



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

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Hintergrund und Motivation . . . . .	4
1.2	Problemstellung . . . . .	4
1.3	Zielsetzung . . . . .	4
<b>2</b>	<b>Sensoren zur Erfassung von 3D-Daten</b>	<b>6</b>
2.1	LiDAR-Sensoren . . . . .	6
2.2	Stereo-Kameras und Tiefenkameras . . . . .	6
2.3	Passive und aktive Sensoren . . . . .	7
2.4	Auswahl von Sensoren für die semantische Segmentierung . . . . .	7
<b>3</b>	<b>Datengrundlage und Vorverarbeitung</b>	<b>9</b>
3.1	3D-Datenformate und Datentypen . . . . .	9
3.2	Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling . . . . .	9
3.3	Datenannotation und Ground Truth-Erstellung . . . . .	9
<b>4</b>	<b>Grundlagen der semantischen Segmentierung</b>	<b>10</b>
4.1	Definition und Bedeutung . . . . .	10
4.2	Methoden und Techniken . . . . .	10
4.3	Herausforderungen und Limitationen . . . . .	10
<b>5</b>	<b>Anwendungszzenarien der semantischen Segmentierung</b>	<b>11</b>
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
<b>6</b>	<b>State-of-the-Art Methoden zur semantischen Segmentierung auf Basis von 3D-Daten</b>	<b>12</b>
6.1	Überblick über aktuelle Forschung und Entwicklungen . . . . .	12
6.2	Vorstellung ausgewählter Methoden und deren Funktionsweise . . . . .	12
<b>7</b>	<b>Bewertung und Vergleich von Methoden zur semantischen Segmentierung</b>	<b>13</b>
7.1	Evaluationsmetriken und -verfahren . . . . .	13
7.2	Vergleich von Methoden anhand von Leistungskriterien . . . . .	13
<b>8</b>	<b>Herausforderungen und zukünftige Entwicklungen</b>	<b>14</b>
8.1	Herausforderungen bei der semantischen Segmentierung von 3D-Daten . . . . .	14
8.2	Potenziale und Trends für zukünftige Entwicklungen . . . . .	14

<b>9 Zusammenfassung und Anwendung</b>	<b>15</b>
9.1 Zusammenfassung der Arbeit . . . . .	15
9.2 Ausblick auf zukünftige Forschungsrichtungen . . . . .	15
<b>10 Literaturverzeichnis</b>	<b>16</b>

# 1 Einleitung

## 1.1 Hintergrund und Motivation

In den letzten Jahren hat die Forschung im Bereich der autonomen Fahrzeuge und der Robotik einen enormen Fortschritt 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 jeder 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.

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.

## 1.3 Zielsetzung

Das übergeordnete Ziel dieser Arbeit ist es, die semantische Segmentierung der Umgebung auf Basis von 3D-Daten zu verbessern und ihre Anwendungsmöglichkeiten in der Praxis zu erweitern. Konkret sollen folgende Ziele erreicht werden:

verschiedener Methoden und Technologien für die semantische Segmentierung von 3D-Daten  
Bewertung der Genauigkeit und Zuverlässigkeit der verschiedenen Methoden  
Identifikation von Herausforderungen und Limitationen der semantischen Segmentierung auf Basis von 3D-Daten

Entwicklung von Empfehlungen für zukünftige Forschung und Anwendungen D. Forschungsfragen

## 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**

### **3.1 3D-Datenformate und Datentypen**

### **3.2 Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling**

### **3.3 Datenannotation und Ground Truth-Erstellung**

## **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