

TECHNISCHE UNIVERSITÄT DRESDEN

FAKULTÄT INFORMATIK

INSTITUT FÜR SOFTWARE- UND MULTIMEDIATECHNIK

PROFESSUR FÜR COMPUTERGRAPHIK UND VISUALISIERUNG

PROF. DR. STEFAN GUMHOLD

## Diplomarbeit

zur Erlangung des akademischen Grades  
Diplom-Informatiker

# Optimierung und Übertragung von Tiefengeometrie für Remote-Visualisierung

Josef Schulz

(Geboren am 20. Oktober 1989 in Naumburg (Saale), Mat.-Nr.: 3658867)

Betreuer: Dr. Sebastian Grottel

Dresden, 24. November 2016

---

## Aufgabenstellung

In Big-Data-Szenarien in der Visualisierung spielt der Ansatz der Remote-Visualisierung eine zunehmende Rolle. Moderne Netzwerktechnologien bieten große Datenübertragungsraten und niedrige Latenzzeiten. Für die interaktive Visualisierung sind aber selbst kleinste Latenzzeiten problematisch. Um diese vor dem Benutzer maskieren zu können, kann eine Extrapolation der Darstellung durchgeführt. Diese Berechnungen erfordern zusätzlich zum normalen Farbbild weitere Daten, beispielsweise ein Tiefenbild und die Daten der verwendeten Kameraeinstellung. Für die Darstellungsextrapolation werden Farb- und Tiefenbild zusammen interpretiert, beispielsweise als Punktwolke oder Höhenfeldgeometrie. Im Rahmen dieser Arbeit soll untersucht werden, wie die Darstellung mittels Höhenfeldgeometrie optimiert werden kann. Ansätze sind hierfür Algorithmen aus der Netzvereinfachung. Zu erwarten sind sowohl harte Kanten als auch glatte Verläufe der Tiefenwerte, welche sich in der Netzgeometrie durch adaptive Vernetzung mit reduziertem Datenaufwand darstellen lassen.

Dem Szenario der Web-basierten Remote-Visualisierung folgend soll der Web-Browser als Klient-Komponente eingesetzt werden. Die einzusetzenden Technologien sind HTML5, Javascript, WebGL und WebSockets. Entsprechende Javascript-Bibliotheken sollen genutzt werden um die Qualität und Wartbarkeit des Quellcodes zu steigern. Für die Server-Komponente darf die Technologie vom Bearbeiter frei gewählt werden.

Zu Beginn der Arbeit wird eine Literatur-Recherche zu Web-basierter Visualisierung und Remote-Visualisierung erfolgen. Schwerpunkte sind hierbei die Bild-Extrapolation, Vernetzung und Rekonstruktion auf Basis von Tiefenbildern und die Netzoptimierung und -Vereinfachung. Im Anschluss an die Literaturrecherche wird ein Konzept für die Implementierung mit dem Betreuer abgesprochen und anschließend als prototypische Software umgesetzt. Folgendes Szenario dient als Grundlage für dieses Konzept:

Als Eingabedaten stehen mehrere Datensätze aus unterschiedlichen Szenarien der wissenschaftlichen Visualisierung zur Verfügung. Für jeden Datensatz sind mehrere Tripel aus Farbbild, Tiefenbild und Kamera-Parameter gegeben. Die Serverkomponente bereitet einen Datensatz auf und bietet ihn dem Klienten an. Diese Aufbereitung ist vor allem die Generierung einer optimierten Tiefennetzgeometrie aus den Tiefenbilddaten. Der Klient fordert Farbbilder, Kameraeinstellungen und Tiefengeometrie von Tripel-Paaren an. Konzeptuell wird ein Tripel als aktueller Zustand und das zweite Tripel als Ground-Truth einer Bildextrapolation verstanden. Diese können daher auch in dieser Reihenfolge angefordert werden. Die Tripel werden zwischen

---

Klient und Server direkt per Sockets/WebSockets übertragen. Die Daten des ersten Tripels werden anschließend genutzt um dessen Farbbild in die Ansicht des zweiten Tripels extrapoliert. Hierbei werden vom zweiten Tripel nur die Kameraeinstellung genutzt. Diese Extrapolation wird Klient-seitig in WebGL implementiert damit alle Berechnungen auf der GPU ausgeführt werden. Anschließend wird das extrapolierte Bild mit dem originalen Ground-Truth-Farbbild aus dem zweiten Tripel verglichen um die Qualität der Extrapolation zu bewerten, z.B. durch SSIM.

Die umgesetzte Lösung wird ausführlich evaluiert. Zentraler Wert ist hierbei die Bildqualität nach der Extrapolation abhängig vom Winkelunterschied zwischen den Kameraeinstellungen und den Parametern der Vereinfachung der Tiefennetzgeometrie. Hierfür werden Tripel-Paare aus den Datensätzen und Variationen der Parameter der Algorithmen systematisch und automatisiert vermessen. Untersuchungen zum Laufzeitverhalten der Netzoptimierung im Server und der Bildextrapolation im Klienten sind optional durchzuführen.

---

## Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die von mir am heutigen Tag dem Prüfungsausschuss der Fakultät Informatik eingereichte Arbeit zum Thema:

*Optimierung und Übertragung von Tiefengeometrie für Remote-Visualisierung*

vollkommen selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Dresden, den 24. November 2016

Josef Schulz

---

## **Kurzfassung**

Zusammenfassung Text Deutsch

## **Abstract**

abstract text english

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
<b>2</b>	<b>Verwandte Arbeiten</b>	<b>4</b>
<b>3</b>	<b>Grundlagen</b>	<b>7</b>
3.1	Datensätze . . . . .	7
3.2	Extrapolation . . . . .	8
3.3	Delaunay-Triangulierung . . . . .	10
3.4	PSNR . . . . .	11
3.5	SSIM . . . . .	11
<b>4</b>	<b>Methodik</b>	<b>15</b>
4.1	Vollvernetzung . . . . .	15
4.2	Delaunay-Triangulierung . . . . .	16
4.2.1	Quadtree . . . . .	18
4.2.2	Floyd-Steinberg . . . . .	20
4.2.3	Erzeugung des Netzes . . . . .	22
4.2.4	Netzoptimierung . . . . .	23
<b>5</b>	<b>Implementierung</b>	<b>27</b>
5.0.5	16 Bit . . . . .	27
<b>6</b>	<b>Ergebnisse</b>	<b>29</b>
<b>7</b>	<b>Diskussion</b>	<b>35</b>
<b>8</b>	<b>Zusammenfassung</b>	<b>36</b>
<b>9</b>	<b>Ausblick</b>	<b>37</b>

# 1 Einleitung

Bei der Remote-Visualisierung, wird die Bildsynthese und die eigentliche Darstellung voneinander getrennt. Der Server-Prozess erzeugt und kodiert jedes Bild zu einem kompakten Datenpaket, welches an den Klient-Prozess gesendet wird. Der Klient empfängt und dekodiert das Datenpaket und gibt das Bild auf einem Bildschirm aus.

Remote-Visualisierung, ist ein insbesondere für mobile Endgeräte interessantes Konzept, weil es die Visualisierung von komplexen Szenen auch auf Leistungsarmen Geräten ermöglicht. Neben Computerspielen, ist die wissenschaftliche Visualisierung ein wichtiges Anwendungsgebiet, da Datensätze Größenordnungen erreichen können, die den Speicher herkömmlicher Desktops, Laptops, Smartphones etc. bei weitem übersteigen. Auch wenn ausreichend Speicher zur Verfügung steht, kann die Übertragung dieser Daten viel Zeit in Anspruch nehmen.

Mit Hilfe der Remote-Visualisierung ist es möglich, dass der Server die Bildsynthese übernimmt und nicht der komplette Datensatz übertragen werden muss. Ein weiterer Vorteil der mit leistungsstarken Serversystemen einhergeht ist der, dass sich komplexe Visualisierungs- und Beleuchtungsmethoden verwenden lassen, die mit normalen Endgeräten nicht zu realisieren sind.

Die Latenz bezeichnet in der Netzwerktechnik die Übertragungszeit von einem zum anderen Gerät. Diese ist in modernen Netzwerken gering, für die Interaktive Visualisierung allerdings immer noch zu groß um eine für den Menschen nicht wahrnehmbare Verzögerungszeit und damit eine interaktive Visualisierung zu gewährleisten. Ein möglicher Ausweg besteht darin, dass der Klient-Prozess ein bereits empfangenes Bild extrapoliert. Das bedeutet, dass das noch nicht empfangene Bild aus den vorhandenen Informationen approximiert wird, wodurch die Bildwiederholfrequenz akzeptabel bleibt.

Bei der Bildsynthese des Server-Prozesses lässt sich, aus dem Tiefenpuffer ein Tiefenbild erzeugen, zusätzlich zum Farbbild. Geometrisch entspricht das Tiefenbild einer 2.5D Ansicht der Szene. Wird es zum Klient gesendet, kann dieser das Tiefenbild exemplarisch als Punktwolke interpretieren. Das approximierte Bild entsteht, indem die Punktwolke aus einer neuen Kameraperspektive gezeichnet wird. Mit zunehmenden Winkelunterschied, zwischen der alten und der neuen Kameraperspektive, werden die Lücken zwischen den Punkten, der Punktwolke, immer

größer. Durch die Erzeugung eines Dreiecksnetzes, aus dem Tiefenbild, lassen sich die Lücken schließen. Das Farbbild wird dabei als Textur über das Netz gelegt.

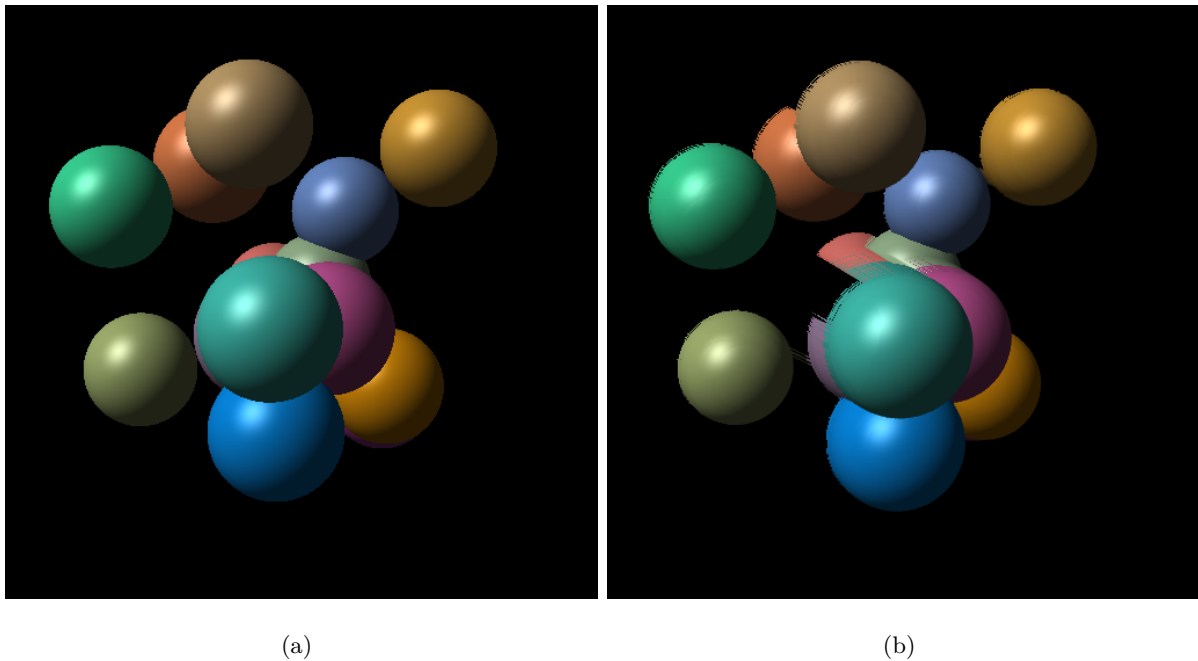


Abbildung 1.1: Die Abbildungen zeigt die Extrapolation des ersten Bildes a, aus dem Datensatz *TestSpheres*. Der Winkelunterschied zum Bild b beträgt 8 Grad.

Die Zentrale Aufgabe dieser Arbeit, ist die Optimierung und die Übertragung der Tiefeninformationen, vom Server zum Klient. Zu diesem Zweck erzeugt der Server, aus dem Tiefenbild ein adaptives Dreiecksnetz. Dieses wird an den Klient übertragen und extrapoliert. Mit Ground-Truth Datensätzen werden, in Abhängigkeit des Kamerawinkels, verschiedene Ansätze hinsichtlich ihrer Güte evaluiert. Um die Algorithmen zu testen wurde eine Server- und eine Klient-Komponente entwickelt. Beide Komponenten tauschen Informationen über das auf TCP basierende WebSocket-Protokoll aus. Der Klient ist ein Browser basierter Web-Klient, damit die implementierten Algorithmen mit den Einschränkungen durch JavaScript und WebGL evaluiert werden können.



