

B

C

Rekonstruktion von Meshes für industrielles Bin-Picking und Depalettieren

S

BACHELORARBEIT

zur Erlangung des akademischen Grades

Bachelor of Science (B. Sc.)

an der

Hochschule Konstanz
Technik, Wirtschaft und Gestaltung

Fakultät Informatik
Studiengang Angewandte Informatik

Thema: **Rekonstruktion von Meshes für industrielles Bin-Picking und Depalettieren**

Bachelorkandidat: Lorenz Bung, Banater Str. 9, 78467 Konstanz

1. Prüfer: Prof. Dr. Georg Umlauf
2. Prüfer: Simon Schmeißer, M.Sc.

Ausgabedatum: 01.05.2020
Abgabedatum: 31.07.2020

Ehrenwörtliche Erklärung

Hiermit erkläre ich, *Lorenz Bung*, geboren am 26.06.1997 in Konstanz, dass ich

- (1) meine Bachelorarbeit mit dem Titel

Rekonstruktion von Meshes für industrielles Bin-Picking und De-palettieren

bei der Isys Vision GmbH unter Anleitung von Prof. Dr. Georg Umlauf selbständig und ohne fremde Hilfe angefertigt und keine anderen als die angeführten Hilfen benutzt habe;

- (2) die Übernahme wörtlicher Zitate, von Tabellen, Zeichnungen, Bildern und Programmen aus der Literatur oder anderen Quellen (Internet) sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet habe;
- (3) dass die eingereichten Abgabe-Exemplare in Papierform und im PDF-Format vollständig übereinstimmen.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Konstanz, 31.07.2020

(Unterschrift)

Abstract

Thema:	Rekonstruktion von Meshes für industrielles Bin-Picking und Depalettieren
Bachelorkandidat:	Lorenz Bung
Betreuer:	Prof. Dr. Georg Umlauf Institut für Optische Systeme Simon Schmeißer, M.Sc. Isys Vision GmbH
Abgabedatum:	31.07.2020
Schlagworte:	Robotik, ROS, PCL, Punktwolke, Geometrisches Modellieren, Machine Vision, Meshrekonstruktion

Describe the objective and results of this thesis in a few words. Typically one page.

In dieser Arbeit werden verschiedene Ansätze zur Rekonstruktion von Meshes aus 3D-Punktwolken analysiert, miteinander verglichen und bewertet. Die erste Methode ist YAK, eine Variante von Kinect Fusion, bei der ein Truncated Signed Distance Field zum Einsatz kommt. Der zweite Ansatz ist, ein einzelnes Teil mithilfe einer beweglichen Kamera am Roboterarm aus mehreren Posen einzuscannen und so ein Gesamtbild zu erhalten. Im dritten Ansatz wird eine

Extended Abstract

Thema:	Rekonstruktion von Meshes für industrielles Bin-Picking und Depalettieren
Bachelorkandidat:	Lorenz Bung
Betreuer:	Prof. Dr. Georg Umlauf Institut für Optische Systeme Simon Schmeißer, M.Sc. Isys Vision GmbH
Abgabedatum:	31.07.2020
Schlagworte:	Robotik, ROS, PCL, Punktwolke, Geometrisches Modellieren, Machine Vision, Meshrekonstruktion

Extended Abstract über 2 Seiten. Beispielhafte Texte aus anderen Teamprojekten oder Abschlussarbeiten können aus dem verlinkten Dokument entnommen werden.

Dieser Text soll als Dokumentation des Teamprojekts für den zukünftigen Jahresbericht des Institut für Optische Systeme dienen. Gerne können auch Bilder eingefügt werden. Ebenso wichtig ist es auch die Referenzen aufzulisten wie z.B. [1].

Literatur

- [1] Radu Bogdan Rusu und Steve Cousins. "3d is here: Point cloud library (pcl)". In: *2011 IEEE international conference on robotics and automation*. IEEE. 2011, S. 1–4.

Danksagung

Ich möchte mich herzlich bei Simon Schmeißer von der Firma Isys Vision GmbH für die Betreuung bedanken - ohne ihn wäre diese Bachelorarbeit nicht möglich gewesen.

Auch allen weiteren Mitarbeitern von Isys Vision möchte ich für das Ermöglichen der Arbeit sowie das freundliche Arbeitsumfeld danken.

Weiterhin gilt mein Dank den Mitgliedern des Instituts für Optische Systeme, da sie mich bei meiner Tätigkeit am Institut maßgeblich auf diese Arbeit vorbereitet haben.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Mikado	1
1.2	Motivation	1
1.3	Zielsetzung	2
2	Grundlagen	3
2.1	Aufnahme und Speicherung von 3D-Bildern	3
2.1.1	Tiefenbild	4
2.1.2	Voxel Grid	5
2.1.3	Punktwolken	6
2.2	Meshrepräsentationen	8
2.2.1	Vertex-Vertex-Mesh	9
2.2.2	Winged Edge	9
2.2.3	Half Edge	11
2.3	Point Cloud Library	11
3	Vorhandene Arbeit	12
3.1	Registrierung von Punktwolken	12
3.1.1	Lokale Registrierung	12
3.1.2	Globale Registrierung	16
3.2	Segmentierung	17
3.3	Triangulierung	17
3.3.1	Marching Cubes	18
3.3.2	Poisson	19
3.3.3	Weitere Ansätze	19
4	Implementierung	20
5	Auswertung	21
6	Fazit	22
	Literatur	VII

Abkürzungsverzeichnis

PCL	Point Cloud Library
ROS	Robot Operating System
KinFu	Kinect Fusion
YAK	Yet Another Kinect Fusion
ICP	Iterative Closest Point
4PCS	Four-Points Congruent Sets
TSDF	Truncated Signed Distance Field
OpenGR	Open Global Registration
VVM	Vertex-Vertex-Mesh
LUT	Lookup Table

1. Einleitung

1.1 Mikado

Isys vision GmbH [1] ist ein Unternehmen aus Freiburg im Breisgau, das sich mit Systemintegration und industrieller Bildverarbeitung beschäftigt. Neben Entwicklungen in der 2D-Bildverarbeitung (zum Beispiel in der Leiterplattenproduktion) spielt auch Machine Vision im 3D-Bereich in Kombination mit Robotik eine große Rolle.

Mikado [2] ist ein Softwarepaket von isys vision, welches zur Robotersteuerung und 3D-Bildverarbeitung in der Industrie eingesetzt wird. Es besteht aus zwei Komponenten: Mikado 3D dient der eigentlichen Bildverarbeitung, während Mikado Adaptive Robot Control (ARC) zusätzlich dazu die Robotersteuerung und Kollisionsplanung beinhaltet. Die Hauptanwendung ist dabei das sogenannte "Bin-Picking", also das Greifen von sortenreinen Teilen aus einer unsortierten Kiste. Zur Erfassung der Bilddaten kommen 3D-Kameras von Ensenso [3] zum Einsatz.

1.2 Motivation

Bei der Bestimmung der 6D-Posen von Objekten wird bei Mikado unter Anderem der "surface based matching"-Algorithmus von MvTec Halcon [4] verwendet. Dafür wird ein CAD-Modell des Objekts benötigt, welches jedoch in vielen Fällen nicht vorhanden ist.

