



UNIVERSITÄT PADERBORN

Die Universität der Informationsgesellschaft

Fakultät für Elektrotechnik, Informatik und Mathematik

Institut für Informatik

Fachgebiet Computergrafik, Visualisierung und Bildverarbeitung

Warburger Straße 100

33098 Paderborn

Ermittlung optimaler Beleuchtungs- und Materialparameter für medizinische Volumenvisualisierungen

Bachelor-Arbeit

im Rahmen des Studiengangs Informatik

zur Erlangung des Grades

Bachelor of Science

von

CHRISTIAN LUCAS

vorgelegt bei

Prof. Dr. Gitta Domik

und

Prof. Dr. Uwe Kastens

Paderborn, Oktober 2011

Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen worden ist. Alle Ausführungen, die wörtlich oder sinngemäß übernommen worden sind, sind als solche gekennzeichnet.

Ort, Datum

Unterschrift

Inhalt

1	Einleitung	1
1.1	Motivation	1
1.2	Zielsetzung	2
1.3	Struktur der Arbeit	2
1.4	Verwandte Literatur	3
2	Grundlagen	5
2.1	Volumenrendering	5
2.2	Style Transfer Functions	9
2.3	Möglichkeiten der Beleuchtung im Volumenrendering	10
2.4	Perzeptuelle Anforderungen medizinischer Volumenvisualisierungen	11
2.4.1	Physikalische Voraussetzungen der Anzeigegeräte	12
2.4.2	Formwahrnehmung und Schattierung	13
2.4.3	Oberflächenbeschaffenheit	14
2.4.4	Farben	15
2.4.5	Top-down-Einflüsse	17
3	Versuchskonstruktion	19
3.1	Vorgehen	19
3.2	unabhängige Variablen	20
3.2.1	nicht-experimenteller Vorversuch	21
3.3	abhängige Variable	23
3.4	Störvariablen	24
3.5	Hypothese	26
3.6	Aufgaben	26
3.7	Versuchsdurchführung	30
4	Versuchsauswertung	31
4.1	Vorgehen	31
4.1.1	Umgang mit den Ergebnistabellen	33
4.2	Auswertung der Aufgabe Surface	34
4.3	Auswertung der Aufgabe Curvature	35
4.4	Auswertung der Aufgabe Depth	36
4.5	Diskussion	37
5	Zusammenfassung und Ausblick	41
5.1	Zusammenfassung	41

5.2 Ausblick	42
Anhang	
A Abbildungen	43
A.1 Beleuchtungsmethoden im Volumenrendering	43
A.2 Beleuchtungssituation für die Lit Spheres	44
A.3 Unabhängige Variablen	44
A.4 Fragebogen für den Vorversuch	46
B Tabellen	49
B.1 Rankings aus dem Vorversuch	49
Literatur	51

Abbildungsverzeichnis

2.1	Erhebung dreidimensionaler medizinischer Daten bei der CT	6
2.2	Direct Volume Rendering mittels Ray Casting	8
2.3	Gerendeter Volumendatensatz eines Herzens	9
2.4	Einfluss der Schattierung auf die Formwahrnehmung	13
2.5	Einfluss der Oberflächenurrisse auf die Beleuchtungswahrnehmung	14
3.1	Kontrolle der unabhängigen Variablen in Maya	20
3.2	Die vier für den Hauptversuch ermittelten Lit Spheres	23
3.3	Zwei Beispieldurchgänge der Aufgabe Surface	27
3.4	Zwei Beispieldurchgänge der Aufgabe Curvature	28
3.5	Zwei Beispieldurchgänge der Aufgabe Depth	29
4.1	Übersicht der Gesamtperformances der Aufgaben (Mittelwerte) .	31
A.1	Vergleich zweier Beleuchtungsmethoden im Volumenrendering . .	43
A.2	Draufsicht auf die verwandte Drei-Punkt-Beleuchtung	44
A.3	Mögliche Ausprägungen der unabhängigen Variablen	45
A.4	Bilderset 3 des Fragebogens	46
A.5	Bilderset 1 und 2 des Fragebogens	47

Tabellenverzeichnis

4.1	Versuchsergebnisse zur Aufgabe Surface	34
4.2	Versuchsergebnisse zur Aufgabe Curvature	35
4.3	Versuchsergebnisse zur Aufgabe Depth	37
B.1	Ranking der besten Rendering-Bedingungen für Bilderset 1	49
B.2	Ranking der besten Rendering-Bedingungen für Bilderset 2	49
B.3	Ranking der besten Rendering-Bedingungen für Bilderset 3	49

1 Einleitung

1.1 Motivation

In der Medizin ist es heutzutage nicht nur üblich Patienten oder angehenden Ärzten im Studium medizinische Sachverhalte anhand von Abbildungen und Visualisierungen verständlich darzustellen, sondern eine Vielzahl fachlicher Diagnosen auf Basis rechnergestützter Visualisierungen durchzuführen. Die meisten Menschen kennen Schlagworte wie „Röntgen“, „CT“ oder „MRT“ und wissen, dass hier von außen nicht sichtbare Bereiche des menschlichen Körpers graphisch zu Papier oder auf den Bildschirm gebracht werden. Dies geschieht nicht- oder minimalinvasiv, etwa durch die Gabe von Kontrastmitteln und einer eventuell aufgenommenen Strahlendosis, womit der Vorteil dieser Verfahren recht klar auf der Hand liegt: Ohne einen operativen Eingriff Einblicke über den physischen Zustand des Körpers eines Patienten zu bekommen.

Bei den Medizinern und Ingenieuren hat dies zu einer stetigen Weiterentwicklung und Nutzung der Techniken geführt, ohne kritische Beurteilung der resultierenden Visualisierungen in Hinblick auf ihre Wirksamkeit. Gerade die vielfältigen Möglichkeiten, die sich durch die Anzahl an Verfahren und deren Unterschiede sowie den Änderungen an Parametern innerhalb der Verfahren ergeben, legen nahe, dass aus den verschiedenartigen Visualisierungen eines Körperbereichs auch unterschiedliche Diagnosen des medizinischen Fachpersonals resultieren könnten. Die Wahl zwischen den Techniken erfolgt dabei noch meist wirkungsvoll aufgabenbezogen; bei der Wahl der oftmals nicht intuitiven Parameter einer ausgewählten Visualisierungstechnik dagegen ist dies nicht so eindeutig. Schlussendlich kann die Wahl der Parameter, sobald sich dadurch unterschiedliche und falsche Diagnosen ergeben, Einfluss auf das Leben der Patienten haben, weshalb es nahe liegt, der Evaluierung von Visualisierungstechniken mehr Aufmerksamkeit zu widmen.

Die vorliegende Arbeit befasst sich mit diesem Thema in einem Teilbereich der medizinischen Visualisierung. In der Arbeitsgruppe „Computergrafik, Visualisierung und Bildverarbeitung“ von Frau Prof. Dr. Gitta Domik an der Universität Paderborn liegt der Fokus momentan vor allem auf der dreidimensionalen Visualisierung von Volumendaten mittels der dafür entwickelten Software „Volume Studio“. Die visualisierten Herzdaten stammen aus dem „Herz- und Diabeteszentrum NRW“ in Bad Oeynhausen, mit dem die Arbeitsgruppe kooperiert. Mit einer Arbeit von Felix Steffen [Ste10] wurde im letzten Jahr an der AG schon die Grundlage zur Durchführung von Benutzertests geschaffen, um systematisch statistische Daten zu erheben, mit denen die Visualisierungen evaluiert werden können.

Die dabei von Steffen durchgeführte Benutzerstudie verglich zwei Techniken, die beim Zuweisen von Farbwerten zu den Volumendaten verwendet werden können: *Phong Shading* und *Style Transfer Functions* (genauer zum Hintergrund dieser findet sich in Kapitel 2). Das Ziel war es, zu ermitteln, ob sich eine der Techniken als visuell vorteilhafter herausstellen würde, indem man Laien einfache Aufgaben zu den gezeigten Bildern stellte, ohne dass diese medizinisches Fachwissen erfordern. Problematisch dabei ist, dass die Parameter innerhalb der beiden Visualisierungstechniken Phong Shading und Style Transfer Functions von vorn herein in je einer konkreten Ausprägung festgelegt wurden und genau nur diese beiden miteinander verglichen. Somit ist das Ergebnis aus diesem Test nicht allgemeingültig anwendbar auf die Fragestellung, ob das Phong Shading oder die Style Transfer Functions der jeweils anderen Technik vorzuziehen ist.

1.2 Zielsetzung

Diese Arbeit entstand nun aus der eben genannten Situation, dass nicht klar ist, ob man denn „gute“ oder „schlechte“ Ausprägungen der beiden Techniken – insbesondere der in Bezug auf die Einstellmöglichkeiten mächtigeren Style Transfer Functions – verwandt hat. Es gilt zunächst also festzustellen, was gute Beleuchtungsparameter für Style Transfer Functions sind. Da es sich aber generell um Techniken handelt, die Beleuchtungssituationen simulieren, die sich durch Eigenschaften der Lichtquellen und der Materialeigenschaften parametrisieren lassen, kann man das Ergebnis dieser Fragestellung allgemein auf jede andere Beleuchtungstechnik übertragen (etwa auf das Phong Shading), weswegen das Ziel dieser Arbeit allgemeiner formuliert ist: Die Durchführung und Auswertung eines Benutzertests zur Feststellung, welche Beleuchtungs- und Materialparameter am geeignetsten sind, dem Betrachter Vorteile bei der Wahrnehmung von medizinischen Volumendaten zu ermöglichen.

1.3 Struktur der Arbeit

Dafür wird festgestellt, welche Beleuchtungssituationen man im Volumenrendering mit Style Transfer Functions simulieren kann und welche dieser für die Wahrnehmung durch den Menschen geeignet sind. Darauf baut das Experimentaldesign auf, denn es werden entsprechende Visualisierungen erzeugt, für die mittels einer von Steffen entwickelten und von mir angepassten Software schließlich Testdaten von Studenten als Versuchspersonen erhoben und auf ihre statistische Signifikanz hin ausgewertet werden. Dies impliziert auch eine Diskussion der Versuchsergebnisse, etwa wenn bestimmte Behauptungen nicht gezeigt werden können. Am Ende steht als Resultat eine Beschreibung solcher Beleuchtungs- und Materialparameter, welche sich am besten für die medizinische Visualisierung eignen - hergeleitet aus den als geeignet ermittelten und dafür verwandten Style Transfer Functions.

1.4 Verwandte Literatur

Es ist bisher weniger der Versuch einer umfassenden Studie zur optimalen Gestaltung dreidimensionaler Computervisualisierungen in Bezug auf eine wirksame Wahrnehmung unternommen worden, sondern man hat vorwiegend ausgewählte basale Faktoren der Bildgebung auf geeignete Ausprägungen überprüft oder versuchte für neu vorgestellte Techniken einer Anwendungsdomäne, wie der medizinischen Bildgebung, diese mit vielen Parametern belegten Techniken durch Benutzerstudien als vorteilhaft herauszustellen (siehe beispielsweise [RDRS10]). Dies schließt allerdings keine systematische Kontrolle der Parameter mit Fokus auf eben jene ein, wie es bei nun folgenden Untersuchungen der Fall ist.

Caniard und Fleming haben zunächst den Einfluss auf die Formwahrnehmung diffuser Oberflächen durch Variationen der Leuchtrichtungen mit gängigen Beleuchtungstechniken betrachtet [CF07]. In den beiden dafür verwandten Experimenten wurden künstlich erzeugte NURBS¹ benutzt, deren Unterschiede in ihrer Form objektiv durch ein mathematisches Maß beschrieben werden konnten. Dabei stellte sich etwa heraus, dass unterschiedliche Beleuchtungseinstellungen zu veränderter Formwahrnehmung führen, selbst bei ein- und derselben Form.

Etwas abstrakter versuchten Weigle und Banks – unter anderem beziehend auf eben genannte Studie – den Einfluss von physikalisch korrekter Beleuchtung gegenüber lokaler Beleuchtung sowie den Einfluss von perspektivischer gegenüber orthographischer Blickrichtung festzustellen [WB08]. Dabei wurde jeweils für beide betrachtete Faktoren in der Untersuchung ein Einfluss auf die Wahrnehmung einer Ansammlung dreidimensional dargestellter Röhren für den Test deutlich; besonders gut unterstützt die physikalisch korrekte Beleuchtung die Wahrnehmung.

Fleming hat zusammen mit Torralba und Adelson neben der oben erwähnten Untersuchung „weicher“ Oberflächen auch den Einfluss von Spiegelungen auf Oberflächen betrachtet [FTA04]. Sie beschreiben, dass Spiegelungen in einer Art und Weise wie Texturen verarbeitet werden können, wobei sich ein fundamentaler Unterschied bei der Bildstauchung bzw. -dehnung offenbart: Bei Spiegelungen hängt diese von der Oberflächenausrichtung ab, bei Texturen jedoch von der Oberflächentiefe. Gezeigt wurde etwa, dass die spiegelnden Flächen einer Form selbst beim Wechsel der darin abgebildeten Umwelt eine recht stabile Wahrnehmung der Form ermöglichen, mehr als es bisherige Ergebnisse (etwa von Todd und Mingolla mit drei Experimenten zur Wahrnehmung von gekrümmten Oberflächen durch die einfache Lichtquellenreflexion [TM83]) vermuten ließen.

Zu den Texturen haben Kim, Hagh-Shenas und Interrante zwei kontrollierte Benutzerexperimente durchgeführt, um die Frage nach der bestmöglichen Textur zur dreidimensionalen Formwahrnehmung zu beantworten [KHSI04]. Bidirektionale Gitter-Muster werden meistens als sehr vorteilhaft angesehen, was wiederum von der Blickrichtung – schräg oder frontal zum betrachteten Objekt – abhängt, sodass auch einige andere Muster die Formwahrnehmung gut unterstützen.

¹nicht-uniforme rationale **B-Splines** sind mathematisch definierte Kurven und Flächen

2 Grundlagen

In diesem Kapitel werden die technischen Grundlagen zum Verständnis der Volumenvisualisierung von medizinischen Daten gelegt, von der physikalischen Entstehung hin zu einem dreidimensionalen Volumenmodell, welches zur Darstellung in ein zweidimensionales Bild auf dem Display umgewandelt wird, abhängig von den gewählten Parametern dieses Umwandlungsprozesses. Abschließend folgen die perceptuellen Grundlagen für die Beleuchtung dreidimensionaler Szenen, womit die daraus resultierenden Anforderungen an die vorher dargelegten technischen Möglichkeiten aufgedeckt werden.

2.1 Volumenrendering

Die moderne medizinische Bildgebung ist mittlerweile sehr vielfältig. Während Conrad Wilhelm Röntgen am Ende des 19. Jahrhunderts mit der Entdeckung der durch ihn so bezeichneten „X-Strahlen“ erste Schattenbilder von Körperpartien auf zweidimensionalen Projektionsfolien erzeugte, gibt es heute nicht nur zusätzlich die Möglichkeit der dreidimensionalen Visualisierung, sondern innerhalb dieser auch verschiedenste Techniken. Die Computertomographie (CT) verwendet dabei ebenso Röntgenstrahlen und war das erste Verfahren zur 3D-Darstellung. Durch verschiedene Dichten der Körpergewebe werden die Röntgenstrahlen unterschiedlich stark abgeschwächt und erzeugen so die „Schatten“, die das Bild ergeben. Da eine gewöhnliche Röntgenaufnahme nur eine Überlappung der durchleuchteten anatomischen Bestandteile darstellen kann¹, werden beim CT aus mehreren verschiedenen Richtungen um das Objekt Röntgenstrahlen ausgesendet und auf der gegenüberliegenden Seite ihre Stärke nach der Durchdringung des Objekts detektiert, sodass mit rechnergestützter Rekonstruktion (daraus folgt auch der Begriff *Computertomographie*) dieser CT-Daten mittels Radontransformation ein Schichtbild für den Schnitt des Körpers an dieser Stelle entsteht. Setzt man diese Schichtbilder nun dicht nebeneinander an und reiht sie zusammen, ergibt sich ein dreidimensionales Bild der betrachteten Körperpartie – ein *Volumen* entsteht.

Es gibt, wie oben angedeutet, verschiedenste Techniken, um solche Schicht- und Volumenbilder zu erzeugen. Bei der Magnetresonanztomographie („MRT“

¹Man kann u.U. Gewebe hinter einem dichten Knochen kaum erkennen, da die Knochen schon vorher einen starken Schatten werfen, d.h. die Röntgen-Strahlen durch die dichten Knochen aufgehalten werden; Umgekehrt kann eine viel breitere Schicht durchleuchtet werden, deren Material zwar weniger Strahlung absorbiert als Knochen, aber dafür eben durch die Dicke genauso viel Schatten wirft wie ein Knochen und nicht davon unterschieden werden kann.

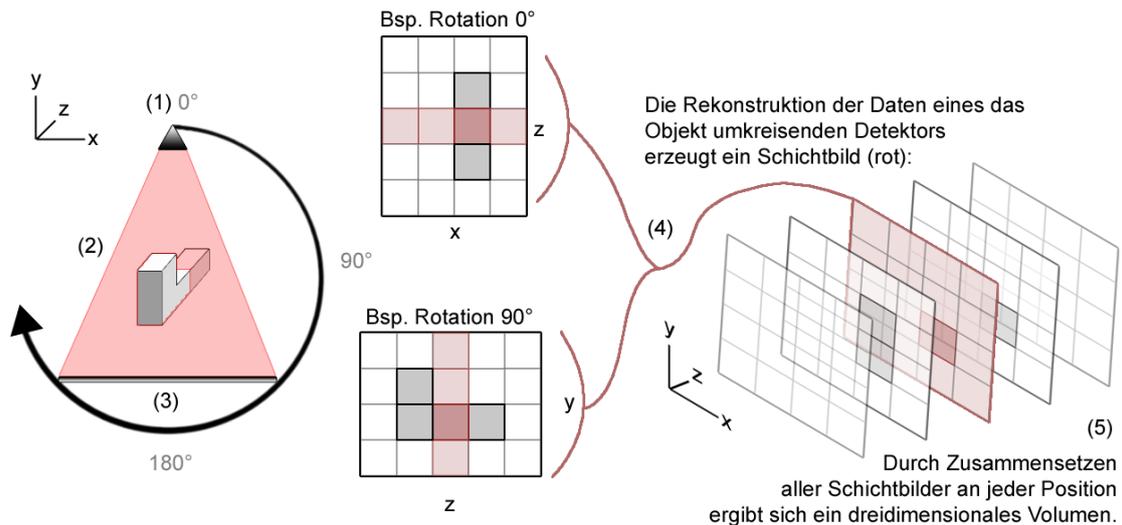


Abbildung 2.1: Erhebung dreidimensionaler medizinischer Daten bei der CT

oder aus dem Englischen „MRI“) werden durch starke Magnetfelder die Atome im Körper angeregt, welche dadurch ihrerseits elektrische Signale induzieren. Durch entsprechend unterschiedliche Zeiten, bis diese Atome wieder ihren Gleichgewichtszustand erreichen, entstehen die Schnittbilder. Daher nennt man diese Technik vereinzelt auch Kernspintomographie.