## 2 Verwandte Arbeiten

Einen Überblick über Architekturen und Methoden der interaktiven Remote-Visualisierung geben Shu Shi *et al.* [?]. Zum zentralen Problem ihrer Arbeit, wird die Latenz und die effiziente Übertragung der Daten vom Server zum Klient. Lösungen hängen vom Anwendungsfall ab. In *THIN*-Systemen besteht die Aufgabe in der Übertragung von 2D Informationen, mit denen sich zum Beispiel Desktop-Anwendungen Fernsteuern lassen. Beispiele sind *SLIM* [?] und *THiNC* [?]. Sie sind für die Übertragung von 2D-Daten optimiert und profitieren davon, dass die meisten Änderungen nur Teilbereiche der eigentlichen Oberfläche betreffen. Ein anderes extrem der Remote-Visualisierung ist die Aufteilung der Bildsynthese, auf verschiedene Host-Systeme wie bei WireGL [?]. Jin Zhefan stellt in diesem Kontext Kompressionsmethoden für Zeicheninstruktionen, Vektoren, Normalen und Texturinformationen vor [?]. Peter Eisert und Philipp Fechteler haben ein Remote-Visualisierungssystem für Computerspiele entwickelt [?]. Ihr System überträgt bei kleinen Auflösungen, für Endgeräte ohne GPU die Bilder kodiert mit h264 oder wahlweise mpeg4. Bei größeren Auflösungen werden Zeicheninstruktionen gesendet und die Bilder werden vom Klienten synthetisiert. Auch wenn ihr System ausschließlich für den Einsatz in lokalen Netzwerken konzipiert wurde, ist die Latenz zu hoch um vom Benutzer nicht registriert zu werden. Mit Hilfe einer Bildextrapolation durch den Klienten kann die Latenz reduziert werden. Zu diesen Zweck schätzen Shu *et al.* in ihrer Arbeit [?], zusätzliche Referenz-Kamerapositionen und erzeugen zusätzlich zum Farbbild der Originalen Kameraposition weitere Tiefenbilder. Durch Warping, werden die Farbinformationen, unter Zuhilfenahme der Tiefenbilder in die neue Kameraperspektive überführt. Durch Verdeckung bedingte Lücken lassen sich durch die Referenztiefenbilder schließen. Das Farbbild wird mit JPEG komprimiert und die Tiefenbilder mit ZLIB. Ein ähnlicher Ansatz wurde im VR-Bereich von Smit *et al.* erprobt [?]. Auch in diesem System werden Lücken mit Hilfe von Tiefenbildern aus Referenz-Kamerapositionen geschlossen. Die Wahl des Warping-Algorithmus ist entscheidend für die Performance der Klient-Komponente, neben der im VR-System zum Einsatz kommenden Variante [?], bieten Mark *et al.* eine Übersicht über verschiedene mögliche Verfahren [?].

Von der Kameraperspektive abhängige betrachtung...

Die Übertragung und Visualisierung von Tiefenbildern spielt auch in der Visualisierung von Sensordaten eine wichtige Rolle. Palomo *et al.* beschäftigen sich mit der Visualisierung und Übertragung, von Tiefenbildern [?], die von Tiefensensoren erhoben werden. Sie kombinieren mehrerer aus verschiedenen Blickwinkeln aufgenommenen Bilder mit Hilfe eines Tiefenpuffers zu einem Gesamtbild und vermindern auf diese Weise entstehende Artefakte im Zusammenhang von Verdeckung bei der Visualisierung der Daten aus einer anderen Perspektive.

Um die Latenz bei der Übertragung von Tiefeninformation zu verbessern, können die Daten auch progressiv zum Klient gesendet werden. Evans *et al.* haben für diesen Zweck ein Datenformat für Punktwolken entwickelt und für den Einsatz mit WebGL optimiert [?].

Banno *et al.* erzeugen mit Hilfe der Delaunay-Triangulierung adaptive Dreiecksnetze aus Tiefenbildern, um die Qualität der Darstellung zu verbessern [?].

Lee *et al.*

Eine Vergleichbaren Ansatz wird von untersucht.

Wessels *et al.* stellen eine Konzeption für den Programmaufbau eines interaktiven Remote-Visualisierungssystem basierend auf dem WebSocket-Protokoll vor [?]. In ihrem System besteht der Server-Prozess aus zwei Hauptkomponenten, der Visualisierungs-Engine und dem Daemon. Während die Visualisierungs-Engine für die Bildsynthese zuständig ist, übernimmt der Daemon die Kommunikation mit dem Klient-Prozess. Der Klient schickt dabei seine Eingabeinformationen von Maus und Tastatur direkt an den Server. Dieser wertet die Daten aus und erzeugt darauf hin ein mit JPEG komprimiertes Bild, das mit Base64 kodiert wird und schließlich an den Klient-Prozess geschickt wird. Dieser kann das Bild nativ mit Hilfe eines HTML5 Canvas dekodieren und darstellen. Ihr System wird zur Grundlage dieser Arbeit.

Das erzeugte Tiefenbild lässt sich Pauly et al. [?]

Simon Stegmaier [?] A Generic Solution for Hardware-Accelerated Remote Visualization

Gabriel Taubin und Jarek Rossignac haben ein Algorithmus zur Erzeugung und effizienten Kodierung von Dreiecksstreifen aus Dreiecksnetzen entwickelt [?]. Dazu konstruiert ihr Algorithmus Spannbäume über dem Netz, die zur Erzeugung möglichst großer Dreiecksstreifen genutzt werden. Die Kompression kann wahlweise verlustfrei oder verlustbehaftet durchgeführt werden. Typische Kompressionsraten werden mit 1:50 angegeben.

Eine weitere Arbeit die sich mit der Kompression von Dreiecksnetzen und einer kompakten Repräsentation von diesen beschäftigt wurde Stefan Gumhold und Wolfgang Straßer geschrieben [?]. Kompression und Dekompression sind echtzeitfähig.

Michael Deering hat ebenfalls ein Geometrisches Kompressionsverfahren entwickelt. [?]

Federico Ponchio und Matteo Dellepiane [?] Fast decompression for web-based view-dependent 3D rendering

Diplomarbeit mädcheninformatiker Effiziente Datenübertragung von Modellen und Texturen für die Verwendung in WebGL Stefan Wagner noch kein cite

WEB-BASED VISUALISATION OF ON-SET POINT CLOUD DATA Alun Evans et al [?]

planarer scheiß [?]

## 3 Grundlagen

In diesem Kapitel werden die Datensätze im Detail vorgestellt und es wird gezeigt, wie die Bildextrapolation anhand der zur Verfügung stehenden Informationen durchgeführt wird. Anschließend werden die zwei Metriken PSNR und SSIM zum Vergleich von Bildern vorgestellt, die für die Evaluation verwendet werden.

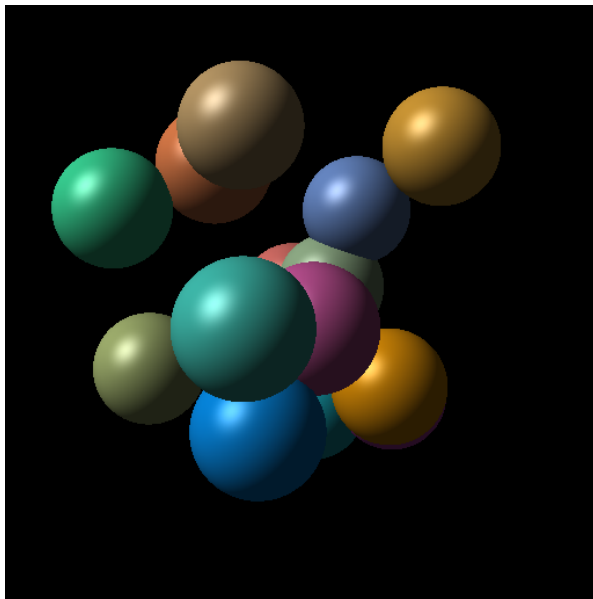
### 3.1 Datensätze

Zur Analyse der verwendeten Methoden stehen zwei Datensätze zur Verfügung. Jeder Datensatz besteht aus einer Sequenz von Tripeln. Dabei setzt sich jedes Tripel aus einem Farbbild  $I$ , einem Tiefenbild  $D$  sowie einem Satz von Kameraparametern zusammen. Die Auflösung der Farb- und Tiefenbilder beträgt für alle Datensätze  $w \times h = 512 \times 512$ . Die Tiefenbilder wurden mit 16 Bit und die Farbbilder mit 24 Bit quantisiert, dabei entfallen jeweils 8 Bit auf die einzelnen Farbkanäle. In der Abbildung 3.1 werden Farb- und Tiefenbild des ersten Tripels aus dem Datensatz *TestSpheres* dargestellt.

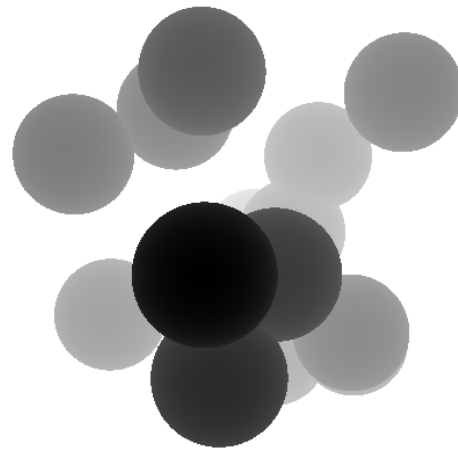
Die Kameraparameter setzen sich aus einem Vektor für die Kameraposition, einem *lookAt*-Vektor, der Punkt, auf den die Kamera schaut und dem *up*-Vektor zusammen. Für die Projektion werden die Positionen der Clippingebenen durch die Werte  $z_{near}$  und  $z_{far}$  beschrieben. Des Weiteren wird der Öffnungswinkel der Kamera benötigt.

Von beiden Szenen stehen 493 Bilder zur Verfügung. Beide Datensätze lassen sich in Abhängigkeit der Winkelschrittweite in 3 Sequenzen unterteilen. Der Winkel gibt dabei den Unterschied zwischen der Kameraposition aus dem ersten und dem gerade betrachteten Frame an. Eine Auflistung der Sequenzen zeigt die Tabelle 3.1.

Die Datensätze *CoolRandom* und *TestSpheres* wurden aus Partikeldatensätzen erzeugt. *TestSpheres* ist ein synthetisch erzeugter Datensatz, der aus wenigen Partikeln geniert wurde. Bei dem Datensatz *CoolRandom* handelt es sich dagegen um einen komplexen Datensatz, der mit Molekül-Datensätzen vergleichbar ist, welche in existierenden Anwendungen visualisiert werden.



(a)



(b)

Abbildung 3.1: Farb und Tiefenbild aus dem Datensatz *TestSpheres*

#	Anzahl Bilder	min Winkel	max Winkel	Winkel Schritt
1	241	0	5	0.25
2	60	6	10	1
3	192	15	90	5

Tabelle 3.1: Die Tabelle zeigt eine Übersicht über die Winkel der aufgenommenen Sequenzen.

## 3.2 Extrapolation

Eine einfache Möglichkeit, um die Extrapolation durchzuführen, besteht darin, das Tiefenbild als Punktwolke zu interpretieren und diese aus einer neuen Kameraperspektive zu zeichnen. Dazu wird aus jedem Pixel  $d \in D$  ein Vertex erzeugt. Mit Hilfe der Rückprojektion lassen sich die Vertices in die gewünschten Positionen transformieren.

Um das Prinzip der Rücktransformation zu erläutern, soll zunächst betrachtet werden, wie ein Vertex  $v$  aus dem Modellkoordinatensystem in normalisierte Gerätekoordinaten transformiert wird. Die Abbildung 3.2 verdeutlicht diesen Vorgang.

Der Vertex  $v = (x, y, z, 1)^T$  liegt zunächst in homogenen Modellkoordinaten vor. Mit Hilfe der Modellmatrix  $M = RT$ , die aus einer Rotations- und einer Translationsmatrix besteht, lässt sich der Vektor in das Weltkoordinatensystem überführen. Die Transformation in das Kamerakoordinatensystem wird durch die Multiplikation der Kameramatrix  $V$  mit dem Vektor  $v$

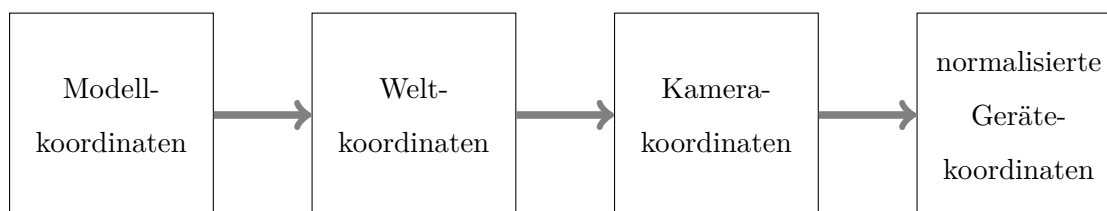


Abbildung 3.2: Die Abbildung zeigt die Teilschritte der Transformation eines Vertices aus den Modellkoordinaten in Bildschirmkoordinaten.

berechnet. Schließlich kann die Projektion aus dem Kamerakoordinatensystem in normalisierte Gerätekoordinaten mit Hilfe der Projektionsmatrix  $P$  berechnet werden. In der Gleichung 3.1 wird die Transformation von dem Vertex  $v$  aus den Modellkoordinaten in die normalisierten Gerätekoordinaten  $v'$  zusammengefasst:

$$v' = M \cdot V \cdot P \cdot v. \quad (3.1)$$

Normalisierte Gerätekoordinaten haben für die x-, y- und z-Komponente den Wertebereich von  $-1$  bis  $1$ . Diese lassen sich in Bildschirmkoordinaten umrechnen, indem die drei Komponenten in das Intervall  $[0, 1]$  übersetzt werden. Anschließend werden die x- und die y-Komponenten auf den Bildbereich gestreckt, indem sie mit  $w - 1$  und  $h - 1$  der gewünschten Auflösung  $w \times h$  multipliziert werden.

