

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 industriell-
 les Bin-Picking und Depalettieren

Kandidat: 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 Isys Vision GmbH unter Anleitung von Prof. Dr. Georg Umlauf und Simon Schmeißer, M. Sc. 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
Kandidat:	Lorenz Bung
Betreuer:	Prof. Dr. Georg Umlauf Institut für Optische Systeme Simon Schmeißer, M. Sc. Isys Vision GmbH
Abgabedatum:	31.07.2020
Schlagworte:	Meshrekonstruktion, Robotik, Geometrisches Modellieren, ROS, PCL, Punktwolke

Es wurden verschiedene Ansätze zur Rekonstruktion von Meshes aus 3D-Punktwolken analysiert, miteinander verglichen und bewertet. Dazu wurde eine Pipeline entwickelt, welche aus Segmentierung, Registrierung und Triangulation der Punktwolke besteht.

Im ersten Ansatz wurde ein einzelnes Teil mithilfe einer beweglichen Kamera am Roboterarm aus mehreren Posen eingescannt und so ein Gesamtbild erstellt. Bei der zweiten Methode wurde zusätzlich eine Registrierung durchgeführt, so dass eine fest montierte Kamera verwendet werden kann. Im dritten Ansatz fand zunächst eine Segmentierung statt, welche sortenreine Teile in einzelne Cluster teilt. Nach der Registrierung dieser Segmente wurde eine Triangulation durchgeführt.

Es hat sich gezeigt, dass die Qualität des generierten Meshs entscheidend von der Segmentierung abhängig ist. Die Rekonstruktion durch Rotation eines einzelnen Objekts lieferte das beste Ergebnis.

Extended Abstract

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

Einleitung

Robotik spielt im industriellen Bereich eine größer werdende Rolle. Insbesondere das sogenannte Bin-Picking, also das autonome Greifen unsortierter Teile gewinnt zunehmend an Relevanz. Zur Erkennung der Objekte werden dabei in vielen Fällen CAD-Modelle benötigt, welche jedoch nicht immer vorliegen. Die Rekonstruktion aus aufgenommenen 3D-Daten bietet daher eine einfache Möglichkeit, diese zu erhalten. Es wurden verschiedene Ansätze dafür analysiert, verglichen und bewertet.

Methodik

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielsetzung	2
1.3	Anwendungsbereich	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	10
2.2.2	Winged Edge	10
2.2.3	Half Edge	12
3	Vorherige Arbeit	13
3.1	Segmentierung	13
3.1.1	Segmentierung im 3D	14
3.1.1.1	Euclidean Cluster Extracion	14
3.1.1.2	Region Growing	15
3.1.2	Watershed-Transformation	17
3.2	Registrierung von Punktwolken	18
3.2.1	Globale Registrierung	18
3.2.2	Lokale Registrierung	19
3.2.2.1	Iterative Closest Point	19
3.2.2.2	Kinect Fusion	21
3.3	Triangulierung	23
3.3.1	Marching Cubes	24
3.3.2	Poisson	25
3.3.3	Weitere Ansätze	26
3.3.3.1	Greedy-Triangulierung	26
3.3.3.2	Delaunay-Triangulierung	27
3.3.3.3	Ball Pivoting	28

4	Implementierung	30
4.1	Aufbau	30
4.2	Interaktion und Parametrisierung	31
4.3	Pipelineablauf	31
4.3.1	Preprocessing	32
4.3.2	Segmentierung	33
4.3.3	Registrierung	33
4.3.4	Integration	33
4.3.5	Triangulierung	34
5	Auswertung	35
5.1	Evaluation der Daten	35
5.1.1	Werte	35
5.1.2	Bewertung	38
5.2	Fehlerbetrachtung	38
5.3	Zukünftige Arbeit	39
6	Zusammenfassung	40
	Literatur	VIII

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
ECE	Euclidean Cluster Extracion
RG	Region Growing
RoI	Region of Interest
MSE	Mean Squared Error
KNN	K Nearest Neighbors

1. Einleitung

1.1 Motivation

Robotik im industriellen Kontext spielt in den letzten Jahren eine immer größer werdende Rolle für die Wirtschaft. Insbesondere das autonome Greifen von unsortierten Teilen (sogenanntes Bin-Picking) durch einen Roboter kann viele Prozesse beschleunigen und somit wertvolle Zeit und Ressourcen sparen.

Zum Greifen der Objekte müssen jeweils Position und Orientierung bekannt sein. Für die dabei durchgeführte Erkennung sind in vielen Fällen CAD-Modelle der Objekte erforderlich, beispielsweise für den “surface based matching”-Algorithmus [1]. Dies stellt häufig ein Problem dar, da diese aus verschiedenen Gründen nicht vorhanden sein können. Etwa können für Objekte im aktuellen Fertigungszustand keine Modelle existieren. Ein weiterer Grund für fehlende CAD-Daten ist, dass diese aus organisatorischen Gründen schwer zu bekommen sind, beispielsweise wenn nur eine Weiterverarbeitung eines zugelieferten Bauteils stattfindet.

Neben fehlenden sind auch häufig falsche Modelle ein Problem. So kann es vorkommen, dass das tatsächlich vorhandene Produkt fertigungsbedingte Abweichungen von den existierenden Daten aufweist und Teile daher nicht korrekt erkannt werden können.

Bei Bin-Picking-Anwendungen sind in vielen Fällen 3D-Kameras am Roboter verbaut. Um schnell eine korrekte Darstellung der Objekte zu erhalten ist es eine Möglichkeit, die aufgenommenen 3D-Daten zur Generierung eines CAD-Modells zu verwenden. Insbesondere bei wechselnden Produktkonfigurationen vereinfacht dies den Anwendungsprozess enorm.

1.2 Zielsetzung

In vielen Fällen liefern 3D-Kameras Daten in Form einer Punktwolke, eine Darstellung, auf die in 2.1.3 genauer eingegangen wird. Zur Rekonstruktion eines CAD-Modells aus einer solchen Punktwolke gibt es drei verschiedene Ansätze:

1. Eine am Roboterarm befestigte Kamera nimmt das Objekt aus verschiedenen Winkeln auf. Die Kameraposition ist durch die Pose des Roboters gegeben.
2. Eine fest montierte, unbewegliche Kamera nimmt mehrere Aufnahmen eines Objekts auf, das zwischen den Aufnahmen bewegt wird. Da die Position des Teils unbekannt ist, muss diese geschätzt werden.
3. Durch eine unbewegliche Kamera wird eine einzelne Aufnahme einer Kiste mit Objekten derselben Art aufgenommen. Die unterschiedliche Orientierung der Objekte innerhalb der Kiste führt zu Daten aus mehreren Perspektiven.

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.

1.3 Anwendungsbereich

Entwurf, Implementierung und Evaluation der Ergebnisse finden mithilfe des proprietären Softwarepakets Mikado [2] statt, einem Projekt von isys vision GmbH [3]. Zur Erfassung der Bilddaten kommen 3D-Kameras von Ensenso [4] zum Einsatz.

2. Grundlagen

Zum Verständnis des Themas der Arbeit ist die Erklärung einiger Grundlagen notwendig. In 2.1 werden zunächst wichtige Datenstrukturen zur Arbeit mit 3D-Daten eingeführt und die verschiedenen Vor- und Nachteile beleuchtet. Weiterhin werden in 2.2 unterschiedliche Methoden und Repräsentationen von Polygonnetzen erklärt. Auch hier wird diskutiert, welche positiven und negativen Aspekte die jeweiligen Ansätze auszeichnet.

2.1 Aufnahme und Speicherung von 3D-Bildern

Zur Aufnahme von 3D-Bilddaten gibt es mehrere verschiedene Möglichkeiten. Ein LIDAR-System sendet beispielsweise mehrere Laserstrahlen in verschiedene Richtungen. So können anschließend Informationen über die Entfernung zum jeweiligen Punkt geliefert werden, in dem die einzelnen Strahlen auf ein Objekt treffen. Eine Stereokamera liefert im Gegensatz dazu zwei Bilder, die anschließend durch spezielle Software (beispielsweise durch Semiglobal Matching [5] oder den PatchMatch-Algorithmus [6]) 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 [7].

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 [8, S. 32]. Dabei werden sowohl das aufgenommene Bild, aber zusätzlich auch ein 2D-Array mit der Tiefeninformation des zugehörigen Pixels gespeichert. Das Tiefenbild dient somit als Lookup Table, um die z-Information zu einem Pixel im 2D-Bild zu erhalten. Ein Beispiel für den Zusammenhang zwischen diesen beiden Komponenten ist in [Abbildung 2.1](#) dargestellt.

So kann das originale 2D-Bild um eine dritte Dimension erweitert werden, wodurch sich beispielsweise dreidimensionale Formen modellieren oder rekonstruieren lassen [9]. Auch in der 3D-Fotografie finden Tiefenbilder Anwendung [10].



Abbildung 2.1: 2D-Farbbild und zugehöriges Tiefenbild. Entnommen aus [11, 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 eine aufwändige Umrechnung stattfinden.
- Da (in den meisten Fällen) nur ein zweidimensionales Array mit einem Kanal für die Tiefeninformation angelegt wird, können nur die Entfernungen zu Objekten gespeichert werden, die der Kamera am nächsten sind. 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.

2. Grundlagen

- Die Bittiefe der Werte im Array kann - je nach gewünschter Auflösung - zu niedrig sein, beziehungsweise muss erhöht werden. Eine typische Farbtiefe eines Grauwertbilds von 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 vorliegenden 2D-Grauwertbild werden die Bilddaten diskretisiert in einem zweidimensionalen Array I gespeichert. Mit einer Bittiefe b , einer Bildhöhe i und einer Bildbreite j ergibt sich damit die in [Gleichung 2.1](#) dargestellte Matrix für I .

$$I = \begin{pmatrix} v_{11} & v_{12} & v_{13} & \dots & v_{1j} \\ v_{21} & v_{22} & v_{23} & \dots & v_{2j} \\ v_{31} & v_{32} & v_{33} & \dots & v_{3j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{i1} & v_{i2} & v_{i3} & \dots & v_{ij} \end{pmatrix}, v_{ij} \in [0; 2^b - 1], b, i, j \in \mathbb{N}^+ \quad (2.1)$$

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 Computertomographie [[12](#), S. 15], Archäologie oder Stadtplanung [[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 beziehungsweise Größe) angelegt und sämtliche Daten kopiert werden.

Zur Reduzierung des großen Speicherbedarfs werden deswegen häufig Baumstrukturen (sogenannte Octrees) eingesetzt [[15](#)]. Dabei wird der Raum in 8 Subvoxel geteilt, welche jeweils ein Blatt des Baumes darstellen. Ist der Voxel gefüllt, wird er wieder ent-

2. Grundlagen

sprechend in 8 Subvoxel geteilt. Dies wird so lange wiederholt, bis die gewünschte Auflösung 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.