Im Gegensatz zur MRT, die vollkommen ohne radioaktive oder Röntgenstrahlung auskommt, setzt die nun von mir als letztes erwähnte Technik der „PET“ (Positronenemissionstomographie) auf Verabreichung und Abbildung der Verteilung eines radioaktiven Kontrastmittels im Körper. Diese Technik wird oft als Ergänzung zu den vorher genannten Techniken genutzt, da sich hiermit besonders gut die funktionellen Gewebeeigenschaften für bestimmte physiologische Prozesse abbilden lassen, z.B. die Zuckerverbrennung von Muskeln wie dem Herzen. Allen Techniken jedoch ist gemein, dass daraus Volumendaten erzeugt werden können.

Abbildung 2.1 zeigt beispielhaft anhand der Computertomographie den Ablauf eines Scan-Vorgangs: Zuerst wird in Schritt (1) die für die CT notwendige Strahlung von einer Röntgenquelle emittiert, wobei die Röntgenstrahlen beim Durchdringen durch den Körper teilweise absorbiert werden (2) und entsprechend geschwächt als Schatten an einem Detektor (3) auftreffen (graue Färbung bedeutet hier hohe Dichte, daher ergibt sich bei den nur mit Luft gefüllten Freiräumen weiß). Die Daten – hier rot markiert – eines oder mehrerer das Objekt umkreisender Detektoren, auf welche die Strahlung gelenkt wird, werden durch Rekonstruktion zu einem Schichtbild zusammengerechnet (4). Der gesamte Vorgang passiert für in bestimmtem Abstand aufeinanderfolgende Positionen entlang der hier als z-Achse bezeichneten Raumdimension und durch die Aneinanderreihung aller Schnittbilder (5) erzeugt man einen Volumendatensatz. Bei der MRT oder der PET gibt es zur CT unterschiedlich stark abweichende Vorgehensweisen, unter anderem kann auch die Gabe von Kontrastmitteln erforderlich sein.

Bestehen die Darstellungen der Schichtbilder aus Pixeln, also den „picture elements“, so ist nun ein Volumen aus *Voxeln*, also „volume elements“, aufgebaut, die in einem dreidimensional angeordneten Gitter einzelne Skalare enthalten. Bei Rastergrafiken sind die Pixel typischerweise Tupel bestehend aus den zwei Positions-Koordinaten sowie deren Abbildung auf einen Graustufen- oder RGB-Farbwert, wohingegen in der medizinischen Volumenvisualisierung die Voxel typischerweise aus den drei Koordinaten im Raum und einer Abbildung auf einen Absorptionswert² bestehen (klassische CT), wobei z.B. Knochen im Körper eine hohe Absorption aufweisen und damit Voxel an diesen Stellen große Absorptionswerte enthalten. Die Werte sind standardisiert durch den Hounsfield-Wert, wobei ein Wert von -1000 für Luft definiert ist, und ein Wert von 0 für Wasser. Menschliches Lungen- oder Fettgewebe ordnet sich dabei zwischen diesen Werten ein, Herz- oder Knochengewebe bei Werten größer 0.

Die nun im dreidimensionalen Gitter vorliegenden Rohdaten müssen auf irgendeine Weise visualisiert werden. Um diese auf dem zweidimensionalen Bildschirm sichtbar zu machen, wäre die bereits naheliegende Methode natürlich das „Scrolen“ durch die einzelnen Schichtbilder, sodass man immer nur eines dieser gleichzeitig betrachten kann. Dabei geht allerdings der räumliche Eindruck verloren, es mag jedoch für bestimmte Anwendungen sinnvoll sein. Unser Ziel sind aber dreidimensionale Volumendarstellungen. Möglichkeiten dafür wären etwa das reine Herausstellen von Dichtegrenzschichten mit dem Marching Cubes Algorithmus, um so die Oberfläche der Organe mittels Isoflächen sichtbar zu machen. Es gibt aber auch die Techniken des Direct Volume Rendering (DVR).

Diese ermöglichen, wie der Name schon nahelegt, eine direkte dreidimensionale Abbildung des Volumendatensatzes durch die Eigenschaft, dass er semi-transparent dargestellt werden kann. Man kann sich das Volumen gasförmig vorstellen – je nach vorher getroffenen Annahmen emittiert, absorbiert oder streut es das Licht (das ist vorwiegend eine Abwägung zwischen Realismus und Rechenaufwand). Der Zusammenhang zwischen diesen Eigenschaften des Gases in Abhängigkeit von der Entfernung auf einem Lichtstrahl zum Betrachter wird mit der Volumenrendering-Gleichung ausgedrückt, bei der die Emission und das Produkt aus der Intensität und der Absorption an einer Position addiert werden [EHK⁺06, S. 8]. Durch Integrieren der Gleichung über den Weg durch das Volumen zum Auge des Betrachters ergibt sich das Volumenrendering-Integral, welches schließlich die Berechnung eines Intensitätswertes für den Lichtstrahl beim Erreichen des Betrachters erlaubt.

Die eben erwähnten Berechnungen auf Basis des gasförmig angenommenen Volumens müssen nun an allen Positionen berechnet werden, sodass sich eine ganzheitliche dreidimensionale Visualisierung auf dem Bildschirm ergibt. Die Positionen, an denen man das Volumenrendering-Integral berechnet, hängen dabei von der Rate ab, mit der das Volumen abgetastet wird. Diese kann man vorgeben: Je höher sie ausfällt (je kleiner die Abstände zwischen den Abtastpunkten), desto

²Dieser Absorptionswert ist fast identisch mit den tatsächlichen Dichten der Gewebe.

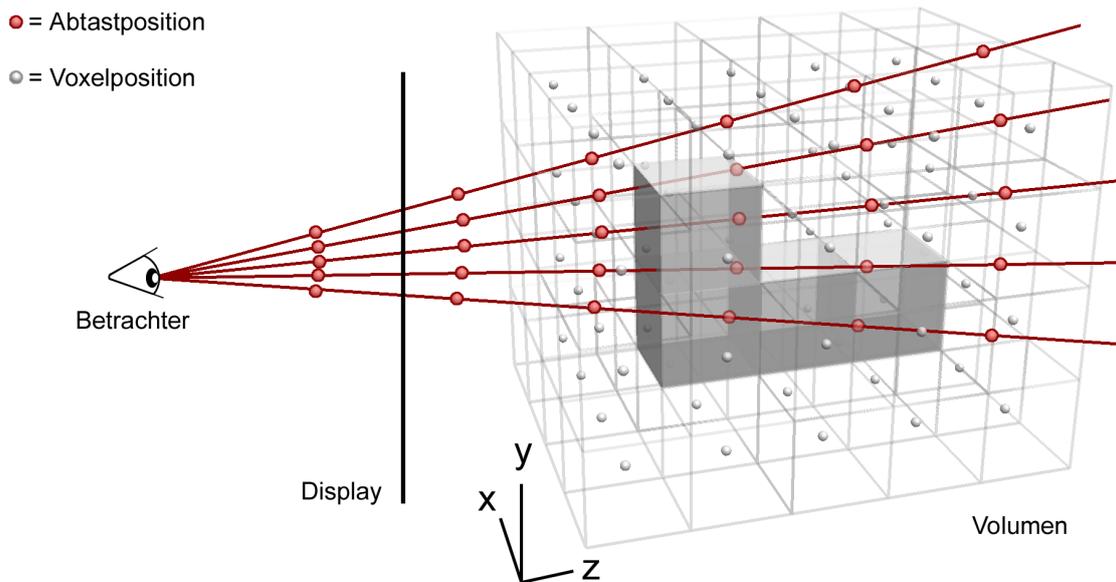


Abbildung 2.2: Ray Casting, wobei hier Abtastpunkt ungleich Voxelposition

feiner wird das resultierende Bild aufgelöst und desto mehr Rechenaufwand muss betrieben werden. Liegt ein Abtastpunkt nicht genau an der Stelle eines Voxels im Volumendatensatz (siehe Abbildung 2.2), so wird für diesen Abtastpunkt aus den Skalaren der umliegenden Voxel ein Wert interpoliert.

Dabei gibt es bild- oder objektbasierte DVR-Techniken. Eine Möglichkeit ist es, auf die Unterstützung der vorhandenen Grafik-Hardware zu setzen und das Volumen parallel zur Ausrichtung des Volumendatensatzes oder parallel zur Betrachtungsebene in Texturen abzubilden. Diese unter Umständen effiziente Methode kann zu sichtbaren Bildfehlern führen. Bessere Bildqualität liefert das Ray Casting als bildbasierte Methode, bei der für jeden Pixel auf dem Bildschirm des Betrachters ein Strahl ausgesandt wird und man direkt darauf intuitiv das Volumenrendering-Integral beim Durchdringen des Volumens berechnen kann.

Aus den verschiedenen an den Abtastpunkten berechneten Intensitätswerten müssen nun schließlich Farb- und Transparenzwerte für die Pixel berechnet werden. Mithilfe von Transferfunktionen werden die skalaren Datenwerte z.B. durch eine Lookup-Tabelle auf RGBA-Werte abgebildet. Dies kann sowohl durch ein-dimensionale Transferfunktionen geschehen, die allein auf Basis eines einzelnen Wertes RGBA-Werte definieren, als auch durch mehrdimensionale Transferfunktionen, wobei verschiedene Werte auf einen RGBA-Wert abgebildet werden. Es ist dabei prinzipiell egal und implementierungsabhängig, ob es sich um Werte einer Modalität handelt, also direkt den gemessenen Daten des Scanners, oder um Werte einer daraus erst zu berechnenden Metrik, wie den „Distanzen zum Herzmittelpunkt“ oder den Hounsfield-Werten eines Herzens. Damit kann man etwa Transparenzwerte (Alpha) zum Freistellen des Herzens definieren, wobei nur jene Voxel mit Werten, die Herzgewebe repräsentieren, sichtbar bleiben; bei den genannten

Metriken also jene mit kleiner Distanz oder Hounsfield-Werten des Herzens.

Mit der Metrik des Gradienten, welcher als Annäherung an die Normale aus den ursprünglichen Daten berechnet wird [EHK⁺06, S. 252], kann man das Herz nun z.B. mittels Phong Shading beleuchten oder wie im Falle der im folgenden Abschnitt beschriebenen Style Transfer Functions geschickt einfärben. Zuletzt seien noch der Vollständigkeit halber neben den Modalitäten und Metriken die Segmentierungen erwähnt, welche mit nominalen Werten auf Basis eines anderen Volumens definieren, was zu dem segmentierten Volumen gehört und was nicht.

2.2 Style Transfer Functions

Die Style Transfer Functions bilden eine intuitive Möglichkeit, dem Volumen Farben zuzuweisen. Dabei wird nur abhängig vom Gradienten am Abtastpunkt ein RGBA-Wert auf einer vorher gerenderten Kugel als Stilvorlage (engl. *style*), der so genannten *Lit Sphere*, referenziert; das Volumen wird also winkelabhängig eingefärbt, wie in Abbildung 2.3 dargestellt. Das Definieren von Transfer Funktionen und deren korrekter Wertepaare in einer Lookup-Tabelle ist für den Benutzer wenig intuitiv und ohne Erfahrung verhältnismäßig aufwendig. Bei Style Transfer Functions hingegen ist von vorn herein klar, welcher Beleuchtungssituation das Volumen ausgesetzt sein wird, denn dieses ist bei Erstellung der Lit Sphere in gängiger Grafiksoftware direkt sichtbar und somit sind Style Transfer Functions sehr geeignet, um im Volumenrendering komplexe Lichtsituationen umzusetzen.

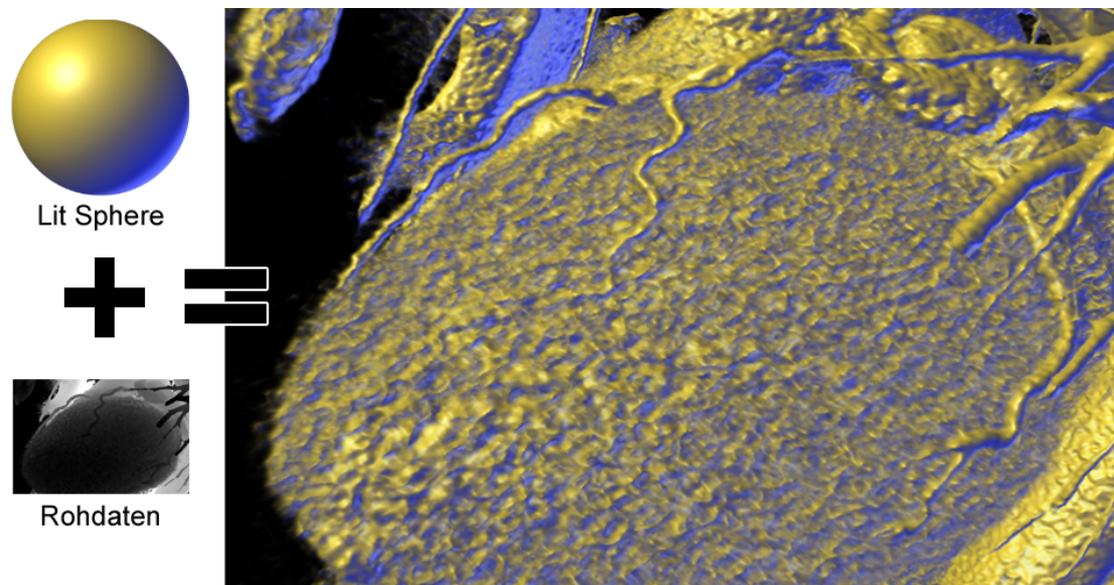


Abbildung 2.3: Ein mit einer Style Transfer Function gerendertes Volumendatensatz eines Herzens, bei dem mittels einer separaten Transferfunktion und den damit definierten Transparenzwerten die Oberfläche des Herzens und dessen Arterien freigestellt wurden.

Was für Transfer-Funktionen allgemein gilt, gilt auch für Style Transfer Functions im Speziellen: Man kann zugleich mehrere verwenden. Dies ist gerade dann sinnvoll, wenn man nur für bestimmte Intensitätswerte einzelne Style Transfer Functions definieren möchte, um etwa für die Werte, die das Herz repräsentieren, eine rote Lit Sphere zu verwenden und für Werte von Knochen eine hellgelbe Lit Sphere. So können Volumen wie Abbildungen in einem Anatomiebuch eingefärbt werden (siehe z.B. [BG07]). Die sich mit der Erfahrung deckende Visualisierung wird jedoch nicht Gegenstand meines Versuchs sein, sondern diejenige, welche bestmögliche Beleuchtungsbedingungen zum Stellen einer Diagnose ermöglicht, und daher werde ich mich auf einzelne Style Transfer Functions beschränken.

Die Style Transfer Functions bieten aufgrund ihrer intendierten Verwendung für Visualisierungen, die bestimmte Details etwa für Lehrzwecke hervorheben, vor allem Vorteile beim nicht-photorealistischen Rendering (NPR), da mit ihnen beispielsweise auch Konturen, Toon Shading und ähnliches in einer Stilvorlage möglich sind. Gerade bei technischen Visualisierungen wird auch die Verwendung von Cool-Warm Shading geschätzt [GGSC98]. So ergeben sich verschiedenste Möglichkeiten, wobei in dieser Arbeit nur realitätsnahe Beleuchtungs- und Materialparameter – Konturierungen oder Toon Shading gehören zweifelsohne nicht dazu – untersucht werden. Gleichzeitig ist die Wahl relevanter Parameter zum Herstellen der Beleuchtungssituation durch die Möglichkeiten der Style Transfer Functions beschränkt, wie im nächsten Abschnitt deutlich wird.

2.3 Möglichkeiten der Beleuchtung im Volumenrendering

Die Notwendigkeit dieser Arbeit ergab sich aus der Frage, welches eigentlich die beste Style Transfer Function ist, um mit einer solchen die ursprüngliche Frage nach der Überlegenheit dieser gegenüber dem Phong Shading³ beantworten zu können. Ein Bild von beiden Techniken kann man sich im Anhang unter A.1 auf Seite 43 machen. Da allen Techniken gemein ist, dass sie der Beleuchtung und Färbung des Volumens dienen, kann man anhand der mit den Style Transfer Functions als geeignet ermittelten Parametereinstellungen diese Konfiguration allgemeingültig als geeignet auf alle Techniken, durch die diese Parameter gesteuert werden können, übertragen. Generell ist nämlich auch im Volumenrendering sehr vieles möglich, was sonst aus der Computergrafik bekannt ist, um eine Beleuchtungssituation herzustellen. Wichtig ist nun zu wissen, welche dieser Parameter in Frage kommen, mit den Style Transfer Functions gesteuert zu werden.

³Vom Phong Shading spricht man meist bei der Schattierung dreidimensionaler Polygonmodelle auf Basis der interpolierten Normalen der Polygone zusammen mit dem Phong-Beleuchtungsmodell, welches zwar ein physikalisch nicht korrektes, aber empirisches Modell natürlicher Beleuchtung im Gegensatz zu den NPR-Techniken darstellt. Dafür werden ambiente, diffuse und spiegelnde Reflexionsanteile des Lichts teils in Abhängigkeit des Winkels beim Auftreffen auf die Oberfläche zu einer Oberflächenbeleuchtung an dieser Stelle addiert.

Als lokale Beleuchtungsmethode können sie nur auf die lokale Information des Gradienten zurückgreifen, von der die gesamte Beleuchtung an einer beliebigen Stelle des Volumens abhängt. Auf der anderen Seite bieten sie den Vorteil, auch Materialparameter der Oberfläche des beleuchteten Objekts zu definieren, da sie direkt die Oberfläche der Lit Sphere übernehmen und damit auch bestimmte, zum Rendern der Lit Sphere verwandter Oberflächenparameter.

Dabei kann die Kugeloberfläche von vorn herein gänzlich unbunt sein oder eingefärbt, auch in Form von Texturen. Der Nutzen von Texturen auf Lit Spheres der Style Transfer Functions ist jedoch sehr begrenzt, da die Vorteile von Texturen in Bezug auf die dreidimensionale Wahrnehmung erst in Verbindung mit globalen Informationen wie der Entfernung zum Betrachter zustande kommen. Das so genannte Stippling ist eine Methode, die dies nicht benötigt und so zwar durch Verwendung einer Punkttextur einen Schattierungseffekt ganz ohne Farbe ermöglicht [EHK⁺06, S. 356], dabei aber auch nur eine geringer aufgelöste Schwarz/Weiß-Variante einer normalen farblichen Schattierung (engl. *shading*) darstellt, bietet also keinen zusätzlichen Hinweisreiz auf die Tiefengestaltung, wie es etwa ein sich in der Tiefe verjüngendes Schachbrettmuster bieten würde.

