Visuelle Odometrie und dichte Rekonstruktion für mobile Roboter[1]

Christian Forster[2]

Abstract: Damit ein autonomer Roboter sich bewegen und den Raum mit Menschen teilen kann, muss er eine interne Repräsentation seiner Umgebung erstellen und sich gleichzeitig darin positionieren. In dieser Dissertation werden Methoden untersucht, wie ein Roboter in Echtzeit eine solche Repräsentation mittels Kameras und Inertialsensoren erstellen kann. Der erste Beitrag ist eine semidirekte Methode zur Bewegungsschätzung welche sehr genau und robust ist und darüber hinaus den aktuellen Stand der Technik im Bereich Rechenzeit signifikant übertrifft. Im zweiten Beitrag wird gezeigt, wie die Genauigkeit der vorgeschlagenen Methode durch die Fusion mit Inertialsensoren nochmals verbessert werden kann. In einem Experiment wurde gezeigt wie die Kameraposition über eine Distanz von 300 Metern auf 0.3 Meter genau bestimmt werden kann. Im dritten Beitrag wird eine probabilistische Methode entwickelt welche es ermöglicht die Oberflächenrekonstruktion zu verdichten. In einem integrierten System erlaubt dieser Algorithmus einem autonomen Mikrodrone selbständig einen Landeplatz zu finden welcher frei von Hinternissen ist. Der letze Beitrag nützt die Tatsache aus, dass der Roboter die Datenaufnahme beeinflussen kann. Es wird ein Algorithmus entwickelt der die optimale Trajektorie berechnet um möglichst schnell die Tiefe jedes Pixels im Referenzbild zu schätzen, d.h. die Oberfläche im Bild dreidimensional rekonstruiert.

1 Motivation

Es scheint problemlos, wie Menschen die Umgebung wahrnehmen und damit interagieren. Die Fähigkeit des visuellen Cortexes, die riesige Masse an Informationen, bereitgestellt von der Retina, zu verarbeiten, ist überwältigend. Sofort können wir unsere Bewegung im Raum beschreiben, die Grösse und Struktur vom Raum, in dem wir uns befinden, charakterisieren und nur schwer werden wir von Reflektionen, Schatten und Verdeckungen getäuscht. Menschenähnliche Wahrnehmung von Raum und Bewegung in künstlichen Systemen zu reproduzieren ist eine riesige Herausforderung für die Forschung. Nichts desto trotz, schon die kleinsten Schritte in diese Richtung haben das Potential, eine Flut an industriellen Anwendungen zu ermöglichen. Beispiele von diesem Potential sind autonome Autos, Serviceroboter und Assistenzsysteme für Blinde. Und tatsächlich haben wir guten Grund optimistisch zu sein: In den letzten zwanzig Jahren wurde ein gigantischer Fortschritt gemacht, unterstützt durch die stetige Verbesserung von Prozessoren und Sensorik. Heute sind Computer besser als Menschen in der Detektion von Strassenschildern und die autonomen Autos von Google sind schon mehr als eine Million Meilen selbst gefahren.

[1] Englischer Titel der Dissertation: "Visual Inertial Odometry and Active Dense Reconstruction for Mobile Robots"

