



TU WIEN
DEPARTMENT OF
GEODESY AND
GEOINFORMATION

Diplomarbeit

Automatisierte Abschätzung von Raum-Dimensionen aus SfM-Punktwolken

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Diplom-Ingenieur/in

ausgeführt am Department für

Geodäsie und Geoinformation

Forschungsbereich Photogrammetrie

der Technischen Universität Wien

unter Anleitung von

Univ.Ass. Dipl.-Ing. Michael Wimmer

und

Univ.Prof. Dipl.-Ing. Dr.techn. Norbert Pfeifer

durch

Georg Lebeth BSc

Wien, im November 2021

Unterschrift (Verfasser/in)

Unterschrift (Betreuer/in)

Kurzfassung

Die Photogrammetrie bietet eine Möglichkeit, aus RGB-Fotos dreidimensionale Modelle zu berechnen. In dieser Arbeit wurde mit einem Teilgebiet der Photogrammetrie namens Structure-from-Motion (**SfM**) gearbeitet. Dabei wird von einer Vielzahl an Fotos desselben Objektes die relative Orientierung bestimmt. Anschließend kann aus den orientierten Fotos eine Punktwolke rekonstruiert werden. SfM-Punktwolken sind von Haus aus nicht korrekt orientiert und skaliert, es gibt jedoch verschiedene Möglichkeiten dafür.

Diese Arbeit beschäftigt sich konkret mit SfM-Punktwolken von Innenräumen und den daraus ableitbaren Raum-Größen. Die Orientierung erfolgt über koordinativ festgelegte „Targets“. Da die Punkte von Innenraum-Punktwolken häufig unregelmäßig verteilt und verrauscht sind, stoßen typische Segmentierungs-Methoden, wie sie für Laser-Scanning-Daten und ToF-Kamera-Daten verwendet werden, hier an ihre Grenzen. Aus diesem Grund wurde ein robuster Algorithmus entwickelt, der in der Lage ist, trotz dieser Erschwernisse die Wände abzuschätzen. Zu jeder geschätzten Wandebene werden statistische Informationen exportiert, um auch nach dem Processing Information über die Zuverlässigkeit der Ebenen-Schätzung zu haben. Abschließend werden die Raummaße über den Verschnitt aller Ebenen und den daraus resultierenden Eckpunkten bestimmt. Bei der praktischen Anwendung der Methodik auf 10 Räume betrug der durchschnittliche Längenfehler im Vergleich zu den Referenzmessungen $\sim 2.5 \text{ cm}$.

Die größte Herausforderung beim Ebenen-Fitting stellen Wände dar, die nur aus großer Entfernung fotografiert wurden und einfarbig ohne jegliche Textur waren. Durch solche Wände traten in dieser Arbeit Längenfehler von maximal 12 cm (= bei einer Raumlänge von $\geq 12 \text{ m}$) auf. Um diese Wände besser rekonstruieren zu können, sollten zusätzliche Fotos mit geringerer Entfernung aufgenommen werden.

Abstract

Photogrammetry offers the opportunity to generate three-dimensional models from RGB-photos. This thesis uses one part of the photogrammetry, named Structure-from-Motion (**SfM**). With this method it is possible to compute the relative orientation of a multitude of photos of the same object. Afterwards a pointcloud can be generated from these oriented photos. Generally SfM-pointclouds are not correctly oriented and scaled, but there are a couple of options to do so.

This thesis is especially about indoor SfM-pointclouds and the deviated room-size. The orientation and scaling is done with dative bound targets. Pointclouds from indoor-SfM are often irregular distributed and affected by random noise, which causes typical segmentation methods, like they are used for laser-scanning-data or ToF-camera-data to fail. For this reason a robust algorithm was developed, which is capable of handling these difficulties. For every classified wall a plane is fitted and statistical information is exported to have information about the fitting, even after the processing has finished. Finally the room dimensions are calculated by miscutting all six planes and calculating the distances between the points of intersection. When the algorithm was tested on 10 different rooms, the average length error was about ~ 2.5 *cm* compared to the ground truth.

The biggest challenge when it comes to plane fitting occurs, if photos are taken from greater distance to a wall, which is also single-colored without any texture. With these walls, length errors of about ~ 12 *cm* (at a room length of about ≥ 12 *m*) occurred. To get better results there, some additional photos from shorter distance should have been made.

Inhaltsverzeichnis

1	Einleitung	5
1.1	State-of-the-Art	5
1.2	Ziel der Arbeit	6
1.3	Struktur der Arbeit	7
2	Theoretischer Hintergrund	8
2.1	Structure-from-Motion (SfM)	8
2.1.1	Aufnahme der Fotos	8
2.1.2	SfM-Workflow	9
2.1.3	Orientierung und Skalierung der Punktwolke	14
2.1.4	CCTags/Targets	15
2.2	RANSAC	16
2.3	LSM	18
3	Methodik	19
3.1	SfM-Softwarevergleich für Punktwolkenerzeugung	19
3.2	Orientierung mit Hilfe von Targets	22
3.3	Grobauswahl der Wände	23
3.4	Bestimmung der Wandebene	26
3.5	Qualitätskriterien	26
3.6	Abschätzung der Raumgröße	27
4	Experimente und Diskussion	29
4.1	Untersuchte Räume	29
4.2	Aufnahme der Fotos	30
4.3	Erzeugung der Punktwolke	31
4.4	Ebenen-Fitting	32
4.5	Qualitätsüberprüfung	36
4.6	Punktwolken mit erhöhter Punktdichte	36
5	Zusammenfassung und Schlussfolgerung	40
6	Literaturverzeichnis	42
7	Anhang	47

1 Einleitung

Automatisierung spielt in der modernen Technik eine tragende Rolle. Häufig ist es dabei notwendig, Objekte der realen Welt zu digitalisieren, Modelle zu erstellen und daraus vollautomatisch Informationen abzuleiten. Mit manuellen Methoden ist dies zwar auch möglich, jedoch ist es meist zeitintensiv und verlangt Vorwissen.

Für den Übergang von der Realität in ein digitales Medium gibt es verschiedenste Möglichkeiten, eine davon führt über den photogrammetrischen Ansatz. Allgemein wurde die Photogrammetrie von der American Society for Photogrammetry and Remote Sensing (ASPRS) folgendermaßen definiert (http://www.asprs.org/wp-content/uploads/2010/12/ProcurementGuidelines_SubmittalforBoardApproval_081409.pdf):

„The art, science, and technology of obtaining reliable information about physical objects and the environment, through processes of recording, measuring, and interpreting images and patterns of electromagnetic radiant energy and other phenomena.“

Ein mögliches Produkt einer photogrammetrischen Auswertung ist eine dreidimensionale Punktwolke, die für räumliche Analysen verwendet werden kann. Diese Punktwolke kann aus verschiedenen Quellen wie zum Beispiel Laser-Scannern, Time-of-Flight-Kameras oder RGB-Bildern stammen.

Ein Teilgebiet der Photogrammetrie beschäftigt sich mit der Rekonstruktion von Gebäuden, sowohl im Innen-, als auch im Außenraum. Dabei gibt es verschiedene Levels-of-Detail (**LoD**) (Fan & Meng, 2009), die erreicht werden können. Diese reichen von einem einfachen Blockmodell bis hin zu einer fein detaillierten Rekonstruktion des Gebäudes. Das Ziel bei der Rekonstruktion eines Gebäudes ist es, ein zuvor festgelegtes LoD zu erreichen. In dieser Arbeit ist es das Ziel, die Raumgröße eines Zimmers zu bestimmen, wodurch sich das LoD auf die Stufe 1, also ein einfaches Blockmodell beschränkt. Die für die Punktwolkenerzeugung verwendete Methode ist Structure-from-Motion (**SfM**), wodurch sich Schwierigkeiten durch schlecht verteilte Punktwolken ergeben. Dies wird in Abschnitt 2.1 näher diskutiert.

1.1 State-of-the-Art

In vorhergehenden Arbeiten wurde bereits die Rekonstruktion von Räumen aus verschiedensten Produkten untersucht.

Liu et al., 2018 verwenden ToF-Kameras, aus deren Video-Aufzeichnungen die Grundrisskarte eines gesamten Gebäudes erstellt wird. Anschließend wird aus dieser Karte ein komplettes dreidimensionales Modell eines Gebäudes erstellt. Dabei wird auch auf die Segmentierung einzelner Räume Rücksicht genommen. ToF-Kameras und deren Eigenschaften werden später in diesem Abschnitt noch näher behandelt.

Reisner-Kollmann, 2013 und Yang et al., 2019 verwenden für die Rekonstruktion Laser-Scanning Daten. Bei Reisner-Kollmann, 2013 werden vollautomatisch Gebäudeflächen segmentiert und anschließend Fensteröffnungen detektiert. Außerdem können komplexe Objekte aus geometrischen Primitiven rekonstruiert werden, dies fordert jedoch User-Input.

In [Yang et al., 2019](#) werden Wände in gekrümmte und runde Gebäude gefittet und anschließend die einzelnen Räume segmentiert. Es werden dabei die Rundungen aus einzelnen glatten Ebenen und aus gekurvten Flächen rekonstruiert und verglichen. Als Endergebnis steht wie bei [Liu et al., 2018](#) eine Grundrisskarte des Gebäudes und ein dreidimensionales Modell zur Verfügung.

In Arbeiten von [Ding et al., 2019](#) und [Dong et al., 2015](#) wird die Rekonstruktion von Innenräumen durch Fotos und Structure-from-Motion behandelt. Das Ziel beider Arbeiten ist ein vollständiges Modell der Räume.

Bei den meisten Arbeiten steht die Segmentierung von größeren Punktwolken aus Laser-Scanning- und ToF-Kamera-Daten im Vordergrund. Beide Systeme sind kostspielig, bieten dafür jedoch eine relativ gleichmäßige Abdeckung über den gesamten aufzeichneten Bereich.

Laser-Scanner bieten eine Möglichkeit zu großräumigen Messungen ([Alkan & Karşidag, 2012](#)), die trotzdem eine hohe Messpräzision, also geringe Streuung der Daten erreichen können ([Arayici et al., 2006](#)). Je nach verwendetem Modell sind unterschiedliche Genauigkeiten und Distanzen erreichbar. Da es sich um ein aktives System handelt, spielt die Textur der gemessenen Objekte keine sehr große Rolle. Der Nachteil dieser Methode sind die hohen Anschaffungskosten für einen Laser-Scanner.

Eine weitere Möglichkeit zur Punktwolkenerzeugung sind Time-of-Flight-Kameras (**ToF**), bei denen zusätzlich zu der RGB-Information einer klassischen Kamera auch noch die Distanz in jedem Pixel bestimmt wird, weshalb sie auch RGB-D-Kameras ($D = \text{Distance}$) genannt werden. Der Nachteil dieser Methode ist jedoch, dass die Ergebnisse meist von geringerer Qualität sind, als die von Laser-Scannern ([Reisner-Kollmann, 2013](#)). Preismäßig liegen ToF-Kameras zwischen RGB-Kameras und Laserscannern.

Die Methode Structure-from-Motion wird in dieser Arbeit verwendet. Auf diese wird in Abschnitt 2.1 noch ausführlich eingegangen. Grundsätzlich beschäftigt sich diese Methode damit, aus einer Vielzahl überlappender Fotos desselben Objektes eine Punktwolke zu berechnen.

1.2 Ziel der Arbeit

Diese Arbeit beschäftigt sich mit der Rekonstruktion und Größenabschätzung einzelner, teils stark möblierter Zimmer. Als Datensatz werden SfM-Punktwolken verwendet, die im Vergleich zu Laser-Scanning-Daten verrauschter sind und einen tendenziell höheren Anteil an falschen Punkten aufweisen. Eine weitere Schwierigkeit im Umgang mit dieser Methode ist, dass die Punktwolken in ihrer Punktdichte nicht näherungsweise konstant sind, sondern an schlecht texturierten Stellen nahezu keine Punkte aufweisen.

Der Grund, wieso trotzdem diese Methode gewählt wurde, ist der, dass sie sehr kostengünstig und einfach zu handhaben ist, da mit den Kameras moderner Smartphones bereits hochwertige Punktwolken generiert werden können. Somit ist kein spezielles Equipment notwendig, lediglich Software zur Berechnung von SfM-Punktwolken.

Die gesamte Arbeit steht unter der Motivation, kosteneffizient und nutzerfreundlich abzulaufen, weshalb für die Segmentierung und Größenabschätzung des Raumes ein vollautomatischer Algorithmus gewählt wurde. Lediglich bei der Generierung der Punktwolke sind User-Inputs notwendig, was später in Kapitel 3 näher diskutiert wird. Das finale Er-

gebnis sind die Abmessungen des Raumes ohne jegliches Mobiliar, Fenster, Türen und ähnlichem, sowie statistische Informationen zu den Berechnungen.

1.3 Struktur der Arbeit

Die Arbeit ist folgendermaßen strukturiert: Zu Beginn wird in Kapitel 2 auf die Theorie von SfM (siehe Abschnitt 2.1), RANSAC (2.2) und LSM (2.3) eingegangen.

Im Kapitel 3 werden zuerst ein kommerzielles und ein Freeware Softwarepaket für die SfM-Punktwolkenberechnung verglichen (3.1) und anschließend auf die verwendete Methode zur Orientierung der Punktwolke eingegangen (3.2). Zuletzt wird auf den verwendeten Workflow zur Segmentierung der Wände (3.3 und 3.4), deren Qualität (3.5) und anschließenden Größenabschätzung (3.6) eingegangen.

In Kapitel 4 wird der ausgearbeitete Workflow an zehn verschiedenen Räumen getestet und die aufgetretenen Probleme werden diskutiert.

Abschließend wird in Kapitel 5 das Ergebnis der Arbeit zusammengefasst und die Schlussfolgerung gezogen.

2 Theoretischer Hintergrund

In diesem Kapitel wird ein Überblick über die verwendeten Methoden gegeben. Bevor die eigentliche Abschätzung der Raumgröße möglich ist, müssen Fotos gemacht und daraus die Punktwolke erzeugt werden. Nach der Skalierung und Orientierung der Punktwolke wird diese gefiltert und es werden mit Hilfe der robusten Methode RANSAC sechs Wände gefittet. Um ein optimales Ergebnis zu erhalten, werden die sechs Ebenen noch einem LSM-Ausgleich unterzogen, bevor die Längen des Raumes geschätzt werden.

2.1 Structure-from-Motion (SfM)

Structure-from-Motion basiert, wie der Name bereits sagt, auf Bildern mit bewegten Projektionszentren. Konkret werden mehrere Bilder desselben Gegenstandes mit unterschiedlichen Projektionszentren benötigt. Für diese Methode ist ausschließlich eine klassische RGB- oder schwarzweiß-Kamera notwendig, um eine Vielzahl an Fotos desselben Objektes aufzuzeichnen (Pollefeys et al., 1998). Die gesamte Punktwolke kann dann ausschließlich aus diesen Fotos generiert werden (Westoby et al., 2012), wobei sie grundsätzlich weder korrekt skaliert noch orientiert ist (Özyeşil et al., 2017). Hierfür sind weitere Schritte notwendig, wie in späteren Abschnitten noch beschrieben wird. Allgemein ist diese Methode sehr günstig, da sogar einfache Handy-Kameras bei einigen Anwendungen zu brauchbaren Ergebnissen führen. Im folgenden Abschnitt wird näher auf den Workflow vom Foto zur Punktwolke eingegangen.

2.1.1 Aufnahme der Fotos

Als „Messinstrument“ dient grundsätzlich eine RGB-Kamera (Pollefeys et al., 1998), wie etwa eine Spiegelreflexkamera oder die Kamera eines Smartphones. Viele der modernen Smartphones haben ausreichend gute Kameras verbaut, um damit aussagekräftige Structure-from-Motion-Punktwolken zu erzeugen. Vorweg genommen, wurden in dieser Arbeit die Fotos mit einer Smartphone-Kamera aufgenommen, nähere Informationen zur Kamera finden sich in Abschnitt 4.2.

Der nächste wichtige Punkt der Datenaufnahme ist die Position und Ausrichtung der Kamera. Für gute Ergebnisse sollten die **Kamera-Positionen** (etwa wie in Abbildung 1) gleichmäßig um das Objekt verteilt gewählt werden, um eine gute Abdeckung aus allen Richtungen zu gewährleisten (Westoby et al., 2012). Da die Kamera-Positionen sich „außerhalb“ des Objektes befinden, bietet es sich an, rund um das Objekt Fotos zu machen.

Im Fall eines Zimmers stellt sich die Frage der besten Kamera-Positionen. Theoretisch bieten sich zwei grundsätzliche Ansätze an:

1. Im Raum-Zentrum
2. Entlang der Wände

Beide Ansätze sind in Abbildung 2 abgebildet. Beim ersten Ansatz links befindet sich die Kamera nahe des Raumzentrums und wird lediglich in kleinen Winkeln weitergedreht. Das Projektionszentrum sollte dabei auch bewegt werden, um keine Baseline mit

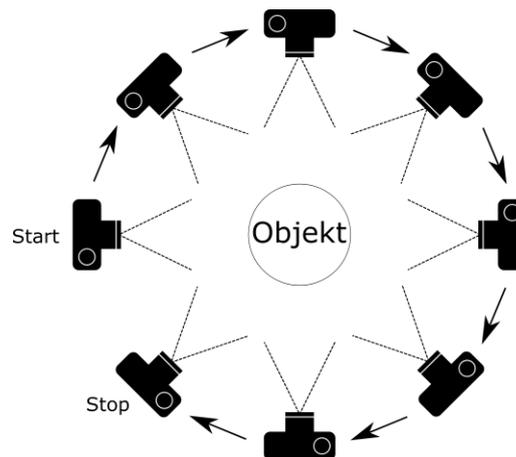


Abbildung 1: Kamera-Positionen für SfM mit Kameras „außerhalb“ des Objektes

Länge Null zwischen zwei Fotos zu erhalten. Dabei kann es vorkommen, dass bestimmte Stellen des Raumes nie abgebildet werden, falls sie durch Objekte verdeckt sind, wie links des Kastens schematisch dargestellt ist. Der zweite Ansatz wird im rechten Bild veranschaulicht. Dabei werden die Kamerapositionen so nahe wie möglich an der Wand gewählt, mit Blickrichtung Raumzentrum. In den Ecken werden zusätzlich kreisförmig Fotos aufgenommen.

In einer Studie von **Pix4D** (Software-Hersteller für SfM-Lösungen: <https://www.pix4d.com>) wurden diese zwei und noch weitere Ansätze getestet und die Ergebnisse verglichen (PIX4D, 2017). Als Qualitätsangabe wurde die Anzahl der gefundenen Verknüpfungspunkte in den Bildern und der Reprojection-Error (siehe Abschnitt 2.1.2) gewählt. Das Ergebnis zeigte, dass der erste Ansatz zu verwerfen war, da die Baselines (Strecke zwischen Projektionszentren) zu gering und daher die Schnittwinkel der Kollineationsstrahlen sehr schleifend waren. Dieser Ansatz wurde in der Studie als Negativ-Beispiel herangezogen. Der zweite Ansatz brachte hingegen die besten Ergebnisse, da die Verzerrung der Objekte am Bildrand in diesem Fall am geringsten war und auch größere Baselines ermöglicht wurden.