Die Lichtreflexion auf der Oberfläche hängt dabei von den definierten Materialeigenschaften ab, man kann also die Glanzeigenschaft (engl. *glossiness*) des Materials variieren, um so für einen starken Lichtreflex des technisch als „specular“ bezeichneten Anteils der Lichtquelle sorgen. Die Lichter können dafür beliebig arrangiert sein. Sowohl Farbe als auch Anteil des diffusen Streulichts und des specular-Anteils sind mit vielen 3D-Rendering-Tools, wie z.B. Autodesk's Maya, beim Rendern der Lit Sphere und damit für Style Transfer Functions einstellbar. Durch das Reduzieren des specular-Anteils kann man verhindern, dass für jede Lichtquelle ein Lichtreflex auf der Oberfläche erzeugt wird. Dies ist vergleichbar mit dem Anbringen eines Diffusors vor einem Blitzlicht im Fotostudio, der harte Lichtreflexe vermeidet und nur weiches diffuses Licht abgibt.

Mit der Flexibilität der Lichtquellen sind auch Cool-Warm Shadings durch die Style Transfer Functions möglich, indem man den Lichtquellen schlicht unterschiedliche Farbwerte zuweist, was letztendlich die Tiefenwahrnehmung fördern soll. Denn neben perspektivischer Texturierung sind auch andere Hinweisparameter der Tiefengestaltung in Style Transfer Functions nicht umsetzbar, wie etwa ein globaler Schattenwurf. Dafür benötigte es wieder Informationen, die über die lokal verfügbaren hinausgehen, z.B. ob zwischen Lichtquelle und aktueller Position ein anderer Teil des Volumens einen Schatten wirft.

2.4 Perzeptuelle Anforderungen medizinischer Volumenvisualisierungen

Leider gibt es für die Anforderungen, die durch die menschliche Wahrnehmung an eine Beleuchtung eines Volumens in der medizinischen Visualisierung gestellt wer-

den, keinen konkreten Leitfaden oder eine umfassende Auflistung entsprechender Kriterien im medizinischen Anwendungsbereich, sodass vor allem allgemeingültige Erkenntnisse bezüglich der dreidimensionalen Visualisierung zusammengetragen wurden. Die Evaluation von medizinischen Visualisierungen beschränkt sich bisher zum Großteil auf (oft zweidimensionale) Graustufenbilder, die z.B. in der Radiologie und vielen anderen Teilbereichen der medizinischen Bildgebung weiterhin eine dominierende Rolle spielen. Außerdem betrachte ich in der vorliegenden Arbeit und dem Versuch nur statische Bilder, womit ich mich auch nicht den perzeptuellen Vorteilen bezüglich der Tiefenwahrnehmung widme, welche erwiesenermaßen aus Bewegungen von Objekten und dem damit zusammenhängenden Wechsel der Perspektive oder des Schattenwurfs in 3D-Szenen entstehen [WFG92].

Die visuelle Sinneswahrnehmung beginnt mit dem optischen System im Auge und geht über einzelne Verarbeitungsstufen zum primären visuellen Cortex im hinteren Teil des Gehirns und von dort wiederum in noch höhere (kognitive) Verarbeitungsstufen. Verschiedenste Phänomene rühren daher von den verschiedenen Mechanismen bei der Verarbeitung her, sind aber hier nicht herzuleiten sondern nur zu erklären. Vorher müssen zunächst die Anzeigeräte bestimmte Voraussetzungen erfüllen, damit überhaupt Informationen zu diesen Stufen gelangen können.

2.4.1 Physikalische Voraussetzungen der Anzeigeräte

Beginnend mit dem auf der Netzhaut auftreffenden Licht gibt es zwei verschiedene Rezeptoren, die abhängig von der Helligkeit zugeschaltet werden. Bei Lichtverhältnissen von über $100\text{cd}/\text{m}^2$ spricht man vom photopischen Sehen⁴, bei dem ausschließlich die für das Farbsehen notwendigen und dank dichter Zusammenschaltung in der Fovea scharfes Sehen ermöglichenden Zapfen aktiv sind, wie es von Lehmann, Oberschelp, Pelikan und Repges beschrieben wird [LOPR97, S. 77]. Daher sollte immer eine Grundhelligkeit von mindestens $100\text{cd}/\text{m}^2$ sichergestellt werden. Tatsächlich liegt sie auch üblicherweise bei den typischen Computerdisplays deutlich darüber, sofern man die Helligkeit in den Monitoreinstellungen nicht stark reduziert hat. Über die benötigte Grundhelligkeit hinaus hat nach Boyce die Helligkeit jedoch keinen weiteren Einfluss mehr auf die Sehschärfe [Boy03, S. 69 f.], sondern nur der Kontrast, der durch Luminanz- oder Farbunterschiede entsteht.

Die Computerdisplays korrigieren außerdem ein weiteres Phänomen des visuellen Systems des Menschen: Der Zusammenhang zwischen der eingegebenen „Intensität“ auf der Seite des Computersystems und der wahrgenommenen „Helligkeit“ auf der Seite des Menschen ist nach Lehmann et al. nämlich nicht linear,

⁴Sinkt die Helligkeit auf unter 100 aber über $0,01\text{cd}/\text{m}^2$ werden sowohl Zapfen als auch Stäbchen zum mesopischen Sehen benutzt, denn nicht der gesamte Helligkeitsbereich ist gleichzeitig wahrnehmbar. Bei unter $0,01\text{cd}/\text{m}^2$, wodurch nur noch Stäbchen aktiv sind, spricht man vom skotopischen Sehen, das uns nur noch Helligkeitsunterschiede und Kontraste bei schwacher Sehschärfe aufgrund der dafür notwendigerweise zusammengeschalteten rezeptiver Felder wahrnehmen lässt [LOPR97, S. 78].

sodass bei einer Verdopplung der Intensität uns die Helligkeit nicht verdoppelt zu sein scheint. So benötigt man eigentlich einen logarithmischen Anstieg der Intensitäten, bei der die Abstände zwischen den Grauwerten mit steigender Intensität zunehmen. Dies erledigt typischerweise die Gammakorrektur, eine Potenzfunktion, die mittels dem Reziproken des Gammawertes der für den Menschen typischen Wahrnehmungskurve trotzdem einen gleichmäßigen Intensitätsverlauf vermittelt.

2.4.2 Formwahrnehmung und Schattierung

Einige Erkenntnisse über Formwahrnehmung durch Krümmung und die daraus resultierende Schattierung wurden mit den populär werdenden und professionell eingesetzten grafischen Computersystemen bereits vor über 20 Jahren veröffentlicht. Die dreidimensionale Wahrnehmung basiert dabei vor allem auf der Schattierung, bei der die dem Licht zugewandte Stellen heller sind als die dem Licht abgewandten.

Ramachandran beschreibt die Formwahrnehmung durch Schattierung als eine globale Operation, sodass in einer Szenendarstellung davon ausgegangen wird, dass alle Objekte von einer Lichtquelle gleich beleuchtet werden [Ram88]. Das bedeutet, dass beim gleichzeitigen Betrachten zweier unterschiedlich beleuchteter Objekte nur eine Form akkurat als dreidimensionales Objekt erkannt würde. Das Gehirn präferiert dabei eine gemeinsame Lichtquelle, im Gegenteil zu einer gemeinsamen Tiefe der Objekte. Wäre letzteres der Fall, würde dies bedeuten, dass entgegengesetzt schattierte Formen als von verschiedenen Lichtquellen beleuchtet wahrgenommen werden. Stattdessen sieht aber die eine Form emporgehoben aus und die andere eingedellt. Das ganze wurde von Ramachandran für einfache konkave/konvexe Flächen gezeigt, losgelöst von anderen Hinweisen auf die Tiefengestaltung. Man kann diese Erkenntnisse also mit wachsender Komplexität einer Szene nicht eins zu eins übernehmen.

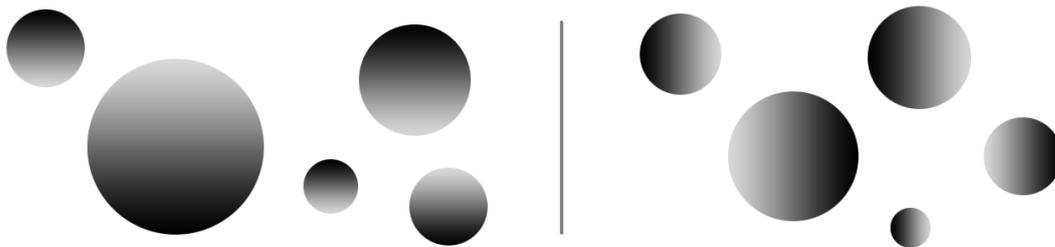


Abbildung 2.4: Einfluss der Schattierung auf die Formwahrnehmung

Auch gibt es dabei eine bestimmte Richtung, aus der das Licht zu scheinen scheint. Eine Schattierung von hell nach dunkel in vertikaler Richtung nach unten führt dazu, dass ein Objekt hervorragt; Die Schattierung von oben dunkel nach unten hell dagegen führt zu einem eingedellten Eindruck – wohl eine Folge aus unserer

natürlichen Lichtsituation, in der die Sonne immer von oben scheint. Erstaunlicherweise haben horizontale Helligkeitsgradienten nicht einen solchen Einfluss auf die Formwahrnehmung, wie in Abbildung 2.4 bei Betrachtung der rechten Kugeln gegenüber den linken deutlich wird. Der vertikale Tiefeneffekt tritt übrigens auch bei farbiger Schattierung auf: Im Anhang auf Seite 45 scheint die erste Kugel hinter 3. im Vergleich zu allen anderen „eingedellt“ zu sein, denn ihre Schattierung deckt sich nicht mit Erfahrung der Lichtes von oben (hierbei ist es nun nämlich genau umgekehrt).

Hinzu kommt, dass die Wölbung oder Tiefenstruktur eines Objekts von den Umrissen dieses abhängt. Wir kennen die zwei nebeneinander liegenden Zylinder, mit dem gleichen Helligkeitsverlauf, die uns ein Buch von einer Lichtquelle beleuchtet wahrnehmen lassen (siehe Abbildung 2.5). Sobald die gekrümmten Kanten der Buchform begradigt werden (man nun direkt auf das Buch von oben drauf schaut), kann man nicht mehr die gleiche Lichtsituation wie vorher erkennen – der Bruch des Helligkeitsgradienten in der Mitte wirkt hier hinderlich [KK91].



Abbildung 2.5: Die Oberflächenform beeinflusst die Beleuchtungswahrnehmung

Formen an sich werden je nach Anordnung und Kombination auch als zusammengehörig oder eben getrennt interpretiert, was die Gestaltgesetze versuchen zu beschreiben. Generell gilt das Prägnanzprinzip: Es setzt sich jene Anordnung durch, die einheitlich, einfach, dicht und geschlossen ist [LOPR97]. Allerdings ist Vorsicht geboten, da sich die dafür beschriebenen Gestaltgesetze des Öfteren auch widersprechen. Weiterhin gelten sie eher für einfache Gestaltkombinationen und sind in Zusammenhang mit visualisierten medizinischen Sachverhalten wohl als Erklärung nur selten anwendbar.

2.4.3 Oberflächenbeschaffenheit

Die in einem computergenerierten Bild entstehende Lichtintensität ist das Produkt aus der Stärke der Reflexion und der Leuchtstärke der reflektierten Lichtquelle [Hor74]; denn schauen wir uns eine 3D-Szene an, so sehen wir selten die Lichtquellen an sich, sondern die Reflexionen ihrer Lichtbestandteile auf der Oberfläche der Objekte. Das Wahrnehmungssystem kann die Lichtmuster auf der Retina nach Mustern von Reflexionen und direkter Beleuchtung unterscheiden, sodass auch eine Szene in einem Raum mit wenig Reflexionen als hell wahrgenommen wird, wie Boyce beschreibt [Boy03, S. 99]. Die Menge des reflektierten Lichts von einer Oberfläche hängt dabei in der Natur von Größen wie der Ausrichtung der Oberfläche,

der Materialbeschaffenheit und den Positionen sowie spektralen Zusammensetzungen der Lichtquellen ab.

Die Glossiness (Glanzeigenschaft) des Oberflächenmaterials verbessert die Wahrnehmung der Oberflächenkrümmung, hat aber keinen weiteren Einfluss auf die wahrgenommene Richtung der Beleuchtung, wie Todd und Mingolla in Experimenten zeigten [TM83]. Dabei schnitten die Versuchspersonen besonders schlecht ab, wenn eine idealisierte Lambert'sche Oberfläche (d.h. die pure diffuse Lichtreflexion ohne Glanz) verwandt wurde. Im Sinne der genauen Wahrnehmung der Form eines Objekts sollte man folglich auf spiegelnde Lichtreflexe setzen.

Ein Vorteil bei der Verwendung von glänzenden Reflexionen zeigt sich dabei auch in vorwiegend hellen Umgebungen. Die Reflexe geben einen Hinweis darauf, welche Stellen direkt vom Licht angestrahlt respektive welche nicht angestrahlt werden (also Hinweis auf die Gestalt des Objekts) und in einem allgemein sehr hellen Bild eigentlich dunkel wären, was jedoch allein durch das helle Umgebungslicht kaum zu unterscheiden wäre, wie Ware schreibt [War04, S. 88]. So lassen sich bei dem Licht zugewandten homogen beleuchteten Flächen durch die Glanzpunkte kleinere Strukturen wie leichte Dellen oder etwas emporragende Stellen besser identifizieren.

Wie eben dargelegt sind Reflexionen des Lichtes zwar generell recht zweckdienlich, dies bezieht sich aber ausschließlich auf das Abbilden des specular-Lichtanteils, dass heißt die direkte Reflexion der Lichtquelle ohne Umwege. Inter-Objekt-Reflexionen sind nach Ware im Sinne der Klarheit nicht mehr sinnvoll [War04, S. 254], ließen sich jedoch allein mit Style Transfer Functions auch gar nicht umsetzen.

Insbesondere bei komplexen Szenen zeigt sich nach Ansorge und Leder zudem ein Vorteil bei der Verwendung nicht nur farbiger (siehe Abschnitt 2.4.4), sondern auch texturierter Oberflächen [AL11, S. 110]. Dies ist, wie bereits dargelegt, mit Style Transfer Functions auf Basis der Lit Spheres nur ungenügend umsetzbar, da die Texturierung unabhängig von der räumlichen Position, aber abhängig vom Winkel überall gleich auf das Objekt aufgetragen wird. Der durch Texturen erreichte Tiefeneffekt rührt aber gerade daher, dass sich z.B. parallele Linien einer Textur in der Tiefe verjüngen oder Strukturen mit größerer Entfernung kleiner werden [WFG92].

2.4.4 Farben

Insbesondere Schumann und Müller [SM99, S. 82 ff.] beschreiben einige Phänomene, die es bei der Verwendung von Farben zu beachten gilt. Zunächst einmal ist es eine längst gesicherte Erkenntnis, dass sich farbige Objekte besser eignen als rein achromatische („unbunte“), da der Mensch für Erstere nicht nur eine wesentlich bessere Kontrastempfindlichkeit, weil geringere Schwellenempfindlichkeit, besitzt, sondern auch circa 300 verschiedene Farbnuancen unterscheiden kann, aber nur maximal 150 Helligkeitsstufen. Ansorge und Leder haben im Zuge der Aufmerk-

samkeitsforschung das Salienz⁵-Modell mit einer getrennten Verarbeitung von Farbe und Luminanz bzw. Kontrast erwähnt [AL11, S. 63]. So wirken nach Schumann und Müller farbige Flächen auch heller als achromatische gleicher Helligkeit.

Es spielt also vor allem der Bildkontrast eine führende Rolle, in positiver, manchmal aber auch negativer Hinsicht. Ware nennt störende Kontrasteffekte bei Graustufenbildern, die zu einer deutlich fehlerhaften Wahrnehmung der Helligkeiten führen [War04, S. 94 f.], etwa bei Helligkeitsverläufen, die in starkem Kontrast zur Umgebung stehen. Dagegen ist ein deutlicher Kontrast generell eine der Wahrnehmung förderliche Eigenschaft: In oben genanntem Salienz-Modell beispielsweise geht ein hoher Merkmalskontrast mit erhöhter neuronaler Aktivität einher, so dass ihnen schlussendlich mehr Aufmerksamkeit geschenkt werden kann.

Aber auch bei farbigen Flächen treten Kontrasteffekte auf, wie sie Schumann und Müller zahlreich beschreiben. Durch den Sukzessivkontrast kann je nach Art und Dauer der vorherigen Darstellung die darauffolgende Visualisierung beeinflusst werden. Darum ist es ratsam, die Farbpalette im Verlauf der Visualisierung klein zu halten, was jedoch bei meinen Visualisierungen im Versuch und den Zeitdauern, für welche diese angezeigt werden, kein Problem darstellen sollte. Farbige Flächen können auch gleichzeitig, wenn sie benachbart liegen, das Wahrnehmungsphänomen des Simultankontrastes bzw. Induktion hervorrufen, wobei z.B. die Komplementärfarben als Saum an den Grenzen beider Farbflächen auftreten. Weiterhin gibt es einen Effekt, bei dem ein kleines rotes Objekt auf schwarzem Hintergrund intensiver wirkt und dunkler als größere Objekte gleicher Farbe, da diese mehr Bereiche außerhalb der Fovea bedecken und sich das Verhältnis zwischen Zapfen und Stäbchen verschiebt.

Weiterhin zitieren Schumann und Müller eine Reihe an Farb- respektive Farbkombinations-Rankings, etwa dass die am besten wahrgenommenen Farben in absteigender Reihenfolge gelb, weiß, rot, grün und blau seien. Zu oben genannter Kontrastempfindlichkeit gibt es ebenfalls eine Reihenfolge der niedrigsten benötigten Schwellenempfindlichkeiten, wobei blau-gelb besser ist als rot-grün. Zuletzt sei noch der kinetische Effekt scheinbar schwebender Flächen erwähnt, der besonders bei gelb auf schwarz, orange auf schwarz und rot auf schwarz auftritt.

Was die Farbfehlsichtigkeit betrifft, so sind nach Lehmann et al. vor allem grüne und roten Farben niedriger Sättigung und Helligkeit für einige Menschen problematisch [LOPR97, S. 79 ff.]. Andere Farbtonkombinationen führen dabei noch deutlich seltener zu Erkennungsproblemen. Es ist generell für alle Menschen auch nicht möglich, bestimmte Farbkombinationen zu sehen, wie rötliches Grün oder bläuliches Gelb, welches man unter dem Begriff Farbantagonismus versteht. Diese jeweiligen Gegenfarben führen dann auch bei dem oben genannten Simultankontrast zum scheinbaren Auftreten der entsprechenden Komplementärfarben.