Die Gründe dafür sind vielfältig. Etwa können die existierenden Modelle im aktuellen Fertigungszustand nicht vorhanden sein. Ein weiterer Grund für fehlende Modelle ist, dass diese aus organisatorischen Gründen schwer zu bekommen sind, beispielsweise wenn nur eine Weiterverarbeitung eines zugelieferten Bauteils stattfindet.

Neben fehlenden Modellen sind auch häufig falsche Modelle ein Problem. So kann es vorkommen, dass die existierenden Daten fertigungsbedingt nicht zum tatsächlichen Produkt passen und Teile daher nicht genau erkannt werden können.

Die Generierung eines CAD-Modells aus den Daten der 3D-Kamera ist somit eine Möglichkeit, um für das tatsächlich vorhandene Produkt eine korrekte Repräsentation zu finden, welche im weiteren Bildverarbeitungsprozess genutzt werden kann. Weiterhin dient dies auch der Vereinfachung der Endanwendung von Mikado ARC. Besonders bei oft variierenden Produktkonfigurationen wird der Anwendungsprozess vereinfacht, wenn nicht erst ein entsprechendes CAD-Modell organisiert werden muss.

1.3 Zielsetzung

Für die Rekonstruktion aus der Punktwolke gibt es drei mögliche Ansätze:

1. Es liegt ein stationäres Objekt vor, welches mit einer am Roboter angebrachten 3D-Kamera erfasst wird. Der Roboter bewegt die Kamera um das Objekt herum, um so Informationen aus mehreren Perspektiven zu erhalten. Diese werden dann miteinander kombiniert; der Hintergrund der erfassten Daten wird anschließend eliminiert.
2. Die Kamera ist stationär angebracht, während das zu erkennende Objekt freihändig demonstriert wird. So kann das Objekt aus mehreren Ansichten aufgenommen werden, welche dann zusammengeführt werden.
3. Es liegen mehrere Objekte der selben Art in verschiedenen Orientierungen vor, beispielsweise in einer Kiste. Nach Segmentierung der Objekte wird aus den unterschiedlichen Teilansichten ein repräsentatives Modell generiert.

Ziel der Arbeit ist es, geeignete Lösungen für die verschiedenen Ansätze zu entwickeln sowie die unterschiedlichen Rekonstruktionsmöglichkeiten zu implementieren.

Weiterhin sollen die Ergebnisse untereinander und mit bereits bestehenden Algorithmen verglichen werden. Dieser Vergleich soll sowohl bezüglich der wichtigsten Eigenschaft der Qualität, als auch auf Basis untergeordneter Faktoren wie Geschwindigkeit der Algorithmen und Nutzungskomfort der Ansätze stattfinden.

2. Grundlagen

Zum Verständnis des Themas der Arbeit ist die Erklärung einiger Grundlagen notwendig. In 2.1 und 2.2 werden zunächst einmal wichtige Grundbegriffe und Datenstrukturen erläutert, die bei der Arbeit mit 3D-Daten auftreten. Anschließend werden die verwendeten Bibliotheken Point Cloud Library (PCL) und Robot Operating System (ROS) erklärt, welche zur Datenverarbeitung bzw. zur Robotersteuerung verwendet wurden. Außerdem sind selbstverständlich die bereits bestehenden Elemente des Mikado-Projekts relevant, da diese Arbeit fundamental darauf aufbaut.

2.1 Aufnahme und Speicherung von 3D-Bildern

Zur Aufnahme von 3D-Bilddaten gibt es mehrere verschiedene Möglichkeiten. Ein LIDAR-System sendet beispielsweise mehrere Lichtstrahlen in verschiedene Richtungen, die anschließend Informationen über die Entfernung zu einem Objekt in diesem Punkt liefern. Eine Stereokamera liefert im Gegensatz dazu zwei Bilder, die anschließend durch spezielle Software zu einem dreidimensionalen Bild zusammengesetzt werden. Eine weitere Möglichkeit besteht darin, ein bestimmtes Muster auf die Umgebung zu projizieren, dieses dann aus einer anderen Perspektive aufzunehmen und aus der räumlichen Verzerrung des Musters die Tiefe zu errechnen.

Die so gewonnenen Informationen lassen sich durch mehrere verschiedene Datenmodelle repräsentieren.

2.1.1 Tiefenbild

Ein Tiefenbild ist eine einfache Möglichkeit, in einem zweidimensionalen Bild zusätzlich Informationen über die Entfernung der Kamera zu Objekten abzuspeichern. Diese Technik ist unter Anderem aus der Computergrafik bekannt, wo sie beim Z-Buffering Anwendung findet [5, S.32]. Dabei werden sowohl das aufgenommene Bild, aber zusätzlich auch ein 2D-Array mit der Tiefeninformation des zugehörigen Pixels gespeichert.

So kann das originale 2D-Bild um eine dritte Dimension erweitert werden, wodurch sich beispielsweise dreidimensionale Formen modellieren oder rekonstruieren lassen [6]. Auch in der 3D-Fotografie finden Tiefenbilder Anwendung [7].



Abbildung 2.1: Depth Map eines 2D-Farbbilds. Entnommen aus [8, S.649]

Tiefenbilder haben im Vergleich zu anderen möglichen Datenmodellen den Vorteil, dass 2D-Bilder sehr einfach um zugehörige Tiefeninformationen erweitert werden können. Jedoch gibt es auch einige Nachteile, die nicht umgangen werden können:

- Die 3D-Daten sind ausschließlich aus der Kameraperspektive vorhanden. Um Informationen aus einer anderen Perspektive zu erhalten, muss erst aufwändig umgerechnet werden.
- Da (in den meisten Fällen) nur ein zweidimensionales Array für die Tiefeninformation angelegt wird, können nur die Entfernungen zu den der Kamera nächsten Objekten gespeichert werden. Verdeckte, reflektierende oder durchsichtige Oberflächen können nicht gespeichert werden.
- Das Array ist an die Auflösung der Kamera gebunden. Insbesondere bei großer Entfernung zu Objekten werden diese aufgrund des Bildwinkels sehr schlecht abgetastet, was später zu Aliasing-Effekten führen kann.
- Die Bittiefe der Werte im Array kann - je nach gewünschter Auflösung - zu niedrig sein, bzw. muss erhöht werden. Eine typische Farbtiefe eines Grauwertbilds von

2. Grundlagen

8 Bit repräsentiert beispielsweise nur $2^8 = 256$ Stufen - was schnell zu wenig wird. Insbesondere bei hoher notwendiger Auflösung im nahen Bereich, aber gleichzeitig vorhandenen weit entfernten Objekten kann dies zum Problem werden.

