BACHELORARBEIT

Vergleich von Neural Radiance Fields und Photogrammetrie für 3D Asset-Creation

vorgelegt am 10. April 2024

Leo Kruse [redacted]

Erstprüfer: Prof. Dr. Eike Langbehn
Zweitprüfer: Simon Dewert

HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG
Department Medientechnik
Finkenau 35
22081 Hamburg

Zusammenfassung

Mit Photogrammetrie lassen sich aus Fotos oder Videos automatisch detailgetreue und texturierte 3D-Meshes generieren. Ob 3D-Meshes von Neural Radiance Fields ebenfalls qualitativ gleich oder bessere Meshes produzieren, wird in dieser Fallstudie überprüft. Für eine objektive Beurteilung werden beide Verfahren mit einem selbstangefertigten Luft-/Nahbild Datensatz durchgeführt und miteinander verglichen. Die Polygon-Meshes der Methoden werden mit spezieller Software analysiert und auf ihre Oberflächenstruktur und Textur geprüft. Zusätzlich wird ein Einblick in die Geschichte und aktuelle Lage der Photogrammetrie gegeben und eine Einführung in Neural Radiance Fields. Die Auswertung der beiden Rekonstruktionsverfahren zeigt, dass die subjektiven Beurteilungen der Meshes mit den objektiveren Messergebnissen grundsätzlich vergleichbar sind. Allerdings bleiben die 3D-Meshes der Photogrammetrie Software qualitativ und quantitativ hochwertiger und unkomplizierter in ihrer Erzeugung.

Abstract

Photogrammetry can be used to automatically generate detailed and textured 3D meshes from photos or videos. Whether 3D meshes from Neural Radiance Fields also produce meshes of the same or better quality is examined in this case study. For an objective assessment, both methods are carried out with a self-created aerial/close-up image data set and compared with each other. The polygon meshes of the methods are analyzed with special software and checked for their surface structure and texture. In addition, an insight into the history and current state of photogrammetry and an introduction to Neural Radiance Fields will be given. The evaluation of the two reconstruction methods shows that the subjective assessments of the meshes are basically comparable with the more objective measurement results. However, the 3D meshes of the photogrammetry software remain qualitatively and quantitatively of higher quality and less complicated to generate.

Danksagung

Die vorliegende Bachelorarbeit entstand im Rahmen des Studiengangs Medientechnik der HAW-Hamburg. Die Arbeit wird dem Bereich der Computergrafik zugeordnet und beinhaltet die praxisbezogene Untersuchung der zwei gängigen 3D-Rekonstruktionsverfahren Photogrammetrie und NeRFs im Bezug auf die Qualität der entstandenen 3D-Polygonnetze.

Zunächst möchte ich mich für meinen bisherigen Bildungsweg an der Hochschule bedanken. Durch die Vorlesungen meines Studiums - unter anderem bei Prof. Dr. Eike Langbehn -

war ich gut auf die inhaltliche Materie vorbereitet und wurde von ihm über die gesamte Bearbeitungszeit als Erstprüfer betreut. Durch Gespräche auf Augenhöhe und wertvolle Inspirationen und hilfreichen Denkansätzen konnte ein schlüssiges Konzept entstehen.

Des Weiteren möchte ich mich bei Simon Dewert für seine Unterstützung bei der Umsetzung der Drohnen-Aufnahmen für den Luftbild Datensatz bedanken.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

1.1 Ziel der Arbeit

Photogrammetrie und Neural Radiance Fields bieten neue Möglichkeiten, reale Objekte in 3D zu rekonstruieren. Die angewandten mathematischen Konzepte der Photogrammetrie werden seit über einem Jahrhundert erforscht und stetig weiterentwickelt. Photogrammetrie hat sich in den letzten Jahren in diversen Anwendungsbereichen etabliert. Darunter auch in der VFX- und Gamesbranche, in der es genutzt wird, um möglichst detailgetreue reale Objekte mit Bildern und Videos einzuscannen und sie als 3D-Assets zu verwenden. Neural Radiance Fields, kurz NeRFs, kommen aus der Computergrafik und sind Modelle, die mithilfe eines Multi-Layer Neuronalen Netzwerks, welches die Geometrie und Beleuchtung einer 3D-Szene lernt, neuartige Blickrichtungen auf ein Objekt generieren. Die daraus resultierende Darstellung ist volumetrisch.

Diese Arbeit zeigt empirisch auf, in welchen Punkten sich die beiden Methoden unterscheiden und welche Merkmale in den Ergebnissen zu erkennen sind, wobei ein Fokus auf die Modellierungsqualität und die der Texturierung der 3D-Meshes gelegt wird. Daneben werden die beiden Methoden hinsichtlich ihrer Anwendbarkeit auf unterschiedlich ausgeprägte Datensätze, der Benutzerfreundlichkeit der benutzten Programme und den Zeitaufwand geprüft.

1.2 Motivation für den Vergleich

In den letzten 20 Jahren war die 3D-Computergrafik ein rasant wachsendes und ständig weiterentwickeltes Gebiet. Die großen technischen Herausforderungen wurden gelöst. Die Erzeugung von 3D-Modellen ist in vielen Bereichen sehr gefragt, unter anderem im Bereich der visuellen Effekte, animierten Filmen, Architektur und Games. Die Modelle werden üblicherweise durch den Prozess der 3D-Modellierung erzeugt. Der Prozess der 3D-Modellierung lässt sich aufgrund der Mischung aus hohem technischen Können und künstlerischem Wissen sehr gut mit dem Handwerk vergleichen. Im Laufe der Zeit wurden die benutzten Werkzeuge verbessert und es hat sich eine Industrie entwickelt, die die benötigten 3D-Modelle erstellen

konnte. Angesichts des rasanten Wandels war ein disruptiver technologischer Wandel in diesem Bereich unausweichlich. Dieser Wandel war die Integration der Photogrammetrie für die Asset-Erstellung von 3D-Modellen. Aus Bildern und Videos kann nun mithilfe eines Algorithmus ein 3D-Modell erstellt werden. Die Asset Darstellung in Form eines 3D-Polygonmesh (im weiteren Verlauf 3D-Mesh genannt) ist immer noch die weitverbreitetste und auch am häufigsten unterstützte Darstellungsform. 3D-Meshes unterstützen Echtzeit Renderings und physikbasierte Simulationen in den gängigen VFX-Programmen oder Game Engines. Aus diesem Grund liegt der besondere Fokus dieser Arbeit auf den Ergebnissen der Methoden als 3D-Mesh.

Photogrammetrie ist heute eine weitestgehend erforschte Technik, die bereits seit ein paar Jahren in der Gamesbranche etabliert ist. Asset Creation mit Photogrammetrie erlaubt es den Spieleentwicklern, größere, lebendigere Welten zu realisieren und den Workflow noch effizienter zu gestalten. Von Hand modellierte und texturierte Assets können meistens nicht dem Fotorealismus gerecht werden. Der Mensch ist besonders gut darin, zu erkennen, ob Objekte nach der Wirklichkeit modelliert oder echte reale Abbilder sind, die durch ihre Willkür und Detailgenauigkeit schwer nachzubilden sind. Objekte, die durch natürliche Prozesse, wie beispielsweise Oxidation oder den Überzug von Pflanzen, über eine detailreiche Textur verfügen, nachzubilden, ist im besonderen Maße zeitaufwändig. An dieser Stelle setzt die Photogrammetrie an. So sind nicht nur fotorealistische Assets ein Vorteil, sondern auch die potenzielle Zeitersparnis im Workflow.

Neural Radiance Fields wurden entwickelt, um neuartige Blickrichtungen zu generieren, die nicht in dem Ursprungsdatensatz zu sehen sind (NVS = Novel View Synthesis). Die neuartige Technik der NeRFs ist recht simpel. Zugehörige Arbeiten, die im Laufe der letzten beiden Jahre auf dem Grundgerüst der NeRFs aufbauen, konzentrierten sich insbesondere auf die Verbesserung der Bildqualität, Robustheit, der Trainings- und Rendergeschwindigkeit. NeRFs werden zuerst volumetrisch dargestellt und sind nicht explizit darauf optimiert, die zugrundeliegende Geometrie als ein präzises 3D-Mesh zu exportieren. Mithilfe von Approximationen wird die Geometrie nur geschätzt und dann zu einem Polygonnetz verbunden. Dieser Teil bleibt also die offene Herausforderung für NeRFs.

Ziel des Vergleichs ist es aufzuzeigen, ob NeRFs Photogrammetrie für die Asset Creation ablösen können, welche Vor- und Nachteile sie mit sich bringen, und in welchen Situationen welche Technik am sinnvollsten sein kann.

2 Theoretische Grundlagen

2.1 Fundamentale Methode der Photogrammetrie

Im Allgemeinen liefern photogrammetrische Systeme dreidimensionale Objektkoordinaten, die aus Bildmessungen abgeleitet werden. Dabei werden nicht nur Daten wie die Farbe einer Hauswand oder die Textur aufgenommen, sondern auch quantitative Daten wie Linien, Abstände, Flächen oder Distanzen berechnet. Die Photogrammetrie kann demnach als „*Wissenschaft der Vermessung in Fotos*“ definiert werden. Photogrammetrie gehört zum Feld der Fernerkundung („*Remote Sensing*“) und kommt traditionell aus der Geodäsie (Acquise von Geoinformationen) (Linder, 2016).

Aktuell ist die Kernaufgabe von Photogrammetrie, eine exakte geometrische 3D-Rekonstruktion aus 2D Bildern zu erschaffen.

Die fundamentale Methode, die sich Photogrammetrie zunutze macht, ist das stereoskopische Sehen, genauso wie bei unserem Sehapparat. Dieser verarbeitet die optische Information beider Augen, die sich in ihrer Position nur leicht unterscheiden, zu der Fähigkeit der räumlichen Wahrnehmung (Linder, 2016).

Bei Photogrammetrie erfolgt dies mit der geometrischen Methode der Triangulation, wie in Abbildung 2.1 dargestellt. Die Berechnung erfolgt mittels trigonometrischer Funktionen. Von den zwei verschiedenen Bildpunkten P’ und P” wird die zu bestimmende Koordinate von Objektpunkt P angepeilt. Die Strahlen der beiden Bilder überschneiden sich genau bei Objektpunkt P. Unter Kenntnis des Basisabstandes, des Winkels der Bilder relativ zum Objektpunkt P (extrinsische Parameter) und der Brennweite der Kamera (intrinsische Parameter) kann die Koordinate von Objektpunkt P relativ zum Koordinatenursprung bestimmt werden.

Bei der Stereo-Photogrammetrie werden zwei Bilder benötigt. Bei der Mehrbild-Photogrammetrie ist die Anzahl der Bilder im Prinzip unbegrenzt.

Die heutige digitale Photogrammetrie verarbeitet ein Kompendium an Bildern. Dieses mathematische Prinzip wird bei mehreren zu verarbeitenden Bildern auf eine Vielzahl von Objektpunkten P angewendet. Die Form und Position des Objekts werden durch die Rekonstruktion der Lichtstrahlen (Rays) ermittelt, indem auf jedem Bild ein Objektpunkt P’ identifiziert wird, mithilfe der Merkmal-Korrespondenz (Feature Matching). Dadurch, dass

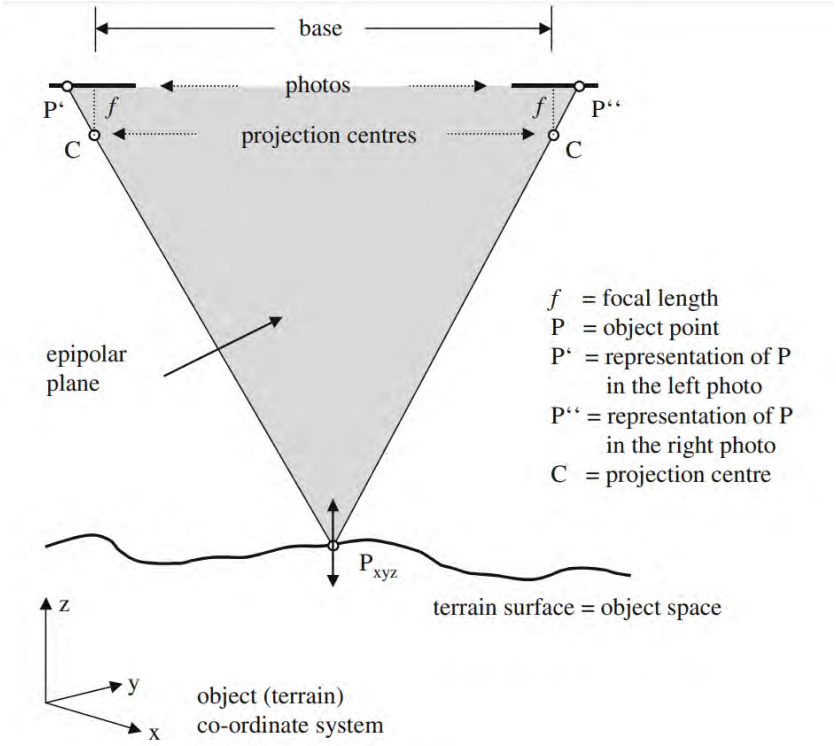


Abbildung 2.1: Stereoskopische Photogrammetrie (Linder, 2016)

die Blickwinkel der Kamera verschieden sind, ist die Position des Objektpunktes P' auf den Bildern ebenfalls unterschiedlich. Unter Kenntnis des exakten Blickwinkels, der intrinsischen Parameter der Kamera und der Koordinaten der Kamera im Raum ist es möglich jeden Lichtstrahl im Raum zu rekonstruieren. In Abbildung 2.2 wird dies visuell dargestellt, (Luhmann, 2023).

Die daraus berechneten Koordinaten der Objektpunkte werden digital zu einem Gesamtmodell vernetzt. Derzeit übernehmen Computer diese intensive Rechenarbeit (Linder, 2016). Die Koordinaten können auch in Form von Pointclouds und in grafischer Form wie Bildern und Plänen dargestellt werden (Luhmann, 2023).

2.2 Geschichte der Photogrammetrie

Die Entwicklung der heutigen Photogrammetrie ist lang und fortschrittsreich, so basieren die Grundbestandteile aller Fortschritte auf der Trigonometrie und der geometrischen Optik. In der geometrischen Optik werden der Verlauf, die Reflexion und Brechung von Lichtstrahlen mithilfe elementargeometrischer Prinzipien („*Fermatsches Prinzip*“) ermittelt.

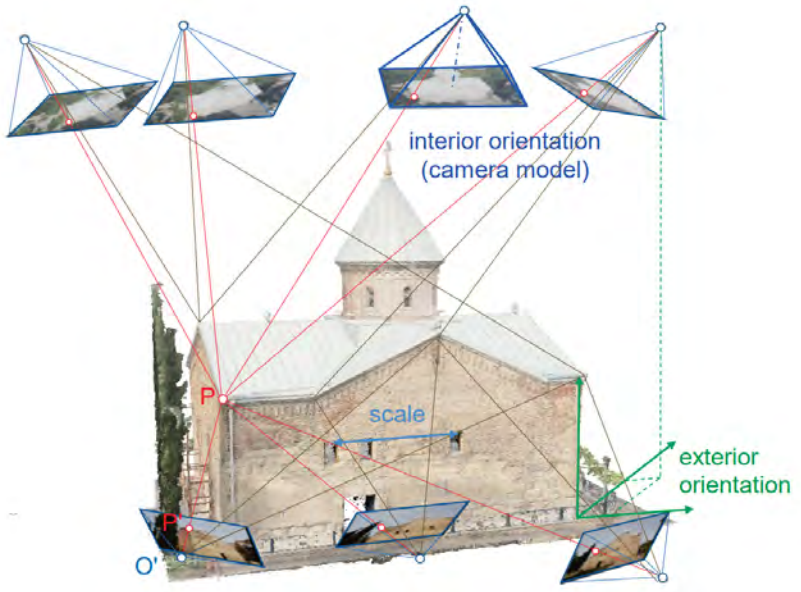
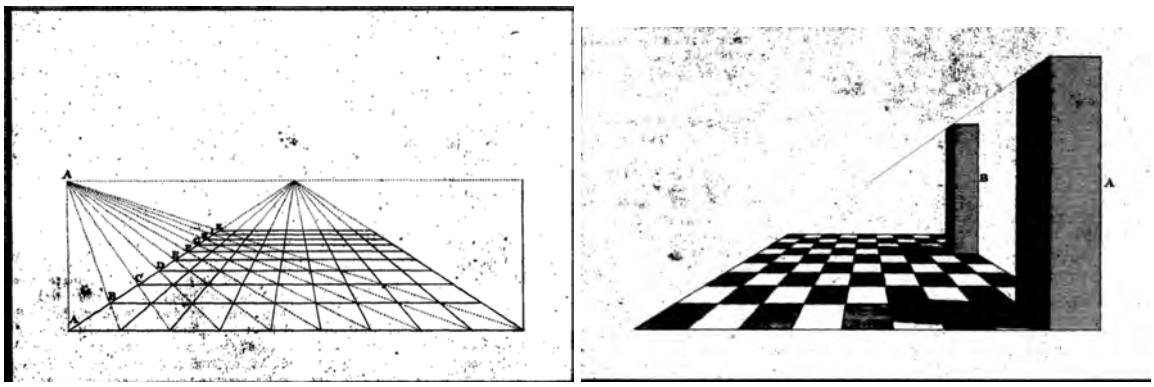


Abbildung 2.2: Prinzip der heutigen Photogrammetrie visualisiert (Luhmann, 2023)

Die Geschichte der Photogrammetrie kann bis ins fünfzehnte Jahrhundert zurückverfolgt werden. Die Vorreiter des perspektivischen Zeichnens waren die Künstler der italienischen Renaissance. Es wurden die Regeln der Euklidischen Geometrie („*Parallel sind gerade Linien, die in derselben Ebene liegen und dabei, wenn man sie nach beiden Seiten ins Unendliche verlängert, auf keiner Seite einander treffen.*“) gebrochen, um lineare Perspektive zu erschaffen (Gentili et al., 2022). Parallele Linien liefen nun zusammen, um einen Horizontpunkt zu erschaffen, der Tiefe vermittelt. Aus heutiger Sicht könnte man sagen, dass Leon Battista Alberti, italienischer Künstler und Architekt, einen wissenschaftlichen und praktischen Algorithmus schuf, den jeder Maler nutzen kann, um sein Werk unter dem Gesichtspunkt der Perspektive korrekt malen zu können (Abbildung 2.3).