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 nicht sichtbar wären. Desweiteren ist der Raum einheitlich diskretisiert. Dies kann je nach Szene und Kameraposition sowohl ein Vorteil 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 in deutlich mehr Messpunkten erfasst als sehr weit entfernte Objekte. Im Voxel Grid werden Objekte jedoch überall gleich abgetastet. Dies führt zu verbesserter Auflösung bei Objekten, die weiter entfernt sind, jedoch zu schlechterer Auflösung bei nahen Objekten.

2.1.3 Punktwolken

Ein weiteres häufig verwendetes Datenmodell ist eine Punktwolke. Hier werden die Daten als Teilmenge eines Vektorraums dargestellt. Im Fall dreidimensionaler Daten ist eine Punktwolke also $M \subseteq \mathbb{R}^3$. In den meisten Fällen repräsentieren die Punkte dabei eine Diskretisierung der Oberfläche. Die in [Abbildung 2.2](#) dargestellte Punktwolke stellt beispielsweise die Oberfläche des Stanford Bunny [16] dar. Zusätzlich zur räumlichen Information können auch noch weitere Daten pro Punkt gespeichert sein, wie RGB-Werte, Normalen, Genauigkeit oder Objektklasse, falls schon eine Segmentierung vorgenommen wurde.

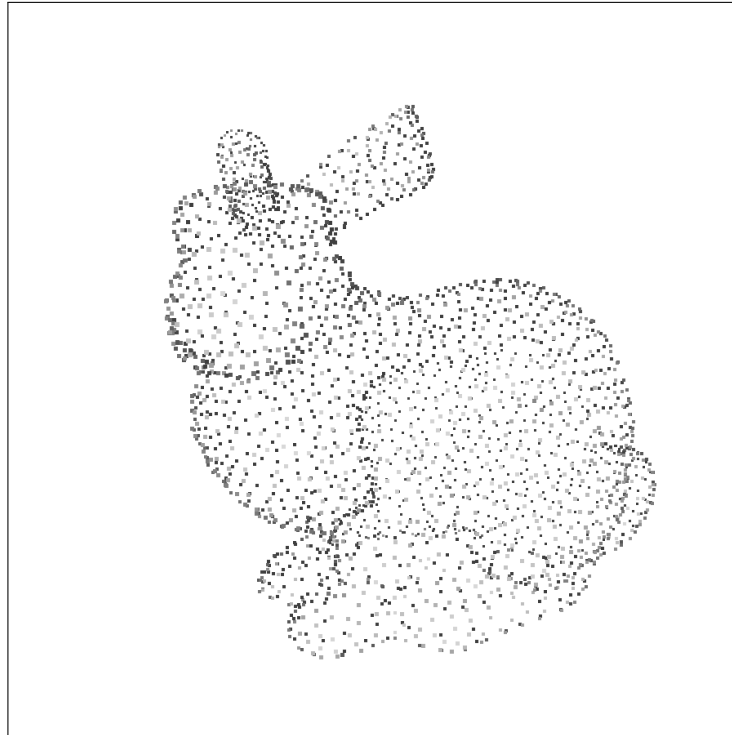


Abbildung 2.2: Punktwolke des Stanford Bunny [16]

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

- Die zugrundeliegende Datenstruktur ist eine Liste. Dies ermöglicht sehr schnelle Operationen, wie zum Beispiel das Hinzufügen eines Punktes oder die Erweiterung um zusätzliche Informationen. 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, das Voxel Grid kann maximal $x * y * z$ Punkte repräsentieren (ein Wert pro Voxel). Eine Kapazitätserweiterung, wie beispielsweise das Hinzufügen neuer Pixelspalten oder -zeilen im Tiefenbild, ist bei der Punktwolke nicht notwendig.

2. Grundlagen

- Die Auflösung ist nur durch die Gleitkommagenauigkeit der Maschine limitiert. Beim Voxel Grid ist sie im Gegensatz dazu durch die Größe der Voxel limitiert und kann aufgrund des Speicherbedarfs nicht beliebig gut gewählt werden. 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 ein eindeutiges Mesh rekonstruieren, beispielsweise mithilfe des Marching-Cubes-Algorithmus [17]. 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 oder Filterung 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 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. Die in [Abbildung 2.2](#) dargestellte Punktwolke ist in [Abbildung 2.3](#) als Mesh zu sehen.

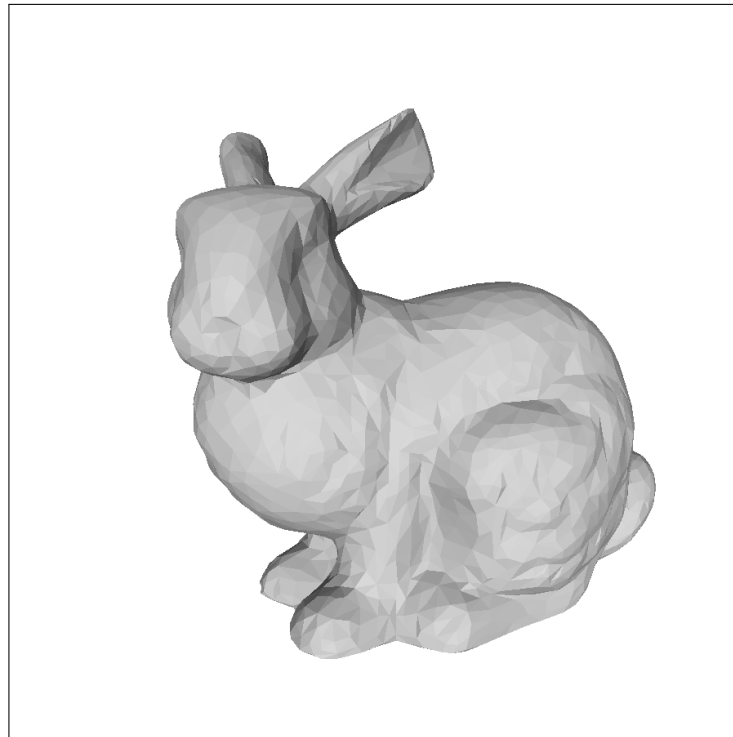


Abbildung 2.3: Mesh des Stanford Bunny [16]

Vertices, Edges und Faces können in verschiedenen Repräsentationen gespeichert werden. Diese haben unterschiedliche Eigenschaften und Vor- und Nachteile, die im folgenden genauer beleuchtet werden. Zum besseren Verständnis wird das in [Abbildung 2.4](#) dargestellte Polygonnetz jeweils in den unterschiedlichen Strukturen codiert.

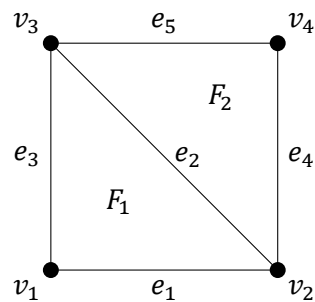


Abbildung 2.4: Beispielhaftes Polygonnetz, bestehend aus Knoten, Kanten und Faces

2.2.1 Vertex-Vertex-Mesh

Beim Vertex-Vertex-Mesh (VVM) werden zu jedem Vertex nur die verbundenen Vertices gespeichert. Kanten und Flächen ergeben sich daraus implizit. Zu den in [Abbildung 2.4](#) dargestellten Dreiecken wird also die Liste in [Tabelle 2.1](#) gespeichert.

Vertex	verbundene Vertices
v_1	v_2, v_3
v_2	v_1, v_3, v_4
v_3	v_1, v_2, v_4
v_4	v_2, v_3

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 [18, 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 [19] 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 relativ komplex ist.

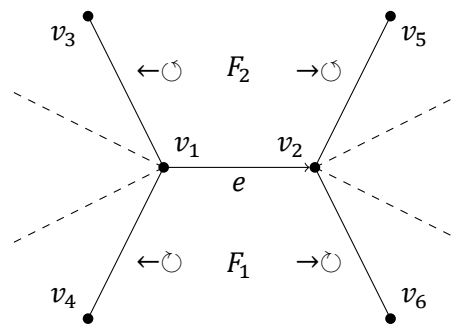


Abbildung 2.5: Vor- und Nachfolgerkanten der Winged-Edge-Datenstruktur nach [19, S. 591]

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.

Um Vor- und Nachfolger jeweils im und gegen den Uhrzeigersinn definieren zu können, muss eine Orientierung vorliegen. Die Kanten in der Winged-Edge-Datenstruktur sind daher immer gerichtet. In [Abbildung 2.5](#) findet sich eine Übersicht über die entsprechenden Kanten.

Die Winged-Edge-Datenstruktur zum Beispiel in [Abbildung 2.4](#) ist in [Tabelle 2.2](#) zu sehen.

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 bei der Winged-Edge-Darstellung

2.2.3 Half Edge

Im Gegensatz zu bisherigen Modellen ist die Idee bei der Half-Edge-Datenstruktur, die Edges in jeweils zwei Halbkanten aufzuteilen. Eine Halbkante hat somit einen Vor- und Nachfolger, sowie einen gegenüberliegenden Nachbarn. Da wie bei der Winged-Edge-Darstellung eine Orientierung notwendig ist, ist die Halbkante gerichtet.

Der Vorteil dieser Art der Speicherung ist, dass sowohl Kantenvorgänger und -nachfolger schnell bestimmt werden können, aber insbesondere auch aneinander angrenzende Faces. Das Hinzufügen, Entfernen oder Ändern von Daten ist jedoch leichter als bei der Winged-Edge-Darstellung, da nicht zu jeder Kante vier Nachbarkanten und alle Faces gespeichert werden müssen.

Die Vertex-, Edge- und Face-Tabellen zu [Abbildung 2.4](#) sind in [Tabelle 2.3](#) aufgelistet. Statt der vollständigen Außen- beziehungsweise Innenzyklen können bei den Faces auch nur einzelne Halbkanten gespeichert werden. Die entsprechenden Vorgänger- und Nachfolgerhalbanten repräsentieren den kompletten Zyklus implizit.

Half-Edges					
Edge	← Edge	→ Edge	Twin	Origin	Face
e_{1a}	e_{3a}	e_{2a}	e_{1b}	v_1	F_1
e_{1b}	e_{4b}	e_{3b}	e_{1a}	v_2	-
e_{2a}	e_{1a}	e_{3a}	e_{2b}	v_2	F_1
e_{2b}	e_{5a}	e_{4a}	e_{2a}	v_3	F_2
e_{3a}	e_{2a}	e_{1a}	e_{3b}	v_3	F_1
e_{3b}	e_{1b}	e_{5b}	e_{3a}	v_1	-
e_{4a}	e_{2b}	e_{5a}	e_{4b}	v_2	F_2
e_{4b}	e_{5b}	e_{1b}	e_{4a}	v_4	-
e_{5a}	e_{4a}	e_{2b}	e_{5b}	v_4	F_2
e_{5b}	e_{3b}	e_{4b}	e_{5a}	v_3	-

Vertices	
Vertex	ausgehend
v_1	e_{1a}, e_{3b}
v_2	e_{1b}, e_{2a}, e_{4a}
v_3	e_{2b}, e_{3a}, e_{5b}
v_4	e_{4b}, e_{5a}

Faces		
Face	Außenzyklus	Innenzyklus
F_1	e_{1b}, e_{3b}, e_{2b}	e_{1a}, e_{2a}, e_{3a}
F_2	e_{4b}, e_{2a}, e_{5b}	e_{2b}, e_{4a}, e_{5a}

Tabelle 2.3: Vertex-, Edge- und Face-Tabellen bei der Half-Edge-Darstellung nach [20]

3. Vorherige Arbeit

Das Kernproblem der Arbeit lässt sich in drei verschiedene Teile gliedern: Segmentierung in einzelne Cluster (3.1), Registrierung von mehreren Punktwolken in einem Koordinatensystem (3.2) und Triangulierung der Wolke zu einem Polygonnetz (3.3).