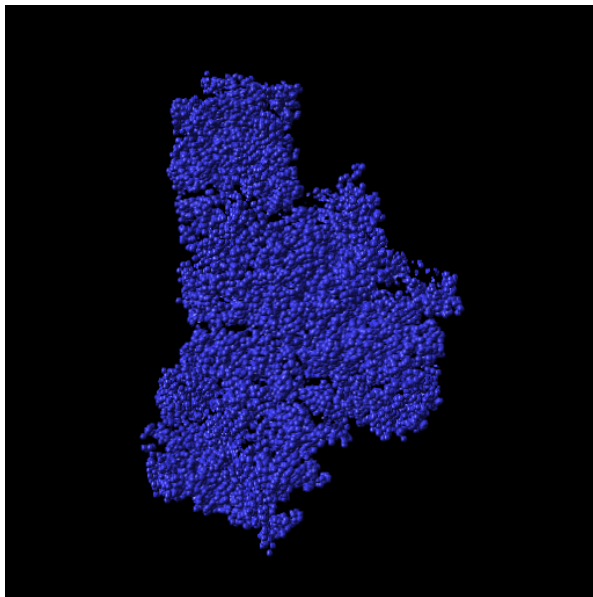
Im Folgenden wird die Menge der Tiefenbilder  $D_i$  und den dazugehörigen Transformationsmatrizen  $T_i$ , mit  $i \in \{0, 1, \dots, N\}$  betrachtet. Jede Transformationsmatrix  $T_i$  setzt sich dabei aus einer Modellmatrix  $M_i$ , einer Kameramatrix  $V_i$  und einer Projektionsmatrix  $P_i$  zusammen:

$$T_i = M_i \cdot V_i \cdot P_i. \quad (3.2)$$

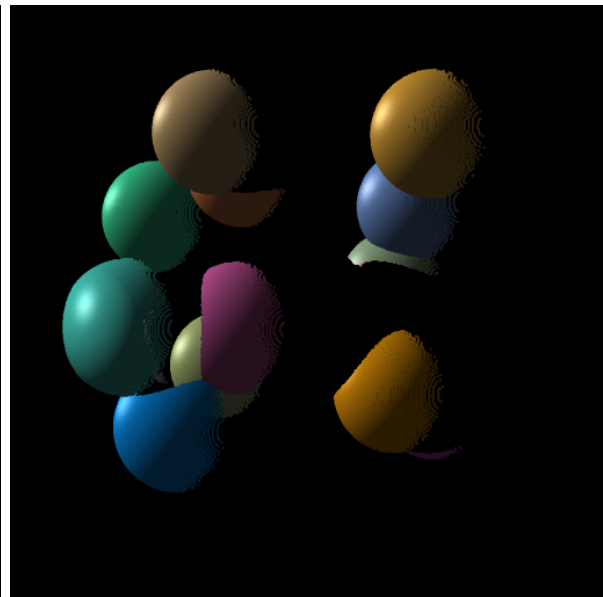
Um die Punktwolke des Tiefenbildes  $D_i$  mit der Transformationsmatrix  $T_j$  auf das Tiefenbild  $D_j$  abzubilden, muss zunächst für jeden Pixel  $d_i \in D_i$  ein Vertex  $v_i$  erzeugt werden, mit  $i, j \in \{0, 1, \dots, N\}$ . Zunächst liegt  $v_i$  in Bildschirmkoordinaten des Pixels  $d_i$  vor und muss in normalisierte Gerätekoordinaten übersetzt werden. Zu diesem Zweck werden zuerst die x- und y-Komponenten durch  $w - 1$  beziehungsweise  $h - 1$  geteilt. Anschließend müssen alle drei Komponenten des Vektors  $v_i$  aus dem Wertebereich  $[0, 1]$  in das Intervall  $[-1, 1]$  transformiert werden. Jetzt liegt  $v_i$  in normalisierten Gerätekoordinaten vor. Durch die Multiplikation der inversen Transformationsmatrix  $T_i^{-1}$  mit  $v_i$ , kann der Vertex in seine Modellkoordinaten überführt und im Anschluss mit Hilfe von  $T_j$  aus einer neuen Kameraperspektive gezeichnet werden:

$$v'_i = T_j \cdot T_i^{-1} v_i. \quad (3.3)$$

Die Interpretation der Tiefenbilder als Punktwolke ist keine optimale Lösung, da bei der Bildsynthese Lücken im Bild entstehen. Eine bessere Lösung ist es, aus dem Tiefenbild ein Dreiecksnetz zu erzeugen und dieses für die Bildsynthese zu verwenden. Mit Hilfe der Delaunay-Triangulierung ist es möglich aus Tiefenbildern adaptive Dreiecksnetze zu erzeugen, welche die Lücke füllen.



(a)



(b)

### 3.3 Delaunay-Triangulierung

Die Delaunay-Triangulierung ist ein Verfahren, um ein Dreiecksnetz aus einer Menge von Punkten  $p \in \mathbb{R}^2$  zu erzeugen. Dabei wird für jedes Dreieck ein Umkreis erzeugt, innerhalb dessen keine Punkte eines anderen Dreiecks enthalten sein dürfen. Jedes Dreieck des zu erzeugenden Netzes muss diese Bedingung erfüllen. Das Resultat dieser Forderung ist die maximierte Innenwinkelsumme aller Dreiecke. Für eine gegebene Punktmenge ist die Lösung nicht eindeutig, es kann verschiedene Netzkonfigurationen geben, welche die Forderung erfüllen.

Es existieren verschiedene Algorithmen die Delaunay-Triangulierung durchzuführen. Die besten erreichen eine Laufzeit von  $O(n \log n)$  und sind damit für den Einsatz in Echtzeitanwendungen tauglich. Beispiele sind der Sweep-Algorithmus und die inkrementelle Konstruktion.

### 3.4 PSNR

Die Abkürzung PSNR in Englisch *Peak signal-to-noise ratio*, gibt das Verhältnis zwischen dem *Peak signal-to-noise ratio*, kurz PSNR, gibt das Verhältnis zwischen dem Maximalwert und der maximalen Störung an. Da die meisten Signale sehr große Skalen haben, wird der PSNR häufig mittels einer logarithmischen Skala angegeben.

Der PSNR wird zur Messung der Qualität von nicht verlustfreien Kompressionsalgorithmen verwendet. Dazu wird das Originalbild als Signal interpretiert und der Fehler, der durch die Kompression eingeführt wird als Rauschen. Ein größerer PSNR-Wert bedeutet eine bessere Qualität des dekomprimierten Bildes.

Bei einer Farbtiefe von 8 Bit pro Kanal, stehen Werte von 30 - 40 dB für ein geringes Störsignal.

Der *mean squared error*, kurz *MSE* summiert in einem Fenster der Größe  $m \times n$  die quadratischen Abstände zwischen dem Original und dem rekonstruierten Bild auf.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (3.4)$$

Beim PSNR wird der maximal mögliche Wert  $MAX_I$  mit dem *MSE* ins Verhältnis gesetzt:

$$PSNR = 10 \times \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (3.5)$$

$$= 20 \times \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (3.6)$$

$$= 20 \times \log_{10}(MAX_I) - 10 \times \log_{10}(MSE) \quad (3.7)$$

### 3.5 SSIM

Eine weitere Metrik, die zum Vergleich von Bildern eingesetzt wird, wurde von Wang *et al.* entwickelt [?]. Diese basiert auf der Idee, dass die Struktur der abgebildeten Objekte von Beleuchtung und Kontrast unabhängig ist.

Beleuchtung und Kontrast können im gesamten Bild variieren, aus diesem Grund werden beide Parameter lokal in einem Fenster bestimmt.

In ihrem System wird die Aufgabe, die Ähnlichkeit zu messen, in drei Teile unterteilt: Beleuchtung, Kontrast und Struktur. Als Erstes wird die Beleuchtung bestimmt. Wenn die Signale, wie



in diesem Fall diskret sind, lässt sich die mittlere Intensität, des Signals  $x$  mit dem Term

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.8)$$

berechnen. Die Vergleichsfunktion für die Beleuchtung zwischen zwei Signalen  $x$  und  $y$  wird  $l(x, y)$  und setzt die werte  $\mu_x$  und  $\mu_y$  wie folgt ins Verhältnis:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}. \quad (3.9)$$

Als Nächstes wird die mittlere Intensität von den Signalen  $x$  und  $y$  abgezogen, so dass  $x' = x - \mu_x$  und  $y' = y - \mu_y$  entstehen.

Der Kontrast des Signals  $x$  wird mit Hilfe der Standardabweichung  $\sigma_x$  approximiert. In diskreter Form lässt sich diese mit der folgenden Formel berechnen:

$$\sigma_x = \left( \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}}. \quad (3.10)$$

Die Vergleichsfunktion für den Kontrast wird mit  $c(x, y)$  bezeichnet und definiert sich wie folgt:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \quad (3.11)$$

An dieser Stelle wird das Signal normalisiert, indem es durch seine eigene Standardabweichung geteilt wird, so dass die Signale  $x'' = x' - \mu_{x'}/\sigma_{x'}$  und  $y'' = y' - \mu_{y'}/\sigma_{y'}$  entstehen. Der Vergleich der strukturellen Eigenschaften  $s(x, y)$  wird mit den normalisierten Signalen  $x''$  und  $y''$  durchgeführt. Für den Vergleich der Struktur sollen folgende Eigenschaften gelten:

Erstens, die Funktion  $s(x, y)$  muss Symmetrisch sein  $s(x, y) = s(y, x)$ .

Zweitens soll die Funktion auf einen Wert kleiner oder gleich 1 beschränkt werden  $s(x, y) \leq 1$ .

Drittens es soll nur ein Maximum  $s(x, y) = 1$ , genau dann und nur dann, wenn gilt das  $x = y$ .

Eine Definition, die diese Forderungen erfüllt ist:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x + \sigma_y + C_3}. \quad (3.12)$$

Die Kovarianz  $\sigma_{xy}$  berechnet sich folgendermaßen:

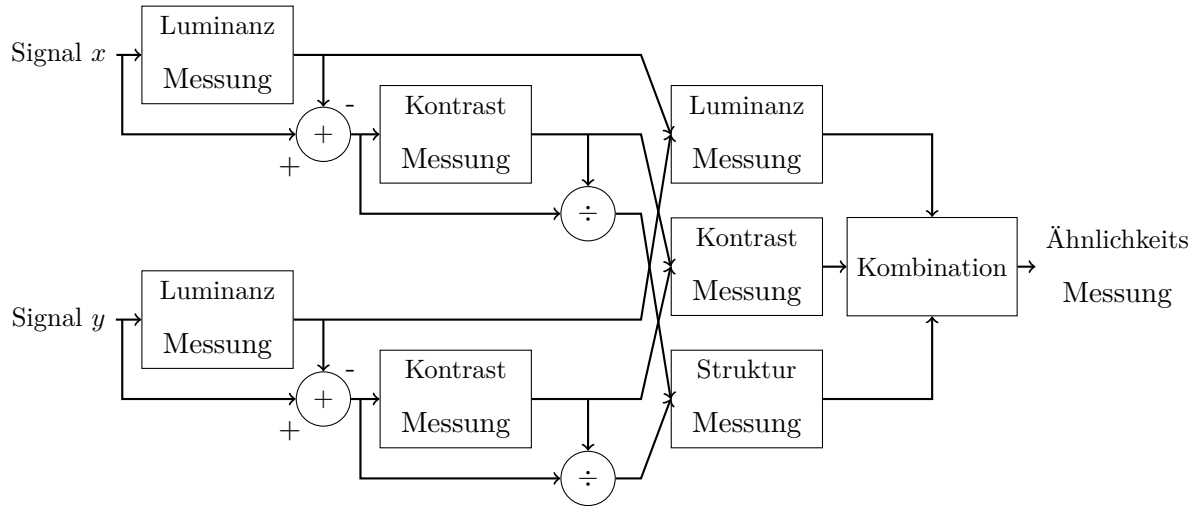


Abbildung 3.3: Das Modell zeigt das Vorgehen von dem SSIM-Algorithmus

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y). \quad (3.13)$$

Sie korreliert mit dem Kosinus des Winkels zwischen den beiden Vektoren  $x - \mu_x$  und  $y - \mu_y$ . Die Konstanten  $C_1, C_2$  und  $C_3$  in den drei Vergleichsfunktionen sorgen dafür, dass eine Division durch 0 nicht möglich wird, sollten die Werte in den Nennern zu klein werden. Letztlich kann die Gesamtqualität gemessen werden, indem Beleuchtung, Kontrast und Strukturvergleich kombiniert werden:

$$SSIM(x, y) = [l(x, y)]^\alpha \times [c(x, y)]^\beta \times [s(x, y)]^\gamma. \quad (3.14)$$

Mit den Parametern  $\alpha > 0, \beta > 0, \gamma > 0$ , kann die Gewichtung zwischen den Vergleichsfunktionen variiert werden. Im Rahmen dieser Arbeit gilt  $\alpha = \beta = \gamma = 1$  und es gilt für die Konstanten:  $C_3 = C_2/2$ , mit  $C_1 = 6.5025$  und  $C_2 = 58.5225$ . Der *SSIM* wird lokal berechnet, in Fenstern der Größe  $11 \times 11$ , dabei werden die Signalwerte  $x_i$  und  $y_i$  mit einer Gaussianfunktion gewichtet. Damit lassen sich die mittlere Intensität, die Standardabweichung und die Kovarianz zu folgenden Gleichungen umschreiben, wobei  $w_i$  die Gewichtung an dem Punkt  $i$  bezeichnet:

$$\mu_x = \sum_{i=1}^N w_i x_i \quad (3.15)$$

$$\sigma_x = \left( \sum_{i=1}^N w_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (3.16)$$

$$\sigma_{xy} = \sum_{i=1}^N w_i (x_i - \mu_x)(y_i - \mu_y). \quad (3.17)$$

Die Gesamtgleichung für den  $SSIM(x, y)$  lässt sich durch die Gleichung

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.18)$$

ausrechnen. Der  $SSIM(x, y)$  misst die Güte jedoch nur lokal, um eine Aussage über das gesamte Bild treffen zu können, kann der Mittelwert über allen  $x \in X$  und  $y \in Y$  berechnet werden:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j). \quad (3.19)$$

## 4 Methodik

In dieser Arbeit werden zwei grundsätzliche Verfahren untersucht, um die Tiefeninformationen zu übertragen und als Dreiecksnetz darzustellen. Der erste Ansatz besteht darin, das Tiefenbild als solches verlustfrei zu komprimieren und zu übertragen. Der Klient empfängt und dekodiert das Tiefenbild und erzeugt daraus ein voll vernetztes Dreiecksnetz. Im zweiten Ansatz wird aus dem Tiefenbild ein adaptives Dreiecksnetz mit Hilfe der Delaunay-Triangulierung vom Server konstruiert und an den Klienten übertragen.