2.1.2 Voxel Grid

Bei einem als Bitmap vorliegendes 2D-Bild werden die Bilddaten diskretisiert in einem zweidimensionalen Array I gespeichert. Mit einer Bittiefe b , einer Bildhöhe i und einer Bildbreite j ergibt sich damit folgende Matrix:

$$I = \begin{pmatrix} v_{11} & v_{12} & v_{13} & \cdots \\ v_{21} & v_{22} & v_{23} & \cdots \\ v_{31} & v_{32} & v_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, v_{ij} \in [0; 2^b) \wedge b, i, j \in \mathbb{N}^+$$

Dieses Prinzip lässt sich einfach auf den dreidimensionalen Raum erweitern. Die so erhaltene Datenstruktur nennt sich Voxel Grid. Als Voxel bezeichnet man somit eine einzelne Datenzelle im 3D-Array, also das dreidimensionale Äquivalent zum Pixel.

Voxel Grids finden in vielen Bereichen Anwendung, zum Beispiel in der Medizin [9–12] oder Geographie [13]. Auch zum Downsampling von Punktwolken werden sie verwendet [14].

Gegenüber einem Tiefenbild hat ein Voxel Grid den inhärenten Nachteil, dass Speicherplatz für jeden Voxel benötigt wird. Dies führt schnell zu großem Speicherbedarf M , da dieser kubisch zur Auflösung r steigt: $M = r_x * r_y * r_z * b$ bei Bittiefe b . Da der Speicher bereits bei der Initialisierung reserviert werden muss, ist die Änderung der Auflösung oder der Größe des abgedeckten Raumes unmöglich. In diesem Fall müssen ein neues Voxel Grid (mit der neuen Auflösung bzw. Größe) angelegt und sämtliche Daten kopiert werden.

Zur Reduzierung des großen Speicherbedarfs werden deswegen häufig Baumstrukturen (sogenannte Octrees) eingesetzt. Dabei wird der Raum in 8 Voxel geteilt, welche jeweils ein Blatt des Baumes darstellen. Ist der Voxel gefüllt, wird er wieder entsprechend in 8 Untervoxel geteilt. Dies wird wiederholt, bis die gewünschte Auflösung erreicht oder eine bestimmte Tiefe im Baum erreicht ist. Szenen, die große Unterschiede in der Auflösung aufweisen oder viele leere Voxel beinhalten, brauchen so deutlich weniger Speicherplatz.

2. Grundlagen

Ein Vorteil im Vergleich zum Tiefenbild ist jedoch, dass die Abhängigkeit von der Kameraperspektive wegfällt. Dadurch können hier auch Objekte modelliert werden, die im Tiefenbild durch eine Verdeckung versteckt wären. Desweiteren ist der Raum einheitlich diskretisiert. Dies kann je nach Szene und Kameraposition sowohl ein Vor- als auch ein Nachteil sein: Bei einem Tiefenbild nimmt der Abstand der Messpunkte mit der Tiefe ab. Ist die Entfernung zwischen Objekt und Kamera gering, werden diese mit einer deutlich besseren Auflösung abgetastet als sehr weit entfernte Objekte. Im Voxel Grid werden Objekte jedoch überall gleich abgetastet. Dies führt zu verbesserter Auflösung bei entfernten und zu schlechterer Auflösung bei nahen Objekten.

2.1.3 Punktwolken

Ein weiteres häufig verwendetes Datenmodell ist eine Punktwolke. Als Punktwolke bezeichnet man eine Menge $M \subset \mathbb{R}^3$ von Punkten im (mindestens) dreidimensionalen Raum. Zusätzlich zur räumlichen Information können auch noch weitere Daten pro Punkt gespeichert sein, wie RGB-Werte, Genauigkeit oder Objektklasse (falls schon eine Segmentierung vorgenommen wurde). Dadurch gilt:

$$M = \begin{pmatrix} p_x^1 & p_y^1 & p_z^1 & \dots \\ p_x^2 & p_y^2 & p_z^2 & \dots \\ p_x^3 & p_y^3 & p_z^3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Die Nutzung von Punktwolken bringt im Vergleich zu anderen 3D-Datenmodellen einige Vorteile.

- Die zugrundeliegende Datenstruktur ist extrem trivial; es handelt sich um eine einfache Liste. Dies ermöglicht sehr schnelle Operationen, wie zum Beispiel das Hinzufügen eines Punktes. Es lässt sich sehr einfach über die Punkte iterieren.
- Der Speicherbedarf einer Punktwolke wächst dynamisch, und zwar linear mit der Zahl der enthaltenen Punkte. Beim Voxel Grid muss dagegen bereits am Anfang Speicher für jede Zelle reserviert werden, egal ob ein Punkt enthalten ist oder nicht.
- Man ist nicht, wie beim Tiefenbild oder Voxelgrid, auf eine feste Anzahl an Punkten limitiert. Beim Tiefenbild können maximal $n * m$ Punkte (ein Wert pro Pixel) mit einer Tiefeninformation versehen werden, beim Voxel Grid maximal $x * y * z$ Punkte (ein Wert pro Voxel). Eine Kapazitätserweiterung, wie beispielsweise das

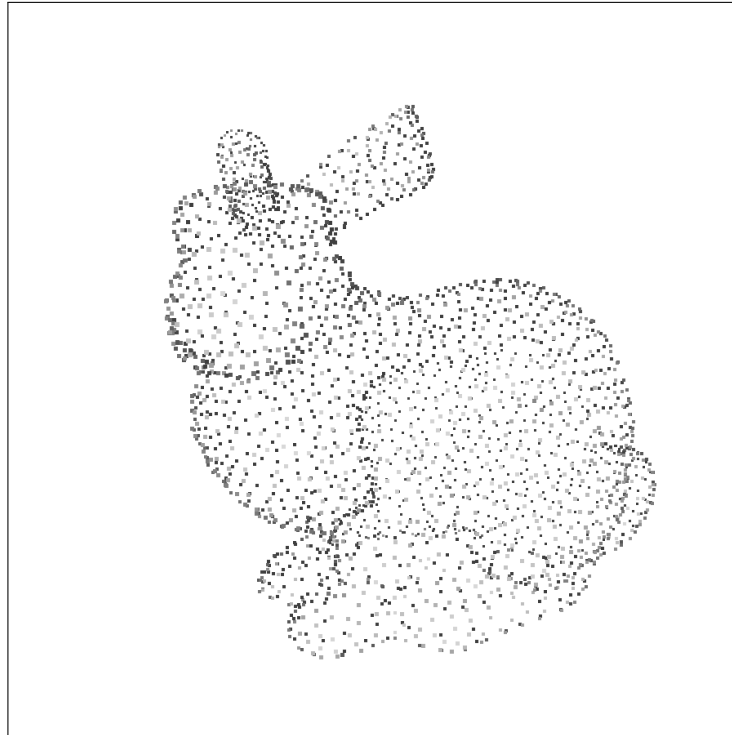


Abbildung 2.2: Punktwolke des Stanford Bunny [15]

Hinzufügen neuer Pixelspalten oder -zeilen im Tiefenbild, ist bei der Punktwolke nicht notwendig.

- Die Auflösung ist nur durch die Gleitkommagenauigkeit der Maschine limitiert. Beim Voxel Grid ist sie im Gegensatz dazu durch die Größe der Voxel limitiert. Beim Tiefenbild ist die Auflösung abhängig von der Bittiefe und der z-Tiefe im Bild - bei weiter entfernten Objekten ist die Auflösung auch niedriger.