Im Folgenden werden die wissenschaftlichen Erkenntnisse zum jeweiligen Thema erläutert. Außerdem werden mögliche Lösungsansätze auf ihren praktischen Nutzen untersucht und die Vor- und Nachteile diskutiert.

3.1 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 lange bekanntes Problem, für das es viele verschiedene Lösungsansätze gibt. Auch im dreidimensionalen Raum ist eine Segmentierung häufig notwendig. In vielen Fällen 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 [21]. Eine Übersicht bieten Nguyen und Le, die verschiedene Varianten vergleichen und die Vor- und Nachteile diskutieren [22]. In den letzten Jahren findet auch verstärkt Forschung zur Segmentierung mithilfe von neuronalen Netzen statt [23].

Für die Segmentierung können viele Informationen über die vorhandene Szene relevant sein. Neben der reinen geometrischen Information können zum Beispiel auch Textur

3. Vorherige Arbeit

(also RGB-Werte), Normalenorientierung oder Topologie hilfreich sein, um zusammenhängende Regionen voneinander unterscheiden zu können.

Die Segmentierung muss dabei keineswegs immer auf 2D-Bildern oder 3D-Punktwolken stattfinden. In [24] wird beispielsweise eine Übersicht über Segmentierungsverfahren auf 3D-Meshes vorgestellt. Weiterhin müssen auch nicht zwingend unterschiedliche Objekte getrennt werden. Das Ziel kann es auch sein, nur eine Trennung verschiedener Partitionen innerhalb eines Objekts zu erzielen, zum Beispiel um eine Hand am menschlichen Körper zu identifizieren [25].

3.1.1 Segmentierung im 3D

Im dreidimensionalen Raum gibt es viele verschiedene Eigenschaften (Features), die zur Segmentierung von Punktwolken herangezogen werden können. Es spielt selbstverständlich die Distanz zwischen den Punkten eine Rolle, aber auch andere Faktoren wie Normalenorientierung, Farbe oder Krümmung bieten in vielen Fällen nützliche Zusatzinformationen.

Im Folgenden werden die zwei getesteten Methoden Euclidean Cluster Extracion (ECE) und Region Growing (RG) zur Segmentierung von dreidimensionalen Punktwolken erklärt. Eine gute Übersicht über die beiden Verfahren bietet auch Rusu in [26, S. 88–93].

3.1.1.1 Euclidean Cluster Extracion

Diese Methode der Segmentierung im dreidimensionalen Raum ist sehr einfach. Sie basiert auf der Annahme, dass Punkte so lange zu einem Cluster gehören, so lange ihr Abstand unter einem Grenzwert liegt.

In Algorithmus 1 wird beschrieben, wie die Cluster $\mathcal{C} \subseteq P$ für die Punktwolke P errechnet werden. Der Threshold δ gibt dabei den minimalen Abstand zwischen zwei Clustern an.

3. Vorherige Arbeit

Algorithm 1 Euclidean Cluster Extracion [26, S. 89–90]

```
 $\mathcal{C} \leftarrow \emptyset$  // list of clusters
 $\mathcal{Q} \leftarrow \emptyset$  // current cluster queue
while  $P \neq \emptyset$  do
   $p \leftarrow P.\text{pop}()$ 
   $\mathcal{Q}.\text{append}(p)$ 
  while  $\mathcal{Q} \neq \emptyset$  do
     $q \leftarrow \mathcal{Q}.\text{pop}()$ 
     $\mathcal{N} \leftarrow \text{pointsInRegion}(q, \delta)$  //  $\delta$ -region around  $q$ 
     $\mathcal{Q}.\text{addAll}(\mathcal{N})$ 
     $P.\text{removeAll}(\mathcal{N})$ 
  end while
   $\mathcal{C}.\text{add}(\mathcal{Q})$  // add processed queue to clusters
   $\mathcal{Q} \leftarrow \emptyset$  // reset the queue
end while
return  $\mathcal{C}$ 
```

Nach der Berechnung der verschiedenen Cluster werden diese noch gefiltert. Dabei werden nur Cluster mit $\theta_{\min} < |\mathcal{C}| < \theta_{\max}$ erhalten, alle anderen werden verworfen. Die untere Grenze θ_{\min} dient dabei der Robustheit gegen Rauschen und statistische Ausreißer, während die obere Grenze θ_{\max} verhindert, dass untersegmentierte Regionen zurückgegeben werden.

3.1.1.2 Region Growing

Ein großer Nachteil von ECE ist, dass nur die räumliche Distanz betrachtet wird. Bei nicht optimal getrennten Regionen wird die Punktwolke daher untersegmentiert und aufgrund der Filterung am Ende werden viele Cluster wieder verworfen. Dies ist auch bei den meisten Anwendungen in der Realität der Fall. Selbst bei gesetzter Region of Interest (RoI) und somit bereits gefiltertem Hintergrund liegen Objekte häufig nahe beieinander.

RG umgeht dieses Problem, indem statt der Distanzmetrik die Krümmung im entsprechenden Punkt betrachtet wird. Unter der Krümmung versteht man die Änderung der Normalen in einem Punkt. Eine gute Approximation liefert die in Gleichung 3.1 dargestellte Metrik. Weiterhin spielt auch die Glattheit der Oberfläche eine Rolle (siehe dazu Definition 2).

Definition 1 Sei $p \in P \subset \mathbb{R}^3$ und $P' = P \setminus \{p\}$ die Menge der aufsteigend nach Abstand

3. Vorherige Arbeit

von p sortierten Punkte. Dann sind die K Nearest Neighbors

$$Q_k(p) = \{q_i(p) | i = 1, \dots, k, i < j \Rightarrow \|p - q_i\| \leq \|p - q_j\|\}.$$

Sei E_k nun eine Ebene, welche die Punkte $Q_k(p)$ nach least squares [27], RANSAC [28] oder anders approximiert. Dann ist

$$\kappa = r^2 = \|p - E_k\|^2, \quad (3.1)$$

also der Mean Squared Error (MSE) zwischen E_k und p , eine Metrik zur Schätzung der Krümmung in p (nach [29, Abs. 2.1.2]).

Definition 2 Sei f eine Funktion. Dann gibt die Glattheit $\mathcal{G}(f)$ an, wie oft f differenzierbar ist:

$$\mathcal{G}(f) = \begin{cases} 0 & f \text{ nicht differenzierbar} \\ n & f^{(n)} \text{ definiert} \\ \infty & f \text{ unendlich oft differenzierbar} \end{cases} \quad (3.2)$$

Ansonsten verhält sich der RG-Algorithmus sehr ähnlich zu ECE. Einzig die Fehlermetrik wird geändert; statt der euklidischen Norm $\|\cdot\|_2$ werden Normalenwinkel und Krümmung betrachtet. Algorithmus 2 beschreibt den Ablauf bei maximaler Krümmung ϕ , Normalendifferenz α und unter Betrachtung der k K Nearest Neighbors (KNN).

Algorithm 2 Region Growing

```
 $\mathcal{C} \leftarrow \emptyset$  // list of clusters
 $\mathcal{Q} \leftarrow \emptyset$  // current cluster queue
sort  $P$  by ascending curvature
while  $P \neq \emptyset$  do
   $p \leftarrow P.\text{pop}()$ 
   $\mathcal{Q}.\text{append}(p)$ 
  while  $\mathcal{Q} \neq \emptyset$  do
     $q \leftarrow \mathcal{Q}.\text{pop}()$ 
     $\mathcal{N} \leftarrow \text{getKNeighbors}(q, k)$  //  $k$  nearest neighbors of  $q$ 
    for all  $c \in \mathcal{N}$  do
      if  $\text{curvature}(c) < \phi \wedge \cos^{-1}(\vec{n}_p, \vec{n}_c) < \alpha$  then
         $\mathcal{Q}.\text{add}(c)$ 
         $P.\text{remove}(c)$ 
      end if
    end for
  end while
   $\mathcal{C}.\text{add}(\mathcal{Q})$  // add processed queue to clusters
   $\mathcal{Q} \leftarrow \emptyset$  // reset the queue
end while
return  $\mathcal{C}$ 
```

3.1.2 Watershed-Transformation

Die bisherigen Ansätze sind Möglichkeiten zur Segmentierung in 3D-Punktwolken. Dies kann häufig deutliche Vorteile gegenüber einer Segmentierung im zweidimensionalen Raum haben, in manchen Fällen aber auch weniger brauchbare Informationen liefern. Beispielsweise können Objekte sehr nah aneinander liegen und ECE damit erschweren. Kanten zwischen Objekten können im 2D-Grauwertbild jedoch deutlich besser sichtbar sein. Die Segmentierung auf zweidimensionalen RGB- oder Grauwerten kann daher in manchen Fällen bessere Ergebnisse liefern als eine Segmentierung auf Basis der Geometrie.

Die Watershed-Transformation ist eine der bekanntesten Methoden, um eine solche Segmentierung vorzunehmen. Der Gedanke dahinter stammt aus der Geographie und beschreibt das Verhalten von ansteigendem Wasser in unebenem Terrain [30, S. 1–2]. Dabei bilden sich Wasserbecken, welche bei steigendem Pegel an bestimmten Linien zusammenlaufen. Diese Linien beschreiben im Fall des Clusterings die Grenzen zwischen unterschiedlichen Bildregionen.

Der Ablauf der Watershed-Transformation ist in Algorithmus 3 beschrieben. Dabei werden zunächst die garantierten Hintergrundpixel einheitlich auf 0 gesetzt und ein Pre-processing in Form eines Laplace-Filters vorgenommen. Weiterhin basiert die Binarisierung des Eingabebilds auf dem Thresholding nach Otsu [31], welches die Varianz σ zwischen den Grauwertbins minimiert. Hier kann auch eine beliebige andere Methode verwendet werden. Dieser Ansatz liefert jedoch bei einer binären Verteilung im Histogramm optimale Ergebnisse.