Die Erzeugung von adaptiven Dreiecksnetzen lässt sich als eine Form der verlustbehafteten geometrischen Kompression bezeichnen. Um die Anzahl der zu übertragenden Informationen weiter zu reduzieren, kann das erzeugte Dreiecksnetz in Dreiecksstreifen zerlegt und sehr kompakt kodiert werden. Durch Valenz-basierte Kodierung lässt sich diese Informationsmenge weiter reduzieren.

Weil die Tiefenwerte des Tiefenbildes komplett und ohne Informationsverlust im voll vernetzten Dreiecksnetz enthalten sind, liefert dieses Verfahren Ergebnisse, die als *Ground-Truth* Information zum Vergleich für die adaptive Vernetzung dienen.

### 4.1 Vollvernetzung

Bei der Vollvernetzung wird für jeden Pixel aus dem Tiefenbild  $D$  ein Vertex erzeugt. Die entstandenen Vertices werden über Kanten zu Dreiecken miteinander verbunden. Die Abbildung 4.1 zeigt drei unterschiedliche Varianten, wie sich die Vertices zu voll vernetzten Dreiecksnetzen verbinden lassen. Alle drei Varianten wurden implementiert, weisen bei entsprechend großen Auflösungen jedoch keine Unterschiede hinsichtlich der Qualität der Darstellung auf.

Die Anzahl der Vertices entspricht der Anzahl der Bildpunkte. Die x- und y-Komponenten lassen sich aus der bekannten Auflösung vorausberechnen und müssen im Falle einer Auflösungsänderung neu bestimmt werden. Für jeden Vertex wird die z-Komponente mit Hilfe der Pixelwerte aus dem Tiefenbild während der Bildsynthese gesetzt. Die Anzahl der Dreiecke lässt sich dabei wie folgt berechnen:  $2(w - 1)(h - 1)$ .

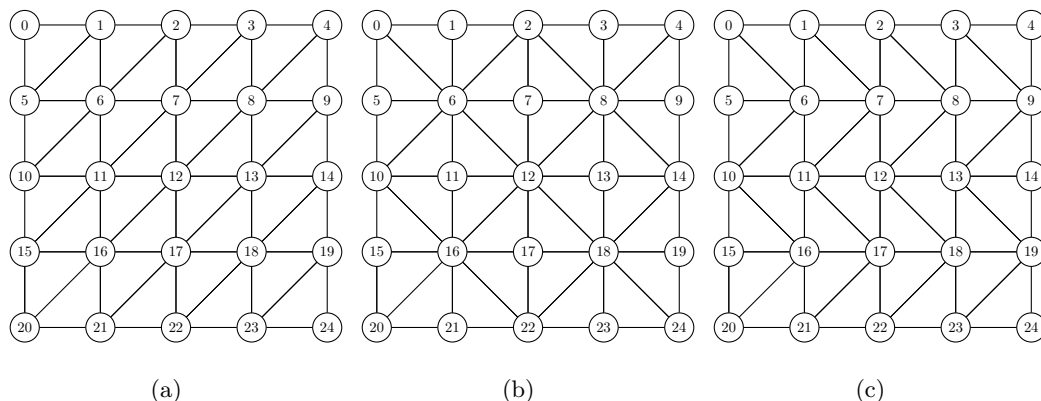


Abbildung 4.1: Darstellung von drei verschiedene Varianten der Vollvernetzung.

Problematisch bei diesem Verfahren ist, dass die Anzahl der Dreiecke von der Auflösung abhängt und der Rechenaufwand der klientseitigen Bildsynthese linear mit der Auflösung steigt.

## 4.2 Delaunay-Triangulierung

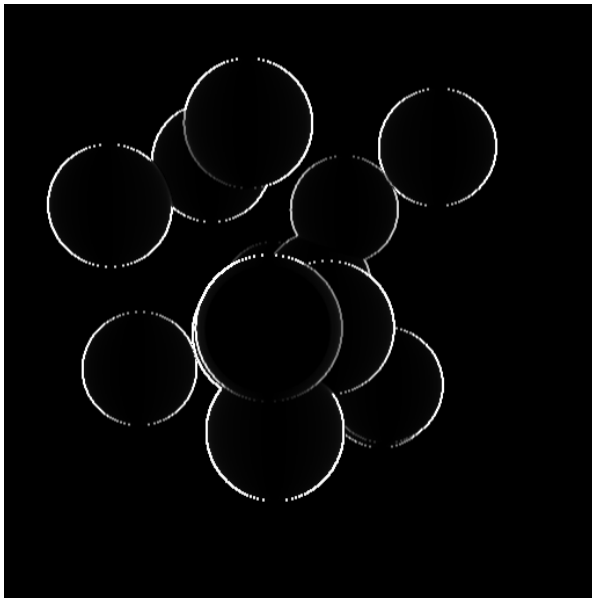
Die zweite in dieser Arbeit untersuchte Variante, besteht in der serverseitigen Erzeugung eines adaptives Dreiecksnetz aus einem Tiefenbild  $D$ . Prinzipiell, lässt sich die Vorgehensweise in drei Schritte unterteilen. Im ersten Schritt muss eine Menge von Punkten  $P \subset D$  definiert werden. Dabei muss die Menge  $P$  so gewählt werden, dass die Werteverläufe des Tiefenbildes möglichst optimal wiedergegeben werden. Der zweite Schritt besteht in der Vernetzung der Punkte  $p \in P$  zu Dreiecken, dass dabei Eingesetzte verfahren ist die Delaunay-Triangulierung. Zum Schluss kann das erzeugte Netz, in einem dritten Schritt, weiter verfeinert werden.

Entscheidend für das Ergebnisnetz und für die Geschwindigkeit des gesamten Verfahrens ist die Wahl der Punktmenge  $P$ . Das Ziel bei der Verteilung der Punktmenge  $P$  ist es, eine hohe Punktdichte im Bereich von Kanten und gekrümmten Flächen zu erzielen. Im Gegensatz dazu sollten planare Regionen möglichst keine Punkte enthalten, weil diese mit wenigen Dreiecken approximiert werden können. Ausgangspunkt für die Erzeugung der Punktmenge  $P$ , sind die Gradienten des Tiefenbildes  $\nabla_x$  und  $\nabla_y$ . Die Gradienten werden mit Hilfe des Sobel-Operators berechnet. Dazu wird die erste Ableitung berechnet und gleichzeitig wird die dazu orthogonale Richtung geglättet. Mit Hilfe einer Faltungsmatrix der Größe  $3 \times 3$  lassen sich die Gradienten berechnen:

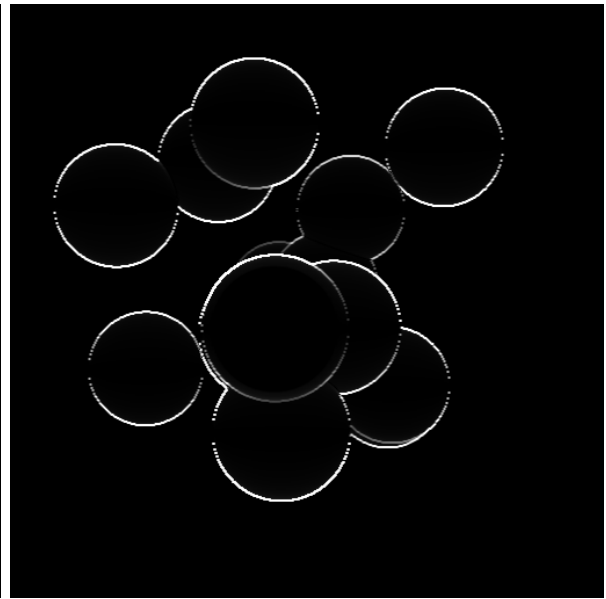
$$\nabla_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * D,$$

$$\nabla_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * D.$$

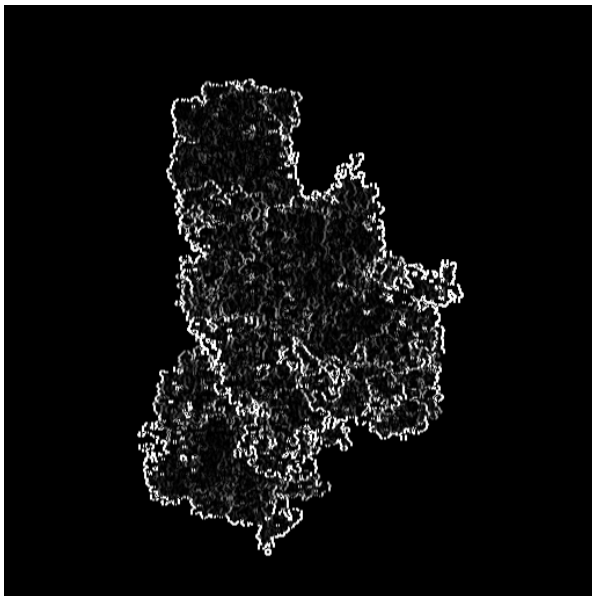
Auf diese Weise entstehen für ein Tiefenbild zwei Gradientenbilder, wie in Abbildung 4.2 dargestellt. Zur Erzeugung der Punktmenge  $P$  wurden zwei Verfahren untersucht, die im Folgenden näher betrachtet werden.



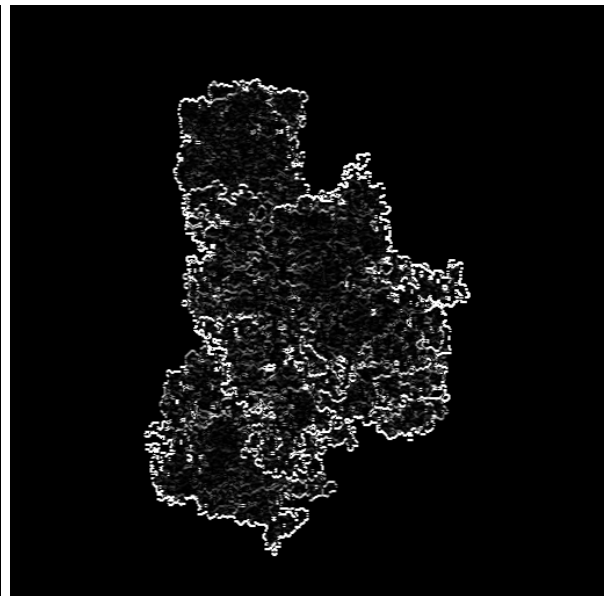
(a)



(b)



(c)



(d)

Abbildung 4.2: Gradientenbilder beider Datensätze vom jeweils ersten Tiefenbild.

### 4.2.1 Quadtree

Die erste in dieser Arbeit vorgestellte Methode, zur Erzeugung der Punktmenge  $P$ , basiert auf einer Quadtree-Datenstruktur. Ein Quadtree ist eine Baumdatenstruktur, in der die Anzahl der Kindknoten, eines Knotens auf vier Kindknoten beschränkt wird. Im Rahmen dieser Arbeit repräsentiert jeder Knoten eine rechteckige Region  $R$  des Tiefenbildes  $D$ . Eine Region  $R$  wird durch seine Eckpunkte parametrisiert. Um die Baumdatenstruktur zu erzeugen, wird ein Wurzelknoten definiert, dessen Region  $R = (0, 0, w - 1, h - 1)$  den Ausmaßen des Tiefenbildes  $D$  entspricht. Im Anschluss wird die Region des Wurzelknotens in vier gleichgroße Teilregionen aufgeteilt. Diese Teilregionen sind die Kindknoten des Wurzelknotens. Alle Kindknoten werden auf diese Weise sukzessiv weiter unterteilt, bis die Tiefe des erzeugten Baumes eine maximale Baumtiefe  $d_{max}$  erreicht. Die Kindknoten, der Baumtiefe  $d_{max}$  enthalten selbst keine weiteren Kindknoten und werden als Blattknoten bezeichnet, die Regionen die sie repräsentieren als Blattregionen. Die maximale Tiefe des Quadrees ist dabei invers proportional zur Breite der Blattregionen. Mit dem Parameter  $d_{max}$ , kann die Qualität und die Kompression des erzeugten Netzes, sowie der Berechnungsaufwand skaliert werden. Die erzeugte Baumdatenstruktur hängt von der Auflösung der Tiefenbilder ab und muss nur neu erzeugt werden, wenn sich die Auflösung oder der Parameter  $d_{max}$  ändert.

