Evaluierung der photogrammetrischen Rekonstruktion für numerische Berechnung von HRTFs

Autorin: Katharina Pollack, BSc.
Betreuer: Doz. Dr. Piotr Majdak
(Institut für Schallforschung)
DI Ph.D. Matthias Frank
(Institut für Elektronische Musik und Akustik)

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

iem

Inhaltsverzeichnis

1 Abstract

Head-related transfer functions (HRTFs) are essential to many applications such as spatialisation of audio signals and virtual acoustics. Their acoustic measurement can be cumbersome and time consuming. But what if the acoustic measurement equipment is not available? Photogrammetry can be a potential answer to that question: a pipeline of algorithms constructs a three-dimensional mesh based on photos of head and pinnae. The mesh is then used for numerical calculation of the HRTFs. Within the scope of this seminar paper that was part of an internship at the Acoustics Research Institute in Vienna, setup of the photo session and parameters of photogrammetry algorithms are investigated, aiming at calculating HRTFs.

2 Zusammenfassung

Außenohrübertragungsfunktionen (in weiterer Folge HRTFs) sind essenziell für etliche Anwendungen wie der Spatialisierung von Audiosignalen und Virtual Reality geworden. Messen kann man sie mit dem heutigen Stand der Technik in mittlerweile nurmehr 16 Minuten; doch was ist, wenn es kein geeignetes Messsystem in unmittelbarer Nähe gibt? Diese Arbeit, die im Zuge eines Praktikums am Institut für Schallforschung in Wien entstand, beschäftigt sich mit einer möglichen Antwort: der Fotogrammetrie. Ein Algorithmus, der die Eigenschaften von Fotos des Kopfes bzw. der Pinnae erfasst, konstruiert dreidimensionale Meshes, woraus die HRTFs numerisch ermittelt werden können. Im Speziellen werden im Rahmen der Projektarbeit Kriterien für die Erfassung der Geometrie ermittelt, damit die berechneten HRTFs näherungsweise den gemessenen entsprechen.

3

Selbstständigkeits-erklärung

Ich, Katharina Pollack, geboren am 04.11.1992, erkläre hiermit ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel "Evaluierung der Fotogrammetrischen Rekonstruktion für numerische Berechnung von HRTFs" um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Sämtliche in der oben genannten Arbeit verwendeten fremden Quellen habe ich als solche kenntlich gemacht. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen beziehungsweise unverändert übernommenen Tabellen, Grafiken u.Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen beziehungsweise von mir abgewandelten Tabellen, Grafiken u.Ä. anderer Autor*innen (Paraphrasen) die Quelle angegeben habe.

Ort, Datum: ..... Unterschrift: .....

4

Außenohrübertragungsfunktionen

4.1 Definition

Die Außenohrübertragungsfunktion (engl. Head Related Transfer Function) fasst mehrere Eigenschaften des Außenohrs zusammen: frequenz- und richtungsabhängige Filterung liefern durch Anregung verschiedener Resonanzen von Torso, Kopf und Pinna in der Horizontal- und Sagittalebene unterschiedliche räumliche Auflösung.

Trifft aus einer mit Azimuth φ und Elevation θ eindeutig definierten Richtung Schall auf ein Subjekt, so beschreiben die interauralen Zeitdifferenzen (engl. Interaural Time Differences) die Laufzeitunterschiede, die eine Schallwelle vom der Schallquelle näheren zum entfernteren Ohr erfährt und die interauralen Pegeldifferenzen (engl. Interaural Level Differences) die Pegelunterschiede. Da ab einer Frequenz von ca. 1600 Hz eine Schallwellenlänge dem Ohrenabstand entspricht und somit die interauralen Zeitdifferenzen überhalb dieser Grenzfrequenz nicht mehr eindeutig sind, werden Pegeldifferenzen deutlicher wahrgenommen.

Es gibt durchaus Punkte im Raum, für die ITD und ILD gleich groß sind. Diese werden zusammengefasst unter dem Begriff des “Cone of Confusion“, einem Doppelkegel mit gegeneinander gerichteten Spitzen (ähnlich einer Sanduhr), dessen Höhenachse in die Gerade zwischen beiden Ohren gelegt werden kann. Alle Schallquellen auf dem Umkreis einer Kegelscheibe weisen dieselbe ITD und ILD auf, was durch Kopfdrehung ausgeglichen werden kann, da die Quellenposition aus dem Kegelschreibenrand tritt und somit die beiden Differenzenparameter erneut geändert werden. Zusätzlich bieten die Augen eine zusätzliche Lokalisationshilfe[1].

Die räumliche Auflösung wird durch die sogenannte Just Noticeable Difference (der gerade noch wahrnehmbare Unterschied) beschrieben, die - in Abhängigkeit der Einfallsrichtung - durchaus extrem unterschiedlich sein kann. So beträgt die JND zum Beispiel direkt frontal vor der Person (0° Azimuth und 0° Elevation) ca. 1° und ganz links bzw. rechts $5° - 10°$. In der Elevation ist die JND stark signalabhängig, beträgt sie bei weißem Rauschen nur 4° kann sie bei durchgehender Sprache einer unbekannten Person bis zu 17° ausmachen.

4.2 Messung und Simulation

Als Beispiel für einen Aufbau zur Messung von Außenohrübertragungsfunktionen ist in Abbildung 4.1 der am Institut für Schallforschung in Wien verwendete zu sehen. Ein aufwändiges Konstrukt aus einem Ring, auf dem Lautsprecher von $-20°$ bis $+200°$ Elevation (bezogen auf die Horizontalebene durch die Pinnae) im kleinstmöglichen Abstand zueinander Platz finden, siehe Abbildung [1]. In der Mitte befindet sich ein Drehteller, auf dem der/die Proband*in platziert wird, sodass der Gehörgang in der Medianebene (bei 0° Elevation) liegt. Messmikrofone werden am Ende der Gehörgänge platziert, das Messsignal besteht aus ineinander verschachtelten Sinussweeps, die durch Entfaltung und Fensterung nach der Messung in Impulsantworten zerlegt werden.

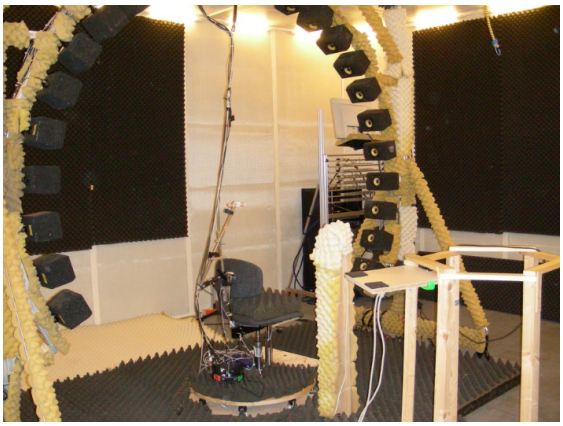

Abbildung 4.1: Aufbau der Messanlage am ISF.

In der Simulation muss das Schallfeld für jede Lautsprecherposition um ein 3D-Büsten-

modell berechnet werden. Hierzu greift man auf die reziproke Messmethode der HRTFs zurück, bei der nach dem Helmholtz'schen Prinzip der Reziprozität von Schallfeldern Quelle und Senke vertauscht werden. Das Prinzip besagt, dass in einem beliebig komplexen linearen und zeitinvarianten Schallfeld der von einer Schallquelle im Punkt B verursachte Schalldruck in Punkt A der gleiche ist, der an Punkt B gemessen würde, wenn die Quelle in Punkt A läge. Damit müssen im Gegensatz zur direkten Methode nurmehr 2 Lautsprecherpositionen - die in den Gehörgängen platzierten Mikrofone - berechnet werden, was die Berechnungsdauer der Simulation von mehreren Wochen auf "nur" ein paar Stunden auf dem Linux Cluster des Instituts (im Folgenden "Denker" genannt) dezimiert. Dieser Cluster besteht aus 8 Maschinen mit Intel i7-3820 Prozessoren mit jeweils 3.6 GHz und 64 GB RAM.

5

Photogrammetrie

5.1 Definition

Photogrammetrie (aus dem Griechischen "Bildmessung") beschreibt den Vorgang, mittels Kombination vieler zweidimensionaler Fotografien aus verschiedenen Winkeln eines dreidimensionalen Objekts ein dreidimensionales Modell zu erstellen[2]. Dabei wird zwischen zwei Anwendungsbereichen unterschieden: der *LuftbildPhotogrammetrie* und der *NahbereichsPhotogrammetrie*. Ersterer beschreibt in einigen hundert Metern bis Kilometern Höhe fliegende Flugzeugen oder Satelliten, deren Fotomaterial zum Beispiel zur Erstellung ortografischer Karten oder dem Erkennen versteinerter Spuren aus der Urzeit verwendet wird. Die NahbereichsFotogrammetrie umfasst die Objektmodellierung von einigen wenigen Zentimetern bis ca. 100 Metern und wird zum Beispiel in der Archäologie verwendet, um Artefakte zu modellieren; oder in der Spiele- und Filmindustrie bei der Animation von Charakteren und Objekten.

5.2 Pinnamodellierung

Diese Projektarbeit befasst sich mit der Modellierung der Pinna mittels Fotogrammetrischer Rekonstruktion. Zunächst werfen wir einen Blick auf das zu modellierende Objekt - das Außenohr (lat. Pinna):

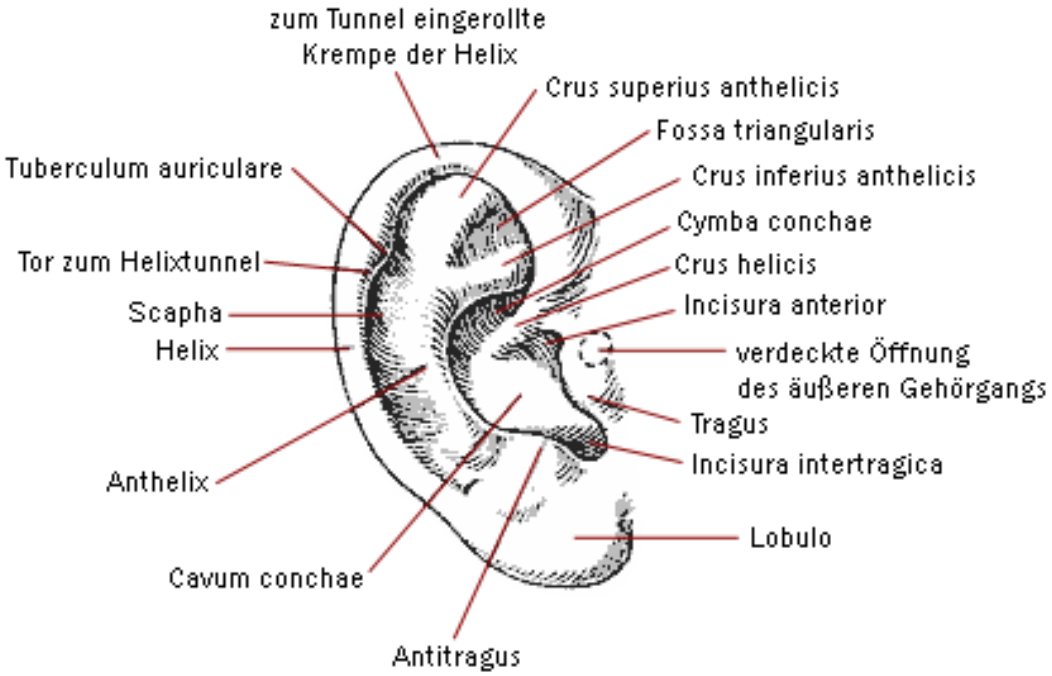

Abbildung 5.1: Schematische Darstellung des menschlichen Ohres.

Schon in der schematischen Darstellung[1] ist deutlich zu sehen, dass Bereiche vorkommen, die Schatten werfen und andere mit einer Kamera sehr unwahrscheinlich zu erfassen sind, wie zum Beispiel die Krempe der Helix und die Tiefe der Cavum Conchae (auch Concha).
Um auch diese Bereiche möglichst gut zu erfassen, wurde die Idee verfolgt, aus drei verschiedenen Höhen bzw. Einfallswinkeln zu fotografieren ($\pm 45°$ und $0°$ Elevation). Auf jeder Elevationsebene wurden ca. 10-15 Fotos im Abstand von $10 - 15°$ aufgenommen und dabei ein Bereich im Azimuth von ca. $150°$ abgebildet.