Eine weiteres Problem bei Farbkombination an Grenzflächen liegt in der chromatischen Aberration begründet. Durch die verschiedenen Wellenlängen der Farben sind Kombination von stark gesättigtem Rot und Blau etwa sehr anstrengend für

⁵Eine geeignete Beschreibung für die Salienz existiert mit dem Wort „Deutlichkeit“.

das Auge, da dieses immer nur auf eine Wellenlänge fokussieren kann. Der Effekt tritt auch an der Grenze zwischen Weiß und Schwarz auf, wo die Farbbestandteile des weißen Lichts nicht mehr überlagert wahrgenommen werden und wieder ein leichter Saum entsteht.

2.4.5 Top-down-Einflüsse

Eben aufgezählte Eigenschaften des Menschen in Bezug auf Farb- und Kontrastwahrnehmung sind stark sensorischer Natur. Es liegen aber noch Besonderheiten vor, die erst bei höheren Verarbeitungsstufen auftreten und bestimmte Vorerfahrung respektive spezifisches Wissen voraussetzen (daher engl. *top-down*: der Einfluss höherer Verarbeitung auf die Wahrnehmung). So wurde etwa bei der Schattierung bereits ein Top-down-Einfluss beschrieben, nämlich die Annahme, dass (Sonnen-)Licht immer von oben scheint und so unsere Wahrnehmung schattierter Flächen beeinflusst.

In Bezug auf Farben habe ich bei den Möglichkeiten der Beleuchtung im Volumenrendering bereits das Cool-Warm Shading angemerkt, bei dem das Volumen abhängig davon, ob es einer Lichtquelle zu- oder abgewandt ist, mit kalten oder warmen Farbtönen eingefärbt wird. Das basiert auf der „Wahrnehmung, dass kalte Farben zurückweichen, während warme Farben hervorrücken“, wie Gooch et al. beschreiben [GGSC98]. Dabei zeigen sie anhand von Beispielbildern auch den Vorteil von zusätzlichem Cool-Warm Shading gegenüber der reinen Phong-Beleuchtung: Bei ausschließlicher Nutzung von diffusem Licht und den Lichtreflexen sind Details u.U. in den hellen und dunklen Bereichen des Bildes kaum sichtbar (etwa in den dem Licht abgewandten Bereichen). Ein zusätzliches Cool-Warm Shading ermöglicht die räumliche Schattierung ohne das Licht, denn dieses kann man nun zugleich auf ein mittleres Niveau regulieren, sodass die hellen und dunklen Bereiche nicht über- oder unterbelichtet werden.

Genauso kann, dass haben Ropinski, Steinicke und Hinrichs in Versuchen mit Angiographien gezeigt [RSH06], die farbkodierte Tiefeninformation genutzt werden und stark zum Tiefenverständnis beitragen, indem weiter entfernte Stellen des Volumens anders eingefärbt werden als dicht beim Betrachter liegende. Dies benötigt allerdings wieder Techniken, denen diese Tiefeninformationen zugänglich sind, was die Style Transfer Functions nicht mit einschließt.

Es kann nicht nur gezielt bestimmten Farben eine Bedeutung zugewiesen werden, sondern ganz generell spielt das Top-down-Vorwissen beim Menschen unbewusst für die einem Objekt zugewiesenen Bedeutungseigenschaften eine Rolle, denn es steht von vornherein zur Verfügung und basiert nicht auf der Auswertung von Form oder Struktur des Objekts [Spa06]. Problematisch scheint in Bezug auf Vorerfahrung die Bildung einer Erwartungshaltung an eine medizinische Visualisierung zu sein. So könnte eine Gestaltungsrichtlinie etwa heißen, dass die Farben rot und blau grundsätzlich für Arterien respektive Venen im Blutkreislauf zu verwenden sind [RDRS10]. Damit ist jedoch noch nicht klar, ob es im Gegenzug von Nachteil sein kann, etwa das gesamte Herz in ein rot-blaues Cool-Warm Shading

zu färben. So nennt Boyce dahingehend auch den Begriff der Wahrnehmungsstabilität, der vor allem gerade dann gegeben ist, je weniger Einfluss an Vorwissen vorherrscht und wenn sich die Wahrnehmung nah am visuellen System orientiert [Boy03, S. 194]. So sind z.B. die Wahrnehmung der Helligkeit oder Farbigkeit eines Reizes im Gegensatz etwa zur Räumlichkeit recht stabil.

Fazit. Abschließend sollte man sich – mit dem nun geschaffenen Überblick über die perzeptuellen Eigenheiten – des folgenden Satzes bewusst werden: „Wie die Integration aller Einzelinformationen zu einem Gesamtbild funktioniert oder welche Rolle das Bewusstsein des Menschen spielt, ist bis jetzt ungeklärt.“ [LOPR97, S. 96] Aus dieser Erkenntnis rührt letztendlich auch die von mir anfangs beklagte Tatsache, dass eben kein allgemeingültiger Leitfaden zum Thema existiert, sondern nur ein Katalog an Phänomenen und Sinnestäuschungen, denen der Mensch im visuellen Kontext erliegt. Die Maxime lautet daher, diesen Katalog abzuprüfen, sodass man die aufgezählten Phänomene (z.B. Kontrasteffekte) ausschließen kann und die Hinweise auf geeignete Visualisierungseigenschaften (z.B. Farbrankings) versucht umzusetzen. Die Versuchskonstruktion im nächsten Kapitel habe ich nun daraufhin in möglichst großer Übereinstimmung der jetzt erworbenen Kenntnisse durchgeführt.

3 Versuchskonstruktion

Die Frage nach der optimalen Beleuchtungssituation für medizinische Volumendaten soll empirisch beantwortet werden, folglich bedarf es nun der Konstruktion eines geeigneten Versuchs. Zunächst müssen dafür theoretisch einige Charakteristiken des Versuchs zur statistischen Hypothesenprüfung festgelegt werden, woraus sich die zu lösenden Aufgaben und auch die später anzuwendenden Tests für die Auswertung ergeben. Diese Charakteristiken sind unter dem Begriff der Versuchskonstruktion zusammengetragen, wobei sich am Ende des Kapitels auch Details zur konkreten Versuchsdurchführung finden.

3.1 Vorgehen

Anhand der Literaturrecherche kann man sich orientieren, welche Parameter bei der Erzeugung der Lit Sphere Stilvorlagen zum Abtesten für den Versuch infrage kommen und welche man stattdessen von vorn herein fixieren kann oder sollte. Das heißt, ich suche unabhängige Variablen, die bei dem Versuch variiert werden, um deren Einfluss auf die weiterhin zu definierende abhängige Variable in Form eines Ergebnismaßes der Versuchsteilnehmer (engl. *performance*) zu messen. Dies geschieht in bestimmten Aufgaben, welche zu definieren sind und von den Teilnehmern gelöst werden müssen. Die unabhängigen Variablen werden in Maya bei Erstellung der Lit Spheres manipuliert (siehe Abbildung 3.1), mit denen dann die Ausschnitte des Herzvolumens zur Verwendung in den Aufgaben des Versuches gerendert werden können (wie in Abschnitt 2.2 erläutert).

Zugleich muss man auch vorher mögliche Störvariablen erkennen und deren Gleichverteilung über die Teilnehmer (gleichmäßiger Einfluss auf die Teilnehmer, sodass sie auf die Messwerte immer in gleicher Weise wirken) sicherstellen respektive sie wie unabhängige Variablen gezielt verändern.

Mittels der unabhängigen und abhängigen Variablen ist eine falsifizierbare Hypothese zu generieren, auf die sich hinterher die Auswertung und dabei ermittelte Signifikanzwerte beziehen. Tiefergehende Informationen zur Versuchskonstruktion hält Steffen im Kapitel „Experimental Evaluation Theory“ bereit [Ste10, S. 20 ff.].

Erste Anhaltspunkte für die Hypothese ergeben sich aus der Literaturrecherche sodass diese wie folgt lauten könnte: „Bilder, die mit einer Beleuchtungssituation erzeugt wurden, welche eine räumliche Schattierung ermöglicht, d.h. nicht alles gleich hell oder dunkel ist, sowie durch Materialeigenschaften in Form von deutlich sichtbaren Lichtreflexen und farbigem Cool-Warm Shading (wobei warme Farben dem anscheinenden Licht zugewandt und kalte Farben abgewandt sind) charak-

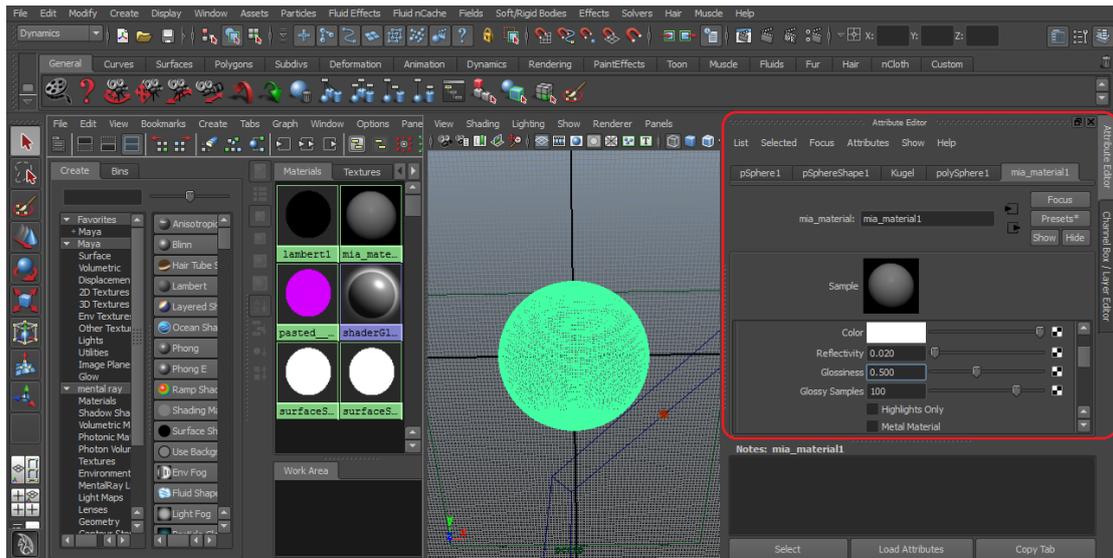


Abbildung 3.1: Änderung der unabhängigen Variablen in Maya durch den „Attribute Editor“ (rot markiert), mit dem sich die entsprechenden Parameter je nach gewähltem Objekt steuern lassen (hier: Glossiness)

terisiert ist, als auch den Erwartungen einer Beleuchtung in der medizinischen Visualisierung entspricht, führen zu den besten Erkennungsleistungen bei räumlichen Darstellungen“. Diese Hypothese wird nach Ermittlung der unabhängigen Variablen mit Verweis auf jene konkretisiert werden.

3.2 unabhängige Variablen

Die Lichtpositionierung in meinem Versuch orientiert sich an der aus der Lichttechnik bekannten Drei-Punkt-Beleuchtung [Mül08], womit sowohl eine gleichmäßige und helle Ausleuchtung als auch durch Ausschalten einzelner Lichtquellen ein besserer Tiefeneindruck durch die dadurch resultierende Schattierung möglich ist, sodass die Lichtpositionierung aufgrund der Möglichkeiten dieser Lichtanordnung als unabhängige Variable ausgeschlossen werden konnte. Der für den Versuch verwandte schematische Aufbau der Beleuchtungssituation ist in Abbildung A.2 auf Seite 44 dargestellt. Dort erkennt man schräg vorn vor dem zu beleuchtenden Objekt das Führungslicht, welches als starker Ausgangspunkt dient und die plastische Schattierung ermöglicht, da an den Seitenrändern des Objekts die Helligkeit stark abfällt. Um diese harte Schattierung zu vermeiden und so die Helligkeit der zu betrachtenden Vorderseite recht konstant zu halten, kann man ebenfalls frontal ein Aufhelllicht zuschalten. Als letztes verwendet man zur Aufhellung der äußersten Ränder seitliche oder ein Hintergrundlicht, sodass keine Partie des Objekts, die potenziell sichtbar ist, im Dunkeln verborgen bleibt. Wie man diese Lichtquellen nun einstellt (Punkt 5. und 6. der jetzt folgenden Auflistung), ist eine Frage der

Abwägung zwischen Schattierung und Sichtbarkeit. Ich habe die beiden letztgenannten Lichtquellen nur diffuses Licht ausstrahlen lassen, sodass keine weiteren Lichtreflexe wie durch das Führungslicht auf der Kugel entstehen.

Folgende unabhängige Variablen, die sich durch Style Transfer Functions (also den Lit Sphere Stilvorlagen) kontrollieren lassen und nach Literaturrecherche sinnvoll sind, habe ich ermittelt sowie dafür infrage kommende Ausprägungen in Mengenschreibweise notiert (siehe Lit Spheres auf S. 45), wobei die Einteilung „Voll - Halb - Aus“ im mathematischen Sinne bedeutet, dass beim Rendern der Lit Spheres unter Maya für den entsprechenden Parameter bei „Halb“ genau nur der halbe Wert gewählt wurde wie für „Voll“, und bei „Aus“ folglich der Wert 0:

1. Cool-Warm Färbung = {gelb-blau, gelb-lila, orange-lila, rot-blau}
2. Cool-Warm Sättigung = {schwarz-weiß, halb, voll}
3. Cool-Warm Richtung = {cool-to-warm, warm-to-cool}
4. Glossiness = {aus, halb, voll}
5. Lichtstärke = {aus, halb, voll}
6. Lichtauswahl = $Pow(\{\text{Führungslicht, Aufhelllicht, Hintergrundlicht}\})$

Dabei ergibt sich insgesamt eine Anzahl an $4 * 3 * 2 * 3 * 3 * 6 = 1296$ möglichen Beleuchtungssituationen (Kardinalität des Kreuzprodukts über alle Mengen). Diese muss man theoretisch alle abtesten, um eine Aussage über die optimale Beleuchtungssituation treffen zu können, was jedoch aufgrund der unhandhabbar großen Anzahl praktisch nicht umsetzbar ist. Daher musste ich die Anzahl reduzieren und eine Auswahl der unabhängigen Variablen vornehmen, wofür es generell keinen korrekten Algorithmus gibt, da dieser logischerweise schon die Ergebnisse eines Versuchs vorwegnehmen würde oder aber es wären eben jene viel zu vielen Versuche notwendig.

Um die Auswahl der wichtigsten und entscheidenden unabhängigen Variablen und deren Ausprägungen dennoch systematisch zu gestalten, blieb mir im zeitlichen Rahmen dieser Bachelorarbeit die Möglichkeit eines nicht-experimentellen Vorversuchs. Das bedeutet, dass dieser weder randomisiert ist, noch ist eine Gleichverteilung gegeben und somit die Zuschreibung einer Ursache möglich. Der Vorversuch kann also höchstens Korrelationen zeigen und soll hier als Beleg für die möglichst systematische und nicht-subjektive Wahl von einer eng begrenzten Anzahl unabhängiger Variablen für den eigentlichen Hauptversuch dienen.

3.2.1 nicht-experimenteller Vorversuch

Den Vorversuch habe ich in Form eines Fragebogens durchgeführt, der auf zwei wichtige Aspekte abzielt, die beim späteren Experiment für die Bearbeitung der dort gestellten Aufgaben relevant werden würden: der Detailliertheit und dem

räumlichen Eindruck der Bilder. Dafür musste ich aufgrund eben jener großen Anzahl an möglichen Beleuchtungssituationen einige unabhängige Variablen parallel abfragen, was nur möglich ist, wenn diese unabhängig voneinander sind. Dafür habe ich folgende Annahmen bezüglich der Unabhängigkeit zueinander getroffen:

- Der Einfluss der Richtung des Cool-Warm Shadings ist unabhängig von den Lichtern und den Reflexionseigenschaften.
- Der Einfluss der Färbung und Stärke des Cool-Warm Shadings ist unabhängig von den Lichtern und den Reflexionseigenschaften.
- Der Beleuchtungseinfluss des restlichen Lichtaufbaus ist somit relativ unabhängig von den Charakteristiken des Cool-Warm Shadings.

Aufgrund dieser Annahmen habe ich drei Bildersets (siehe Seite 46 f.) für den Fragebogen erstellt, die nun teilweise die unabhängigen Variablen parallel abfragen:

1. Cool-Warm Färbung x Cool-Warm Richtung
bei neutralem Licht, d.h. gleichmäßig ausgeleuchtetes Bild, keine Glossiness
ergibt $4 * 2 = 8$ Bilder
2. Lichtstärke x Cool-Warm Färbung x Cool-Warm Sättigung
bei neutralem Licht, d.h. gleichmäßig ausgeleuchtetes Bild, keine Glossiness
ergibt $3 * 4 * 3 = 36$, abzüglich gleicher Bilder¹ = 24 Bilder
3. Lichtstärke x Glossiness x Lichtanzahl
bei einer festen Färbung und Richtung des Cool-Warm Shadings
ergibt $3 * 3 * 6 = 54$, abzüglich gleicher Bilder² = 31 Bilder

Aus diesem online durchgeführten Vorversuch, welcher den Vorteil bietet, dass die Teilnehmer den Fragebogen am Bildschirm ausfüllen müssen wie bei dem eigentlichen Versuch später, ergibt sich durch subjektive Rankings der 23 Teilnehmer bezüglich der Kriterien Detailgrad und räumlicher Tiefe für das erste Set eine der beiden Richtungen als vorteilhafter und für Set 2. und 3. zwei der besten aber am weitesten entfernten Rendering-Bedingungen als Wahl für den Hauptversuch. Die Teilnehmer aus dem Vorversuch haben dann auch nicht mehr am eigentlichen Experiment teilgenommen, da sie nun vorbelastet waren.

Die besten definieren sich dabei durch die Wertung der Rangfolgen für die beiden oben genannten Kriterien, indem ein erster Platz bedeutet, dass diese Rendering-Bedingung mit +5 Punkten gezählt wird, bis hin zum fünften Platz, welcher noch zu +1 Punkt führt. Damit sind die besten Rendering-Bedingungen diejenigen, welche die höchsten Punktzahlen erhalten. Die Distanz wiederum lässt sich durch

¹Bei fehlender Sättigung ergeben sich für jede Lichtstärke oder Cool-Warm Färbung die gleichen Bilder.

²Bei fehlender Glossiness bzw. ausschließlich diffusem Licht ergeben sich ebenfalls identische Bilder für jede Beleuchtung nur mit „specular“-Licht respektive für jede Glossiness.

die Anzahl an Unterschieden in den Parameterausprägungen zweier Renderings definieren (Beispiel zum dritten Bilderset mit drei variablen Parametern: Eine unterschiedliche Lichtstärke und Glossiness, aber eine gleiche Lichtanzahl zweier gewählter Renderings ergibt eine Entfernung zwischen diesen Renderings von 2).