Algorithm 3 Watershed-Transformation [32]

```
 $\mathcal{I} \leftarrow \text{input}$ 
Set guaranteed background pixels to 0
 $L \leftarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ 
 $\text{filter}(\mathcal{I}, L)$  // sharpening using laplace filter
 $\text{binarize}(\mathcal{I}, \text{"Otsu"})$  // binarize (e.g. Otsu's thresholding)
 $\text{distanceTransform}(\mathcal{I})$  // distance to next background pixel
 $M \leftarrow \text{local minima}$ 
 $W \leftarrow \text{local maxima}$ 
 $\text{applyMask}(\text{input}, W)$  // Add region borders to original image
for all  $\text{marker} \in M$  do
     $\text{floodFill}(\text{input}, \text{marker})$ 
end for
```

3.2 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. Weiterhin können so Reflektionen und spiegelnde Oberflächen wie auch Hintergrundrauschen gemittelt werden, sodass die ermittelten Werte weitgehend robust sind.

In vielen Fällen ist die zugehörige Aufnahmeposition nicht bekannt, unzureichend genau oder es liegt kein kalibriertes System vor. Aus diesem Grund müssen die erfassten Punktwolken erst in ein globales Koordinatensystem gebracht werden. Diesen Prozess nennt man Registrierung.

3.2.1 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 [33–35]. Die in dieser Arbeit verwendete Methode ist Super4PCS [36], eine optimierte Version von Four-Points Congruent Sets (4PCS) [37]. Dieser Ansatz wird verwendet, da die vorhandene Implementierung Open Global Registration (OpenGR) [38] über einen Wrapper bereits zur Point Cloud Library (PCL) kompatibel ist.

4PCS basiert auf der Erkenntnis, dass die Verhältnisse der Distanzen zwischen Punkten bei einer affinen Transformation erhalten bleiben. Wählt man nun also vier koplanare Punkte (vgl. Definition 3), überschneiden sich mindestens zwei der Geraden, die die Eckpunkte verbinden. Da es in der Realität jedoch sehr unwahrscheinlich ist, vier koplanare Punkte zu finden, wird eine leichte Abweichung akzeptiert.

Definition 3 Seien $p, q, r, s \in \mathcal{P} \subseteq \mathbb{R}^3$. Dann heißen p, q, r, s *koplanar*, wenn $\vec{a} = \overline{pq}$,

3. Vorherige Arbeit

$\vec{b} = \overline{pr}$ und $\vec{c} = \overline{ps}$ linear abhängig sind, also eine Lösung für

$$\alpha \vec{a} + \beta \vec{b} + \gamma \vec{c} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \alpha, \beta, \gamma \in \mathbb{R} \setminus \{0\}$$

existiert.

Seien (p, q) und (r, s) nun die Paare, deren jeweilige Verbindungsgeraden sich im Punkt e schneiden. Die Verhältnisse $r_1 = \frac{\|p-e\|}{\|p-q\|}$ und $r_2 = \frac{\|r-e\|}{\|r-s\|}$ bleiben nach affiner Transformation T gleich. Da es eine solche (gesuchte) Transformation von \mathcal{P} nach \mathcal{S} geben muss, gilt [Gleichung 3.3](#).

$$\forall p, q, r, s \in \mathcal{P} \exists p', q', r', s' \in \mathcal{S} : r_1 = r'_1 \wedge r_2 = r'_2 \quad (3.3)$$

Also werden die Punkte p', q', r', s' mit den Verhältnissen $r'_1 = \frac{\|p'-e'\|}{\|p'-q'\|}$ und $r'_2 = \frac{\|r'-e'\|}{\|r'-s'\|}$ gesucht, indem $d_1 = r'_1 - r_1$ und $d_2 = r'_2 - r_2$ minimiert werden. Weiterhin muss $\|\mathcal{P} * T - \mathcal{S}\| < \delta$ für die gefundene Transformation T gelten, also der [MSE](#) unter einem Grenzwert δ liegen.

3.2.2 Lokale Registrierung

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

3.2.2.1 Iterative Closest Point

[ICP](#) [39] ist der wohl bekannteste Algorithmus zur lokalen Registrierung von Punktwolken. Im Laufe der Zeit wurden zahlreiche Varianten und Optimierungen dafür entwickelt, die beispielsweise auch die Normalen in den Punkten in Betracht ziehen [40], besonders resistent gegen statistische Ausreißer sind [41] oder die Effizienz des Algorithmus

3. Vorherige Arbeit

erhöhen [42]. Unter anderem existiert auch eine vorhandene Implementierung in der PCL [43].

Algorithmus 4 beschreibt, wie Punktwolke \mathcal{P} an Punktwolke \mathcal{Q} registriert werden kann. Der Ablauf lässt sich folgendermaßen grob erklären:

1. \mathcal{P} wird iterativ durch Anwendung von Translationen und Rotationen in eine neue Position gebracht, zum Beispiel mithilfe von Quaternionen [44].
2. Zu jedem Punkt $p \in \mathcal{P}$ wird der Punkt $q \in \mathcal{Q}$ gesucht, der den geringsten Abstand $d = \|p - q\|$ von p hat.
3. Berechne den Fehler $\varepsilon = \sum d^2$, Standard ist der MSE.
4. Wiederholung, bis eine der Abbruchbedingungen eintritt.

Die Abbruchbedingungen sind dabei je nach Implementierung unterschiedlich. Meist terminiert der Algorithmus jedoch, wenn

- $\varepsilon < \alpha$ mit Grenzwert $\alpha > 0$ für den Restfehler
- $\Delta d > \delta$, mit $\delta > 0$ als Grenzwert für die Änderung des Fehlers
- eine festgelegte Zahl von β Iterationen abgelaufen ist.

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

Algorithm 4 Iterative Closest Point [45, S. 56]

```
 $C_{\mathcal{P}} \leftarrow \text{getCenterOfMass}(\mathcal{P})$ 
 $C_{\mathcal{Q}} \leftarrow \text{getCenterOfMass}(\mathcal{Q})$ 
 $\text{translate}(C_{\mathcal{P}} \text{ into } C_{\mathcal{Q}})$  // align cloud position by center of mass
while  $\sum \| \mathcal{P} - \mathcal{Q} \|^2 > \alpha \wedge n < \beta$  do
  for all  $p_i \in \mathcal{P}$  do // find tuples  $(p_i, s_i)$  for alignment
     $s_i \leftarrow \text{nearestNeighbor}(p_i, \mathcal{Q})$ 
  end for
   $T \leftarrow \min_T \sum_{k=0}^i \|q_k - T(p_k)\|^2$  // translation that minimizes MSE
   $\mathcal{P} \leftarrow \text{applyTransformation}(T, \mathcal{P})$ 
   $n \leftarrow n + 1$ 
end while
```

3.2.2.2 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 [46, 1:55]. Die Kinect ist eine 3D-Kamera, welche ursprünglich für die Nutzung im Entertainment- und Gamingbereich entwickelt worden ist. Bestehend aus einem Lichtemitter, einem Infrarotsensor, einer 2D-RGB-Kamera und weiteren Sensoren, liefert sie 3D-Daten bei einer Bildwiederholfrequenz von 30 Hz. Neben dem ursprünglichen Anwendungsgebiet findet sie heute auch in vielen anderen, auch wissenschaftlichen Bereichen Anwendung, beispielsweise bei der Aufzeichnung geomorphologischer Daten [47].

Kinect Fusion (**KinFu**) ist das Ergebnis einer Forschungsarbeit von Microsoft Research [48, 49] 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. Normalerweise enthält dieses Informationen über die Daten im betreffenden Voxel. Hier werden jedoch mithilfe einer signed distance function die Entfernungen zur nächsten Oberfläche gespeichert [50]. Somit stehen innerhalb des Objekts negative Werte in den Voxeln und außerhalb positive Werte. Die Oberfläche wird damit implizit durch den Nulldurchgang der Distanzen in den Voxeln definiert.

Dies ermöglicht bei einer relativ geringen Auflösung dennoch eine recht genaue Rekonstruktion der vorhandenen Oberflächen. Um dies zu erreichen, werden die Distanzen in den Voxeln interpoliert und Fehler durch die geringe Auflösung oder Sensorrauschen minimiert.

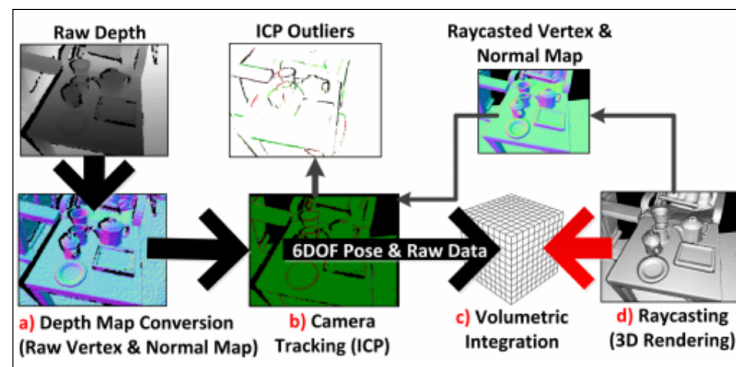


Abbildung 3.1: Integration eines Tiefenbilds in **KinFu**. Entnommen aus [48]

3. Vorherige Arbeit

Durch eine schnelle GPU-Implementierung auf Basis von CUDA erreicht **KinFu** dabei eine Integration neuer Tiefenbilder bei 30 Hz, der Bildwiederholffrequenz der Kinect. 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 [51]. Eine optimierte Version ist beispielsweise Chisel [52], bei der 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 Robot Operating System (**ROS**) angepasste Version von **KinFu**, um Trajectory Waypoints für die Robotersteuerung zu errechnen [53]. 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 werden, 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.

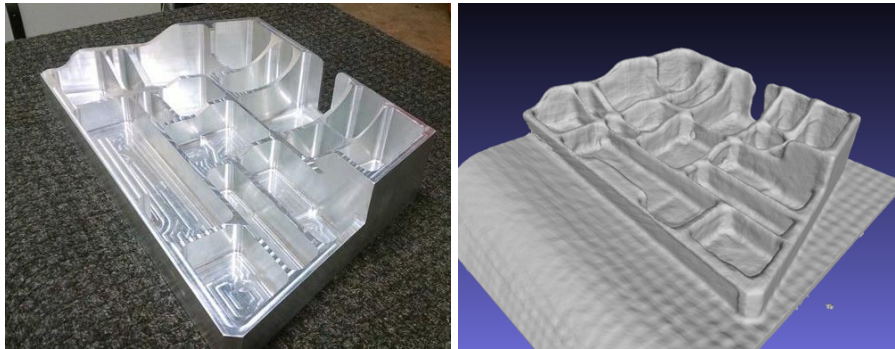


Abbildung 3.2: Szene und Rekonstruktion eines reflektierenden Objekts durch YAK. Entnommen aus [53]

Der Vorteil der Echtzeitintegration ist jedoch gleichzeitig auch ein großes Problem. In industriellen Robotikanlagen sind nur selten Grafikkarten verbaut. Eine Hochleistungs-GPU-Einheit ist dabei aber unerlässlich, da sonst nur ein Bruchteil der Geschwindigkeit erreicht werden kann. Weiterhin ist CUDA ein proprietäres Softwarepaket, was die Nutzung mit Grafikkarten von anderen Herstellern unmöglich macht.