[1] https://parlograph.files.wordpress.com/2015/06/schema_ohr_mit_beschriftung1.jpg

5.3 Rekonstruktion und Simulation

Im nächsten Schritt werden die Fotos auf “Features“ überprüft, indem jedes Foto mit jedem verglichen wird. Je mehr Fotos gemacht werden, desto höher ist der Rechenaufwand. Aus diesen gefundenen Features kann dann eine Punktwolke erstellt werden, die die Qualität des 3D-Modells (auch “Mesh“) bestimmt. Zwischendurch können Punkte aus der Wolke, die nur in wenigen Fotos eindeutig gesetzt wurden, aus dem weiteren Berechnungsvorgang ausgeschlossen werden, da sie das Mesh verrauschen. Abschließend werden die Punktdaten auf dreieckige Flächen übertragen, in der Geostatistik wird dieser Vorgang “Regionalisierung“ genannt. In den untersuchten Programmen wird dieser letzte Schritt als “Interpolation“ bezeichnet.

Die Weiterverarbeitung des Meshes erfolgt in einer 3D-Grafiksoftware wie zum Beispiel Meshlab oder Blender, in der dann eventuell vorhandene Löcher gestopft (Verbinden von Punkten und Definieren von dreieckigen Flächen) oder Korrekturen vorgenommen werden können. Wichtig dabei ist, dass die Büste eine geschlossene Oberfläche darstellt, da sonst Randeffekte auftreten können. Des Weiteren muss händisch eine Ohrkanalkorrektur vorgenommen, also das Mikrofon am Eingang des Gehörganges platziert werden. Mit der .ply-/.obj- oder .stl-Datei (übliche 3D Objekt-Formate) kann dann die Berechnung *mesh2hrtf* in der Kommandozeile gestartet werden, die die HRTFs mittels reziproker Methode berechnet[3].

Es werden zwei mögliche Software-Pipelines zur Fotogrammetrischen Realisierung von Fotos zu Mesh betrachtet: Multi View Environment und Agisoft.

6 Multi View Environment

6.1 Beschreibung

Multi-View Environment beschreibt einen Algorithmus, der auf der TU Darmstadt entwickelt wurde und als OpenSource Projekt verfügbar ist [4][5]. Dabei werden Features mittels "Structure from Motion" und "Depth Maps Reconstruction"[6] extrahiert und zu einer Punktewolke mit Tiefeninformationen zusammengefügt. Durch Interpolation mittels Floating Scale Surface Reconstruction[7][8] können diese Punkte verdichtet und zu dreieckigen Flächen verbunden werden und ergeben somit ein dreidimensionales *Mesh* [9]-[10].

6.2 Pipeline der Unterprogramme

MVE besteht aus einzelnen Programmen, die nacheinander aufgerufen werden. Um eine Berechnung des gesamten Prozedere lückenlos durchzuführen zu können, wurden die

Programme mittels Shellskript hintereinander ausgeführt. Die folgenden Unterkapitel beschreiben die einzelnen Programme und geben Einblick in die verwendeten Parametersets. Der Vollständigkeit halber sei an dieser Stelle erwähnt, dass bei jeder der Funktionen ein Aufruf ohne Parameter die verfügbaren Optionen anzeigt.
Die Berechnungsdauer auf den Denkern belief sich bei der Durchführung der beschriebenen Experimente 2016 auf einige Stunden. 2019 wurde erneut mit einem großen Fotoset von 340 Fotos gerechnet, auf einem Laptop mit Intel i7-8550U bei 1.8 GHz und 16 GB RAM betrug die Rechenzeit in etwa 12 Stunden.

makescene

Die erste Funktion erzeugt zur Weiterverarbeitung nötige "MVE-Scenes" aus Fotos oder aus dem Output anderer Structure from Motion-Software (z.B. Noah's Bundler, Fotosynther, Visual SfM, Open MVG) und erstellt einen Ordner *views* mit .mve Dateien.

Im Skript werden folgende Optionen verwendet:

`-i` lädt nur Bilder ohne Kamerainformationen aus dem Inputverzeichnis

sfmrecon

Führt Structure from Motion durch und rekonstruiert dabei die einzelnen Kamerapositionen, Ausrichtungen, Brennweiten aller Fotos, verändert dabei die .mve Dateien und gibt als Ergebnis eine spärliche Punktewolke zurück. Diese Funktion wird ohne zusätzliche Parametersets aufgerufen.

dmrecon

Diese Funktion sorgt für die Erzeugung der sogenannten "depth map"[6], ein Bild, das Informationen über die Oberflächenabstände eines Objekts aus einem Blickpunkt enthält. Sie verändert ebenso die .mve Dateien und erzeugt die Datei *synth_0.out*

Im Skript werden folgende Optionen verwendet:

`-s2`	setzt Scalewert auf 2, setzt Auflösung zur Berechnung herunter 0 wäre original 2 ist laut Autoren der Software oft sinnvoll[4]
`--progress='silent'`	unterdrückt unnötig detaillierte Fortschrittsinformationen
`--nocolorscale`	unterdrückt verbesserte Farbdarstellung (weil die Farbinformation später ohnehin gelöscht wird)

scene2pset

Nun kann aus der depth map eine dichte Punktewolke erzeugt werden. Das Ergebnis ist in der Datei *pset-L2.ply* festgehalten.

Im Skript werden folgende Optionen verwendet:

`-F2` in Übereinstimmung mit s2 von oben

fssrecon

Mittels Floating Scale Surface Reconstruction[7][8] werden Skalierungsinformationen zur Tiefenstaffelung aus der Punktwolke extrahiert und das Mesh kostruiert. Diese Technik kommt aus dem Computergrafik-Bereich und beschreibt den Prozess, die Punkte eines realen Objekts in einem realen Raum in Punkte eines virtuellen Objekts in einem virtuellen Raum umzuwandeln. Das ist keine triviale Prozedur, da Features aus der depht map eine finite Oberfläche und nicht nur einzelne Punkte (infinitesimal kleine Flächen) repräsentieren. Die diesen finiten Oberflächen innewohnenden Skalierungsinformationen sind maßgeblich beteiligt an der Qualität der Rekonstruktion. Die Funktion wird ohne zusätzliche Parametersets aufgerufen; Das Ergebnis der Berechnung ist das Mesh *surface-L2.ply*

meshclean

Abschließend wird das Mesh noch "aufgeräumt". Das bedeutet, dass Punkte mit geringem *confidence*-Wert - ein Wert, der größer ist, je mehr Fotos denselben Punkt bzw. dasselbe Feature erkennen; unzusammenhängende Bereiche unterhalb einer vom Benutzer festzulegenden Anzahl an Punkten und eventuell vorhandene Farbinformationen für das Mesh gelöscht werden. Als "unzusammenhängenden Bereiche" werden andere Objekte im Raum bezeichnet, zum Beispiel Wände oder Tische, die nicht relevant für das gewünschte Objekt sind. Das Ergebnis *surface-L2-clean.ply* ist nun bereit zur Weiterverarbeitung in zum Beispiel Meshlab oder Blender.

Im Skript werden folgende Optionen verwendet:

`-t50` Knoten mit einem confidence-Wert von unter 50 werden gelöscht.
Entfernt unzuverlässig erstellte Teilbereiche.
Bei niedrigeren Werten bleiben unerwünschte Artefakte übrig,
bei höheren entstehen mehr Löcher.

`-c10000` Entfernt unzusammenhängende Objekte mit unter 10.000
Knoten, damit bleibt nur das Kopfmesh übrig

`--delete-color` löscht Farbinformationen für die Mesh-Knoten
(weil für HRTF Berechnung irrelevant)

6.3 Ergebnisse

MVE

Eine der ersten Aufgaben im Zuge des Praktikums war, mit den vorhandenen Tools ein bestimmtes Ergebnis zu erhalten. Mit den diskutierten Parametern der Unterprogramme sollte von einer Fotosession mit mehr als 300 Fotos das folgende Mesh generiert werden:

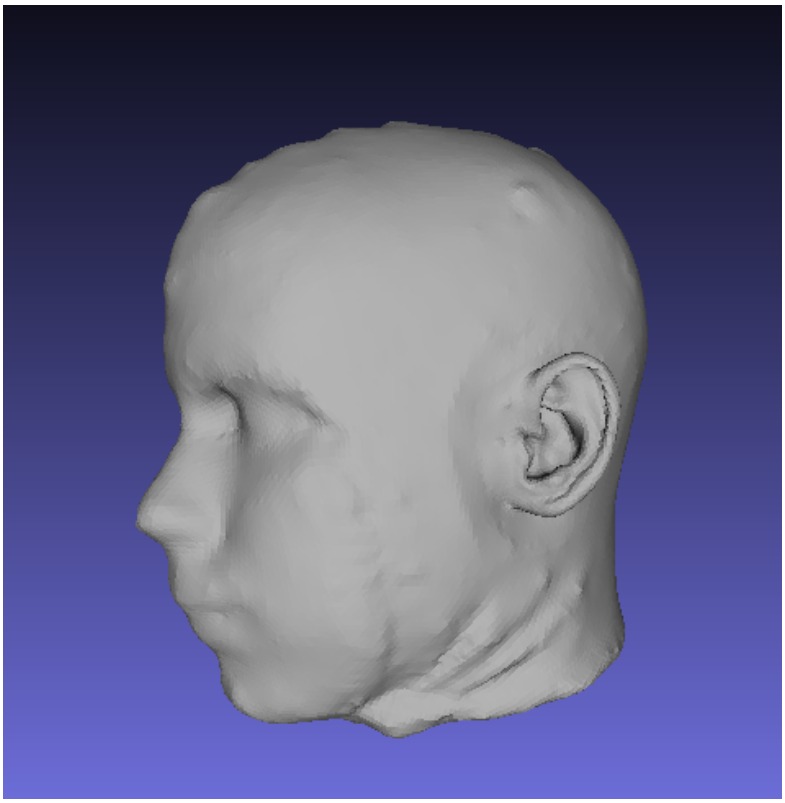

Abbildung 6.1: Vorlage mit gegebenem Parameterset.

Allerdings wurde mit dem selben Input und dem angegebenen Parameterset folgendes Ergebnis erzielt:

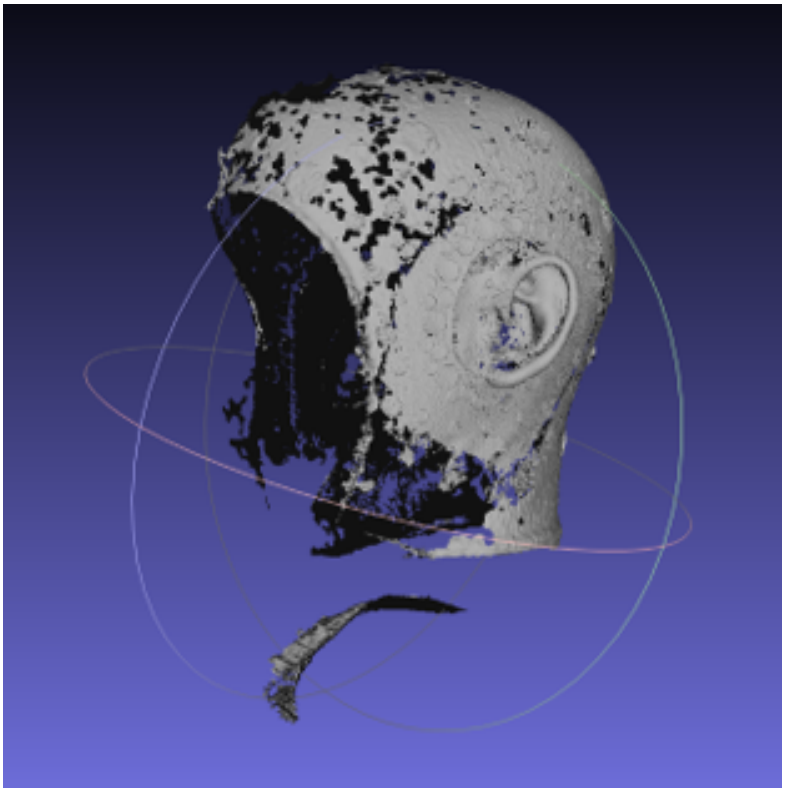

Abbildung 6.2: Missglückte Nachbildung des Status Quo vor dem Praktikum.