Es gibt jedoch auch einige Nachteile gegenüber den anderen Repräsentationen:

- Die Rekonstruktion eines Meshs ist nicht eindeutig. Beim Voxel Grid lässt sich, beispielsweise mithilfe des Marching-Cubes-Algorithmus [16], ein eindeutiges Mesh rekonstruieren. Bei einer Punktwolke ist im Gegensatz dazu nicht festgelegt, welche Punkte miteinander verbunden werden sollen.
- Viele bekannte Techniken aus der 2D-Bildverarbeitung (wie zum Beispiel Segmentierung, Filterung usw.) lassen sich auf Tiefenbilder direkt übertragen. Dies ist bei Punktwolken nicht möglich.

2.2 Meshrepräsentationen

Die bisher angesprochenen Datenstrukturen beinhalten ausschließlich Informationen über Punkte im dreidimensionalen Raum (Geometrie), nicht jedoch aber über die Relation, in der diese zueinander stehen (Topologie). Das Polygonmesh erweitert diese Daten durch Kanteninformationen: Punkte (sogenannte Vertices) sind durch Kanten (Edges) miteinander verbunden. Die so entstehenden Flächen der Polygone werden als Faces bezeichnet. Da jedes Polygon in Dreiecke zerlegt werden kann, werden meist nur diese gespeichert.

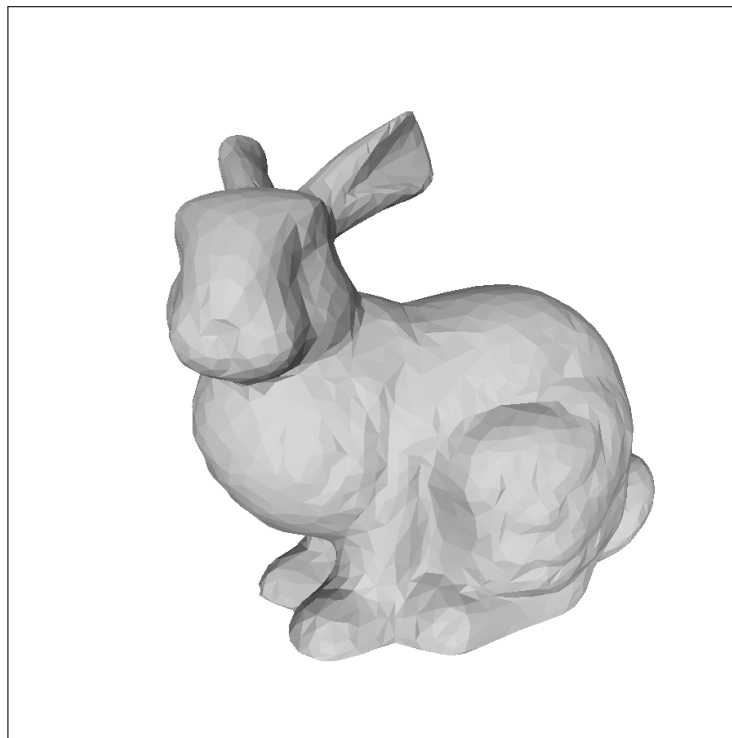


Abbildung 2.3: Mesh des Stanford Bunny [15]

Vertices, Edges und Faces können in verschiedenen Repräsentationen gespeichert werden. Diese haben unterschiedliche Eigenschaften und Vor- und Nachteile, die im folgenden genauer beleuchtet werden.

2.2.1 Vertex-Vertex-Mesh

Beim Vertex-Vertex-Mesh (VVM) werden nur Vertices gespeichert, Kanten und Flächen ergeben sich implizit. Existiert beispielsweise eine Verbindung zwischen den Vertices v_1 und v_2 , v_2 und v_3 sowie v_3 und v_1 (ein einfaches Dreieck), werden die Werte in [Tabelle 2.1](#) in einer Liste gespeichert:

Vertex	verbundene Vertices
v_1	v_2, v_3
v_2	v_1, v_3
v_3	v_1, v_2

Tabelle 2.1: Vertextabelle beim VVM

Der Speicherbedarf dieser simplen Darstellung ist sehr gering. Die Einträge in der Tabelle können als Indizes der Vertexliste gespeichert werden. Weiterhin sind Vertex-basierte Operationen sehr schnell. Zum Hinzufügen eines Knotens muss beispielsweise nur ein neuer Eintrag in der Liste angelegt werden, sowie der neue Knoten zu den verbundenen Vertices hinzugefügt werden.

Um jedoch Informationen zu bestimmten Kanten oder Flächen zu bekommen, muss über die gesamte Liste iteriert werden. Dies ist sehr langsam und schränkt den praktischen Nutzen von VVMs stark ein.

Ein Vergleich und eine Abwägung der Vor- und Nachteile mit anderen Datenstrukturen findet sich in [17, Kap. 11]. Neben dem VVM kann man in beliebigen Kombinationen auch Listen für Kanten oder Faces hinzufügen. Dies erhöht zwar den Speicherbedarf und die Komplexität der Datenstruktur, bestimmte Zugriffe werden dadurch jedoch stark beschleunigt.

2.2.2 Winged Edge

In der Winged-Edge-Datenstruktur [18] werden sowohl Vertices als auch Edges und Faces explizit abgespeichert. Diese Repräsentation ermöglicht es, die Geometrie des Meshes im Vergleich zu anderen Strukturen leichter zu verändern. Der Nachteil ist jedoch, dass der Speicheraufwand sehr hoch und die Datenstruktur im Vergleich zu den anderen Darstellungen sehr komplex ist.

2. Grundlagen

Bei der Winged-Edge-Darstellung werden in der Kantentabelle jeweils die vorhergehenden und nachfolgenden Kanten im und gegen den Uhrzeigersinn gespeichert. Dies ermöglicht die schnelle Bestimmung von angrenzenden Kanten, Vertices oder Faces, bedeutet aber gleichzeitig auch viel Verwaltungsaufwand der Indizes. In [Abbildung 2.4](#) findet sich eine Übersicht über die entsprechenden Kanten.

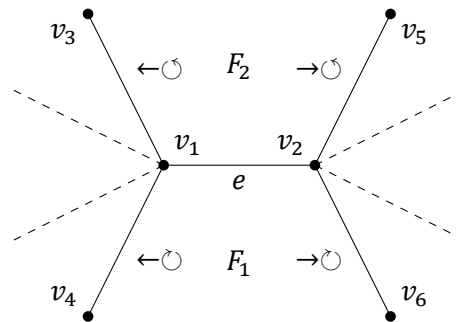


Abbildung 2.4: Vor- und Nachfolgerkanten der Winged-Edge-Datenstruktur (nach [18, S.591])

Die in [Abbildung 2.5](#) gezeigte Darstellung von zwei beispielhaften Dreiecken (v_1, v_2, v_3) und (v_2, v_4, v_3) ist in [Tabelle 2.2](#) dargestellt.