Damit folgt dank einer festzulegenden Cool-Warm-Richtung aus dem ersten Bilderset und nur zwei auszuwählenden Rendering-Bedingungen bei Bilderset 2. und 3. für den Hauptversuch ein 2x2-Design, das heißt zwei Faktoren mit je zwei Faktorstufen (Ausprägungen). Ein Faktor ist dabei die Cool-Warm Färbung aus Bilderset 2., ein weiterer die Lichter- und Materialkomposition aus dem dritten Bilderset. Die Ergebnisse aus dem Fragebogen sahen nun wie folgt aus:

Aus dem ersten Set (siehe Tabelle B.1 auf Seite 49) reihte sich dabei die Bedingung „warm-to-cool“ im direkten Vergleich gegenüber „cool-to-warm“ bei gleicher Färbung immer höher ein. In Bilderset 2. (Tab. B.2) wurden vor allem die Färbungen in gelb-blau und gelb-lila hoch eingeschätzt, dabei führen jeweils die gesättigten und hellen Ausprägungen. Mit dem letzten Bilderset (Tab. B.3) kristallisierten sich die Bedingungen ohne das frontale Aufhelllicht heraus³. Aufgrund der größeren Distanz, aber immer noch guten Punktzahlen, wurden die Rendering-Bedingungen mit schwachem Führungslicht bei starker Glossiness sowie mit starkem Führungs- und Hintergrundlicht bei schwacher Glossiness für den Hauptversuch gewählt. Es ergaben sich also konkret die vier Rendering-Bedingungen, wie in Abbildung 3.2 zu sehen, mit den folgenden zwei Faktoren für den Hauptversuch:

1. Cool-Warm Färbung = {gelb-blau, gelb-lila}
2. Lichter = {halbes Führungslicht links vorn mit starker Glossiness, volles Führungs- & Hintergrundlicht rechts hinten mit schwacher Glossiness}

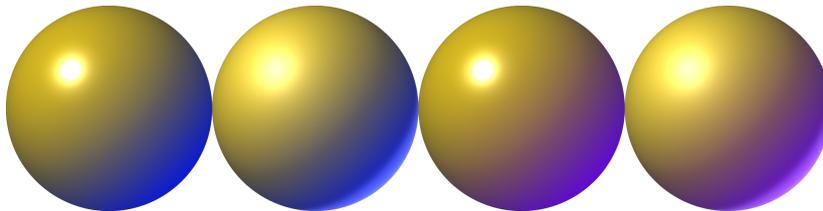


Abbildung 3.2: Die vier für den Hauptversuch ermittelten Lit Spheres

3.3 abhängige Variable

Die abhängige Variable ist ein von den Aufgaben abhängiges Performance-Maß, welches sich im Fall dieser Arbeit durch die Anzahl der gemachten Fehler bei

³In mündlichen Äußerungen haben die Teilnehmer den Bildern mit Aufhelllicht eher weniger Zeichnung und eine schlechtere Tiefenwirkung zugesprochen.

den einzelnen Aufgaben ergibt: Gemessen wird für jede Rendering-Bedingung der Anteil der richtigen Antworten. Dieses habe ich direkt von den Versuchen von Steffen übernommen, da zwei Aufgaben identisch sind und die dritte sich ebenfalls durch eine richtige gegen eine falsche Antwort charakterisiert. Messbar wäre auch die Fehleranzahl unter unterschiedlichen Zeitdauern, was dem Problem der festen Wahl einer Anzeigedauer zwar entgegenwirken würde, aber nur mit für jede Zeitbedingung ausreichenden Versuchsdurchgängen durchzuführen wäre, womit dies nicht infrage kam; das Einführen einer zweiten Zeitdauer hätte somit bereits eine Halbierung der Durchgänge pro Bedingung zur Folge.

3.4 Störvariablen

Störvariablen sind Variablen, die Auswirkung auf die abhängige Variable haben, aber nicht gezielt manipuliert werden (können). Es ergeben sich verschiedene Möglichkeiten mit ihnen umzugehen, wie es auch Steffen schon in seiner Arbeit beschreibt. Ein wichtiges Mittel ist die Randomisierung, um den Einfluss der vor allem unbekanntesten Störvariablen reduzieren zu können. Die für meinen Versuch mir soweit bekannten Variablen habe ich nach ihrer Behandlung eingeteilt:

Störfaktoren, die durch Trainingsdurchgänge ausgeschaltet werden können.

Üblicherweise haben Versuchspersonen auch bei eindeutigen und kurzen Einweisungen nicht immer gleich die Aufgabe korrekt verstanden und die Tastenbelegung verinnerlicht; zudem ist es oft nicht möglich, dass der Versuchsleiter bei jeder Aufgabe zum Instruieren anwesend sein kann. Daher werden die ersten Durchgänge der Aufgaben nicht gewertet, sofern dort noch Missverständnisse oder Benutzungsprobleme aufgetreten sein könnten (Details zu den Aufgaben in Abschnitt 3.6).

Störfaktoren, die als unabhängige Variablen behandelt werden. Sollten sich die Störfaktoren nicht ausschalten lassen, so ist es einerseits möglich, sie durch Zusatzfragen zu kontrollieren und als unabhängige Variablen zu behandeln, was die Versuchsauswertung verkompliziert, denn es sollte immer Ziel sein, so wenige unabhängige Variablen wie möglich zu verwenden, um später klare Schlussfolgerungen ziehen zu können. Von mir wurde zu jedem Teilnehmer notiert, ob ein Vorwissen und somit Trainingseffekt durch bereits absolvierte andere Versuche dieser Art besteht (beispielsweise der Versuch von Steffen oder aus der Veranstaltung „Computer Graphics II“ im Sommersemester 2011), eine Farbfehlsichtigkeit bekannt ist und welche der Rendering-Bedingungen persönlich präferiert wurde. Bei letzterem kann es etwa passieren, dass man eine Rendering-Bedingung für geeigneter hält und deshalb die Antworten zugunsten dieser ausfallen, obwohl man nach Aufgabenstellung dies eigentlich nicht so oft gewählt hätte.

Ziel der Notiz der eben genannten Störvariablen ist es, bei der Auswertung nachzuvollziehen, ob es wahrscheinlich war, dass von ihnen ein Einfluss auf die

Versuchsergebnisse ausgegangen ist und somit die Ergebnisse in Form von Ursachenzuschreibungen verfälschen könnten. Ebenso wurde von mir festgehalten, in welcher Kabine und mit welcher Aufgabenreihenfolge sowie mit welcher Block-Bedingung der Versuch durchgeführt wurde (näheres dazu findet sich in den Abschnitten 3.6 und 3.7).

Störfaktoren, die gleich verteilt sind. Sollte eine Kontrolle nicht infrage kommen, muss versucht werden die Störvariable für jeden Teilnehmer gleich zu verteilen, sodass der Einfluss in jeder Rendering-Bedingung gleich stark ausgeprägt ist.

Der Einfluss, der pro Versuchsperson von der Störvariablen ausgeht, ist dadurch gleich verteilt, da alle Teilnehmer die gleichen Beleuchtungssituationen vorgelegt bekommen und die selben Entscheidungen treffen müssen; variiert sind diese nur in ihrer Reihenfolge des Auftretens, das Auftreten jedoch ist sichergestellt. Man muss beachten, dass die Gleichverteilung natürlich nur die Validität der Ergebnisse unter diesen Einflüssen sicherstellt und man die Ergebnisse nicht zu einer allgemein gültigen Aussage umformuliert (hohe interne, aber niedrige externe Validität).

Für meinen Versuch nahm ich an, dass etwa die subjektiv geschätzte Genauigkeit der gegebenen Antworten bei jeder Person für jede Aufgabe konstant gleich ist. Das Auftreten persönlicher Assoziationen⁴ sollte auch in jedem Durchgang gleichbleibend sein und so die Antworten gleichmäßig beeinflussen.

Weiterhin darf nicht unerwähnt bleiben, dass es sich bei der Cool-Warm Richtung, die durch den Vorversuch festgelegt wurde, auch um eine Störvariable handelt, da sie nicht wie die anderen beiden unabhängigen Variablen manipuliert, sondern kontrolliert und konstant gehalten wird.

Störfaktoren, die schlecht zu kontrollieren sind. Es verbleiben immer Unabwägbarkeiten, ob eine Störvariable nun gleich verteilt ist, oder ob im Rahmen der zur Verfügung stehenden Ressourcen eine Kontrolle nicht möglich ist. Dies wäre etwa bei der Frage der Fall, ob intuitiv geantwortet oder viel Zeit zum Nachdenken aufgebracht wurde. Das ist eigentlich durch eine vorgegebene Antwortzeit kontrollierbar, welche allerdings von Teilnehmer zu Teilnehmer unterschiedlich ausfällt und schlecht zu bestimmen ist.

Außerdem ist gerade bei meinem Versuch problematisch – wie man in Abschnitt 3.6 sehen wird –, dass eine (unbewusste) Vorauswahl der Bildersets durch den Autor des Experiments stattgefunden haben könnte. Dies ist allein schon deshalb nicht auszuschließen, da keine objektiven oder messbaren Kriterien zum Kategorisieren der Bilder verwandt wurden, um Charakteristiken wie die Krümmung einer Arterie oder Rauheit einer Oberfläche festzustellen. Diese müssten zudem auch unter bestimmten Bedingungen (etwa Vergrößerungsstufe) betrachtet werden.

⁴Ein Beispiel aus der persönlichen Erfahrung: Im Experiment von Steffen wurde das braune Phong-Shading von mir als Darstellung von Schokolade und damit tendenziell mit einer positiveren Assoziation wahrgenommen.

3.5 Hypothese

Die Hypothese kann nun aus dem Vorversuch, den dort ermittelten Rankings und mit den bisherigen Erkenntnissen der Literaturrecherche anhand der Wechselwirkung der unabhängigen und abhängigen Variablen formuliert werden und lautet:

Die gelb-bläulichen Rendering-Bedingungen und jene mit beiden zugeschalteten Lichtquellen erzielen höhere Performance-Werte als die Rendering-Bedingungen mit gelb-lila Cool-Warm Shading oder nur dem einen Führungslicht. Demzufolge erreicht die Rendering-Bedingung mit gelb-bläulichem Cool-Warm Shading und beiden zugeschalteten Lichtern (in Kapitel 4 als „blau2“) die höchsten Performance-Werte.

Die Hypothese ist falsifizierbar (sollte sich zeigen, dass die genannte Rendering-Bedingung keine signifikant besseren Performance-Werte liefert) und spricht den unabhängigen Variablen als Ursache eine bestimmt geartete Wirkung auf die abhängige Variable zu.

3.6 Aufgaben

Die idealen Aufgaben zum Testen der Eignung einer Rendering-Technik in der medizinischen Visualisierung wären natürlich jene, wie sie im klinischen Alltag auch vorkommen. Bei Herzvolumen wäre dies z.B. typischerweise die Suche nach Ablagerung an den Koronararterien (so genannte Plaques). Es handelt sich jedoch fast ausschließlich um Aufgaben mit hohem kognitiven Aufwand, was sie für einen Benutzertest eher disqualifiziert. Das Ziel müssen hierbei einfache Aufgaben sein, um so viel Streuungseinfluss wie möglich zu vermeiden, der durch die Denkprozesse und Verwendung von Vorwissen entsteht. Außerdem ist es den zur Verfügung stehenden Versuchsteilnehmern (Studenten der Informatik aus einer Veranstaltung für das Masterstudium) gar nicht möglich, ohne Fachwissen und Erfahrung anatomische Anomalien in medizinischen Visualisierungen zu identifizieren. Im Gegenzug erlaubt das Zurückgreifen auf gewöhnliche Studenten aber eine hinreichend große Zahl an homogenen Versuchspersonen, um statistische Aussagen treffen zu können.

Daher muss die eigentliche Aufgabe in kleinere und einfachere Teilaufgaben aufgeteilt werden, die nicht auf dem Erfahrungsschatz eines Experten beruhen, aber notwendig sind, um die übergeordnete Aufgabe zu bearbeiten. Steffen nannte bei seiner Versuchskonstruktion eine Reihenfolge der folgenden Aufgaben um zur Erkennung eines Plaques zu gelangen: Rauheit einschätzen, Arterien-Krümmung einschätzen, deren Arterienwandgrenzen feststellen und die Lokalisierung eines Punktes auf der Arterie. Das bedeutet nicht, dass dies alle Teilaufgaben wären, die zur Bearbeitung der übergeordneten Aufgabe notwendig sind. Eine Verbesserung der Visualisierung, die zunächst nur zu einer Verbesserung bei einer Teilaufgabe beiträgt, ermöglicht jedoch automatisch eine bessere Bearbeitung der eigentlichen

Aufgabe durch das medizinische Fachpersonal (sofern zeitgleich keine Verschlechterung bei den restlichen Aufgaben auftritt).

Es wurden im Zuge der Ergebnisse von Steffens Experiment bereits Aufgaben als ungeeignet ausgeschlossen und so übernahm ich nur zwei dieser Aufgaben zu Rauheit und Krümmung, auf die nach meiner Einschätzung die Rendering-Bedingung noch den größten Einfluss haben dürfte. Ich behielt hier auch die englische Bezeichnung bei. Zusätzlich, um die räumliche Orientierung als Teil der übergeordneten Aufgabe zu testen, habe ich als dritte Aufgabe eine neu geartete hinzugefügt.

Aufgabe „Surface“. Die erste Aufgabe diente der Erkennung von Oberflächen zu Objekten und deren Unterscheidung gegenüber dem Rauschen in den Daten. Denn ein medizinischer Volumendatensatz enthält für gewöhnlich nie solche idealisierten Daten wie es etwa das L-förmige Objekt auf Seite 8 vermuten lassen könnte, sondern die Werte schwanken und werden als Rauschen beim Freistellen des Volumens sichtbar. Die Versuchspersonen mussten zwei verschieden raue Ausschnitte eines 3D-Renderings eines Herzens miteinander vergleichen und entscheiden, welcher von beiden rauher ist. Diese Konstruktion folgt wieder dem Paradigma, die Aufgaben so einfach wie möglich zu gestalten. Hätte man stattdessen den Teilnehmer um eine konkrete Werteangabe bezüglich der Rauheit der Ausschnitte gebeten, wären beispielsweise Fragen der Skalierung aufgetreten die sich beim paarweisen Vergleich vermeiden lassen.

So bekam jede Versuchsperson in 80 Durchgängen (siehe beispielhaft die zwei Durchgänge in Abbildung 3.3) insgesamt 40 verschiedene Ausschnitte zu sehen⁵, wobei jedem Versuchsteilnehmer wieder die gleichen Ausschnitte gezeigt wurden und jeder Ausschnitt immer in jeder der vier Rendering-Bedingungen geprüft wurde. Die Reihenfolge war jedoch randomisiert, um verschiedenste Störeffekte, wie z.B. Erinnerung oder Müdigkeit, auszuschließen. Dabei wurde in jedem Durchgang aber immer nur eine Rendering-Bedingung abgefragt, das heißt die beiden gezeigten Ausschnitte waren in derselben Rendering-Bedingungen gerendert worden. Das Antworten für jeden Durchlauf erfolgte intuitiv mit den Pfeiltasten als Angabe für den linken oder rechten Ausschnitt. Diese Bedienung wurde für alle Aufgaben beibehalten.

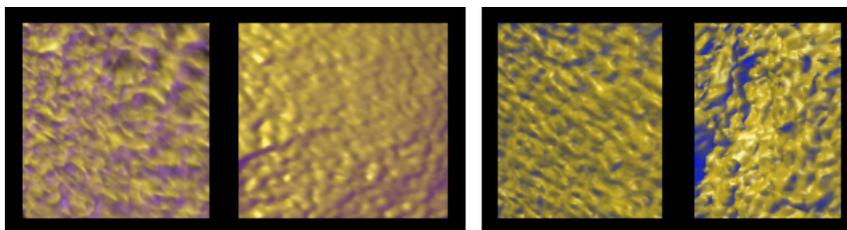


Abbildung 3.3: Zwei Beispieldurchgänge der Aufgabe Surface

⁵80 Durchgänge ergeben sich bei 40 Ausschnitten durch $4 \frac{\text{Rendering-Bedingungen}}{\text{Ausschnitt}} * 2 \frac{\text{Ausschnitte}}{\text{Durchgang}}$.

Die Referenzeinteilung der Rauheit erfolgte durch den Autor in vier verschiedene Kategorien, von „wenig rau“ bis „sehr rau“. So wurden auch bei dem Versuch Vergleiche zwischen Ausschnitten der gleichen Rauheits-Kategorie ausgeschlossen, ebenso wie Vergleiche zwischen der ersten und letzten Kategorie, die sehr leicht zweifelsfrei hätten beantwortet werden können, obwohl schon die Anzeigedauer für jeden Ausschnitt auf 205 Millisekunden (arithmetisches Mittel aus den beiden in Steffens Versuch verwandten Zeiten) begrenzt wurde. Die Ausschnitte eines Durchgangs wurden nacheinander angezeigt: Dabei wird der rechte Ausschnitt nach Ausblenden des linken eingeblendet und aufgrund der notwendigen Augenbewegung von links nach rechts 20 Millisekunden⁶ länger angezeigt.

Aufgabe „Curvature“. Eine zweite Aufgabe hatte die Krümmung von Arterien zum Gegenstand. Die Konstruktion orientiert sich stark an der Surface-Aufgabe, wie man auch an den zwei Beispieldurchgängen in Abbildung 3.4 erkennt. Die Versuchspersonen wurden diesmal gebeten anzugeben, in welchem der zwei Ausschnitte die jeweils gezeigte Arterie stärker gekrümmt ist. Dabei wurde natürlich im Voraus darauf geachtet, dass immer nur eine Arterie in jedem Ausschnitt sichtbar ist. Die Relevanz dieser Teilaufgabe für die übergeordnete Aufgabe des Mediziners erklärt sich dadurch, dass deren Krümmung ein indirekter Hinweis darauf sein kann, dass sich Plaques in der Arterie gebildet haben.