3.3 Triangulierung

Die Konvertierung von punktbasierten Objektinformationen zu einem aus zusammenhängenden Dreiecken bestehenden Mesh nennt sich Triangulierung. Gerade bei der Triangulierung von Punktwolken gibt es jedoch nicht immer eine einzelne korrekte Lösung dieses Problems. Es ist beispielweise 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 [17] handelt es sich um einen Algorithmus, der ursprünglich zur Triangulierung medizinischer Computertomographiescans entwickelt wurde. Diese werden in den meisten Fällen als Voxel Grid dargestellt. Für die Anwendung auf einer Punktwolke muss diese also zunächst konvertiert werden. Der Ablauf ist in [Algorithmus 5](#) beschrieben.

Das Ziel des Algorithmus ist es, die sogenannte Isofläche aus den 3D-Daten zu extrahieren, welche dann durch die einzelnen Dreiecke approximiert wird. Die Isofläche repräsentiert im Falle einer Punktwolke die Oberfläche eines Objekts. Die Punkte liegen also bei fehlerfreien Daten exakt darauf. Um festzulegen, ab wann ein Punkt innerhalb der Oberfläche gesehen wird, muss sein Wert unter dem Parameter α liegen.

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 ermittelt. Dieses Vorgehen wird nun für jeden Voxel wiederholt, was zu einer vollständigen Rekonstruktion der Oberfläche führt.

Algorithm 5 Marching Cubes [17, Abs. 4]

```
 $\mathcal{L} \leftarrow \emptyset$  // The list of output vertices (Faces implicit)
 $\mathcal{V} \leftarrow Voxels$ 
for all  $v \in \mathcal{V}$  do
     $i \leftarrow calculateIndex(v)$ 
     $\mathcal{E} \leftarrow edgeTable[i]$ 
    for all  $e \in \mathcal{E}$  do
         $p \leftarrow interpolate(e[0], e[1])$  // interpolate corners
         $\mathcal{L}.add(p)$ 
    end for
end for
return  $\mathcal{L}$ 
function  $calculateIndex(v)$ 
     $x \leftarrow 0$ 
    for  $corner \in \{0, \dots, 7\}$  do
        if  $v[corner] < \alpha$  then // corner inside isosurface
             $x \mid = 1 \ll corner$  // Set the i-th bit to 1
        end if
    end for
    return  $x$ 
end function
```

3. Vorherige Arbeit

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. Insgesamt kann bei Marching Cubes die Topologie des Objekts verfälscht dargestellt werden [54, S. 2].

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

Die 2006 von Kazhdan, Bolitho und Hoppe präsentierte Poisson-Rekonstruktion [55] basiert auf der impliziten Darstellung der Oberfläche als Funktion. Die mathematische Darstellung dieser Funktion lässt sich Definition 4 entnehmen.

Definition 4 Sei M die zu rekonstruierende Oberfläche. Dann ist

$$\chi_M(p) = \begin{cases} 0 & p \text{ außerhalb von } M \\ 1 & p \text{ innerhalb von } M \end{cases}, \quad \forall q \in M : \chi_M(q) = 0.5 \quad (3.4)$$

die implizite Funktion, die die Oberfläche im Punkt $p \in M$ darstellt.

Der Gradient $\nabla\chi_M$ dieser Funktion nimmt also im optimalen Fall überall den Wert 0 an, außer in der Oberfläche selbst. Dieser Gradient entspricht exakt den nach innen gerichteten Normalen in den Oberflächenpunkten [55, Abs. 3]. Somit gilt für die Indikatorfunktion:

$$\begin{aligned} \nabla\chi_M &= \vec{V} \\ \Leftrightarrow \nabla^2\chi_M &= \nabla\vec{V} \\ \Leftrightarrow \Delta\chi_M &= \nabla\vec{V} \end{aligned}$$

3. Vorherige Arbeit

Dies ist eine Poisson-Gleichung [56, Kap. 2], deren Lösung sich beispielsweise mit dem Gauß-Seidel-Verfahren iterativ berechnen beziehungsweise annähern lässt. Da eine genaue Lösung der Poissongleichung nur in der Nähe der Oberfläche erforderlich ist, kann nun ein Octree aufgestellt werden.

Mit der Poisson-Rekonstruktion generierte Polygonmeshes weisen die Eigenschaft auf, immer wasserdicht zu sein. Dies ist in der Darstellung der Oberfläche als mehrdimensionale, kontinuierliche Funktion begründet.

Die Auflösung des Meshes lässt sich durch die maximale Baumtiefe einstellen. Mit der Tiefe und damit der Qualität steigen jedoch auch Laufzeit und Speicherbedarf drastisch an. Weiterhin kann konfiguriert werden, wie viele Punkte einem Blatt im Baum zugewiesen werden. Hohe Werte führen zu einer verringerten Auflösung, aber auch zu einem gleichzeitigen Glättungseffekt. Außerdem verringern sich aufgrund der niedrigeren Zahl an Blättern der benötigte Speicher und die Laufzeit.

3.3.3 Weitere Ansätze

Neben den bereits genannten Rekonstruktionsalgorithmen Marching Cubes und Poisson gibt es noch weitere unterschiedliche Ideen, um ein Polygonnetz zu triangulieren.

3.3.3.1 Greedy-Triangulierung

Bei der Greedy-Triangulierung der Wolke \mathcal{P} werden zunächst alle Punkte $p, q \in \mathcal{P}$ miteinander verknüpft. So entstehen Kantenkandidaten, welche nun nach ihrer Länge sortiert werden. Diese werden nun nacheinander akzeptiert oder verworfen, wenn ein Schnittpunkt mit einer bereits akzeptierten Kante auftritt [57, S. 235–237].

3. Vorherige Arbeit

Algorithm 6 Greedy-Triangulierung

```
 $\mathcal{C} \leftarrow \emptyset$ 
for all  $p, q \in \mathcal{P}, p \neq q$  do           // add edge candidates to  $\mathcal{C}$ 
     $\text{add}(\mathcal{C}, \overline{pq})$ 
end for
 $\text{sortByLength}(\mathcal{C})$ 
 $\mathcal{T} \leftarrow \emptyset$            // the accepted edges
for all  $e \in \mathcal{C}$  do
    for all  $f \in \mathcal{T}$  do
        if  $e \cap f$  then           //  $e$  intersects  $f$ , remove from candidates
             $\mathcal{C} \leftarrow \mathcal{C} \setminus \{e\}$ 
        end if
    end for
     $\mathcal{T} \leftarrow \mathcal{T} \cup \{e\}$ 
end for
return  $\mathcal{T}$ 
```

Diese Art der Triangulierung funktioniert nur im zweidimensionalen Raum, da der Schnittpunkt zwischen zwei Geraden in höheren Dimensionen nicht als Kriterium ausreicht. Im \mathbb{R}^3 werden daher die Punkte über ihre Normale auf eine Ebene projiziert. Zur Anwendung auf 3D-Punktwolken ist daher eine vorherige Normalenschätzung notwendig. Anschließend kann der in [Algorithmus 6](#) dargestellte Ablauf regulär durchgeführt werden.

Ein enormer Vorteil der Greedy-Triangulierung ist ihre Geschwindigkeit. So zeigen Berg, Cheong, Kreveld und Overmars, dass ein Polygon mit n Vertices in $\mathcal{O}(n \log n)$ rekonstruiert werden kann [20, S. 56–58]. Dies ist beispielsweise für Echtzeitanwendungen relevant, wie die in [58] vorgestellte Methode demonstriert.

3.3.3.2 Delaunay-Triangulierung

Dreiecke mit mindestens einem kleinen Innenwinkel sind häufig problematisch in der Computergrafik. Beim Renderingprozess können aufgrund von Präzisions- und Rundungsfehlern Artefakte und Aliasingeffekte auftreten.

Die Delaunay-Triangulierung behebt dieses Problem weitestgehend, da sie die minimalen Innenwinkel der generierten Dreiecke maximiert [20, S. 199]. Eine weitere Eigenschaft ist, dass kein Umkreis eines der Dreiecke einen Knoten beinhaltet, der nicht eine Ecke des Dreiecks ist [20, S. 196]. Der Ablauf der Delaunay-Triangulierung ist in [Algorithmus 7](#) zu sehen.

Algorithm 7 Delaunay-Triangulierung [20, S. 200–201]

```
 $\mathcal{P} \leftarrow$  list of  $n$  vertices  $p_1, \dots, p_n$   
 $\mathcal{T} \leftarrow$  a triangle  $p_0qr$  with  $p_0 \in \mathcal{P}$  so that  $\forall p \in \mathcal{P} : p$  inside  $p_0qr$   
 $\mathcal{P} \leftarrow \mathcal{P} \setminus \{p_0\}$   
while  $\mathcal{P} \neq \emptyset$  do  
   $p_r \leftarrow \mathcal{P}.\text{pop}()$   
  find triangle  $t \in \mathcal{T}$  so that  $p_r$  is inside  $t$   
   $\mathcal{T} \leftarrow \mathcal{T} \setminus \{t\}$   
  if  $p_r$  is on edge  $t_1t_2 \in t$  then  
    find vertices  $v_1$  and  $v_2$  of the two triangles separated by  $t_1t_2$   
    create triangles  $t_1p_rv_1$ ,  $t_1v_2p_r$ ,  $t_2v_1p_r$  and  $t_2p_rv_2$   
    check the untouched edges using  $\text{checkFlipped}(p_r, \text{edge})$   
    add the triangles to  $\mathcal{T}$   
  else  
     $p_r$  is inside  $t = t_1t_2t_3$   
    split  $t$  into three triangles with  $p_r$  as vertex  
    check the old edges using  $\text{checkFlipped}(p_r, \text{edge})$   
    add them to  $\mathcal{T}$   
  end if  
  remove all triangles  $t \in \mathcal{T}$  which have  $q$  or  $r$  as vertex  
end while  
return  $\mathcal{T}$   
function  $\text{checkFlipped}(\text{new vertex } p, \text{edge } t)$   
  if the smallest angle is bigger after a flip then  
    get two triangles  $a, b \in \mathcal{T}$  separated by  $t$   
    connect vertices  $p_1 \in a$  and  $p_2 \in b$  not connected by  $t$   
    remove  $t$   
    check old edges not ending at  $p$  using  $\text{checkFlipped}(p, \text{edge})$   
  end if  
end function
```

3.3.3.3 Ball Pivoting

Die Ball-Pivoting-Rekonstruktion liefert eine sehr anschauliche und intuitive Möglichkeit, eine Punktwolke zu triangulieren. Dabei wird ein imaginärer Ball mit Radius ρ auf ein Seed-Dreieck gelegt und nacheinander über die Kanten gerollt [59]. Trifft er dabei auf einen weiteren Punkt, wird aus diesem und der bereits existierenden Kante ein neues Dreieck gebildet. Falls er auf keinen Punkt trifft, handelt es sich um eine Außenkante, die nicht weiter betrachtet wird. Der genaue Ablauf ist in [Algorithmus 8](#) dargestellt.

