gesamten Stadt) liefert die 2D-3D Punktkorrespondenzsuche viele mögliche Übereinstimmungen, allerdings nur mit einem geringen Prozentsatz an wirklich korrekten Korrespondenzen. Wir veranschaulichen die Verwendung von einfachen, jedoch effektiven geometrischen Filtern zur Detektion von falschen Korrespondenzen und formulieren die Schätzung der Kamerapose als Abstimmungsverfahren. Daraus resultiert eine lineare Laufzeit des Algorithmus, eine multimodale Posenschätzungen welche auf repetitive Szenenstrukturen anwendbar ist, sowie eine erhöhte Genauigkeit im Vergleich zu aktuellen Methoden.