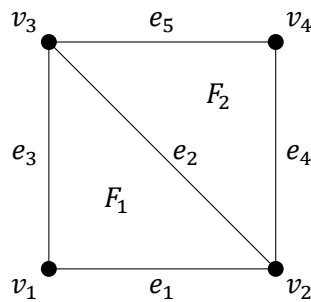


Abbildung 2.5: Beispieldreiecke zur Winged-Edge-Datenstruktur

Edges						
Edge	Vertices	Faces	$\leftarrow \circ$	$\leftarrow \circ$	$\rightarrow \circ$	$\rightarrow \circ$
e_1	v_1, v_2	F_1	e_3	e_3	e_2	e_4
e_2	v_2, v_3	F_1, F_2	e_1	e_4	e_3	e_5
e_3	v_3, v_1	F_1	e_2	e_5	e_1	e_1
e_4	v_2, v_4	F_2	e_2	e_1	e_5	e_5
e_5	v_4, v_3	F_2	e_4	e_4	e_2	e_3

Vertices	
Vertex	Edges
v_1	e_1, e_3
v_2	e_1, e_2, e_4
v_3	e_2, e_3, e_5
v_4	e_4, e_5

Faces	
Face	Edges
F_1	e_1, e_2, e_3
F_2	e_4, e_5, e_2

Tabelle 2.2: Vertex-, Edge- und Face-Tabellen zu [Abbildung 2.5](#)

2.2.3 Half Edge

2.3 Point Cloud Library

Bei der PCL [19] handelt es sich um eine Bibliothek, welche die Arbeit mit 3D-Punktwolken immens vereinfacht. Sie bietet Möglichkeiten zur Filterung, Segmentierung, Oberflächenrekonstruktion und Visualisierung von Punktwolken, sowie weitere Module.

Neben den zahlreichen unterstützten Anwendungsgebieten und implementierten Algorithmen bietet die PCL den weiteren entscheidenden Vorteil, dass sie direkt zu ROS kompatibel ist. Die Kompatibilität wird durch ROS-Nodelets im Paket `perception_pcl` hergestellt, sodass Punktwolken, Meshes oder andere Datenstrukturen direkt per Message versendet werden können.

3. Vorhandene Arbeit

In diesem Kapitel wird der aktuelle Stand der Forschung genannt sowie bereits existierende Lösungsansätze erklärt.

3.1 Registrierung von Punktwolken

Beim Erfassen von Objekten in der Realität, beispielsweise mithilfe eines Laserscanners, werden häufig Punktwolken aus verschiedenen Perspektiven aufgenommen. Der Grund dafür ist, dass aus einer Scanposition nicht immer die gesamte Oberfläche sichtbar ist, etwa aufgrund von Verschattungen durch das Objekt selbst.

In vielen Fällen kann die zugehörige Kameraposition nicht bestimmt werden oder ist unzureichend genau. Aus diesem Grund müssen die erfassten Punktwolken erst in dasselbe Koordinatensystem gebracht werden. Diesen Prozess nennt man die Registrierung von zwei Punktwolken.

3.1.1 Lokale Registrierung

Bei der lokalen Registrierung von zwei Punktwolken müssen diese bereits grob aneinander ausgerichtet sein. Der Registrierungsalgorithmus verfeinert diese Ausrichtung dann weiter. Einer der bekanntesten Algorithmen zur lokalen Registrierung ist Iterative Closest Point (ICP) und seine vielen verschiedenen Varianten.

Iterative Closest Point

ICP [20] ist der wohl bekannteste Algorithmus zur lokalen Registrierung von Punktwolken. Im Laufe der Zeit wurden zahlreiche Varianten und Optimierungen dafür entwickelt [21, 22]. Unter anderem existiert auch eine vorhandene Implementierung in der PCL [23].

Der Algorithmus zur Registrierung der Punktwolke A an der Punktwolke B ist in Algorithmus 1 beschrieben. Der Ablauf lässt sich folgendermaßen grob erklären:

1. A wird iterativ durch Anwendung von Translationen und Rotationen in eine neue Position gebracht, zum Beispiel mithilfe von Quaternionen [24].
2. Zu jedem Punkt $p \in A$ wird der Punkt $q \in B$ gesucht, der den geringsten Abstand ε von p hat.
3. Berechne den Fehler e , Standard ist die Summe der Residuenquadrate.
4. Wiederholung, bis eine der Abbruchbedingungen eintritt.

Die Abbruchbedingungen sind dabei je nach Implementierung unterschiedlich. Häufig kommen aber folgende Kriterien zum Einsatz:

- der Fehler e unter einen Grenzwert fällt
- e bei erneuter Wiederholung nicht um einen bestimmten Wert sinkt
- eine festgelegte Zahl von Iterationen abgelaufen ist.

Selbstverständlich lassen sich beliebig viele zusätzliche Kriterien definieren, beispielsweise eine festgelegte maximale Laufzeit.

Kinect Fusion

Durch die Veröffentlichung der Microsoft Kinect im Jahr 2010 wurde aufgrund ihres niedrigen Preises erstmals der breiten Masse der Zugang zu Tiefenkameras ermöglicht [25, 1:55]. Die Kinect ist eine 3D-Kamera, welche ursprünglich für die Nutzung im Entertainment- und Gamingbereich entwickelt worden ist. Bestehend aus einem Lichtemitter, Infrarotsensor, einer 2D-RGB-Kamera und weiteren Sensoren, liefert sie

Algorithm 1 Iterative Closest Point

```
function ICP(cloud1, cloud2)
  PointMap  $\leftarrow \emptyset$ 
  for all  $p \in \textit{cloud1}$  do           // Find corresponding points
     $q \leftarrow \text{findNearestNeighbor}(\textit{cloud2}, p)$ 
    if  $\text{dist}(p, q) < \textit{maxDist}$  then
      PointMap.add( $p, q$ )
    end if
  end for
  while termination Criteria not met do           // Now, align the clouds
     $t \leftarrow \text{estimateTransformation}()$ 
    cloud1.transform( $t$ )
     $e \leftarrow \text{calculateError}(\textit{PointMap})$ 
  end while
end function
function calculateError(PointMap)
   $e \leftarrow 0$ 
  for all  $p, q \in \textit{PointMap}$  do
     $\varepsilon \leftarrow \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2}$ 
     $e \leftarrow e + \varepsilon^2$ 
  end for
  return  $e$ 
end function
```

3D-Daten bei einer Bildwiederholffrequenz von 30 Hz. Neben dem ursprünglichen Anwendungsgebiet findet sie heute auch in vielen anderen, auch wissenschaftlichen Bereichen Anwendung, beispielsweise bei der Aufzeichnung geomorphologischer Daten [26].