[2] Robotics and Perception Group, Universität Zürich, ch.forster@gmail.com

2 Kamerabasierte Bewegungsschätzung und Rekonstruktion

Damit ein autonomer Roboter sich bewegen und den Raum mit Menschen teilen kann, muss er eine interne Repräsentation seiner Umgebung erstellen. Die ideale Repräsentation hängt von der Aufgabe ab, die der Roboter ausführen soll. Muss der Roboter zum Beispiel ein Paket transportieren, dann ist eine nützliche Repräsentation eine Karte von Merkmalen der Umgebung, die der Roboter mit seiner Sensorik wiedererkennen kann. Ist diese Karte nicht von Anfang an verfügbar, muss sie der Robot selbst erstellen, während er sich gleichzeitig darin lokalisiert. Dieses berühmte Problem der Robotik ist bekannt unter der Abkürzung SLAM, was für "*Simultaneous Localization and Mapping*" steht. Kameras sind sehr nützliche Sensoren für mobile Roboter, da sie sehr klein, günstig und allgegenwertig sind. Weil jedes einzelne Kamerabild aber aus hunderttausenden Pixel-Messungen besteht, ist es eine grosse Herausforderung, aus dieser Datenflut die Kamerabewegung und gleichzeitig die Umgebung dreidimensional zu rekonstruieren. Dieser Prozess wird auch als visual-SLAM bezeichnet, oder als visuelle Odometrie. Noch schwieriger wird es, wenn dies in Echtzeit auf einer Recheneinheit mit beschränkter Kapazität, wie sie in Robotern eingesetzt wird, geschehen soll. Ausserdem wird die Robustheit des Systems ein sehr wichtiger Faktor, wenn der mobile Roboter sich in einer unkontrollierten Umgebung bewegt. In diesem Fall treten Verdeckungen, Beleuchtungsänderungen und wenig texturierte Oberflächen auf, was das Wiedererkennen von visuellen Merkmalen im Bild verhindert und deshalb die kamerabasierte Bewegungsschätzung erschwert.

3 1. Beitrag: Visuelle Odometrie

Der erste Beitrag dieser Dissertation [Fo16] ist ein effizienter, robuster und sehr genauer Algorithmus für die visuelle Odometrie. Dieser Algorithmus schätzt die Bewegung einer einzelnen Kamera ausschliesslich anhand der von der Kamera aufgenommenen Bildern. Dazu wurden *direkte Methoden*, welche mit den Intensitätswerten der Pixel operieren, untersucht. Dies steht im Gegensatz zu merkmalbasierten Methoden, welche als erster Schritt visuelle Merkmale im Bild detektieren und daraufhin die restliche Bildinformation verwerfen [Fo12, Fo13, FPS13]. Ein Vorteil von direkten Methoden ist, dass die Pixel-Korrespondenz von Bild zu Bild durch die Geometrie des Problems gegeben ist und durch die Minimierung von Pixel-Intensitätsunterschieden weiter optimiert werden kann. Die Optimierung der Kamerapositionen und der 3D Geometrie der Umgebungsstruktur wird jedoch sehr rechenintensiv, wenn die Karte wächst. Daher wird ein halb-direkter (*semi-direct*) Algorithmus vorgeschlagen [Fo17b, FPS14b], der in einem ersten Schritt die Pixel-Korrespondenz mittels direkten Ansätzen ermittelt und daraufhin auf bewährten merkmalbasierten Methoden aufbaut, um die Geometrie des Problems zu optimieren.

Die bedeutendste Innovation dabei ist der vorgeschlagene *Sparse-Image-Align* Algorithmus, der zwei Bilder robust und effizient anhand der Intensitätswerten von ausgewählten Pixeln aligniert. Im Zusammenspiel mit einem direkten Ansatz für die Schätzung der Tiefenwerte von den ausgewählten Pixeln, erlaubt diese Methode die Bewegungsschätzung in Umgebungen mit sehr wenig Textur. Abbildung 1 zeigt die visuellen Merkmale, sowohl

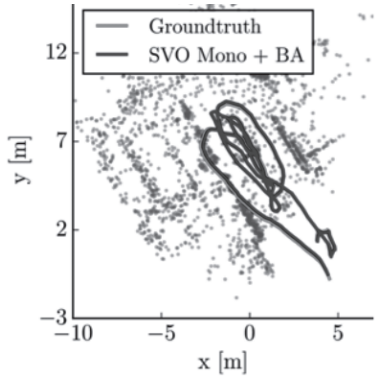

Abb. 1: Die Abbildung links zeigt die visuellen Merkmale, die der Algorithmus detektiert und kartographiert. Durch die dreidimensionale Rekonstruktion dieser Eck- und Kantenpunkte über mehrere Bilder, wird die Bewegung der Kamera geschätzt. Die Abbildung rechts zeigt die rekonstruierten Merkmale im dreidimensionalen Raum und die berechnete Trajektorie, die mit der wahren Trajektorie verglichen wird.

Eckpunkte als auch Kantenpunkte, die der Algorithmus nutzt, um die Bewegung und die dreidimensionale Karte der Umgebung zu schätzen.