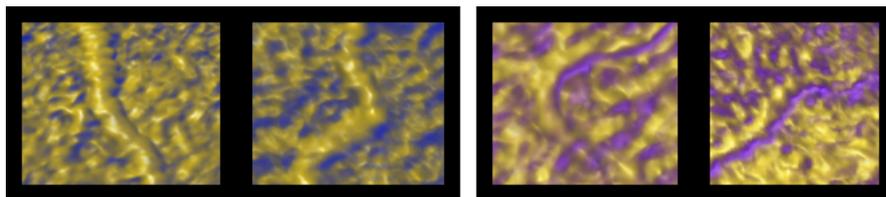


Abbildung 3.4: Zwei Beispieldurchgänge der Aufgabe Curvature

Auch die Anzahl an Durchgängen und Ausschnitten war identisch mit der ersten Aufgabe, jedoch wurden die Ausschnitte diesmal in sechs statt vier Kategorien eingeteilt (von „nicht gekrümmt“ bis „größtmöglich gewunden“), wobei hier ebenfalls bei der Präsentation der Paare darauf geachtet wurde, nur solche Ausschnitte in einem Durchgang zu vergleichen, die sich mindestens um eine, aber maximal um zwei Kategoriestufen unterscheiden, um sowohl die sehr einfachen als auch die nicht lösbaren Paarungen auszuschließen.

Bei der Aufgabe Curvature habe ich auch Konsequenzen aus einem Vortest ziehen können, welcher mit neun Personen durchgeführt wurde, um das Versuchsprogramm zu testen und gegebenenfalls auf bisher nicht erkannte Probleme reagieren zu können. Die Präsentationsdauer der Ausschnitte habe ich hierbei auf 250 Millisekunden in der Folge erhöht, da die Ergebnisse der neun Personen bei dieser Auf-

⁶Die Zeit von 20 Millisekunden orientiert sich an der Geschwindigkeit von Sakkaden. Deren Latenz bis zur Ausführung wird mit Beginn der Anzeige des linken Bildes überbrückt, sodass man in etwa die gleiche Zeit für den rechten Ausschnitt zur Verfügung hat wie für den linken.

gabe stärker Richtung Ratequote tendierten als bei den anderen beiden Aufgaben. Interessanterweise zeigte sich im Vortest noch ein Mittelwert-Vorsprung der gelbbläulichen Rendering-Bedingungen mit beiden zugeschalteten Lichtquellen um bis zu neun Prozent Unterschied gegenüber den anderen Rendering-Bedingungen (also wie in der Hypothese behauptet). Ähnliches galt auch für die letzte Aufgabe mit teilweise bis zu fünf Prozentpunkten, während die Ergebnisse bei der Surface-Aufgabe bereits untereinander auf einem ähnlichen Niveau lagen.

Aufgabe „Depth“. Die letzte Aufgabe wurde aus dem Grund entwickelt, dass gerade bei dreidimensionaler Visualisierung die Tiefengestaltung der Objekte essentiell ist, um etwa Arterien zu lokalisieren. Denn diese liegen immer an ähnlicher Position und sind dementsprechend so zu finden, indem man die Ausrichtung des Herzvolumens identifiziert, um dann an der richtige Stelle nach ihnen zu suchen. Des Weiteren kann bereits eine abnorme Form in der Tiefe Hinweise auf Beeinträchtigungen in einer Arterie liefern, wie sie sonst erst mit der direkten Draufsicht auf die Arterien wie in der Curvature-Aufgabe möglich werden. So ist gezielteres Suchen dieser Stellen möglich, indem etwa eine stärker werdende kalte Färbung durch das Cool-Warm Shading ein Hinweis auf die Krümmung eines Objekts in die Tiefe ist, wie man es z.B. beim rechten Beispiel von Abbildung 3.5 erkennt.

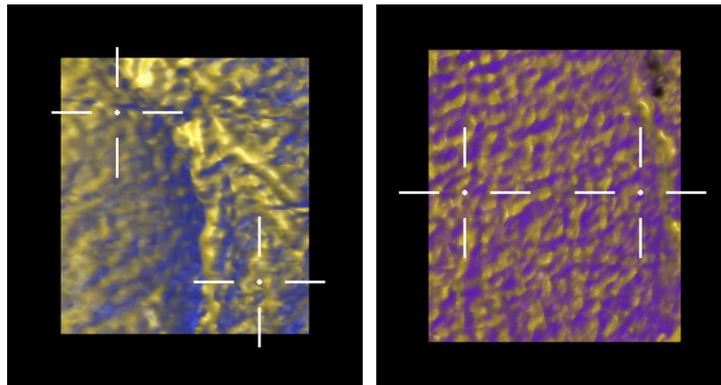


Abbildung 3.5: Zwei Beispieldurchgänge der Aufgabe Depth

Die Versuchspersonen hatten nun anzugeben, welche der zwei mit einem Kreuz markierten Positionen der Oberfläche näher am Betrachter liegt. Dabei wurde in einem Durchgang im Gegensatz zu den beiden vorherigen Aufgaben nur ein Ausschnitt gezeigt, schließlich müssen beide Positionen in einem zusammenhängenden Bereich liegen. So gab es in dieser Aufgabe nur eine halbierte Anzahl an 40 Durchgängen, dennoch wurden dabei wie in den anderen Aufgaben 40 Ausschnitte gezeigt. Der Grund dafür liegt an den Erinnerungseffekten, die in den Vortests spürbar wurden. Um diese auszuschließen wurde jeder Ausschnitt in dieser Aufgabe für jede Versuchsperson diesmal nur in einer Rendering-Bedingung gezeigt. Dabei griff ich auf ein Block-Design zurück, sodass vier Versuchspersonen nacheinander die vier verschiedenen Rendering-Bedingungen gezeigt bekamen, bevor sich

dies wiederholte. Dabei ist wieder sichergestellt worden, dass auch jeder vierte Teilnehmer exakt die gleichen Rendering-Bedingung der jeweiligen Ausschnitte präsentiert bekam, wie zuvor. Natürlich waren dennoch alle vier Rendering-Bedingungen gleich oft in den 40 Durchläufen einer Person vertreten (jeweils zehn mal).

Ebenfalls reduziert wurde die Anzahl der vorangestellten Probedurchgänge: In den ersten beiden Aufgaben waren dies vier, in der Aufgabe Depth nur noch zwei. Denn die Bilder wurden statt der sehr kurzen Zeit, die hier nicht möglich war, zehn Sekunden lang angezeigt, bevor sie ausgeblendet wurden, um eine Antwort zu erwarten. Es braucht bei dieser Aufgabe schließlich eine umfangreichere Betrachtung des Ausschnitts, um dabei eine Entscheidung zu bilden; gleichzeitig ist das Verständnis zu Beginn auch schneller gegeben.

3.7 Versuchsdurchführung

Wie in der Aufgabe Depth mit den Rendering-Bedingungen wurde auch die Reihenfolge der drei Aufgaben geblockt, sodass sie sich nach dem sechsten Teilnehmer wiederholte (sechs mögliche Permutationen bei drei Aufgaben). Die Versuchspersonen wurden in vier Kabinen in der zufälligen Reihenfolge ihrer Ankunft aufgeteilt, wobei die Belegung der Kabinen natürlich von der Dauer für die Bearbeitung bei jeder einzelnen Versuchsperson abhing. Dabei muss angemerkt werden, dass der Fragebogen aus dem Vorversuch im Anschluss an das Versuchsprogramm auch noch einmal ausgefüllt wurde, um zu ermitteln, ob sich die Daten aus dem Vorversuch mit einer diesmal größeren Anzahl an Personen bestätigen würden.

Zu den Kabinen ist anzumerken, dass in einer dieser das Versuchsprogramm nicht automatisch beendete und somit ein Eingreifen des Versuchsleiters erforderte und in einer anderen der vier Kabinen sich statt eines CRT-Monitors ein LCD-Display befand. In allen wurde jedoch eine ausreichende Grundhelligkeit durch Zuschalten von Klemmlampen sichergestellt und nach Versuchsdurchführung durch keine Kabine ein Einfluss als Störfaktor auf die Versuchsergebnisse festgestellt.

Schlussendlich haben 67 Versuchspersonen teilgenommen, fast ausschließlich aus dem Kurs „Daten- und Informationsvisualisierung“, denen die Aufgaben im Voraus natürlich nicht bekannt waren. Dabei wäre eine ideale Anzahl an Personen ein Vielfaches von 24 gewesen, da die Blockbildung von Reihenfolge und in der Depth-Aufgabe zu 24 ($= 6 * 4$) möglichen Kombinationen führte. Es gab jedoch insgesamt genügend Teilnehmer pro Kombination, sodass dies vernachlässigbar ist. Auch die mehr als doppelt so vielen Teilnehmer wie bei Steffen sind mit Blick auf doppelt so viele Rendering-Bedingungen notwendig geworden.

Mit einer Ausnahme wurde der Versuch mit allen Versuchspersonen innerhalb zweier Tage im PsyLab der Universität Paderborn durchgeführt. Die Teilnehmer brauchten mit dem Versuchsprogramm und dem anschließenden Ausfüllen des Fragebogens etwa 20 Minuten, bevor sie die Teilnahme als Bonus für die oben genannte Veranstaltung angerechnet bekamen.

4 Versuchsauswertung

4.1 Vorgehen

Die Auswertung der während des Versuchs gesammelten Daten erfolgte in zwei Schritten. Zunächst wurde die Performance einer jeden Versuchsperson pro Aufgabe und Rendering-Bedingung berechnet, das heißt für die je 80 Durchgänge bei den Aufgaben Surface und Curvature sowie die 40 Durchgänge der Depth-Aufgabe das arithmetische Mittel gebildet, sodass sich pro Versuchsperson $3 * 4 = 12$ Mittelwerte ergeben (Aufgabe x Rendering-Bedingung). Aus diesen Daten kann man nun pro Rendering-Bedingung und Aufgabe über alle Versuchspersonen mitteln, um einen Überblick über die erreichten Gesamt-Performances der Rendering-Bedingungen zu bekommen, denen schließlich mein Interesse gilt. Man betrachtet dabei alle drei Aufgaben parallel aber getrennt voneinander, da natürlich unterschiedliche Aufgaben auch unterschiedlich profitieren können und die Konstruktion der Aufgaben nicht von vornherein zwischen ihnen vergleichbare Performance-Werte ermöglicht. In Abbildung 4.1 sind außerdem die Konfidenzintervalle, in denen sich mit einer Wahrscheinlichkeit von 95% der tatsächliche Mittelwert der zugrunde liegenden Grundgesamtheit befindet, als Fehlerbalken abgebildet.

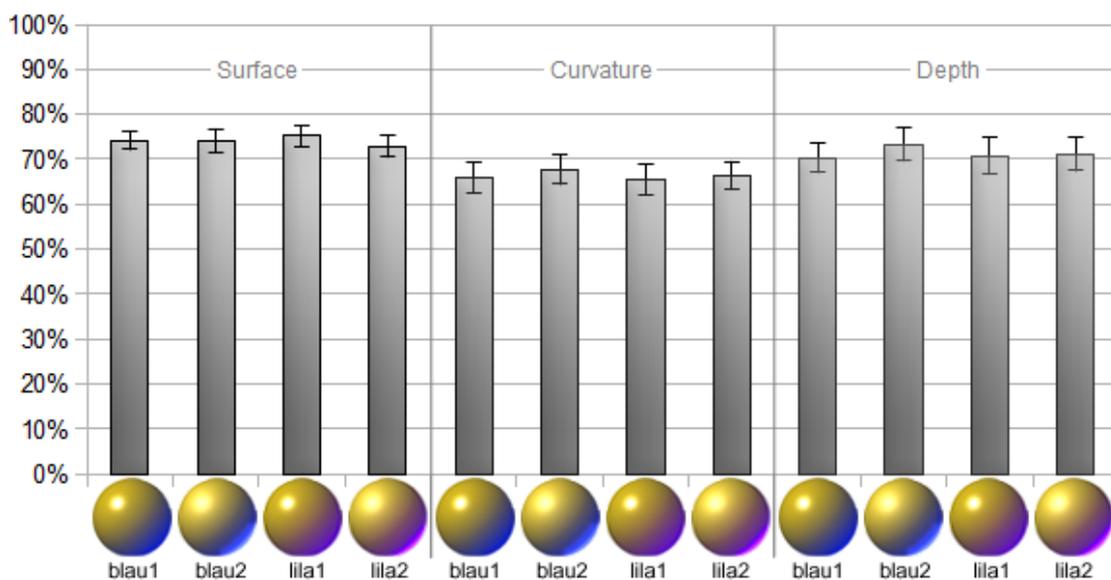


Abbildung 4.1: Übersicht der Gesamtperformance pro Aufgabe und Rendering-Bedingung mit jeweiligen 95%-Konfidenzintervall als Fehlerbalken

In der gesamten Auswertung wurden die vier Rendering-Bedingungen des Versuchs aufgrund der leichteren Wiedererkennung mit „blau1“, „blau2“, „lila1“ sowie „lila2“ bezeichnet, wobei „blau“ für das Cool-Warm Shading von gelb nach blau steht (analog „lila“ für gelb nach lila), und die Zahl „1“ für die Konfiguration mit einem Führungslicht halber Leistung bei voller Glossiness des Materials, sodass die Zahl „2“ auf der anderen Seite für die beiden voll leuchtenden Lichtquellen des Führungs- und Hintergrundlichts bei halber Glossiness steht (siehe Seite 23).

Im zweiten Schritt erfolgte die statistische Auswertung, um festzustellen, ob sich signifikante Unterschiede zwischen den Rendering-Bedingungen ergeben haben. Mein Vorgehen dabei orientierte sich vor allem an dem Kapitel „Result Analysis“ in der Arbeit von Steffen sowie bei der Auswahl weiterer Tests an dem Leitfaden auf den Webseiten der Arbeitsgruppe „Computergrafik, Visualisierung und Bildverarbeitung“ [Lei]. Bei der Anwendung der Tests habe ich auf eine zeitlich beschränkte Version der Statistik-Software „SPSS Statistics 17.0“ zurückgegriffen.

Durch die Konstruktion des Versuchs ergibt sich auch der Test, der zum Auswerten anzuwenden ist. Es handelt sich in dem Versuch um eine abhängige Stichprobe, da jede Versuchsperson jede der Rendering-Bedingungen und folglich jede der Faktorausprägungen absolviert hat, womit vor allem Versuchspersonen-bezogene Effekte bei den einzelnen Rendering-Bedingungen ausgeschlossen werden konnten. Zu diesem sogenannten „Within-Subjects“-Design kann man in SPSS-Notation einen gepaarten t-Test für abhängige Stichproben machen, wie ihn Steffen bei seinem Experiment benutzte.

Da wir nun aber mehr als nur zwei Rendering-Bedingungen vorliegen haben (aufgrund der zwei gekreuzten Faktoren des zweifaktoriellen Designs), reicht der Student'sche t-Test hierbei aber nicht aus, um auch Interaktionen der zwei im Versuch verwandten Faktoren feststellen zu können. Somit weiche ich bei der Auswertung von Steffens Vorgehen ab und greife auf eine Varianzanalyse mit Messwiederholung zurück, die ähnlich dem t-Test die Mittelwerte vergleicht und deren Varianzen zwischen und innerhalb der Rendering-Bedingungen auf Signifikanz hin untersucht – als mehrfaktorielle Varianzanalyse eben auch für mehr als nur zwei Rendering-Bedingungen (= zwei Ausprägungen eines einzelnen Faktors). Daher sind auch die Signifikanzwerte, welche man für die beiden einzeln betrachteten Faktoren in der Varianzanalyse erhält, mit denen aus einem paarweisen t-Test jedes Faktors identisch und man kann sich die Anwendung dieses Tests ersparen.

Folgende Prozedur wurde unter Beachtung der eben genannten Argumentation zur Auswertung der vier Rendering-Bedingungen für jede Aufgabe durchlaufen:

1. Student'scher t-Test zur Überprüfung zufällig geratener Versuchsergebnisse
2. Kolmogorov-Smirnov-Test zur Überprüfung auf Normalverteilung
3. Varianzanalyse zur Überprüfung auf Unterschiede innerhalb der beiden Faktoren sowie der vier Rendering-Bedingungen (Interaktion der Faktoren)
4. Varianzanalyse zur Überprüfung des Einflusses der einzelnen Störfaktoren

Da fast alle Maße und Tests, wie Konfidenzintervall, Student'scher t-Test oder die Varianzanalyse, annähernd normalverteilte Daten voraussetzen, muss dies auch vor der Interpretation der Ergebnisse sichergestellt sein. Daher habe ich in der eben genannten Prozedur auch den Kolmogorov-Smirnov-Test auf Normalverteilung mit einbezogen¹. So kann es Hinweise geben, dass bei keinen signifikanten Unterschieden zwischen den Rendering-Bedingungen dies nicht allein daran liegt, dass es keine gibt, sondern von vornherein die Aufgabe nicht normalverteilte Performancewerte liefern kann, was der Aufgabe die Eignung zur Beantwortung der Hypothese abspricht. Denn bei dem Versuch sollten annähernd normalverteilte Daten herauskommen, wie schon Steffen im Kapitel „Experimental Evaluation Theory“ beschrieb.

4.1.1 Umgang mit den Ergebnistabellen

In der folgenden Auswertung der einzelnen Aufgaben sind der Mittelwert der Stichprobe, der Standardfehler sowie die Signifikanzwerte zu den Tests der Punkte 1. bis 3. der auf der vorherigen Seite dargelegten Prozedur aufgelistet. Ich habe mich dabei auf die Angabe des Standardfehlers als Streuungsmaß der Stichprobe beschränkt, da sich daraus sowohl die Standardabweichung als auch die Grenzen eines Konfidenzintervalls, wie sie in Abbildung 4.1 als Fehlerbalken dargestellt sind, leicht berechnen lassen.

Innerhalb des Standardfehlers bewegen sich etwa 68% der Mittelwerte, das heißt innerhalb jener Grenzen, die sich ergeben, wenn man den Standardfehler dem Mittelwert hinzuaddiert oder ihn davon abzieht. Dies ist natürlich nur unter Voraussetzung der Normalverteilung korrekt. Beim Konfidenzintervall zu 95% sind es dann 95 von 100 ermittelten Mittelwerten, die sich innerhalb eines Bereichs von circa $\pm 1,96$ Standardfehlern um den Mittelwert bewegen (vgl. Abb. 4.1). Außerdem ergibt sich die Standardabweichung der Population aus etwas mehr als dem Achtfachen des Standardfehlers, was sich mit $\sqrt{67}$ als Multiplikator berechnet (Teilnehmeranzahl $N = 67$). Diese Hinweise sollen hier nur als grobe Umrechnungsregel dienen, sofern benötigt; Details zu den Berechnungen sind bei Bedarf dem „Experimental Evaluation Theory“-Kapitel von Steffen zu entnehmen.