Zum Zeitpunkt des Praktikums war unklar, warum es bei gleichen Input zu einem unterschiedlichen Output kommen konnte; das Problem konnte auch nicht in Zusammenarbeit mit Simon Fuhrmann von der TU Darmstadt gelöst werden. Es ist zu vermuten, dass für den Kopf und die Ohren verschiedene Parametersets verwendet und dann zu einem Mesh zusammengefügt wurden.

Eine weitere Software zur Punktwolkenrekonstruktion wäre VisualSFM und CMPVMS gewesen, allerdings gab es Probleme mit dem Auffinden von notwendigen Windows .dll-Dateien und NVIDIA CUDA[11]. In beiden Fällen musste nach der Mesh-Rekonstruktion in einem 3D-Grafikprogramm weiterbearbeitet werden: Löcher waren zu stopfen und die Oberfläche evtl. zu glätten, dabei traten ebenso Probleme auf.

6.3.1 Meshlab

Zum Stopfen von Löchern und der Oberflächenglättung standen die Meshlab-Filterskripts “FakeMeshDoctor“ und “Iso-Parametrization“ zur Verfügung, allerdings ist das Mesh nach Anwenden dieser Skripts gelöscht - aus der Benutzeroberfläche unwiederbringlich entfernt - worden, wodurch ein Großteil des Workflows wiederholt werden musste. Da diese Prozedur nicht zielführend war, wurde zur Software *Blender* gewechselt.

6.3.2 Blender

Das Löcherstopfen hat viel Zeit in Anspruch genommen, weil jedes Loch händisch gestopft wurde. Einen Algorithmus, der verlässlich genug funktioniert, war zu diesem Zeitpunkt nicht bekannt. Da die Ohren im Bereich der Cavum Conchae nicht sehr gut abgebildet wurden, war es durchaus eine Herausforderung, eine Ohrkanalkorrektur durchzuführen bzw. das virtuelle Mikrofon am Eingang des Ohrkanals zu platzieren. Es wurde abschließend der Ansatz formuliert, nur die Pinnae zu exportieren und in eine vorgegebene Standard-Büste einzubinden. Da allerdings das Ergebnis des Status Quo mit MVE nicht reprodzuziert werden konnte, lag der Schwerpunkt danach auf einer anderen Software: Agisoft.

7 Agisoft

7.1 Programmüberblick

Als weitere Software zur Modellerstellung sei *Agisoft Metashape* der Firma Agisoft LLC in St. Petersburg vorgestellt, die auf Microsoft Windows, macOS und Linux läuft; Allerdings mit dem Nachteil, dass diese nicht OpenSource ist und durch das User Interface nicht direkt auf den Denkern ausgeführt werden kann. Kurz nach der Umstellung auf Agisoft war klar, dass der Praktikumsrechner in diesem Fall eine schnellere Festplatte brauchte. Auch mit einem Upgrade auf Windows 10 und einer SSD belief sich die Berechnungszeit einer Szene (von Fotos bis Mesh) trotzdem noch auf ca. 12 Stunden. Im Folgenden soll der Programmablauf im Detail erläutert werden.
Nach Programmstart müssen wie bei MVE zuerst alle Fotos geladen werden. Dazu wählt man im Menüpunkt *Workflow* "Add Folder..." aus, im kommenden Dialogfenster lässt sich bestimmen, ob ein *Chunk* aus Fotos (Kamerapositionen) erstellt werden soll oder ob es für jede Kamerapositionen mehrere Frames zur Verarbeitung gibt. Letzteres bedeutet, dass ein einzelner Ordner eine Kameraposition repräsentiert und für jede Kameraposition mehrere Fotos verschiedener Zeitpunkte herangezogen werden, ein Objekt also verfolgt werden kann. In dieser Arbeit wurde ausschließlich mit der ersten Erstellungsmethode gearbeitet, da es sich um ein statisches Objekt handelt.

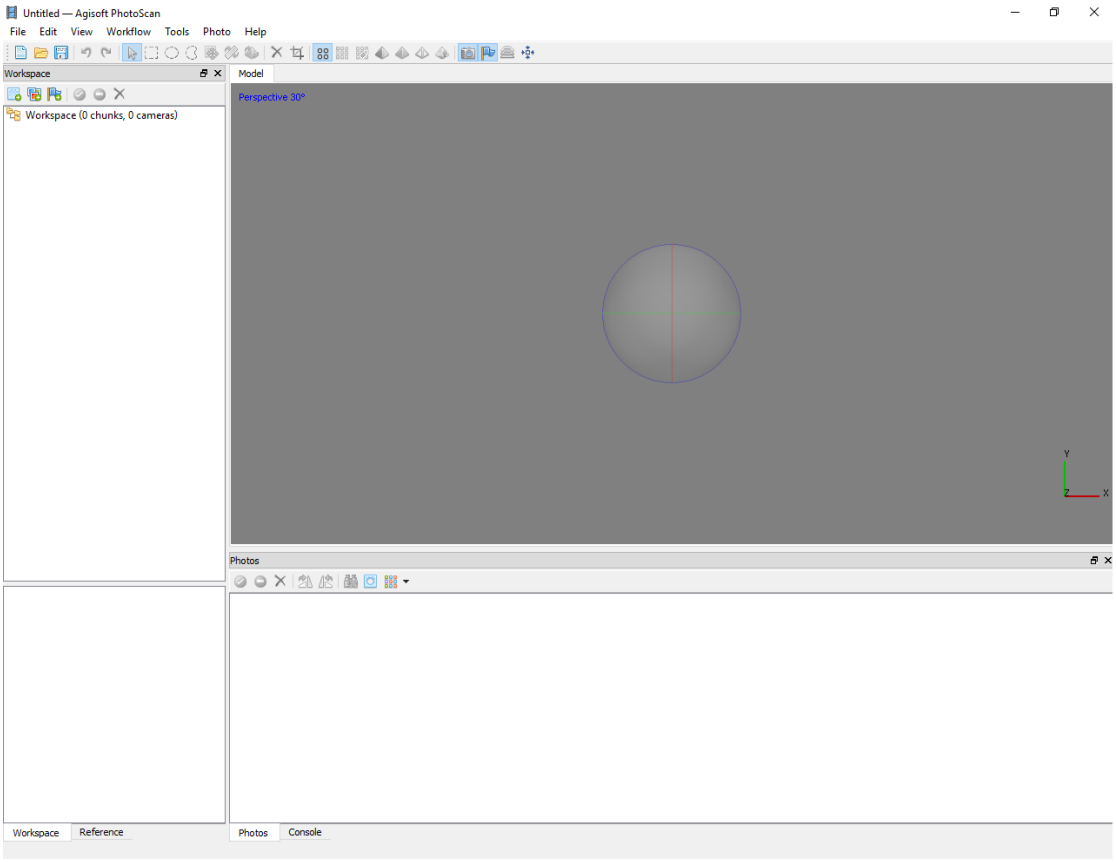

Abbildung 7.1: Grafische Benutzeroberfläche bei Programmstart.

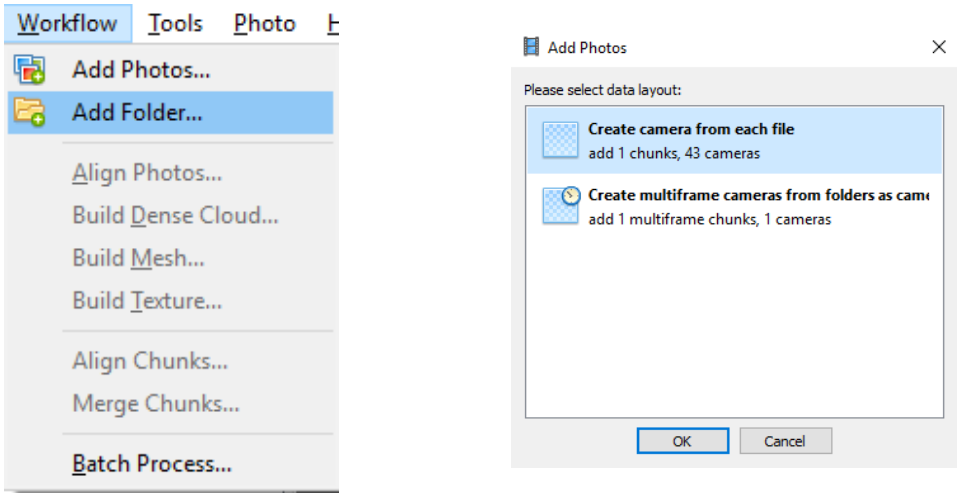

Abbildung 7.2: Erstellen eines Chunks aus einem Fotoordner (links), Auswahl eines Chunks aus N Fotos (rechts).

Sobald die Fotos in den *Chunk* geladen wurden, werden sie wie in Abbildung 7.3 dargestellt.

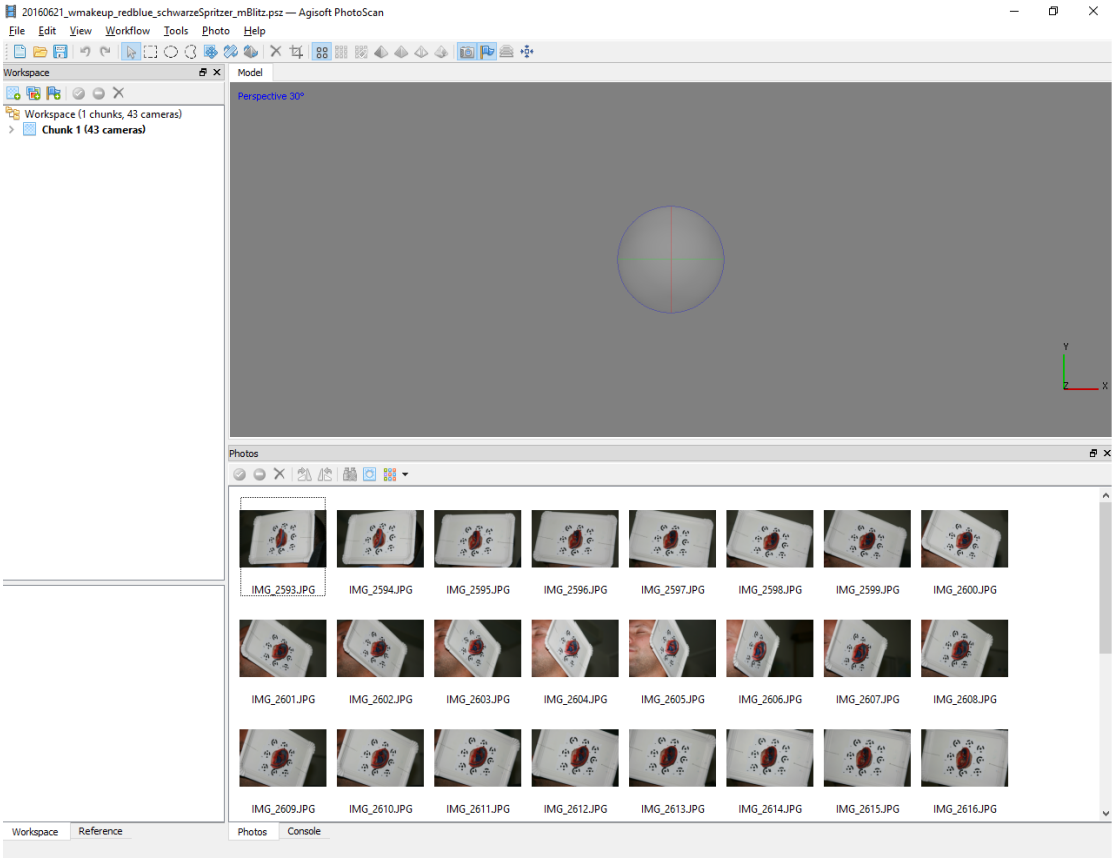

Abbildung 7.3: Grafische Benutzeroberfläche nach Erstellen eines Chunks.