Experimentelle Resultate zeigen, dass der vorgeschlagene Algorithmus sehr genaue Schätzungen in Echtzeit erzielt, wobei insbesondere in Bezug auf die Rechenzeit signifikant bessere Resultate als im aktuellen Stand der Technik erreicht werden. In einer Erweiterung wird ausserdem gezeigt, wie der Algorithmus auf Kamerasysteme mit mehreren Kameras sowie auch Fischaugen-Kameras erweitert werden kann [Zh16]. Unter folgendem Link kann ein Video mit einer Zusammenfassung der Experimente und Resultate abgerufen werden: https://youtu.be/hR8uq1RTUfA

Seine Robustheit hat dieser Algorithmus durch seinen täglichen Einsatz zur Bewegungsschätzung von kamerabasierten Mikrodrohnen in der Robotics and Perception Group bewiesen [Fa15, Gi15]. Über 300 Mal wurde das System schon an Messen, Konferenzen und internen Demonstrationen vorgeführt. Ausserdem wurde die entwickelte Software an Firmen lizenziert, die sie zum Beispiel für die 3D Rekonstruktion mit dem Smartphone und für Anwendungen der Virtuellen Realität einsetzen. Eine Open-Source Version der Software[3] wurde der Forschungsgemeinschaft frei zur Verfügung gestellt und wird heute zum Beispiel vom Autonomy and Robotics Center der NASA Langley eingesetzt.

[3] Die Software is verfügbar unter `https://github.com/uzh-rpg/rpg_svo`

3.1 2. Beitrag: Fusion von Kameras mit Intertialsensoren

In einer Erweiterung wird gezeigt, wie Inertialmessungen von Gyroskop und Beschläunigungssensor in die vorgeschlagene visuelle Odometrie integriert werden können [Fo15a, Fo17a]. Dies stellt den zweiten Beitrag dieser Dissertation dar. Die Anwendung von Inertialsensoren, welche die Beschleunigung und Rotationsgeschwindigkeit messen, ist ideal, da die Messungen komplementär zur visuellen Sensorik sind und darüber hinaus preislich so günstig sind, dass sie in vielen mobilen Geräten bereits verbaut werden. Inertialsensoren werden nicht durch visuelle Störungen beeinträchtigt und bieten über einen kurzen Zeithorizont (im Rahmen von Bruchteilen von Sekunden) eine gute Schätzung der Bewegung. Über längere Zeit hingegen weist die Integration von Inertialmessungen einen starken Drift auf (getrieben durch die zweifache Integration von Messrauschen), was wiederum mit der Detektion von visuellen Messungen kompensiert werden kann.

In der Dissertation wird gezeigt, wie die Messungen dieser beiden Sensorsysteme modelliert und in einer Optimierung fusioniert werden können, um die optimale Trajektorie zu berechnen. Im Gegensatz zu früheren Arbeiten, wird eine effizientere Methode für die Optimierung vorgeschlagen. Dies resultiert aus der formalen Beschreibung der Messungenauigkeiten der beiden Sensorsysteme, was eine analytische Berechnung der Ableitungen der Kostenfunktionen ermöglicht. In einem Experiment wurde gezeigt, dass dieser Ansatz über 360 Meter Distanz nur 30 cm Fehler akkumuliert, was halb so viel ist wie die besten Vergleichsmethoden. In der Abbildung 2 werden zwei weitere Experimente veranschaulicht, in welchen die Kamera mit den Intertialsensoren um ein Gebäude und in einem Treppenhaus bewegt wird.