Abschließend ist zu sagen, dass die Fotos so stark wie möglich überlappen, dabei jedoch trotzdem eine größtmögliche Baseline aufweisen sollten. In Peterson et al., 2015 wird eine Überlappung von über 60 % empfohlen, um möglichst viele Verknüpfungspunkte in den Bildern zu erreichen, in dieser Arbeit wurde die Überlappung auf über 80 % gewählt.

2.1.2 SfM-Workflow

Nach der Fotoaufnahme folgt die Punktwolkenableitung mittels Structure-from-Motion. Wie schon zuvor erwähnt, ist eine Grundbedingung für eine erfolgreiche Berechnung, dass sich die Projektionszentren der verschiedenen Bilder unterscheiden müssen. Die Orientierung der Bilder ist frei wählbar, es muss nur einen Überlappungsbereich zu den anderen Fotos geben. Im Vergleich dazu sind beim photogrammetrischen Normalfall (Heimonen et al., 2001) die Bilder zwingend gleich ausgerichtet. Der SfM-Workflow lässt sich in folgende Schritte unterteilen:

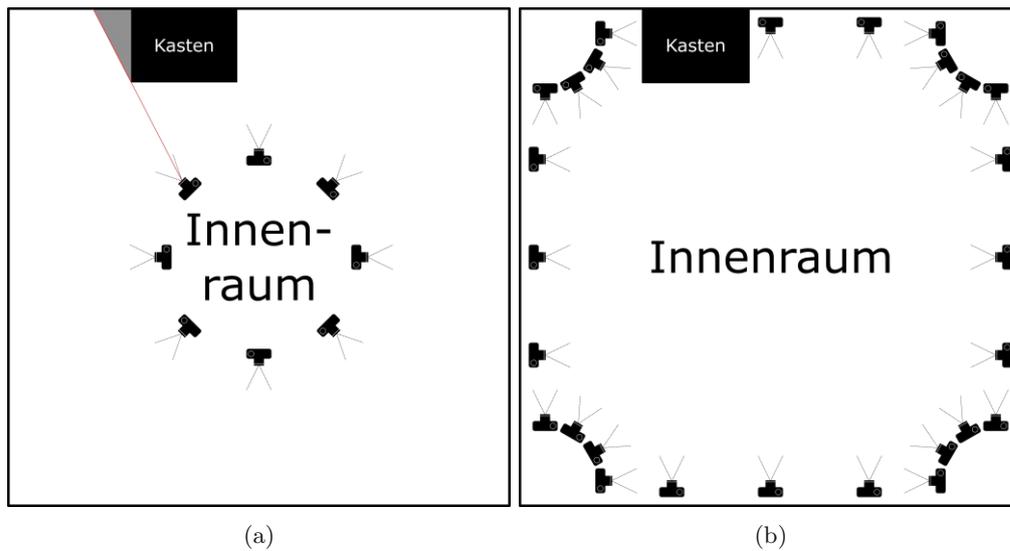


Abbildung 2: Anordnung der Fotos aus dem Zentrum (Ansatz 1, Abbildung a) oder entlang der Wände (Ansatz 2, Abbildung b)

1. Feature-Extraction
2. Feature-Matching
3. Bündelblockausgleich \rightarrow Sparse Cloud
4. Dense Matching \rightarrow Dense Cloud

Die **Feature-Extraction** wird für jedes Bild getrennt durchgeführt (Jiang et al., 2020). Dabei werden bestimmte markante Keypoints und deren Umgebung durch sogenannte lokale Deskriptoren beschrieben. Für die genaue Position jedes einzelnen Features kann Sub-Pixel-Genauigkeit erreicht werden (Peleg et al., 1987). Ein möglicher Deskriptor ist der **SIFT**-Operator (Scale Invariant Feature Transform) (Lowe, 2004), der die Eigenschaften der näheren Umgebung in einem sogenannten Feature-Vektor beschreibt. Das Ergebnis der Feature-Extraction ist eine Liste von markanten Keypoints und deren Feature-Vektor für jedes Bild (Lowe, 2004). Es stehen auch andere Deskriptoren wie zum Beispiel KAZE (Alcantarilla et al., 2012) oder Variationen von SIFT (Wu et al., 2013) zur Verfügung, auf diese wird an dieser Stelle jedoch nicht näher eingegangen.

In vielen SfM-Software-Paketen ist es möglich die maximale Dichte der Features zu definieren. Je höher die Dichte an Features, desto mehr Punkte können im Anschluss in der Sparse-Cloud rekonstruiert werden. Der Preis für eine höhere Punktdichte ist dabei die Rechenzeit, die vor allem beim Feature-Matching (siehe übernächsten Absatz) enorm ansteigt, da sehr viele Features miteinander verglichen werden müssen (Jiang et al., 2020). Somit ist die Wahl der Feature-Dichte eine klassische Kosten-Nutzen-Abschätzung.

Ein wichtiger, erwähnenswerter Punkt ist, dass die Feature-Dichte von der Textur der Objekte in den Fotos abhängig ist (Oats et al., 2017 und Khaloo und Lattanzi, 2016). Je stärker ein Objekt texturiert ist, desto mehr Information kann daraus gewonnen werden. Dies wurde auch in einer Arbeit von Ding et al., 2019 behandelt. Die Features

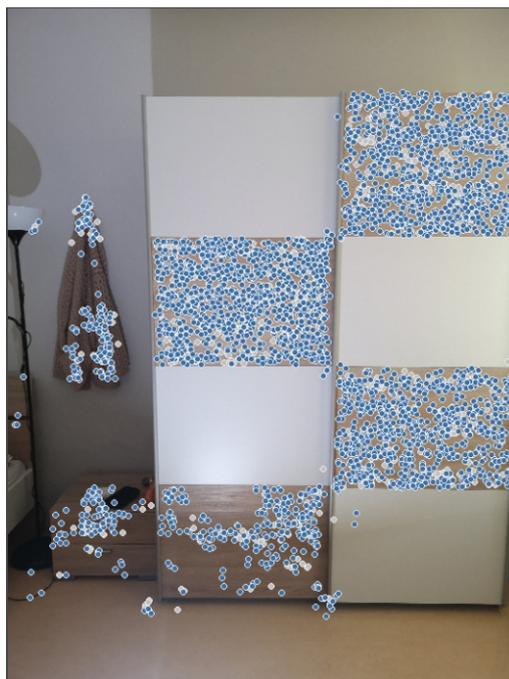


Abbildung 3: Ein Beispiel für extrahierte Features aus dieser Arbeit. Auf dem stark texturierten Kasten (Raum 3) sind sehr viele Punkte zu finden, während an den Wänden so gut wie keine Punkte vorhanden sind

befinden sich demnach bei weißen Wänden hauptsächlich an den Kanten des Raumes und an texturierten Gegenständen wie Bildern, gemasertem Holz oder Schattenwürfen (vgl. Abbildung 3). Diese Tatsache spielte vor allem in dieser Arbeit eine große Rolle, da bei den Innenaufnahmen ein Großteil der Fotos ausschließlich glatte, weiße Wände zeigte.

Das **Feature-Matching** beschreibt den Prozess, in dem die Features verschiedener Bilder miteinander verglichen und Verknüpfungspunkte gesucht werden (Jiang et al., 2020). Können zwischen jeweils zwei Fotos ausreichend viele Verknüpfungspunkte gefunden werden, ist es möglich, sie anhand dieser Verknüpfungspunkte relativ zueinander zu orientieren. Dieser Prozess ist besonders bei größeren Foto-Datensätzen sehr rechenintensiv, da viele Millionen mögliche Punktpaare miteinander verglichen werden müssen (Jiang et al., 2020). Je mehr Punkte in der Feature-Extraction detektiert wurden, desto länger dauert die Berechnung, da im uneingeschränkten Fall jedes Feature jedes Bildes mit jedem anderen verglichen werden muss. Um nicht alle Features miteinander vergleichen zu müssen, wurden in Jiang et al., 2020 effizientere, zum Beispiel räumlich oder geometrisch eingeschränkte Ansätze untersucht. In vielen SfM-Software Paketen werden diese Ansätze verwendet, da sie eine enorme Zeitersparnis bringen.

Nach dem Feature-Matching muss noch eine Grobfehlersuche durchgeführt werden, da nicht alle Matches auch wirklich korrekt sind (Pollefeys et al., 1998), wie in Abbildung 4 ersichtlich. Dafür bietet sich ein robuster Algorithmus wie RANSAC (siehe Abschnitt 2.2) an. Die genaue Vorgehensweise für RANSAC beim Feature-Matching ist in der Arbeit von (Pollefeys et al., 1998) nachzulesen.

Die relative Orientierung der Bilder geschieht anschließend durch einen **Bündelblock-**

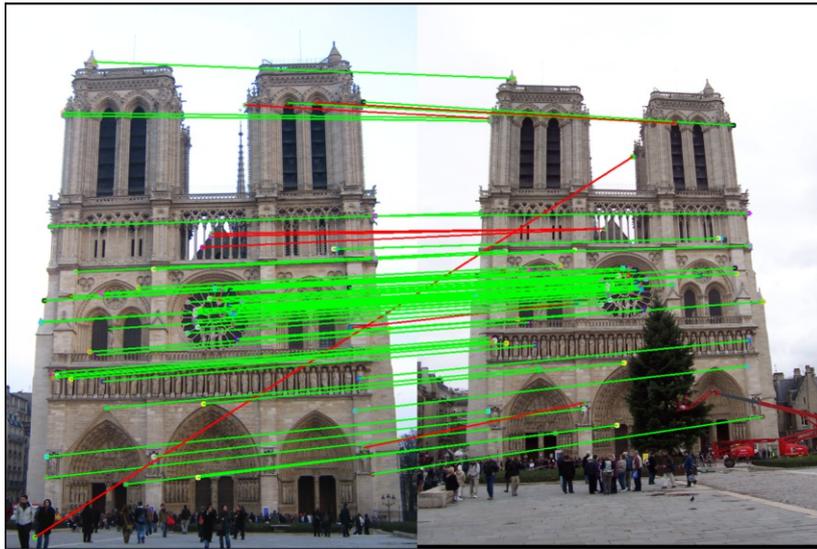


Abbildung 4: Ein Beispiel für Matches (grün) und Mismatches (rot) aus einer Arbeit von [Hays et al., 2018](#). Der Anteil falscher Matches muss durch robuste Methoden gefiltert werden.

ausgleich, wobei die relative Lage der einzelnen Projektionszentren der Fotos, sowie die 3D-Koordinaten der Verknüpfungspunkte bestimmt wird. Dabei wird der sogenannte Reprojection-Error (**RE**) minimiert ([Özyeşil et al., 2017](#) und [Triggs et al., 2000](#)). Der Reprojection-Error beschreibt dabei den Abstand zwischen dem projizierten und dem beobachteten Bildpunkt (Feature) ([Nguyen et al., 2012](#)). Er ist somit ein Qualitäts-Maß für die Feature-Matches und die Genauigkeit der Orientierung der Kameras untereinander ([Özyeşil et al., 2017](#)).

Der ursprüngliche Ansatz nach [Longuet-Higgins, 1987](#) sieht ausschließlich die relative Orientierung **zweier** Bilder vor. Daher wurden in den Anfängen die Fotos Schritt für Schritt einzeln zu den bereits orientierten Fotos hinzugefügt und dabei immer iterativ der Workflow der ersten drei Arbeitsschritte von SfM (Feature-Extraction, Feature-Matching und Bündelblockausgleich) wiederholt ([Jiang et al., 2020](#) und [Westoby et al., 2012](#)). Da dadurch nach jedem hinzugefügten Bild ein Bündelblockausgleich berechnet werden muss, wird bei größeren Datensätzen die Berechnung sehr rechenintensiv. Die Arbeit von [Jiang et al., 2020](#) zeigt auch hier effizientere Ansätze, die zum Beispiel mehrere Sub-Szenen vorsehen, die anschließend verknüpft werden, oder eine Bild-Vorauswahl, um nicht jedes einzelne Bild im Feature-Matching vergleichen zu müssen. Die Ergebnisse sind eine grobe Punktwolke (**Sparse Cloud**) aus Verknüpfungspunkten sowie die relativen Orientierungsparameter der einzelnen Kameras zueinander ([Jiang et al., 2020](#) und [Westoby et al., 2012](#)).

Der letzte Schritt in SfM ist die Verdichtung der Punktwolke, um die sogenannte **Dense Cloud** zu erhalten. Hier wird mit verschiedenen Ansätzen versucht, die bestehenden Verknüpfungspunkte zu verdichten. In einem Ansatz muss zuerst eine sogenannte „Dense Depth Map“ für jedes Bild erstellt werden ([Ernst et al., 2002](#)), wie in Abbildung 5 dargestellt. Einige Möglichkeiten dafür werden in [Pollefeys et al., 1998](#) präsentiert. Die einfachste Methode wäre eine simple Interpolation zwischen den Verknüpfungspunkten

des Bündelblockausgleichs, dabei würde jedoch auf kleine Details verzichtet werden. Aus diesem Grund können zum Beispiel Algorithmen verwendet werden, die bereits aus Stereobild-Ansätzen (Heimonen et al., 2001) bekannt sind und in beinahe jedem Pixel des Bildes Korrespondenzen suchen (Pollefeys et al., 1998). Ein anderer Ansatz namens „Segment Matching“ von Ernst et al., 2002 segmentiert jedes Bild nach Farbregionen und matcht anschließend diese einzelnen Regionen, wodurch für alle Pixel einer Region ein Tiefenwert bestimmt wird. Eine weitere Möglichkeit ist „Semi-Global-Matching“ (Hirschmüller, 2008 und Stentoumis et al., 2015), wobei ebenfalls ein Stereo-Matching-Ansatz verwendet wird.



Abbildung 5: In Metashape berechnete Depth-Map für das Foto aus Abbildung 3. Grüne Punkte befinden sich näher, blaue weiter hinten.

Abschließend ist zu sagen, dass bei SfM eine Vielzahl an Variablen einen Einfluss auf die Genauigkeit der 3D Modelle hat. Neben den Eigenschaften des Kamerasensors spielen die Bild-Parameter, Kamerapositionen, Textur der abgebildeten Oberfläche und die Objektgröße eine große Rolle (Khaloo & Lattanzi, 2016). Liegen viele glatte, einfarbige Flächen in den Bildern vor, können nur wenige Keypoints (siehe Abschnitt 2.1.2) beschrieben werden. Auch beim Dense Matching ergeben sich dadurch an diesen Stellen weniger dichte Punktwolken. Trotzdem können durch diese Methode sehr kostengünstig aussagekräftige Punktwolken generiert werden.

Wurde die Punktwolke erzeugt, besteht die nächste Aufgabe darin, diese zu orientieren und zu skalieren, da in den Fotos grundsätzlich keinerlei Information diesbezüglich enthalten ist. Für diese Problemstellung gibt es mehrere Methoden, die im folgenden Abschnitt beschrieben werden.

2.1.3 Orientierung und Skalierung der Punktwolke

Ein großes Problem bei der SfM-Auswertung liegt in der Orientierung und Skalierung der Punktwolke, da diese in einem Fotodatensatz normalerweise nicht enthalten oder ableitbar ist. Somit muss dieser Schritt separat oder parallel zur Punktwolkenerzeugung gemacht werden. In dieser Arbeit wird auf drei mögliche Methoden sowie deren Vor- und Nachteile eingegangen. In vorhergehenden Arbeiten wie [Zhang et al., 2019](#) und [Froideval et al., 2019](#) wird auf direkte Georeferenzierung mithilfe von GNSS sowie auf Referenzierung über georeferenzierte Targets (siehe Abschnitt 2.1.4) eingegangen. Da in Gebäuden die direkte Georeferenzierung nicht so leicht möglich ist, ist dieser Ansatz hier zu verwerfen. Für die Referenzierung mit und ohne Targets gibt es jedoch mehrere Möglichkeiten von manuell bis vollautomatisch.

Die erste Methode ist vollkommen manuell und benötigt lediglich die fertige Punktwolke und einen graphischen Punktwolkeneditor wie zum Beispiel „**CloudCompare**“ (<https://www.danielgm.net/cc/>). Mithilfe der Software, kann die Lage der Punktwolke im Raum verändert (im Falle dieser Arbeit achsenparallel), sowie die Skalierung der Punktwolke adaptiert werden. Dafür muss ein Referenzmaß im Objektraum bekannt sein, auf das anschließend die Punktwolke angepasst wird. Im Fall von Targets können die Abstände der Target-Zentren zur Skalierung herangezogen werden, es ist aber auch jedes andere, bekannte Referenzmaß möglich. Dabei ist jedoch zu beachten, dass es in der Punktwolke relativ schwer ist, exakte Punkte zu bestimmen, da diese (zum Beispiel die Targets) in der Punktwolke nicht so gut aufgelöst werden, wie in Fotos. Dies ist in Abbildung 6 erkennbar. Die Targets in den Fotos sind deutlich klarer definiert als in der 3D-Rekonstruktion, wodurch das Target-Zentrum im Foto besser bestimmt werden kann. Wird also in CloudCompare die Punktwolke anhand der Targets skaliert, ist mit einem Skalierungsfehler zu rechnen.



(a) Foto der Targets



(b) Dense-Cloud Rekonstruktion der Targets

Abbildung 6: Darstellung der Targets in Fotos und der Dense-Cloud

Die zweite halb-automatische Methode zur Punktwolkenorientierung erfordert bereits spezielle Tools innerhalb der SfM-Software. Einige kommerzielle Programme wie **Agisoft Metashape** (<https://www.agisoft.com>) oder **Pix4D** (<https://www.pix4d.com>) bieten

die Möglichkeit, vor der eigentlichen SfM-Berechnung innerhalb der Fotos Targets manuell zu identifizieren und diesen Koordinaten im Objektraum zuzuweisen. In Agisoft Metashape können die Targets nur direkt mithilfe der Software erzeugt und gedruckt werden. Da die Targets nach dem Bündelblockausgleich innerhalb der Punktwolke (lokales Koordinatensystem) ebenfalls koordinativ bekannt sind, ist eine 7-Parameter-Transformation (Mataija et al., 2014) der Punktwolke ins übergeordnete System möglich (laut User-Manual Agisoft Metashape Pro 1.5). Die Koordinaten der Targets im übergeordneten System können dabei theoretisch willkürlich gewählt werden.

Die letzte Methode funktioniert vollautomatisch. Viele Softwarehersteller für SfM bieten spezielle codierte Targets (siehe Abschnitt 2.1.4) an, die von der Software automatisch erkannt werden. Somit ist kein User-Input für die Auswahl der Targets notwendig. Es müssen den jeweiligen Target-IDs nur Koordinaten im Objektraum zugeordnet werden, die Transformation erfolgt anschließend wie im vorigen Absatz beschrieben. Diese Methode wird auch von Freeware-Tools wie **Alicevision Meshroom** (<https://alicevision.org/>) unterstützt, das ebenfalls in dieser Arbeit verwendet wurde. Ein klarer Nachteil der automatischen Target-Erkennung ist, dass nicht immer alle Targets erkannt werden. Vor allem bei sehr flachen Blickwinkeln auf das Target, haben die Programme Schwierigkeiten dieses eindeutig zu identifizieren. In manchen Fällen kann es deswegen möglich sein, manuell nachzubessern. Im getesteten Freeware-Tool Alicevision Meshroom (Version 2021.1.0) ist die manuelle Target-Wahl jedoch nicht möglich (siehe 3.1).