Algorithm 8 Ball Pivoting [59, Abs. 4]

```
 $\rho \leftarrow$  the ball radius  
 $\mathcal{F} \leftarrow$  seed triangle  
while true do  
  while  $e_{(i,j)} \leftarrow \text{getActiveEdge}(\mathcal{F})$  do  
    if  $\sigma_k \leftarrow \text{ballPivot}(e_{(i,j)}) \wedge (\text{notUsed}(\sigma_k) \vee \text{onFront}(\sigma_k))$  then  
      outputTriangle( $\sigma_i, \sigma_k, \sigma_j$ )  
      join( $e_{(i,j)}, \sigma_k, \mathcal{F}$ )  
      if  $e_{(k,i)} \in \mathcal{F}$  then  
        glue( $e_{(i,k)}, e_{(k,i)}, \mathcal{F}$ )  
      end if  
      if  $e_{(j,k)} \in \mathcal{F}$  then  
        glue( $e_{(k,j)}, e_{(j,k)}, \mathcal{F}$ )  
      end if  
    else  
      markAsBoundary( $e_{(i,j)}$ )  
    end if  
  end while  
  if  $(\sigma_i, \sigma_j, \sigma_k) \leftarrow \text{findSeedTriangle}()$  then  
    outputTriangle( $\sigma_i, \sigma_j, \sigma_k$ )  
    insertEdge( $e_{(i,j)}, \mathcal{F}$ )  
    insertEdge( $e_{(j,k)}, \mathcal{F}$ )  
    insertEdge( $e_{(k,i)}, \mathcal{F}$ )  
  else  
    return  
  end if  
end while
```

Ein Nachteil des Ball-Pivoting-Algorithmus ist, dass bei zu groß gewählten Radii einige Details verloren gehen können. Dies tritt insbesondere dann auf, wenn die rekonstruierte Oberfläche stark konkav ist und die Abstände zwischen den Punkten deutlich kleiner als der gewählte Radius sind.

Um diesem Fall vorzubeugen, kann man mehrere Triangulierungen mit unterschiedlichen Werten für ρ kombinieren. Aufgrund der mehrfachen Ausführung führt dies jedoch zu einer deutlich erhöhten Laufzeit.

4. Implementierung

Nachdem nun die theoretischen Grundlagen gelegt sind, wird im Folgenden die Implementierung der Pipeline erklärt. Zunächst wird in 4.1 der Aufbau des Programms beschrieben. Die Interaktion und Parametrisierung wird in 4.2 erläutert. In 4.3 wird dann der gesamte Ablauf der Pipeline und das Zusammenspiel der verschiedenen Komponenten noch einmal zusammengefasst.

4.1 Aufbau

Die Pipeline ist in C++ auf Basis der PCL [60] geschrieben. Dabei handelt es sich um eine Open-Source-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 [61] kompatibel ist. Die Kompatibilität wird durch ROS-Nodelets im Paket `perception_pcl` hergestellt, sodass Punktwolken, Meshes oder andere Datenstrukturen direkt zwischen ROS-basierten Komponenten versendet werden können.

Um eine modulare Verwendung des Tools zu gewährleisten wurde ein einzelnes ROS-Paket erstellt. Die 3D-Datenverarbeitung findet hauptsächlich auf Basis der PCL statt, mit einer Grobregistrierung über OpenGR. Zur 2D-Segmentierung über den Watershed-Algorithmus wird OpenCV verwendet.

Weiterhin werden eine aktuelle Version des Ensenso SDKs und die ROS-Serviceheader zu den von Mikado versendeten Nachrichten benötigt. Mit ihrer Hilfe können die aktuellen 2D-Grauwertbilder ausgelesen und zur Segmentierung verwendet werden.

4.2 Interaktion und Parametrisierung

Die Interaktion zwischen Mikado und dem Rekonstruktionstool findet über **ROS** statt. Dabei handelt es sich um einen `ros::Subscriber` auf das von Mikado veröffentlichte Topic `/point_cloud`. Bei Aufnahme einer neuen Punktwolke wird somit das Durchlaufen der Pipeline automatisch angestoßen.

Weiterhin lassen sich über `rosparam set` und `rosparam get` Parameter setzen oder auslesen. Diese werden dann im Programm geparkt und – falls nicht vorhanden – durch Standardwerte ersetzt. Dies bietet mehrere Vorteile gegenüber anderen Optionen:

- Parameter können optional über die Konsole, auf Programmebene oder auch gar nicht gesetzt werden.
- Das Programm muss nicht für jede Konfiguration neu kompiliert werden und kann bei Laufzeit parametrisiert werden.
- Die Interaktion mit anderen Programmen ist durch die Verwendung der **ROS**-Schnittstelle erheblich vereinfacht.

Auch die Triangulation wird durch einen externen Trigger angestoßen. Dieser wird als `rosservice` zur Verfügung gestellt, welcher mit dem Output-Dateipfad aufgerufen wird. Auch hier ist durch die Verwendung der **ROS**-basierten Schnittstelle eine hohe Flexibilität gegeben.

4.3 Pipelineablauf

Der Ablauf der Pipeline ist stark abhängig von der durch den Nutzer gewählten Konfiguration. Diese kann jedoch auch nach Start des Programms noch geändert werden. Ohne vorherige Wahl eines Modus werden weder Segmentierung noch Registrierung vorgenommen.

4. Implementierung

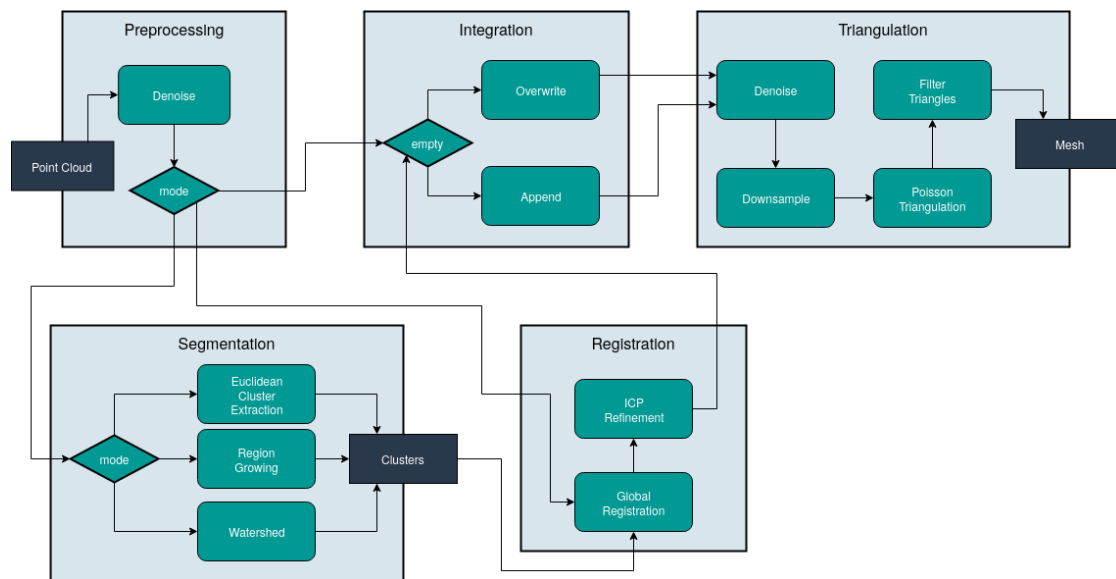


Abbildung 4.1: Ablauf der Pipeline

Der gesamte Ablauf des Prozesses vom Empfang einer Punktwolke bis zur Ausgabe des rekonstruierten Meshs ist in [Abbildung 4.1](#) sichtbar. Der Programmablauf kann in mehrere verschiedene Blöcke unterteilt werden, welche im Folgenden erläutert werden.

4.3.1 Preprocessing

Das Preprocessing findet unabhängig von der Wahl des Modus statt. Nach Empfang einer Punktwolke wird diese zunächst durch ein `pcl::StatisticalOutlierRemoval` gefiltert. Das Ziel ist die Eliminierung von unvermeidbarem Sensorrauschen.

In diesem Schritt könnten auch noch andere Preprocessingmethoden durchgeführt werden. Dabei kann es sich sowohl um ein Downsampling zur Datenreduktion handeln als auch um Prozesse wie eine Normalenschätzung, falls diese nicht bereits vorher stattgefunden hat.

4.3.2 Segmentierung

Dieser Block wird nur ausgeführt, wenn der entsprechende Modus gewählt ist. In diesem Schritt findet die Segmentierung der Punktwolke statt. Je nach gewählter Methode lässt sich dabei zwischen **ECE**, **RG** und Watershed-Segmentierung wählen. Die Punktwolke wird dabei in einen Vektor gespalten, der die einzelnen Cluster beinhaltet.

Im 2D-Bild der Kamera sind Daten vorhanden, die unter Umständen im 3D bereits durch eine **RoI** herausgefiltert wurden. Dies kann bei einer Segmentierung mithilfe des Watershed-Algorithmus zu verschlechterten Ergebnissen führen. Um bereits gefilterte Regionen also nicht zu beachten, wird den entsprechenden Pixeln im 2D-Bild zunächst der Grauwert 0 zugewiesen. Weiterhin kann eine algorithmische Kontrasterhöhung durchgeführt werden.

4.3.3 Registrierung

Hier wird die Eingabepunktwolke an der bereits integrierten Punktwolke registriert. Dabei wird zunächst eine globale Schätzung mithilfe von **4PCS** durchgeführt. Diese Registrierung wird anschließend durch **ICP** verfeinert.

Der Registrierungsprozess wird abgebrochen, falls einer der beiden Registrierungen nicht konvergiert. Weiterhin kann ein Grenzwert gewählt werden, unter dem der Score liegen muss. Bei **ICP** beschreibt dieser Score den **MSE**.

Die Registrierung wird durch Parametrisierung aktiviert bzw. deaktiviert. Hat zuvor eine Segmentierung stattgefunden, wird jeder Cluster separat diesem Block zugeführt.

4.3.4 Integration

Dieser Schritt beschreibt das Zusammenführen der beiden Punktwolken. Es wird vorausgesetzt, dass vorher bereits eine Registrierung stattgefunden hat.

In der Praxis ist dieser Block sehr einfach gehalten. Liegt die Zahl der Punkte unter ei-

4. Implementierung

nem Grenzwert, werden diese überschrieben. Ist dies nicht der Fall, müssen die beiden Vektoren konkateniert werden.