Nachdem der Chunk erstellt wurde und sofern sie erkannt werden sollen, gibt es die Möglichkeit, die Präsenz von Markern (eindeutig definierte Muster, die bei der Ausrichtung der Fotos helfen) anzugeben, damit die Erkennung gestartet werden kann. Sind diese Marker zu über 90% gefunden worden, ist davon auszugehen, dass die Berechnung der Kamerapositionen ein ausgezeichnetes Ergebnis liefert.

Bei MVE wurde mit den vorhandenen Fotos gearbeitet, auf denen Marker zu finden waren, allerdings bestanden diese nicht aus eindeutig erkennbaren Mustern, sondern dienten vielmehr der Gleichverteilung stochastischer Elemente am Probanden. Nun war es so, dass bei der Umstellung auf Agisoft die Verwendung von speziellen Kreisringsegment-Markern in einer Ebene liegend um das Objekt empfohlen war und daher die Marker bei Fotosessions eine große Rolle spielten. Dabei entsprach ein Kreisringsegment einer bestimmten Nummer, die beim Drucken von Markern auch im Bereich des Markers zu sehen war, siehe Abbildung 7.6. Es wird sich allerdings am Ende der Untersuchungen zeigen, dass sogar ein besseres Mesh konstruiert werden kann, wenn die Marker weggelassen werden.

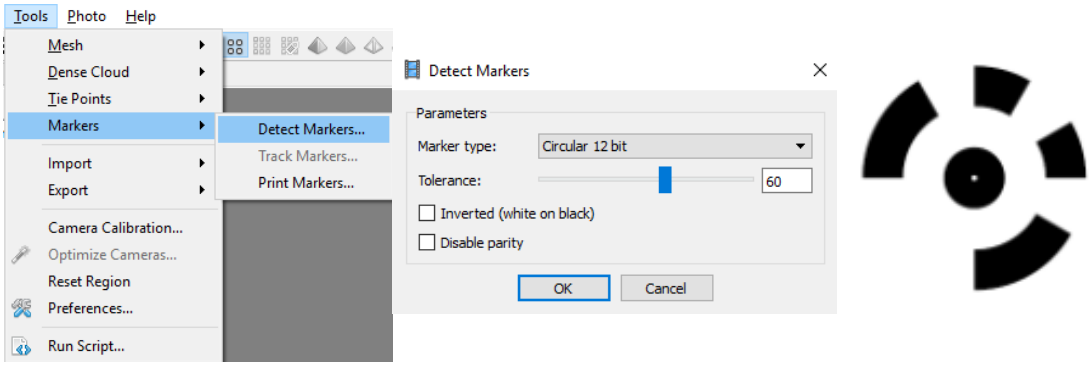

Abbildung 7.4: Detektion von Markern starten (links). Dialogfenster *Detect Markers* (Mitte). Beispiel eines Kreisringsegment-Markers (rechts).

Die Erkennung der Marker ist kaum zeitaufwändig, es dauert bei einem Set von unter 50 Fotos wenige Sekunden, bis alle Fotos geprüft sind. Sind alle Marker gefunden worden, so wird das Foto mit einem grünen Fähnchen markiert; wenn auf die Position von Markern aus anderen Fotos geschlossen wurde, ist dieses Foto zusätzlich mit einem grauen Fähnchen markiert. Werden unzureichend wenig Marker erkannt und das Foto nicht mit einem Fähnchen versehen, kann es mit einem Einbahnzeichen markiert und aus der Berechnung ausgeschlossen werden.

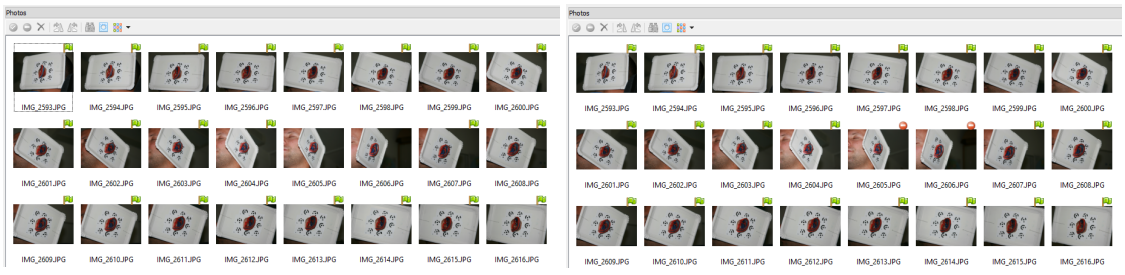

Abbildung 7.5: Markererkennung fertiggestellt (links). Fotos mit nicht erkannten Markern werden mit einem Einbahnschild gekennzeichnet und aus der Berechnung genommen (rechts).

Wählt man nach der Markererkennung ein Foto aus, kann man überprüfen, welche Marker (auch *Targets*) erkannt wurden. Hier wird klar, dass die neben dem Marker platzierte Nummer nicht zur Erkennung desselbigen beiträgt, da die Punkt-Ringsegment-Kombination für die Zuordnung ausreicht. Somit können auch Marker erkannt werden, deren Nummer aus bestimmten Kamerapositionen von der Pinna verdeckt werden.

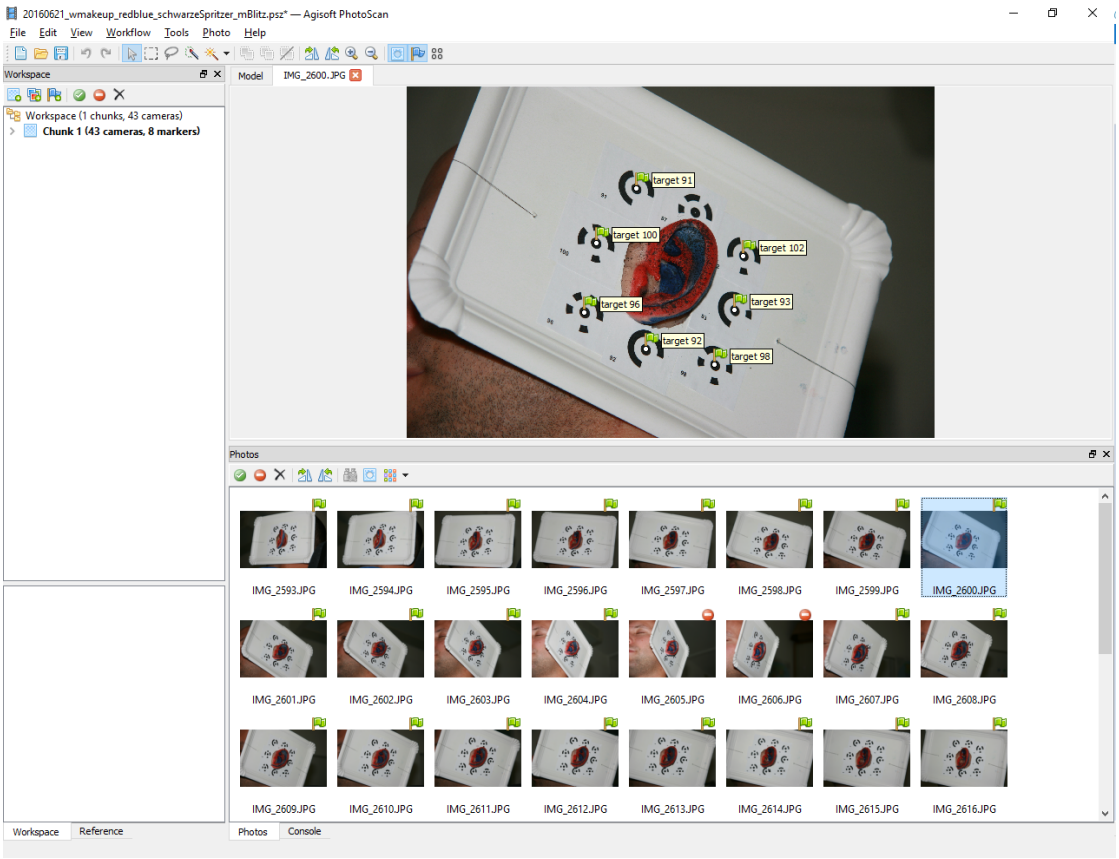

Abbildung 7.6: Die gefundenen Targetnummern stimmen mit den Markern überein.

Die nächsten Schritte sind nun sehr ähnlich zu MVE, laufen aber nur im Hintergrund ab. In Agisoft lautet die diese Schritte zusammenfassende Aktion “Align Fotos“, sie ist ebenso im Menüpunkt *Workflow* zu finden.
Im nachfolgenden Dialogfenster wurde stets mit den Parametern “hoher Genauigkeit“ und eine “keine paarweise Vorselektion“ gerechnet. Letzteres bedeutet, dass - wie bereits bei MVE - jedes Foto mit jedem verglichen und nicht wie bei Erstem die ähnlichsten Fotopaare gefunden und dann erst auf Features untersucht werden.

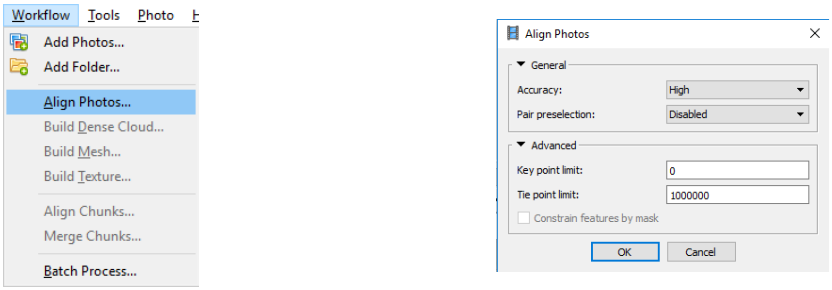

Abbildung 7.7: Menüpunkt Auswahl (links). Optionen zu Align Fotos (rechts).

Sind nun die Kamerapositionen rekonstruiert und die Punktwolke erstellt, besteht vor der Berechnung des Meshes die Möglichkeit, potentiell falsch erkannte Positionen für die weitere Berechnung zu deaktivieren.

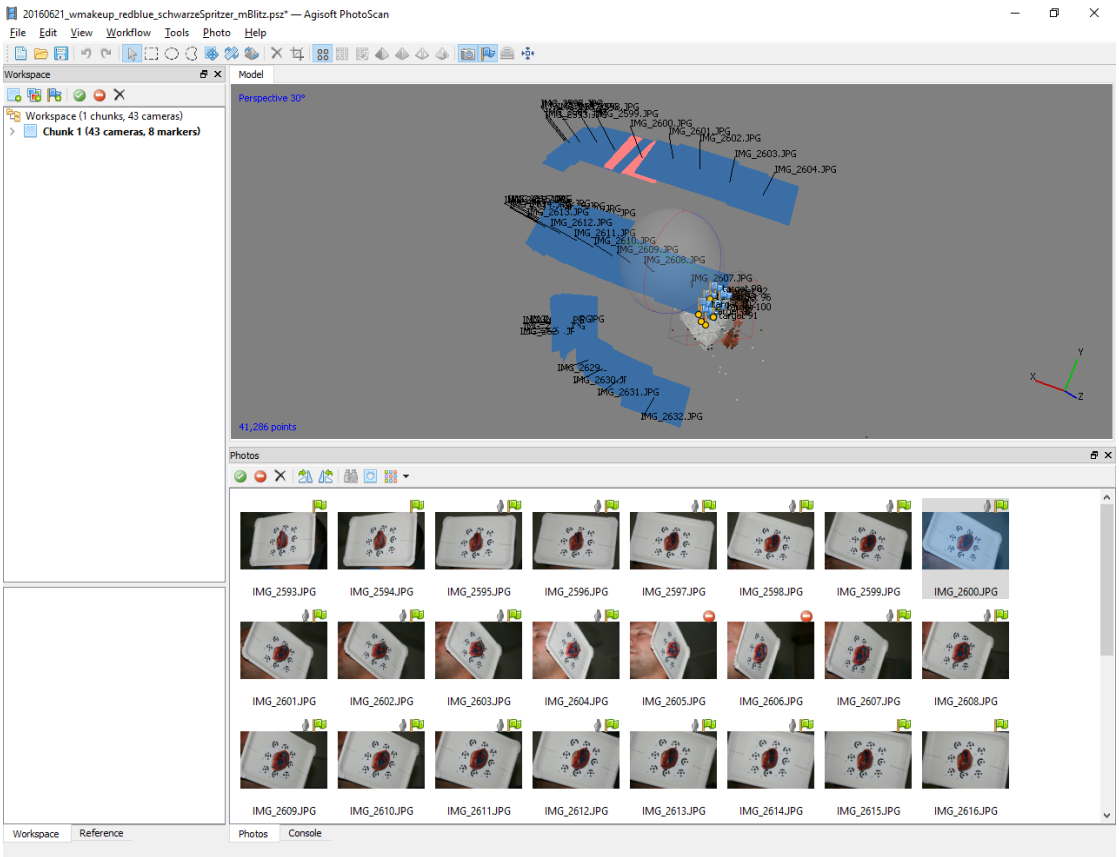

Abbildung 7.8: Die berechnete Kameraposition des ausgewählten Fotos aus dem Fotoordner wird rosa eingefärbt.