2.1.4 CCTags/Targets

Wie zuvor erwähnt, gibt es verschiedene Arten von Targets. Verschiedene Softwarelösungen bieten ihre eigenen Targets an.

So verwendet etwa **Alicevision Meshroom** sogenannte **CCTags** (Concentric Circles Tags), wie in Abbildung 7a gezeigt. Die Targets sind durch ihre Strichstärken eindeutig einer Identifikationsnummer zugeordnet. Wie in Calvet et al., 2016 beschrieben, funktioniert die Detektion auch unter (nicht extremen) Verzerrungen. Bei der Detektion werden zuerst Ellipsen an jeder Kante zwischen schwarz und weiß gefittet, anschließend das Zentrum der Ellipsen bestimmt und zuletzt die ID des Targets aus den Abständen der Ellipsen bestimmt.

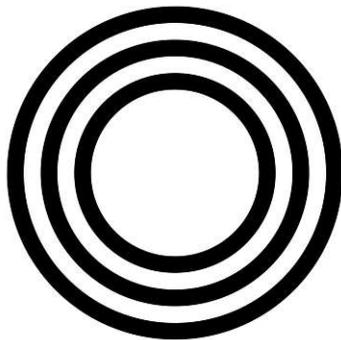
Ein wichtiges Kriterium ist die Größe der Targets im Bild. Für eine erfolgreiche Detektion von CCTags muss jedes Target mit einem Radius von mindestens 30 *px* abgedeckt sein (laut CCTag User-Manual). Damit ist die Mindestgröße der Targets abhängig von der Größe des Raumes und dementsprechend dem Abstand der Kamera zum Target. Für die Größenwahl der Targets kann Faustformel 1 verwendet werden (Gasparini, 2020):

$$R = \frac{mu}{f} d \quad (1)$$

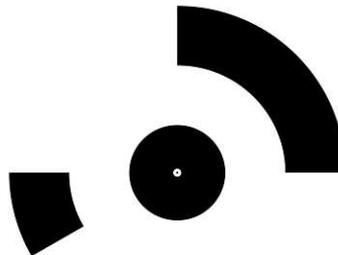
Dabei gelten folgende Definitionen:

- R : zumindest benötigter Radius der Targets (mm)
- m : Mindestgröße in Pixel für den Radius der Targets (px)
- u : Pixelgröße der Kamera (mm)

- f : Brennweite der Kamera (mm)
- d : Abstand der Kamera zu den Targets



(a) CCTag - AliceVision Meshroom



(b) CT - Agisoft Metashape

Abbildung 7: Targets der verwendeten Software

Agisoft Metashape bietet ebenfalls ein eigenes Set an Targets (**CT** - Coded Targets), wie in Abbildung 7b ersichtlich, an. Die Detektion erfolgt ebenfalls über Kantendetektion und Graustufenbilder. Zuerst wird der Kreis im Zentrum gesucht, anschließend mithilfe der Codes im zweiten Kreis die ID des Targets bestimmt (Wijenayake et al., 2014). Ein Vorteil von Agisoft Metashape gegenüber Meshroom ist, dass für die gefundenen Targets Genauigkeitsinformation in Pixel angegeben wird. Dadurch ist bereits vor der Berechnung der Dense Cloud eine Abschätzung der möglichen Genauigkeit der Ergebnisse möglich.

Für die Größe der Targets liegt in Metashape keine konkrete Information vor, weshalb mit denselben Größen wie bei den CCTags gearbeitet wurde.

2.2 RANSAC

Unter der Voraussetzung, dass die Raumwände ausreichend planar sind, werden sie durch ein Ebenen-Fitting rekonstruiert. Dafür wird eine robuste Methode zur Unterteilung in Outlier und Inlier benötigt. Aufgrund der leichten Handhabbarkeit und Robustheit (Ruzgiene & Förstner, 2005), wurde der sogenannte **RAN**dome **SAM**ple **CON**sensus (RANSAC) (Fischler & Bolles, 1981) gewählt. Da bei RANSAC der Anteil grober Fehler weit über 50 % erreichen kann (Ruzgiene & Förstner, 2005), bietet er sich speziell für diese Arbeit gut an, weil aufgrund der vielen schwach texturierten Wände (siehe Abschnitt 2.1.2) die Punktwolken stark verrauscht sind. Zusätzlich führen „störende“ Objekte wie Möbel und andere Einbauten zu Erschwernissen.

Grob gesagt, entspricht RANSAC einem Trial- und Error-Verfahren, da **zufällig** einzelne **Samples** aus dem gesamten Datensatz entnommen werden, für jedes eine eindeutige Lösung berechnet und anschließend das Ergebnis bewertet wird. Der Workflow entspricht somit folgenden Arbeitsschritten (Kaspi et al., 2017):

1. Zufälliges Sample wählen
2. Modell aus Sample berechnen

3. Modell bewerten

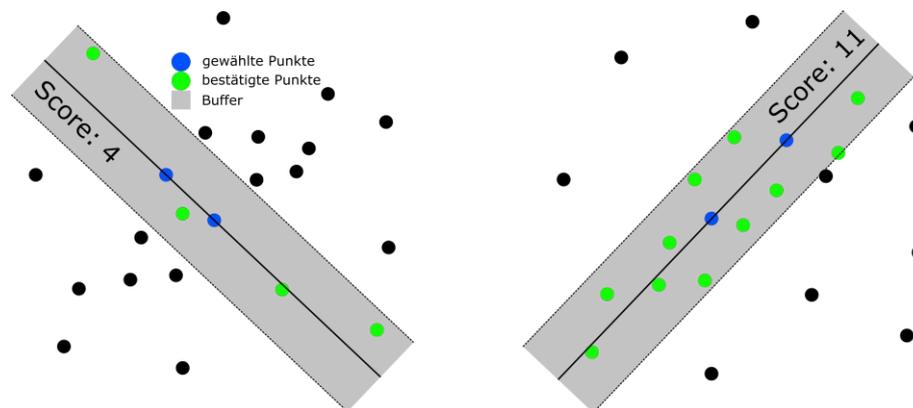
→ Iterative Wiederholung von Schritt 1-3

4. Modell mit bester Bewertung wählen

Im ersten Schritt werden zufällig genau so viele Daten gewählt, dass eine eindeutige Lösung bestimmt werden kann. Im Falle einer ausgleichenden Ebene sind dafür 3 Punkte notwendig. Da dies jedoch nicht leicht zu veranschaulichen ist, wird im Folgenden auf ein zweidimensionales Problem zurückgegriffen. Im Beispiel einer Ausgleichsgeraden sind nur zwei Punkte nötig, durch die eine Gerade gelegt wird. In Abbildung 8a wurden die zwei blau markierten Punkte gewählt.

Der nächste Schritt besteht darin, aus den beiden gewählten Punkten eine Geradengleichung zu bestimmen. Die berechnete Gerade ist ebenfalls in Abbildung 8a als schwarze Linie erkenntlich.

Zur Bewertung des Ergebnisses wird ein Buffer δ mit vordefiniertem Abstand um die Gerade gelegt, der in der Abbildung grau hinterlegt ist. Anschließend werden alle Punkte innerhalb des Buffers gezählt (Kaspi et al., 2017). Im Fall von Abbildung 8a ist der sogenannte **Score** 4. Der Score kann auch relativ zur Gesamtzahl aller Punkte angegeben werden. Bei insgesamt 22 Punkten (abzüglich zwei Punkte zur eindeutigen Lösung), wäre der prozentuelle Score (auch Inlier-Anteil) 20 %.



(a) Zufällig gewähltes Sample mit einem Score von 4 (b) Zufällig gewähltes Sample mit einem Score von 11

Abbildung 8: Vergleich zweier Samples und deren erreichter Score bei RANSAC

Da dieses einzelne Sample nicht aussagekräftig ist, werden N verschiedene, zufällige Samples gewählt. Schritte 1-3 werden somit N -mal durchgeführt und jeweils der Score bestimmt. In Abbildung 8b wurde ein anderes Sample gewählt, das einen Score von 11, oder 55 % erreicht und damit die Gerade „besser“ beschreibt. Nun stellt sich die Frage nach einer sinnvollen Anzahl von Samples oder einem Abbruchkriterium, da ansonsten theoretisch Samples in einer Endlosschleife gewählt werden würden.

In Navratil und Staudinger, 2019 werden verschiedene Ansätze zur Bestimmung eines Abbruchkriteriums vorgeschlagen. Einer davon beinhaltet den Abbruch nach einem vordefinierten Inlier-Anteil. In dieser Arbeit wurde als Abbruchkriterium ein Inlier-Anteil von 80 % gewählt. Da aufgrund der starken Punkt-Streuung normal zu den Wänden

nicht immer gewährleistet ist, dass der Algorithmus in keiner Endlos-Schleife endet, wurde zusätzlich eine Obergrenze für die maximale Anzahl an Samples gesetzt. Diese wurde durch ein Trial- and Error-Verfahren auf 10 % der Punktzahl eines vorausgewählten Subsets (näheres zum Auswahlverfahren des Subsets in Abschnitt 3.3) festgesetzt, um den gesamten Algorithmus möglichst effizient zu gestalten.

Nach einem vollständigen Durchlauf des Algorithmus erhält man ein Sample von Punkten, deren Buffer die meisten Inlier (aller getesteten Samples) aufweist. Gleichzeitig kann aus diesem Sample eine Ebenengleichung bestimmt werden, deren Parameter als Startwerte für einen Ausgleich dienen können. Die Inlier des „besten“ Samples werden ebenfalls für den Ausgleich benötigt und deshalb als eigenes Subset gespeichert.

Allgemein ist ein Nachteil an RANSAC, dass nicht alle möglichen Varianten getestet werden, wodurch es passieren kann, dass kein „richtiges“ Sample gefunden wird (Navratil & Staudinger, 2019). Im konkreten Anwendungsfall dieser Arbeit ist dies jedoch ausreichend, da zuvor bereits eine Vorauswahl (vgl. Abschnitt 3.3) getroffen wird und im Anschluss noch durch einen Ausgleich nach Methode der kleinsten Quadrate die gefittete Ebene optimiert wird.

2.3 LSM

Die Methode der kleinsten Quadrate wurde erstmals von Carl Friedrich Gauss beschrieben (Gauss & Davis, 1857) und beschreibt eine Möglichkeit zur Bestimmung der optimalen Ergebnisse (bezüglich der Verbesserungsquadratsumme) aus normalverteilten Daten (Navratil & Staudinger, 2006). Zusätzlich muss gewährleistet sein, dass die Daten frei von groben Fehlern sind und die Näherungswerte für die unbekanntes Funktionswerte ausreichend genau sind. Beim Ausgleich werden die Verbesserungsquadratsummen, also die quadrierten Abstände zwischen einem realen Datenpunkt und einem Funktionswert minimiert. Als Ergebnis erhält man die unbekanntes Funktionsparameter, die Verbesserungen und ausgeglichenen Beobachtungen.

Für einen Ebenenausgleich kann nach Pfeifer, 2002 auch mithilfe der Eigenwerttheorie gearbeitet werden. Diese Vorgehensweise wurde auch in Islam, 1989 herangezogen. Da dieser Ansatz eine schnelle Lösung ohne Linearisierung bietet, wurde er auch in dieser Arbeit herangezogen. Ein wichtiger Unterschied zur zuvor beschriebenen Methode nach Gauss ist, dass keine Startwerte für den Ausgleich benötigt werden.

Als Ergebnis erhält man die Ebene in Hesse-Normalform (Konstante und Normalvektor) und den **Root Mean Square Error** (RMSE) (Chai & Draxler, 2014) aller Punkte. Dieser ergibt sich aus dem kleinsten der drei Eigenwerte und Formel 2 (Pfeifer, 2002).

$$\sigma = \sqrt{\frac{\lambda_s}{n-3}} \quad (2)$$

mit $\sigma = RMSE$, $\lambda_s = \text{kleinster Eigenwert}$ und $n = \text{Anzahl Freiheitsgrade}$.

3 Methodik

Dieses Kapitel beschäftigt sich mit dem generellen Ablauf dieser Arbeit. Allgemein kann die Arbeit in folgende Schritte unterteilt werden:

1. Punktwolkenerzeugung und Orientierung
2. Statistische Extraktion von achsenparallelen Wänden
3. Bestimmung der Wandebenen durch RANSAC und LSM
4. Abschätzung der Raumgröße

3.1 SfM-Softwarevergleich für Punktwolkenerzeugung

Die Datenaufnahme wurde mit einem Smartphone der Marke Huawei durchgeführt. Die Details zur Foto-Aufnahme und Kamera sind in Abschnitt 4.2 zu finden. Nach diesem ersten Schritt folgt die Punktwolkenerzeugung durch spezielle SfM-Software.

Wie bereits in Abschnitt 2.1 angemerkt, wurden zwei verschiedene Software-Lösungen für die Punktwolkenerzeugung verwendet. Einerseits die kommerzielle Softwarelösung Agisoft Metashape, andererseits das Freeware-Tool Alicevision Meshroom (Version 2021.1.0). Beide Programme wurden vor den eigentlichen Experimenten unter dem Vorsatz der einfachen Handhabbarkeit und Effizienz an zwei Räumen (Raum 9 und Raum 10 aus Kapitel 4) mit den Standardeinstellungen getestet. Der verwendete Rechner war ein Windows 10 PC mit AMD FX(tm)-4100 Quad-Core-Prozessor und einer Nvidia GeForce 9600 GT Grafikkarte.

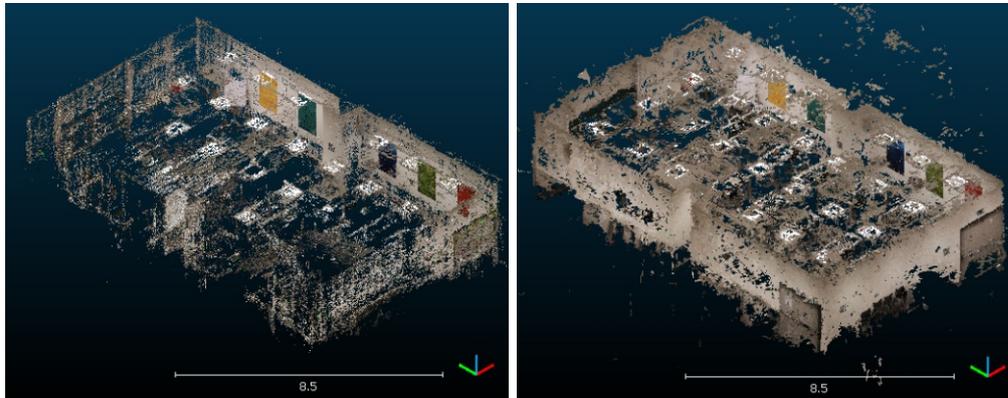
Agisoft Metashape bietet sowohl automatische, als auch manuelle Targeterkennung an. In den getesteten Datensätzen wurde der Großteil der Targets automatisch erkannt, einige konnten aufgrund sehr flacher Blickwinkel nicht korrekt identifiziert werden. In diesen Fällen konnte mit der manuellen Targeterkennung nachgeholfen werden. In beiden Vergleichsräumen konnten alle Fotos problemlos orientiert werden. Der Reprojection-Error betrug in Raum 9 2.50 *px*, in Raum 10 2.09 *px*. In der Sparse-Cloud konnten 16.3 *Tsd.* und 19.7 *Tsd.* Features gematcht werden. In der Dense-Cloud konnten anschließend 2.73 *Mio.* und 2.15 *Mio.* 3D-Punkte mit einem Wandanteil von 30% und 57 % rekonstruiert werden.



Abbildung 9: Fotos aus Raum 9: Links: in Meshroom nicht korrekt orientiertes Foto. Rechts: Gegenüberliegende Wand, wurde korrekt orientiert

Alicevision Meshroom bietet ausschließlich automatische Targeterkennung an. Somit kann nicht garantiert werden, dass alle sichtbaren Targets in allen Fotos auch verwendet wurden. In beiden Vergleichsräumen wurden nur wenige Fotos korrekt orientiert, wodurch eine manuelle Anpassung der Parameter in Meshroom notwendig war. Der Nachteil daran ist, dass jedes angepasste Parameter-Set nur für einen einzigen Datensatz passend ist. Nach der Anpassung konnten in Raum 9 29/42, in Raum 10 36/39 Fotos korrekt orientiert werden. Alle nach wie vor fehlenden Fotos zeigten dieselbe Wand, die nur sehr wenig Textur aufweist, wie in Abbildung 9 erkennbar. Dadurch war eine Wand sehr schlecht durch orientierte Fotos repräsentiert und wies nur eine sehr geringe Punktdichte auf. Mit den verbleibenden Fotos konnten insgesamt 22.7 *Tsd.* und 71.6 *Tsd.* Features mit einem Reprojection-Error (RE) von $\sim 0.95 px$ und $\sim 1.02 px$ gematcht werden. In der Dense-Cloud konnten insgesamt 0.62 *Mio.* und 0.78 *Mio.* 3D-Punkte rekonstruiert werden. Auffällig ist somit, dass bei den gleichen Datensätzen vom Freeware-Tool deutlich weniger 3D-Punkte in der Dense-Cloud rekonstruiert werden konnten als von der kommerziellen Lösung.

In Abbildung 10 sind die gesamten, berechneten Punktwolken aus beiden Programmen zu sehen. Man erkennt deutlich, dass in Meshroom die Punkte dünner und vor allem unregelmäßiger verteilt liegen als in Metashape. Zur Qualitätskontrolle der Punktwolken wurden die Wandpunkte pro Quadratmeter Wandfläche bestimmt. Die Ergebnisse sind in Tabelle 1 abzulesen. Die Punktdichten bestätigen, dass die Punktwolken aus Meshroom auch an den Wänden deutlich dünner sind als die aus Metashape.