Das Ziel der Photogrammetrie ist jedoch das komplette Gegenteil. Statt Geometrie zu nutzen, um Malereien Perspektive zu verleihen, wird hier Mathematik genutzt, um quantitative Informationen aus Bildern zu extrahieren.

Als Schöpfer der Photogrammetrie wird Erich Meydenbauer gehalten. Er wurde 1858 damit beauftragt, den Wetzlarer Dom zu vermessen. Er nutzte Fotografien des Doms als alternative Methode, um die Fassade des Doms nicht manuell messen zu müssen. Er entwickelte eine eigene Kamera, um 1300 Bilder von bekannten preußischen Monumenten aufzunehmen. Durch das direkte Vermessen des Orientierungswinkels der aufgenommenen Bilder war er in der Lage, die Fassade im Verhältnis 1:5000 quantitativ zu vermessen. Abbildung 2.5 zeigt das gleiche Prinzip angewendet an der Fassade der Marburger Elisabethkirche (Luhmann, 2023).



(a) Abbildung aus ‘Book I’ (b) Abbildung aus ‘Book II’

Abbildung 2.3: Leon Battista Alberti, *Of Painting in three books*, ‘Book II’ und ‘Book I’, zitiert nach (Gentili et al., 2022)

Abbildung 2.4: Bild der Marburger Elisabethkirche (Luhmann, 2023)

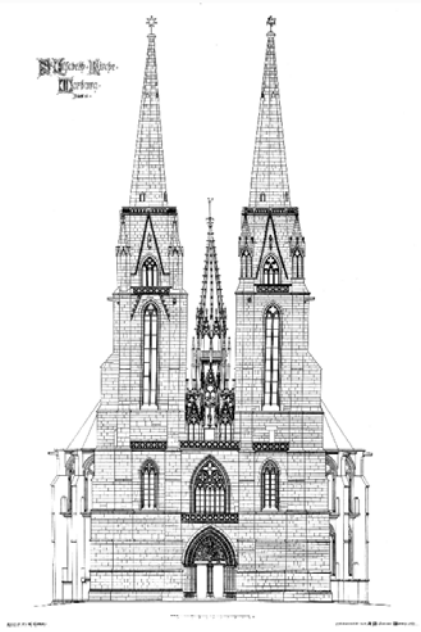


Abbildung 2.5: Fassadenzeichnung der Marburger Elisabethkirche (Luhmann, 2023)

Zudem waren Meydenbauer und Kersten, ein Geograph, die ersten, die den Begriff „*Photogrammetrie*“ in ihrer Veröffentlichung verwendeten. Der französische Offizier Aimé Laussedat entwickelte zur selben Zeit eine Methode namens „Métrophotographie“, um Terrain topographisch aufzunehmen. Meydenbauer schloss sich an und entwickelte weitere Methoden, um zu kartografieren. Dennoch wurden die Messungen manuell mit mathematischen Prinzipien und Regeln der geometrischen Optik auf Messtischen berechnet.

Mit der Erfindung des Stereoplotters 1909 von Orel und der Zusammenarbeit mit der Firma Carl Zeiss wurde die Phase der analogen Photogrammetrie eingeläutet.

Stereoplotter sind Geräte, die Objekte durch zwei minimal überlappende stereoskopische Bilder automatisiert messen können.

In den 1930er Jahren wurden diese Technologien weiterentwickelt und verfeinert. Zudem war rapide Entwicklung der Luftfahrt für die Effektivität von Photogrammetrie für die Topografie von Vorteil. Vertikale Bilder der Erde aus erhöhter Position sind ideal für die photogrammetrische Kartografie. Vor diesem Hintergrund etablierte sich die Luftbildphotogrammetrie zur Hauptmethode der Kartografie.

Die weitere Verfeinerung der analogen Geräte für Stereoplotting und Fotografie, zum Beispiel hochpräzise Servomotoren oder Drehimpulszähler, die die Koordinaten der Objektpunkte registrieren, ermöglichte eine noch größere Genauigkeit. Mit dem Aufkommen der Computer war die Idee, die Orientierung nicht mehr analog, sondern algorithmisch über Formeln, deren Parameter berechnet und im Computer gespeichert werden, zu konstruieren. Trotz dieser Meilensteine der Photogrammetrie wurde die Arbeit weiterhin mit echten analogen Fotos verrichtet. Deswegen wurde diese Phase der Photogrammetrie auch als analytische Photogrammetrie bezeichnet (Luhmann, 2023).

Die seit Mitte der 80er Jahre und bis dato andauernde Phase wird als digitale Photogrammetrie bezeichnet. Die unglaublichen Entwicklungen der Digitalfotografie ermöglichen feinere und schärfer aufgelöste Informationen in Bildern.

Beispiele hierfür sind „Pan-chromatic Imagery“ und „Near-infrared Imagery“. Ersteres kommt ohne RGB-Farbinformation aus, enthält aber gegenüber multispektraler Auflösung mehr räumliche Informationen auf einem individuellen Spektralband. Letzteres ist dazu in der Lage, Objekte jenseits des für das menschliche Auge sichtbaren Lichts darzustellen, um so Unebenheiten auf komplexen Oberflächen aufzunehmen (EOS, 2021) (Kumar, 2022). Vor diesem Hintergrund ist auch die exponentielle Entwicklung der Computerchips nicht zu vergessen. Moderne Computertechnologie ermöglicht die Verarbeitung digitaler Bilder, insbesondere zur automatischen Erkennung und Messung von Bildmerkmalen, einschließlich Musterkorrelation zur Bestimmung von Objektoberflächen. Viele der physischen Plotter wie der Stereoplotter wurden durch Licht- und Abstandsbestimmung (LiDAR: Light Detecting

and Ranging) ersetzt, bei der ein Laser anstelle von Bildern verwendet wird, um Tiefen- und Entfernungsinformationen zu erfassen (Luhmann, 2023).

Viele der Photogrammetrietechniken werden neben der Kartographie allmählich auch in anderen Forschungsbereichen eingesetzt und der breiten Bevölkerung für kreative, Bildungs- oder Hobbyzwecke zugänglich gemacht. Es werden beispielweise photogrammetrisch Avatare realer Künstler erschaffen, um in Musikvideos verwendet zu werden (Foster, 2014).

2.3 Heutige Anwendungsbereiche

Die Anwendungsbereiche von Photogrammetrie lassen sich verschieden kategorisieren. Überzeugend ist es vor dem Hintergrund dieser Arbeit der ersten und einfachsten Unterscheidung zu folgen. Hier werden Objekte mit der zugehörigen Kamera-Position zwischen Nahbereichsphotogrammetrie, also terrestrischer Position der Kamera, und Luftbildphotogrammetrie (die Kamera macht Luftaufnahmen) getrennt.

2.3.1 Luftbildphotogrammetrie

Die Luftbildphotogrammetrie ist weiterhin das gängige Vorgehen der modernen Geodäsie und Kartografie (Erstellung von Landkarten). Üblicherweise liegt der Objektabstand bei unbemannten fliegenden Kamerasystemen bei einem Abstand von über 120 m Höhe und bei bemannten Kamerasystemen bei über 300 m Höhe über dem Objekt. Darunter fallen auch Satelliten. Apple Karten und Google Maps sammeln seit nun einem Jahrzehnt hochaufgelöste Drohnenbilder in diversen Städten. Beide Softwareentwickler halten sich jedoch zurück, wie die 3D Rekonstruktionen ihrer respektiven Software funktioniert. Unter dem Namen „Flyover“ bietet Apple in ihrer Karten App den Benutzern an, bestimmte Gebiete in 3D darstellen zu können (Flyover, 2024).

2.3.2 Nahbereichsphotogrammetrie

Bei einem deutlich geringeren Objektabstand gibt es einen fließenden Übergang in die Nahbereichsphotogrammetrie. Folgende Anwendungsbereiche gehören zu den wichtigsten:

- Dokumentation und Rekonstruktion von Architektur und kulturellen Denkmälern
- Terrestrische und Unterwasserarchäologie
- Fertigungstechnik und Analyse von Crashtests in der Luft-, Raumfahrt- und Automobilindustrie

- Zahnmedizin, orthopädische Medizin und Biomechanik (Dentalmessungen, Messung von Deformation, plastische Chirurgie und Bewegungsanalyse)
- Ingenieursarbeit (Messung von großflächigen Baustellen, Verlegungsmessung bei Rohren oder Eisenbahnstrecken)
- Polizeiarbeit und forensische Analyse (Unfallrekonstruktion und z. B. Personenmessung anhand von Kameras)

(Luhmann, 2023)

2.3.3 Photogrammetrie im Zusammenhang mit der Unterhaltungsindustrie

Die großen Marktführer der Unterhaltungsindustrie haben das Potenzial der Photogrammetrie vor über zehn Jahren entdeckt und verfeinern es seitdem stetig weiter. Insbesondere Microsoft forschte daran, Computergrafik und Photogrammetrie für die Unterhaltungsindustrie zu nutzen. Eine der ersten großen Anwendungen für den Consumer-Bereich war die Microsoft Kinect, die 2010 als Zubehörgerät zur Xbox 360 verkauft wurde (Foster, 2014). Die Kinect besteht aus einer normalen Kamera, einem Mikrofon und einem Infrarotsensor. Durch die Analyse der visuellen Information und der Tiefeninformation der Infrarotkamera gibt es eine Vielzahl an Möglichkeiten, die Information zu interpretieren und ohne Gamepad mit der Konsole zu interagieren. Neben der Tiefenerkennung wurde auch die Gestenerkennung in die Software implementiert. Microsoft verbesserte in den Folgemodellen besonders die Hardware. Die „Kinect One“ wurde parallel zur neuen Konsolengeneration 2013 auf dem Markt veröffentlicht. Sie konnte nunmehr sechs Personen gleichzeitig tracken und die Kamera hatte eine vierfach gesteigerte Kameraauflösung. Die aufgenommenen 3D-Modelle, größtenteils sind das die Körper der Spieler vor der Kamera, wurden viel präziser und detaillierter wiedergegeben (Foster, 2014). Die Produktion der Kinect One wurde 2017 von Microsoft eingestellt, nachdem wenige Spieltitel nach der neuen Konsolengeneration die Technologie nutzen wollten (Kinect, 2024). Trotzdem erforschte Microsoft noch die Möglichkeiten der Kinect als akademisches und kommerzielles Gerät. Das Resultat davon war die Kinect Azure. Neben erhöhter Performance durch die Unterstützung von Künstlicher Intelligenz war die Azure im Vergleich zu anderen Tiefensensorgeräten preiswerter (400$) und robuster. Das dazu veröffentlichte Software Development Kit (SDK) ermöglicht es, die Sensorinformation auf jede mögliche Art zu benutzen. Im Oktober 2023 beendete Microsoft die Produktion der Kinect Azure (Kinect, 2024).

Besonders in der Filmindustrie werden heutzutage auch Laserscanner eingesetzt. Die bereits in 2.2 erwähnte LiDAR Methode wird verwendet, um Drehorte und Filmkulissen aufzunehmen. Die Scans werden in der Post-Production genutzt, um die Sets virtuell zu erweitern oder

um Kamerawinkel zu rekonstruieren. Der Vorteil von LiDAR in diesem Anwendungsbereich ist die extrem hohe Genauigkeit dieses Systems. Es können mit einem Scan eine Vielzahl von Punkten aufgenommen werden und die resultierende Pointcloud hat dann eine präzise Auflösung. Außerdem basiert LiDAR für die Rekonstruktion der Koordinaten nicht auf Feature Matching. Bei Texturen mit wenig markanten Eigenschaften versagt die Photogrammetrie oftmals. Dennoch bringt LiDAR auch Nachteile mit sich. Die Geräte sind nicht nur hochpreisig (und Maschine Deutschland GmbH, n. d.), sondern erfordern für den Scan geschultes Personal. LiDAR Scans sind sehr zeitaufwändig und können keine sich schnell bewegenden Objekte aufnehmen. Daher kann LiDAR bezüglich der Präzision als Alternative zur Photogrammetrie angesehen werden.

Dieser Text befasst sich im Folgenden speziell mit einem relativ neuen Anwendungsbereich, der zu der Liste hinzugefügt wird. Der Bereich der 3D-Rekonstruktion, insbesondere der Integration digitalbildbasierter Photogrammetrie in die 3D-Objekt Pipeline, die für Games und Animationen in der Filmindustrie genutzt wird. Besonders der Bereich der Computergrafik hat im Zusammenhang mit Photogrammetrie viele neue Innovationen entwickelt. Da der Oberbegriff „Photogrammetrie“ nicht besonders stark repräsentiert wird, wird der Prozess der 3D-Rekonstruktion in der Computergrafik auch „Multi-View Stereo“ genannt. Der Bereich der 3D-Computergrafik hat damit begonnen, Photogrammetriedaten und Algorithmen zur schnellen Erzeugung von 3D Modellen zu benutzen, um visuelle Abbilder auf der Grundlage des wirklichen Lebens zu erstellen. Diese generierten 3D-Modelle, auf Englisch „Assets“ genannt, werden dann für virtuelle Welten in der Virtual Reality, Augmented Reality oder der Mixed Reality benutzt.

Die heutigen Arbeitsschritte der Computergrafik für Photogrammetrie (Multi-View-Stereo) setzen sich aus vier Schritten zusammen:

1. Structure-from-Motion wird auf den Bilderdatensatz angewendet. Dieser Algorithmus sucht nach Korrespondenzen von Objektpunkten in den versetzt aufgenommenen Bildern. Daraus kann die genaue Kameraposition im Koordinatensystem ermittelt werden (Siehe Abschnitt 3.3).
2. Eine dichte Point-Cloud wird erstellt. Die Oberfläche der Punkte wird texturiert mit den Farbinformationen der Pixel aus den Bildern.
3. Die Punkte werden trianguliert vernetzt und es entsteht ein 3D-Mesh mit Oberflächen.
4. Die neuen Oberflächen des Meshes werden mit den Bildinformationen texturiert.

(Grechnyev, 2023)

Das kleine Indiestudio „The Astronauts“ brachte 2014 das Computerspiel „The Vanishing of Ethan Carter“ heraus. Die in dem Spiel enthaltenen Assets in Form von Gebäuden, Steinen, Texturen und anderen Objekten wurden mittels Photogrammetrie erstellt (Poznanski, 2014).

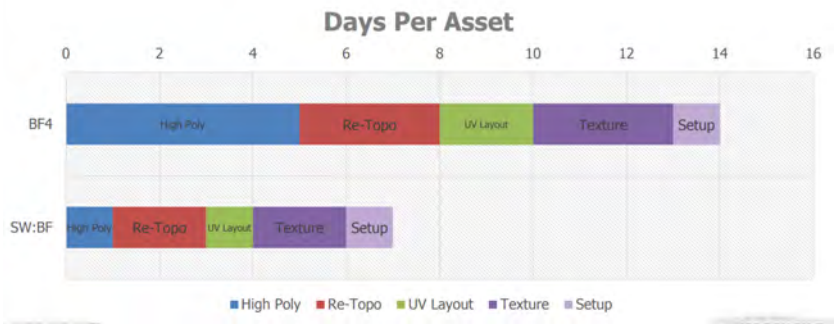


Tabelle 2.1: Geschätzte Zeit der Photoscanned Asset-Creation von SW Battlefront verglichen mit der klassischen Pipeline von Battlefront 4 (Hamilton, 2016)

Die Assets beeindruckten auch die großen Spielentwickler. Electronic Arts kündigte im darauf folgenden Jahr an, ihr neues Star-Wars Flagship Game mit Photogrammetry Assets umsetzen zu wollen (Hamilton, 2016). Bisher wurde Photogrammetrie nur zur Erstellung von Gesichtern eingesetzt, wie z. B. in der FIFA-Serie.