Damit vom Zielobjekt entfernte Punkte aus der Punktwolke nicht in die weitere Berechnung miteinfließen, wird nun ein Quader mit Hilfe der Werkzeugleiste erstellt, der den

Kalkulationsbereich einschränken soll.

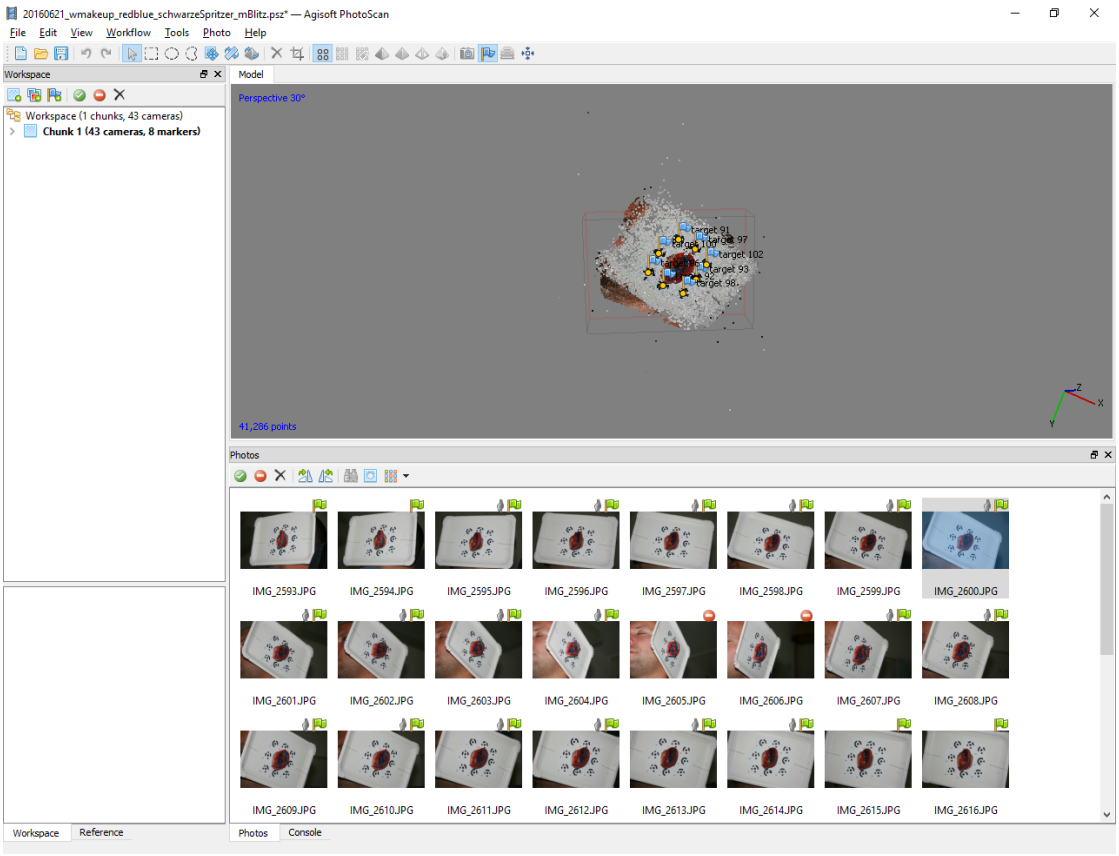

Abbildung 7.9: Quader zur Einschränkung des Berechnungsraums.

Abschließend wird aus der spärlichen Punktwolke eine dichte Punktwolke berechnet. Erneut gibt es in einem Dialogfenster weitere Einstellungen zu bestimmen, wobei die Qualität auf die höchste Stufe und das *depth filtering* auf mild gestellt wird. Bei dieser Einstellung wird zwar die geringste Anzahl an Punkten, die außerhalb einer geschlossenen Fläche liegen, entfernt; allerdings liefert diese Einstellung das beste Ergebnis im Vergleich zu *moderate* oder *aggressive*. Eine komplette Deaktivierung dieser Option ist nicht ratsam, da es durchaus noch fehlerhaft erkannte Punkte geben kann, die das Modell unnötig verrauschen.

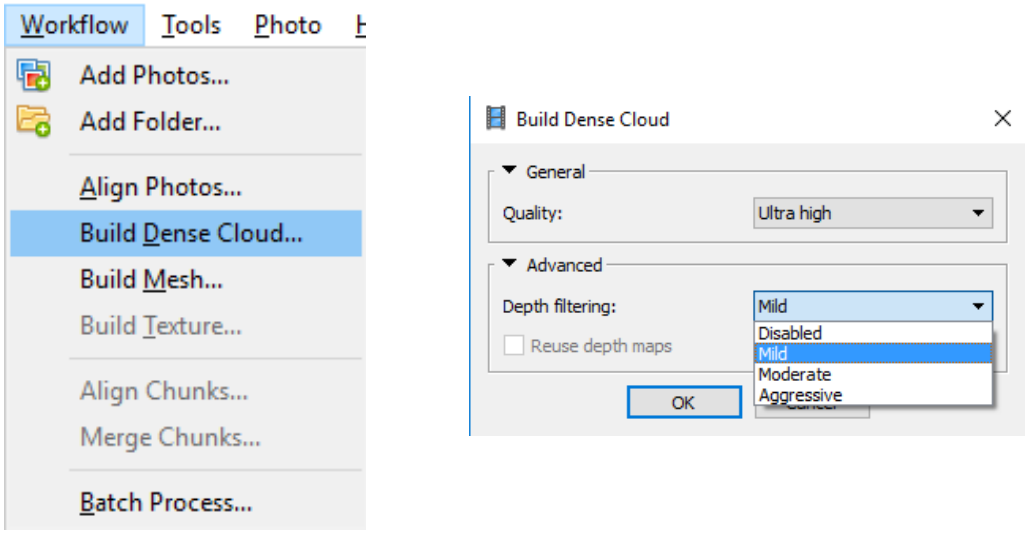

Abbildung 7.10: Vor der Mesherstellung muss noch eine dichtere Punktwolke erstellt werden (links). Einstellungen bei der dichten Punktwolkenberechnung (rechts).

7.2 Fotosessions

Zunächst wurde bei Agisoft der vorhandene Input (340 Fotos von Kopf und Pinnae) herangezogen, um das Ergebnismesh mit MVE und dem status Quo vergleichen zu können.

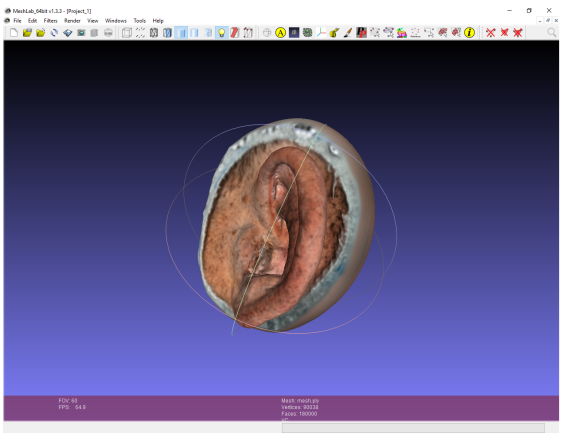

Abbildung 7.11: Das erste Ergebnis ist zufriedenstellend: nur die Tragusinnenseite ist verrauscht, kann aber geglättet werden. Allerdings ist der Schnittpunkt der Crus Inferius Anthelicis (die *Gabel*) mit der Helix nicht eindeutig, das könnte eventuell zu Problemen führen.

Im nächsten Schritt sollte die Anzahl der Fotos reduziert werden. Die Wahl fiel wie

bereits erwähnt auf drei verschiedene Elevationswinkel (ca. ±45° und 0°) und einen Azimuthschritt von ca. 15°, womit ich auf ca. 40 Fotos pro Pinna kam. Im Allgemeinen wurde bei allen Fotosession nur eine Pinna bearbeitet, da für die zweite Pinna derselbe Arbeitsablauf gilt.

Da die Kopfbedeckung bei der ursprünglichen Fotosession verwendet wurde, um Haare vom Ohr zurückzuhalten, wurde eine ähnliche erneut verwendet. Unglücklicherweise konnte nicht dieselbe benutzt werden, da sie bei der Lagerung die Marker geknickt hatte, allerdings wurde versucht, mittels Kajal-Stift Punkte zufällig am Ohr verteilt zu zeichnen. Damit die Härchen am Außenohr eine möglichst geringe (visuelle) Rauschquelle darstellten, wurden sie mit Makeup (unter den Kajal-Punkten) am Außenohr “festgeklebt“.

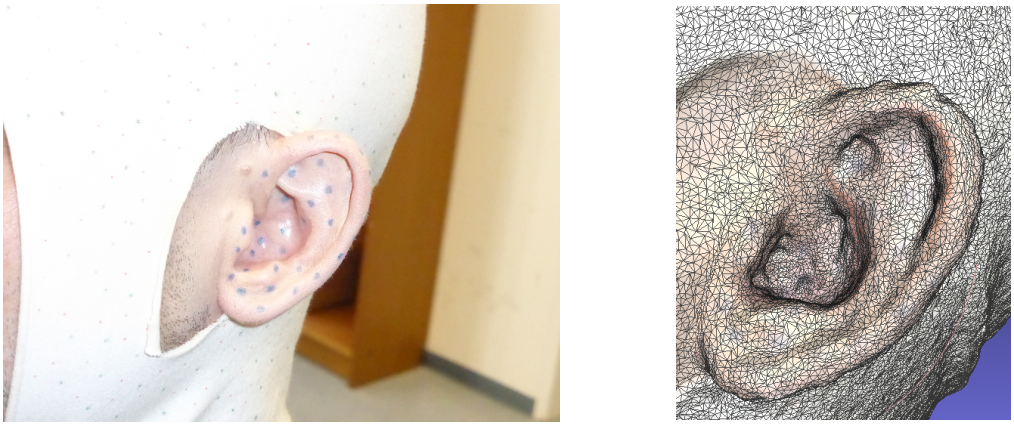

Abbildung 7.12: Makeup zur Härchenglättung und blaue zufällig verteilte Punkte mit Kajal. Fotos mit Blitz, ohne Marker.

Hier wurde klar, dass der Einsatz von Blitz unangebracht ist, da er für zusätzliches Rauschen sorgt, wenn auf benachbarten Fotos ähnliche Regionen dunkel sind, die aus einem anderen Winkel gut ausgeleuchtet werden. Die ursprüngliche Idee, verdeckte Bereiche wie die Krempe der Helix so zu erfassen, wurde somit zunichte gemacht.

Nun, da ein zufriedenstellendes Ergebnis mit den ursprünglichen Fotos erzielt wurde und die erste Fotosession bestätigte, dass Marker verwendet werden sollten, war der nächste Punkt die Bestimmung der Markergröße. Da ein Versuch mit dem Probanden erst nach Evaluierung der Marker in Frage kam, wurde mir zu ersten Fotografieversuchen ein Silikonohr zur Verfügung gestellt. Begonnen wurde mit runden Markern (Kreis-Ringsegment-Kombination) mit einem Durchmesser von ca. 1 cm, allerdings wurde dabei der Untergrund besser konstruiert als das Ohr, siehe Abbildung 7.13.