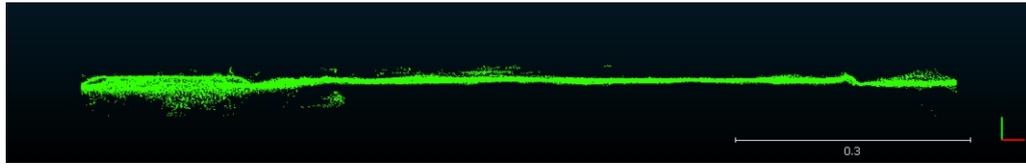


(a) Punktwolke aus AliceVision Meshroom:
0.62 *Mio.* Punkte

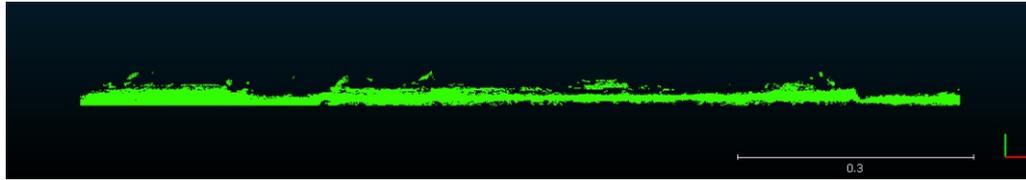
(b) Punktwolke aus Agisoft Metashape:
2.78 *Mio.* Punkte

Abbildung 10: Vergleich der Punktwolken aus Raum 9 mit beiden Softwarelösungen

Beide Softwarepakete bieten höhere Punkt-Dichten bei der Berechnung der Dense-Cloud an. Testweise wurden die Testräume in beiden Programmen mit erhöhten Dichten berechnet. In Metashape erhöhte sich die Berechnungszeit von etwa einer Stunde auf über zwölf Stunden, dafür hatten die berechneten Dense-Clouds anschließend etwa 14.30 *Mio.* und 14.98 *Mio.* Punkte mit einem Wandanteil von etwa 39 % und 52 %. Die Punktzahlen hatten sich also etwa um den Faktor 5 und 6 vergrößert. Die höhere Punktzahl bedeutet jedoch nicht unbedingt, dass Wände dadurch „besser“ rekonstruiert werden. In Abbildung 11 sind Ausschnitte der zwei Ergebnisse derselben Punktwolke abgebildet.



(a) **Normale** Punktdichte in der Dense-Rekonstruktion. Gesamte Punktwolke: 2.78Mio. Punkte



(b) **Hohe** Punktdichte in der Dense-Rekonstruktion. Gesamte Punktwolke: 17.40Mio. Punkte

Abbildung 11: Ausschnitte von Punktwolken aus Metashape mit unterschiedlicher Punktdichte aus Raum 9

Wie man im Vergleich der beiden Bilder erkennen kann, ist zwar die Punktdichte in Abbildung 11b deutlich höher, die Streuung normal der Wand nimmt jedoch eindeutig zu. Dies ist vor allem im rechten Teil der Bilder gut erkennbar, wo in der lichtereren Punktwolke die Wand eindeutig als schmaler Strich repräsentiert ist, während in der dichten Punktwolke die Streuung normal zur Wand zunimmt. Die Punktwolke mit hoher Punktdichte wurde ebenfalls zur Bestimmung der Raumgröße verwendet, die Ergebnisse stellten sich jedoch nicht als besser (siehe Abschnitt 4.6) heraus, weshalb auf die höhere Punktdichte verzichtet wurde.

In Meshroom erhöhte sich die Berechnungszeit auf über fünf Stunden, die Punktzahl der Dense-Cloud verdoppelte sich etwa, der prozentuelle Wandanteil verschlechterte sich jedoch. Zudem blieben große Teile des Raumes aufgrund fehlender Kameras unbestimmt, wodurch auch mit den dichten Punktwolken keine sinnvolle Rekonstruktion möglich war (vgl. Abschnitt 4.6).

In Tabelle 1 sind alle Ergebnisse der beiden Testräume in verschiedenen Parametern gegenübergestellt.

Tabelle 1: Verknüpfungspunkte (VP), Ground-Resolution (GR), Target-RMSE (TR), RE, Punktzahlen und Wandpunkte, sowie Punktdichten pro Quadratmeter Wand (PD) der Kameras der beiden Testräume. **In Meshroom konnten nie alle Fotos korrekt orientiert werden. Raum 9: 29/42, Raum 10: 36/39** Weiters ist keine Information über GR und TR verfügbar

Raum	Metashape		Meshroom		Metashape		Meshroom	
	9	9 HP	9	9 HP	10	10 HP	10	10 HP
VP (Tsd.)	16.3	16.3	22.7	22.7	19.7	19.7	71.6	71.6
GR [mm/px]	1.52	1.52	-	-	1.04	1.04	-	-
TR [px]	0.47	0.47	-	-	0.57	0.57	-	-
RE [px]	2.50	2.50	0.95	0.95	2.09	2.09	1.02	1.02
Punkte (Mio.)	2.73	14.30	0.32	0.75	2.15	14.98	0.39	0.68
davon Wand (%)	0.39	0.39	0.49	0.47	0.57	0.52	0.76	0.64
∅ PD (Tsd./m ²)	10.19	18.91	0.59	0.89	11.85	50.01	1.19	1.52
Zeit (h:min)	1:10	≥ 12 h	2:40	≥ 5 h	1:05	≥ 13 h	2:24	≥ 5 h

Abschließend ist zu sagen, dass aufgrund nicht orientierter Fotos, sowie weiteren Faktoren, wie der deutlich längeren Berechnungszeit und der geringeren, unregelmäßigeren Punktdichte, Meshroom mit den Standardeinstellungen keine optimalen Ergebnisse für die Auswertung in dieser Arbeit bringt. Durch Anpassung der Parameter speziell für jeden Datensatz, konnten gleichmäßigere Ergebnisse (siehe Abbildung 12) erreicht werden, dies bedeutet jedoch einen relativ hohen zusätzlichen Zeitaufwand durch ein Trial- und Error-Verfahren, da nicht jede spezielle Einstellung für alle Datensätze passend ist. Trotz des zeitlichen Mehraufwandes, konnten nie alle Fotos orientiert werden. Die Spezialisierungen widersprechen jedoch dem Vorsatz der Automatisierung und einfachen Handhabbarkeit. Meshroom stellt sich somit zwar als brauchbares Tool heraus, jedoch mit dem Nachteil, dass deutlich mehr manuelle Arbeit in jede einzelne Punktwolke gesteckt werden muss. Aus genannten Gründen wurde für die Punktwolkenerzeugung für alle weiteren Räume ausschließlich Agisoft Metashape gewählt. Für die Räume 9 und 10 wurden die Rekonstruktionen aus Meshroom ebenfalls weiterverarbeitet, die Raum-Abmessungen wiesen jedoch merklich größere Abweichungen von den wahren Maßen auf, als die aus Metashape, da aufgrund der unausgeglicheneren Punktverteilung Wände schlecht gefittet wurden. Die Ergebnisse der Raumgrößenabschätzung sind in Abschnitt 4.6 dargestellt. In Metashape wurden die Standard-Punktdichten verwendet, da diese in der Zeit-Leistungs-Abschätzung die besten Ergebnisse brachten.

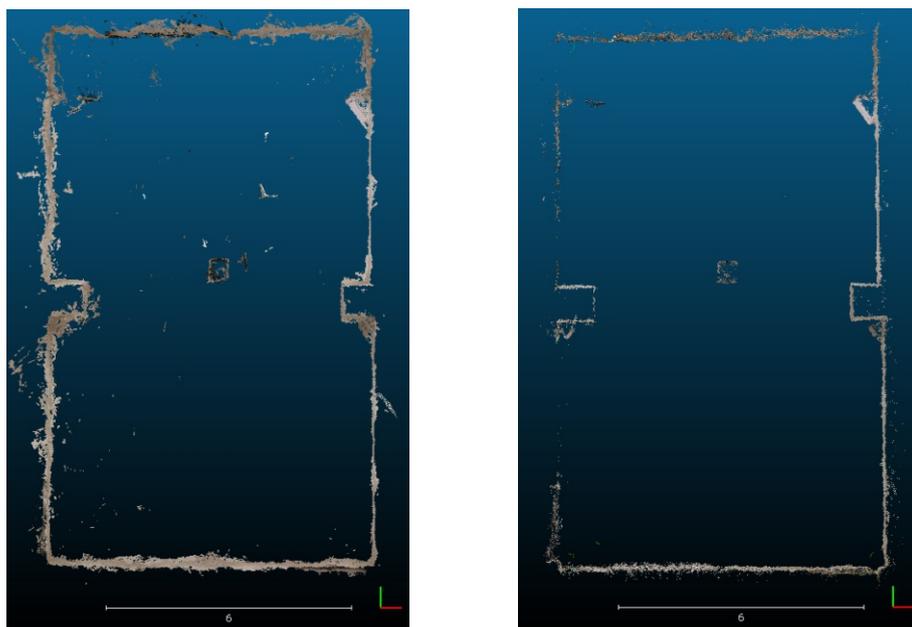


Abbildung 12: Punktwolke aus Raum 9: **Links:** Extrahierte Höhengleichung aus einer Auswertung durch Metashape mit gleichmäßiger Punktverteilung im gesamten Raum. **Rechts:** aus Auswertung durch Meshroom. Die linke Wand weist sehr wenige Punkte auf, da einige Fotos nicht orientiert werden konnten (vgl. Abbildung 9 links)

3.2 Orientierung mit Hilfe von Targets

Für die Orientierung wurden vier Targets (sowohl für Meshroom, als auch Metashape) verwendet, die in fixen, bekannten Abständen auf einem Plakat angebracht waren. Der

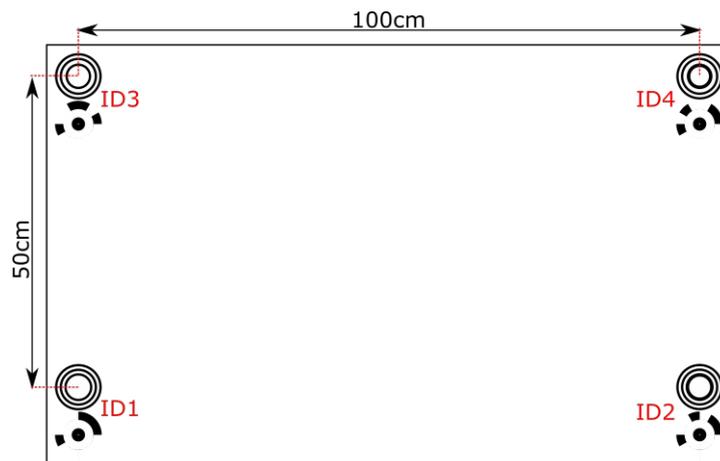


Abbildung 13: Schematische Darstellung des Plakates mit inkludierten Targets aus Meshroom und Metashape

Grund dafür war, dass es dadurch möglich war, alle vier beziehungsweise acht Targets sehr schnell und mit hoher Abstandsgenauigkeit anzubringen.

Wie in Abbildung 13 erkennbar, wurden für die Marker Abstände von 50 cm und 100 cm gewählt. Diese Referenzabstände waren somit verglichen zu den Raumgrößen sehr klein. Dadurch ergaben sich durch die Fehlerfortpflanzung kleine Probleme in der Skalierung größerer Räume. Trotzdem wurden auch für diese Räume diese Abstände gewählt, da die gesamte Arbeit sich mit der zeit- und arbeitseffizienten Bestimmung der Parameter beschäftigt. Ein Plakat mit 50 – 60 cm Breite kann sehr leicht gerollt und verstaut werden, während alles darüber hinaus schlicht zu groß für den „alltäglichen“ Gebrauch ist. Für diese Arbeit wurden Targets mit etwa 34 mm Radius verwendet, da die meisten Räume mit unter 5 m Länge dafür gut geeignet waren (Formel zur Bestimmung der Target-Größe: siehe Abschnitt 2.1.4). Bei langen Räumen wurden die Targets an der Längsseite der Wände fixiert, um die Kameradistanz auf die Breite des Raumes zu reduzieren.

3.3 Grobauswahl der Wände

Nach der Generierung der Punktwolke müssen die Wände grob vorausgewählt werden. Es existieren einige Ansätze zur automatischen Segmentierung von Punktwolken, wie sie etwa in [Rabbani et al., 2006](#) oder [Hoover et al., 1996](#) behandelt wurden. Die verwendeten Punktwolken waren aus ToF-Kameras oder Laserscannern, wodurch die Punktverteilung sehr gleichmäßig war. Im aktuellen Anwendungsfall war etwa die Segmentierung nach einem „Region-Growing-Algorithmus“ nicht möglich, da die SfM-Punktwolken der Räume dafür zu ungleichmäßig verteilt und lückenhaft waren. Somit musste auf andere Methoden wie RANSAC zurückgegriffen werden.

Ein einziger Durchlauf des RANSAC Algorithmus über die gesamte Punktwolke wäre nicht sinnvoll, weil so nur eine einzige Ebene durch den ganzen Raum gefittet würde. Aus diesem Grund muss eine Grobauswahl der Gesamtpunktwolke getroffen werden, die mit hoher Wahrscheinlichkeit mehrheitlich eine Wand enthält. Wie schon in Abschnitt 2.1.3 erwähnt, wurde die Punktwolke (allgemein durch die entsprechende Definition der

Target-Koordinaten) so orientiert, dass alle Wände etwa achsenparallel liegen. Wird für jede Achse ein Punktdichte-Histogramm ausgegeben, ergeben sich an den Positionen der Wände lokale Maxima, wie in Abbildung 14.

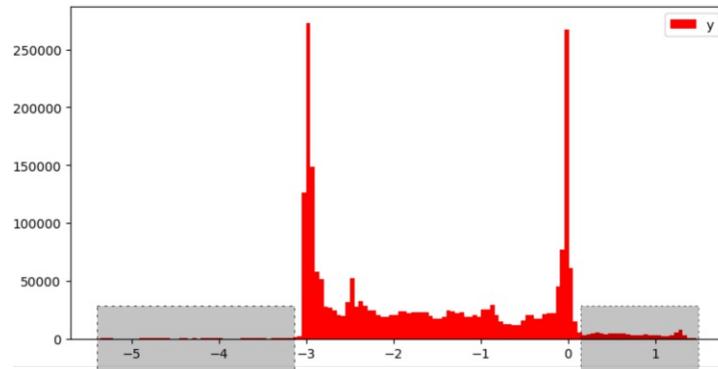


Abbildung 14: Histogramm mit vielen weit entfernt liegenden Punkten (Bin-width 5 cm)

Außerhalb der Wände liegen einige Ausreißer (in Abbildung 14 grau markiert), die eliminiert werden sollten. Deshalb wurde als erster Schritt die 1 % Perzentile der Punkte in jeder Achse entfernt. Dadurch wird das Histogramm nicht unnötig weit gespreizt.

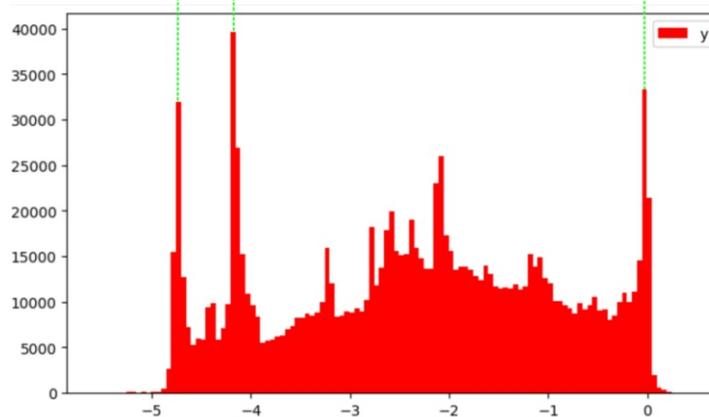
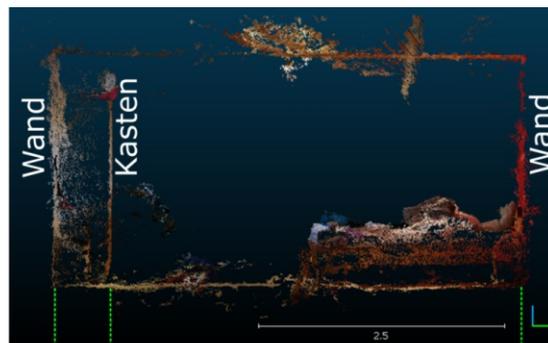


Abbildung 15: Ausschnitt der Punktwolke eines Raumes und dazugehörige Histogramm in y-Achse (Bin-width 5 cm)

In Abbildung 15 ist erkennbar, dass das globale Maximum nicht immer einer Wand entsprechen muss, da sehr stark texturierte Objekte wie Kästen sehr hohe Punktdich-

ten aufweisen können. Es muss also automatisch entschieden werden, welches Maximum tatsächlich einer Wand entspricht. Aus diesem Grund wird in den aktualisierten Histogrammen beiderseits in den äußersten 25 cm nach einem lokalen Maximum gesucht. Dieses wird anschließend mit dem globalen Mittelwert des Histogramms verglichen, um zu gewährleisten, dass das gefundene Maximum ausreichend deutlich ist. Das Maximum wird als relevant erachtet, wenn es größer als $\frac{2}{3}$ des globalen Mittelwerts ist. Dieser Vergleichswert ergab sich durch ein Trial- and Error-Verfahren. Ist die Bedingung nicht erfüllt, oder in anderen Worten, ist das lokale Maximum zu klein, wird der abgesuchte Bereich um 25 cm erweitert. Zum besseren Verständnis ist im Folgenden der Pseudocode dargestellt. Die gewählten Breiten der einzelnen Bins wurden auf 5 cm gesetzt. Die Schrittweite von 25 cm (entspricht 5 Bins) wurde gewählt, weil die meisten Kästen weiter als 25 cm aus der Wand ragen.

```
bins = 5
mean = mean(bin_values)
max = max(bin_values[1:bins])
while mean/max > 1.5:
    bins += 5
    max = max(bin_values[1:bins])
    extrema = position of found maxima
```

Das Ergebnis der Berechnung sind jeweils zwei Maxima in jeder Achse, in deren Umgebung mit hoher Wahrscheinlichkeit die Wand liegt. Um jedes Maximum wird anschließend ein Buffer von ± 10 cm gelegt und so eine Schicht von 20 cm aus der gesamten Punktwolke extrahiert, wie in Abbildung 16 ersichtlich.

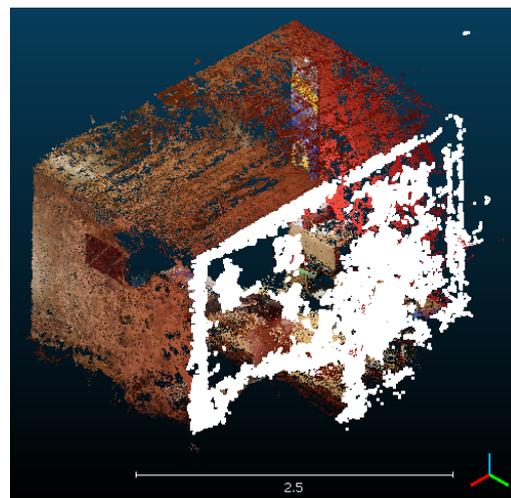


Abbildung 16: Durch Maximum extrahiertes Subset (weiß) der gesamten Punktwolke (eingefärbt)

Man erkennt gut, dass diese extrahierte Schicht zum Großteil Wandpunkte und nur wenige Ausreißer enthält. Die sechs extrahierten Subsets eignen sich anschließend gut für eine Ebenen-Schätzung mithilfe von RANSAC.

3.4 Bestimmung der Wandebene

Die sechs extrahierten Subsets der Wände wurden einzeln durch einen RANSAC Algorithmus (vgl. Abschnitt 2.2) gefiltert, um daraus jeweils eine Ebene zu fitten.