Kinect Fusion (KinFu) ist das Ergebnis einer Forschungsarbeit von Microsoft Research [27] und bietet eine Möglichkeit, mithilfe der Kinect 3D-Rekonstruktionen in Echtzeit durchzuführen. KinFu kombiniert dabei die lokale Registrierung durch den ICP-Algorithmus mit einem Truncated Signed Distance Field (TSDF). Ein TSDF ist im Grunde genommen nichts anderes als ein Voxel Grid. Während ein Voxel im ursprünglichen Voxel Grid entweder gefüllt ist oder nicht, wird hier die Entfernung zur nächsten Oberfläche gespeichert [28]. Dies ermöglicht bei einer relativ geringen Voxelauflösung dennoch eine recht genaue Rekonstruktion der vorhandenen Oberflächen. Um dies zu erreichen, werden die eingetragenen Werte interpoliert, und Fehler durch die geringe Auflösung so minimiert.

Durch eine schnelle GPU-Implementierung auf Basis von CUDA erreicht KinFu dabei eine Integration neuer Tiefenbilder bei 30 Hz, der Bildwiederholffrequenz der Kinect.

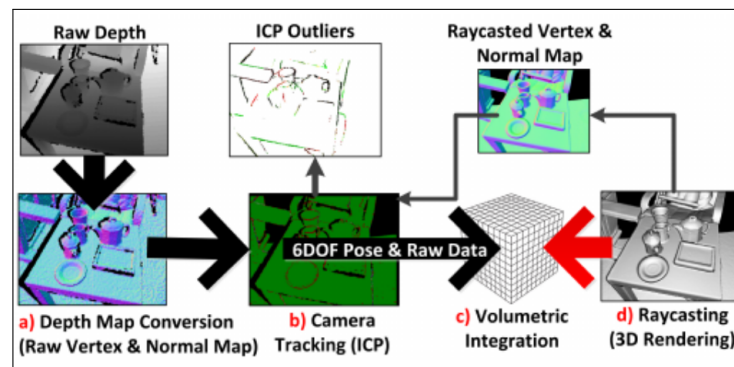


Abbildung 3.1: Integration eines Tiefenbilds in KinFu. Entnommen aus [27]

Somit ist die zeitliche Differenz zwischen zwei Aufnahmen sehr niedrig. Es wird davon ausgegangen, dass die Kamera handgeführt (bzw. mit geringer Geschwindigkeit bewegt) wird, daher hält sich auch die räumliche Distanz zwischen einzelnen Frames in Grenzen. Aufgrund dieser Gegebenheiten reicht bei KinFu eine lokale Registrierung aus, es kann also direkt ICP verwendet werden.

In [Abbildung 3.1](#) ist der Ablauf der Integration eines neuen Tiefenbilds ins TSDF bei KinFu dargestellt. Das Tiefenbild wird zunächst in eine Punktwolke konvertiert. Diese wird anschließend durch ICP registriert, um dann unter Bildung eines Mittelwerts in das TSDF integriert zu werden. Zur Darstellung der Szene wird dieses anschließend mithilfe eines Raycasters gerendert.

Es gibt zahlreiche Optimierungen und veränderte Versionen von KinFu. Unter anderem gibt es in der PCL eine freie Implementierung [\[29\]](#). Eine optimierte Version ist beispielsweise Chisel [\[30\]](#), wo eine effizientere Voxelstrategie verwendet wird. Weiterhin ist Chisel eine reine CPU-Implementierung, was die Nutzung auf Mobilgeräten ohne Grafikkarte ermöglicht.

Die in dieser Arbeit verwendete Implementierung ist Yet Another Kinect Fusion (YAK). YAK ist eine auf ROS angepasste Version von KinFu, um Trajectory Waypoints für die Robotersteuerung zu errechnen [\[31\]](#). Die Verwendung von KinFu in industriellen Robotikanlagen ist in vielerlei Hinsicht sinnvoll:

- Objekte können aus mehreren Perspektiven gescannt werden, was eine bessere Darstellung der Szene liefert.
- Rauschen durch den Bildsensor wird aufgrund der Mittelwertbildung minimiert. Das Ergebnis ist ein glatteres und realistischeres Modell.
- Durch die GPU-Implementierung können Tiefenbilder in Echtzeit integriert wer-

3. Vorhandene Arbeit

den, was insbesondere bei Robotikanwendungen ein großer Vorteil ist.

- Ein häufig in der Bildverarbeitung auftretendes Problem sind reflektierende bzw. spiegelnde Oberflächen. Die betroffenen Regionen können oft nur falsch, verzerrt oder gar nicht modelliert werden. Da hier eine Aufnahme aus mehreren Perspektiven möglich ist, können derartige Fehler weitestgehend vermieden werden. In [Abbildung 3.2](#) ist dies an einem Beispiel besonders gut sichtbar.

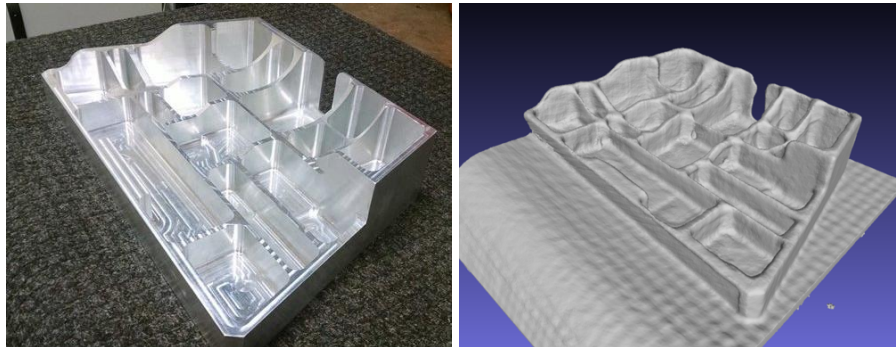


Abbildung 3.2: Szene und Rekonstruktion eines reflektierenden Objekts durch YAK. Entnommen aus [\[31\]](#)

3.1.2 Globale Registrierung

Bei der globalen Registrierung müssen sich die beiden Punktwolken nicht nahe der endgültigen Ausrichtung befinden - Translation und Rotation können beliebig sein. Ein Nachteil ist jedoch, dass die globale Registrierung oft nur eine Grobregistrierung bietet, also keine optimalen Ergebnisse liefert. Aus diesem Grund bietet es sich meist an, anschließend noch eine lokale Registrierung zur Verbesserung der Ergebnisse durchzuführen.

Zur globalen Registrierung von Punktwolken existieren verschiedene Ansätze [\[32–34\]](#). Die in dieser Arbeit verwendete Methode ist Super4PCS [\[35\]](#), eine optimierte Version von Four-Points Congruent Sets (4PCS) [\[36\]](#). Dieser Ansatz wird verwendet, da die vorhandene Implementierung Open Global Registration (OpenGR) [\[37\]](#) über einen Wrapper bereits zur PCL kompatibel ist.

3.2 Segmentierung

Ziel der Segmentierung ist es, zusammengehörende Bildregionen zu identifizieren und durch Zuweisung verschiedener Klassen voneinander zu trennen. Dies ist, insbesondere in der 2D-Bildverarbeitung, ein altes Problem, für das es viele verschiedene Lösungsansätze gibt. Auch im dreidimensionalen Raum ist eine Segmentierung häufig notwendig. Häufig gibt es dabei nicht eine beste Methode - die Objektart, -größe, -form und viele weitere Faktoren beeinflussen die Wahl eines geeigneten Ansatzes.