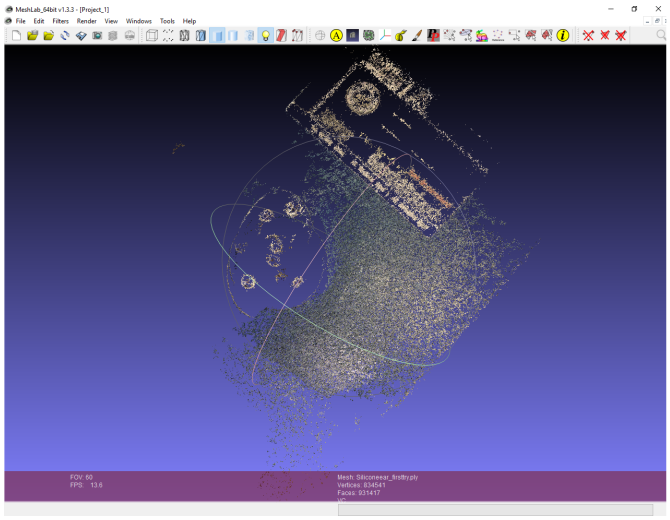

Abbildung 7.13: Die Marker müssen vielleicht in einer Ebene liegen, bzw. dürfen sie nicht geknickt werden (zum Beispiel in der Concha).

Der neu aufgesetzte PC ließ die Einstellung “ultrahohe Auflösung“ zu, wie Abbildung 7.14 deutlich zeigt. Ebenso ermöglichte diese Einstellung eine ausreichend gute Abbildung der Rückseite der Pinna, siehe Abbildung 7.15. Die Krempe der Helix bleibt allerdings weiterhin eine Problemstelle.

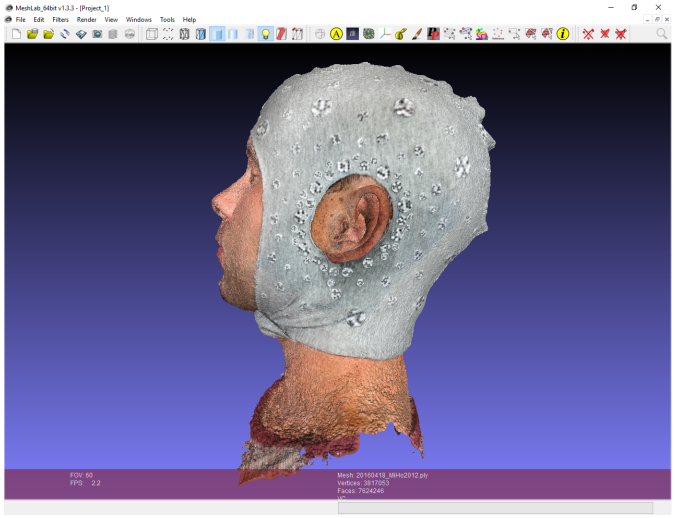

Abbildung 7.14: Dank wesentlich genauerer Berechnung ist nun auch der Schnittpunkt der Crus Inferius Anthelicis mit der Helix besser abgebildet.

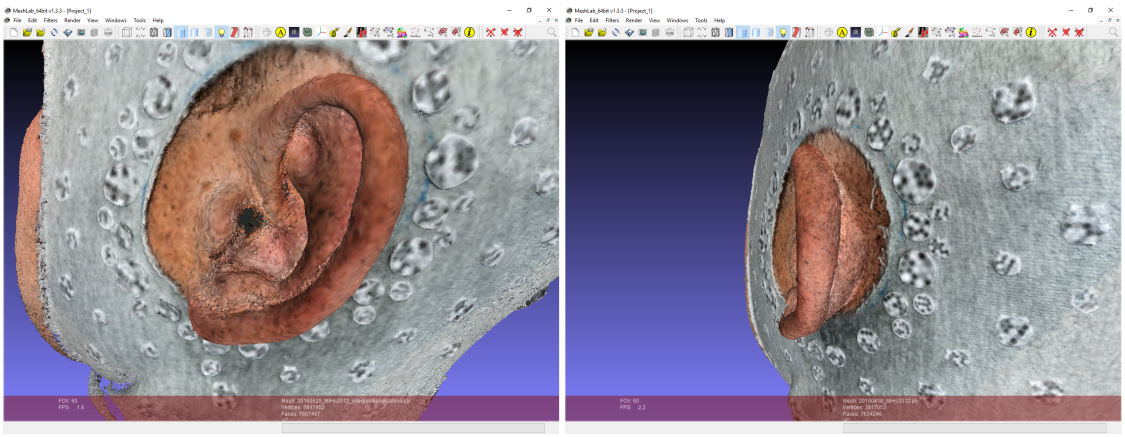

Abbildung 7.15: Die einzig verbleibende Problemstelle: Krempe der Helix (links). Hinterer Bereich gut erfasst (rechts).

Im Folgenden wurde mit dem Silikonohr getestet, welche Kombination von Markergröße und -anzahl das genaueste Ergebnis lieferte, damit mit dem Probanden zuverlässige Fotos gemacht werden konnten.

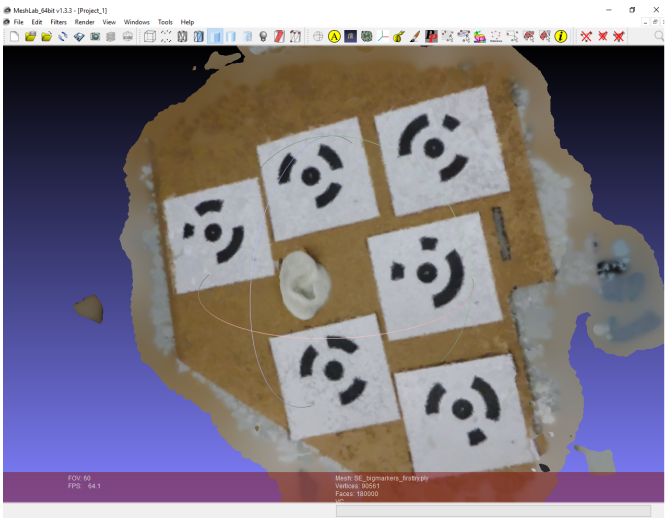

Abbildung 7.16: An die Anleitung von Agisoft[12] gehalten, stimmt das Verhältnis von Maker- zu Objektgröße nicht.

Die Markergröße mit 1 cm Innenringradius war zu klein (siehe Abbildung 7.13), in Abbildung 7.16 waren sie mit $2,5$ cm zu groß. Der anfangs geglaubte Nachteil der Marker, es müssen die Zahlen auf den Markern für die Kamera erkennbar sein, erwies sich dann als Irrtum.

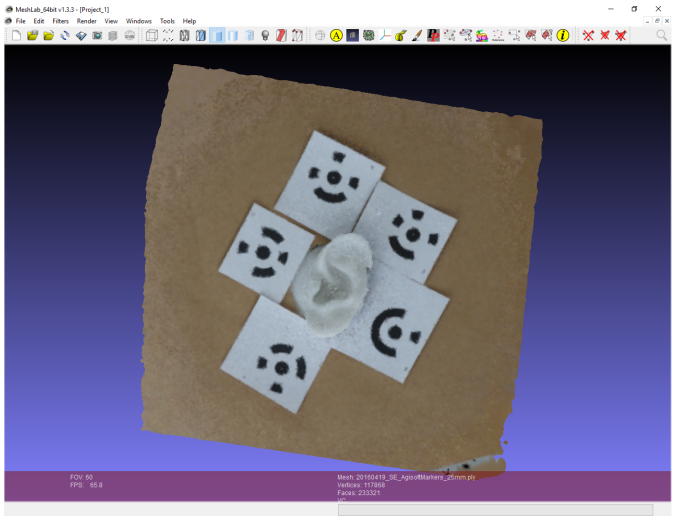

Abbildung 7.17: Marker in engerem Abstand, aber immer noch zu groß. Helixkrempe nur Hügel, teilweise durchlöchtert.

Trotzdem die Helix stark durchlöchert und die Tiefenschärfe der Pinna praktisch nicht vorhanden war, wurde eine Fotosession mit den großen Markern am Proband ausgeführt. Sie lieferte ungefähr das gleiche unzufriedenstellende Ergebnis wie der Versuch mit dem Silikonohr und zusätzlich war es für den Probanden unangenehm, einen so großen Karton mit der Pinna und Schulter halten zu müssen.

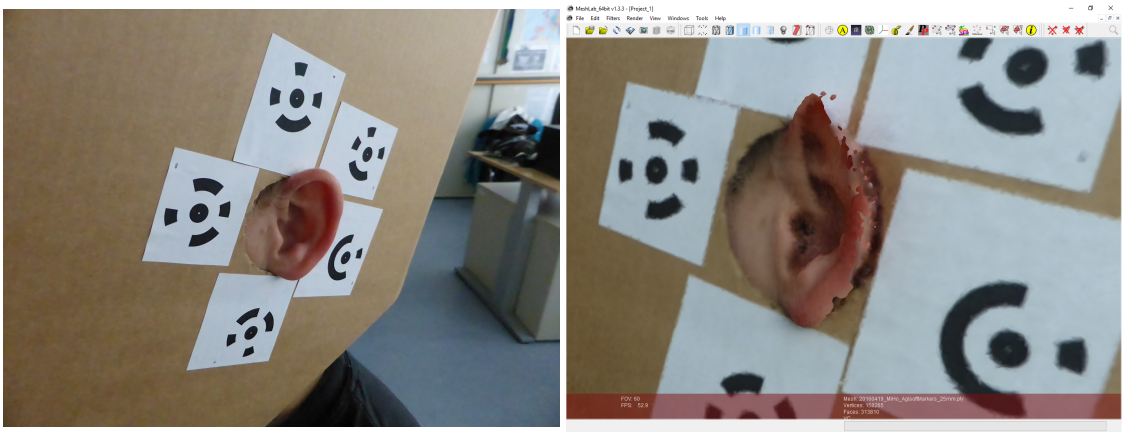

Abbildung 7.18: Auch beim Probandenversuch ist die Helix durchlöchert und die Pinna besitzt allgemein wenig Tiefenschärfe.

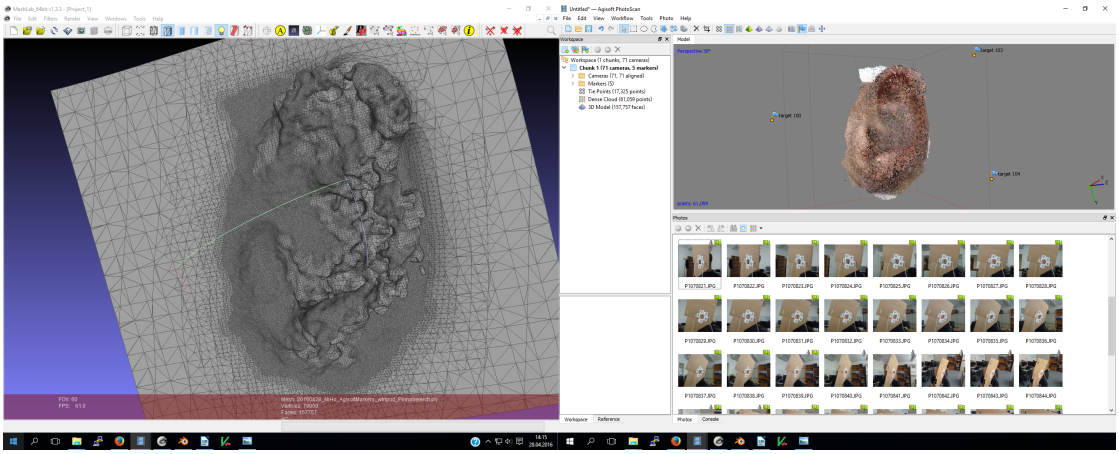

Abbildung 7.19: Selbst zugeschnittener Berechnungsquader auf Pinna ergibt kein genaueres Ergebnis.

Damit der Proband nicht unnötig während der Fotosession physisch belastet werden muss, wurde der Karton kleiner geschnitten und mit einer Schnur um den Kopf befestigt.