Die Settings für den Algorithmus wurden folgendermaßen gewählt:

- Bufferabstand (normal zur Ebene) für Inlier: $\pm 5 \text{ cm}$
- Abbruchkriterium:
 - mindestens 80 % Inlier-Anteil
 - maximale Iterationsanzahl: 10 % der Subsetpunktzahl

Nach Durchlauf des Algorithmus lagen ein Set von Inlier-Punkten, sowie Näherungswerte für die Ebenengleichung in Hesse-Normalform vor.

Diese Produkte wurden anschließend einem Ausgleich nach Methode kleinster Quadrate (vgl. Abschnitt 2.3) unterzogen, um die Ebene zu optimieren. Ergebnis des Ausgleichs waren die neuen Ebenenparameter und der RMSE der Punkte. Zusätzlich wurde eine Schicht der Punkte mit einem Abstand von $\pm 5 \text{ cm}$ um jede Ebene extrahiert, die anschließend als finale Wandpunkte klassifiziert wurden.

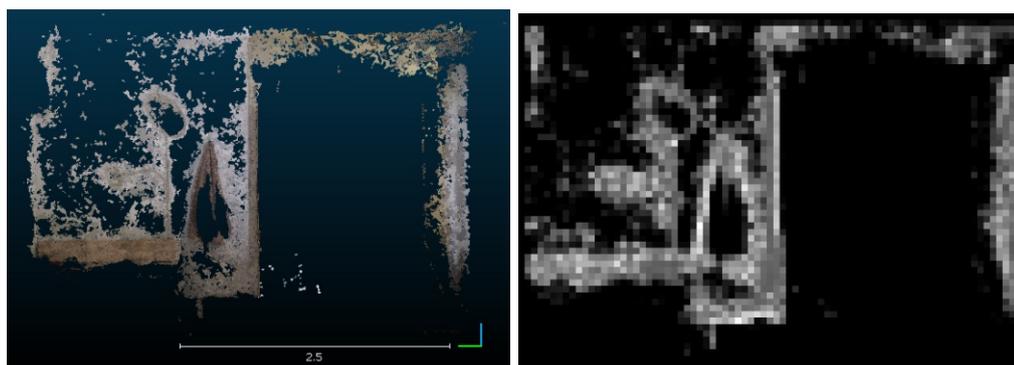
3.5 Qualitätskriterien

Für die Prüfung der Ergebnisse wird einerseits der RMSE der Ebenenausgleichung (siehe Abschnitt 2.3), andererseits Abdeckungskarten der gefitteten Wände verwendet. Durch diese ist erkenntlich, wie viel Prozent der Wand durch klassifizierte Wandpunkte abgedeckt sind und wie viele Punkte pro Pixel enthalten sind. Zusätzlich werden die Winkel der einzelnen Wände untereinander bestimmt, um mögliche falsch gefittete Wände aufzudecken.

Für die Erstellung der Abdeckungskarten werden jene Punkte, die als Wand klassifiziert wurden, für jede einzelne Wand extrahiert und in ein Raster-Bild mit Pixelgröße 5 cm umgewandelt, wie in Abbildung 17 dargestellt. Als Projektionsebene wurden die Koordinatenparallelen Ebenen gewählt, da die Punktwolken der Räume achsenparallel orientiert wurden. Die Graustufen wurden zwischen minimaler (0 Punkte pro Pixel) und maximaler Punktanzahl aufgespreizt, wie in Abbildung 17b erkenntlich. Anschließend wurden die Raster-Bilder für jede Wand als .png-Datei gespeichert.

Zusätzlich wurden die Winkel der Ebenen untereinander untersucht. Dafür wurde von jeder Ebene der Winkel zu allen anderen Ebenen bestimmt, wie in Tabelle 2 dargestellt. Bei normal aufeinander stehenden Ebenen wurden 90° subtrahiert, um leichter vergleichbare Werte zu erhalten. Zur Ausreißerdetektion wurde zusätzlich der Durchschnitt aller Winkel für jede Ebene berechnet. Ein Nachteil der Winkeluntersuchung ist bei älteren Gebäuden zu finden, deren Wände aufgrund der Bauweise oder des Alters tatsächlich nicht exakt normal oder parallel zueinander stehen. Dies muss in der Bewertung der einzelnen Winkel berücksichtigt werden.

Alle Qualitätsmaße (RMSE, Abdeckungskarten, prozentuelle Abdeckung und Winkel) werden für jeden Datensatz in separaten Dateien exportiert und sind frei ersichtlich. Dadurch kann auch nach der Verarbeitung die Genauigkeit und Zuverlässigkeit der Ergebnisse überprüft werden.



(a) Punktwolke vor der Umwandlung

(b) Rasterbild nach Umwandlung

Abbildung 17: Abdeckungskarte von Raum 3 mit Pixelgröße 5 cm

Tabelle 2: Raum 1: Winkel der einzelnen Wände untereinander, rechte Winkel wurden um 90° verringert! 1 + 2 lange Wand, 3 + 4 kurze Wand, 5 + 6 Boden und Decke

	1 [°]	2 [°]	3 [°]	4 [°]	5 [°]	6 [°]		∅ [°]
1	0.0	0.7	0.6	0.3	0.6	0.0		0.4
2	0.7	0.0	0.5	0.3	0.0	0.7		0.4
3	0.6	0.5	0.0	0.8	0.5	0.2		0.4
4	0.3	0.3	0.8	0.0	0.2	0.5		0.3
5	0.6	0.0	0.5	0.2	0.0	0.8		0.4
6	0.0	0.7	0.2	0.5	0.8	0.0		0.4

3.6 Abschätzung der Raumgröße

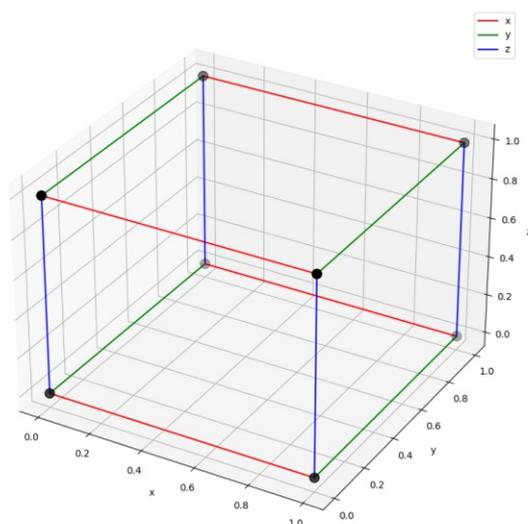


Abbildung 18: Raumkanten in allen Achsen

Als Ergebnis des Ausgleichs liegen die einzelnen sechs Wand-Ebenen in Hesse-Normalform vor. Um die Raumparameter zu schätzen, werden jeweils drei Ebenen verschnitten und daraus insgesamt acht Schnittpunkte (entsprechend den Raum-Eckpunkten) berechnet. Aus diesen acht Punkten können dann die Abstände in den drei Achsrichtungen be-

stimmt werden, wodurch sich zwölf unabhängige Längenmaße ergeben, wie in Abbildung 18 ersichtlich. Um ein eindeutiges Ergebnis zu erhalten, werden die vier Längen einer Achse anschließend noch gemittelt.

Die Standardabweichung des Mittels aller vier Längen ist in Kapitel 4, Tabelle 7 abzulesen.

4 Experimente und Diskussion

4.1 Untersuchte Räume

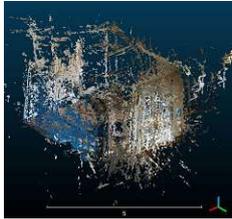
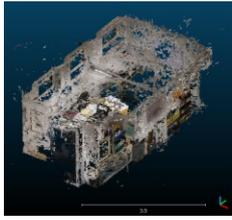
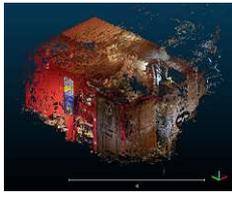
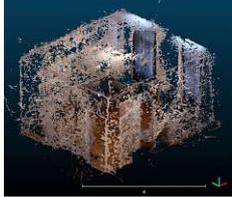
Zur Überprüfung des oben beschriebenen Workflows zur Ermittlung von Raumabmessungen wurden zehn verschiedene Räume photogrammetrisch rekonstruiert und anschließend vollautomatisch deren Dimensionen bestimmt.

Die Räume unterschieden sich sowohl in Größe, als auch Ausstattung, um möglichst viele Eventualitäten abzudecken. Der kleinste Raum hatte Abmessungen von $2.94 \times 2.83 \times 2.76 \text{ m}$ ($L \times B \times H$), der größte $13.24 \times 7.83 \times 2.95 \text{ m}$ ($L \times B \times H$). Sieben der Räume hatten glatte, weiße Wände, die drei restlichen hatten mindestens eine farbige Wand, die Böden waren von unterschiedlicher Beschaffenheit wie Holz, Teppich oder PVC, während die Decken meist weiß und schlecht bis gar nicht texturiert waren. In den Tabellen 3 und 4 sind alle Räume, sowie die daraus generierten Punktwolken abgebildet. Zusätzlich sind die Soll-Maße abzulesen, die mithilfe von Maßbändern und Laser-Distanzmessern bestimmt wurden.

Tabelle 3: Räume 1 - 4 mit gemessenen Größen. Alle Räume mit Fenstern und natürlicher und künstlicher Beleuchtung

Nr.	Raum	Punktwolke	L [m]	B [m]	H [m]	Fotos
1			3.90	3.43	2.65	48
2			2.83	2.94	2.77	43
3			4.04	3.38	2.76	57
4			4.11	3.05	2.44	61

Tabelle 4: Räume 5 - 9 mit gemessenen Größen. Raum 9 und 10 hatten ausschließlich künstliche Beleuchtung

Nr.	Raum	Punktwolke	L [m]	B [m]	H [m]	Fotos
5			3.88	3.41	2.34	42
6			5.00	2.98	2.15	49
7			4.70	2.91	2.33	60
8			4.70	4.12	2.83	50
9			13.24	7.83	2.95	42
10			13.24	6.27	2.95	39

4.2 Aufnahme der Fotos

Als Kamera wurde die Handykamera des **Huawei P20 Pro** (Datasheet: <https://www.devicespecifications.com/de/model/6fe84965>) verwendet. Dieses Smartphone verwendet ein Triple-Sensor-System aus einem ~ 40 MP Hauptsensor, einem ~ 8 MP Telefoto-Sensor und einem ~ 20 MP Monochrom-Sensor. Die Aufzeichnungen aller drei Sen-

soren werden durch die Smartphonesoftware kombiniert und im Automatik-Modus direkt optimiert und nachgeschärft. Der genaue Vorgang ist nicht bekannt, laut Hersteller (<https://consumer.huawei.com/at/>) wird jedoch eine AI-Unterstützung verwendet.

Die Kamera wurde im Automatik-Modus betrieben, wodurch sich klarerweise für jedes Bild unterschiedliche Blenden und Schärfentiepen-Werte ergaben und sich auch die innere Orientierung der Kamera änderte. Ein grober Überblick über die Spezifikationen der Kamera ist in Tabelle 5 zu finden.

Tabelle 5: Spezifikationen der verwendeten Kamera (entnommen aus dem oben angeführten Datasheet)

	MP	f (Blende)	Brennweite	ISO	Sensorformat
Huawei P20 Pro	~39.9	1.8 - 2.4	27 mm	50 - 102400	1/1.78"

Einer der wichtigsten Schritte ist die Aufnahme der Fotos. Je höher die Auflösung am Objekt im Bild ist, desto mehr Features können gefunden und gematcht werden, da feinere Texturen unterschieden werden können. Auch die Wahl der Kamerapositionen spielt eine wichtige Rolle, um schleifende Schnitte zu vermeiden. Aus diesem Grund wurde die in Abschnitt 2.1.1 empfohlene und in Abbildung 2b dargestellte Konstellation gewählt. Zumindest an den Wänden wird durch diese Konstellation eine relativ große Baseline zwischen den einzelnen Fotos erreicht und dadurch schleifende Schnitte vermieden. In den Ecken lassen sich bei dieser Konstellation kleine Baselines nicht vermeiden. Die Baselines wurden so gewählt, dass jedes Foto eine Überlappung von etwa 70 - 90 % (vgl. Abschnitt 2.1.1) zu seinen „Nachbarn“ aufwies. Dadurch ergaben sich entlang der Wände Baselines von etwa 50 – 150 cm, und zwischen 40 - 61 Fotos für jeden Raum.

Die Fotos wurden alle bei Tageslicht und zusätzlicher künstlicher Beleuchtung aufgenommen, nur Raum 9 und 10 hatten ausschließlich künstliche Beleuchtung, da diese Räume keine Fenster hatten.

4.3 Erzeugung der Punktwolke

Die Datensätze aller zehn Räume wurden anschließend in Metashape verarbeitet. Die erreichten Parameter der Berechnung aus Metashape sind in Tabelle 6 zu finden. Folgende Parameter wurden zum Vergleich herangezogen:

- Verknüpfungspunkte:
 - Punktzahl der Sparse-Cloud
- Ground-Resolution:
 - durchschnittliche Größe der Pixel am Target
- Target-RMSE:
 - RMSE der einzelnen Targets, die für die Orientierung verwendet wurden
- Reprojection-Error
- Punktzahlen:
 - Punktzahl der Dense-Cloud

- Wandanteil:
 - Prozentueller Anteil an Wandpunkten innerhalb der Dense-Cloud
- Punktdichte an den Wänden:
 - Anzahl der Wandpunkte dividiert durch geschätzte Wandfläche = durchschnittliche Punktzahl pro Quadratmeter Wand
- Benötigte Rechenzeit

Tabelle 6: Verknüpfungspunkte (VP), Ground-Resolution (GR), Target-RMSE (TR), RE, Punktzahlen und Wandpunkte, sowie mittlere Punktdichten pro Quadratmeter Wand (PD) der Datensätze aller Räume.

Raum	1	2	3	4	5	6	7	8	9	10
VP (Tsd.)	40.6	15.6	50.0	57.4	31.2	27.8	21.5	12.3	16.3	19.7
GR [mm/px]	0.52	0.48	0.43	0.42	0.51	0.46	0.89	0.57	1.15	1.04
TR [px]	1.02	1.22	2.29	2.03	1.02	0.48	1.72	1.39	0.47	0.57
RE [px]	2.58	4.33	2.52	2.74	8.78	2.65	5.80	7.09	2.50	2.09
Punkte (Mio.)	3.83	2.38	4.64	5.56	3.03	3.76	1.23	3.20	2.73	2.15
davon Wand (%)	0.59	0.71	0.54	0.58	0.67	0.51	0.51	0.59	0.39	0.57
∅ PD (Tsd.)	35.97	35.14	38.98	54.47	33.48	32.16	9.51	21.65	3.94	8.17
Zeit (h:min)	1:30	0:48	1:32	1:42	0:55	1:06	0:19	1:01	1:10	1:05

Auffällig ist, dass die Punktzahl in Raum 4 sehr hoch ist. Dies ist auf den Umstand zurückzuführen, dass große Teile der Wände und Decke des Raumes mit Holz verkleidet sind und demnach sehr stark texturiert sind. Die höchsten Wandanteile liegen in den beiden Räumen 2 und 5 vor, was darauf zurückzuführen ist, dass diese Räume nur sehr schwach möbliert waren. Die Punktdichte an den Wänden hängt stark von deren Beschaffenheit ab. Raum 4 hatte zum Beispiel gut texturierte Wände, weshalb auffällig viele Punkte pro Quadratmeter Wand rekonstruiert werden konnten. In Raum 9 und 10 waren die Wände nahezu vollkommen weiß und glatt, weshalb die durchschnittlichen Punktdichten pro Quadratmeter auffällig gering ausfielen.

4.4 Ebenen-Fitting

Nach der Punktwolkenerzeugung wurden die bereits orientierten und skalierten Punktwolken in den in Abschnitt 3.4 beschriebenen Workflow zur Bestimmung der Raumgröße eingebracht. Es wurden, wie beschrieben, die einzelnen Achsen separat untersucht, da sich durch die achsenparallele Orientierung der Punktwolke die Wände parallel zu den jeweiligen Hauptebenen befinden. Nach der Entfernung der 1 %-Perzentile wurden die Punktdichten der jeweiligen Koordinaten auf den zugehörigen Achsen untersucht.

Durch den vorgestellten Algorithmus konnten in allen Datensätzen alle Wandpositionen entlang der Achsen richtig abgeschätzt werden. In Abbildung 19 sind einige herausfordernde Histogramme abgebildet. Abbildung 19a, 19b, 19c und 19e zeigen Histogramme von Punktwolken aus Metashape, bei denen die Abschätzung korrekt funktionierte. Konkret handelt es bei 19a um den Fall, der auch in Abbildung 3 und 15 gezeigt wird. In diesen ist der Kasten sehr stark texturiert und weist viele Features auf, während die Wand vergleichsweise kahl ist. Ein ähnliches Bild zeigt sich auch in Abbildung 19b und 19c aus

den beiden Testräumen des Softwarevergleichs. Abbildung 19d zeigt anschließend noch die x-Achse der Punktwolke aus Meshroom von Raum 9, die nicht korrekt abgeschätzt werden konnte. Wie man gut erkennen kann, ist die linke Wand nur durch sehr wenige Punkte abgedeckt und kann demnach vom Algorithmus auch nicht erkannt werden. Dies war der Hauptgrund, die Punktwolken aus Meshroom zu verwerfen. Vergleichsweise ist in Abbildung 19e das Histogramm aus Metashape abgebildet, bei dem eindeutig die Wände erkennbar sind.