2015 wurde Photogrammetrie noch wenig in Zusammenhang mit Assets für Videospiele gebracht. Der damalige Lead Environment Artist Andrew Svanberg Hamilton war dafür zuständig, Photogrammetrie bei EA Dice, dem Studio für *Star Wars Battlefront*, voranzutreiben. Der Zeitpunkt, um Photogrammetrie eine Chance bei der Asset Erstellung zu geben, war ideal. Zunächst hatte das Art-Team ausreichend Zeit, um vorher alle Workflows bezüglich der Zeit und Qualität der Assets zu testen. Anschließend wurde das Spiel für die neuere und leistungsstärkere achte Konsolengeneration entwickelt. Trotzdem wurde das Art-Team besonders auf der Managementebene herausgefordert, inwiefern mit Photogrammetrie Zeit eingespart werden könnte (Hamilton, 2016).

Mit der „Photoscanned Pipeline“ konnte besonders bei der High-Poly 3D-Modellierung Zeit eingespart werden, wie in Tabelle 2.1 zu sehen ist. Die anderen Arbeitsschritte blieben zeitlich ungefähr gleich. Das Art-Team erwartete sogar, die Pipeline noch weiter zu verkürzen, indem sie bestimmte Schritte automatisieren.

Für die Aufnahmen machte sich jeweils ein kleines Team auf den Weg zu den Drehorten der Star-Wars Filme. Das Aufnahme-Kit bestand aus einer Canon 6D mit einem 24mm Objektiv und einem Stativ. Die Teams waren je für einen Drehort von Aufnahme der Bilder bis zum fertigen Asset zuständig. Konzentriert wurde sich auf bestimmte vorher ausgesuchte Objekte. Geplant war, einen möglichst großen Korpus an Assets zu erstellen, die dann als Baukasten für ein bestimmtes Level benutzt werden, ein sogenanntes Level Construction Kit wie in Abbildung 2.7 gezeigt. Die ausgewählten Objekte waren generisch, sodass sie für die Baukastennutzung auch einfach zu drehen oder zu verändern sind, um nicht repetitiv zu wirken (Hamilton, 2016).

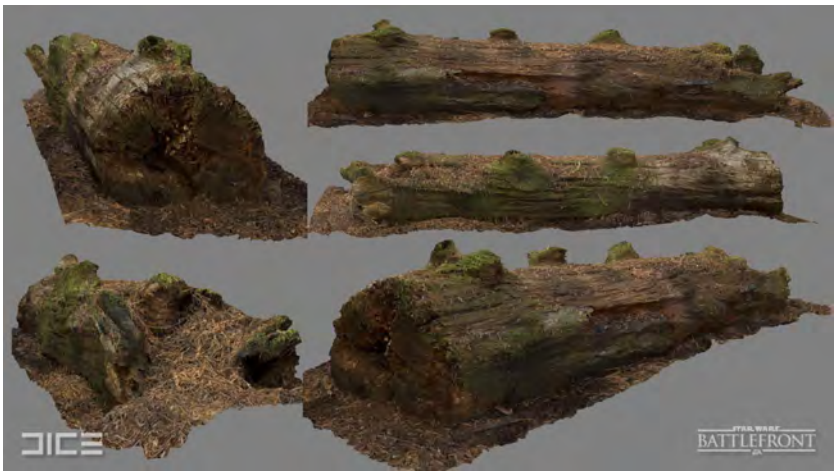


Abbildung 2.6: SW Battlefront Photogrammetrie Game-Asset Baum (Hamilton, 2016)

Abbildung 2.7: SW Battlefront Level Construction Kit (Hamilton, 2016)

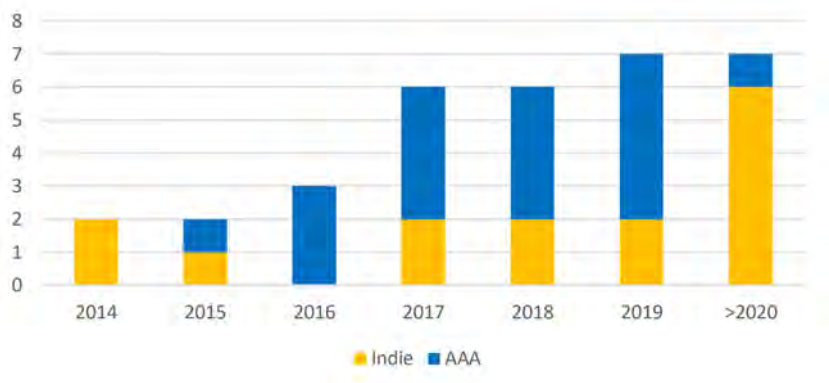


Abbildung 2.8: Veröffentlichte Spiele mit Photogrammetrie Assets 2014-2020 (Statham et al., 2020)

Neben generischen Objekten wurden auch Texturen aufgenommen. In verschiedenen Winkeln wurde ein Abschnitt von 3 × 3 Metern aufgenommen. Durch Photogrammetrie konnten nun auch detailreiche Displacement Maps erzeugt werden, die den Texturen im Spiel noch mehr Realität verleihen. Durch Displacement Maps können Texturen auf einem 3D-Mesh mit einer Height-Map verschoben werden. Dabei wird die Geometrie des 3D-Mesh tatsächlich verschoben, was zu mehr Detailtreue führt.

Außerdem wurde das umliegende Umgebungsterrain Mesh der Level aus echten Topographiedaten der Orte erstellt. Diese Daten wurden mit Luftbildphotogrammetrie aufgenommen. Um den Realismus zu unterstützen, hatten diese Meshes einen Radius von über 40 km.

Die wichtigsten Erkenntnisse des Teams lagen jedoch darin, dass Photogrammetrie keine „*Wunderwaffe*“ ist, sondern die eigentliche Handwerkskunst der 3D-Modellierung immer noch essenziell ist. Trotzdem war Photogrammetrie ein wichtiger Bestandteil der finalen Ergebnisse innerhalb von *Star Wars Battlefront* (Hamilton, 2016).

Mit diesem Titel begann die Integrierung von Photogrammetrie in allen großen Studios. Seit 2014 bis 2019 wurden 26 Titel veröffentlicht, in denen Photogrammetrie für die Asset Erstellung genutzt wurde (2.8). Während 26 Spiele im Zeitraum von fünf Jahren einen geringen Prozentsatz des Spielemarktes ausmachen, spricht das Kaliber dieser Spiele eine andere Sprache. Vor allem unter den AAA-Titeln, die auf Photogrammetrie zurückgreifen, finden sich bekannte und etablierte Franchises wie *Battlefield, Call of Duty, Elder Scrolls, Far Cry, Final Fantasy* und *Forza*. Die Entscheidung zu dieser vorher unbekannten Technologie zeugt von der Überzeugung aller großen Studios, Photogrammetrie in die Asset-Pipeline zu integrieren (Statham et al., 2020).

Die interne Verfeinerung der Photogrammetrieprozesse kann exemplarisch an der Entwicklung bei DICE erkannt werden („Crafting Environments for Battlefield V with DICE LA“, 2019). Photogrammetrie wurde 2019 auch für Battlefield V eingesetzt. Die Aufnahmeteams

wurden jeweils einem Ort zugeteilt, für den sie dann auch im Spiel zuständig waren. Die Reisen wurden im Vorhinein effizienter geplant. Dabei enthielt der Katalog an Assets natürlich 3D-Objekte wie Häuser oder kleinere Requisiten, aber auch Hintergrundlandschaften oder Pflanzen.

An einem Beispielhaus, welches in Griechenland eingescannt wurde, referieren die Entwickler, dass sie alles vorher systematisch geplant haben. Eine Rundaufnahme mit einer Spiegelreflexkamera, anschließend mit ein bisschen räumlichem Abstand eine Drohnenaufnahme. Außerdem wurde darauf geachtet, die nötigen Genehmigungen zu haben, um Drohnen fliegen zu dürfen („Crafting Environments for Battlefield V with DICE LA“, 2019).

Nach dem Scannen der Objekte werden diese meist runterskaliert, weil die Qualität der Meshes und der Texturen über den Qualitätsanforderungen liegt. Ein großer Anteil der dargestellten Assets im Spiel waren photogescannt. Dazu zählen Objekte, Häuser, Texturen, Hintergrundlandschaft und Pflanzen. Photogrammetrie diente den Entwicklern dazu, eine hochwertige und detaillierte Grundlage der Zielobjekte als Mesh zu haben, um weitere Anpassungen vornehmen zu können. Ein Ziel in Battlefield V war die Umsetzung von teilweiser Zerstörung von Gebäuden durch das Spielgeschehen, welches mit detaillierten Grundmeshes sehr leicht umsetzbar war. Die Aufnahme Workflows wurden somit verbessert und die Qualität erhöht. Trotzdem bleibt Photogrammetrie nur ein Teil der Kette, um hochwertige Game-Assets zu erstellen.

Bezüglich der rechtlichen Grundlage von eingescannten Objekten bleibt es undurchsichtig. Die Autoren der Studie „*Photogrammetry for Game Environments 2014-2019*“ meinen : „Die Gesetze bezüglich digitaler Repräsentation von realem Eigentum sind überholt. Dazu kommt, dass die Spieleindustrie keine Erfahrung hat, wie man die Genehmigung einholt für die Darstellung realer Orte im Spiel. Man kann nur spekulieren, inwiefern Gesetze bezüglich Privateigentums oder Urheberrechts greifen“ (Statham et al., 2020).

Interessant ist der Blick auf die Verteilung der Assets bezüglich der Typisierung. Es wird unterschieden in:

- Natural Props: Kleinere natürliche Objekte wie Steine, Äste oder Moos
- Man-Made Props: Kleinere handgemachte Objekte wie Fässer, Kisten, Müll oder Möbel
- Natural Environment: Größere natürliche Objekte wie Steinformationen, Bäume oder Höhlen
- Man-Made Environment: Größere gebaute Objekte wie Häuser, Burgen oder Schuppen
- Modulare Assets: Alleinstehende Objekte, die als Bausteine genutzt werden können, um andere Assets zu unterstützen. Ein Beispiel für Modulare Assets sind beispielsweise

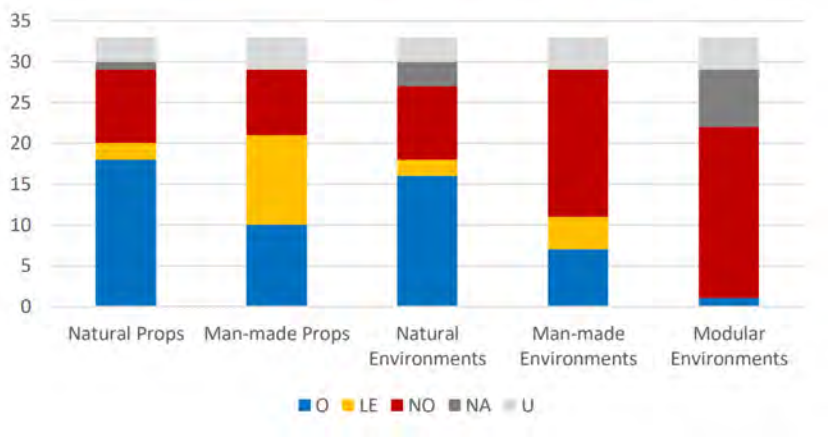


Abbildung 2.9: Verteilung der verschiedenen Asset-Typen (Statham et al., 2020)
O = Observed, LE = Limited Extent, NO = Not Observed, NA = Not Applicable, U = Unavailable

Holzbalken. Das Holzbalken Asset kann nach Belieben rotiert und ausgerichtet werden, um die Statik eines Hauses glaubwürdiger zu gestalten.

(Statham et al., 2020)

Die Abbildung 2.9 beruht nicht auf exakten Angaben der Entwickler, welche und wie viele Objekte mit Photogrammetrie erstellt worden sind, sondern auf Beobachtungen und Nachfragen. Aus den Daten ist abzuleiten, dass größtenteils kleinere und größere natürliche Objekte aufgenommen und als Assets in den Spielen genutzt worden sind. Des Weiteren stellte man fest, dass Entwickler in der Regel sowohl große als auch kleine Objekte als Assets verwenden, wenn sie Photogrammetrie für natürliche Objekte einsetzten. Die hohe Anzahl von natürlichen Props ist auch nachvollziehbar, denn kontrollierte Lichtverhältnisse innerhalb extra eingerichteter Studios machen die Aufnahmen einfacher. Außerdem gibt es keine veränderten Wetterverhältnisse oder andere Einflüsse wie Verkehr oder Passanten (Statham et al., 2020).

In dem Blog-Post von Activision über Photogrammetrie für Call of Duty Modern Warfare 3 aus dem Jahr 2019 wird genau von speziellen Studios berichtet. Die Entwickler bauten einen Raum mit insgesamt 160-200 Kameras, die simultan Bilder machten. Diese wurden dann sechs Stunden automatisch prozessiert und am Ende hatten die Entwickler ein Asset. Neben dem Scanner wurden auch natürliche und handgemachte Objekte wie Metallrohre oder gestapelte Ziegel eingescannt. Der Art-Director Joel Emslie sagte: „We'd done this for years [...]", die aufgenommenen Assets waren fotorealistisch. Die Herausforderung war nun, wenn „[...] we have an environment that's as real [as a photograph], everything else has to be photographically real too." (Hodgson, 2019).

Nicht nur große Studios nutzen Photogrammetrie. Besonders hervorzuheben ist Quixel.

Quixel bietet eine Bibliothek mit über 16.000 eingescannten Assets und Texturen an. Mit der Quixel-Bridge können die Assets direkt in die gewünschte Game-Engine oder 3D-Software importiert werden. Quixel ist auch für Privatpersonen im Abo-Modell zugänglich („Quixel | 3D world-building made easy“, n. d.).

Zusammenfassend lässt sich sagen, dass Photogrammetrie seit 2014 vollkommen in die Game-Industrie integriert ist. Über die Jahre wurden die Prozesse automatisiert und verfeinert. Die Aufnahmetechniken wurden so optimiert, dass die Endergebnisse möglichst fotorealistisch sind.

2.4 Vorteile, Herausforderungen und derzeitige Grenzen der Photogrammetrie

Speziell für Anwendungen in der Unterhaltungsbranche werden üblicherweise hochauflösende Video- oder Fotokameras eingesetzt, um Bilder von Objekten zu schießen. Hochauflösende Fotosensoren sind heute meist kostengünstig und eine Standard DSLR wiegt meist weniger als 1,5 kg. Dadurch ist die Kamera problemlos im Freien einsetzbar. Generell sind die Vorteile der Aufnahme von Digitalbildern die Praktikabilität und Schnelligkeit der Aufnahmen, und der Prozess ist weitestgehend nicht aufdringlich (Bzgl.: Aufnahmen in Museen oder kulturellen Orten) (Foster, 2014).

Der Workflow und die eingesetzten Programme für die Photogrammetrie werden im Kapitel 3 vorgestellt. Trotzdem gibt es bei der Aufnahme der Zielobjekte bekannte Probleme, die nach der Verarbeitung die Qualität des 3D-Modells verschlechtern können.
Diese Herausforderungen beziehen sich auf die Zielobjekte und den Prozess der Aufnahme der Bilder. Einige von ihnen werden hier erläutert:

- **Verdeckung**
  Umgebungsverdeckung (Occlusion) beschreibt, wenn ein ungewolltes Objekt zwischen der Kamera und dem Zielobjekt kommt. Bei Nahbereichsphotogrammetrie kann dies eine Brille vor dem Gesicht sein, bei Luftbildphotogrammetrie zum Beispiel Wolken. Die meisten Zielobjekte, die rekonstruiert werden, zeigen eine gewisse Selbstverdeckung auf, meist in Form von Überlappungen, Verformungen oder komplexen Formen. Die einzige Möglichkeit zur Minimierung dieses Problems ist, die Anzahl der aufgenommenen Bilder zu erhöhen und den Winkelunterschied zu jedem Bild zu verkleinern. Als Faustregel bei starker Verdeckung sind Bilder im Abstand von 5° bis 10° mit Überlappung des Bildinhaltes von 50%. Bei weniger verdeckten Objekten reicht ein Winkelunterschied von 20° (Foster, 2014).

- **Merkmale Features**
  Die Algorithmen der Photogrammetrie suchen nach Parallaxverschiebungen von Merkmalen. Werden nun Bilder von weißen Wänden oder leeren Oberflächen gemacht, werden keine Koordinaten generiert und folglich nicht im Gesamtnetz dargestellt. Merkmale wie Linien oder markante Eigenschaften sind perfekt für den Algorithmus. Um dieses Problem zu umgehen, wird versucht, Merkmale durch Klebebänder und Sticker aufzukleben oder auch Muster auf die Oberfläche des Objekts zu projizieren (Foster, 2014).