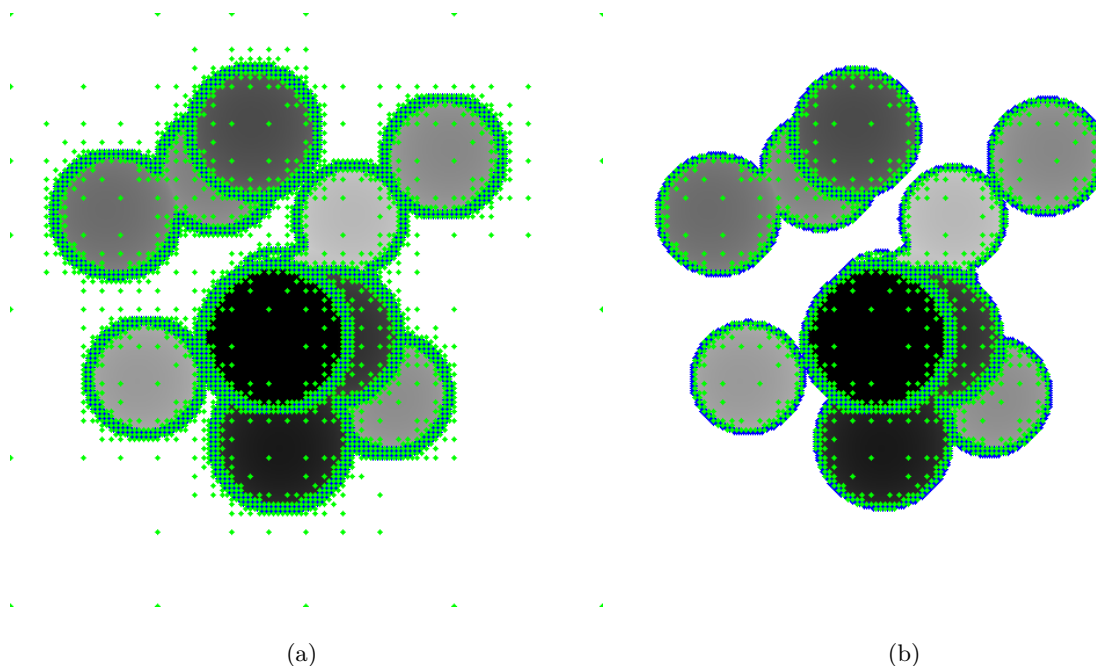
Im nächsten Schritt wird mit Hilfe der Baumdatenstruktur die Punktmenge  $P$  erzeugt. Zu diesem Zweck wird der Baum, beginnend mit den Blattknoten, zum Wurzelknoten traversiert. Mit Hilfe der Gradienten kann die Region  $R$  eines Knoten auf planarität getestet werden. Ist eine Region  $R$  nicht planar, dann werden die Eckpunkte von  $R$  zur Menge  $P$  hinzugefügt.

Um zu überprüfen, ob eine Region  $R$  als planar bezeichnet werden kann, lässt sich mit Hilfe der Differenz, zwischen dem maximalen und dem minimalen Wert der Gradienten innerhalb von  $R$  bestimmen:

$$c_x = \max_R \nabla_x - \min_R \nabla_x \quad (4.1)$$

$$c_y = \max_R \nabla_y - \min_R \nabla_y. \quad (4.2)$$

Wenn entweder  $c_x$  oder  $c_y$  größer als ein Schwellwert ist, dann ist die Region  $R$  nicht planar und die Eckpunkte werden der Menge  $P$  hinzugefügt. Die Werte für  $c_x$  und  $c_y$  der inneren Knoten lassen sich effizient aus den bereits berechneten maximalen und minimalen Gradienten seiner Kindknoten berechnen.

Abbildung 4.3: Szene *TestSpheres*, Baumtiefe 9,  $l=5$ ,  $i=4$ 

Die Wahl des Schwellwerts ist entscheidend für die Anzahl der eingefügten Punkte in  $P$ . Dieser kann adaptiv gewählt werden, um die Komplexität des Dreiecksnetzes anzupassen an die zur Verfügung stehende Bandbreite anzupassen.

Für den eigentlichen Planaritäts-Test unterscheidet sich dieser Schwellwert, in Abhängigkeit, ob es sich um einen Inneren oder einen Blattknoten handelt. Der Schwellwert für die Blattknoten wird mit  $l$  bezeichnet und der für die inneren Knoten mit  $i$ . Blattknoten sollten vorrangig Tiefenuntersprünge, Kanten detektieren und somit die Kontur der dargestellten Objekte. Innere Knoten dagegen, repräsentieren die Oberfläche der Objekte und werden gesetzt, um deren Verläufe von nicht planaren Regionen abzubilden zu können. Weil Tiefensprünge größere Gradienten bedeuten als gekrümmte Oberflächen, sollte der Schwellwert  $i$ , für die inneren Knoten kleiner sein, als der Schwellwert  $l$  für die Blattknoten,  $i < l$ . Die Werte von  $i$  und  $l$  liegen im Intervall  $[0, 1]$ .

Das Ergebnis der Traversierung des Baums ist eine Menge von Punkten  $P$ , die sich aus den Eckpunkten der getesteten Regionen zusammensetzt, in denen die Gradientendifferenzen den entsprechenden Schwellwert überschreitet.

Diese Punktmenge  $P$  kann weiter reduziert werden, indem getestet wird, ob ein Punkt  $p \in P$  zum Hintergrund gehört. Das ist der Fall, wenn der Tiefenwert des Tiefenbildes an stelle  $p$  dem maximalen Tiefenwert entspricht. Zur Erinnerung die Tiefenwerte liegen linear zwischen den



beiden Clippingebenen, wobei ein großer Wert bedeutet, dass dieser nahe der Bildebene liegt.

### 4.2.2 Floyd-Steinberg

Der Zweite in dieser Arbeit vorgestellte Ansatz zur Erzeugung der Punktmenge  $P$ , basiert auf einem Fehler-Diffusionsverfahren. In diesem Abschnitt wird die Bildfunktion  $f(x)$  über der zweidimensionalen Bilddomain definiert. Das bedeutet, dass die Tiefenwerte des Bildes  $D$  mit der Funktion  $f$  angegeben werden. Im ersten Schritt wird aus den Gradientenbildern ein Merkmalsbild  $\sigma$  erzeugt:

$$\sigma(x) = \left( \frac{\|\nabla f(x)\|}{A} \right)^\gamma. \quad (4.3)$$

Dabei entspricht der Nenner von 4.3 der Länge des Gradientenvektors, an der Stelle  $x$ . Dieser wird mit  $A$ , dem größtmöglichen Wert von  $\|\nabla f(x)\|$  normiert. Der Parameter  $\gamma \in [0, 1]$  wird genutzt, um die Ausprägung von schwachen Kanten im Merkmalsbild  $\sigma$  zu erhöhen.

Die beiden Gradientenbilder  $\nabla_x$  und  $\nabla_y$  werden durch die Gleichung 4.3 auf ein einzelnes Merkmalsbild  $\sigma$  abgebildet. Um aus dem Merkmalsbild  $\sigma$  die Punktmenge  $P$  zu erzeugen, kommt der Floyd-Steinberg Algorithmus zum Einsatz. Bei dem Floyd-Steinberg Algorithmus handelt es sich, um ein einfaches und effizientes Fehler-Diffusion-Verfahren. Zu diesem Zweck wird das Merkmalsbild  $\sigma$  Pixel für Pixel Zeilenweise durchlaufen. Der betrachtete Pixel wird mit einem Schwellwert  $\delta$  verglichen. Wenn der Wert des Pixels größer als der Schwellwert ist, dann wird der Wert des Pixels mit  $\delta$ , andernfalls mit 0 ersetzt. Zusätzlich wird der neu gesetzte Wert und der daraus resultierende Quantisierungsfehler auf die benachbarten Pixel verteilt. Das folgende Codebeispiel 4.1 verdeutlicht das Verfahren:

```

1  for each y
2      for each x
3          oldpixel      := pixel[x][y]
4          newpixel      := (pixel[x][y] > δ) ? δ : 0
5          pixel[x][y]    := newpixel
6          quant_error    := oldpixel - newpixel
7          pixel[x+1][y]  := pixel[x+1][y] + quant_error * 7 / 16
8          pixel[x-1][y+1] := pixel[x-1][y+1] + quant_error * 3 / 16
9          pixel[x][y+1]  := pixel[x][y+1] + quant_error * 5 / 16
10         pixel[x+1][y+1] := pixel[x+1][y+1] + quant_error * 1 / 16

```

Listing 4.1: Floyd-Steinberg Algorithmus [?]

Immer wenn die Bindung erfüllt wird und der Wert eines Pixels auf  $\delta$  gesetzt wird, dann wird die Menge  $P$  mit dem gerade bearbeiteten Pixel erweitert. Die Abbildung 4.4 zeigt verschiedene bereits auf  $\delta$  und 0 reduzierte Merkmalsbilder  $\sigma$ .

Ein Vorteil dieses Verfahrens, ist die effizienten Erzeugung der Menge  $P$ . Mit Hilfe der beiden Parameter  $\gamma, \delta$  lässt sich die Anzahl der Punktmenge  $P$  variieren. Dabei spielt insbesondere  $\delta$  eine entscheidende Rolle, da durch diesen die Punktdichte angepasst werden kann. Wenn  $\delta$  kleine Werte annimmt, dann wird die Punktdichte erhöht und im Fall großen Werten reduziert.

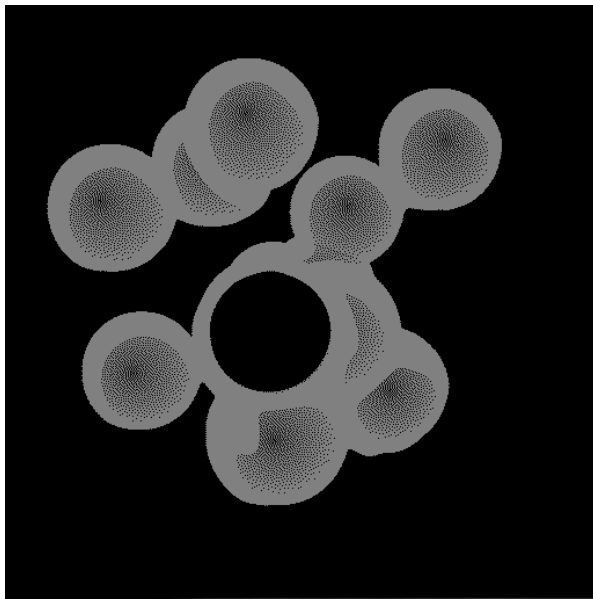
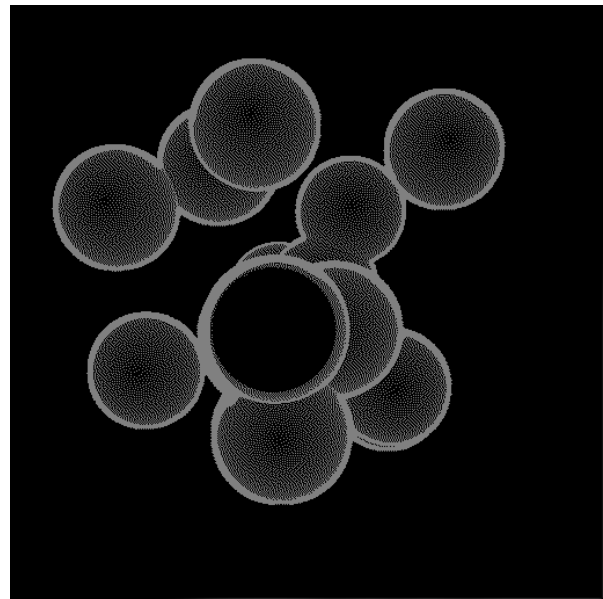
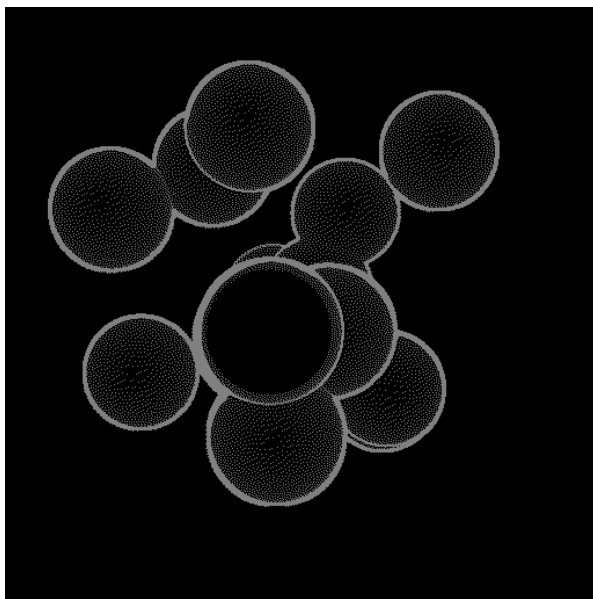
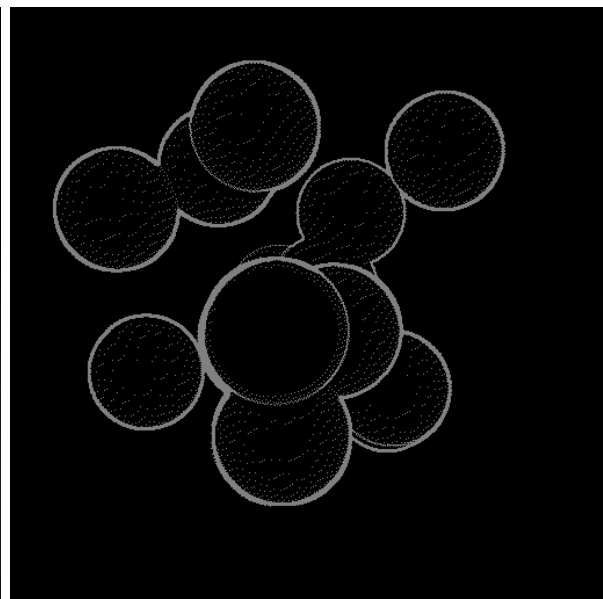
(a)  $\gamma = 0.2$ (b)  $\gamma = 0.4$ (c)  $\gamma = 0.6$ (d)  $\gamma = 1.0$ 

Abbildung 4.4: Die Bilder a bis d zeigen das Merkmalsbild, in Abhängigkeit des Parameters  $\gamma$  bei einem festgesetzten Schwellwert von 0.5.

### 4.2.3 Erzeugung des Netzes

Die aus den beiden erzeugten Methoden Punkte  $p \in P$  können jetzt mit Hilfe der Delaunay-Triangulierung zu einem Dreiecksnetz miteinander verbunden werden. In dieser Arbeit wird zu diesem Zweck die Delaunay-Triangulierung aus dem opencv-Framework verwendet.

