



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
1.3	Zielsetzung	4
2	Sensoren zur Erfassung von 3D-Daten	6
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
3	Datengrundlage und Vorverarbeitung	9
3.1	3D-Datenformate und Datentypen	9
3.2	Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling .	9
3.3	Datenannotation und Ground Truth-Erstellung	10
4	Grundlagen der semantischen Segmentierung	12
4.1	Definition und Bedeutung	12
4.2	Methoden und Techniken	12
4.3	Herausforderungen und Limitationen	12
5	Anwendungszszenarien der semantischen Segmentierung	13
5.1	Autonomes Fahren	13
5.2	Robotik in der Industrie	13
5.3	Augmented Reality	13
5.4	Stadtplanung	13
5.5	Umweltüberwachung	13
6	State-of-the-Art Methoden zur semantischen Segmentierung auf Basis von 3D-Daten	14
6.1	Überblick über aktuelle Forschung und Entwicklungen	14
6.2	Vorstellung ausgewählter Methoden und deren Funktionsweise	14
7	Bewertung und Vergleich von Methoden zur semantischen Segmentierung	15
7.1	Evaluationsmetriken und -verfahren	15
7.2	Vergleich von Methoden anhand von Leistungskriterien	15
8	Herausforderungen und zukünftige Entwicklungen	16
8.1	Herausforderungen bei der semantischen Segmentierung von 3D-Daten	16
8.2	Potenziale und Trends für zukünftige Entwicklungen	16

9 Zusammenfassung und Anwendung	17
9.1 Zusammenfassung der Arbeit	17
9.2 Ausblick auf zukünftige Forschungsrichtungen	17
10 Literaturverzeichnis	18

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

Die Auswahl von Sensoren für die semantische Segmentierung erfordert eine sorgfältige Betrachtung der verwendeten 3D-Datenformate und Datentypen. Diese bilden die Grundlage für die Erfassung, Speicherung und Verarbeitung von 3D-Daten, die für die semantische Segmentierung verwendet werden. In diesem Kapitel werden verschiedene 3D-Datenformate und Datentypen untersucht, die in der Forschung und Praxis weit verbreitet sind.

Ein wichtiges 3D-Datenformat ist das Punktwolkenformat, das häufig von LiDAR-Sensoren erzeugt wird. Punktwolken sind Sammlungen von 3D-Punkten, die die Oberfläche von Objekten in der Umgebung darstellen. Sie können in verschiedenen Dateiformaten gespeichert werden, wie beispielsweise dem ASCII-Format oder dem binären LAS-Format (LASer File Format), das speziell für LiDAR-Daten entwickelt wurde. Diese Formate ermöglichen die Speicherung von großen Mengen an Punkten mit 3D-Koordinaten, Intensitätsinformationen und weiteren Attributen, die zur semantischen Segmentierung verwendet werden können.

Neben Punktwolken werden auch 3D-Gitter oder Voxel-Daten oft für die semantische Segmentierung verwendet. Voxel sind volumetrische Elemente, die den Raum in einem dreidimensionalen Gitter unterteilen. Jeder Voxel kann eine semantische Bedeutung, wie zum Beispiel Straße, Gebäude oder Vegetation, zugeordnet werden. Voxel-Daten können in verschiedenen Formaten gespeichert werden, wie zum Beispiel das binäre OctoMap-Format oder das ASCII-Format.

Darüber hinaus können auch 3D-Modelle, wie Polygonnetze oder Punktwolken mit Texturinformationen, für die semantische Segmentierung verwendet werden. Polygonnetze sind Sammlungen von 3D-Polygonen, die die Oberfläche von Objekten repräsentieren. Punktwolken mit Texturinformationen enthalten zusätzlich zu den 3D-Koordinaten auch Farbinformationen, die aus Bildern oder Kameras gewonnen werden. Diese 3D-Datenformate ermöglichen eine detailliertere Darstellung von Objekten und können für die semantische Segmentierung von komplexeren Szenen verwendet werden.

Die Wahl des richtigen 3D-Datenformats und Datentyps hängt von den spezifischen Anforderungen der semantischen Segmentierung ab, wie zum Beispiel der benötigten Genauigkeit, der Verarbeitungsgeschwindigkeit, dem Speicherbedarf und der Verfügbarkeit von Sensordaten. Es ist wichtig, die Vor- und Nachteile der verschiedenen Formate und Typen zu berücksichtigen und das geeignete Format für die spezifische Anwendung auszuwählen, um eine effektive und effiziente semantische Segmentierung der Umgebung zu ermöglichen.

3.2 Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling

Die Vorverarbeitung von 3D-Daten ist ein wichtiger Schritt in der semantischen Segmentierung, um die Qualität und Genauigkeit der Segmentierungsergebnisse zu verbessern. In diesem

Kapitel werden verschiedene Vorverarbeitungsschritte wie Filterung, Normalenberechnung und Downsampling betrachtet, die oft in der Praxis angewendet werden.

Filterung von 3D-Daten: Die Filterung von 3D-Daten beinhaltet das Entfernen von unerwünschtem Rauschen oder Ausreißern, um die Qualität der Daten zu verbessern. Dies kann durch verschiedene Filtertechniken erfolgen, wie zum Beispiel Medianfilter, Gaußscher Filter oder Bilateralfilter. Diese Filter können angewendet werden, um Rauschen oder Ausreißer in den 3D-Punktwolken oder Voxel-Daten zu reduzieren und somit die Qualität der Daten für die semantische Segmentierung zu verbessern.