- **Transparenz und Reflektion**
  Bei glänzenden oder transparenten Oberflächen kann auch ein projiziertes Muster keine Abhilfe leisten. Erneut sucht der Algorithmus nach Korrespondenz in Bildern. Sollte nun eine Punktreflexion auf einem Spiegel oder ein heller Punkt auf der Oberfläche des gleichen Spiegels sein, kann der Algorithmus es nicht mehr unterscheiden und es kommt zu inkompletten oder ganz misslungenen Meshes. Spiegelnde und transparente Oberflächen bleiben folglich eine der größten Herausforderungen der Photogrammetrie, sodass das Ergebnis abhängig ist von der Oberflächeneigenschaft des Objekts (Foster, 2014).

- **Belichtung**
  Es sollte stets darauf geachtet werden, die Lichtquelle nicht zu verändern. Die Position und Intensität der Lichtquelle sollten auf jedem Bild gleich sein, um konsistente Ergebnisse zu erzielen. Der Algorithmus funktioniert demnach am besten bei konstanter diffuser Beleuchtung. Aufnehmen der Fotos draußen bei bewölktem Himmel kreieren die besten Voraussetzungen. Aufnahmen in einem Haus benötigen Diffuser, um gleichmäßige Lichtverhältnisse zu schaffen. Von der Verwendung des Blitzes wird abgeraten, da bei jedem Auslöser eine neue einzigartige direktionale Lichtsituation geschaffen wird (Foster, 2014).

- **Dünne Objekte**
  Bei dünnen, amorphen oder filigranen Objekten scheitert der Algorithmus meist dabei, Details voneinander zu unterscheiden. Koordinaten werden falsch gesetzt, sodass die resultierenden Meshes fragmentierte Oberflächen aufweisen, die nicht mehr viel mit dem Objekt gemeinsam haben. Besonders bei Blättern oder Ästen, die sich im Aufnahmeprozess durch Wind bewegen, kommt dies häufig vor. Wie in Abbildung 2.10 werden Details ignoriert und nur Umrisse erkannt. Erneut schaffen hier nur mehr Detailaufnahmen Abhilfe (et al., 2022).

Zusammengefasst kann man sagen, dass Photogrammetrie bei der Darstellung als 3D-Mesh besonders mit einheitlichen Oberflächen, reflektierenden und dünnen Objekten Probleme aufweist. Demgegenüber weisen Neural Radiance Fields durch ihren fundamental anderen Darstellungsansatz einen Vorteil auf (et al., 2022).

Abbildung 2.10: Photogrammetrie Probleme bei dünnen Objekten (et al., 2022)

2.5 Neural Radiance Fields

Neural Radiance Fields ist eine Technik, die an 2D Bildern ein neuronales Netzwerk trainiert, welches als Ausgabe eine volumetrische 3D Darstellung neuartiger Sichten ausgibt. Neural Radiance Fields, kurz NeRFs, wurden 2020 von Ben Mildenhall und anderen Forschern von der UC Berkeley und Google Research präsentiert (Mildenhall, 2020).

Zunächst ist es wichtig zu verstehen, welches Problem Neural Radiance Fields löst. Sowohl bei Neural Radiance Fields als auch Photogrammetrie (Multi-View-Stereo) ist das Ziel eindeutig: Die Rekonstruktion eines Objektes, welches aus neuen Blickrichtungen angeschaut werden kann, die nicht in den Trainingsblickwinkeln sichtbar sind. Das Produkt der Multi-View-Stereo Pipeline (Photogrammetrie) ist das texturierte Mesh. Die Geometrie und Texturierung des Meshes geschieht nur auf Grundlage der Trainingsbilder und deren zugehörigen Blickwinkel. Das Mesh kann zwar nun von neuen Blickwinkeln betrachtet werden, diese repräsentieren aber nicht akkurat die speziellen ansichtsabhängigen Informationen wie z. B. Transparenz oder Spiegelungen. Dieses Problem lösen NeRFs. NeRFs können ansichtsabhängige Informationen von Blickwinkeln rendern, die nicht in den Traingsblickwinkeln zu sehen sind.

Um das Verfahren grundlegend zu verdeutlichen, wird chronologisch durch die Arbeitsschritte geführt. Beginnend mit den Grundlagen der Repräsentation der Szene als Funktion. Anschließend die Inputs des Multi-Layer-Perceptron, die konzeptuelle Stütze von NeRFs und schließlich das resultierende volumetrische Rendering neuer Blickwinkel.

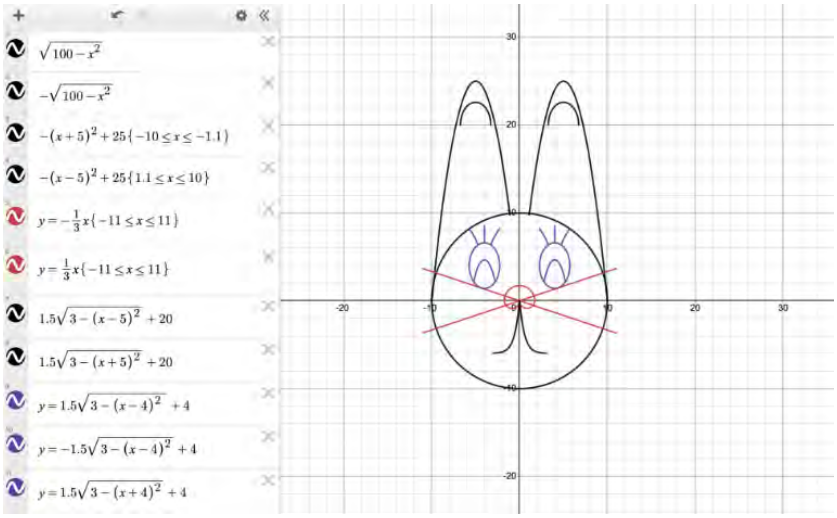


Abbildung 2.11: Beispieldarstellung einer Kontur durch Funktionen (et al., 2022)

2.5.1 Szenenrepräsentation als Funktion

Um Szenenrepräsentation als Funktion zu verstehen, lohnt es sich zunächst, im 2D-Bereich zu bleiben. Es gibt verschiedene Möglichkeiten, 2D Bilder im Computer zu repräsentieren. Die gängigste Form ist durch Pixel in Dateiformaten wie JPEG oder PNG. Außerdem gibt es die vektorielle Darstellung. Bilder werden durch Linien, Kurven oder Kreise dargestellt. Eine Repräsentation durch eine Punktwolke (Pointcloud) ist ebenso möglich. Diese ist dann in Form einer Liste von Punkten mit zugehörigen Koordinaten repräsentiert (Grechnyev, 2023) (et al., 2022).

Zuletzt gibt es auch die Möglichkeit, das Bild mit Formeln bzw. Funktionen zu repräsentieren. Soll die Farbe C eines bestimmten Punktes (x,y) eines Bildes bestimmt werden, kann dies als Funktion C=f(x,y) parametrisiert werden. Dies ist die volumetrische Darstellung in 2D. Die Oberfläche, beziehungsweise Kontur im 2D-Bild, wird als implizite Funktion f(x,y)=0 dargestellt. Die Farbe und Kontur ergeben dann ein Bild.

Die Kontur des Zielbildes lässt sich in 2D näherungsweise durch ein paar nichtlineare Funktionen repräsentieren. Sollte die Ziel-Geometrie jedoch komplexer werden, nimmt demnach auch die Anzahl und die Komplexität der Formeln zu. Um diese Funktionen zu approximieren, wird ein neuronales Netzwerk benutzt. Dem Input-Bild wird für jeden Pixel eine Koordinate (x,y) zugeordnet und die Farbe mit C=f(x,y), als auch die Kontur mit f(x,y)=0 werden approximiert.

Die Approximation passiert im neuronalen Netzwerk. Dieses wird Multi-Layer-Perceptron (MLP) genannt und hat die Fähigkeit, Trainingsinstanzen iterativ vorwärts und rückwärts zu durchlaufen, um den Fehler der Verlustfunktion durch Anpassung der Gewichtungswerte

mittels Gradientenabstieg zu reduzieren. Dies wird Backpropogation genannt und ist die Kernfunktion, welche die Trainingsfähigkeit des MLP ermöglicht. Ist das MLP nun trainiert, kann die Farbe und die Kontur vorhergesagt werden (Grechnyev, 2023).

Obwohl das MLP mehr Parameter als die Anzahl der Pixel im Bild hat, ist die Darstellung unscharf und nicht genau, siehe Abbildung 2.12a. Die einfachen Koordinaten im Bild sind zu ungenau und kleine Details, die wichtig für die Gesamtdarstellung sind, gehen verloren.

Abhilfe schafft hier, den Input zu verändern. Bisher wurden die Koordinaten (x,y) in das MLP gespeist. (Tancik et al., 2020) zeigen, dass eine Vorverarbeitung der Input Koordinaten durch Fourier-Analyse zielführend ist. Die Input Frequenz, hier die beiden Koordinaten, werden durch Fourier Spektralanalyse in Einzelteile zerlegt. Statt dem einen Koordinatenpaar werden die Spektralanteile des Paares einzeln eingespeist. MLPs sind bekannt dafür, besonders bei koordinatenbasierten Grafikaufgaben Schwierigkeiten zu haben beim Lernen von hochfrequenzen Funktionen. Dieses Phänomen wird in der Literatur „Spectral Bias“ genannt. Das MLP hat einen sogenannten Kernel. Dieser Kernel ist eine Funktion, die die Anwendung linearer Methoden auf reale Probleme nichtlinearer Natur ermöglicht(„Designing of different kernels in Machine Learning and Deep Learning.“ n. d.). Die Theorie hinter MLPs legt nahe, dass der Kernel auf Koordinaten Input mit rapidem Frequenzabfall antwortet, was ihn daran hindert, den hochfrequenten Inhalt natürlicher Bilder und Szenen darstellen zu lassen. Die Forscher haben beobachtet, dass, wenn die Input Koordinaten in ihre Spektralanteile aufgeteilt und einzeln eingespeist werden, der Kernel mit einer stabilen Frequenz antwortet. Das Spektrum des Kernels kann nun durch Modifikation der Frequenzvektoren $\mathbf{b}_j$ gesteuert werden, sodass der Frequenzbereich vergrößert wird. Mit mehr Input der gleichen kodierten Koordinaten kann das MLP nun feinere Geometrie-Details in Bildern, die hohe Frequenzen haben, besser darstellen und akkurater Farbinformation in Form von RGB-Werten vorraussagen. Dieser Prozess wird im Englischen „Positional Encoding“ genannt. Das MLP ist dann in der Lage die Farbe der Koordinaten vorherzusagen und die Kontur in einem gerenderten Bild wiederzugeben 2.12b (Tancik et al., 2020).

Im Gegensatz zum 2D-Bereich wird im 3D-Bereich eine weitere Koordinate hinzugefügt (x,y,z) – die Theorie dahinter bleibt bestehen. Um die Szene zu generieren, wird Structure-from-Motion auf die Aufnahmebilder angewandt und an den Bildern mit bekannter Position im Raum (x,y,z) kann das MLP trainieren.

Identisch wie bei 2D, lässt sich die Szene in 3D auch als eine Reihe von Funktionen repräsentieren. Die „Kontur“ im dreidimensionalen Raum, jetzt als Oberfläche oder Geometrie gemeint, kann wieder als implizite Funktion von f(x,y,z)=0 dargestellt werden.

Hier ist es entscheidend zu verstehen, dass nicht nur Oberflächen, sondern das komplette Volumen repräsentiert werden. Wird beispielsweise das MLP an Bildern einer Metalltür trainiert, gibt es natürlich wenig Tiefeninformationen, die dargestellt werden können, da die Bilder

(a) MLP Output (b) MLP Positional Encoding (c) Ziel-Bild

Abbildung 2.12: MLP Training an Ziel-Bild (et al., 2022)

nicht „in“ das Metall sehen können. Sollte in der Szene aber ein Objekt mit leichter Transparenz zu sehen sein, werden die Tiefeninformationen, die von der Kamera aufgenommen werden, auch volumetrisch dargestellt.

Es wird im MLP iterativ durch die Ebenen gegangen, um die Szene mit Gradientenabstieg und Backpropogation möglichst real zu repräsentieren. Farbe wird mit C=f(x,y,z) gegeben. Zu jedem Punkt im 3D-Raum (f=(x,y,z)) wird mit C Farbe hinzugefügt. Das Resultat: Eine komplette dreidimensionale volumetrische Darstellung der teils transparenten Punkte mit zugehörigen approximierten RGB-Farbwerten. Dieses Konzept der Szenenrepräsentation bildet das Grundgerüst von NeRFs.

Die dreidimensionale volumetrische Darstellung lässt sich am besten als kontinuierliches, blickrichtungsabhängiges Volumen mit Opazität beschreiben oder noch bildlicher als durchsichtiges, farbiges Gelee (et al., 2022).

2.5.2 Herleitung der Formel und Input von NeRFs

Bevor neue Blickwinkel gerendert werden können, lohnt es sich noch, die zugrunde liegende Funktion anzuschauen. Als konzeptuelle Stütze lässt sich der Input mit der Plenoptischen Funktion herleiten. 1991 wurde in einem einflussreichem Paper von Anderson und Bergen (Chan, 2014) die Plenoptische Funktion vorgestellt, um die visuelle Wahrnehmung der Menschen zu parametrisieren, um sie dann computergrafisch darzustellen. Die Plenoptische Funktion setzt sich aus sieben Variablen zusammen. θ und ϕ bilden hier zusammen die Blickrichtung. Lambda ist die Wellenlänge. Diese drei Variablen alleine bilden einen Farb-„Schnappschuss“ einer bestimmten Blickrichtung. Nun kommt noch mit t Zeit dazu. Die letzten drei Variablen sind Koordinaten. Diese sieben Variablen zusammen bilden eine

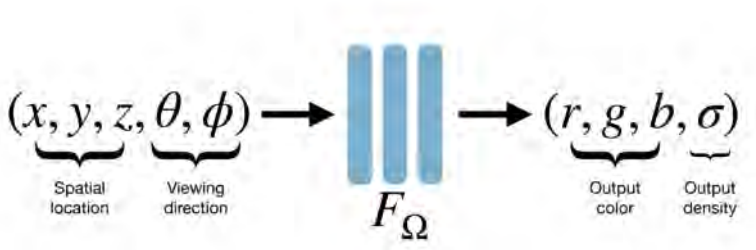


Abbildung 2.13: Formel der Szenenrepräsentation, F_Ω = MLP (et al., 2022)

farbliche Darstellung der Welt zu jeder Zeit, jeder Position und jedem Blickwinkel. So soll diese Funktion unsere gesamte visuelle Realität darstellen (et al., 2022).

$$P(\theta, \phi, \Lambda, t, V_x, V_y, V_z)$$

Der Input der NeRFs ist hier von eine simplifizierte Version. Zeit t und Wellenlänge Lambda werden ignoriert. Nun müssen die Blickrichtung und die Position im Raum aus 2D Bildern gewonnen werden. Identisch wie bei Photogrammetrie wird die Kameraposition der 2D-Bilder durch den Structure-from-Motion Algorithmus ermittelt. Deswegen lassen sich diese beiden Verfahren auch sehr gut vergleichen, da nicht nur der Datensatz identisch ist, sondern auch, weil sie beide mit den gleichen zugrunde liegenden Informationen der extrinsischen (Kamerawinkel und -position) und intrinsischen Parametern (Brennweite, ISO, Verschlusszeit, Vignette, Körnung und Chromatische Aberration) arbeiten.

Der Input sind die Koordinaten r=(x,y,z) und die Blickrichtung und Position der Kamera werden durch θ und ϕ dargestellt. Wie im Abschnitt 2.5.1 erklärt, werden diese Koordinaten durch Fourier-Transformation kodiert. Der Output des MLPs F ist dann die Farbe C=(r,g,b) und die Dichte σ. Die Dichte σ beschreibt in der volumetrischen Darstellung, wie transparent oder opak dieser bestimmte Punkt im Volumen ist.

Die Plenoptische Funktion geht vom Betrachter aus. Er nimmt mit seinen Augen die visuellen Informationen um sich herum wahr. Der Ansatz der NeRFs unterscheidet sich, da er nicht vom Betrachter ausgeht. Hier wird vorausgesetzt, dass sowohl jeder Punkt direktionales Licht mit C=(r,g,b) emmitieren als auch Licht mit Dichte σ absorbieren kann. Diese beiden Eigenschaften bilden die Strahldichte (Radiance) (et al., 2022).

2.5.3 Volumetrisches Rendering

3D-Rendering an sich kann als eine Funktion definiert werden, die eine 3D-Szene als Eingabe annimmt und ein 2D-Bild ausgibt. Übersetzt ist es zu beschreiben als Bildgebungsverfahren.

Wie in Abschnitt 2.5.1 erklärt, ist die 3D-Szene, die mit dem MLP trainiert wurde, eine komplette kontinuierliche volumetrische Darstellung.