#### 4.2.4 Netzooptimierung

Das auf diese Weise entstandene Netze können, in Abhängigkeit der Eingabedaten, zwei Arten von Artefakten enthalten. Zum einen können Dreiecke falsche Tiefenregionen approximieren, weil nur die Eckpunkte der jeweiligen Regionen  $R$  zur Konstruktion der Dreiecke genutzt werden. Zum anderen können Dreiecke über Tiefensprüngen liegen, wodurch Kanten nicht korrekt vom Netz abgebildet werden. Diese Art von Artefakte kommen zustande, wenn die Blattregionen zu große Bildregionen repräsentieren. Um die Artefakte zu reduzieren, werden nicht valide Dreiecke nochmals unterteilt oder verworfen.

Die Validität eines Dreiecks wird anhand seiner Kanten bestimmt. Eine Kante wird als nicht valide Kante bezeichnet, wenn die Tiefenwerte einiger von ihr überspannten Pixel abweicht, oder die 3D Richtung der Kante sich dem Lot der Bildebene nähert, während gleichzeitig eine bestimmte Länge in der  $xy$ -Ebene überschritten wird.

Die Eckpunkte der zu prüfenden Kante, werden im folgenden mit  $p_1$  und  $p_2$  bezeichnet. Der Punkt  $p_m$  bezeichnet dabei den Mittelpunkt der Kante  $p_1p_2$ .

$$\frac{d_{p_1} + d_{p_2}}{2} - d_{p_m} < \varepsilon \quad (4.4)$$

Eine Kante ist nicht valide, wenn der eigentliche Tiefenwert  $D(p_m)$  In der Gleichung 4.5 approximiert der erste Term das Verhältnis, zwischen dem Betrag des Tiefenunterschieds von  $p_1$  und  $p_2$  zu der Länge der Kante projiziert auf die  $xy$ -Ebene.

$$\frac{|d_{p_1} - d_{p_2}|}{\|p_1 - p_2\| (d_{p_1} + d_{p_2})} < \alpha \quad (4.5)$$

Ist der Wert des ersten Terms der Gleichung 4.5 einer Kante größer als der Schwellwert  $T_{angle}$ , dann steht diese Kante nahezu senkrecht auf der Bildebene und sie liegt mit hoher Wahrscheinlichkeit über einen Tiefensprung. Eine Kante auf die das zutrifft wird als nicht valide bezeichnet. Enthält ein Dreieck mindestens zwei Kanten die valide sind, wird es direkt zum endgültigen Dreiecksnetz hinzugefügt. Wenn ein Dreieck dagegen mehr als zwei nicht valide Kanten enthält, wird es weiter unterteilt.

Zwei Bildpunkte  $p_1$  und  $p_2$  werden als verbindbar bezeichnet, wenn alle Bildpunkte zwischen  $p_1$  und  $p_2$  valide Tiefenwerte besitzen und ihre Tiefe sich zum größten Teil linear ändert. Um Rechenzeit zu sparen wird empfohlen nur den Median  $p_m$  zwischen  $p_1$  und  $p_2$  zu testen. Eine Strecke wird als Verbindbar bezeichnet, wenn sie die folgende Gleichung erfüllt:

$$| (d(p_2) - d(p_m)) - (d(p_m) - d(p_1)) | < \beta \quad (4.6)$$

.

Um ein Dreieck zu unterteilen muss eine Fallunterscheidung durchgeführt werden. Hat das Dreieck nur eine valide Kante  $p_1p_2$ , dann müssen die anderen beiden Kanten wie in der Abbildung 4.5 geteilt werden und es entstehen aus dem ursprünglichen Dreieck drei neue Dreiecke.

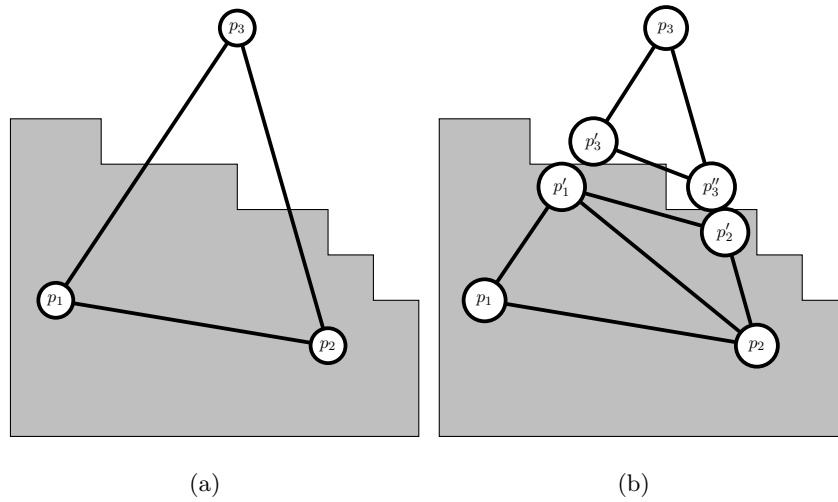


Abbildung 4.5: Die Bilder a und b zeigen die Unterteilung des Dreiecks  $p_1p_2p_3$ , wenn die zwei Kanten  $p_1p_3$ ,  $p_2p_3$  nicht valide sind.

Dazu werden vier neue Vertices iterativ entlang der alten Strecken eingefügt:

Der Vertex  $p'_1$  ist der von Punkt auf der Strecke  $p_3p_1$ , der am weitesten von  $p_1$  entfernt liegt und mit dem Punkten  $p_1$  und  $p'_2$  verbindbar ist.

Der Vertex  $p'_2$  ist der von Punkt auf der Strecke  $p_3p_2$ , der am weitesten von  $p_2$  entfernt liegt und mit dem Punkten  $p_2$  und  $p'_1$  verbindbar ist.

Der Vertex  $p'_3$  ist der von Punkt auf der Strecke  $p_3p_1$ , der am weitesten von  $p_3$  entfernt liegt und mit dem Punkt  $p_3$  verbindbar ist.

Der Vertex  $p''_3$  ist der von Punkt auf der Strecke  $p_3p_2$ , der am weitesten von  $p_3$  entfernt liegt und mit dem Punkt  $p_3$  verbindbar ist.

Anschließend wird das Dreieck  $p_3p'_3p''_3$  zum endgültigen Dreiecksnetz hinzugefügt. Um die anderen beiden Dreiecke einzufügen muss eine weitere Fallunterscheidung durchgeführt werden. Wenn die Strecke  $p_1p'_2$  kleiner ist als  $p_2p'_1$ , dann werden die Dreiecke  $p_1p_2p'_2$ ,  $p_1p'_2p'_1$  zu dem finalen Netz hinzugefügt, andernfalls die beiden Dreiecke  $p_1p_2p'_1$  und  $p_2p'_2p'_1$ .

In dem Fall, dass alle drei Kanten nicht valide sind, werden sechs neue Vertices eingefügt, die Abbildung 4.6 verdeutlicht diesen Fall. Die Berechnung aller Vertices geschieht analog zu dem Vertex  $p'_3$  aus der vorhergehenden Betrachtung mit einer validen Kante.

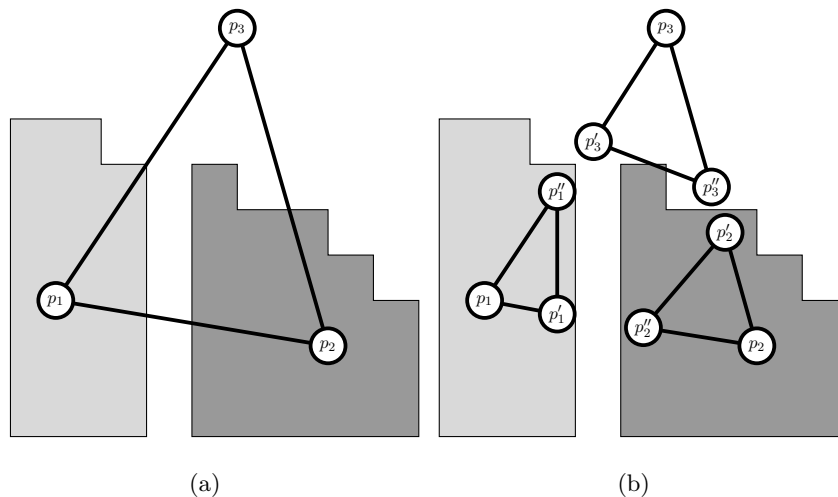


Abbildung 4.6: Die Bilder a und b zeigen die Unterteilung des Dreiecks  $p_1p_2p_3$ , wenn keine Kante valide ist.

Zu beachten ist, dass jedes neu erzeugte Dreieck auf Kollinearität zu überprüfen und gegebenenfalls zu verwerfen.

Um die Schwellwerte unabhängig von der Auflösung und Farbtiefe angeben zu können, werden die Streckenlängen in den Gleichungen 4.6... mit Texturkoordinaten berechnet und die Farbwerte der Pixel auf das Intervall  $[0, 1]$  transformiert.

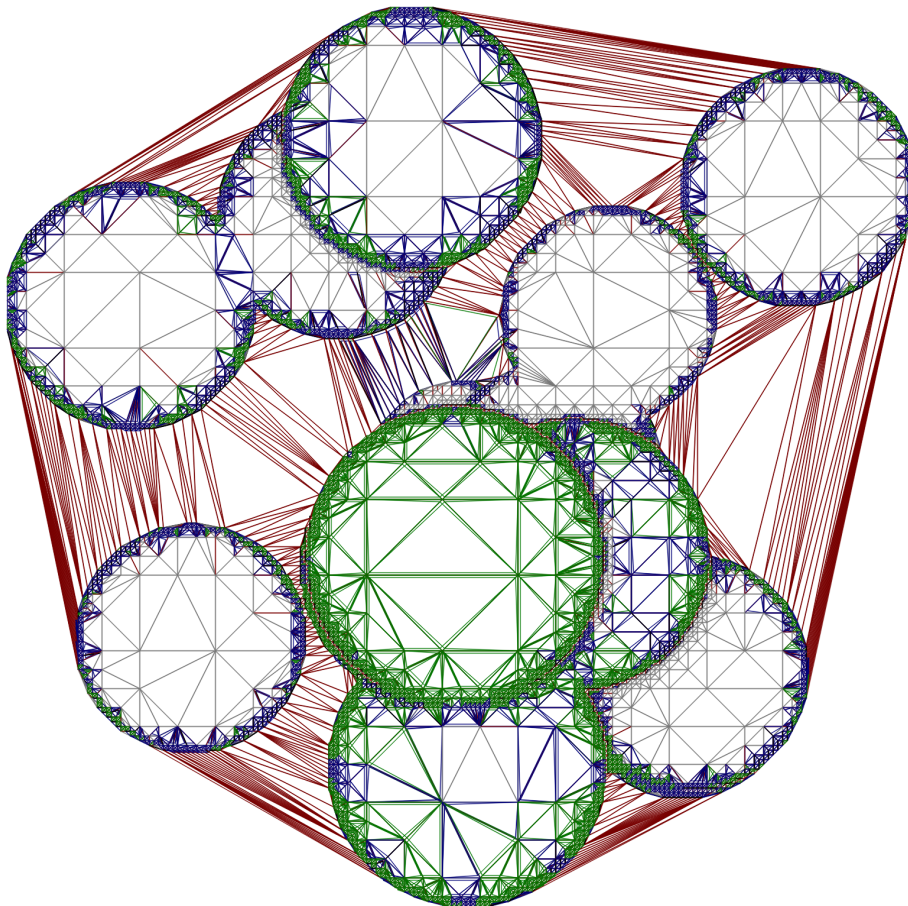
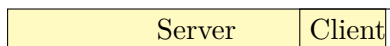


Abbildung 4.7: Nachbearbeitetes Netz

## 5 Implementierung

Die Klient-Anwendung ist eine Browser basierte Web-Anwendung, die mit dem Server über das WebSocket-Protokoll kommuniziert. Dabei handelt es sich um ein auf TCP basierendes Netzwerkprotokoll, das eine Bidirektionale Verbindung zwischen den Verbindungsteilnehmern erlaubt. Die klientseitige Extrapolation der Bilddaten wird mit Hilfe von WebGL durchgeführt. Bei WebGL handelt es sich um eine Bibliothek die von modernen Browsern zur Verfügung gestellt wird, um eine Hardware beschleunigte Bildsynthese zu ermöglichen. Der Vorteil dieser Technologie ist die Unabhängigkeit der Anwendung im Bezug, zur Plattform und dem Gerät, in Kombination mit einer hohen Rechengeschwindigkeit.

Beide Komponenten tauschen Informationen mit Hilfe des kompakten vom Menschen lesbaren JSON



### 5.0.5 16 Bit

Farb- und Tiefenbild werden mit base64 kodiert übertragen und lassen sich vom Browser nativ dekodieren. Die Farbtiefe pro Kanal ist Browserseitig auf 8 Bit beschränkt. Um 16 Bit Tiefeninformationen im Vertex-Shader nutzen zu können, müssen diese 16 Bit auf zwei 8 Bit Kanäle aufgeteilt werden. Die Abbildung 5.1 verdeutlicht dieses Verfahren. In den roten-Farbkanal werden die ersten 8 Bit und in den grünen-Farbkanal die restlichen 8 Bit aufgeteilt. Die Aufteilung der 16 Bit auf die Farbkanäle erfolgt Server-seitig.