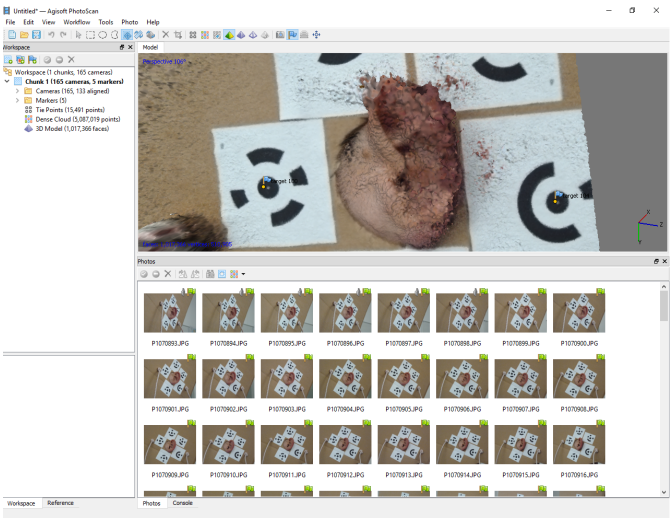

Abbildung 7.20: Pinna ist nun nicht mehr erkennbar. Schnur vielleicht unscheinbarer machen? Karton kleiner? Marker kleiner?

Allerdings wurde zu diesem Zeitpunkt entgültig klar, dass der Karton und die darauf befindlichen Marker zu groß gewählt wurden (siehe Abbildung 7.20). Weiters wurde zusätzlich nach einer Möglichkeit gesucht, das Ohr kontrastreicher zu machen, um die Tiefenstaffelung zu verbessern; die Wahl fiel dabei auf Wasserfarben.

Abbildung 7.21: Helle (grün) und dunkle (blau) Farbe, damit die Pinna kontrastreicher wird (links). Dreiecke des Meshs sind etwas groß, aber die Pinna ist zufriedenstellend vollständig (rechts).

Da der Wasserfarben-Ansatz das bisher beste Ergebnis mit wenigen Fotos lieferte, wurde er weiterverfolgt; es wurde jedoch schnell ein Fehler klar: Kontrast ist nicht durch helle und dunkle Farben gegeben, sondern durch auf der Graustufenskala (engl. *grayscale*) möglichst weit auseinanderliegenden Lichtintensitäten, wie es zum Beispiel bei rot und blau der Fall ist.
Vor dieser Erkenntnis wurden allerdings noch folgende zwei Fotosessions durchgeführt: Bei der ersten (siehe Abbildung 7.22) wurden die Konturen der Pinna, interpretiert als die nach außen gewölbten Bereiche (Tragus, Helix, Antitragus bis Crus Inferius Antihelicis) mit schwarzen Wasserfarben hervorgehoben. Das hatte allerdings zur Folge, dass es zu einer Links-Rechts-Verwechslung kam und die Punktwolke vor allem im Bereich der Helix durchlöchert und somit unbrauchbar zur weiteren Bearbeitung war. Die anschließende Idee, das Ohr komplett grau zu bemalen, lieferte ebenso eine mehr als unzufriedenstellendes Mesh (siehe Abbildung 7.23).

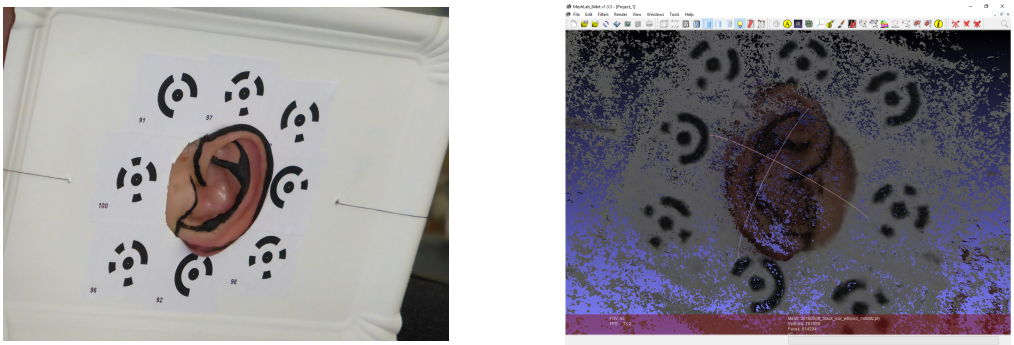

Abbildung 7.22: Konturen alleine reichen nicht aus, Verwechslungen gleicher Punkte aus verschiedenen Perspektiven treten auf.

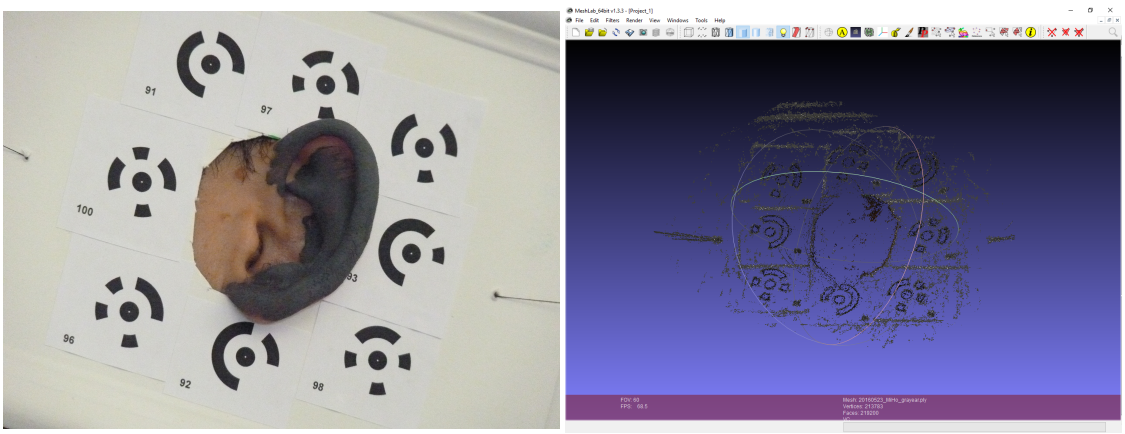

Abbildung 7.23: Das gewählte Dunkelgrau lieferte ein schlechtes Ergebnis am Mesh.

Nach der Erkenntnis über Kontrast wurden drei letzte Fotosessions durchgeführt. Bei allen drei galten folgende Überlegungen:

- Verwendung von Makeup, damit Härchen an der Pinna anliegen
- Farben, die auf der Graustufenskala so weit wie möglich auseinanderliegen
- stochastisches Element: schwarze Spritzer

Wie bereits erwähnt, war die nächste Idee weit auseinanderliegender Farben die Verwendung von Rot und Blau, welche zu folgendem Ergebnis führte:

Abbildung 7.24: Rotblau mit schwarzen Punkten. Fotos (links), Punktwolke (rechts).

Das Ergebnis der Punktwolke war zufriedenstellend, aber es waren noch größere Löcher vorhanden. Daher wurden in einer letzten Fotosession die auf der Graustufenskala am weitesten voneinander entfernten Farben - schwarz und weiß - ausprobiert.

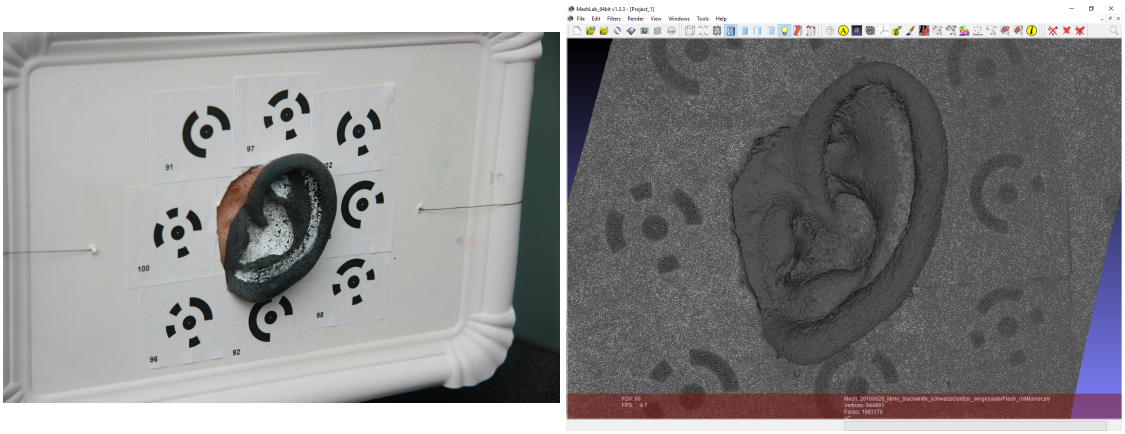

Abbildung 7.25: Schwarzweiß mit Punkten. Fotos (links), Mesh (rechts).

Hier sind nun die Löcher geschlossen, die noch bei der Rot-Blau-Kombination vorhanden waren. In Abbildung 7.26 sind nur noch zwei Bereiche deutlich, die nicht ausreichend erfasst wurden: der obere Übergang der Concha zur Crus Inferius Antihelicis und der Übergang der oberen Crus Inferius Antihelicis zur Helix.

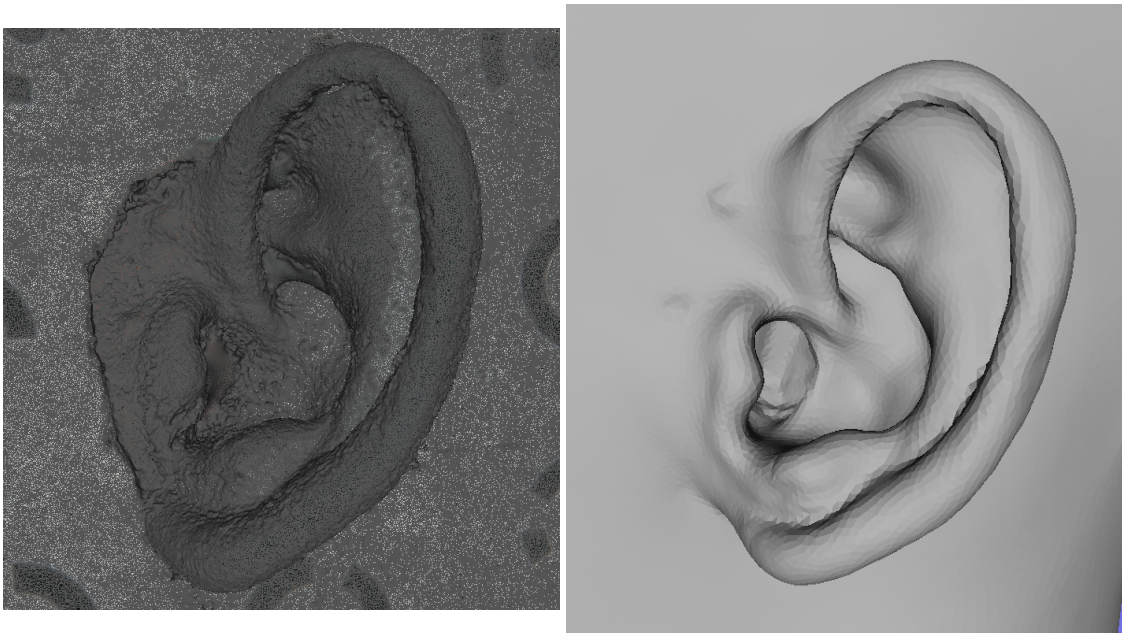

Abbildung 7.26: Ergebnis Schwarz-Weiß-Kombination (links). Referenz aus CT-Scan (rechts).

Abschließend wurde noch die Sinnhaftigkeit der Marker in Frage gestellt, da die Verwendung selbiger durchaus einen sowohl unkonventionellen als auch unangenehmen Messaufbau für den Probanden darstellen, wobei beides möglichst verhindert werden will.