Die Forschung ist in diesem Bereich weiterhin sehr aktuell: Beispielsweise zeigen Ückermann, Haschke und Ritter einen Versuch, in Echtzeit und ohne vorher bekannte Objekte zu segmentieren [38]. Eine Übersicht bieten Nguyen und Le, die verschiedene Varianten vergleichen und die Vor- und Nachteile diskutieren [39]. In den letzten Jahren findet auch verstärkt Forschung zur Segmentierung mithilfe von Neuronalen Netzen statt [40].

Survey [41], SDF [42]

3.3 Triangulierung

Die Konvertierung einer Datenstruktur zu einem aus Polygonen bestehenden Mesh nennt sich Triangulierung. Gerade bei der Triangulierung von Punktwolken gibt es jedoch nicht immer eine einzelne korrekte Lösung dieses Problems. Es ist beispielsweise nicht immer sichergestellt, dass jeder Punkt auch tatsächlich die zu rekonstruierende Oberfläche darstellt - Rauschen, Reflektionen oder sonstige Fehler können auftreten. Weiterhin stellt sich die Frage, aus welchen der Punkten in der Wolke sich ein bestimmtes Polygon zusammensetzt.

Diese und andere Probleme führen dazu, dass es viele verschiedene Ansätze zur Triangulierung gibt.

3.3.1 Marching Cubes

Bei Marching Cubes [16] handelt es sich um einen Algorithmus, der die Triangulation einer Punktwolke auf ein Voxel Grid zurückführt. Der Ablauf ist in [Algorithmus 2](#) beschrieben.

Die eingegebene Punktwolke wird also zunächst diskretisiert, was bei geringer Kantenlänge der Voxel zu Datenverlust führt. Anschließend wird jeder Voxel getrennt betrachtet und anhand der Eckkonfiguration die zu verbindenden Kanten mithilfe einer Lookup Table (LUT) ermittelt. Dieses Vorgehen wird nun für jeden Voxel wiederholt, was zu einer vollständigen Rekonstruktion der Oberfläche führt.

Algorithm 2 Marching Cubes

```
function marchingCubes
    vertexList  $\leftarrow \emptyset$            // The list of output vertices (Faces implicit)
    V  $\leftarrow$  Voxels
    for all v  $\in V$  do
        index  $\leftarrow$  calculateIndex(v)
        edgeList  $\leftarrow$  edgeTable[index]
        for all edge  $\in$  edgeList do
            vertex  $\leftarrow$  interpolate(edge[0], edge[1])           // interpolate corners
            vertexList.add(vertex)
        end for
    end for
end function

function calculateIndex(voxel)
    voxelIndex  $\leftarrow$  0
    for cornerIndex  $\in$  [0..8] do
        if voxel[cornerIndex] < isolevel then           // corner inside isosurface
            voxelIndex | = 1 << cornerIndex           // Set the i-th bit of the index
        end if
    end for
    return voxelIndex
end function
```

Einer der entscheidenden Nachteile bei diesem Algorithmus ist, dass die Qualität des Meshs direkt abhängig von der gewählten Kantenlänge der Voxel ist. Dies macht viele Vorteile der Punktwolke gegenüber einem Voxel Grid zunichte. Bei zu geringer Kantenlänge nimmt die Zahl der möglichen rekonstruierten Polygone ab, was zu einer geringeren Qualität führt. Wählt man die Auflösung jedoch zu hoch, treten Oversampling-Effekte auf und es entstehen Lücken im Mesh. Insbesondere bei großen Differenzen der Distanzen zwischen Punkten in der Punktwolke führt dies zu Problemen.

3. Vorhandene Arbeit

Weiterhin weisen die Meshes in der Standardimplementierung zackenartige Artefakte auf. Dies passiert dann, wenn die Vertices zu den entsprechenden Kanten mittig auf diesen gewählt werden. Durch eine Interpolation zwischen den Werten der beiden Eckpunkte des Voxels lässt sich dieser Effekt minimieren.

3.3.2 Poisson

Poisson [43]

3.3.3 Weitere Ansätze

Greedy Projection [44], Ball Pivoting [45]

4. Implementierung

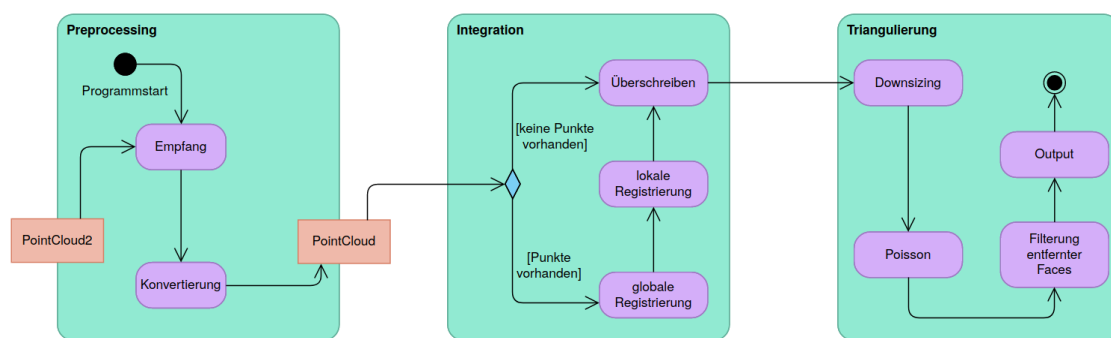


Abbildung 4.1: Ablauf der Pipeline

5. **Auswertung**

6. Fazit

Literatur

- [1] *Isys Vision GmbH*. url: <https://www.isys-vision.de/> (besucht am 30. 03. 2020).
- [2] *Mikado - 3D Bin Picking System for all applications*. url: <https://www.mikado-robotics.com> (besucht am 30. 03. 2020).
- [3] *Ensenso - Stereo 3D Cameras*. url: <https://www.ensenso.com> (besucht am 30. 03. 2020).
- [4] Bertram Heinrich Drost und Markus Ulrich. *Recognition and pose determination of 3D objects in 3D scenes*. US Patent 8,830,229. Sep. 2014.
- [5] Edwin Catmull. *A subdivision algorithm for computer display of curved surfaces*. Techn. Ber. UTAH UNIV SALT LAKE CITY SCHOOL OF COMPUTING, 1974.
- [6] Amir Arsalan Soltani u. a. "Synthesizing 3d shapes via modeling multi-view depth maps and silhouettes with deep generative networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, S. 1511–1519.
- [7] Andre Redert u. a. "Philips 3D solutions: From content creation to visualization". In: *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. IEEE. 2006, S. 429–431.
- [8] Karsten Muller, Philipp Merkle und Thomas Wiegand. "3-D video representation using depth maps". In: *Proceedings of the IEEE* 99.4 (2010), S. 643–656.
- [9] Fedde van der Lijn u. a. "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts". In: *Neuroimage* 43.4 (2008), S. 708–720.
- [10] Stefan Klein u. a. "Elastix: a toolbox for intensity-based medical image registration". In: *IEEE transactions on medical imaging* 29.1 (2009), S. 196–205.
- [11] Manoranjan Mohanty, Pradeep Atrey und Wei Tsang Ooi. "Secure cloud-based medical data visualization". In: *Proceedings of the 20th ACM international conference on Multimedia*. 2012, S. 1105–1108.
- [12] Alexis Roche u. a. "Towards a better comprehension of similarity measures used in medical image registration". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1999, S. 555–566.