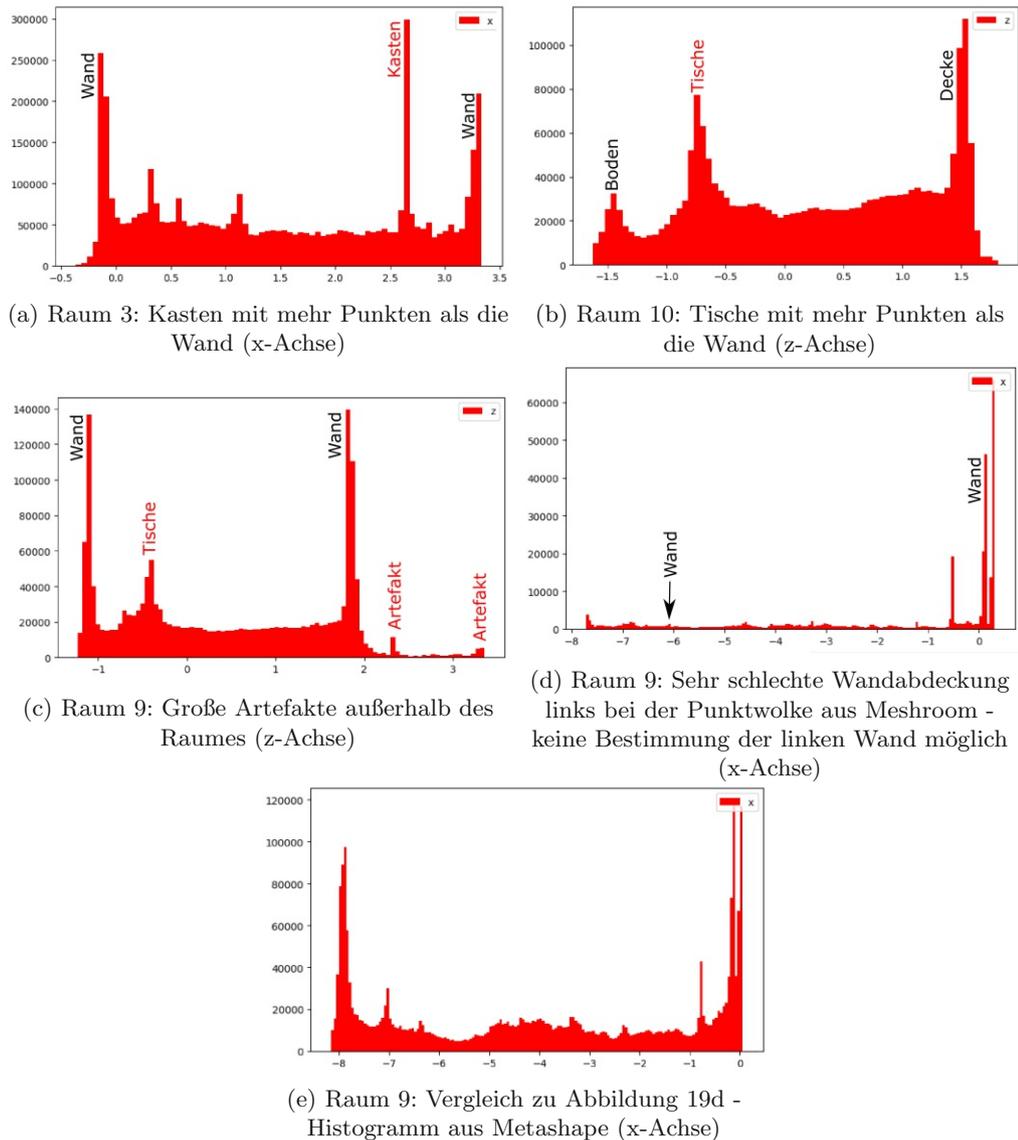


Abbildung 19: Schwierigkeiten bei der Bestimmung der Wandposition aus Histogrammen mit Bin-width 5 cm

Die extrahierten Schichten mit 20 cm Stärke (± 10 cm um den Näherungswert) wurden anschließend der RANSAC-Ebenenschätzung weitergegeben. Die Ebenenschätzung aus RANSAC wurde zu Vergleichszwecken ebenfalls für eine Raumgrößen-Abschätzung herangezogen (Ergebnisse in Tabelle 7). Anschließend wurden die bestimmten Werte dem LSM-Ausgleich übergeben. Das Ergebnis des LSM-Ausgleichs sind die Ebenen-Parameter

in Hesse-Normalform, der RMSE und eine neue Punktwolke, bestehend aus allen Punkten innerhalb eines Buffers von 5 *cm*. Diese wurden als Wandpunkte klassifiziert. Die genauen Abläufe bei der RANSAC-Ebenenschätzung und dem LSM-Ausgleich wurden bereits in den vorherigen Kapiteln diskutiert. Die Ergebnisse der Abschätzung sind in Tabelle 7 abzulesen.

Zusätzlich zum RMSE wurde die prozentuelle Abdeckung der Inlier jeder Wand berechnet. Wie schon in Abbildung 17 gezeigt, wurde eine Abdeckungskarte erstellt (alle Abdeckungskarten sind im Anhang zu finden), aus der auch das Verhältnis zwischen abgedeckten und leeren Bereichen bestimmt werden konnte. Im Durchschnitt über alle Datensätze gerechnet waren etwa 46% der Wände mit Punkten bedeckt. Tendenziell wiesen die Decken die wenigsten Punkte auf, was durch die fehlende Textur und die Orientierung der Kameras begründet werden kann. Da die Fotos immer von der Wand mit Blick ins Raumzentrum und horizontal aufgenommen wurden, lag die Decke immer am Bildrand. Zwei Ausnahmen waren Raum 4 und 7, bei denen die höchsten Punktzahlen an den Decken lagen. Der Grund dafür ist, dass beide Räume eine Holzvertäfelung an der Decke hatten, die dementsprechend viel Textur aufwies. Raum 10 wies an einer Wand nur eine Abdeckung von etwa 9 % auf, was unter anderem zu einem etwas falschen Fitting der Wandebene führte. In Abschnitt 4.5 wird noch näher auf den Fall eingegangen.

Beim abschließenden Verschnitt der Ebenen (siehe Abschnitt 3.6), wurden jeweils die Längen der vier parallelen Kanten gemittelt. Die erhaltenen Abweichungen zur manuell gemessenen Raumgröße lagen dabei im Bereich von 1 *cm* bis 12 *cm*. Tendenziell wies die *z*-Achse die geringsten Abweichungen zur Ground-Truth auf. Die ebenfalls bestimmten Standardabweichungen der Längenmittelung sind ebenso wie die Abweichungen in Tabelle 7 präsentiert.

Tabelle 7: Klassifizierte Wandpunkte pro Wand (P/W) nach LSM, RMSE der Ebenenausgleiche, prozentuelle Abdeckung der Wände, Standardabweichung der vier parallelen Raumkanten σ Raumlängen, Längenabweichungen Δ und Durchschnitt \varnothing vor LSM und nach LSM zu den wahren Raumgrößen (Soll) aller Räume.

Raum	1	2	3	4	5	6	7	8	9	10
\varnothing P/W (Tsd.)	376.46	282.26	420.76	534.14	341.58	320.42	104.94	324.80	178.22	205.99
min. P/W (Tsd.)	310.81	103.85	125.71	403.34	82.27	207.85	36.03	142.79	67.23	26.67
max. P/W (Tsd.)	510.96	455.20	627.91	641.75	899.05	477.83	189.66	702.03	296.65	463.23
\varnothing RMSE [m]	0.022	0.023	0.024	0.020	0.025	0.022	0.023	0.024	0.026	0.024
min. RMSE	0.018	0.019	0.020	0.018	0.020	0.019	0.020	0.020	0.023	0.022
max. RMSE	0.028	0.027	0.028	0.022	0.027	0.022	0.025	0.026	0.028	0.027
\varnothing Abdeckung (%)	0.60	0.49	0.49	0.62	0.57	0.49	0.49	0.51	0.39	0.40
min. Abd. (%)	0.46	0.12	0.15	0.45	0.18	0.31	0.30	0.34	0.18	0.09
max. Abd. (%)	0.77	0.80	0.75	0.86	0.91	0.66	0.75	0.72	0.52	0.79
σH [m]	0.018	0.055	0.044	0.044	0.061	0.037	0.028	0.008	0.073	0.034
σL [m]	0.021	0.040	0.038	0.024	0.046	0.030	0.065	0.062	0.042	0.022
σB [m]	0.028	0.021	0.023	0.008	0.021	0.051	0.023	0.033	0.022	0.015
vor LSM										
ΔH [m]	0.021	0.026	0.019	0.049	0.129	0.016	0.004	0.040	0.011	-0.002
ΔL [m]	0.036	0.002	0.105	0.022	0.017	0.006	0.063	0.028	0.079	0.135
ΔB [m]	0.033	0.045	0.001	0.046	-0.043	0.028	0.041	0.011	0.029	0.023
$\Delta \varnothing$ [m]	0.030	0.024	0.042	0.039	0.063	0.016	0.036	0.026	0.040	0.053
nach LSM										
ΔH [m]	0.006	0.014	0.011	0.021	0.032	0.010	0.000	0.031	0.000	-0.004
ΔL [m]	0.022	-0.023	0.087	0.031	0.016	-0.003	0.054	0.031	0.079	0.123
ΔB [m]	0.016	0.026	-0.011	0.008	-0.022	0.009	0.024	0.013	0.014	0.010
$\Delta \varnothing$ [m]	0.015	0.021	0.036	0.020	0.023	0.007	0.026	0.025	0.031	0.046
H Ist [m]	2.644	2.751	2.749	2.424	2.308	2.140	2.330	2.799	2.950	2.954
L Ist [m]	3.886	2.963	3.953	3.024	3.857	5.003	4.651	4.674	13.161	13.117
B Ist [m]	3.414	2.811	3.388	4.097	3.431	2.971	2.886	4.102	7.816	6.260
H Soll [m]	2.650	2.765	2.760	2.445	2.340	2.150	2.330	2.830	2.950	2.950
L Soll [m]	3.908	2.940	4.040	3.055	3.873	5.000	4.705	4.705	13.240	13.240
B Soll [m]	3.430	2.837	3.377	4.105	3.409	2.980	2.910	4.115	7.830	6.270

4.5 Qualitätsüberprüfung

Wie man in Tabelle 7 erkennen kann, ist die durchschnittliche Längenabweichung $\Delta\emptyset$ vor dem LSM-Ausgleich in allen Räumen merklich größer, als nach dem Ausgleich. Somit ist offensichtlich, dass der Ausgleich sinnvoll ist.

Zwei der größten Abweichungen zur Ground-Truth nach dem LSM-Ausgleich entstanden durch sehr verrauschte Punktwolken. Diese sind in Abbildung 20 dargestellt. Wie man erkennt, sind die oberen Wände in der Abbildung in beiden Räumen sehr stark verrauscht. In Raum 9 wurde die Raumlänge deshalb um etwa 8 cm, in Raum 10 um 12.3 cm zu kurz geschätzt. Dies ist auch am RMSE erkennbar. Der maximale RMSE (von etwa ~ 3 cm) in beiden Räumen befindet sich in beiden Fällen exakt bei der betroffenen Wand. Um die Qualität der Punktwolke in diesen Bereichen zu erhöhen, könnten mehr Detailfotos des betroffenen Bereichs gemacht werden. Durch Hinzunahme zusätzlicher Fotos kann sowohl die Punktdichte, als auch die Genauigkeit erhöht werden.

Eine weitere etwas größere Abweichung von 8.7 cm entstand in Raum 3 durch einen Heizkörper. Durch Objekte wie einen Kasten, einen Tisch und Sessel konnten nur relativ wenige Punkte im unteren Bereich der Wand abgeschätzt werden, wodurch das Fitting durch den Heizkörper nach vorne „gezogen“ wurde.

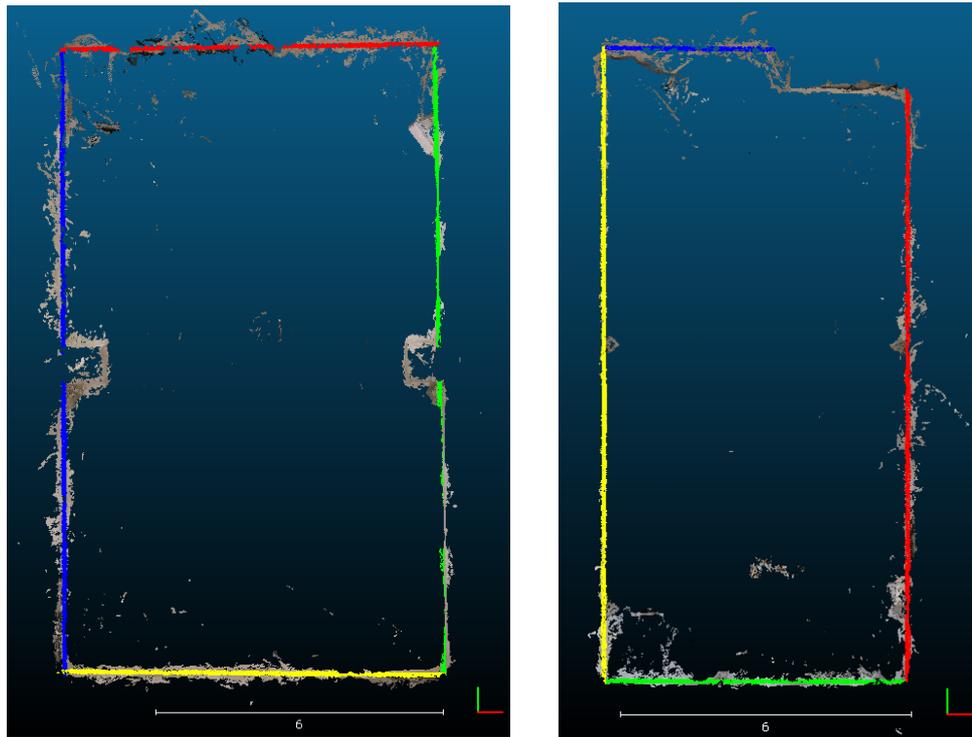
Die schief liegende Wand kann auch in der Untersuchung der Winkel der Wände (siehe Abschnitt 3.5) zueinander nachgewiesen werden. In Tabelle 8 ist erkennbar, dass die beiden kurzen Wände nicht parallel liegen. Außerdem erkennt man gut, dass Ebene 3 auch zu Boden und Decke relativ große Winkelabweichungen aufweist. In den gemittelten Winkeln findet sich das Maximum ebenfalls exakt in Ebene 3. Die genauen Umstände sind in Abbildung 21 dargestellt.

Tabelle 8: Raum 3: Winkel von Wand 3 zu den einzelnen Wände untereinander, rechte Winkel wurden um **90° verringert!** 1 + 2 lange Wand, 3 + 4 kurze Wand, 5 + 6 Boden und Decke

	1 [°]	2 [°]	3 [°]	4 [°]	5 [°]	6 [°]	\emptyset [°]
1	-	0.6	1.0	0.9	1.0	1.0	0.8
2	0.6	-	0.9	0.7	0.6	0.6	0.7
3	1.0	0.9	-	1.4	1.8	1.2	1.0
4	0.9	0.7	1.4	-	0.4	0.2	0.6
5	1.0	0.6	1.8	0.4	-	0.6	0.7
6	1.0	0.6	1.2	0.2	0.6	-	0.6

4.6 Punktwolken mit erhöhter Punktdichte

Zuletzt sind noch die zusätzlichen Ergebnisse (Punktwolken aus Meshroom und Metashape mit hoher Punktdichte) der beiden Testräume aus dem Softwarevergleich abgebildet. In den Spalten sind jeweils die Ergebnisse des Ebenen-Fittings aus Metashape mit hoher Punktdichte und Meshroom dargestellt. Bei den Fittings aus Metashape fällt auf, dass die Punkte pro Wand zwar deutlich höher ausfallen, die prozentuelle Wandabdeckung jedoch gleich oder sogar schlechter ausfällt (siehe Tabelle 10). Auch die Längenabschätzung der Räume verschlechtern sich verglichen zur Standard-Punktdichte. Dies liegt daran, dass ein großer Anteil der Wände (siehe Abbildung 20) durch stark verrauschte Punkte

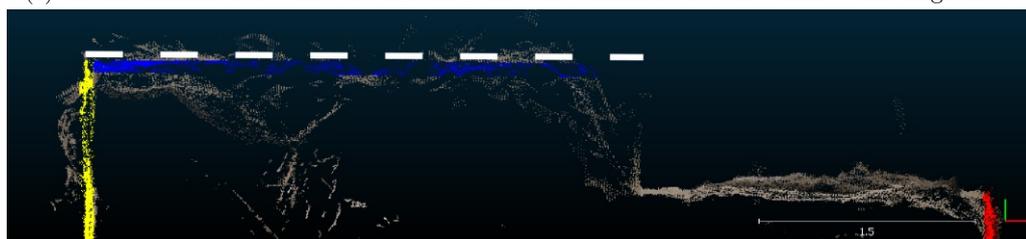


(a) Raum 9 mit sehr stark verrauschter Wand (rot)

(b) Raum 10 mit sehr stark verrauschter Wand (blau)



(c) Großaufnahme der Wand aus Raum 9. In weiß-strichliert ist die wahre Wand abgebildet



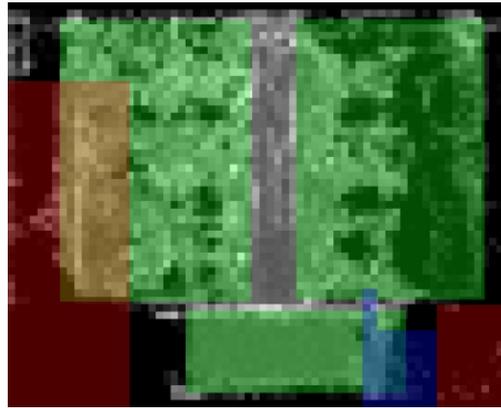
(d) Großaufnahme der Wand aus Raum 10. In weiß-strichliert ist die wahre Wand abgebildet

Abbildung 20: Schwierigkeiten bei der Bestimmung der Wandposition durch verrauschte Punktwolken. Eingefärbte Punkte entsprechen einem Buffer von 5 cm um die gefittete Wand

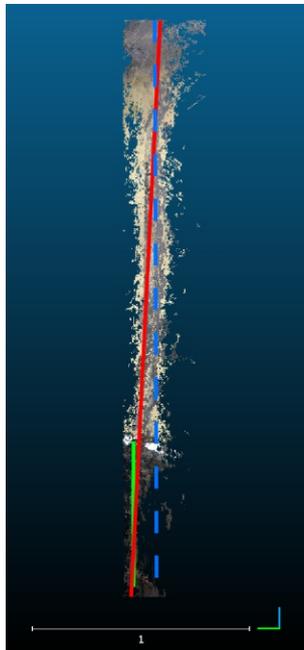
repräsentiert wird. Durch die erhöhte Punktdichte wird nicht die Genauigkeit erhöht, sondern es werden lediglich mehr Punkte erzeugt. Daher wird das Fitting noch weiter in den Raum „gezogen“, als bei normaler Punktdichte. (Anmerkung: Für mit Laserscanner aufgezeichnete Datensätze wäre eine dichtere Punktwolke in der Regel vorteilhaft, da ein Laserscanner nicht auf die Textur des Objektes angewiesen ist, sondern auch an



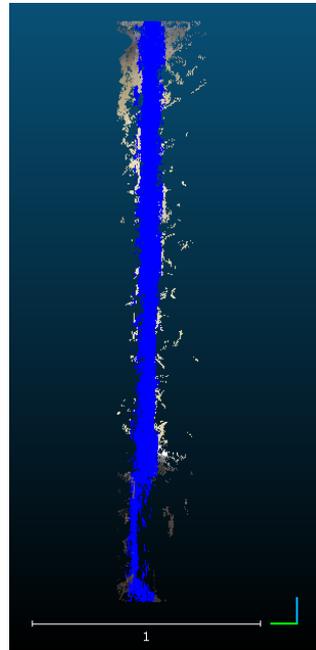
(a) Raum 3 mit vielen, abschattenden Objekten



(b) Abdeckungskarte von Raum 3 mit eingeblendeten, abschattenden Elementen; grün: knapp vor der Wand liegende Objekte, rot: Kasten/Tisch, blau: Sessel



(c) wahre Wand (blau-strichliert), Heizkörper (grün) und gefittete Wand (rot)



(d) Punkte-Buffer von 5 cm um die gefittete Wand (blau) vor der wahren Wand (beige)

Abbildung 21: Schwierigkeiten bei der Bestimmung der Wandposition durch Objekte knapp vor der Wand

einfärbigen nicht texturierten Wänden Punkte messen kann).