4.3.5 Triangulierung

Entspricht die zusammengeführte Punktwolke den Erwartungen des Nutzers, kann dieser nun die Triangulierung anstoßen. Dies bedeutet, dass dieser Block nicht automatisch ausgeführt wird. Der Grund dafür ist, dass es sich um einen sehr zeit-, speicher- und rechenintensiven Vorgang handelt.

Zunächst findet eine weitere Filterung statt. Dies ermöglicht selbst dann ein brauchbares Ergebnis, falls im Registrierungsschritt keine fehlerfreie Ausrichtung erreicht worden ist. Im Optimalfall werden die fehlerhaften Daten aufgrund der hohen Punktdichte an der Oberfläche somit größtenteils entfernt.

Nach dem anschließenden Downsampling findet eine Poisson-Triangulierung statt. Da die Poisson-Triangulierung immer wasserdichte Polygone erzeugt, können hierbei auch inkorrekte Dreiecke generiert werden. Im letzten Schritt werden daher nun Dreiecke aus dem Polygonnetz entfernt, die eine festgelegte Distanz von der Punktwolke überschreiten.

5. Auswertung

In diesem Teil der Arbeit werden die Ergebnisse der Pipeline anhand von Referenzdaten validiert und bewertet. Nach der Evaluation in 5.1 werden mögliche Fehlerquellen in 5.2 genannt. In 5.3 folgt ein Ausblick auf denkbare Erweiterungen in der Zukunft.

5.1 Evaluation der Daten

Zur Evaluation der Pipeline wurden verschiedene Teile rekonstruiert und die Ergebnisse mit vorhandenen CAD-Modellen verglichen. Zunächst wurden mit einer Kamera am Roboterarm bereits registrierte Punktwolken aufgenommen und zusammengeführt. Anschließend wurden mit einer fest montierten Kamera mehrere Aufnahmen eines Teils gemacht, zwischen denen dieses leicht rotiert wurde. Abschließend wurden mit der unbeweglichen Kamera jeweils zwei Aufnahmen einer Kiste gemacht, in der sich mehrere Teile gleicher Art befanden. Im ersten Fall waren diese räumlich gut getrennt, im zweiten Fall handelte es sich um eine volle Kiste.

Da die generierte Rekonstruktion nicht am Referenzmodell ausgerichtet ist, wurde zum Vergleich eine manuelle Grobregistrierung durchgeführt. Diese wurde anschließend durch eine lokale Registrierung durch ICP auf Basis der Vertices verfeinert.

5.1.1 Werte

Als Qualitätsmetrik wurde die Cloud-Mesh-Distanz zwischen Rekonstruktion und Ground Truth gewählt. Dabei wird die durchschnittliche Distanz zwischen Vertices und den

5. Auswertung

nächsten Faces des Vergleichobjekts betrachtet.

Distanzen [mm]	Rad	Zahnrad	Gehäuse	T-Stücke
Fix	0.855	0.462	0.364	0.486
Rotation	0.921	0.558	0.751	0.333
Seg. 1 ECE	0.527	0.318	0.238	0.294
Seg. 1 RG	0.842	0.331	0.819	0.316
Seg. 1 Watershed	0.586	0.315	0.756	0.305
Seg. 2 ECE	1.322	0.393	0.549	0.488
Seg. 2 RG	1.089	0.383	0.491	0.326
Seg. 2 Watershed	0.703	0.411	0.701	0.297

Tabelle 5.1: Cloud-Mesh-Distanzen von Rekonstruktion zu Ground Truth

In [Tabelle 5.1](#) sind die Distanzen von den Vertices der Rekonstruktion zu den Faces des Referenzmodells zu sehen. Diese können als Qualitätsmaß für die Regionen angesehen werden, in denen eine Oberfläche rekonstruiert wurde.

Bei Betrachtung der Werte fällt auf, dass die Segmentierung mit [ECE](#) bei räumlich getrennten Objekten für alle getesteten Teile die geringste Distanz liefert. Weiterhin kann beobachtet werden, dass [ECE](#) bei geringer räumlicher Entfernung häufig zu einer höheren Distanz führt.

Die Segmentierung führt im Fall einer vollen Kiste außerdem je nach Objekt zu einer Verringerung oder Erhöhung der Distanz. [ECE](#) liefert beispielsweise für das Einkaufswagenrad sehr hohe Werte, während beim Zahnrad geringere Entfernungen als die Rekonstruktion über die mobile Kamera erzielt werden können. Bei Wahl der Watershed-Segmentierung sind die Distanzen beim Rad jedoch deutlich niedriger.

Distanzen [mm]	Rad	Zahnrad	Gehäuse	T-Stücke
Fix	10.414	2.468	1.512	2.398
Rotation	10.184	2.513	2.351	3.224
Seg. 1 ECE	11.097	2.734	2.692	3.840
Seg. 1 RG	18.414	2.963	5.831	3.828
Seg. 1 Watershed	25.140	7.005	6.205	4.449
Seg. 2 ECE	19.703	3.253	7.711	3.656
Seg. 2 RG	19.744	3.084	4.675	3.679
Seg. 2 Watershed	16.056	4.647	3.536	5.071

Tabelle 5.2: Cloud-Mesh-Distanzen von Ground Truth zu Rekonstruktion

[Tabelle 5.2](#) zeigt die Distanz von den Vertices des Referenzobjekts zu den Faces der

5. Auswertung

Rekonstruktion. Diese ist insbesondere in jenen Regionen sehr hoch, in denen Lücken in der Rekonstruktion vorhanden sind.

Distanzverhältnis	Rad	Zahnrad	Gehäuse	T-Stücke
Fix	0.08	0.19	0.24	0.20
Rotation	0.09	0.22	0.32	0.10
Seg. 1 ECE	0.05	0.12	0.09	0.08
Seg. 1 RG	0.05	0.11	0.14	0.08
Seg. 1 Watershed	0.02	0.04	0.12	0.07
Seg. 2 ECE	0.07	0.12	0.07	0.13
Seg. 2 RG	0.06	0.12	0.11	0.09
Seg. 2 Watershed	0.04	0.09	0.20	0.06

Tabelle 5.3: Verhältnis der Distanzen in [Tabelle 5.1](#) und [Tabelle 5.2](#)

Das Verhältnis der beiden Distanzmetriken ist in [Tabelle 5.3](#) sichtbar. Im optimalen Fall, also einer vollständigen Rekonstruktion mit perfekter Qualität, konvergiert dieses Verhältnis gegen 1, da dann die Distanz in beide Richtungen etwa gleich ist. Ein höherer Wert steht hier also für eine höhere Vollständigkeit des rekonstruierten Meshs.

Aus den Werten erschließt sich, dass in den meisten Testfällen die Rotation das Mesh mit der höchsten Vollständigkeit geliefert hat. Der Watershed-Algorithmus konnte im Vergleich mit den anderen Tests bei geringer räumlicher Distanz zwischen den Objekten nur den geringsten Teil der Oberfläche herstellen.

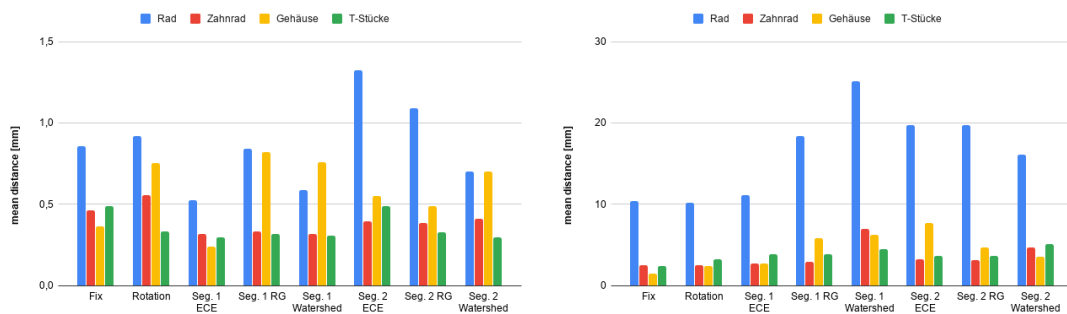


Abbildung 5.1: Diagramme der Werte in [Tabelle 5.1](#) und [Tabelle 5.2](#)

5.1.2 Bewertung

Es zeigte sich, dass die Distanz von Rekonstruktion zu Ground Truth sich je nach Objekt bei unterschiedlichen Segmentierungen erhöht bzw. verringert. Dies lässt darauf schließen, dass die Wahl der Segmentierungsmethode einen erheblichen Einfluss auf die Qualität der Rekonstruktion hat. Jedoch gibt es offensichtlich nicht einen optimalen Ansatz, der in allen Fällen die besten Ergebnisse liefert.

Weiterhin wurde klar, dass die Rekonstruktion über Rotation eines einzelnen Objekts in den meisten Fällen die höchste Vollständigkeit liefert. Dies lässt auf schlechte Ergebnisse bei der Segmentierung schließen, da die Objekte hier aus deutlich mehr Perspektiven zu sehen sein müssten.

Für die schlechten Ergebnisse bei der Segmentierung gibt es verschiedene Erklärungen. Ist die Szene untersegmentiert, also nicht zusammengehörende Objekte noch in einem Cluster, kann das entsprechende Segment nicht integriert werden. Die Information der darin enthaltenen Objekte gehen somit verloren. Ein anderer Fall ist die Übersegmentierung der Szene. Hier kann es vorkommen, dass nur wenige Punkte in einem Cluster enthalten sind und die Registrierung des Objekts somit schwierig ist. Unter Umständen kann keine eindeutige Position gefunden werden und somit gehen wieder wertvolle Informationen verloren.

5.2 Fehlerbetrachtung

Bei Aufnahme der Testdatensätze können mögliche Fehler entstanden sein, die das Ergebnis der Evaluation verfälschen. Die folgende Liste stellt einen Auszug verschiedener Fehlerquellen dar.

- unvermeidbares Sensorrauschen
- Spiegelungen durch einfallendes Sonnenlicht
- Ground Truth CAD-Modelle passen nicht zu tatsächlich vorhandenen Objekten
- tatsächlich vorhandene Objekte sind nicht exakt gleich
- unbeabsichtigte Bewegung des Objekts bei Rekonstruktion über Roboter
- fehlerhafte Ausrichtung zwischen generiertem Mesh und Referenz

5.3 Zukünftige Arbeit

In der **Bewertung** hat sich gezeigt, dass die Qualität des rekonstruierten Meshs abhängig vom Erfolg der Segmentierung ist. Hier ließe sich eine deutliche Verbesserung erzielen, indem intelligentere Verfahren eingesetzt würden. Eine Option wäre ein künstliches neuronales Netz, welches speziell auf industrielle Bauteile trainiert ist.

Eine andere Erweiterungsmöglichkeit bietet die Triangulation. Hier könnten die Verschattungen der Objekte genutzt werden, um die Qualität der Rekonstruktion zu verbessern. Diesen Ansatz verfolgen beispielsweise auch Kazhdan, Chuang, Rusinkiewicz und Hoppe in [62].

