



Technische Hochschule
Ingolstadt
Fakultät Elektro-
und Informationstechnik

Semantische Segmentierung der Umgebung auf Basis von 3D-Daten

Simon Kuhn

Wissenschaftliche Arbeit im Zuge des Fachwissenschaftlichen Seminares

Erstprüfer: Prof. Dr. Christian Pfitzner

Betreuer: Prof. Dr. Christian Pfitzner

Ausgabedatum: 23.03.2023

Abgabedatum: 31.08.2023

Inhaltsverzeichnis

1	Einleitung	4
1.1	Hintergrund und Motivation	4
1.2	Problemstellung	4
2	Sensoren zur Erfassung von 3D-Daten	5
2.1	LiDAR-Sensoren	5
2.2	Stereo-Kameras und Tiefenkameras	5
2.3	Passive und aktive Sensoren	6
2.4	Auswahl von Sensoren für die semantische Segmentierung	6
3	Datengrundlage und Vorverarbeitung	8
3.1	3D-Datenformate und Datentypen	8
3.2	Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling .	8
3.3	Datenannotation und Ground Truth-Erstellung	8
4	Grundlagen der semantischen Segmentierung	10
4.1	Definition und Bedeutung	10
4.2	Methoden und Techniken	10
4.3	Herausforderungen und Limitationen	10
5	Anwendungszzenarien der semantischen Segmentierung	11
5.1	Autonomes Fahren	11
5.2	Robotik in der Industrie	11
5.3	Augmented Reality	11
5.4	Stadtplanung	11
5.5	Umweltüberwachung	11
6	State-of-the-Art Methoden zur semantischen Segmentierung auf Basis von 3D-Daten	12
6.1	Überblick über aktuelle Forschung und Entwicklungen	12
6.2	Vorstellung ausgewählter Methoden und deren Funktionsweise	12
7	Bewertung und Vergleich von Methoden zur semantischen Segmentierung	13
7.1	Evaluationsmetriken und -verfahren	13
7.2	Vergleich von Methoden anhand von Leistungskriterien	13
8	Herausforderungen und zukünftige Entwicklungen	14
8.1	Herausforderungen bei der semantischen Segmentierung von 3D-Daten	14
8.2	Potenziale und Trends für zukünftige Entwicklungen	14

9 Zusammenfassung und Anwendung	15
9.1 Zusammenfassung der Arbeit	15
9.2 Ausblick auf zukünftige Forschungsrichtungen	15
10 Literaturverzeichnis	16

1 Einleitung

1.1 Hintergrund und Motivation

In den letzten Jahren hat die Forschung im Bereich der autonomen Fahrzeuge und der Robotik enorme Fortschritte gemacht. Ein wichtiger Faktor für die Entwicklung dieser Technologien ist die Fähigkeit, die Umgebung ausreichend genau zu erkennen und zu verstehen. In diesem Zusammenhang hat die semantische Segmentierung der Umgebung auf Basis von 3D-Daten eine immer größere Bedeutung erlangt. Die semantische Segmentierung ist ein Verfahren zur automatischen Klassifizierung von Objekten und Strukturen in der Umgebung. Dabei werden jedem Pixel oder jedem Voxel in einem 3D-Modell eine bestimmte semantische Bedeutung zugeordnet, z.B. Straße, Gebäude, Bäume oder Fahrzeuge. Eine präzise semantische Segmentierung ist eine wesentliche Voraussetzung für eine zuverlässige Navigation von autonomen Fahrzeugen und Robotern. In dieser Arbeit wird die semantische Segmentierung der Umgebung auf Basis von 3D-Daten untersucht. Dabei sollen verschiedene Methoden zur semantischen Segmentierung untersucht und bewertet werden. Ziel ist es, die Genauigkeit und Zuverlässigkeit der semantischen Segmentierung zu verbessern und die Anwendungsmöglichkeiten in der Praxis zu erweitern.

1.2 Problemstellung

Obwohl die semantische Segmentierung der Umgebung auf Basis von 3D-Daten in den letzten Jahren deutliche Fortschritte gemacht hat, gibt es noch viele offene Fragen und Herausforderungen. Zum Beispiel sind viele der bestehenden Methoden für die semantische Segmentierung nur bedingt skalierbar und erfordern viel Rechenleistung. Zudem sind sie oft sehr empfindlich gegenüber Veränderungen in der Umgebung, wie z.B. Änderungen der Lichtverhältnisse oder der Perspektive[1]. Um diese Herausforderungen zu meistern, ist es notwendig, neue Methoden und Technologien zu entwickeln, die eine robuste und skalierbare semantische Segmentierung ermöglichen. Diese Arbeit trägt dazu bei, indem sie verschiedene Methoden und Technologien für die semantische Segmentierung der Umgebung auf Basis von 3D-Daten untersucht und bewertet[2]