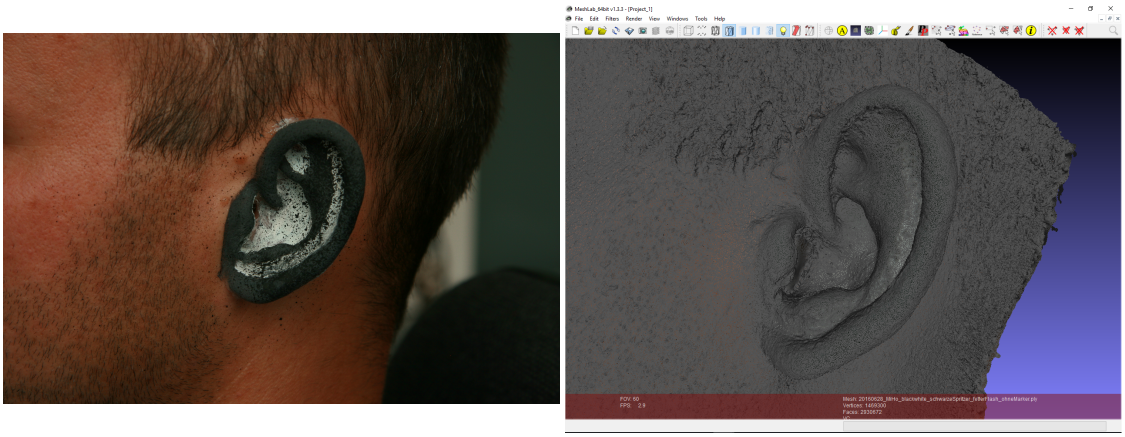

Abbildung 7.27: Ohne Marker, mit Makeup, Schwarz-Weiß-Kombination und schwarzen Spritzern. Fotos (links), Mesh (rechts).

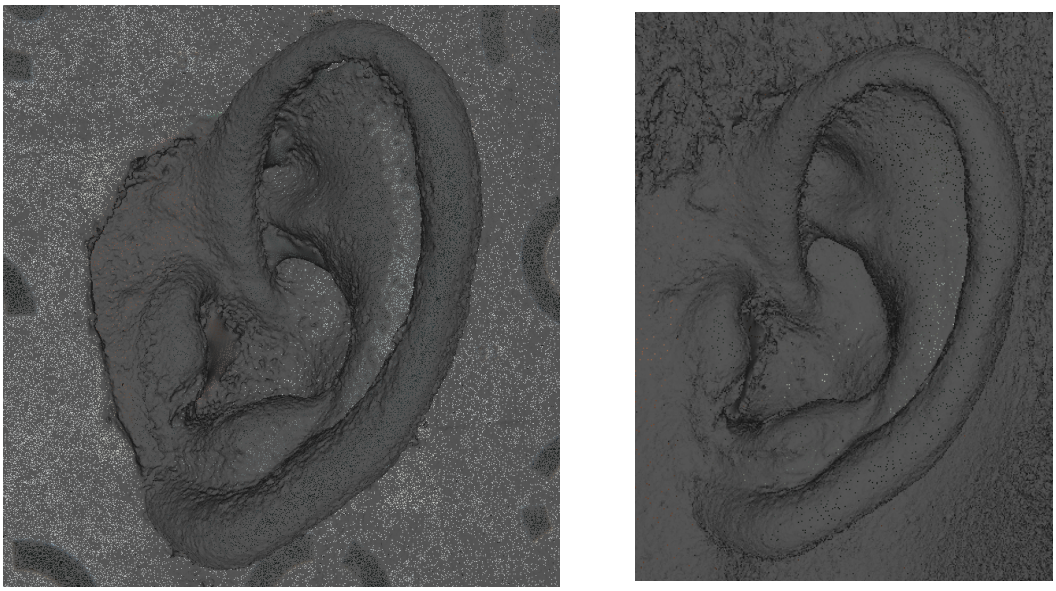

Abbildung 7.28: Vergleich mit (links) und ohne (rechts) Marker.

Das 3D-Modell mit Schwarz-Weiß-Kombination wurde ausgewählt, eine Simulation der HRTF damit durchzuführen. Dazu wurde das Ohr in eine Kunstkopfbüste eingearbeitet und mittels *mesh2hrtf*-Funktion die Außenohrübertragungsfunktion ermittelt. Im Folgenden sind das auf Fotogrammetrie basierende Simulationsergebnis für Elevation der rechten Pinna über eine lineare Frequenzachse und die Referenzmessung geplottet:

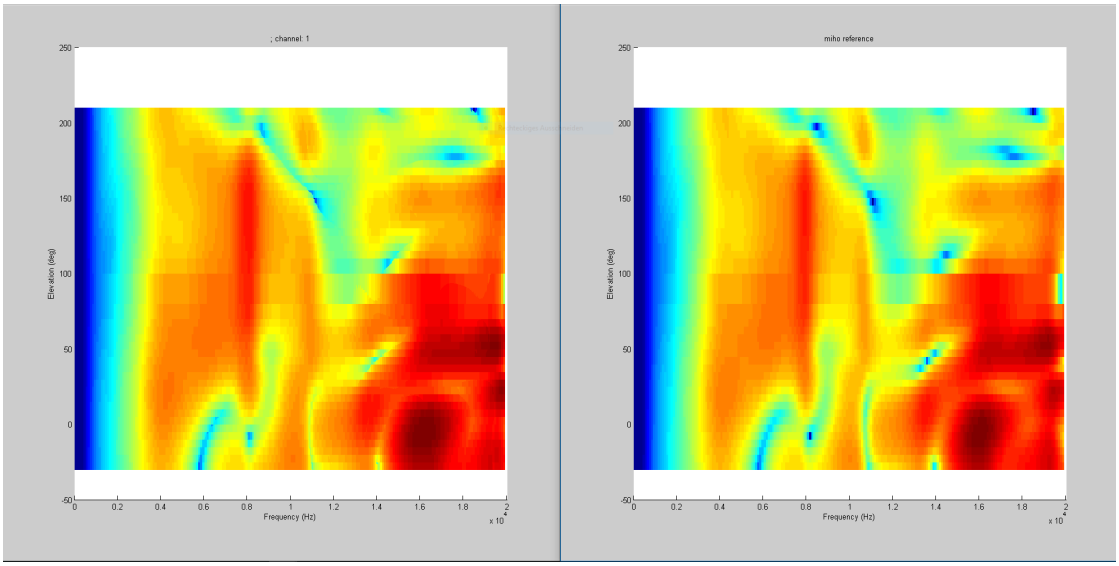

Abbildung 7.29: Vergleich mittels Fotogrammetrie simulierter HRTF (links) zu gemessener Referenz (rechts)

Leider ist die Farbskala aus Abbildung 7.29 verlorengegangen, allerdings konnte der Dynamikbereich von ca. 60 dB(FS) (von blau (−60 dB) über grün und geld nach rot (0 dB)) nachträglich eruiert werden. Deutlich erkennbar ist die Ähnlichkeit der mittels Photogrammetrie berechneten Simulation zur simulierten Referenz aus dem Computertomographie Scan. Die Auflösung eines Meshes wird durch die durchschnittliche Kantenlänge der Dreiecksflächen - Average Edge Length - bestimmt. Bei der Referenz und dem Photogrammetrie-Mesh beträgt sie jeweils ca. 1 mm, was darauf schließen lässt, dass die aus beiden Meshes berechneten HRTFs sehr ähnlich sind. Bereits 2013 wurde psychoakustisch untersucht[9], dass das Referenzmesh mit einer Auflösung von 1 mm zu perzeptiv validen Ergebnissen führt, was darauf schließen lässt, dass ein durch Photogrammetrie konstruiertes Mesh einer Pinna zu einem ausreichend ähnlichen Ergebnis führt wie eines mittels CT-Scan konstruiertes.

8

Diskussion und Zukunftsaussichten

Ziel dieser Arbeit war, den Input einer bereits bestehenden Programmpipeline zu evaluieren und eine möglichst exakte Rekonstruktion der Pinna zu ermöglichen.
Nach einer missglückten Rekreierung des vorhandenen Arbeitsablaufes von Fotos bishin zum Mesh mittels des Ansatzes von Multi-View Environment konnten mit der zweiten Software Agisoft sehr gute Ergebnisse erzielt werden. Des Weiteren wurde gezeigt, dass die ursprüngliche Anzahl an Fotos von 340 für beide Ohren auf 40 pro Pinna reduziert werden konnte.
Abschließend lässt sich sagen, dass von Agisoft empfohlene Marker für die Pinnamodellierung unangebracht sind und dass das Festkleben der Härchen an der Pinna mittels Makeup und Kontrastanreicherung durch Wasserfarben ein weniger verrauschtes Modell ergibt als nur Pinselspritzer auf der Pinna. Letzteres ist darauf zurückzuführen, dass

mittels Pinsel auf die Pinna applizierte Spritzer an den Härchen hängen bleiben und somit eine Rauschquelle darstellen.

Die Konstruktion eines Ohrenmodells und Erhalt der eigenen HRTFs sollte einem größtmöglichen Publikum eröffnet werden, da die wenigen existierenden Messräume eine begrenzte Kapazität haben und des Weiteren nur für im Umkreis wohnende Personen örtlich günstig erreichbar sind. Ein langfristiges Ziel ist - in Anbetracht dessen, dass Kameras in mobilen Geräten stetig bessere Qualität aufweisen - dass der/die Endnutzer*in seine/ihre Ohren ausreichend fotografiert, diese Fotos zur Berechnung auf einen Server lädt und wenige Stunden später die HRTFs erhält.
Im Nachhinein betrachtet wäre die Frage auch interessant gewesen, wie schlecht die Fotoqualität bzw. das Mesh sein muss, um HRTFs mit ausreichender Auflösung zu erhalten. Dazu hätte von allen Zwischenmodellen der Pinna ebenso die HRTF berechnet werden müssen, wofür im Zuge des Praktikums zu wenig Zeit verfügbar war.

Ein weiterer Ansatz zur Modellierung stammt aus dem medizinischen Bereich und wird bei der Erstellung von 3D-Tumormodellen aus CT Scans angewendet. Dabei gibt es ein Modell eines “Standard-Tumors“, das in Abhängigkeit der gefundenen Punktwolke verzerrt wird. So wäre es möglich, ein bereits vorhandenes, sehr gut abgebildetes Pinnamodell an die Proband*innenohren anzupassen, womit der Qualitätsstandard des Meshes gesichert wäre.

Bibliography

[1] P. Majdak. *Räumliches Hören*. Nov. 13, 2014. URL: http://piotr.majdak.com/alg/VO/spatial1.pdf.

[2] Wikipedia. *Photogrammetrie*. Apr. 26, 2016. URL: https://de.wikipedia.org/wiki/Photogrammetrie.

[3] Sourceforge. *mesh2hrtf*. July 29, 2019. URL: http://mesh2hrtf.sourceforge.net/.

[4] S. Fuhrmann, F. Langguth, and M. Goesele. "MVE - A Multi-View Reconstruction Environment". In: *EUROGRAPHICS Workshops on Graphics and Cultural Heritage* (2014).

[5] Simon Fuhrmann. *Multi View Environment*. Nov. 4, 2015. URL: https://github.com/simonfuhrmann/mve.

[6] Wikipedia. *Depth Map*. July 26, 2019. URL: https://en.wikipedia.org/wiki/Depth_map.

[7] S. Fuhrmann and M. Goesele. "Floating Scale Surface Reconstruction". In: *Association for Computing Machinery* (Aug. 10–14, 2014). DOI: http://dx.doi.org/10.1145/2601097.2601163.

[8] S. Furhmann and M. Goesele. "Floating Scale Surface Reconstruction Supplemental Material". In: *Association for Computing Machinery* (2014).

[9] H. Ziegelwanger, A. Reichinger, and P. Majdak. "Calculation of listener-specific head-related transfer functions: Effect of mesh quality". In: *ICA 2013 Montreal* (June 2–7, 2013).

[10] P. Mücke, R. Klowsky, and M. Goesele. "Surface reconstruction from multi-resolution sample points". In: *Vision, Modeling, and Visualization* (2011).

[11] Wikipedia. *Compute Unified Device Architecture*. URL: https://de.wikipedia.org/wiki/CUDA.

[12] Agisoft. *Coded Targets Scale Bars in Agisoft PhotoScan Pro 1.1*. URL: http://www.agisoft.com/pdf/PS_1.1_Tutorial%20(IL)%20-%20Coded%20Targes%20and%20Scale%20Bars.pdf.