Um aus *heigh* und *low* die 16 Bit *v* zu berechnen, genügt es den Wert von *heigh* zuerst mit 255 zu multiplizieren und dann den Wert von *low* zu addieren. Die Gleichung 5.1 zeigt genau diesen Zusammenhang:

$$v = low + heigh \times 255. \quad (5.1)$$

Beim Laden einer Textur auf die Grafikkarte, werden alle Farbkanäle normiert, im Fall eines 8



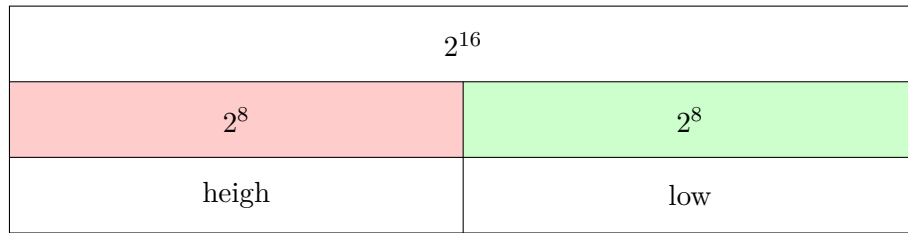


Abbildung 5.1: Hier wird die Aufteilung einer 16 Bit Zahl auf die Farbkanäle, links rot und rechts grün, visuell verdeutlicht. Das Schlüsselwort *heigh* bezeichnet die ersten 8 Bit und *low* die zweiten 8 Bit.

Bit Bildes, wird jeder Wert durch den Maximalwert 255 geteilt, so dass der Wert im Intervall von  $[0, 1]$  liegt.

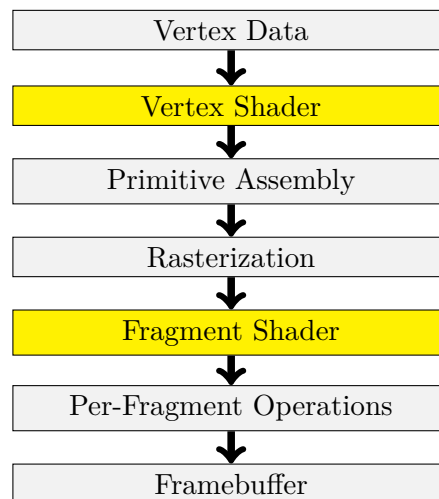


Abbildung 5.2: OpenGL ES 2.0 Darstellungspipeline. Grau unterlegt sind statischen und gelb die programmierbaren Elemente.

## 6 Ergebnisse

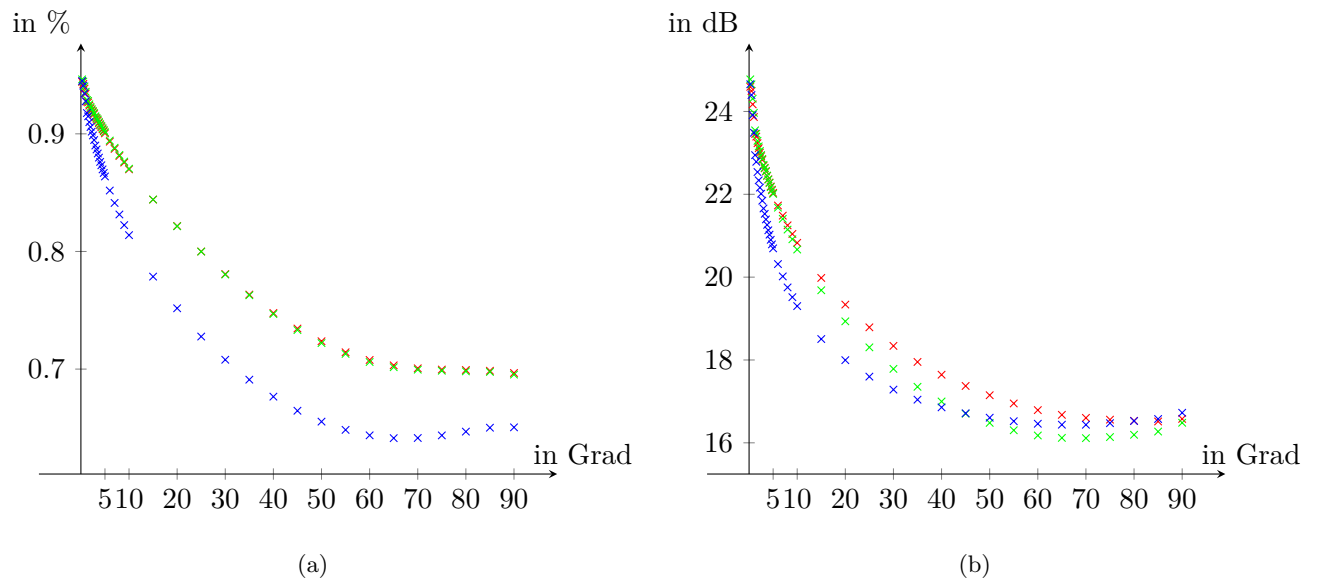


Abbildung 6.1: *CoolRandom* FULL, FLOYD, QUAD

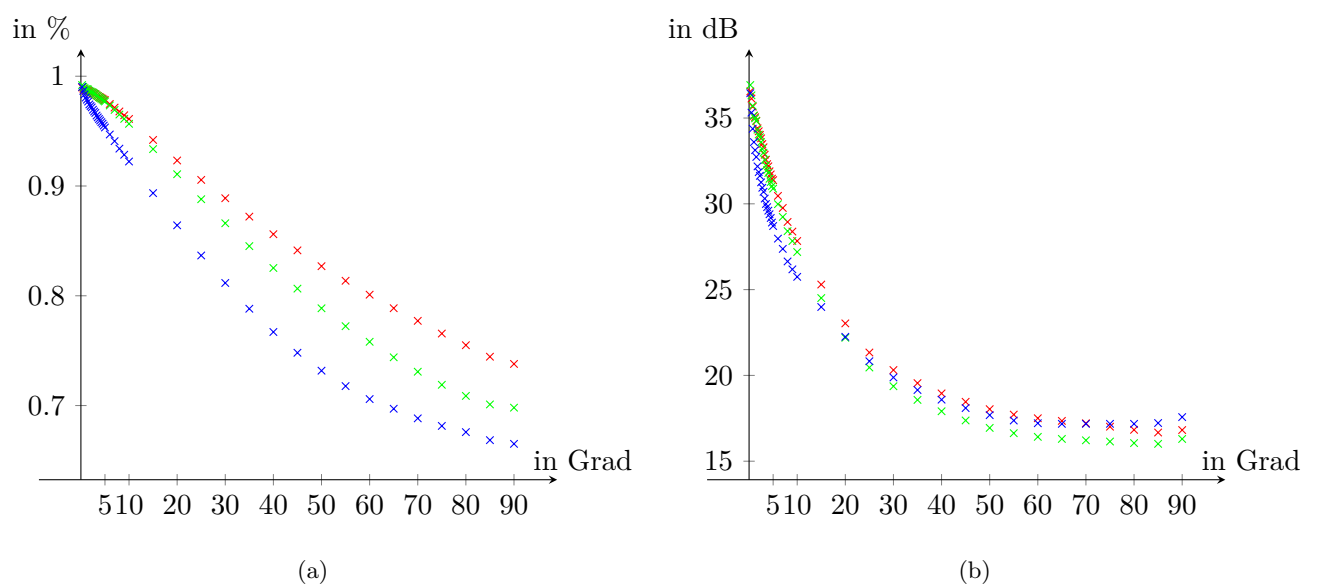
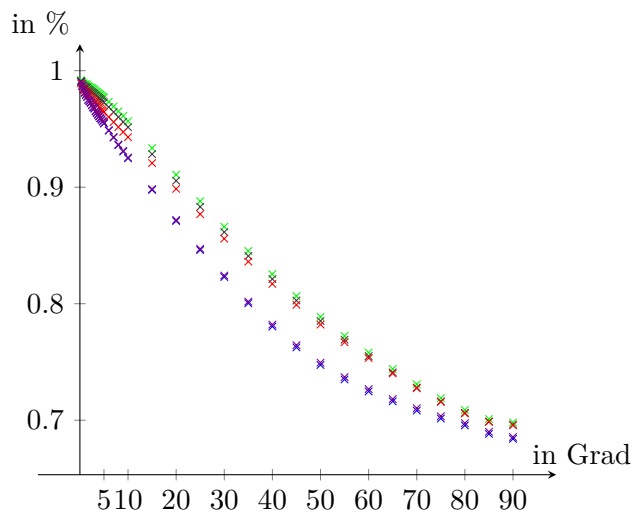
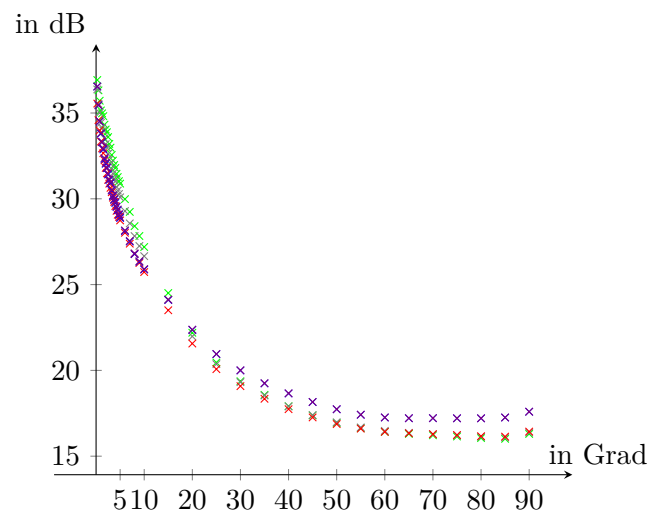


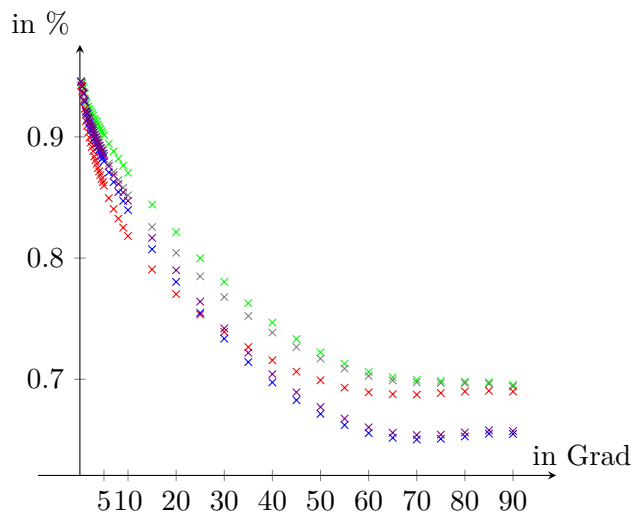
Abbildung 6.2: *TestSpheres* FULL, FLOYD, QUAD



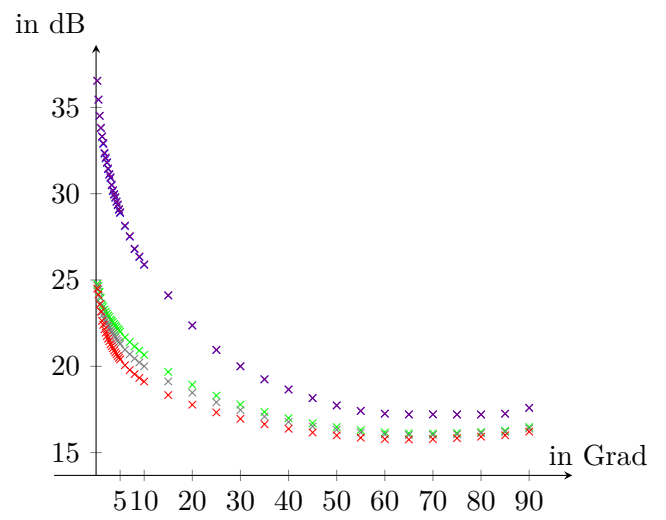
(a)



(b)

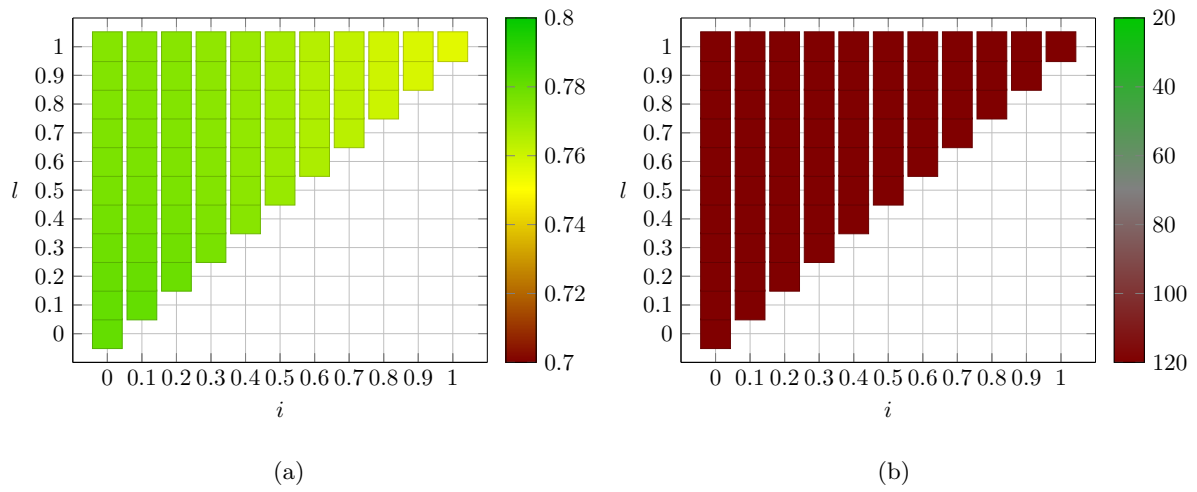
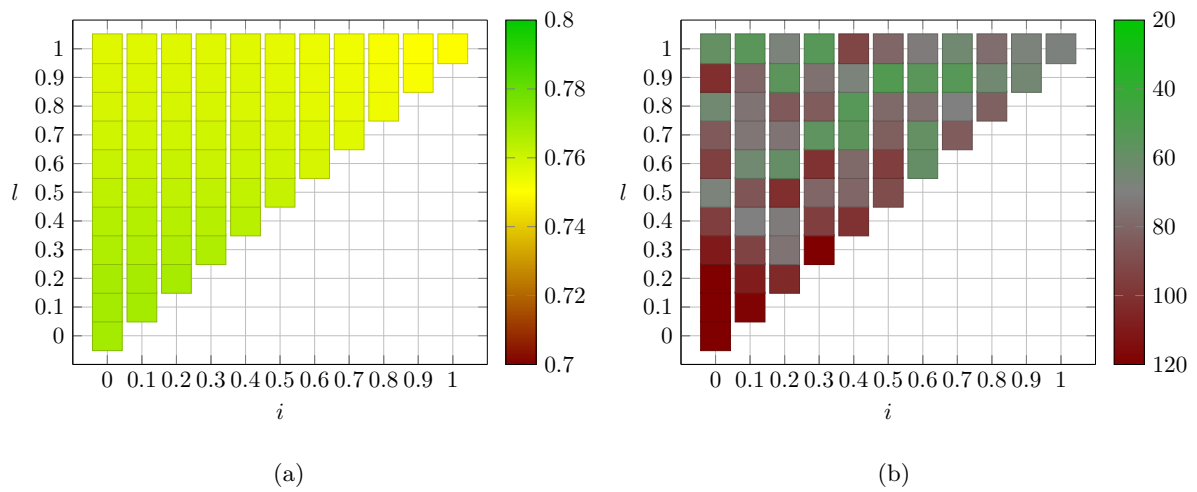
Abbildung 6.3: *TestSpheres* D10, D9, D8