2 Sensoren zur Erfassung von 3D-Daten

2.1 LiDAR-Sensoren

LiDAR-Sensoren, auch bekannt als Light Detection and Ranging-Sensoren, sind eine weit verbreitete Technologie zur Erfassung von 3D-Daten. Sie arbeiten auf Basis von Laserstrahlen, die ausgesendet und von Objekten in der Umgebung reflektiert werden. Durch die Messung der Laufzeit des Laserstrahls können LiDAR-Sensoren die Entfernungen zu den reflektierenden Objekten hochpräzise erfassen. LiDAR-Sensoren können sowohl in der Luft als auch am Boden eingesetzt werden und ermöglichen die Erfassung von detaillierten 3D-Punktwolken, die die Geometrie und die räumliche Verteilung von Objekten in der Umgebung repräsentieren. LiDAR-Sensoren bieten eine hohe Genauigkeit und Auflösung, was sie für die semantische Segmentierung von 3D-Daten besonders geeignet macht. In diesem Abschnitt werden verschiedene Typen von LiDAR-Sensoren vorgestellt, ihre Funktionsweise erläutert und ihre Vor- und Nachteile diskutiert. Zudem werden Auswahlkriterien für die Verwendung von LiDAR-Sensoren in semantischen Segmentierungssystemen behandelt.

2.2 Stereo-Kameras und Tiefenkameras

Moderne Sensortechnologien wie Stereo-Kameras und Tiefenkameras sind ebenfalls von großer Bedeutung für die Erfassung von 3D-Daten zur semantischen Segmentierung der Umgebung. Stereo-Kameras verwenden zwei räumlich getrennte Kameras, die gemeinsam Bilder von derselben Szene aufnehmen. Durch die Berechnung von disparitätsbasierten Tiefeninformationen aus den Unterschieden zwischen den Bildern können Stereo-Kameras die räumliche Tiefe von Objekten in der Umgebung schätzen. Tiefenkameras hingegen verwenden spezielle Sensoren wie Time-of-Flight (ToF)- oder Structured-Light (SL)-Sensoren, um Tiefeninformationen direkt zu erfassen.

Stereo-Kameras und Tiefenkameras bieten einige Vorteile wie ihre vergleichsweise geringen Kosten und ihre kompakte Bauweise, die sie für viele Anwendungen attraktiv machen. Sie können in Echtzeit arbeiten und liefern hochauflösende 3D-Daten, die für die semantische Segmentierung verwendet werden können. Allerdings haben sie auch einige Einschränkungen, wie zum Beispiel die Empfindlichkeit gegenüber Beleuchtungsbedingungen, die Sichtbarkeit von Textur und die Reichweite der Tiefenmessung. Diese Einschränkungen müssen bei der Auswahl und Verwendung von Stereo-Kameras und Tiefenkameras in semantischen Segmentierungssystemen berücksichtigt werden.

In diesem Abschnitt werden Stereo-Kameras und Tiefenkameras als wichtige Sensortechnologien zur Erfassung von 3D-Daten für die semantische Segmentierung behandelt. Es werden ihre Funktionsweise, Vor- und Nachteile sowie Anwendungsbereiche diskutiert. Zudem werden Verfahren zur Vorverarbeitung der von diesen Sensoren erfassten Daten vorgestellt, um sie für die semantische Segmentierung vorzubereiten.

2.3 Passive und aktive Sensoren

Moderne Sensortechnologien wie Lidar, Stereo-Kameras und Tiefenkameras spielen eine entscheidende Rolle bei der Erfassung von 3D-Daten für die semantische Segmentierung der Umgebung. Dabei können diese Sensoren in passive und aktive Sensoren unterteilt werden.

Passive Sensoren, wie zum Beispiel Stereo-Kameras und Tiefenkameras, erfassen die Umgebung, indem sie das von natürlichen oder künstlichen Lichtquellen reflektierte Licht messen. Sie nutzen dabei die Eigenschaften des einfallenden Lichts, um Informationen über die räumliche Tiefe von Objekten in der Szene zu berechnen. Diese Sensoren sind häufig kostengünstig, kompakt und können in Echtzeit arbeiten. Allerdings sind sie auch empfindlich gegenüber Beleuchtungsbedingungen und können Einschränkungen in der Reichweite und Sichtbarkeit von Textur aufweisen.