Die Software zur Datenfusion wurde im Open-Source Projekt *GTSAM* integriert [4] und ein Video mit einer Zusammenfassung der Resultate kann unter folgendem Link abgerufen werden: https://youtu.be/CsJkci5lfco

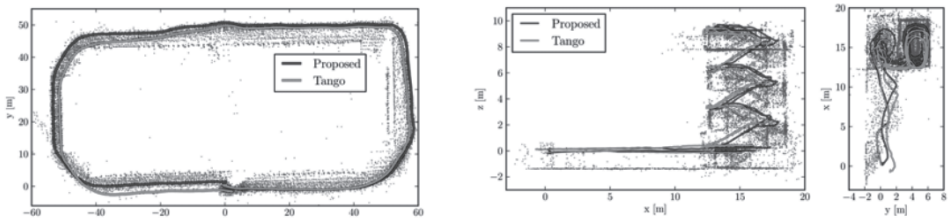

Abb. 2: Vergleich des vorgeschlagenen Algorithmus, welcher Bild und Inertialmessungen kombiniert, gegenüber dem Google Tango Sensor — einem kommerziellen Produkt, das ebenfalls auf Bild und Inertialmessungen basiert. Während im Bild links der Tango Sensor 2.2 Meter Drift akkumuliert, erreicht der vorgeschlagene Ansatz weniger als 1.0 Meter Drift.

[4] Die Software is verfügbar unter `https://bitbucket.org/gtborg/gtsam`

3.2 3. Beitrag: Probabilistische Oberflächenrekonstruktion

Das Rekonstruieren von visuellen Merkmalen in Video-Bildern resultiert in dünn besetzten Punktwolken (siehe Abbildung 1 und 2). Ein Roboter hingegen braucht für die Manipulation, die Bewegungsplanung oder für das Ausweichen von Hindernissen eine dichte Repräsentation der Oberfläche (siehe Abbildung 3).

Das Ziel von früheren Arbeiten im Bereich der dichten Rekonstruktion von Oberflächen mittels Bildern ist meistens das Erzielen von möglichst hoher Genauigkeit [FPS13]. In der Robotik ist es hingegen sehr wichtig, dass man auch ein Mass für die Unsicherheit der Rekonstruktion schätzt, was als Mass für das Risiko bei der Bewegungsplanung oder für das optimale Fusionieren mit anderen Sensormodalitäten benutzt werden kann. Aus diesen Überlegungen entstand der dritte Beitrag dieser Dissertation: Ein Echtzeit Algorithmus für die probabilistische Rekonstruktion der Umgebung mittels einer einzelnen Kamera [PFS14]. Abbildung 3 zeigt exemplarisch die Rekonstruktion einer Oberfläche mittels dem vorgeschlagenen Algorithmus. Wichtig ist anzumerken, dass für jeden Oberflächenpunkt auch ein Mass der Unsicherheit geschätzt wird.

Als Demonstration wurde der Algorithmus auf einer Mikrodrohne implementiert, welche mit einer nach unten schauenden Kamera ausgestattet ist. Der Algorithmus rekonstruiert in Echtzeit die Oberfläche unterhalb der Drohne. In einem Experiment wurde gezeigt, wie diese Information es der Drohne erlaubt, autonom einen sicheren Landeplatz zu finden und daraufhin zu landen [Fo15b]. Ein Video von diesem Experiment is unter folgendem Link abrufbar: https://youtu.be/phaBKFwfcJ4.