Da die 3D-Szenenrepräsentation nicht auf einem Bildschirm direkt wahrnehmbar ist, wird sie in der Regel von einem bestimmten Standpunkt oder Blickwinkel aus dargestellt, der durch die Kameraparameter festgelegt ist: intrinsisch (Brennweite, Bildgröße) und extrinsisch (Kameraposition und -richtung). Das Ergebnis ist ein 2D-Bild. Dieses Bild wird aus der volumetrischen Darstellung synthetisiert.

Jeder Pixel auf dem „To-be synthesized Picture“ wird zu einem Strahl in der 3D Szene. Dieser Strahl sampelt kontinuierlich entlang des Strahls die Farbe und Dichte (Radianz) der Punkte, die er durchdringt. Trifft der Strahl auf einen Punkt im Abstand t, so wird die Farbe als Pixelfarbe zurückgegeben. Dabei wird jedoch nicht die Farbe des ersten Treffers zurückgegeben, sondern ein Beitrag aller Punkte, die der Strahl trifft (et al., 2022).

$$C(\mathbf{r}) = \int_{t_1}^{t_2} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot \mathbf{c}(\mathbf{r}(t), \mathbf{d}) \cdot dt$$

Dabei lässt sich C(r) vielmehr als Integral darstellen. Die Integrationsgrenzen t_1 und t_2, sind die Anfangs- und Enddistanz des ersten und letzten Sampels. Die Beiträge der Sampels werden durch die Dichte σ und Durchlässigkeit („Transmittance“) T multipliziert und dadurch gewichtet. Es ist zu beachten, dass das Rendering volumetrisch ist und es kein Konzept wie eine „Oberfläche“ gibt. In der praktischen Anwendung wird das Ergebnis bei opaken Objekten jedoch von dem kleinen Bereich nahe der Oberfläche dominiert, abhängig von der Dichte σ und Durchlässigkeit T.

$$T(t) = \exp\left(-\int_{t_1}^{t} \sigma(\mathbf{r}(u)) \cdot du\right)$$

Durchlässigkeit lässt sich als Integral aus Dichte σ herleiten.

In dem Beispiel der Metalltür heißt das, dass die Farbe auf der Oberfläche der Metalltür den höchsten Einfluss der Farbe auf dem Pixel hat.

Im Idealfall gibt es auf dem Strahl so viele Sampels wie möglich, um möglichst akkurat die Farbinformationen darzustellen. Da die Anzahl der Sampels aber direkt mit dem rechnerischen Bedarf korreliert, gibt es verschiedene Sampel Ansätze. Im Original-NeRF gibt es 265 uniformierte Sampelpunkte. Sollte die Szene durch die Bounding Box begrenzt sein, lohnt sich Uniform-Sampling, um gleichmäßig abzutasten. Sollte die Szene jedoch unbegrenzt sein, lohnt sich das Uniform-Sampling nicht, da die Sampling-Informationen für sehr nahe Objekte

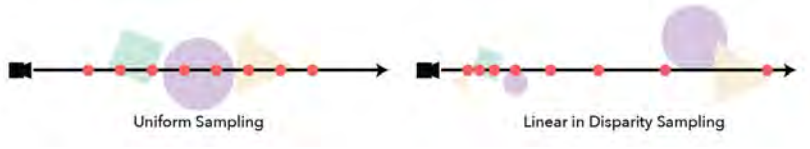


Abbildung 2.14: NeRF Sampling-Typen (Tancik et al., 2023)

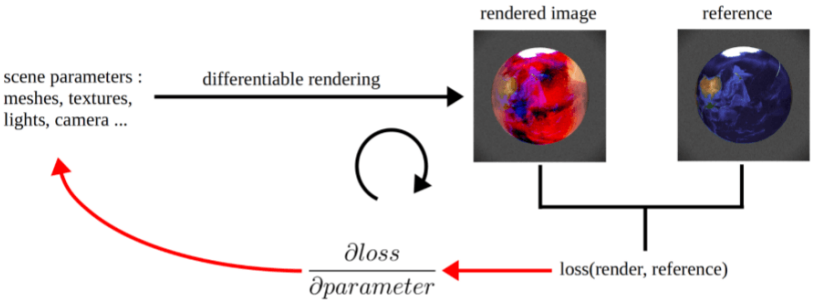


Abbildung 2.15: Differenzierbares Rendering (Robineau, 2021)

spärlich ausfallen würden. Hier kann „Uniform in Disparity“-Sampling eventuell bessere Ergebnisse liefern (siehe Abbildung 2.14) (et al., 2022).

Ein wichtiger Aspekt bei diesen Formeln ist, dass sie differenzierbar sind. Das heißt, dass das volumetrische Rendering mit der obigen Funktion ableitbar ist 2.15. Dies ermöglicht die Ableitung nach den verschiedenen Szeneparametern. Das MLP vergleicht also stetig das Gerenderte mit den Aufnahmebildern (Ground-Truth). Die abgeleiteten Szeneparameter werden zurück in das MLP gespeist, um die Szene noch weiter zu optimieren, sodass die Diskrepanz zwischen dem Gerenderten und der Ground-Truth so minimal wie möglich ist (Robineau, 2021).

Das Ergebnis des Renderings ist anschließend die Szene, betrachtbar aus allen möglichen Blickwinkeln und Positionen.

2.5.4 Vorteile, Herausforderungen und derzeitige Grenzen von Neural Radiance Fields

Wie soeben beschrieben, liegt die Stärke in der Repräsentation des gleichen Punktes aus anderen Blickrichtungen. NeRFs sind somit in der Lage, die verschiedenen Lichteffekte wie Reflektionen und Transparenz darzustellen. Auf diese Weise ist es möglich, eine deutlich realistischere Darstellung zu bekommen als die Repräsentation durch ein Mesh mit Multi-View-Stereo (Pepe et al., 2023). Der Vorteil dieser realistischen Darstellung bleibt aber nach exportieren in ein Mesh nicht bestehen. Die Blickrichtungsabhängigen Lichteffekte werden sozusagen eingefroren und sind anschließend nicht mehr veränderbar. Sie sind „baked“. Sollten

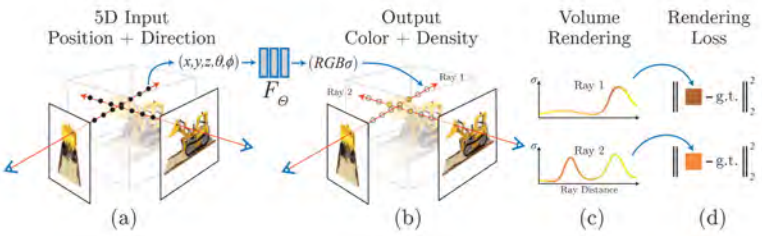


Abbildung 2.16: NeRF Ablauf (Mildenhall, 2020), (a) Zeigt den Aufbau der Szenerepräsentation aus den Aufnhamebildern mit Punktkoordinaten und Kamerapositionen, (b) Nach Einspeisen dieser Werte in das MLP werden Farbe und Dichte ausgegeben. (c) Anhand der Sampelinformationen der Rays wird das Volumen als 2D-Bild gerendert. (d) Da die Formel differenzierbar ist kann die NeRF Szene durch die Diskrepanz zu den Aufnahmebildern optimiert werden.

die Oberflächen nun andere Lichtverhältnisse repräsentieren, müssen die Lichtinformation auf der Oberfläche von dem Objekt separiert werden. Das nennt man in der Computergrafik „Delighting“ und wird üblicherweise nach dem Erstellen der Meshes gemacht. NeRFs Renderings können schwer in Szenen implementiert werden, welche andere Lichtverhältnisse haben, ohne dass sie nach der Transformation in ein Mesh gedelighted wurden.

Für die Aufnahmen der Objekte sind NeRFs, verglichen zu Photogrammetrie, etwas robuster. Die Aufnahmebilder sollten einen Punkt von mindestens fünf Blickrichtungen einfangen, mit circa 30% Disparität. Besonders geeignet sind hier Weitwinkelobjektive (Fisheye-Lens), da sie bei den Aufnahmen mehr Informationen bei gleichen Blickwinkeln sammeln.

Die intrinsischen Einstellungen der Kamera können das Ergebnis auch verändern. Bei Einzelbildaufnahmen des Objekts wird die Verschlusszeit der Kamera oftmals sehr kurz gewählt, um möglichst gleichmäßig um das Objekt zu rotieren und Aufnahmen zu machen. Der Nachteil kurzer Verschlusszeit ist erhöhtes Bildrauschen auf den Bildern. NeRFs können mit Rauschen gut umgehen, da die Bildinformationen über mehrere Bilder gemittelt werden, sodass Rauschen herausgemittelt wird (et al., 2022).

NeRFs bieten eine Alternative zur traditionellen Photogrammetrie, welche vielversprechende Ergebnisse liefern. Trotzdem gibt es in vielen Bereichen immer noch große Hindernisse und Probleme:

- Die gerenderte Szene ist stets statisch mit statischen Lichtverhältnissen (et al., 2022).
- Das Original NeRF von 2020 hatte extreme Probleme mit Unschärfe und Aliasing-Artefakten. Dadurch, dass jedem Pixel ein Strahl zugeteilt wurde, hatten Objekte im Vordergrund höhere Auflösungen als Objekte mit etwas Entfernung. Diese Auflösungsdiskrepanz machte sich durch das Aliasing bemerkbar (Barron et al., 2021).

Abbildung 2.17: Bild aus einem trainiertem Neural Radiance Field aus neuer Blickrichtung

- Das Trainieren des neuronalen Netzwerks mit NeRF ist extrem rechenintensiv und erfordert erhebliche Mengen an Rechenleistung und Arbeitsspeicher (Remondino et al., 2023) (et al., 2022).
- NeRFs können bis zum aktuellen Zeitpunkt nur mit Nvidia Grafikkarten, welche die CUDA Programmierschnittstelle haben, gerendert werden (Grechnyev, 2023).
- Es gibt kein standardisiertes Dateiformat für NeRFs. Die gängigen 3D Modell „Exchange Formats“ wie OBJ oder FBX unterstützen keine NeRFs (Grechnyev, 2023).
- NeRFs erfordern in gleicher Wiese wie Photogrammetrie die genauen Kamerapositionen. Dadurch sind sie immer noch angewiesen auf den Structure-From-Motion Algorithmus (et al., 2022).
- Die NeRFs sind keine Meshes. Genauso wie bei Photogrammetrie ist die Qualität des volumetrischen Renderings und der nachfolgenden Übersetzung in ein Mesh abhängig von der Auflösung der Aufnahmebilder. Hier gilt die gleiche Faustregel. Je höher die Auflösung der Aufnahmebilder, desto feiner die Geometrie und Textur des Meshes. Von Nachteil ist der damit verbundene erhöhte Rechenaufwand (Remondino et al., 2023). Außerdem sind NeRFs, sofern sie nicht in ein Mesh exportiert sind, nicht editierbar, animierbar und deformierbar (Grechnyev, 2023).

Das Extrahieren der Oberflächen ist weiterhin eine der größten Herausforderungen für NeRFs. Moderne Game-Engines und VFX-Programme sind auf Polygonrendering ausgelegt und können die NeRF-Rekonstruktionen von sich aus nicht rendern, bevor sie nicht in eines der Standardformate übersetzt wurden.

Viele dieser Probleme werden in den Weiterentwicklungen der NeRFs bereits aufgenommen. Es ist stets zu beachten, dass diese Technik auch in der Zukunft sehr viel Weiterentwicklungspotenzial hat.

3 Methodik und Vorbereitung des Vergleichs

3.1 Methodik

Das Hauptziel ist eine kritische Bewertung von NeRF-basierten Methoden im Vergleich zu der konventionellen Photogrammetrie durch objektive Messung der Qualität der resultierenden Polygonmeshes. Die Analyse beschränkt sich dabei auf die resultierenden Meshes, sie bilden die Baseline für jede weitere Aktion, die im Asset-Workflow folgt. Retopologie der Vertices, UV-Mapping, Texturbearbeitung und Delighting spielt für diesen Vergleich keine Rolle. Es wird evaluiert, wie genau die Methoden das Zielobjekt rekonstruieren können und welche Zeit dies in Anspruch nimmt. Um dies zu erreichen, werden drei Bilder Datensätze mit unterschiedlichen Umgebungen, Zielobjektgrößen und Oberflächencharakteristika berücksichtigt. Es werden Oberflächencharakteristika wie Transparenz, Reflektion, metallische und feine Strukturen untersucht. Zwei der drei Bilder Datensätze wurden eigens angefertigt, um zu beleuchten, welche Ergebnisse für Laien und Interessierte bei weniger professioneller Hardware oder beschränktem Zugang zur Software zu erwarten sind. Des Weiteren soll dies aufzeigen welche Probleme bei den Aufnahmeprozessen entstehen können und inwiefern sich diese in den Ergebnissen widerspiegeln.

Es wird eine NeRF Methode angewendet, sowie zwei verschiedene Photogrammetrie Softwareprogramme. Jedes Mesh wird mit maximal 100.000 Punkten, maximal 50.000 Faces und einer Texturauflösung von 4096 Pixel gerendert. Diese Werte entsprechen einem typischen detailliertem Umgebungsasset vor der Retopologie (Tkalych, 2024). Retopologie ist der Prozess der Reduzierung der Faces mit Ziel der Simplifizierung des Meshes im Allgemeinen. Bei der Retopologie werden fehlerhafte Faces, falsch positionierte Vertices oder Oberflächen geglättet.

Die vorgeschlagene Bewertungsstrategie beinhaltet einen objektiven Vergleich der resultierenden Meshes mit Meshlab und einen quantitativen Pointcloud Vergleich mit CloudCompare. Die NeRF-Rekonstruktion basiert auf dem Nerfstudio Framework (Tancik et al., 2023). Es sei daran erinnert, dass die NeRF Rekonstruktionen ein neuronales Rendering sind, welche kein Standard-Format haben. Daher wird die Poisson Surface Reconstruction (Kazhdan et al., 2006)

angewendet, um ein Mesh aus jeder Rekonstruktion zu erstellen. So kann die volumetrische Darstellung als Mesh exportiert werden. Diese Methode ist im Nerfstudio Framework bereits implementiert.

Die resultierenden Meshes werden in Meshlab zurück nach ihren realen Abmessungen skaliert. Daraufhin werden diese durch das Alignment übereinandergelegt. Schließlich können die Meshes quantitativ miteinander verglichen werden. Letztendlich folgt noch ein Vergleich der Texturen.

3.2 Vorbereitung und Equipment

Zunächst wurden die beiden Zielobjekte für die Bilderdatensätze ausgewählt.

Datensatz 1: Gebäude

Das Zielobjekt des ersten Datensatzes ist das Gebäude der ehemaligen Miami-Ad-School lokalisiert inmitten des HAW Campus Finkenau. Für die Aufnahmen wurde eine DJI Mini 3 Pro Drohne verwendet. Es wurden rund um das Gebäude, aus verschiedenen Winkeln, mehrere Videoaufnahmen mit 4K Auflösung (3840 x 2160) mit 30 Bilder die Sekunde aufgenommen. Diese Video-Clips wurden danach mit der LosslessCut (Finstad, 2024) Software ohne Qualitätsverlust verknüpft. Das vollständige Video des Rundfluges um das Gebäude hat eine Laufzeit von 2:59 Minuten. Es ist zu erwähnen, dass der Kamera Sensor für Videoaufnahmen gecropped wird. Die Kamera für Videoaufnahmen hat folgende Spezifikationen:

- Sensor: 1/1,3″ CMOS-Sensor , 9.7 x 7.3 mm
- Effektive Pixel: 48 MP
- Sichtfeld: 75°
- Brennweite: 6,72 mm
- 35mm Äquivalente Brennweite: 25 mm
- Blende: f/1.7

Für die Aufnahmen wurde jedoch ein Clip-On Weitwinkelobjektiv, welches vom Hersteller DJI verkauft wird, aufgesetzt. Das Sichtfeld wird von 75° auf 100° erweitert (DJI, n. d.). Blendenzahl,Belichtungszeit und Iso-Wert sind festeingestellt und nicht in den Metadaten sichtbar. Die Sensorgröße wurde seitens DJI nicht genau metrisch angegeben, sodass die grobe Berechnung der neuen Brennweite auf einer Sensorgröße einer Sekundärquelle beruht („DJI Drone Sensor Size Comparison Page“, 2018). Die Sensordiagonale bei den Abmessungen

9.7 x 7.3 mm beträgt 12,1640 mm. Schätzungsweise beläuft sich die Brennweite dann auf 5,1 mm.

$$100° = 2 * \arctan(0,0121640/(2 * X) = 0,0051034039m \approx 5,103mm$$

Der Wert für die Brennweite kann in den Metadaten der Bilderdatensätze bei Windows nicht verändert werden. Da sowohl für Photogrammetrie und NeRF Methoden die Brennweiten durch Structure-from-Motion geschätzt werden, bleibt das für diesen Datensatz irrelevant.