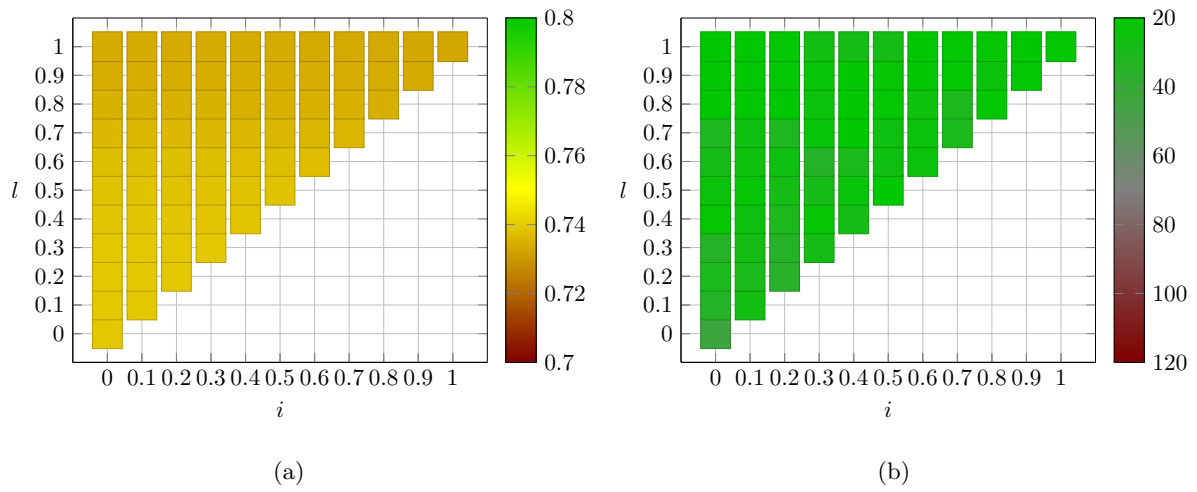
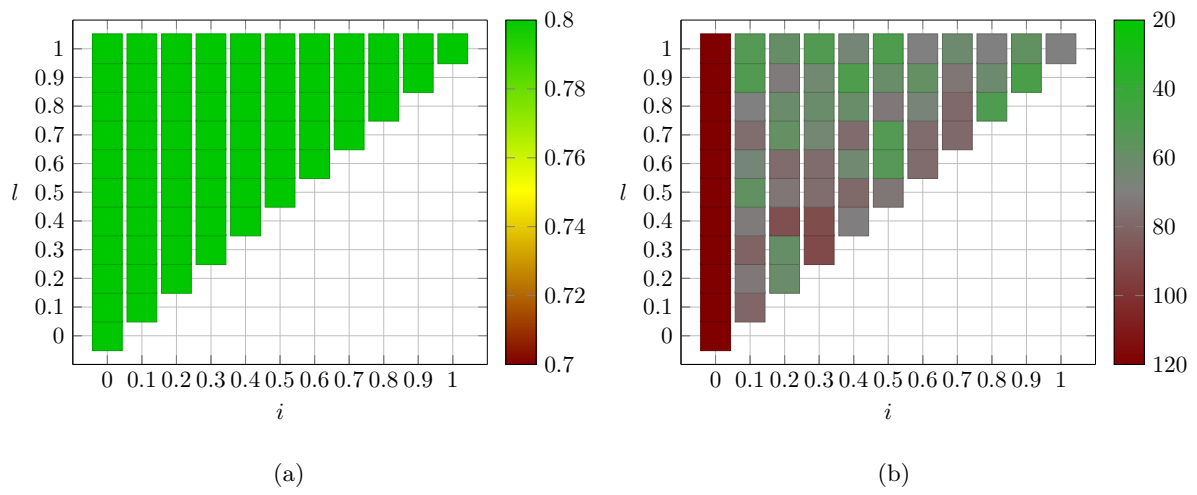
(a)

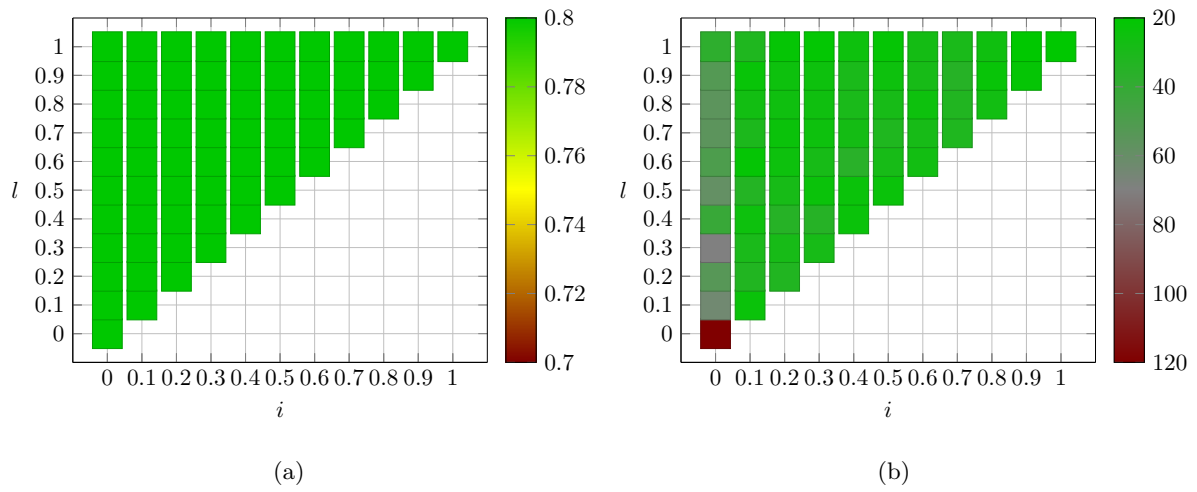
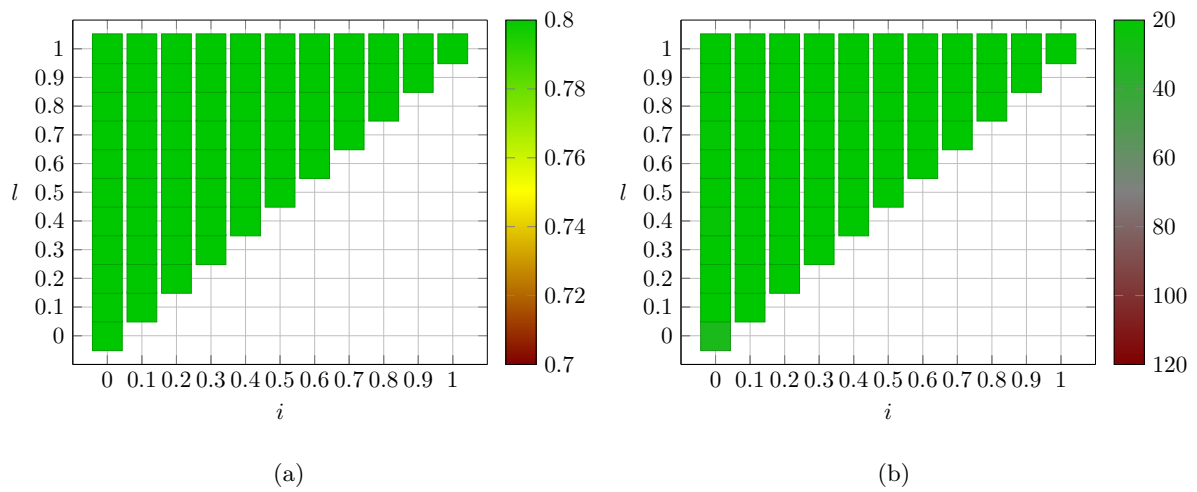


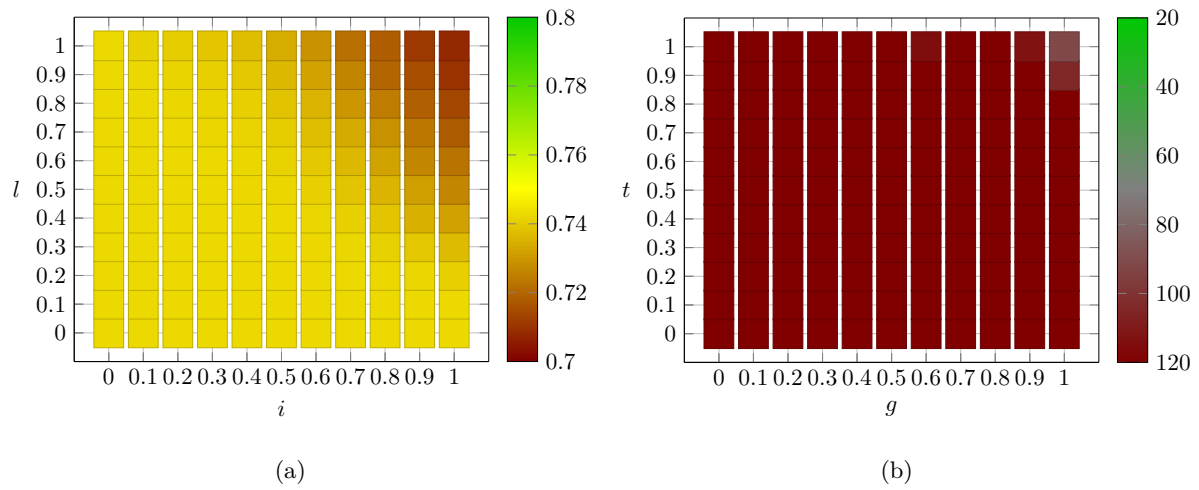
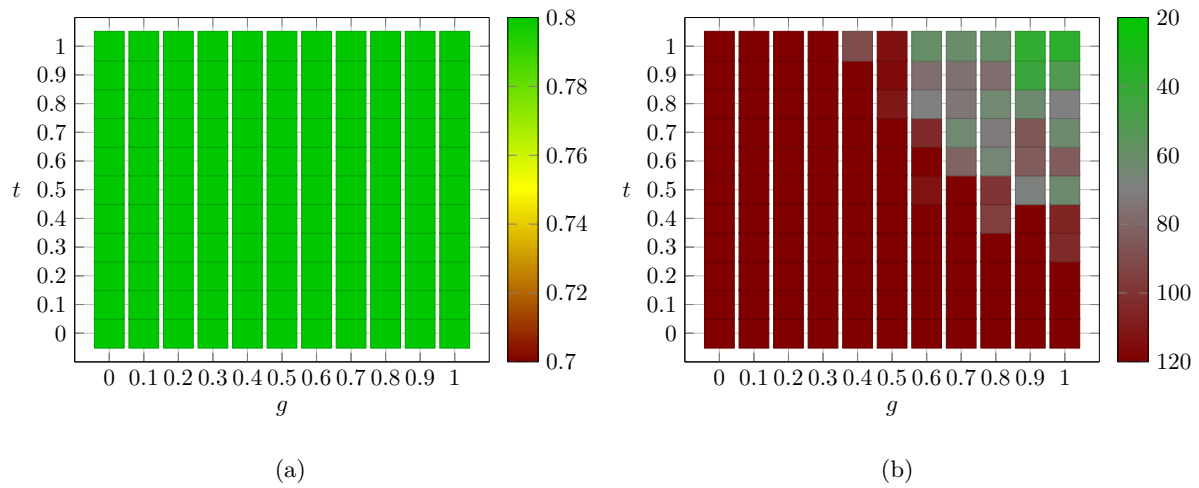
(b)

Abbildung 6.4: *CoolRandom* D10, D9, D8

Abbildung 6.5: *CoolRandom* D10Abbildung 6.6: *CoolRandom* D9

Abbildung 6.7: *CoolRandom* D8Abbildung 6.8: *TestSpheres* D10

Abbildung 6.9: *TestSpheres* D9Abbildung 6.10: *TestSpheres* D8

Abbildung 6.11: *CoolRandom* FloydSteinbergAbbildung 6.12: *TestSpheres* FloydSteinberg

## 7 Diskussion



## **8 Zusammenfassung**

## **9 Ausblick**

## **Danksagung**

Die Danksagung...

## **Erklärungen zum Urheberrecht**

Hier soll jeder Autor die von ihm eingeholten Zustimmungen der Copyright-Besitzer angeben bzw. die in Web Press Rooms angegebenen generellen Konditionen seiner Text- und Bildübernahmen zitieren.