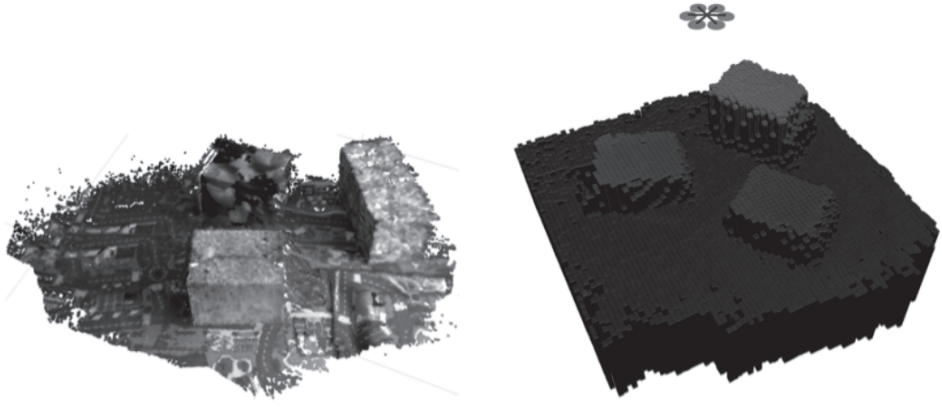

Abb. 3: Resultate des Projekts “Fliegender 3D Scanner”. Eine Drohne rekonstruiert die dreidimensionale Oberfläche mittels Bildinformationen einer nach unten schauenden Kamera. Links: Rekonstruktion der Oberfläche in Echtzeit auf einem leistungsstarken Grafikprozessor. Rechts: Rekonstruktion in Echtzeit auf dem Smartphone Prozessor der Drohne. Diese Rekonstruktion erlaubt es der Drohne, einen idealen Landeplatz zu finden.

3.3 4. Beitrag: Aktive Rekonstruktion mit einer Mikrodrohne

Eine spannende Anwendung von Computervision in der Robotik ist die Tatsache, dass der Roboter die Datenaufnahme beeinflussen kann. Daraus resultiert folgende Frage: Gegeben ist ein Bild der Umgebung; Was ist die optimale Trajektorie, welche eine Kamera durchlaufen muss, um möglichst schnell die Tiefe jedes Pixels im Referenzbild zu schätzen, d.h. die Oberfläche im Bild dreidimensional zu rekonstruieren? Diese Frage wird im letzten Beitrag dieser Dissertation untersucht. Dazu wurde eine Methode vorgeschlagen [FPS14a], um die Messungenauigkeit und dadurch den Informationsgewinn jeder Kameraposition zu berechnen. Die Innovation dieser Arbeit ist, dass für diese Berechnung nicht nur die relative Bewegung der Kamera und die geschätzte 3D Struktur der Oberfläche benutzt wird, sondern auch die Textur der Oberfläche im Referenzbild. Dies führt in Experimenten dazu, dass Roboter Trajektorien wählen, welche bildliche Doppeldeutigkeiten auflösen. Ein Beispiel davon ist in Abbildung 4 illustriert. Die Textur der Oberfläche in diesem Experiment weist eine dominante Gradientenrichtung auf. Wenn sich die Kamera in der Mitte befindet, dann ist der berechnete Informationsgewinn orthogonal zum Gradienten grösser als parallel dazu (im farblich kodierten Feld bedeutet Rot mehr Informationsgehalt als Grün oder Blau). Die Intuition dahinter ist, dass eine Bewegung parallel zum Gradienten die Tiefenschätzung erschwert, da Pixel entlang des Gradienten in den einzelnen Kamerabildern einander nicht eindeutig zugewiesen werden können.

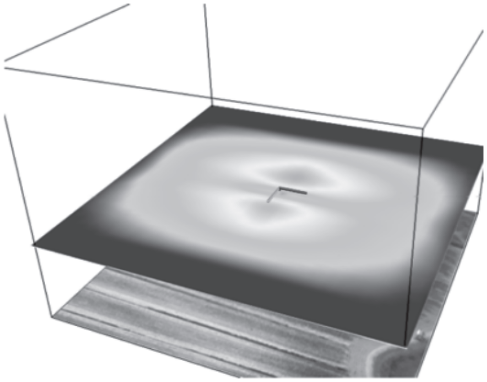

Abb. 4: Informationsgewinn für die Rekonstruktion einer Oberfläche als Funktion der Bewegungsrichtung der Kamera. Der Informationsgewinn in der Ebene horizontal zur aktuellen Kameraposition (in der Mitte) ist als Hitzekarte dargestellt. Wärmere Farben deuten auf einen höheren Informationsgewinn. Das Experiment zeigt, dass eine Bewegungsrichtung orthogonal zur dominanten Gradientenrichtung der Oberflächentextur optimal für die Rekonstruction ist.

Literaturverzeichnis

[Fa15] Faessler, M.; Fontana, F.; Forster, C.; Mueggler, E.; Pizzoli, M.; Scaramuzza, D.: Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor MAV. J. of Field Robotics, S. 1556–4967, 2015.

[Fo12] Forster, C.; Lynen, S.; Kneip, L.; Siegwart, R.: Centralized Multi-Robot Monocular SLAM. In: Robotics: Science and Systems (RSS). Workshop on Integration of Perception with Control and Navigation for Resource-limited, Highly dynamic, Autonomous Systems. 2012.