Im Gegensatz dazu verwenden aktive Sensoren, wie zum Beispiel Lidar (Light Detection and Ranging), eine eigene Lichtquelle, um die Umgebung zu erfassen. Lidar-Sensoren senden Laserstrahlen aus und messen die Zeit, die benötigt wird, um die reflektierten Strahlen zurückzuerhalten. Dadurch können sie präzise Tiefeninformationen mit hoher Genauigkeit und Reichweite erfassen, unabhängig von den Beleuchtungsbedingungen und Textur der Umgebung. Allerdings sind Lidar-Sensoren in der Regel teurer und können größer und schwerer sein als passive Sensoren.

In diesem Abschnitt werden passive und aktive Sensoren als wichtige Sensortechnologien zur Erfassung von 3D-Daten für die semantische Segmentierung behandelt. Es werden ihre Funktionsweisen, Vor- und Nachteile sowie Anwendungsbereiche diskutiert. Zudem werden Verfahren zur Vorverarbeitung der von diesen Sensoren erfassten Daten vorgestellt, um sie für die semantische Segmentierung vorzubereiten.

2.4 Auswahl von Sensoren für die semantische Segmentierung

Die Auswahl der geeigneten Sensoren für die semantische Segmentierung hängt von verschiedenen Faktoren ab, wie den Anforderungen der spezifischen Anwendung, den Umgebungsbedingungen, dem Budget und den gewünschten Ergebnissen. In diesem Abschnitt werden Kriterien und Überlegungen zur Auswahl von Sensoren für die semantische Segmentierung auf Basis von 3D-Daten diskutiert.

Anforderungen der Anwendung: Die Anforderungen der spezifischen Anwendung, in der die semantische Segmentierung durchgeführt werden soll, sind entscheidend für die Auswahl der geeigneten Sensoren. Hierbei können Aspekte wie die benötigte Genauigkeit, die räumliche Auflösung, die Reichweite, die Echtzeitfähigkeit und die Umgebungsbedingungen eine Rolle spielen. Zum Beispiel erfordern Anwendungen im Bereich der autonomen Fahrzeuge möglicherweise Sensoren mit hoher Reichweite und Genauigkeit, während Anwendungen im Innenbereich möglicherweise Sensoren mit höherer räumlicher Auflösung und Echtzeitfähigkeit benötigen.

Umgebungsbedingungen: Die Umgebungsbedingungen, in denen die semantische Segmentierung durchgeführt werden soll, können die Auswahl der Sensoren beeinflussen. Beispielsweise können schlechte Beleuchtungsbedingungen, wie Dunkelheit oder Blendung durch Sonnenlicht, die Leistung von passiven Sensoren wie Kameras beeinträchtigen, während Lidar-Sensoren unabhängig von den Beleuchtungsbedingungen arbeiten können. Ebenso können Umgebungen mit vielen Hindernissen oder komplexen Geometrien die Leistung von Sensoren beeinflussen und die Wahl von geeigneten Sensoren beeinflussen.

Budget: Das Budget ist ein wichtiger Faktor bei der Auswahl von Sensoren. Unterschiedliche Sensoren können unterschiedliche Kosten haben, sowohl in der Anschaffung als auch in der Wartung. Lidar-Sensoren sind in der Regel teurer als Kameras oder Tiefenkameras, aber auch die Kosten für diese Sensoren haben sich in den letzten Jahren reduziert. Bei der Auswahl von Sensoren ist es wichtig, das Budget im Auge zu behalten und eine sorgfältige Kosten-Nutzen-Analyse durchzuführen.

Gewünschte Ergebnisse: Die gewünschten Ergebnisse der semantischen Segmentierung können ebenfalls die Auswahl der Sensoren beeinflussen. Je nachdem, welche Arten von Objekten oder Strukturen in der Umgebung segmentiert werden sollen, können bestimmte Sensoren besser geeignet sein als andere. Zum Beispiel können Lidar-Sensoren aufgrund ihrer präzisen Tiefeninformationen und Reichweite gut geeignet sein, um Objekte wie Straßen, Gebäude oder Bäume zu segmentieren, während Kameras oder Tiefenkameras besser für die Segmentierung von Fußgängern oder Fahrzeugen geeignet sein können.

In diesem Abschnitt werden Kriterien und Überlegungen zur Auswahl von Sensoren für die semantische Segmentierung auf Basis von 3D-Daten erläutert. Es werden verschiedene Aspekte wie Anforderungen der Anwendung, Umgebungsbedingungen, Budget und gewünschte Ergebnisse diskutiert.