Die angegebenen Signifikanzwerte beziehen sich auf die Wahrscheinlichkeit des Zutreffens der jeweiligen Nullhypothese H_0 , wie sie in der Tabelle hinter „Signifikanz:“ vermerkt ist. Ab einem Wert kleiner 0,05 (Signifikanzlevel von 95%) kann also die Nullhypothese verworfen und die Alternativhypothese H_1 angenommen werden. Beispielsweise weiß man beim t-Test auf geratene Ergebnisse, dass in jeder Aufgabe mit 50% Wahrscheinlichkeit korrekt geraten werden kann. Daher untersucht man, ob die Werte der gemessenen Stichprobe zu einer Verteilung mit dem Erwartungswert von 0,5 gehören, was H_0 entspräche. Man „erhofft“ sich hierbei eine Signifikanz von $< 0,05$ zu messen, sodass H_0 verworfen und H_1 ($\hat{=}$ „keine zufällige Verteilung“) angenommen werden darf.

¹Die Wahl fiel aufgrund der großen Anzahl an Messungen nicht auf den Shapiro-Wilk-Test.

4.2 Auswertung der Aufgabe Surface

In der folgenden Tabelle 4.1 finden sich die Versuchsergebnisse der ersten Aufgabe. Ich möchte dabei anmerken, dass hier nicht wie bei Steffen eine Unterscheidung der Schwierigkeiten vorgenommen wurde (also eine Berücksichtigung der bei Aufgabenkonstruktion getätigten Kategorisierung der Bildausschnitte in Form eines weiteren Faktors), was ebenso auf die zweite Aufgabe zutrifft. Denn die Kategorisierungen sind versuchsimmanent, da in der Praxis nicht nach bestimmten Krümmungs- oder Rauheitsgraden unterschieden wird, sondern diese gemischt auftreten. Die Einteilung diente einzig der effizienten Versuchsdurchführung und des Ausschlusses nicht sinnvoller Bildpaare (wie schon in Abschnitt 3.6 erläutert).

Tabelle 4.1: Versuchsergebnisse zur Aufgabe Surface

Surface	Cool-Warm Shading * Lichtaufbau				Cool-Warm Sh.		Lichtaufbau	
	blau1	blau2	lila1	lila2	blau	lila	1	2
Mittelwert	0,741	0,742	0,752	0,729	0,741	0,741	0,747	0,735
Standardfehler	0,01	0,013	0,012	0,012	0,01	0,01	0,009	0,011
Signifikanz: gleiche Population	0,162				0,941		0,183	
Signifikanz: geratene Ergebnisse	< 0,001	< 0,001	< 0,001	< 0,001				
Signifikanz: keine Normalverteilung	0,002	< 0,001	< 0,001	0,001				

Auffällig bei der Betrachtung der Mittelwerte ist der Fakt, dass sich diese zunächst ziemlich gut in der Mitte einreihen – weit entfernt von geratenen Ergebnissen bei 50% oder den vollen 100%. Die Unterschiede zwischen den Rendering-Bedingungen sind recht klein und auch nicht signifikant, obwohl hierbei noch der beste Signifikanzwert aller drei Aufgaben ermittelt wurde. Ebenso liegen die Konfidenzintervalle recht eng um die Mittelwerte (erkennbar an den kleinen Standardfehlern), sodass davon ausgegangen werden kann, dass sich die tatsächlichen Mittelwerte für diese Aufgabe dicht bei den hier gemessenen befinden. Auch können wir, wie erwartet, sicher von einer Normalverteilung ausgehen, um etwa die vorherige Aussage überhaupt treffen zu können.

Interessanterweise erreicht der Faktor des Cool-Warm Shadings in der Einzelbetrachtung in beiden Ausprägungen exakt die gleichen Mittelwerte und eine Signifikanz, von der fast mit hoher Wahrscheinlichkeit behauptet werden kann, dass zwischen den Cool-Warm Färbungen in der Aufgabe Surface kein Unterschied auszumachen ist, also einer „Signifikanz: gleiche Population“ von $p > 0,95$.

Betrachtet man die Lichtkonfiguration, das heißt die aufgestellten Lichter und den Reflexionsfaktor (Glossiness) des Oberflächenmaterials, so ergibt sich ein ähnlicher Signifikanzwert wie für eine Interaktion beider Faktoren. Dabei liegt die erste Lichteinstellung ungefähr um einen Prozentpunkt vorn; auch bei der Interaktion beider Faktoren weist die Rendering-Bedingung „lila1“ ebenfalls nur einen marginalen Vorsprung von genau einem Prozentpunkt gegenüber dem nachfolgenden „blau2“ auf.

Die Störvariableneinflüsse wurden mit einer einfaktoriellen Varianzanalyse einzeln überprüft, ob sich durch sie ein offensichtlicher Einfluss auf die als abhängig betrachteten Rendering-Bedingungen feststellen lässt; demnach wurde nur die Auswirkung eines einzelnen Störfaktors innerhalb einer Rendering-Bedingung überprüft. Eine Behandlung der Störvariablen als weitere Faktoren des Versuchs kam aufgrund der rapide steigenden Anzahl dafür benötigter Versuchsteilnehmer nicht in Betracht. Insbesondere wurde zunächst bei allen Störvariablen – sofern denn möglich – eine Gleichverteilung sichergestellt.

Obwohl die Aufgaben-Reihenfolge gleichmäßig verteilt war ergab sich einzig bei der Aufgabe Surface in der Rendering-Bedingung „blau2“ ein einzelner schwach signifikanter Unterschied ($p < 0,05$) durch die Reihenfolge. Dieser „Ausreißer“ ist aber allein nicht aussagekräftig; schließlich wurden eben alle Teilnehmer auch gleichmäßig über die Bedingungen verteilt und dabei, bis auf eine Permutation, alle Reihenfolgen von 11 Teilnehmern bearbeitet.

Es sind im Übrigen auch keine signifikanten Unterschiede zwischen den Ergebnissen von Teilnehmern, welche eine Farbschwäche angaben, dies verneinten oder dies nicht bekannt war, in keiner der Aufgaben aufgetreten, sodass sowohl in der Surface-Aufgabe als auch den beiden folgenden alle Ergebnisse diesbezüglich zusammen betrachtet wurden.

4.3 Auswertung der Aufgabe Curvature

In der zweiten Aufgabe reihen sich die Mittelwerte um 66% ein, maximal ein Prozentpunkt mehr oder weniger, wobei „blau2“ die Rendering-Bedingung mit dem höchsten Wert ist. In Tabelle 4.2 kann man auch ablesen, dass dieser Vorsprung überhaupt nicht auf einem signifikanten Unterschied beruht. Erkennbar liegen die Mittelwerte ungefähr zehn Prozentpunkte unter denen der ersten oder auch dritten Aufgabe bei bereits größeren Standardfehlern als in der Surface-Aufgabe. Obwohl die Mittelwerte näher an der Ratequote liegen, sind sie mit hoher Sicherheit nicht geraten. Allerdings scheinen sie damit schwieriger gewesen zu sein. Und tatsächlich habe ich ja auch bei der Kategorisierung bezüglich der Anzahl an Kategorien (hier sechs statt vier, wie in der ersten Aufgabe) „feiner“ unterschieden.

Tabelle 4.2: Versuchsergebnisse zur Aufgabe Curvature

Curvature	Cool-Warm Shading * Lichtaufbau				Cool-Warm Sh.		Lichtaufbau	
	blau1	blau2	lila1	lila2	blau	lila	1	2
Mittelwert	0,66	0,677	0,655	0,664	0,668	0,66	0,657	0,671
Standardfehler	0,017	0,016	0,017	0,016	0,015	0,014	0,015	0,014
Signifikanz: gleiche Population	0,702				0,388		0,218	
Signifikanz: geratene Ergebnisse	< 0,001	< 0,001	< 0,001	< 0,001				
Signifikanz: keine Normalverteilung	0,013	< 0,001	0,175	0,01				

Auch ist bei der Curvature-Aufgabe der Signifikanzwert für Unterschiede zwischen dem Lichtaufbau näher an der Signifikanzschwelle, wobei er mit $p > 0,2$ immer noch weit entfernt von einer schwachen Signifikanz um 0,05 ist.

Es fällt auf, dass bei der Bedingung „lila1“ der Test auf Normalverteilung scheitert, was man jedoch als einzelnen Ausrutscher betrachten kann. Allerdings haben auch die anderen Bedingungen in dieser Aufgabe nicht unbedingt zu stark signifikanten Werten bezüglich einer Normalverteilung geführt, was zusammen mit den schwächeren Leistungswerten der Teilnehmer zur Einschätzung führt, dass diese Aufgabe unter allen dreien die generell am schlechtesten konstruierte ist. Dies geht einher mit der Auffälligkeit, dass zwar die Aufgabe nicht konsequent von einzelnen Versuchspersonen falsch verstanden wurde, aber bei einigen wenigen Teilnehmern in einzelnen Rendering-Bedingungen die Ratequote knapp unterschritten wird, wohingegen sie bei den anderen Rendering-Bedingungen darüber liegt. Es wurde mit Blick auf die fehlende objektive Einschätzung der Korrektheit des Ergebnisses (subjektive Kategorisierung) auch kein einziger Teilnehmer aus der Auswertung ausgeschlossen. Ein vermehrtes Auftreten solcher „Ausreißer“ in einer Aufgabe wäre schließlich ein Hinweis darauf, dass sie nicht geeignet ist.

Auch in der Aufgabe Curvature gab es bei der Rendering-Bedingung „blau1“ einen schwach signifikanten Unterschied ($p < 0,05$) zwischen den Reihenfolgen der Aufgaben, der allerdings allein keine große Aussagekraft besitzt. Außerdem soll die Vorerfahrung in dieser Aufgabe bei den Bedingungen „blau2“ ($p < 0,001$) und „lila1“ ($p < 0,05$) zu Unterschieden geführt haben, wobei 29 Personen bereits an ähnlichen Tests teilgenommen haben (38 dagegen noch nie).

4.4 Auswertung der Aufgabe Depth

Die letzte Aufgabe weist unter allen drei Aufgaben die größte Streuung auf; teilweise ist sie doppelt so groß wie bei der Surface-Aufgabe. Das bedeutet, dass die Einschätzung nach Lage einer Position in der Tiefe nicht für jede teilnehmende Person gleich schwierig war sondern es dort größere Unterschiede gibt, wobei die Mittelwerte wieder höher sind als in der vorherigen Aufgabe und sich knapp über 70% einreihen. Dabei sticht die Rendering-Bedingung „blau2“ hervor, welche 73% erreicht. Man kann jedoch auch bei dieser Aufgabe von keinen signifikanten Unterschieden sprechen ($p = 0,38$), wie an Tabelle 4.3 abzulesen ist.

Wie in den Aufgaben zuvor liegen hier wieder die Signifikanzwerte des Lichtaufbaus näher an der Grenze zur schwachen Signifikanz als das für das Cool-Warm Shading der Fall ist, jedoch mit $p = 0,3$ auch weit davon entfernt. Dafür ergeben die Tests auf geratene Ergebnisse und Normalverteilung deutliche Werte, um gesichert von nicht geratenen und normalverteilten Daten ausgehen zu können. Mit dieser Erkenntnis einer guten Verteilung nicht geratener Ergebnisse erhärtet sich der Gedanke, dass die gewählten Rendering-Bedingungen bei den dicht beieinander liegenden Mittelwerten und der Streuung zu keinem entscheidenden Unterschied in der Erkennungsleistung bei einer Aufgabe dieser Form führen können.

Tabelle 4.3: Versuchsergebnisse zur Aufgabe Depth

Depth	Cool-Warm Shading * Lichtaufbau				Cool-Warm Sh.		Lichtaufbau	
	blau1	blau2	lila1	lila2	blau	lila	1	2
Mittelwert	0,703	0,733	0,708	0,712	0,718	0,71	0,705	0,722
Standardfehler	0,016	0,018	0,021	0,019	0,014	0,016	0,014	0,015
Signifikanz: gleiche Population	0,384				0,655		0,299	
Signifikanz: geratene Ergebnisse	< 0,001	< 0,001	< 0,001	< 0,001				
Signifikanz: keine Normalverteilung	< 0,001	< 0,001	< 0,001	< 0,001				

Wie bei der Curvature-Aufgabe gab es nun auch bei „blau2“ ($p < 0,01$) und „lila1“ ($p < 0,01$) eine Signifikanz für den Vorerfahrungseinfluss, bei den verbliebenen anderen beiden Rendering-Bedingungen aber nicht, sodass auch daraus keine Schlussfolgerung gezogen werden kann. Die persönliche Präferenz konnte nicht von allen Personen erhoben werden wie noch bei der Versuchskonstruktion geplant und ist somit lückenhaft und folglich nicht Gegenstand der Auswertung der drei Aufgaben gewesen.

4.5 Diskussion

Wie in jeder einzelnen Aufgabe festgestellt, kommt es zu keinen signifikanten Unterschieden in der Performance zwischen den Rendering-Bedingungen. Es gibt weder einen Effekt aus der Interaktion beider Faktoren noch für einen der beiden Faktoren allein. Selbst bei der ersten Aufgabe, die schon recht gut konstruiert zu sein scheint um gute Daten zu liefern (da normalverteilt und mit kleinem Standardfehler), sind die Signifikanzwerte zwar die besten aller drei Aufgaben, aber dennoch immer noch nicht nah an dem Level, welches zumindest schwache Signifikanz ausdrücken würde ($p < 0,05$). Immerhin liegen sie in allen drei Aufgaben für den Lichtaufbau dichter an diesem Wert als beim Cool-Warm Shading, so dass tendenziell eher der Lichtaufbau hier Einfluss ausüben dürfte. Insgesamt aber weisen die Rendering-Bedingungen in den gewählten Aufgaben zu wenig Unterschiede zueinander auf.

Es ist jedoch sehr stark davon auszugehen, dass Unterschiede in der Performance zwischen den zu Beginn betrachteten verschiedenen Ausprägungen der unabhängigen Variablen existieren, schließlich wurden im Vorversuch bestimmte Rendering-Bedingungen gegenüber anderen aufgrund besserer Sichtbarkeit deutlich bevorzugt. Die Sichtbarkeit ist Voraussetzung zur Beantwortung der Aufgaben, sodass bei sich stärker unterscheidenden Ausprägungen als den in dem Versuch gewählten auch eher messbare Unterschiede in der Erkennungsleistung bei den Aufgaben ergeben könnten.

Folglich lägen die Rendering-Bedingungen, welche ich für den Versuch gewählt habe, zu dicht beieinander, das heißt sie sind sich zu ähnlich. Auf der anderen

Seite war es aber gerade das Ziel, aus den guten Parameterausprägungen (die eben dichter zusammen liegen) die optimalen herauszufinden, und nicht zu zeigen, dass schlechte Renderings aus dem Vortest tatsächlich ungeeignet sind. Dies wäre auch nicht sinnvoll gewesen, denn damit hätte man die Fragestellung der Arbeit von vorn herein nicht beantworten können.

In der vorliegenden Art und Weise war es nun aber ebenfalls nicht gelungen, die Fragestellung zu beantworten, was an der Testkonstruktion gelegen haben kann. Es ist jedoch im Nachhinein nicht möglich festzustellen, ob dies der Fall war oder einfach daran lag, dass diese vier gewählten Rendering-Bedingungen tatsächlich überhaupt keine Unterschiede in der Erkennungsleistung verursachen. So verbleibt nur die Erkenntnis aus dem Vorversuch, was gute Parameterausprägungen für die medizinische Volumenvisualisierung sind, aber nicht, welches davon die optimalen. Das mag nun eventuell auch zu der Erkenntnis führen, dass dieser Nachweis im Rahmen einer Bachelorarbeit nicht möglich war, da er zu viele zeitliche und personelle Ressourcen in einer systematischen Folge von mehreren Versuchen in Anspruch nehmen würde – abgesehen von den Fragen der objektiven Metriken in den Aufgaben, die zuvor noch erörtert werden müssten.

Läge das Nicht-Erreichen signifikanter Unterschiede an der Testkonstruktion, so könnten verschiedenartige Probleme dazu geführt haben. Eventuell hat die viermalige Benutzung eines Ausschnitts in den ersten beiden Aufgaben des Versuchs auch Erinnerungseffekte ausgelöst, die bei der Depth-Aufgabe ausgeschlossen wurden, wobei auch dort eine große Streuung auftrat. Die letzte Aufgabe unterschied sich insgesamt stärker im Aufbau von den anderen beiden als diese sich untereinander.

Bei den ersten beiden könnte es auch zu schwierig gewesen sein, eine korrekte Antwort zu geben, weil sich die Ausschnitte bezüglich der Aufgabe zu sehr ähnelten (fast gleiche Rauheit respektive Krümmung). Dies wiederum wäre ein Problem bei der Kategorisierung durch den Autor gewesen, sodass eventuell sogar nicht nur zu ähnliche Ausschnitte in zwei Kategorien aufgeteilt wurden sondern vielleicht sogar „falsch“ – vorausgesetzt, man könnte dies anhand objektiver Kriterien belegen. Solche Metriken könnten etwa Amplitude und Frequenz vom Signal oder Rauschen für die Surface-Aufgabe sein; für die Krümmungsaufgabe käme eventuell das Bogenmaß in Frage, wofür aber wohl zur praktikablen und zeitnahen Ermittlung zusätzliche Funktionen direkt in die Volumenvisualisierungs-Software eingebaut werden müssten. Außerdem sollte man sich bei den Renderings Gedanken darüber machen, wie stark man in jeden Ausschnitt hineinzoomt, wobei manche Ausschnitte nur schwer aus bestimmten Entfernungen zu erzeugen sind.

Bei der Depth-Aufgabe wurde jedoch ein objektives Maß angelegt, da die Entfernung der markierten Positionen zum Betrachter in der Software gemessen werden konnte. Dies spräche in Verbindung mit der größten gemessenen Streuung für schlecht wahrnehmbare Tiefenverhältnisse, was an dem kleinen Bildausschnitt und dem Fehlen anderer Tiefenhinweisreize in den Style Transfer Functions gelegen haben kann, sodass zwischen den mit Style Transfer Functions steuerbaren Parameterausprägungen kaum ein Einfluss bezüglich des Tiefenempfindens möglich sein könnte.

Zuletzt möchte ich noch das mögliche Problem der wohl zu ähnlich gewählten Rendering-Bedingungen aufgreifen. Zwar weisen diese rein formal größere Distanzen als die anderen Bedingungen auf, die sich im Vorversuch auf den vorderen Rängen platziert haben, aber optisch scheint es, dass Licht und Reflexion sich zu einem Intensitätsempfinden multiplizieren, wie aus der Literatur bereits bekannt ist. In welchem Umfang dies geschieht, war dort aber nicht vermerkt; hier sieht es nun so aus, dass beide ungefähr in gleich großem Maße zu dem Produkt beitragen, denn die volle Lichtstärke bei halb starker Glossiness führt zu einer recht ähnlichen Beleuchtungssituation wie die halbe Lichtstärke bei voller Glossiness – wie an den Rendering-Bedingungen des Versuchs recht gut sichtbar wird („blau1“ gegenüber „blau2“ und „lila1“ gegenüber „lila2“), wenn man diese auf Seite 23 noch einmal betrachtet.