[Fo13] Forster, C.; Lynen, S.; Kneip, L.; Scaramuzza, D.: Collaborative Monocular SLAM with Multiple Micro Aerial Vehicles. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). S. 3962–3970, 2013.

[Fo15a] Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D.: IMU Preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation. In: Robotics: Science and Systems (RSS). 2015.

[Fo15b] Forster, C.; Faessler, M.; Fontana, F.; Werlberger, M.; Scaramuzza, D.: Continuous On-Board Monocular-Vision–based Aerial Elevation Mapping for Quadrotor Landing. In: IEEE Int. Conf. on Robotics and Automation (ICRA). S. 111–118, 2015.

[Fo16] Forster, C.: Visual Inertial Odometry and Active Dense Reconstruction for Mobile Robots. Dissertation, University of Zurich, April 2016.

[Fo17a] Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D.: On-Manifold Preintegration Theory for Fast and Accurate Visual-Inertial Navigation. IEEE Transactions on Robotics, Februar 2017.

[Fo17b] Forster, C.; Zhang, Z.; Gassner, M.; Werlberger, M.; Scaramuzza, D.: SVO 2.0: Semi-Direct Visual Odometry for Monocular and Multi-Camera Systems. IEEE Transactions on Robotics, April 2017.

[FPS13] Forster, C.; Pizzoli, M.; Scaramuzza, D.: Air-Ground Localization and Map Augmentation Using Monocular Dense Reconstruction. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). S. 3971–3978, 2013.

[FPS14a] Forster, C.; Pizzoli, M.; Scaramuzza, C.: Appearance-based Active, Monocular, Dense Depth Estimation for Micro Aerial Vehicles. In: Robotics: Science and Systems (RSS). 2014.

[FPS14b] Forster, C.; Pizzoli, M.; Scaramuzza, D.: SVO: Fast Semi-Direct Monocular Visual Odometry. In: IEEE Int. Conf. on Robotics and Automation (ICRA). S. 15–22, 2014.

[Gi15] Giusti, A.; Guzzi, J.; Ciresan, D.; He, F. Lin; Rodriguez, J. P.; Fontana, F.; Faessler, M.; Forster, C.; Schmidhuber, J.; Caro, G. A. Di; Scaramuzza, D.; Gambardella, L.: A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots. IEEE Robotics and Automation Letters, 2015.

[PFS14] Pizzoli, M.; Forster, C.; Scaramuzza, D.: REMODE: Probabilistic, Monocular Dense Reconstruction in Real Time. In: IEEE Int. Conf. on Robotics and Automation (ICRA). S. 2609–2616, 2014.

[Zh16] Zhang, Z.; Rebecq, H.; Forster, C.; Scaramuzza, D: Benefit of Large Field-of-View Cameras for Visual Odometry. In: IEEE Int. Conf. on Robotics and Automation (ICRA). 2016.

Christian Forster (1986, Schweizer) studierte Maschinenbau im Bachelor Studiengang an der ETH Zürich (2009), Robotik mit Auszeichnung im Masterstudiengang an der ETH Zürich (2012) und promovierte 2016 mit Auszeichnung bei Prof. Davide Scaramuzza mit der Dissertation *Visual Inertial Odometry and Active Dense Reconstruction for Mobile Robots* an der Universität Zürich. Im Jahr 2011 war er Gastforscher am CSIR in Südafrika und im Jahr 2014 war er Gast am Georgia Institute of Technology in der Gruppe von Prof. Frank Dellaert. Im Jahr 2015 war er Mitgründer des Zurich Eye Projekts am Wyss Institut Zürich. Zurich Eye hat sich zum Ziel gesetzt einen Sensor zu entwickeln der es Maschinen erlaubt sich im Inneren von Räumen Millimeter genau zu lokalisieren. Das Projektteam von Zurich Eye wurde Ende 2016 von Oculus (eine Untergruppe von Facebook) übernommen. Seit 2016 arbeitet Christian Forster bei Oculus an Computer Vision Systemen für die virtuelle Realität.