3 Datengrundlage und Vorverarbeitung

In diesem Kapitel werden die Grundlagen der Datengrundlage und Vorverarbeitung von 3D-Daten für die semantische Segmentierung erläutert. Es werden verschiedene 3D-Datenformate und Datentypen vorgestellt sowie Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling diskutiert. Zudem wird auf die Bedeutung von Datenannotation und Ground Truth-Erstellung für die Semantische Segmentierung eingegangen.

3.1 3D-Datenformate und Datentypen

Es gibt verschiedene 3D-Datenformate, die für die Verarbeitung von 3D-Daten in der Semantischen Segmentierung verwendet werden können. Dazu gehören beispielsweise das STL-Format, das OBJ-Format oder das PLY-Format. Jedes Format hat seine spezifischen Eigenschaften und wird für bestimmte Anwendungen bevorzugt eingesetzt. Darüber hinaus können 3D-Daten in verschiedenen Datentypen vorliegen, wie zum Beispiel als Punktwolken oder als Mesh-Modelle. Die Wahl des Datentyps hängt von der Art der zu segmentierenden Objekte ab und kann auch Auswirkungen auf die Wahl der Algorithmen für die semantische Segmentierung haben.

3.2 Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling

Vor der semantischen Segmentierung ist oft eine Vorverarbeitung der 3D-Daten erforderlich, um eine höhere Qualität der Daten zu erreichen und unerwünschte Informationen zu entfernen. Ein wichtiger Vorverarbeitungsschritt ist die Filterung, die zur Entfernung von Rauschen und zur Verbesserung der Datenqualität eingesetzt wird. Ein weiterer wichtiger Schritt ist die Normalenberechnung, die zur Bestimmung der Ausrichtung der Flächen in den 3D-Daten dient. Diese Information kann für die semantische Segmentierung verwendet werden, um eine bessere Unterscheidung zwischen verschiedenen Objekten zu erreichen. Zusätzlich kann auch das Downsampling eingesetzt werden, um die Anzahl der Datenpunkte zu reduzieren und die Verarbeitungszeit zu verkürzen.

3.3 Datenannotation und Ground Truth-Erstellung

Für eine erfolgreiche semantische Segmentierung ist eine genaue Annotation der 3D-Daten und die Erstellung von Ground Truth-Labels von entscheidender Bedeutung. Datenannotation bezieht sich auf den Prozess, bei dem bestimmte Merkmale oder Eigenschaften von 3D-Daten manuell markiert oder etikettiert werden. Dieser Prozess ist notwendig, um eine Referenzgrundlage für den Trainingsprozess von Algorithmen zu schaffen. Die Ground Truth-Erstellung bezieht sich

auf die manuelle Zuordnung von semantischen Labels zu den markierten Objekten in den 3D-Daten. Diese Labels dienen als Referenz für die spätere semantische Segmentierung von neuen, unannotierten Daten.

Insgesamt ist die Datenvorverarbeitung ein kritischer Schritt für die semantische Segmentierung von 3D-Daten. Die Wahl des richtigen Datenformats und Datentyps, sowie die Auswahl geeigneter Vorverarbeitungsschritte und eine präzise Datenannotation und Ground

4 Grundlagen der semantischen Segmentierung

4.1 Definition und Bedeutung

4.2 Methoden und Techniken

4.3 Herausforderungen und Limitationen

5 Anwendungsszenarien der semantischen Segmentierung

5.1 Autonomes Fahren

5.2 Robotik in der Industrie

5.3 Augmented Reality

5.4 Stadtplanung

5.5 Umweltüberwachung

6 State-of-the-Art Methoden zur semantischen Segmentierung auf Basis von 3D-Daten

6.1 Überblick über aktuelle Forschung und Entwicklungen

6.2 Vorstellung ausgewählter Methoden und deren Funktionsweise

7 Bewertung und Vergleich von Methoden zur semantischen Segmentierung

7.1 Evaluationsmetriken und -verfahren

7.2 Vergleich von Methoden anhand von Leistungskriterien

8 Herausforderungen und zukünftige Entwicklungen

8.1 Herausforderungen bei der semantischen Segmentierung von 3D-Daten

8.2 Potenziale und Trends für zukünftige Entwicklungen

9 Zusammenfassung und Anwendung

9.1 Zusammenfassung der Arbeit

9.2 Ausblick auf zukünftige Forschungsrichtungen

10 Literaturverzeichnis

Literatur

- [1] Seung-Jun Shin, Suk-Hwan Suh und Ian Stroud. "Reincarnation of G-code based part programs into STEP-NC for turning applications". In: *Computer-Aided Design* 39.1 (2007), S. 1–16.
- [2] *Have I Been Pwned: Check if your email has been compromised in a data breach*. 2023. URL: <https://haveibeenpwned.com/>.