In Meshroom konnten die Längen der Räume nicht korrekt abgeschätzt werden, da (wie in Abschnitt 3.1 beschrieben) eine Wand nahezu gar nicht rekonstruiert wurde. Dies führte auf ein nahezu „zufälliges“ Ebenen-Fitting, das durch einige Artefakte entstand. Der RMSE steigt zwar im Vergleich zu den anderen Werten aus Tabelle 7 nicht nennenswert an, dies kann jedoch dadurch begründet werden, dass für den LSM-Ausgleich nur eine sehr begrenzte Schicht der Gesamtpunktwolke herangezogen wurde. Wie schon zuvor in Raum 3 (Tabelle 8) präsentiert, konnte auch in Raum 10 die falsch gefittete Wand aufgrund der Winkelabweichung identifiziert werden. Dies ist in Tabelle 9 dargestellt. In

rot sind die größten Winkelabweichungen markiert, die den Schluss zulassen, dass Ebene 2 falsch gefittet wurde.

Tabelle 9: Winkel der einzelnen Wände aus Raum 10 untereinander, rechte Winkel wurden um **90° verringert!** 1 + 2 lange Wand, 3 + 4 kurze Wand, 5 + 6 Boden und Decke

	1 [°]	2 [°]	3 [°]	4 [°]	5 [°]	6 [°]	∅ [°]
1	0.0	2.6	0.1	0.0	0.3	0.4	0.6
2	2.6	0.0	2.6	2.6	0.3	0.4	1.4
3	0.1	2.6	0.0	1.0	0.6	0.2	0.7
4	0.0	2.6	1.0	0.0	1.6	1.2	1.0
5	0.3	0.3	0.6	1.6	0.0	0.4	0.5
6	0.4	0.4	0.2	1.2	0.4	0.0	0.4

Tabelle 10: Klassifizierte Wandpunkte pro Wand (P/W) nach LSM, RMSE der Ebenenausgleiche, prozentuelle Abdeckung der Wände und Längenabweichungen zu den wahren Raumgrößen der Testräume aus Abschnitt 3.1.

Raum	Metashape 9 HP	Meshroom 9	Metashape 10 HP	Meshroom 10
∅ P/W (Tsd.)	930.40	26.65	1309.14	49.43
min. P/W (Tsd.)	390.44	5.11	60.71	5.89
max. P/W (Tsd.)	1896.07	76.13	3720.06	192.76
∅ RMSE [m]	0.029	0.020	0.025	0.021
min. RMSE	0.026	0.011	0.021	0.012
max. RMSE	0.036	0.026	0.029	0.029
∅ Abdeckung (%)	0.26	0.27	0.27	0.34
min. Abd. (%)	0.18	0.12	0.09	0.11
max. Abd. (%)	0.47	0.43	0.61	0.71
ΔX [m]	0.037	-0.010	0.018	0.010
ΔY [m]	0.246	2.218	0.350	1.074
ΔZ [m]	-0.035	-0.088	0.199	0.026

Abschließend ist zu sagen, dass die Änderungen in der Punktdichte an den Wänden und die Streuung normal zu den Wänden ausschlaggebend für das erfolgreiche Fitting der Wandebenen ist. Jene Wände mit starker Streuung wiesen die höchsten Abweichungen von der Ground-Truth auf. Der Grund dafür ist in der schwachen Textur der weißen Wände zu finden. Ein weiterer erschwerender Faktor sind wandnahe Objekte wie Heizkörper, die das Fitting der Wand beeinflussen.

Auch die Punktwolken mit erhöhter Punktdichte brachten an diesen Wänden keine besseren Ergebnisse, da zwar mehr Punkte rekonstruiert wurden, diese jedoch ebenfalls sehr stark gestreut waren. Somit bedeutet eine dichtere Punktwolke nicht zwingend, dass die Qualität auch steigt.

5 Zusammenfassung und Schlussfolgerung

In dieser Arbeit wurde ein Programm entwickelt, um aus SfM-Punktwolken die Raumgrößen einzelner Räume robust und effizient abzuschätzen. Das Programm wurde auf zehn verschiedenen Räume angewandt. Dafür mussten zuerst Fotos von den Räumen aufgenommen und daraus mithilfe von SfM (vgl. Abschnitt 2.1) eine Punktwolke berechnet werden. Nach der achsenparallelen Orientierung und Skalierung der Punktwolke wurden im Programm zuerst die Wände grob abgeschätzt und anschließend durch ein RANSAC-Ebenen-Fitting und einen LSM-Ausgleich in jede Wand eine Ebene gefittet. Zuletzt wurden die Ecken des Raumes durch den Verschnitt aller Ebenen bestimmt und daraus die Raumgröße abgeleitet.

Das eigentliche Programm zur Abschätzung der Raumgröße benötigt ausschließlich eine korrekt skalierte und achsenparallel orientierte Punktwolke. Nach der Punktwolkenableitung und -orientierung erfolgt vollautomatisch die Abschätzung der Raumgröße und einiger Zusatzinformationen. Die Outputs dieses Programms sind folgende:

- Dimensionen des Raumes in Länge/Breite/Höhe
- Eckpunkte des Raumes in XYZ
- Ebenenparameter in Hesse-Form
- RMSE jeder gefitteten Ebene
- Winkel aller gefitteten Ebenen untereinander
- Punktdichtekarte jeder gefitteten Ebene als Rasterdatei
- Prozentuelle Abdeckung jeder Wand durch Punkte
- Punktwolke mit klassifizierten Wänden

Somit ist auch nach der Prozessierung ein Einblick in die Qualität der Abschätzung möglich.

Bei der Grobauswahl der Wände wurden in allen Punktwolken, die mit der Software Agisoft Metashape generiert wurden, die Wandpositionen korrekt abgeschätzt. Dabei wurden erfolgreich Artefakte und andere Punktanhäufungen, die nicht von Wänden stammten, eliminiert.

Das darauf folgende RANSAC-Ebenen-Fitting erreichte in fast 90 % aller Fälle Längenabweichungen von weniger als 5 cm zu den wahren, manuell gemessenen Raummaßen. Die Inlier des RANSAC-Algorithmus wurden anschließend einem LMS-Ausgleich übergeben, um die Ebene zu optimieren. Nach diesem Schritt erreichten die Längenabweichungen in 80 % aller Fälle weniger als 3 cm Längenabweichung zu den wahren Raummaßen.

Die Räume mit größerer Längenabweichung wiesen jeweils eine Wand auf, die „schlecht“ gefittet wurde. Die entsprechenden Wände konnten jeweils im RMSE des Ausgleichs und den Winkeln aller gefitteten Ebenen untereinander als Ausreißer nachgewiesen werden. Die Gründe für „falsche“ Ebenen sind einerseits in starken Punktstreuungen normal zur Wand zu finden. Andererseits wurden Ebenen verzerrt, wenn Objekte nur einige cm aus der Wand ragten, oder die Wände nur in sehr geringem Maß sichtbar waren. Letzterer

Fall trat zum Beispiel durch viele Möbelstücke und Fensteröffnungen auf, wodurch sich der wahre, „sichtbare“ Teil der Wand stark reduzierte.

Zusammenfassend lässt sich sagen, dass die Raumgrößen-Abschätzung durch SfM-Punktwolken von vielen Einflüssen erschwert wird. Die Schwierigkeiten bestanden hauptsächlich darin, die richtigen Schlüsse aus den unregelmäßig verteilten Punktwolken zu ziehen. Durch den stufenweisen Aufbau des Programms werden zuerst grobe Fehler ausgeschlossen und anschließend schrittweise die Ebenen angenähert. Dadurch ist trotz der erschwerten Bedingungen eine relativ gute Abschätzung der Raum-Maße möglich.

Eine zusätzliche Herausforderung sind größere Räume, da in diesen große Distanzen zwischen Projektionszentrum und Objekt auftreten. Dadurch kann es passieren, dass einige Wandregionen nur sehr verrauscht rekonstruiert und dementsprechend die Ebenen nicht optimal gefittet werden können. Um dem vorzubeugen, könnte in weiteren Arbeiten eine erweiterte Konstellation für die Fotoaufnahme untersucht werden, wie in Abbildung 22 dargestellt.

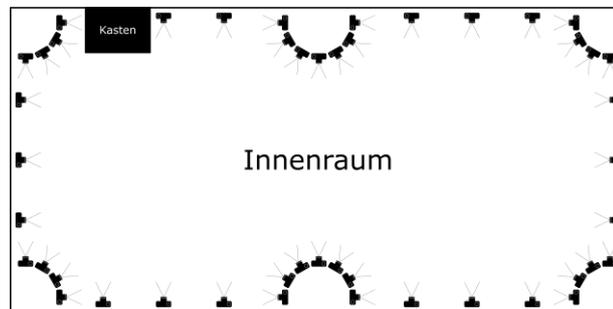


Abbildung 22: Erweiterte Konstellation für große Räume

Durch diese erweiterte Anordnung werden überdurchschnittlich lange Objekt-Distanzen vermieden und gleichzeitig die erzeugten Punktwolken verdichtet.

Allgemein lässt sich sagen, dass die Verarbeitung von SfM-Punktwolken im Innenraum eine große Herausforderung darstellt. Die Erzeugung ist zwar kostengünstiger als zum Beispiel die einer Laserscanning-Punktwolke, dafür sind die Ergebnisse qualitativ nicht optimal. Durch die meist fehlenden Texturen an den Wänden wird eine akkurate Abschätzung der Wände erschwert. Zusätzlich spielt die Skalierung der gesamten Punktwolke eine große Rolle. Mit dieser Arbeit wurde lediglich ein einzelner, simpler Raum aus 6 Ebenen untersucht. Es konnten jedoch in 80 % aller Testräume die Dimensionen mit einer Längenabweichung von weniger als 3 cm bestimmt werden. Für größere Projekte aus mehreren, zusammenhängenden Räumen mit nicht-rechteckigem Grundriss ist der derzeitige Algorithmus nicht weit genug entwickelt. In zukünftigen Arbeiten könnte dieser Fall untersucht werden.

6 Literaturverzeichnis

Literatur

- Alcantarilla, P. F., Bartoli, A. & Davison, A. (2012). KAZE Features. *European Conference on Computer Vision (ECCV)*, 7577. https://doi.org/10.1007/978-3-642-33783-3_16
- Alkan, R. & Karsidag, G. (2012). Analysis of the Accuracy of Terrestrial Laser Scanning Measurements [TS07A - Laser Scanners I]. *Proceedings of FIG Working Week, 6097*.
- Arayici, Y., Hamilton, A. & Gamito, P. (2006). Modelling 3D Scanned Data to Visualise and Analyse the Built Environment for Regeneration. *Surveying and Built Environment*, 17(2), 7–28.
- Calvet, L., Gurdjos, P., Griwodz, C. & Gasparini, S. (2016). Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 562–570. <https://doi.org/10.1109/CVPR.2016.67>
- Chai, T. & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Ding, Y., Zheng, X., Zhou, Y., Xiong, H. & Gong, J. (2019). Low-Cost and Efficient Indoor 3D Reconstruction through Annotated Hierarchical Structure-from-Motion. *Remote Sensing*, 11(1). <https://doi.org/10.3390/rs11010058>
- Dong, J., Xiao, Y., Ou, Z. & Ylä-Jääski, A. (2015). Utilizing Internet Photos for Indoor Mapping and Localization -Opportunities and Challenges. <https://doi.org/10.1109/INFCOMW.2015.7179457>
- Ernst, F., Wilinski, P. & Overveld, C. V. Dense Structure-from-Motion: An Approach Based on Segment Matching. In: *European Conference on Computer Vision (ECCV)*. 2351. 2002, April. https://doi.org/10.1007/3-540-47967-8_15.
- Fan, H. & Meng, L. (2009). Automatic Derivation of Different Levels of Detail for 3D Buildings Modeled by CityGML. *Proceedings of 24th International Cartography Conference (ICC2009)*.
- Fischler, M. A. & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the Association for Computing Machinery (ACM)*, 24(6), 381–395. <https://doi.org/10.1145/358669.358692>
- Froideval, L., Pedoja, K., Garestier, F., Moulon, P., Conessa, C., Pellerin Le Bas, X., Traore, K. & Benoit, L. (2019). A Low-cost Open-source Workflow to Gene-

- rate Georeferenced 3D SfM Photogrammetric Models of Rocky Outcrops. *The Photogrammetric Record*, 34(2), 365–384. <https://doi.org/10.1111/phor.12297>
- Gasparini, S. (2020). Markers usage [Accessed: July 12, 2021]. <https://cctag.readthedocs.io/en/latest/markers/markers.html>
- Gauss, C. F. & Davis, C. H. (1857). *Theory of the motion of the heavenly bodies moving about the sun in conic sections. A translation of Gauss's „Theoria motus“, with an appendix.* Boston, Little, Brown; Company.
- Hays, J., Brahmabhatt, S. & Lambert, J. (2018). Project 2: Local Feature Matching [Accessed: September 1, 2021]. <https://www.cc.gatech.edu/~hays/compvision2018/proj2>
- Heimonen, T., Hannuksela, J., Heikkilä, J., Leinonen, J. & Manninen, M. (2001). Experiments in 3D measurements by using single camera and accurate motion. *Proceedings of the 2001 IEEE International Symposium on Assembly and Task Planning (ISATP2001). Assembly and Disassembly in the Twenty-first Century.* (Cat. No.01TH8560), 356–361. <https://doi.org/10.1109/ISATP.2001.929051>
- Hirschmüller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328–341. <https://doi.org/10.1109/TPAMI.2007.1166>
- Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P., Bunke, H., Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A. & Fisher, R. (1996). An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7), 673–689. <https://doi.org/10.1109/34.506791>
- Islam, S. (1989). Suhail a Islam - Math Notes [Accessed: February 03, 2021]. <http://www.sbg.bio.ic.ac.uk/~islam/plane.html>
- Jiang, S., Jiang, C. & Jiang, W. (2020). Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 230–251. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.04.016>
- Kaspi, O., Yosipof, A. & Senderowitz, H. (2017). RANdom SAmple Consensus (RAN-SAC) algorithm for material-informatics: application to photovoltaic solar cells. *Journal of Cheminformatics*, 9(1), 34–48. <https://doi.org/10.1186/s13321-017-0224-0>
- Khaloo, A. & Lattanzi, D. (2016). Hierarchical Dense Structure-from-Motion Reconstructions for Infrastructure Condition Assessment. *Journal of Computing in Civil Engineering*, 31, 04016047-1 - 04016047-13. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000616](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000616)

- Liu, C., Wu, J. & Furukawa, Y. (2018). FloorNet: A Unified Framework for Floorplan Reconstruction from 3D Scans. *European Conference on Computer Vision (ECCV)*, 6, 201–219.
- Longuet-Higgins, H. (1987). A computer algorithm for reconstructing a scene from two projections. In M. A. Fischler & O. Firschein (Hrsg.), *Readings in Computer Vision* (S. 61–62). Morgan Kaufmann. <https://doi.org/https://doi.org/10.1016/B978-0-08-051581-6.50012-X>
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Mataija, M., Pogarčić, M. & Pogarcic, I. (2014). Helmert Transformation of Reference Coordinating Systems for Geodesic Purposes in Local Frames. *Procedia Engineering*, 69, 168–176. <https://doi.org/10.1016/j.proeng.2014.02.217>
- Navratil, G. & Staudinger, M. (2006). *Ausgleichsrechnung I - Die Anwendung statistischer Methode in Vermessungswesen und GIS*. <http://matrix.ur.krakow.pl/~aborowiecki/skrypty/Navratil/skrypt/Ausgleich1.pdf>
- Navratil, G. & Staudinger, M. (2019). *Ausgleichsrechnung Vertiefung*. <http://matrix.ur.krakow.pl/~aborowiecki/skrypty/Navratil/skrypt/Ausgleich2.pdf>
- Nguyen, H., Wünsche, B., Delmas, P. & Lutteroth, C. (2012). 3D Models from the Black Box: Investigating the Current State of Image-Based Modeling. *Proceedings of the 20th International Conference on Computer Graphics, Visualisation and Computer Vision (WSCG), 2012*, 249–258.
- Oats, R. C., Escobar-Wolf, R. & Oommen, T. (2017). A Novel Application of Photogrammetry for Retaining Wall Assessment. *Infrastructures*, 2. <https://doi.org/10.3390/infrastructures2030010>
- Özyeşil, O., Voroninski, V., Basri, R. & Singer, A. (2017). A Survey on Structure from Motion. *Acta Numerica*, 26, 305–364. <https://doi.org/10.1017/S096249291700006X>
- Peleg, S., Keren, D. & Schweitzer, L. (1987). Improving image resolution using subpixel motion. *Pattern Recognition Letters*, 5(3), 223–226. [https://doi.org/10.1016/0167-8655\(87\)90067-5](https://doi.org/10.1016/0167-8655(87)90067-5)
- Peterson, E., Klein, M. & Stewart, R. (2015). *Whitepaper on Structure from Motion (SfM) Photogrammetry: Constructing Three Dimensional Models from Photography* (Techn. Ber. ST-2015-3835-1). U.S. Bureau of Reclamation.
- Pfeifer, N. (2002). *3D Terrain Models on the Basis of a Triangulation* (Diss.). TU Wien. https://publik.tuwien.ac.at/files/PubDat_119332.pdf

- PIX4D. (2017). Indoor mapping in 3D using photogrammetry software [Accessed: February 01, 2021]. <https://www.pix4d.com/blog/indoor-mapping-game-plan>
- Pollefeys, M., Koch, R., Vergauwen, M. & Van Gool, L. (1998). Metric 3D Surface Reconstruction from Uncalibrated Image Sequences. In R. Koch & L. Van Gool (Hrsg.), *3D Structure from Multiple Images of Large-Scale Environments* (S. 139–154). Springer Berlin Heidelberg.
- Rabbani, T., van den Heuvel, F. & Vosselman, G. (2006). Segmentation of point clouds using smoothness constraints. In H. Maas & D. Schneider (Hrsg.), *ISPRS 2006 : Proceedings of the ISPRS commission V symposium Vol. 35, part 6 : image engineering and vision metrology, Dresden, Germany 25-27 September 2006* (S. 248–253). International Society for Photogrammetry; Remote Sensing (ISPRS).
- Reisner-Kollmann, I. (2013). *Reconstruction of 3D Models from Images and Point Clouds with Shape Primitives* (Diss.). TU Wien - Institute of Computer Graphics und Algorithms. Favoritenstrasse 9-11/E193-02, A-1040 Vienna, Austria. https://www.cg.tuwien.ac.at/research/publications/2013/Reisner_Irene.2013_R3D/
- Ruzgiene, B. & Förstner, W. (2005). Ransac for outlier detection. *Geodesy and Cartography*, 31(3), 83–87. <https://doi.org/10.1080/13921541.2005.9636670>
- Stentoumis, C., Karkalou, E. & Karras, G. (2015). A review and evaluation of penalty functions for Semi-Global Matching. *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 167–172. <https://doi.org/10.1109/ICCP.2015.7312624>
- Triggs, B., Mclauchlan, P., Hartley, R. & Fitzgibbon, A. (2000). Bundle Adjustment - A Modern Synthesis. In B. Triggs, A. Zisserman & R. Szeliski (Hrsg.), *International Workshop on Vision Algorithms* (S. 298–372). Springer-Verlag. https://doi.org/10.1007/3-540-44480-7_21
- Westoby, M., Brasington, J., Glasser, N., Hambrey, M. & Reynolds, J. (2012). ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300–314. <https://doi.org/https://doi.org/10.1016/j.geomorph.2012.08.021>
- Wijenayake, U., Choi, S.-I. & Park, S.-Y. (2014). Automatic detection and decoding of photogrammetric coded targets. *2014 International Conference on Electronics, Information and Communications (ICEIC)*, 1–2.
- Wu, J., Cui, Z., Sheng, V., Zhao, P., Su, D. & Gong, S. (2013). A Comparative Study of SIFT and its Variants. *Measurement Science Review*, 13, 122–131. <https://doi.org/10.2478/msr-2013-0021>