Die Aufnahmen wurden am 15.1.24 13:00 Ortszeit mit bewölktem Himmel gemacht. Die Belichtung war konstant diffus. Der Datensatz ist 678 MB groß, bestehend aus 360 jpg. Bildern.

Datensatz 2: Figur

Das Zielobjekt des zweites Datensatzes ist eine Figur im Innenhof des HAW Campus Finkenau. Die Aufnahmen wurden mit einer Sony Alpha 6300 mit einem Sony Kit Objektiv SELP1650 bei einer Auflösung von 6024 x 4024 aufgenommen. Die Kamera hat folgende Spezifikationen:

- Sensor: Sony APS-C CMOS Sensor , 23,5 x 15,6 mm
- Effektive Pixel: 24,2 MP
- Brennweite: 16-55mm

Folgende Aufnahmeeinstellungen wurden ausgewählt:

- ISO-Filmempfindlichkeit: ISO-800
- Blende: f/5.6
- Belichtungszeit: 1/80 Sek.
- Gammakurve: S-log2
- Brennweite: 20mm

Die Bilder wurden am 1.3.24 um 14:42 Ortszeit aufgenommen. Die Beleuchtung war erneut konstant diffus bei bewölktem Himmel. Es wurden 147 RAW-Aufnahmen mit der Serienaufnahmeoption gemacht. Diese schießt 3 Bilder die Sekunde. Die Dateigröße pro Bild beläuft sich auf 24 MB. Der RAW-Datensatz hat eine Größe von 3,44 GB. Der RAW-Datensatz wurde für Metashape und die NeRF Methode in jpg. konvertiert.

Alle Tests wurden auf einem privaten Computer mit folgenden Spezifikationen durchgeführt:

- CPU: Intel Core i5-12400f 2,5 GHz

	D1 : Gebäude	D2: Figur	D3: Garten
Bilderanzahl	360	147	185
Auflösung	3840 x 2160 px	6024 x 4024 px	5187 x 3361
Kamera Meta-Daten	Nur Brennweite	Ja	Nein
Charakteristika	draußen, Luftbilder, diffuses Licht Fisheyelens, teilweise spiegelnde oder transparente Texturen	draußen, diffuses Licht, wenig Helligkeit, Feine Textur,	draußen, diffuses Licht, Verdeckung, spiegelnde Texturen

Tabelle 3.1: Datensätze, mit verschiedenen Größen, Auflösungen und Oberflächencharakteristika, die für Photogrammetrie und NeRF Methoden evaluiert werden.

- RAM: 32GB DDR4
- GPU: NVIDIA GeForce RTX 3060 12GB VRAM
- Speicher: NVMe M.2 1TB SSD

Datensatz 3: Garten

Dieser Bilder-Datensatz stammt aus dem Test Datensatz der Mip-NeRF 360 Dokumentation (Barron et al., 2022). Der Datensatz besteht aus 185 Bildern mit einer Auflösung von 5187 x 3361. Intrinsische Kameraparameter sind in diesem Datensatz, wie in Datensatz 1, nicht in den Metadaten repräsentiert.

3.3 Structure-from-Motion

Für alle gesammelten Bilder aus den Datensätzen werden die Kamerapositionen benötigt, um die 3D-Rekonstruktion zu erstellen. Die Kamerapositionen werden sowohl für die Photogrammetrie, als auch für die NeRF-basierten Methoden benötigt. SfM ist in den beiden Photogrammetrie Programmen bereits integriert. Für die NeRF Methoden wird SfM zur Verarbeitung vorausgesetzt. Eine kostenfreie SfM Software ist COLMAP. COLMAP ist eine Open-Source, end-to-end, vielseitig einsetzbare Bild-basierte Rekonstruktionssoftware, welche von Johannes L. Schönberger entwickelt wurde. COLMAP beginnt zuerst mit der Feature Extraction. Die eingespeisten Bilder werden bezüglich der intrinsischen Kameraparameter analysiert. Sollten diese fehlen, kann die Software durch Bunde Adjustment eine Abschätzung der Parameter machen. Danach werden die Bilder mit dem Scale-Invariant-Feature-Transform

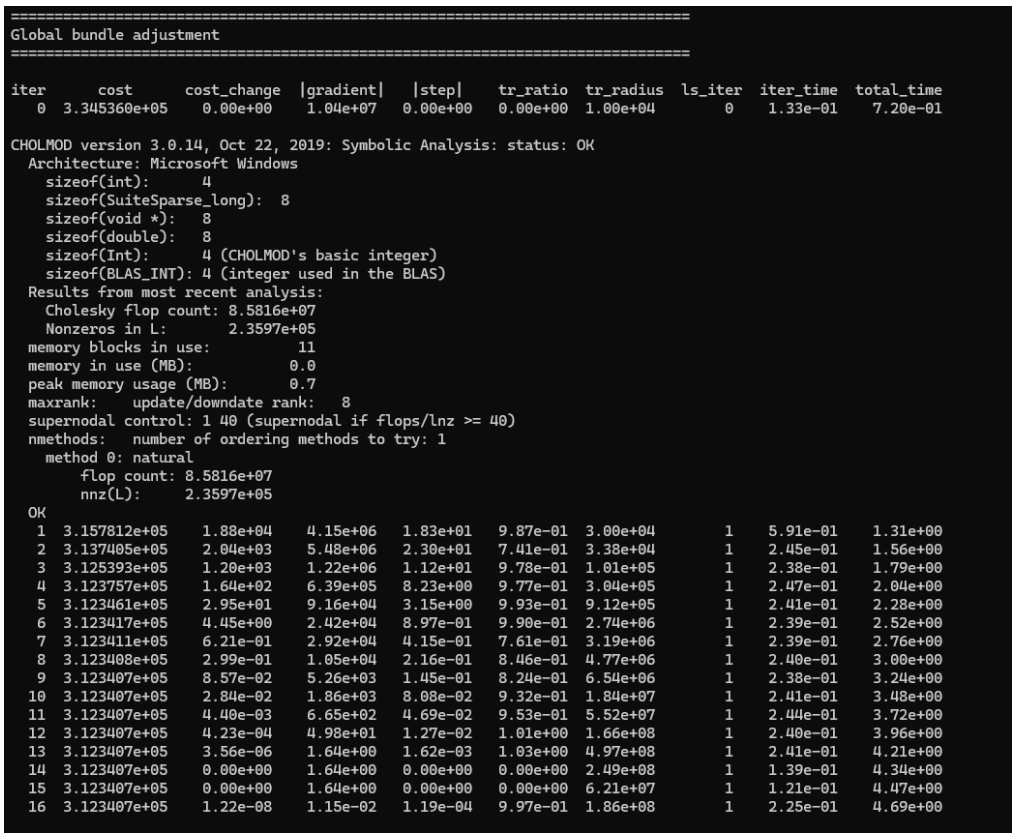


Abbildung 3.1: COLMAP

Algorithmus, ein Algorithmus welcher schon 2005 veröffentlicht wurde, auf Features untersucht. Für diese Arbeit sind Grafikeinheiten vorteilhaft, da sie benutzerdefinierte Modis zur Erkennung von solchen Merkmalen haben. Bei konstrastreichen Bildern ist die Qualität der Merkmale dadurch oft von höherer Qualität. Alle extrahierten Merkmale werden zunächst in einer einfachen Tabelle innerhalb einer Datenbank gespeichert.

Im nächsten Schritt werden die Features „gematched“. Es wird nach Korrespondenzen der Merkmale in der Tabelle der Merkmale gesucht. Dieser Schritt ist der rechenintensivste Schritt, abhängig von der Anzahl der Merkmale und der Gesamtanzahl der zu vergleichenden Bilder. COLMAP lädt daraufhin alle korrespondierenden Daten aus der Datenbank in den Speicher und startet die Rekonstruktion anhand des ersten Merkmalpaares und der zugehörigen Bilder. Die Szene wird iterativ erweitert, indem neue Bilder registriert werden und neue Merkmalpunkte trianguliert werden (Schönberger & Frahm, 2016). Daraus wird auch die Position und Blickrichtung der Bilder im Raum ermittelt, was die Grundlage für die NeRF-Methoden ist. Die COLMAP Rekonstruktion kann in verschiedene Dateiformate exportiert werden, für NeRFs werden nur die Textdateien bestehend aus der COLMAP Datenbank, die geschätzten Kameraparameter, die Kamerapositionen der einzelnen Bilder, die Merkmalliste und der Bild-Datensatz benötigt.

3.4 Photogrammetrie-Software

Meshroom

Meshroom ist die kostenlose Open-Source-Sofware von AliceVision. AliceVision ist ein Verbund von Forschern aus der Industrie, zum Beispiel aus dem französischen VFX-Studio MPC, welche die Meshroom Software entwickeln. Das Projekt ist ein Non-Profit Projekt welches von der EU, Epic Games und dem nationalen Französischen Forschungsinstitut subventioniert wird (Association, n. d.).

Meshroom bietet einen vollautomatisierten One-Click Photogrammetrie Prozess auf Basis eines Node-Systems. In den Nodes kann der Prozess beliebig angepasst werden. Für die Tests wurde Meshroom Version 2023.3.0 verwendet.

Agisoft Metashape

Agisoft Metashape ist die kommerzielle Photogrammetrie Software von Agisoft. Die vormals als "Photoscan" bekannte Software ist mit einer Lizenz für 179 Dollar erhältlich. Agisoft Metashapes Oberfläche ist nicht Node basiert, besitzt aber auch einen Automatisierungsprozess (Batch processing). Es lassen sich sowohl Videos als auch Fotos importieren. Für die Tests wurde Agisoft Metashape Professional Version 2.1.0 in der 30-tägigen Testversion verwendet.

3.5 Neural Radiance Field

Nerfstudio ist ein Framework der NeRF Forscher Matthew Tancik, Ethan Weber und Evonne Ng u. a. Die simple API bietet einen vereinfachten End-to-End-Prozess für die Erstellung, das Training und das Testen von NeRFs. Nerfstudio wurde ursprünglich als Open-Source-Projekt von Berkeley-Studenten im KAIR-Labor des Berkeley AI Research (BAIR) Zentrum im Oktober 2022 als Teil eines Forschungsprojekts gestartet. Es wird derzeit von Berkeley-Studenten und Community-Mitarbeitern weiterentwickelt und verwaltet. Nerfstudio stellt Lernressourcen, Tutorials sowie Dokumentationen zur Verfügung, um den Umgang mit den Methoden und der API zu erleichtern (Tancik et al., 2023).

Für die Installation wird ein Python Environment in Miniconda3 erschaffen. Wie in Abschnitt 2.5.4 erwähnt, können NeRFs zurzeit nur auf Nvidia Grafikkarten erstellt werden. Der NeRF Code basiert hauptsächlich auf dem CUDA Toolkit von Nvidia. Das CUDA Toolkit bietet eine Entwicklungsumgebung für die Erstellung leistungsstarker, GPU-beschleunigter Anwendungen. Neben dem CUDA Toolkit muss außerdem Pytorch installiert werden. Nach erfolgreichem Installieren und Aufsetzen der Umgebung können die Daten nach SfM durch COLMAP, mit folgenden Befehlen verarbeitet werden:

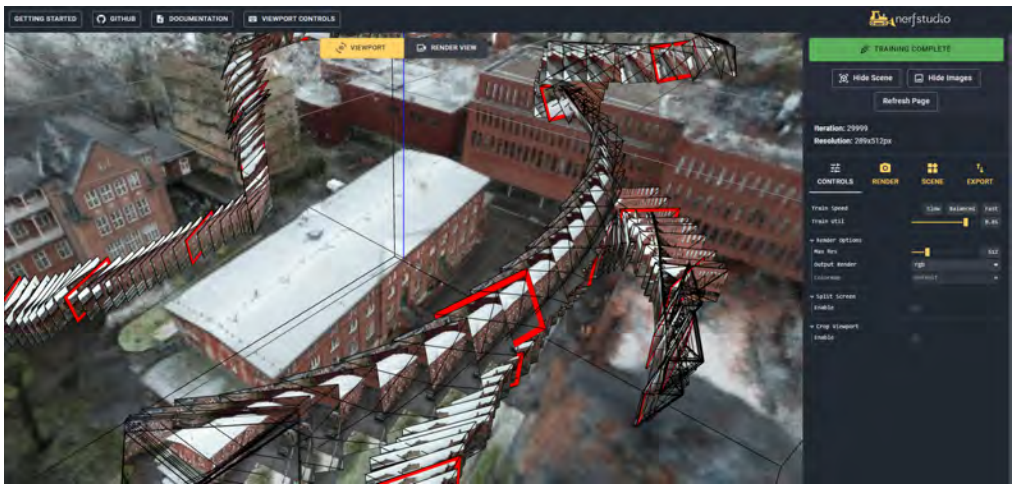


Abbildung 3.2: Nerfstudio Browser-Viewer

```
ns-train nerfacto --data C:\Users\Leo\Pictures\Datasets\Drohnenaufnahme
```

Anschließend trainiert das System eigenständig die angegebene NeRF Methode. Neben dem angezeigten Trainingsfortschritt gibt es außerdem einen Viewer 3.2, der mehrere Funktionen auf einer grafischen Oberfläche im Browser bietet. Im Viewer kann außerdem die Bounding Box für das Mesh skaliert, die Anzahl der Faces und die der Vetices limitiert werden. Ein Beispiel-Befehl für das Exportieren in ein Mesh :

```
ns-export poisson --load-config outputs\Figur_COL_jpg\nerfacto\2024-04-01
_222634/config.yml --output-dir exports/mesh/ --target-num-faces 50000 --num-
pixels-per-side 4096 --normal-method open3d --num-points 100000 --remove-
outliers True --use-bounding-box True --bounding-box-min -0.32 -0.25
-0.30000000000000004 --bounding-box-max 0.08000000000000002 0.25 0.1
```

Die vorgeschlagene NeRF-Methode Nerfacto basiert auf einer anderen NeRF-Methode namens Instant NGP.

Instant NGP

Instant NGP beruht in erster Linie auf dem Vanilla NeRF (Mildenhall, 2020). Direkt nach Einspeisen der Daten baut sich die Szene schnell auf. Dies wird durch eine verbesserte Sampling Methode bewerkstelligt, welche die Trainingsgeschwindigkeit um das 10-100fache erhöht. Die Sampling Methode ignoriert Raum, welcher nicht mit Punkten belegt ist und leeren Raum hinter Bereichen mit hoher Punktdichte. Das Occupancy Grid, ist ein Algorithmus der Hindernisse im umliegenden Raum probabilistisch vorhersagt. Neben einem verbessertem Neuronalen Netzwerk wird auch das Positional Encoding verändert. Die Koordinaten, die in

das MLP eingespeist werden, werden Hash-based kodiert. Die Idee besteht darin, Koordinaten auf trainierbare Merkmalvektoren zu kodieren, die im Standardfluss des NeRF-Trainings optimiert werden können. Durch die Verbesserungen auf allen Ebenen erreicht Instant NGP Geschwindigkeitssteigerungen, die das Training von NeRF-Szenen in wenigen Sekunden ermöglichen (Tancik et al., 2023). Instant NGP wurde 2022 offiziell von Nvidia veröffentlicht (Müller et al., 2022).

Nerfacto

Nerfacto benutzt auch das verbesserte Neuronale Netzwerk und den Hash-based Kodierer von Instant NGP, dazu kommt ein verbesserter Kamerapositions Algorithmus. Dieser hilft dabei eventuell falsch ausgerichtete Posen zu dem Input rückpropagieren (Backpropagation). Diese Information wird zur Optimierung und Verfeinerung der Posen verwendet, sodass wolkige Artefakte oder unscharfe Details verringert werden (Tancik et al., 2023).

3.6 Evaluierungstools der Daten

Die Evaluierung der 3D-Meshes wird in Meshlab vorgenommen (Ranzuglia et al., 2013). Meshlab ist eine Open-Source-Software zur Verarbeitung und Bearbeitung von 3D-Meshes. Es bietet eine Reihe von Werkzeugen zum Bearbeiten, Bereinigen, Heilen, Prüfen, Rendern, Texturieren und Konvertieren von Meshes. Außerdem bietet es eine Reihe an Funktionen zum Vergleichen von 3D-Meshes. Nach Skaliren und Alignment der Meshes in der Software können verschiedene Evaluationen vorgenommen. Für diesen Vergleich wird der Abstand aller Vertices im Mesh, bezogen auf das Referenz Mesh, berechnet. Grafisch lässt sich diese Evaluation auch darstellen.

4 Durchführung und Auswertung der Ergebnisse

4.1 Dauer der Rekonstruktionen

Zuerst wurden auf die Datensets der Structure-from-Motion Algorithmus angewendet. Bei Meshroom ist SfM der erste Automatisierungsschritt. Bei Metashape heißt dieser „Align Photos“. Für NeRF wurde COLMAP angewendet. Mit einer automatisierten Batch-Datei (bat.) lassen sich nach Angabe des Ziel-Dateipfads die Datensets unkompliziert automatisiert verarbeiten.