Normalenberechnung: Die Berechnung von Normalen ist ein wichtiger Schritt, um die geometrische Information der 3D-Daten zu erfassen. Normalen sind Vektoren, die senkrecht zur Oberfläche von Objekten in der Umgebung stehen und die Orientierung der Oberfläche angeben. Die Normalenberechnung kann auf Basis von Punktwolken oder Polygonnetzen durchgeführt werden und ermöglicht es, die Richtung und Orientierung der Objektoberflächen zu erfassen, was für die semantische Segmentierung von Objekten von Bedeutung ist.

Downsampling: Das Downsampling von 3D-Daten beinhaltet die Reduzierung der Datenmenge, um die Verarbeitungsgeschwindigkeit und den Speicherbedarf zu reduzieren. Dies kann durch verschiedene Techniken wie das Subsampling von Punktwolken oder das Voxel-Downsampling bei Voxel-Daten erfolgen. Downsampling ermöglicht es, die Datenmenge zu reduzieren, während wichtige strukturelle und semantische Informationen erhalten bleiben.

Die Anwendung von Vorverarbeitungsschritten wie Filterung, Normalenberechnung und Downsampling ist von großer Bedeutung, um die Qualität und Genauigkeit der 3D-Daten für die semantische Segmentierung zu verbessern. Durch die Reduzierung von Rauschen, die Erfassung von Normalen und die Reduzierung der Datenmenge können genauere und effizientere Segmentierungsergebnisse erzielt werden, die die Grundlage für die nachfolgenden Schritte der semantischen Segmentierung bilden. Es ist wichtig, geeignete Vorverarbeitungsschritte entsprechend den spezifischen Anforderungen der Anwendung und der verwendeten 3D-Daten auszuwählen und anzuwenden, um eine effektive und präzise semantische Segmentierung zu gewährleisten.

3.3 Datenannotation und Ground Truth-Erstellung

Die Datenannotation und die Erstellung einer Ground Truth sind entscheidende Schritte in der semantischen Segmentierung, um die Trainingsdaten für maschinelles Lernen bereitzustellen und Modelle für die Segmentierung von 3D-Daten zu trainieren und zu evaluieren. In diesem Kapitel werden verschiedene Aspekte der Datenannotation und der Ground Truth-Erstellung betrachtet.

Datenannotation: Die Datenannotation beinhaltet das manuelle oder automatische Hinzufügen von semantischen Labels oder Klasseninformationen zu den 3D-Daten. Dies kann durch das Markieren von Regionen oder Objekten in den Punktwolken oder Voxel-Daten erfolgen, um sie bestimmten Klassen oder Kategorien zuzuordnen. Die Datenannotation kann von menschlichen Annotatoren durchgeführt werden oder mit Hilfe von automatisierten Algorithmen, die auf maschinellem Lernen oder Regelbasierten Methoden basieren. Die Qualität und Genauigkeit der Datenannotation sind von entscheidender Bedeutung für die Qualität und Leistungsfähigkeit der semantischen Segmentierungsalgorithmen.

Ground Truth-Erstellung: Die Ground Truth-Erstellung beinhaltet die Erstellung von referenzbasierten Segmentierungsergebnissen, die als Grundlage für das Training und die Evaluierung von semantischen Segmentierungsalgorithmen dienen. Die Ground Truth kann manuell oder au-

tomatisch erstellt werden, indem die annotierten Daten als Referenz verwendet werden, um die Leistung von Segmentierungsalgorithmen zu bewerten. Die Ground Truth-Erstellung ist ein kritischer Schritt, um die Zuverlässigkeit und Vergleichbarkeit von Segmentierungsergebnissen zu gewährleisten und die Qualität von trainierten Modellen zu überprüfen.

Herausforderungen bei der Datenannotation und Ground Truth-Erstellung: Die Datenannotation und Ground Truth-Erstellung können mit verschiedenen Herausforderungen verbunden sein, wie zum Beispiel Unsicherheiten in der Klassen- oder Objektabgrenzung, Unvollständigkeit oder Inkonsistenz in den Annotationsdaten, Schwierigkeiten bei der Handhabung großer Datenmengen und das Management von Zeitaufwand und Ressourcen. Es ist wichtig, diese Herausforderungen zu berücksichtigen und geeignete Methoden und Werkzeuge einzusetzen, um die Qualität und Genauigkeit der Datenannotation und Ground Truth-Erstellung zu gewährleisten.

Die Datenannotation und Ground Truth-Erstellung sind kritische Schritte in der semantischen Segmentierung von 3D-Daten, um Trainingsdaten bereitzustellen, Modelle zu trainieren und zu evaluieren. Die Qualität, Genauigkeit und Vergleichbarkeit der Datenannotation und Ground Truth-Erstellung sind von großer Bedeutung, um zuverlässige und präzise semantische Segmentierungsergebnisse zu erzielen und die Leistungsfähigkeit von Segmentierungsalgorithmen zu bewerten. Es ist wichtig, geeignete Methoden und Werkzeuge einzusetzen, um den Herausforderungen bei der Datenannotation und Ground Truth-Erstellung erfolgreich zu begegnen und die Qualität der verwendeten Trainingsdaten zu gewährleisten.

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