- Yang, F., Zhou, G., Su, F., Zuo, X., Tang, L., Liang, Y., Zhu, H. & Li, L. (2019). Automatic Indoor Reconstruction from Point Clouds in Multi-room Environments with Curved Walls. *Sensors*, 19(17). <https://doi.org/10.3390/s19173798>
- Zhang, H., Aldana Jague, E., Clapuyt, F., Wilken, F., Vanacker, V. & Oost, K. (2019). Evaluating the potential of post-processing kinematic (PPK) georeferencing for UAV-based structure- from-motion (SfM) photogrammetry and surface change detection. *Earth Surface Dynamics*, 7, 807–827. <https://doi.org/10.5194/esurf-7-807-2019>

7 Anhang

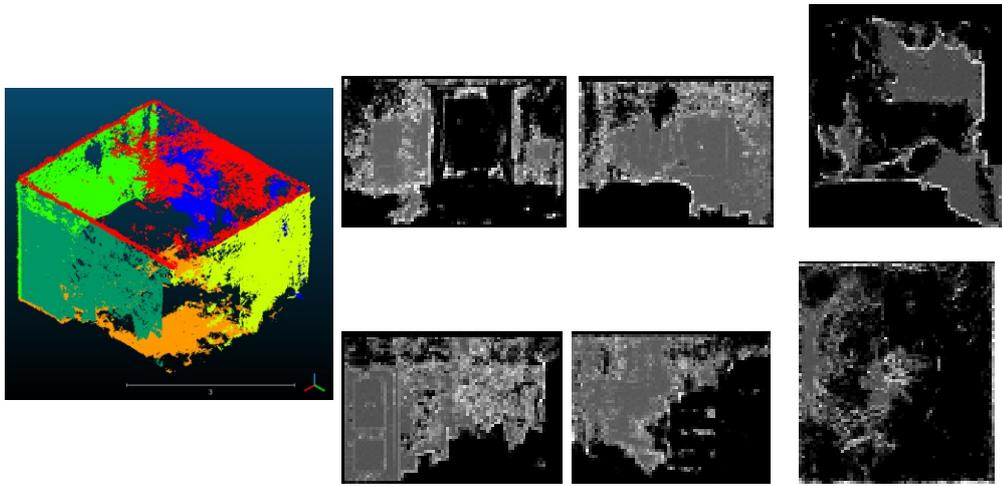


Abbildung 23: Raum 1: Klassifizierung und Abdeckungskarten

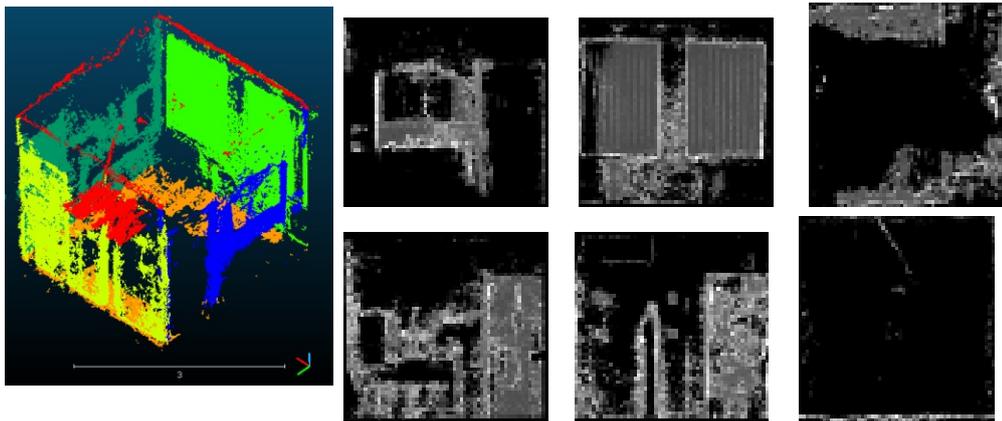


Abbildung 24: Raum 2: Klassifizierung und Abdeckungskarten

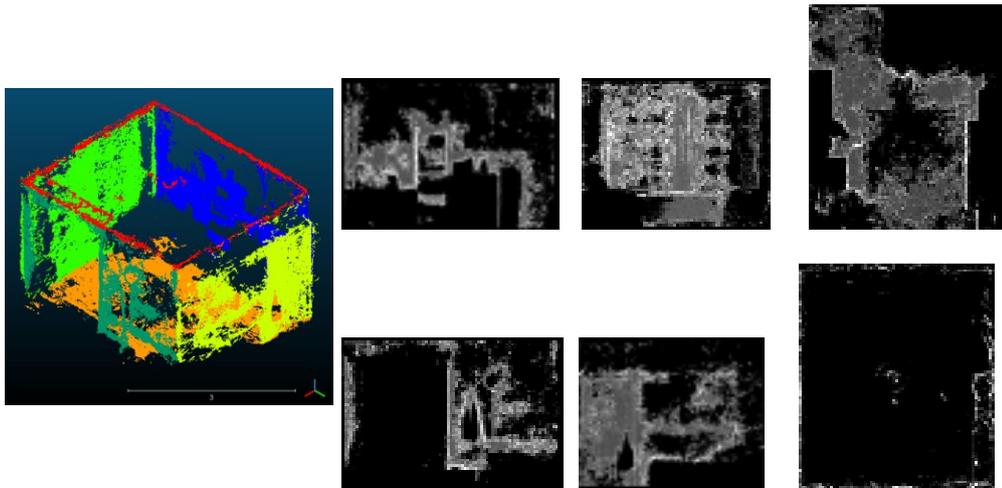


Abbildung 25: Raum 3: Klassifizierung und Abdeckungskarten

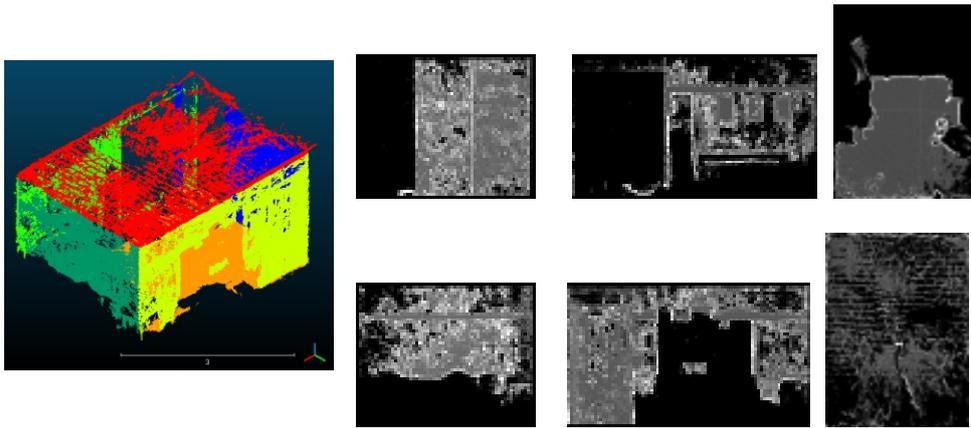


Abbildung 26: Raum 4: Klassifizierung und Abdeckungskarten

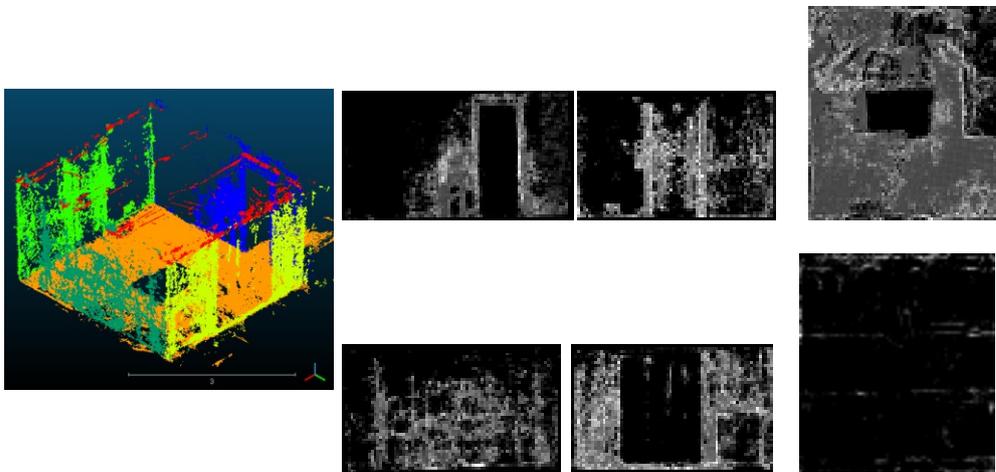


Abbildung 27: Raum 5: Klassifizierung und Abdeckungskarten

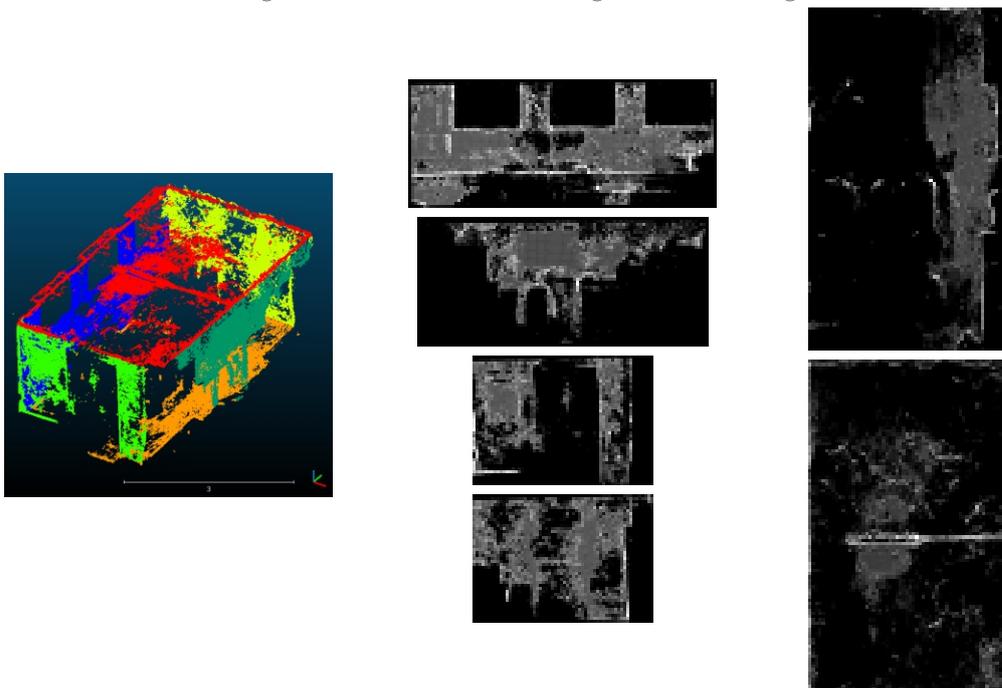


Abbildung 28: Raum 6: Klassifizierung und Abdeckungskarten

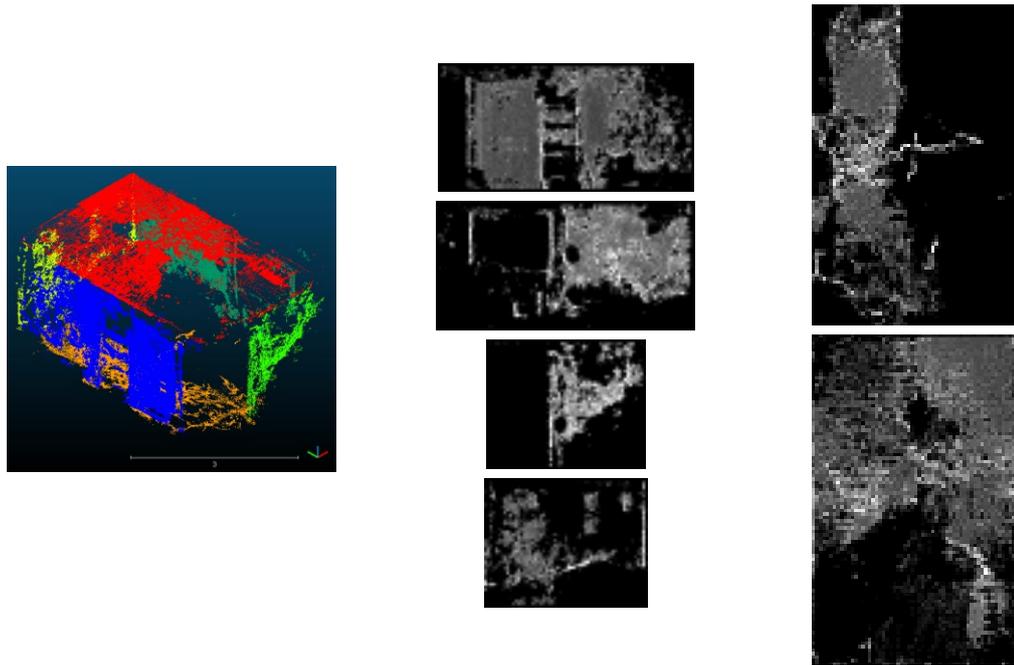


Abbildung 29: Raum 7: Klassifizierung und Abdeckungskarten

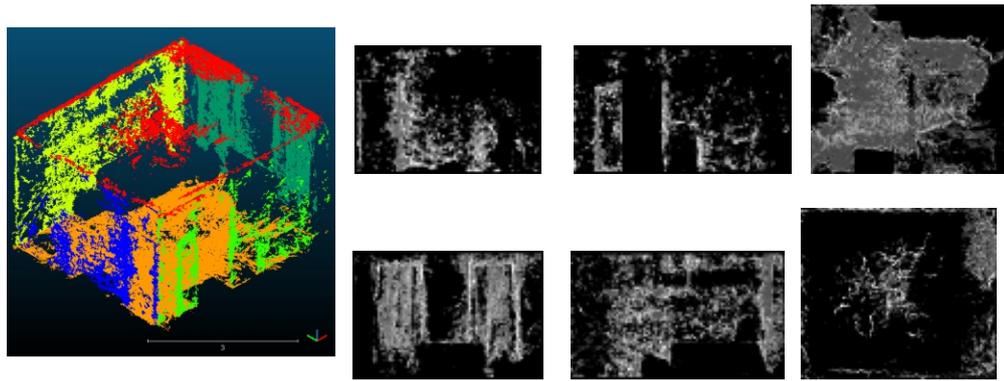


Abbildung 30: Raum 8: Klassifizierung und Abdeckungskarten

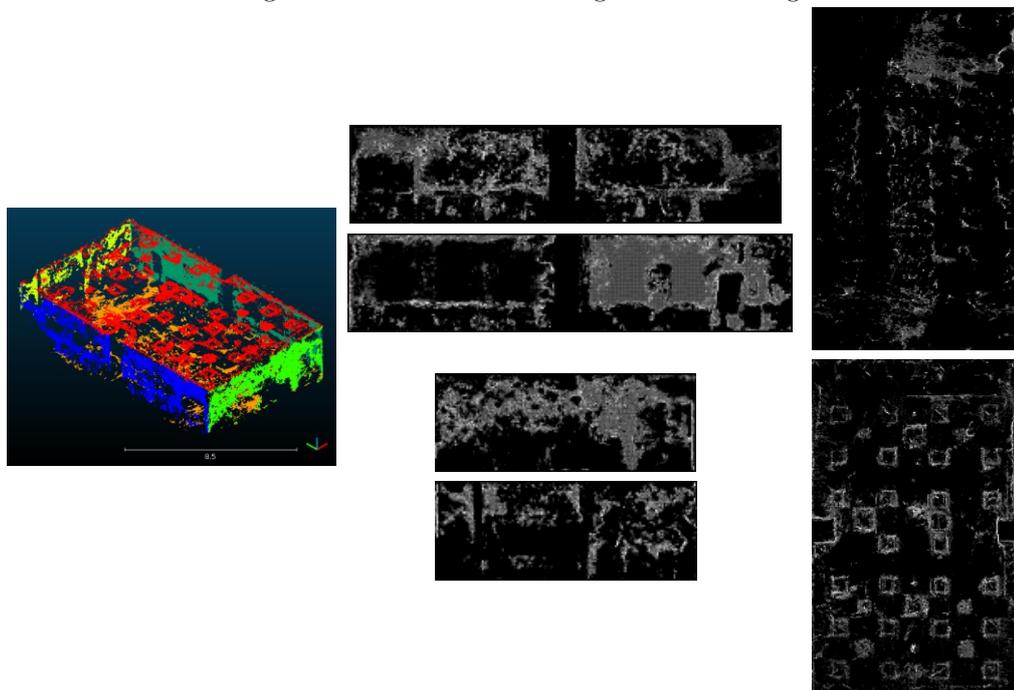


Abbildung 31: Raum 9: Klassifizierung und Abdeckungskarten

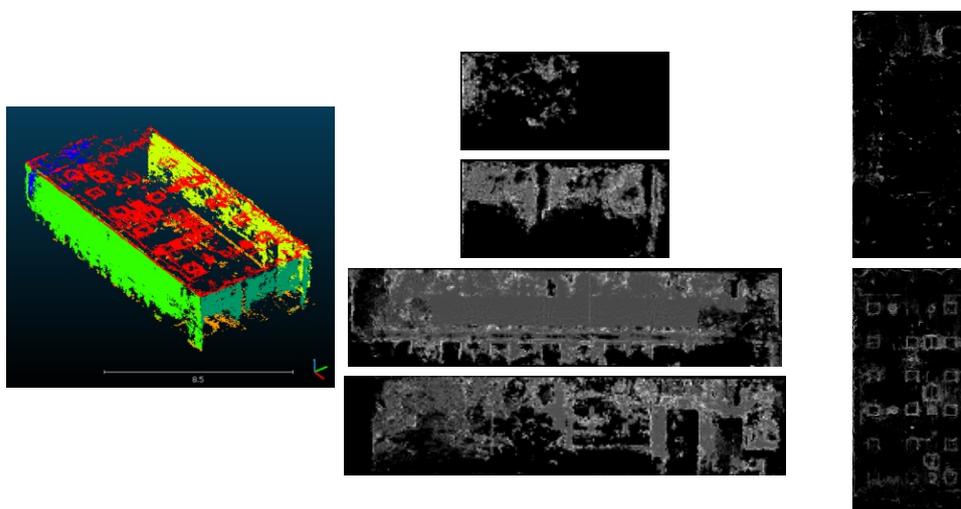


Abbildung 32: Raum 10: Klassifizierung und Abdeckungskarten