Daraufhin folgt das Erstellen des Meshes aus der Point Cloud. Bei NeRFs ist dieser Schritt das Trainieren der Volumetrischen Darstellung. Als nächstes folgt die Texturierung und zuletzt das Exportieren in ein Mesh. Alle Meshes wurden als obj. Datei exportiert, die Texturen als jpg. Datei.

Abbildung 4.1 zeigt die Dauer der Rekonstruktionen in der respektiven Software.

Auffallend ist zunächst die unverhältnismäßig lange Dauer des SfM Algorithmus in Meshroom. Bei der Rekonstruktion in Meshroom konnte beobachtet werden, dass die GPU-Auslastung sehr niedrig war (10%-25%), die CPU-Auslastung wiederum maximal. Dies war konträr zu den anderen beiden Programmen. Hier war die GPU stets maximal ausgelastet. Es ist fraglich, ob dies der Grund dafür ist. Die längeren Export Zeiten der Meshes bei Nerfacto sind nachvollziehbar dadurch, dass die Oberfläche der Volumetrischen Darstellung rekonstruiert werden muss bevor das Mesh exportiert werden kann. In Nerfacto und Meshroom lässt sich eine deutliche Korrelation der Dauer mit der Gesamtzahl der Bilder in den respektiven Datensätzen ziehen. Beide Methoden brauchen am längsten bei dem größten Datensatz. Die beiden kleineren, von der Gesamtzahl der Bilder ähnlichen, Datensätze benötigen ungefähr gleich lang, mit 16 und 36 Minuten Diskrepanz. Metashape liefert im Datensatz 1 und 2 die schnellste Rekonstruktion mit durchschnittlich 75 Minuten Abstand für D1 und 59 Minuten Abstand für D2. Im Datensatz 3 dauert Metashape jedoch 29 Minuten länger als Nerfacto.

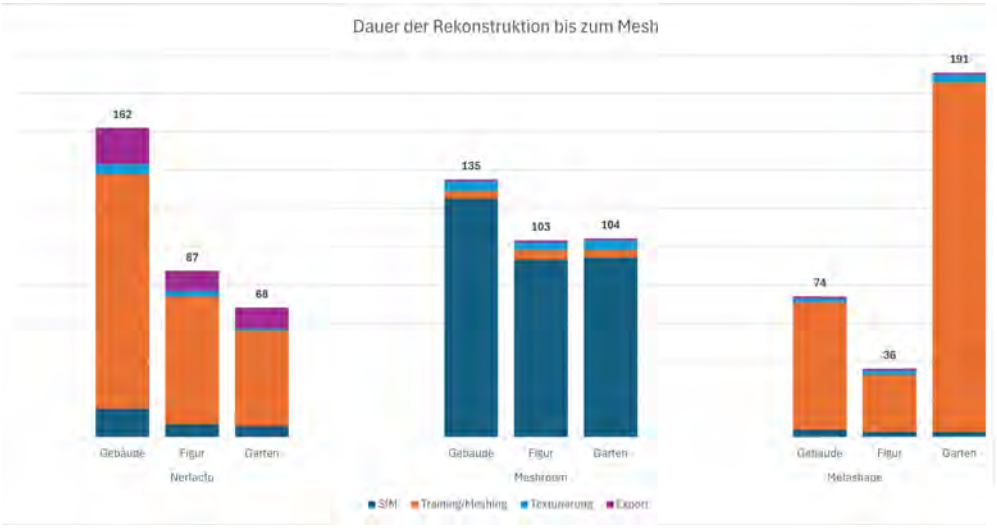


Tabelle 4.1: Vergleichsergebnisse der Dauer der Rekonstruktionen vom Input des Datensatzes bis zum fertigen Mesh

4.2 Quantitativer Vergleich

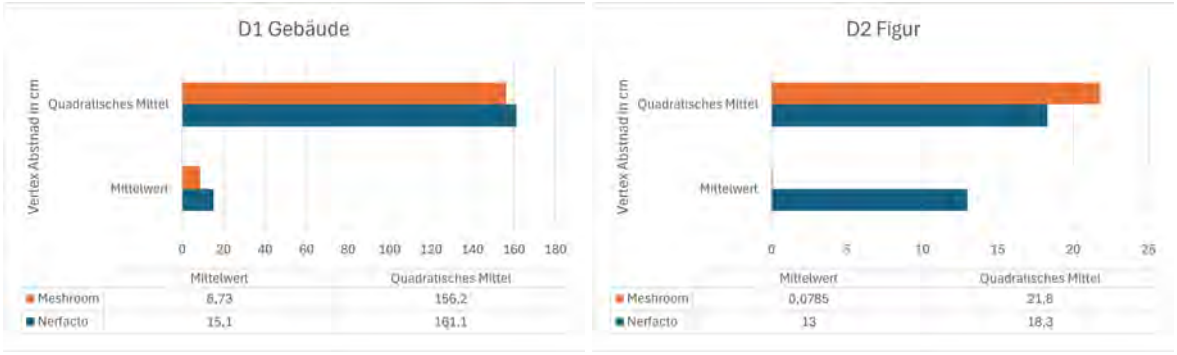
Trotz der Limitierung des Vertex und Face Counts sind die Meshes mit abweichendem Count exportiert worden, siehe 4.2.

Anschließend wurden die Meshes skaliert. Die exportierten Meshes haben unterschiedliche Skalierungen und Rotationen, die in den respektiven Programmen auftreten. Das Measuring Tool von Meshlab wird benutzt um eine Länge zu messen. Bei dem ersten Datensatz (Gebäude) wurde die Treppenstufenbreite gemessen. Die reale Breite der Stufe wird durch die gemessene Länge der Stufe des Meshes aus Meshlab geteilt. Daraus ergibt sich der Skalierungsfaktor, der daraufhin auf das Mesh angewendet wird. Nach Skalierung folgt das Alignment der Meshes. Mit dem Point Based Gluing Tool werden die Meshes durch Auswählen von gleichen Merkmalen übereinander gelegt.

Das Metashape Meshes wird in allen drei Datensätzen, aufgrund ihrer Qualität, stets als Referenzmesh des Vergleichs verwendet **??**. Nach Übereinanderlegen der Meshes lässt sich die Vorzeichenbehaftete Abstandsfunktion auf das Mesh im Verhältnis zum Referenzmesh berechnen. Die Vorzeichenbehaftete Abstandsfunktion (Signed Distance Funktion) berechnet die Abstände, in diesem Fall den Abstand der übereinanderliegenden Vertices der beiden Meshes. Das Vorzeichen gibt an, ob sich der Vertex innerhalb oder außerhalb des Referenzmesh befindet. Durch die bereits durchgeführte Skalierung lassen sich diese Distanzen metrisch wiedergeben. Eine Ausnahme ist Datensatz 3. Da dieser aus dem MipNeRF 360 Datensatz stammt, gibt es keine Messdaten und somit keine genauen metrischen Angaben zum Abstand der Vertices. Näherungsweise wird der Durchmesser des Tischs mit 150cm geschätzt.

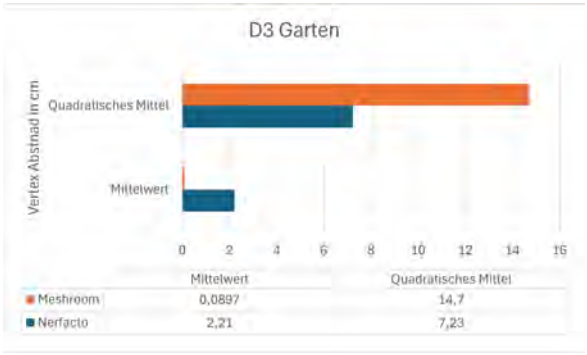
D1/D2/D3	Vertices	Faces	Durchschnittliche Abweichung der Vertices aller Meshes	Durchschnittliche Abweichung der Faces aller Meshes
Nerfacto	52404	99995		
Meshroom	38645	76786		
Metashape	45375	88445		
Nerfacto	38645	76786	9,50%	8,53%
Meshroom	53505	106690	4750	8532
Metashape	51961	100240		
Nerfacto	52443	99995		
Meshroom	50572	100863		
Metashape	45467	88995		

Tabelle 4.2: Topologische Messungen der Meshes



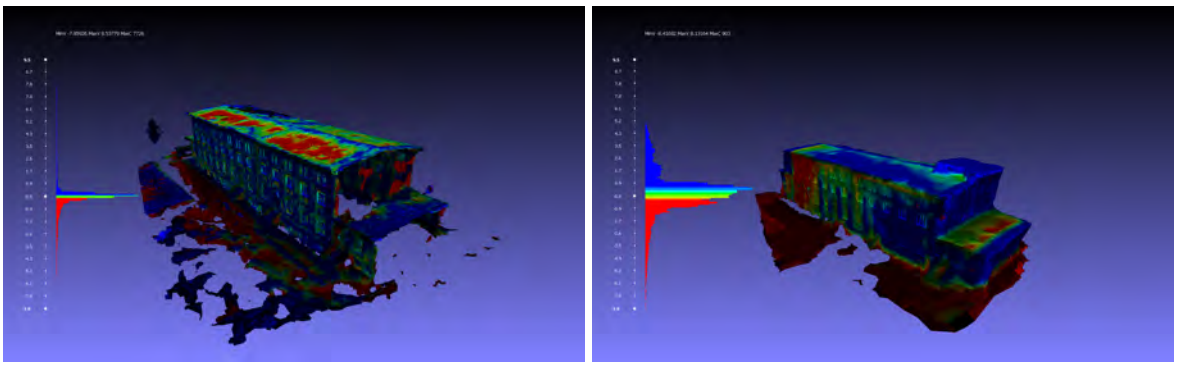
(a) D1 Gebäude

(b) D2 Figur



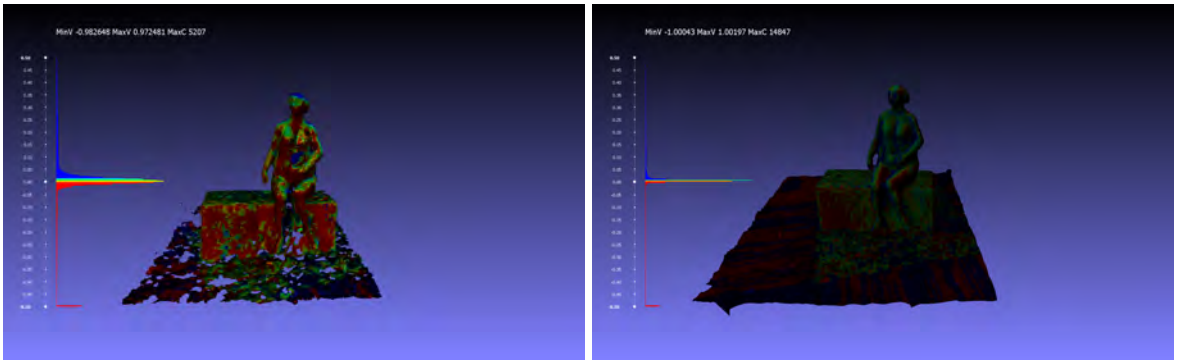
(c) D3 Garten

Abbildung 4.1: Mittelwerte und Quadratisches Mittel der Vertex Abstände zum Referenzmesh



(a) Nerfacto (b) Meshroom

Abbildung 4.2: Gebäude, Vertex Abstands Histogramm auf RBG-Werte gemappt

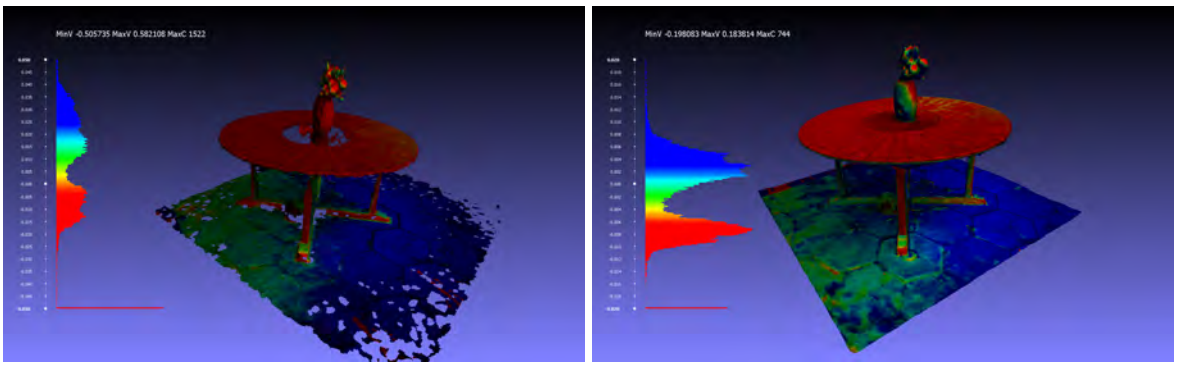


(a) Nerfacto (b) Meshroom

Abbildung 4.3: Figur, Vertex Abstands Histogramm auf RBG-Werte gemappt

Die Abstände jeder einzelnen Vertices lassen sich auch grafisch darstellen. Abbildung 4.2, 4.3 und 4.4 zeigen die jeweiligen Histogramme der Datensätze.

Generell lässt sich ein Zusammenhang zwischen dem errechneten Mittelwert der Vertexabstände und der generellen Qualität der Meshes ziehen. Beispielsweise zeigt das Nerf Mesh des Gebäudes Vertexabstände auf dem Dach von bis zu 90cm auf. Das Meshroom Mesh wiederum weitaus immensere strukturelle Fehler in der Rekonstruktion auf, welche schon im SfM-Algorithmus entstanden sind. Die Rekonstruktion zeigt mehr Vertices an, die relativ zum Referenzmesh eine fehlerhafte Position haben. Bei der Figur zeigt Nerfacto hauptäschlich fehlerhafte Vertexabstände im Bereich von 2-5cm auf. Das Meshroom Mesh zeigt deutlich weniger fehlerhafte Vertexabstände an, welche maximal in einem Abstand von 3cm zum Referenzmesh positioniert sind. Dieser Zusammenhang ist auch bei dem Datensatz Garten zu erkennen.



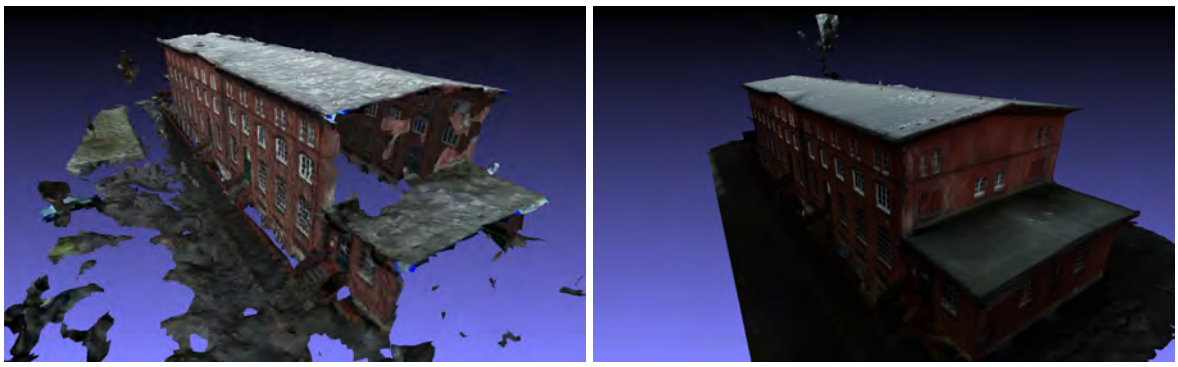
(a) Nerfacto (b) Meshroom

Abbildung 4.4: Garten, Vertex Abstands Histogramm auf RBG-Werte gemappt

4.3 Oberflächen

Die rekonstruierten Oberflächen von Meshroom und Nerfacto weisen in allen drei Datensätzen Löcher und unebene Oberflächen auf. Weitesgehend gerade Oberflächen, wie das Dach im Datensatz 1 und die gräserne Oberfläche des Bodens in Datensatz 2, sind unvollständig und strukturell falsch rekonstruiert. Bei gleichen Vertex und Face-Counts kann Metashape die korrespondierenden Oberflächen wesentlich akkurater rekonstruieren. Beispielsweise bleiben die Fensterrahmen des Gebäudes strukturell konsistent, während sich die Fläche der Rahmen im Nerfacto mesh von Fenster zu Fenster unterscheidet. Die Gebäudeecken sind bei Metashape markant zu erkennen, während Nerfacto bauchige Übergänge kreiert.