- [13] Szymon Chmielewski und Piotr Tompański. “Estimating outdoor advertising media visibility with voxel-based approach”. In: *Applied Geography* 87 (2017), S. 1–13.
- [14] *Downsampling a PointCloud using a VoxelGrid filter*. Website URL changed in may 2020. url: http://pointclouds.org/documentation/tutorials/voxel_grid.php (besucht am 06. 04. 2020).
- [15] Stanford University Computer Graphics Laboratory. *Stanford Bunny*. url: <http://graphics.stanford.edu/data/3Dscanrep/> (besucht am 30. 04. 2020).
- [16] William E Lorensen und Harvey E Cline. “Marching cubes: A high resolution 3D surface construction algorithm”. In: *ACM siggraph computer graphics* 21.4 (1987), S. 163–169.
- [17] Colin Smith. *On vertex-vertex systems and their use in geometric and biological modelling*. University of Calgary, 2006.
- [18] Bruce G Baumgart. “A polyhedron representation for computer vision”. In: *Proceedings of the May 19-22, 1975, national computer conference and exposition*. 1975, S. 589–596.
- [19] Radu Bogdan Rusu und Steve Cousins. “3d is here: Point cloud library (pcl)”. In: *2011 IEEE international conference on robotics and automation*. IEEE. 2011, S. 1–4.
- [20] Paul J Besl und Neil D McKay. “Method for registration of 3-D shapes”. In: *Sensor fusion IV: control paradigms and data structures*. Bd. 1611. International Society for Optics und Photonics. 1992, S. 586–606.
- [21] Szymon Rusinkiewicz und Marc Levoy. “Efficient variants of the ICP algorithm”. In: *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. IEEE. 2001, S. 145–152.
- [22] Sofien Bouaziz, Andrea Tagliasacchi und Mark Pauly. “Sparse iterative closest point”. In: *Computer graphics forum*. Bd. 32. 5. Wiley Online Library. 2013, S. 113–123.
- [23] Dirk Holz u. a. “Registration with the Point Cloud Library - A Modular Framework for Aligning in 3-D”. In: *IEEE Robotics & Automation Magazine* 22 (Dez. 2015), S. 110–124. doi: [10.1109/MRA.2015.2432331](https://doi.org/10.1109/MRA.2015.2432331).
- [24] Berthold KP Horn. “Closed-form solution of absolute orientation using unit quaternions”. In: *Josa a* 4.4 (1987), S. 629–642.
- [25] David Kim. *28c3: KinectFusion*. url: <https://www.youtube.com/watch?v=RvrCAw1IFG0> (besucht am 09. 04. 2020).
- [26] Kenneth David Mankoff und Tess Alethea Russo. “The Kinect: a low-cost, high-resolution, short-range 3D camera”. In: *Earth Surface Processes and Landforms* 38.9 (2013), S. 926–936.

- [27] Shahram Izadi u. a. "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera". In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 2011, S. 559–568.
- [28] Brian Curless und Marc Levoy. "A volumetric method for building complex models from range images". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996, S. 303–312.
- [29] Michele Pirovano. "KinFu—an open source implementation of Kinect Fusion+ case study: implementing a 3D scanner with PCL". In: *Project Assignment* (2012).
- [30] Matthew Klingensmith u. a. "Chisel: Real Time Large Scale 3D Reconstruction Onboard a Mobile Device using Spatially Hashed Signed Distance Fields". In: *Robotics: science and systems*. Bd. 4. 2015, S. 1.
- [31] Joseph Schornak. *YAK: 3D Reconstruction in ROS2*. Aug. 2019. url: <https://rosindustrial.org/news/2019/8/7/yak-3d-reconstruction-in-ros2> (besucht am 27. 05. 2020).
- [32] Kunal N Chaudhury, Yuehaw Khoo und Amit Singer. "Global registration of multiple point clouds using semidefinite programming". In: *SIAM Journal on Optimization* 25.1 (2015), S. 468–501.
- [33] Qian-Yi Zhou, Jaesik Park und Vladlen Koltun. "Fast global registration". In: *European Conference on Computer Vision*. Springer. 2016, S. 766–782.
- [34] Radu Bogdan Rusu, Nico Blodow und Michael Beetz. "Fast point feature histograms (FPFH) for 3D registration". In: *2009 IEEE international conference on robotics and automation*. IEEE. 2009, S. 3212–3217.
- [35] Nicolas Mellado, Dror Aiger und Niloy J. Mitra. "Super 4PCS Fast Global Point-cloud Registration via Smart Indexing". In: *Computer Graphics Forum* 33.5 (2014), S. 205–215. issn: 1467-8659. doi: [10.1111/cgf.12446](https://doi.org/10.1111/cgf.12446). url: <http://dx.doi.org/10.1111/cgf.12446>.
- [36] D. Aiger, N. J. Mitra und D. Cohen-Or. "4-points Congruent Sets for Robust Surface Registration". In: *ACM Transactions on Graphics* 27.3 (2008), #85, 1–10.
- [37] Nicolas Mellado. *OpenGR: a 3D Global Registration Library*. <https://github.com/STORM-IRIT/OpenGR>. 2018.
- [38] André Ückermann, Robert Haschke und Helge Ritter. "Real-time 3D segmentation of cluttered scenes for robot grasping". In: *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*. IEEE. 2012, S. 198–203.
- [39] Anh Nguyen und Bac Le. "3D point cloud segmentation: A survey". In: *2013 6th IEEE conference on robotics, automation and mechatronics (RAM)*. IEEE. 2013, S. 225–230.
- [40] Gusi Te u. a. "Rgcnn: Regularized graph cnn for point cloud segmentation". In: *Proceedings of the 26th ACM international conference on Multimedia*. 2018, S. 746–754.

- [41] Ariel Shamir. "A survey on mesh segmentation techniques". In: *Computer graphics forum*. Bd. 27. 6. Wiley Online Library. 2008, S. 1539–1556.
- [42] Lior Shapira, Ariel Shamir und Daniel Cohen-Or. "Consistent mesh partitioning and skeletonisation using the shape diameter function". In: *The Visual Computer* 24.4 (2008), S. 249.
- [43] Michael Kazhdan, Matthew Bolitho und Hugues Hoppe. "Poisson surface reconstruction". In: *Proceedings of the fourth Eurographics symposium on Geometry processing*. Bd. 7. 2006.
- [44] Zoltan Csaba Marton, Radu Bogdan Rusu und Michael Beetz. "On Fast Surface Reconstruction Methods for Large and Noisy Datasets". In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Kobe, Japan, Mai 2009.
- [45] Fausto Bernardini u. a. "The ball-pivoting algorithm for surface reconstruction". In: *IEEE transactions on visualization and computer graphics* 5.4 (1999), S. 349–359.