Weiterhin wäre es eine Möglichkeit, Informationen durch Nutzereingaben zu erlangen. Dies könnte sowohl den Bereich der Segmentierung als auch der Registrierung umfassen.

Im Moment können Symmetrien und Spiegelungen in den Objekten nur sehr eingeschränkt betrachtet werden. Würden diese berücksichtigt, könnte eine deutlich vollständigeres Mesh generiert werden.'

6. Zusammenfassung

Es wurde eine Pipeline zur Rekonstruktion von 3D-Polygonnetzen aus Punktwolken entwickelt.

Literatur

- [1] B. H. Drost und M. Ulrich, *Recognition and pose determination of 3D objects in 3D scenes*, US Patent 8,830,229, Sep. 2014.
- [2] *Mikado - 3D Bin Picking System for all applications*. Adresse: <https://www.mikado-robotics.com> (besucht am 30. 03. 2020).
- [3] *Isys Vision GmbH*. Adresse: <https://www.isys-vision.de/> (besucht am 30. 03. 2020).
- [4] *Ensenso - Stereo 3D Cameras*. Adresse: <https://www.ensenso.com> (besucht am 30. 03. 2020).
- [5] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, Jg. 30, Nr. 2, S. 328–341, 2007.
- [6] C. Barnes, E. Shechtman, A. Finkelstein und D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," in *ACM Transactions on Graphics (ToG)*, ACM, Bd. 28, Aug. 2009, S. 24.
- [7] J. Geng, "Structured-light 3D surface imaging: a tutorial," *Advances in Optics and Photonics*, Jg. 3, Nr. 2, S. 128–160, 2011.
- [8] E. Catmull, "A subdivision algorithm for computer display of curved surfaces," Diss., 1974.
- [9] A. Arsalan Soltani, H. Huang, J. Wu, T. D. Kulkarni und J. B. Tenenbaum, "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.
- [10] A. Redert, R.-P. Berretty, C. Varekamp, O. Willemsen, J. Swillens und H. Driesen, "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.
- [11] K. Muller, P. Merkle und T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE*, Jg. 99, Nr. 4, S. 643–656, 2010.

- [12] J. Carr, "Surface reconstruction in 3D medical imaging," Diss., 1996.
- [13] S. Chmielewski und P. Tompalski, "Estimating outdoor advertising media visibility with voxel-based approach," *Applied Geography*, Jg. 87, S. 1–13, 2017.
- [14] *Downsampling a PointCloud using a VoxelGrid filter*. Adresse: https://pcl.readthedocs.io/projects/tutorials/en/latest/voxel_grid.html (besucht am 08. 07. 2020).
- [15] D. Meagher, "Efficient synthetic image generation of arbitrary 3-D objects," in *Proceeding of the IEEE Conference on Pattern Recognition and Image Processing*, Bd. 473, 1982.
- [16] Stanford University Computer Graphics Laboratory, *Stanford Bunny*. Adresse: <http://graphics.stanford.edu/data/3Dscanrep/> (besucht am 30. 04. 2020).
- [17] W. E. Lorensen und H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM siggraph computer graphics*, Jg. 21, Nr. 4, S. 163–169, 1987.
- [18] C. Smith, "On vertex-vertex systems and their use in geometric and biological modelling," Diss., 2006. doi: [10.11575/PRISM/439](https://doi.org/10.11575/PRISM/439).
- [19] B. 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.
- [20] M. de Berg, O. Cheong, M. van Kreveld und M. Overmars, *Computational Geometry: Algorithms and Applications*. Springer-Verlag, Berlin / Heidelberg / New York, 2008, S. 31–32, isbn: 978-3-540-77974-2. doi: [10.1007/978-3-540-77974-2](https://doi.org/10.1007/978-3-540-77974-2).
- [21] A. Ückermann, R. Haschke und H. 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.
- [22] A. Nguyen und B. Le, "3D point cloud segmentation: A survey," in *2013 6th IEEE conference on robotics, automation and mechatronics (RAM)*, IEEE, 2013, S. 225–230.
- [23] G. Te, W. Hu, A. Zheng und Z. Guo, "Rgcnn: Regularized graph cnn for point cloud segmentation," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, S. 746–754.
- [24] A. Shamir, "A survey on mesh segmentation techniques," in *Computer graphics forum*, Wiley Online Library, Bd. 27, 2008, S. 1539–1556.
- [25] L. Shapira, A. Shamir und D. Cohen-Or, "Consistent mesh partitioning and skeletonisation using the shape diameter function," *The Visual Computer*, Jg. 24, Nr. 4, S. 249, 2008.
- [26] R. B. Rusu, "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments," Diss., Computer Science department, Technische Universität München, Germany, Okt. 2009.

- [27] V Schomaker, J Waser, R. t. Marsh und G Bergman, "To fit a plane or a line to a set of points by least squares," *Acta crystallographica*, Jg. 12, Nr. 8, S. 600–604, 1959.
- [28] M. A. Fischler und R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Jg. 24, Nr. 6, S. 381–395, 1981.
- [29] T. Rabbani, F. Van Den Heuvel und G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *International archives of photogrammetry, remote sensing and spatial information sciences*, Jg. 36, Nr. 5, S. 248–253, 2006.
- [30] J. B. Roerdink und A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta informaticae*, Jg. 41, Nr. 1, 2, S. 187–228, 2000.
- [31] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Jg. 9, Nr. 1, S. 62–66, 1979.
- [32] OpenCV, *Image Segmentation with Distance Transform and Watershed Algorithm*, 24. Juni 2020. Adresse: https://docs.opencv.org/master/d2/dbd/tutorial_distance_transform.html (besucht am 10. 07. 2020).
- [33] K. N. Chaudhury, Y. Khoo und A. Singer, "Global registration of multiple point clouds using semidefinite programming," *SIAM Journal on Optimization*, Jg. 25, Nr. 1, S. 468–501, 2015.
- [34] Q.-Y. Zhou, J. Park und V. Koltun, "Fast global registration," in *European Conference on Computer Vision*, Springer, 2016, S. 766–782.
- [35] R. B. Rusu, N. Blodow und M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *2009 IEEE international conference on robotics and automation*, IEEE, 2009, S. 3212–3217.
- [36] N. Mellado, D. Aiger und N. J. Mitra, "Super 4PCS Fast Global Pointcloud Registration via Smart Indexing," *Computer Graphics Forum*, Jg. 33, Nr. 5, S. 205–215, 2014, issn: 1467-8659. doi: [10.1111/cgf.12446](https://doi.org/10.1111/cgf.12446). Adresse: <http://dx.doi.org/10.1111/cgf.12446>.
- [37] D. Aiger, N. J. Mitra und D. Cohen-Or, "4-points Congruent Sets for Robust Surface Registration," *ACM Transactions on Graphics*, Jg. 27, Nr. 3, #85, 1–10, 2008.
- [38] N. Mellado, *OpenGR: a 3D Global Registration Library*, 2018. Adresse: <https://github.com/STORM-IRIT/OpenGR> (besucht am 06. 07. 2020).
- [39] P. J. Besl und N. D. McKay, "Method for registration of 3-D shapes," in *Sensor fusion IV: control paradigms and data structures*, International Society for Optics und Photonics, Bd. 1611, 1992, S. 586–606.
- [40] D. Münch, B. Combès und S. Prima, "A modified ICP algorithm for normal-guided surface registration," in *Medical Imaging 2010: Image Processing*, International Society for Optics und Photonics, Bd. 7623, 2010, 76231A.

- [41] S. Bouaziz, A. Tagliasacchi und M. Pauly, "Sparse iterative closest point," in *Computer graphics forum*, Wiley Online Library, Bd. 32, 2013, S. 113–123.
- [42] S. Rusinkiewicz und M. Levoy, "Efficient variants of the ICP algorithm," in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, IEEE, 2001, S. 145–152.
- [43] D. Holz, A. Ichim, F. Tombari, R. Rusu und S. Behnke, "Registration with the Point Cloud Library - A Modular Framework for Aligning in 3-D," *IEEE Robotics & Automation Magazine*, Jg. 22, S. 110–124, Dez. 2015. doi: [10.1109/MRA.2015.2432331](https://doi.org/10.1109/MRA.2015.2432331).
- [44] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Josa a*, Jg. 4, Nr. 4, S. 629–642, 1987.
- [45] I. Elkhachy, "Towards an Automatic Registration for Terrestrial Laser Scanner Data," Diss., Feb. 2008. doi: [10.13140/RG.2.1.3547.7605](https://doi.org/10.13140/RG.2.1.3547.7605).
- [46] D. Kim, *28c3: KinectFusion*. Adresse: <https://www.youtube.com/watch?v=RvrCAw1IFG0> (besucht am 09.04.2020).
- [47] K. D. Mankoff und T. A. Russo, "The Kinect: a low-cost, high-resolution, short-range 3D camera," *Earth Surface Processes and Landforms*, Jg. 38, Nr. 9, S. 926–936, 2013.
- [48] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison 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.
- [49] M. Research, *KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera*. Adresse: <https://www.youtube.com/watch?v=KOUSS1KUJ-A> (besucht am 09.04.2020).
- [50] B. Curless und M. 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.
- [51] M. Pirovano, "KinFu—an open source implementation of Kinect Fusion+ case study: implementing a 3D scanner with PCL," *Project Assignment*, 2012.
- [52] M. Klingensmith, I. Dryanovski, S. Srinivasa und J. Xiao, "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.
- [53] J. Schornak, *YAK: 3D Reconstruction in ROS2*, Aug. 2019. Adresse: <https://rosindustrial.org/news/2019/8/7/yak-3d-reconstruction-in-ros2> (besucht am 27.05.2020).
- [54] E. Chernyaev, "Marching cubes 33: Construction of topologically correct isosurfaces," *Techn. Ber.*, Juli 1995.

- [55] M. Kazhdan, M. Bolitho und H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the fourth Eurographics symposium on Geometry processing*, Bd. 7, 2006.
- [56] A. C. Ponce, *Elliptic PDEs, Measures and Capacities*. European Mathematical Society, 2016, isbn: 978-3-03719-140-8.
- [57] F. P. Preparata und M. I. Shamos, *Computational geometry: An Introduction*. Berlin, Springer-Verlag, 1985, isbn: 978-0-387-96131-6.
- [58] Z. C. Marton, R. B. Rusu und M. 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.
- [59] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva und G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE transactions on visualization and computer graphics*, Jg. 5, Nr. 4, S. 349–359, 1999.
- [60] R. B. Rusu und S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*, IEEE, 2011, S. 1–4.
- [61] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler und A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA workshop on open source software*, Kobe, Japan, Bd. 3, 2009, S. 5.
- [62] M. Kazhdan, M. Chuang, S. Rusinkiewicz und H. Hoppe, "Poisson Surface Reconstruction with Envelope Constraints," *Computer Graphics Forum*, 2020, issn: 1467-8659. doi: [10.1111/cgf.14077](https://doi.org/10.1111/cgf.14077).