Alle Rekonstruktionsverfahren können im Datensatz 3 die Unterseite des Tisches nur ungenau rekonstruieren. Es sollter jedoch zur Sprache gebracht werde, dass in diesem Datensatz keine Kameraposition die Unterseite aufnimmt. So vielversprechend die Premisse der NeRFs ist, solange gar keine Informationen über einen bestimmten Abschnitt in den Bildern zu sehen sind, können NeRFs diesen Abschnitt auch nicht richtig rekonstruiert werden. Dieses Problem liegt auch bei den Photogrammetrie Methoden vor. Meshroom und Metashape rekonstruieren den Tisch strukturell genau, während Nerfacto diverse Löcher aufweist. Auffällig ist hier das der spiegelnde Kreis aus Metall inmitten des Tisches nicht dargestellt wurde. Im trainierten NeRF dieses Datensatzes zeigt der Kreis bei rotieren um den Tisch die bereits erwähnten Blickrichtungsabhängigen Spiegelungen auf. Die Spiegelungen können nicht im Mesh dargestellt werden, trotzdem fehlt bei diesem Mesh der komplette metallische Kreis.



(a) Nerfacto (b) Metashape

Abbildung 4.5: Garten, Bild der Meshes aus Meshlab

4.4 Textur

Die Metashape Texturen weisen hochaufgelöste markante Details auf. In Abbildung 4.5b deutlich zu erkennen die Witterungsverhältnisse auf der Ziegelwand. Trotz der unebenen Oberfläche des Meshes ist die Textur des Nerfacto Meshes auch detailliert und deutet auf der Ziegelwand die gleichen Witterungsverhältnisse an. Auffallend bleiben bei den NeRF Meshes ungewollte Farbfragmente, in Abbildung 4.5b an der Kante des Daches zu sehen, welche markante Farben darstellen, die nicht in den Datensätzen zu sehen sind. Nerfacto weist im Datensatz Garten unscharfe Texturen auf. Die Holzpaneele des Tisches scheinen verschwommen zu sein und zeigen nicht die Details auf, welche Metashape wiederum deutlich darstellt. Trotz gleicher Auflösung bleiben die Texturen in Datensatz 2 Figur auch deutlich verschwommen.

Die Meshroom Texturen des Gebäudes sind sehr gut aufgelöst. Umrisse der Ziegelsteine sind in 4.6a klar erkennbar, Text ist fast lesbar und es sind farbliche Übergänge von Tür zu Fenster sichtbar. Im Kontrast dazu wirken die Metashape Texturen verschwommen und ungenau.

Bei der Figur sind die Metashape Texturen genauer. Die Meshroom Texturen stellen Details auf der Figuroberfläche und des Steins nicht detailliert dar. Noch ungenauer sind die Nerfacto Texturen, sie stellen nur die Farbe ohne etwaige Details dar.

4.5 Limitationen des Vergleichs

Der Vergleich der beiden Methoden ist von entscheidender Bedeutung für die Bewertung ihrer Leistungsfähigkeit und Genauigkeit. Dennoch gibt es einige Einschränkungen, die berücksichtigt werden müssen, um die Ergebnisse angemessen zu interpretieren. Eine wesentliche Limitation besteht darin, dass die exakte Anzahl der Vertices und Faces in einem

(a) Meshroom (b) Metashape

Abbildung 4.6: Vergleich Textur D1 Gebäude, Meshroom mit Metashape

rekonstruierten 3D-Mesh nur grob limitierbar sind und eine exakte Skalierung mit darauf folgendem Alignment fast unmöglich ist. Ideal wäre ein exakter Scan der Objekte mit LiDAR, um präzisere Ground Truth Datensätze zu erstellen. Besonders im Hinblick darauf, dass das Referenzmesh selber mit Photogrammetrie Software erstellt wurde. Allerdings ist ein LiDAR-Scan nicht immer möglich oder praktikabel.

In diesem Zusammenhang wäre es außerdem lohnenswert, in zukünftiger Forschung zu untersuchen, wie die Ergebnisse sich verändern, wenn die Limitierung der Mesh-Paramter erst nach Exportieren in maximaler Qualität angewandt werden. Hier muss jedoch berücksichtigt werden, dass das Erhöhen der Anzahl der Vertices und Faces in einem Mesh zusätzliche Rechenleistung erfordert und den Prozess intensiviert. Dies wiederum stellt eine Hürde dar, insbesondere für weniger erfahrene Anwender*innen, die möglicherweise nicht über die erforderlichen Kenntnisse oder Ressourcen verfügen, um diesen Prozess effektiv durchzuführen.

Selbst wenn potente NeRFs bereits zugänglich sind, besteht oft nicht die Möglichkeit, die Ergebnisse als Mesh zu exportieren, um sie mit Photogrammetrie Meshes zu vergleichen. Spannend ist hier der Blick auf NeRFMeshing (Rakotosaona et al., 2023). NeRFMeshing stellt eine Architektur vor, welche eine einfache 3D-Oberflächenrekonstruktion aus jeglichen NeRF Methoden ermöglicht. Nachdem das NeRF trainiert wurde, wird die volumetrische

3D-Darstellung in ein Signed Surface Approximation Network gespeist, das eine einfache Extraktion des 3D-Meshes ermöglicht. Das finale 3D-Mesh ist strukturell akkurat und kann in den gängigen 3D-Software Programmen in Echt-Zeit gerendert werden. NeRFMeshing stellt somit eine spannende Architektur vor, welche das Potenzial hat Photogrammetrie abzulösen. Ob hier der Vergleich mit Photogrammetrie bessere Ergebnisse liefert, bleibt jedoch zum aktuellen Zeitpunkt unerforscht. Nach Kontakt mit Michael Niemeyer, einem Co-Autor des Papers, wurde bestätigt, dass NeRFMeshing zurzeit noch proprietär ist und es keine Pläne für einen Open Source Release gibt.

5 Fazit und Implikationen für die Praxis

Zielsetzung der vorliegenden Arbeit war es, im Vergleich aufzuzeigen, ob NeRFs Photogrammetrie für die Asset Creation ablösen können, welche Vor- und Nachteile sie mit sich bringen, und in welcher Situation welche Technik am sinnvollsten sein kann. Konkludierend lässt sich festhalten, dass Photogrammetrie seit 2014 in der Games-Industrie vollständig integriert ist und die Prozesse im Laufe der Zeit automatisiert und verfeinert wurden, um möglichst fotorealistische Ergebnisse zu erzielen. Allerdings zeigen sich bei der Darstellung als 3D-Mesh nach wie vor Herausforderungen, insbesondere bei einheitlichen Oberflächen, reflektierenden und dünnen Objekten. In diesem Kontext bieten Neural Radiance Fields (NeRFs) aufgrund ihres grundlegend anderen Ansatzes Vorteile, insbesondere in Bezug auf die Darstellung verschiedener Lichteffekte wie Reflektionen und Transparenz. Die Vergleichbarkeit der beiden Verfahren wird durch die Verwendung identischer Datensätze sowie gemeinsamer zugrunde liegender Informationen zu Kamera- und Objektparametern erleichtert. Obwohl NeRFs vielversprechende Ergebnisse liefern und eine Alternative zur traditionellen Photogrammetrie darstellen, bestehen weiterhin Hindernisse und Probleme.

Wie der Vergleich gezeigt hat, bringt die klassische Photogrammetrie qualitativ und quantitativ bessere Ergebnisse. Auch bei limierten Mesh Parameter sind die NeRF Meshes strukturell fehlerhaft und bieten wenig bis gar keine Grundlage für weitere Retopologie. Die Texturen, verglichen mit den Photogrammetrie Texturen, sind unscharf. Die beiden Photogrammetrie Programme liefern jedoch unterschiedlich gute Ergebnisse. Die Meshroom Meshes weisen detaillierte Texturen auf, teilweise besser als Metashape, sind aber strukturell fehlerbehaftet. Die Metashape Meshes liefern die beste Basis für den weiteren Verlauf in der Asset-Pipeline.

Die Dauer der Rekonstruktionen war in allen Programmen unterschiedlich. Die NeRF Rekonstruktion mit Nerfacto erforderte am meisten Zeit, während Metashape am wenigsten Zeit in Anspruch nahm. Der statistische Ausreißer bei Metashape mit Datensatz 3 bleibt ungeklärt. Meshroom brilliert mit der simplen Oberfläche und der iterativen Rekonstruktion. Die Node-basierte Oberfläche bleibt stets übersichtlich, jedoch weist Meshroom ungewollte CPU-Auslastung auf, was eventuell zu einer längeren Dauer der Rekonstruktion führt. Der Metashape Workflow war zunächst kompliziert gestaltet, da keine klare iterative Struktur für die Asset-Creation vorgegeben wird. Nach Testen der Software erwies sie sich als übersichtlich und effizient. Die Python Installation von Nerfstudio erwies sich als ermüdender Prozess,

da neben der Grundinstallation noch andere Abhängigkeiten installiert und eingestellt werden mussten. Die Installationen von verschiedenen Plattformen und Abhängigkeiten (C+, Cmake, MSYS2 etc.) und Hinzufügen von Systemvariablen bieten für technisch unversierte Anfänger*innen eine mögliche Hürde. Letztendlich bietet Nerfstudio eine solide Umgebung, um sich mit der neuartigen Technik der NeRFs vertraut zu machen.

Die Meshes der NeRFs zeigen noch keine reale Alternative zu der etablierten Photogrammetrie. Die Photogrammetrie Programme sind ausgereift und ermöglichen weitestgehend problemloses Arbeiten für Photogrammetrie Interessierte und für die Beschäftigten in der VFX- und Gamesbranche. Schlussendlich muss erneut erwähnt werden, dass NeRFs nicht primär auf das Exportieren als Mesh spezialisiert sind, sondern auf die Synthese von neuartigen Blickrichtungen optimiert sind.

Literaturverzeichnis

Association, A. (n. d.). AliceVision Association. https://alicevision.org/association/

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., & Srinivasan, P. P. (2021, August). *Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields* (Techn. Ber.) (arXiv:2103.13415 [cs] type: article). arXiv. https://doi.org/10.48550/arXiv.2103.13415

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., & Hedman, P. (2022). Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *CVPR*.

Chan, S. C. (2014). Plenoptic Function. In K. Ikeuchi (Hrsg.). Springer US. https://doi.org/10.1007/978-0-387-31439-6_7

Crafting Environments for Battlefield V with DICE LA. (2019, Juli). Verfügbar 8. April 2024 unter https://www.youtube.com/watch?v=AfLrkL6PoB0

Designing of different kernels in Machine Learning and Deep Learning. (n. d.). Verfügbar 7. März 2024 unter https://www.turing.com/kb/designing-of-different-kernels-in-machine-learning-deep-learning

DJI. (n. d.). DJI Mini 3 Pro - Technische Daten - DJI. Verfügbar 28. Februar 2024 unter https://www.dji.com/de/mini-3-pro/specs

DJI Drone Sensor Size Comparison Page. (2018, Dezember). Verfügbar 1. April 2024 unter https://www.djzphoto.com/blog/2018/12/5/dji-drone-quick-specs-amp-comparison-page

EOS. (2021, September). Panchromatic Imagery And Its Band Combinations In Use. Verfügbar 28. Februar 2024 unter https://eos.com/make-an-analysis/panchromatic/

et al., M. T. (2022). NeRF Tutorial ECCV 2022. Article. Verfügbar 28. Februar 2024 unter https://sites.google.com/berkeley.edu/nerf-tutorial/home

Finstad, M. (2024, März). mifi/lossless-cut [original-date: 2016-10-30T10:49:56Z]. Verfügbar 25. März 2024 unter https://github.com/mifi/lossless-cut

Flyover. (2024, Februar). Flyover (Apple Maps) [Page Version ID: 1210507980]. Verfügbar 28. Februar 2024 unter https://en.wikipedia.org/w/index.php?title=Flyover_(Apple_Maps)&oldid=1210507980

Foster, S. (2014). *Integrating 3D Modeling, Photogrammetry and Design* (D. Halbstein, Hrsg.; 1st ed.) [Description based on publisher supplied metadata and other sources.]. Springer London, Limited.

Gentili, G., Simonutti, L., & Struppa, D. C. (2022, Oktober). *The Mathematics of Painting: the Birth of Projective Geometry in the Italian Renaissance* (Techn. Ber.) (arXiv:2210.13295 [math] type: article). arXiv. https://doi.org/10.48550/arXiv.2210.13295

Grechnyev, O. (2023). What Is NeRF (Neural Radiance Fields)? *IT-Jim Blog*. https://www.it-jim.com/blog/nerf-in-2023-theory-and-practice/

Hamilton, A. S. (2016, August). Star Wars: Battlefront and the Art of Photogrammetry. Verfügbar 28. Februar 2024 unter https://www.youtube.com/watch?v=U_WaqCBp9zo

Hodgson, D. (2019, Juni). Initial Intel: How Photogrammetry is Helping to Shape Call of Duty: Modern Warfare into a new high watermark for graphics in gaming. Verfügbar 7. März 2024 unter https://blog.activision.com/call-of-duty/2019-06/Initial-Intel-How-Photogrammetry-is-Helping-to-Shape-Call-of-Duty-Modern-Warfare-into-a-new-high-watermark-for-graphics-in-gaming

Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson Surface Reconstruction. In A. Sheffer & K. Polthier (Hrsg.), *Symposium on Geometry Processing*. The Eurographics Association. https://doi.org/10.2312/SGP/SGP06/061-070

Kinect. (2024, Februar). Kinect [Page Version ID: 1201920600]. Verfügbar 28. Februar 2024 unter https://en.wikipedia.org/w/index.php?title=Kinect&oldid=1201920600

Kumar, P. (2022, Januar). What is Near-Infrared Imaging and how do NIR cameras work? Verfügbar 28. Februar 2024 unter https://www.e-consystems.com/blog/camera/technology/what-is-nir-imaging-and-how-do-nir-cameras-work/

Linder, W. (2016). *Digital Photogrammetry: A Practical Course* (4. 4th ed. 2016). Springer Berlin Heidelberg.

Luhmann, T. (2023). *Close-Range Photogrammetry and 3D Imaging* (S. Robson, S. Kyle & J. Böhm, Hrsg.; 4th edition). De Gruyter.

Mildenhall, e. a., Ben. (2020, August). *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis* (Techn. Ber.) (arXiv:2003.08934 [cs] type: article). arXiv. https://doi.org/10.48550/arXiv.2003.08934
Comment: ECCV 2020 (oral). Project page with videos and code: http://tancik.com/-nerf.

Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, *41*(4), 102:1–102:15. https://doi.org/10.1145/3528223.3530127

Pepe, M., Alfio, V., & Costantino, D. (2023). Assessment of 3D Model for Photogrammetric Purposes Using AI Tools Based on NeRF Algorithm. *Heritage*, *6*, 5719–5732. https://doi.org/10.3390/heritage6080301

Poznanski, A. (2014). Visual Revolution of The Vanishing of Ethan Carter. https://www.theastronauts.com/2014/03/visual-revolution-vanishing-ethan-carter/

Quixel | 3D world-building made easy. (n. d.). Verfügbar 19. März 2024 unter https://quixel.com/

Rakotosaona, M.-J., Manhardt, F., Arroyo, D. M., Niemeyer, M., Kundu, A., & Tombari, F. (2023, März). *NeRFMeshing: Distilling Neural Radiance Fields into Geometrically-Accurate 3D Meshes* (Techn. Ber.) (arXiv:2303.09431 [cs] type: article). arXiv. https://doi.org/10.48550/arXiv.2303.09431

Ranzuglia, G., Callieri, M., Dellepiane, M., Cignoni, P., & Scopigno, R. (2013). MeshLab as a complete tool for the integration of photos and color with high resolution 3D geometry data. *CAA 2012 Conference Proceedings*, 406–416. http://vcg.isti.cnr.it/Publications/2013/RCDCS13

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., & Qin, R. (2023). A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sensing*, *15*(14), 3585. https://doi.org/10.3390/rs15143585

Robineau, A. (2021, Oktober). An overview of Differentiable Rendering. https://blog.qarnot.com/article/an-overview-of-differentiable-rendering

Schönberger, J. L., & Frahm, J.-M. (2016). Structure-from-Motion Revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Statham, N., Jacob, J., & Fridenfalk, M. (2020). Photogrammetry for Game Environments 2014-2019: What Happened Since The Vanishing of Ethan Carter.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., & Ng, R. (2020, Juni). *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains* (Techn. Ber.) (arXiv:2006.10739 [cs] type: article). arXiv. https://doi.org/10.48550/arXiv.2006.10739
Comment: Project page: https://people.eecs.berkeley.edu/~bmild/fourfeat/.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., & Kanazawa, A. (2023). Nerfstudio: A Modular Framework for Neural Radiance Field Development. *ACM SIGGRAPH 2023 Conference Proceedings*.

Tkalych, D. (2024, Januar). Does Polygon Count Matter in 3D Modeling for Game Assets? Verfügbar 1. April 2024 unter https://3d-ace.com/blog/polygon-count-in-3d-modeling-for-game-assets/

und Maschine Deutschland GmbH, M. (n. d.). Leica RTC360 3D-Laserscanner. https://eshop.mum.de/p-2487-leica-rtc360-3d-laserscanner-paket.aspx

Anhang

Datensätze, Projektdateien und gesammelte Daten befinden sich beigefügt auf den Datenträgern.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel

Vergleich von Neural Radiance Fields und Photogrammetrie für 3D Asset-Creation

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Hamburg, 10. April 2024