5 Zusammenfassung und Ausblick

5.1 Zusammenfassung

In dieser Arbeit wollte ich anhand der Beschaffenheit von Beleuchtungs- und Materialparametern ermitteln, wie eine medizinische Volumenvisualisierung in Bezug auf die optimale Wahrnehmung durch den Menschen auszusehen hat. Dafür standen Herzdaten zur Verfügung, die auf verschiedene Arten visualisiert werden konnten. Ich musste aber zunächst feststellen, was die perzeptuellen Anforderungen an eine Visualisierung sind (vgl. Abschnitt 2.4) und inwieweit man die Parameter mit den Style Transfer Functions steuern kann (vgl. Abschnitt 2.2). Bei diesen handelt es sich um eine Technik, die das Erzeugen recht komplexer Beleuchtungssituationen mit einfachsten Mitteln ermöglicht.

Daraufhin konstruierte ich einen Versuch und führte ihn mit vier – aus einem zuvor online realisierten Vorversuch – als geeignet ermittelten Parameterkombinationen durch, um die beste dieser herauszustellen. Ich nahm an, dass dabei eine Beleuchtung mit einer vorderen gelben Lichtquelle, welche einen deutlichen Glanzpunkt auf der Oberfläche des Volumens erzeugt, zusammen mit einer hinteren blauen Lichtquelle zur Aufhellung der schattigen Randbereiche zu dem besten Ergebnis führen würde (vgl. Abschnitt 3.5).

Der Versuch hat jedoch keine signifikanten Unterschiede bei sehr ähnlichen Mittelwerten in Bezug auf diese Hypothese offenbart (vgl. Kapitel 4), womit ich weder eine Aussage darüber treffen kann, ob eine meiner gewählten Rendering-Bedingungen als optimal anzusehen ist, noch dass im Gegenzug eine der Rendering-Bedingungen nicht optimal sein kann, denn „ein nicht signifikantes Ergebnis darf nicht als Beleg für die Richtigkeit der Nullhypothese interpretiert werden.“ (vgl. [BD06, S. 27]) Der Versuch ist somit streng genommen nicht geeignet gewesen, die Frage zu beantworten, welches optimale Beleuchtungs- und Materialparameter sind, da man bei der gescheiterten Hypothese aus rein wissenschaftstheoretischer Sicht keine induktiven Schlussfolgerungen ziehen darf.

Die Auswertung des Fragebogens, der zusätzlich von jedem Versuchsteilnehmer ausgefüllt wurde, bestätigte die Ergebnisse aus dem identischen Fragebogen des Vorversuchs. Er liefert wie auch die Literatur Hinweise, dass die Verwendung eines Cool-Warm Shadings mit gelb-violetter Färbung (gelb mit blau oder lila) sowie sichtbare Glossiness, also die deutliche Reflexionseigenschaft des Oberflächenmaterials, der räumlichen Wahrnehmung zuträglich sind. Genauso führt die Verwendung eines primären Führungslichtes zur stärkeren Schattierung des Objekts, was durch die Rankings der Teilnehmer im Vorversuch gegenüber einer Aufhellung der

Schattierung durch ein zusätzliches, frontal positioniertes Licht präferiert wurde (vgl. Abschnitt 3.2.1). Letztendlich wurde hierbei aber nur auf eine gängige feste Positionierung dreier Lichtquellen zurückgegriffen.

5.2 Ausblick

Die genannte Einschränkung zeigt, wie schwierig es ist, aus der schier unendlich großen Zahl an Möglichkeiten der Parametergestaltung die essentiell wichtigen herauszufiltern. Bei einigen ist dies klar, bei anderen wiederum nicht. Im Rahmen einer Bachelorarbeit ist das Thema zu umfangreich, da die bisherige Basis an Erkenntnis nicht ausreicht, um darauf aufbauend die optimalen Parameter einer medizinischen Volumenvisualisierung zu finden. Gerade auch die Veränderung der Wahrnehmung durch die Abhängigkeit der Parameter untereinander erschwert dies zusätzlich. Für sich betrachtet gibt es bereits zahlreiche Untersuchungen, die einzelne basale Faktoren getestet und für diese entsprechende Empfehlungen ausgesprochen haben (vgl. Abschnitt 1.4 und 2.4). Darum kann man sich in einer Folge systematisch aufeinander aufbauender Versuche dem Thema sicherlich besser nähern, als es ein umfassender Versuch allein vermag.

Nun gibt es eine weitere Einschränkung des Versuchs durch die Style Transfer Functions, mit welchen sich einige Hinweisreize für die räumliche Wahrnehmung nicht umsetzen lassen, da von den Metriken nur der Gradient als Annäherung der Oberflächennormale als Information zur Verfügung steht. Daher kann z.B. kein Schattenwurf oder eine entfernungsabhängige Texturierung berechnet werden. Dennoch wäre eine Aussage über optimale Parameter allein für Style Transfer Functions schon wünschenswert, denn die eben genannten zusätzlichen Hinweisreize würden vermutlich nichts an der Eignung von Beleuchtungs- oder Materialparametern ändern. Auch ist bei einer anderen Einschränkung, nämlich dass in diesem Versuch nur unfarbige Objekte untersucht wurden, davon auszugehen, dass dies bei Berücksichtigung farbiger Objekte kaum andere Ergebnisse zu den geeigneten Beleuchtungsparametern offenbaren dürfte, was aber zu prüfen wäre.

Einen entscheidenden Einfluss stellt dagegen sicherlich die Versuchskonstruktion mit den Aufgaben dar. Eine grundlegende Überprüfung dieser ist ratsam, um objektiv messbare Metriken zur Einschätzung der zu verwendenden Bildausschnitte anlegen und die Distanz untereinander quantifizieren zu können. Die in diesem Versuch wie auch bei Steffen verwandte subjektive Einteilung erlaubt dies nicht, weder ist klar, ob die Kategorisierung korrekt ist, noch ist anhand der Kategorien ein Maß ableitbar, wie stark ein Merkmal konkret ausgeprägt ist (etwa Krümmung oder Rauheit). Die Depth-Aufgabe griff bereits auf ein objektives Maß zurück, da die Entfernung per Software gemessen werden konnte. Sie litt aber unter der Einschränkung fehlender Tiefenhinweisreize der Style Transfer Functions, bei denen das Cool-Warm Shading zur Unterscheidung zweier Positionen in der Tiefe offenbar nicht ausreicht, um bei allen Teilnehmern ähnliche Versuchsergebnisse zu erzielen, denn die Streuung war hierbei die größte der drei Aufgaben.

Anhang A

Abbildungen

A.1 Beleuchtungsmethoden im Volumenrendering

In folgender Abbildung sieht man in der ersten Reihe ein unfarbiges Volumen auf Basis der gemessenen Scanner-Daten (1), sodass die Helligkeiten kein Licht repräsentieren sondern eher unterschiedlichen Dichten entsprechen. Dank Berechnung des Gradienten zur Beleuchtung mit dem Phong Shading (2) wird ein 3D-Eindruck möglich. Die zweite Reihe zeigt ein bereits farbiges Volumen (3) mit rotem Herzgewebe, was einmal durch Phong Shading (4) beleuchtet wurde und einmal durch Verwendung von Style Transfer Functions mit einer farbigen Lit Sphere (5), wobei deren Farben mit den schon vorhandenen multipliziert wurden.

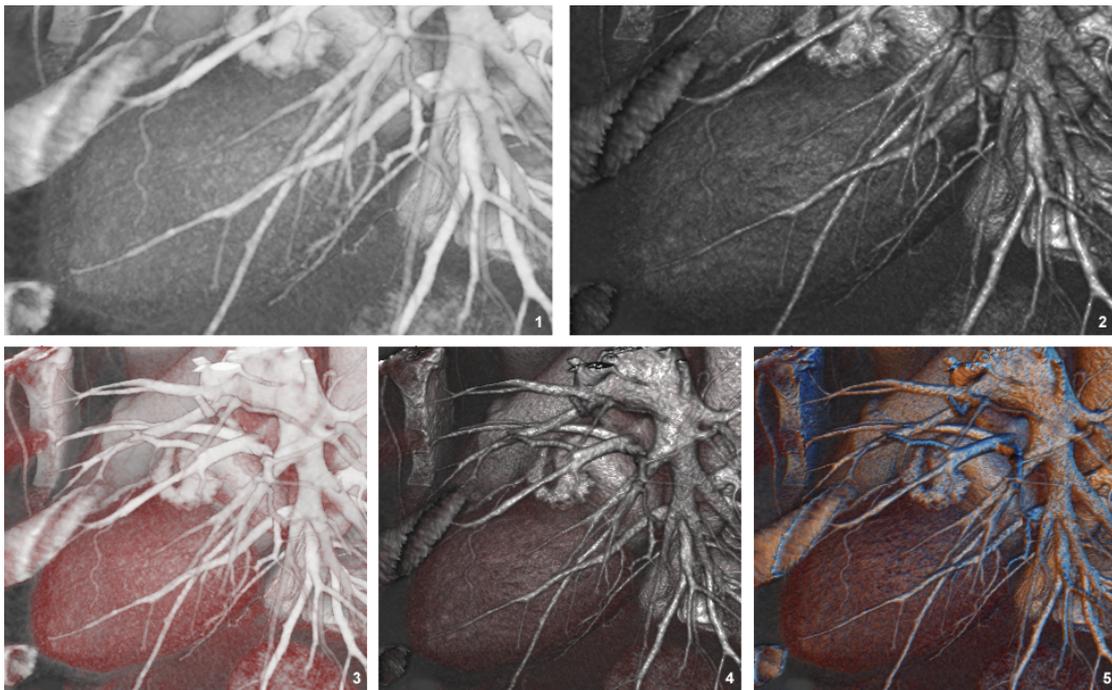


Abbildung A.1: Beleuchtung eines unfarbigen Volumens mittels Phong Shading sowie Vergleich zwischen einem Phong Shading und einer Style Transfer Function zur Beleuchtung eines farbigen Volumens.

A.2 Beleuchtungssituation für die Lit Spheres

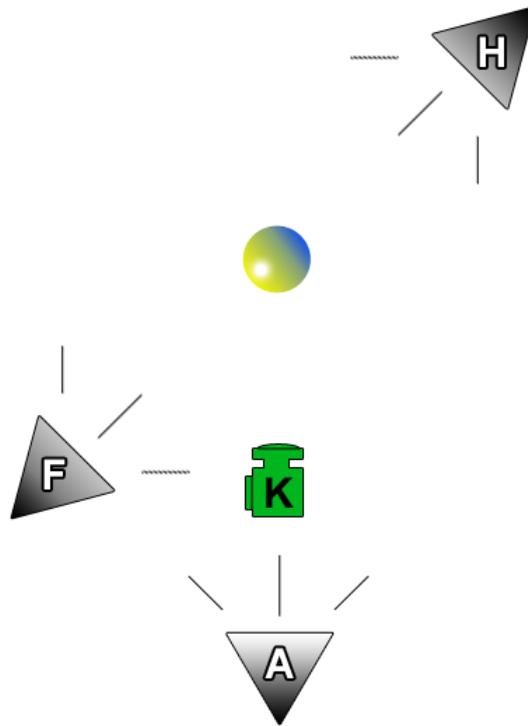


Abbildung A.2: Draufsicht auf die verwandte 3-Punkt-Beleuchtung: K = Kamera, F = Führungslicht, A = Aufhelllicht, H = Hintergrundlicht

A.3 Unabhängige Variablen

Die folgenden Lit Spheres sind vertikal in der Reihenfolge der auf dieser Seite aufgezählten unabhängigen Variablen und horizontal in der jeweiligen Reihenfolge der Ausprägungen abgebildet, wobei die jeweils nicht betrachteten Variablen konstant gehalten wurden. Für die Variable „Lichtauswahl“ (die untersten sieben Lit Spheres zu 6.) siehe zur Bedeutung der Buchstaben Abbildung A.2.

1. Cool-Warm Färbung = {gelb-blau, gelb-lila, orange-lila, rot-blau}
2. Cool-Warm Sättigung = {schwarz-weiß, halb, voll}
3. Cool-Warm Richtung = {cool-to-warm, warm-to-cool}
4. Glossiness = {aus, halb, voll}
5. Lichtstärke = {aus, halb, voll}
6. Lichtauswahl = $\{F, A, H\} \cup \{F+A, F+H, A+H, F+A+H\}$

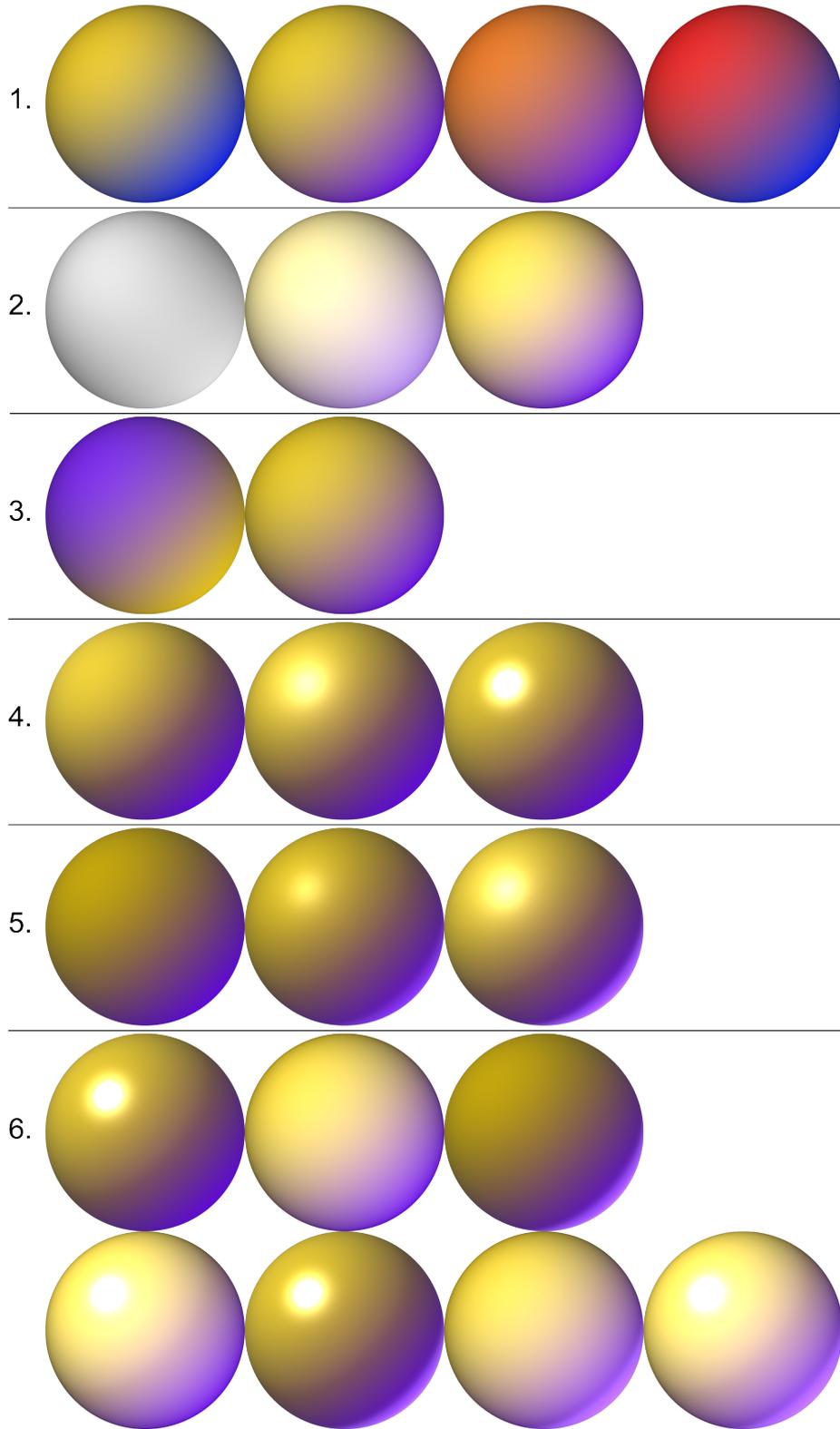
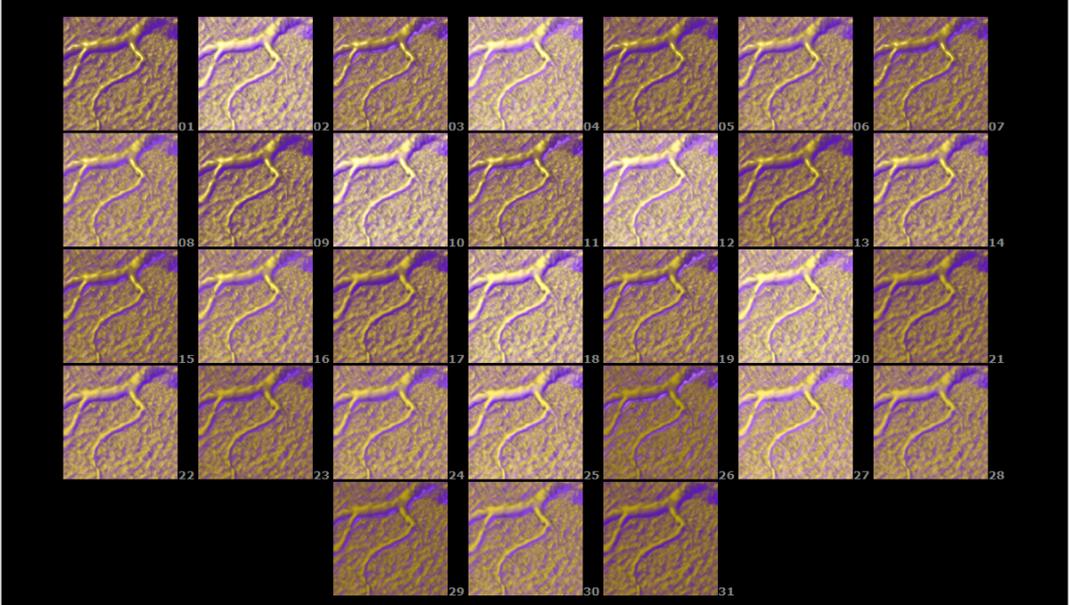


Abbildung A.3: Mögliche Ausprägungen der unabhängigen Variablen anhand einer Auswahl von für den Vorversuch verwandten Lit Spheres

A.4 Fragebogen für den Vorversuch



In welchen der oberen 31 Bilder sind die meisten Details erkennbar? 1. 2. 3. 4. 5.

Welche der oberen 31 Bilder haben die beste räumliche Wirkung? 1. 2. 3. 4. 5.

Fragebogen abschließen

Absenden

Abbildung A.4: Bilderset 3 des Fragebogens

Information

Vielen Dank für die Teilnahme an diesem Fragebogen!

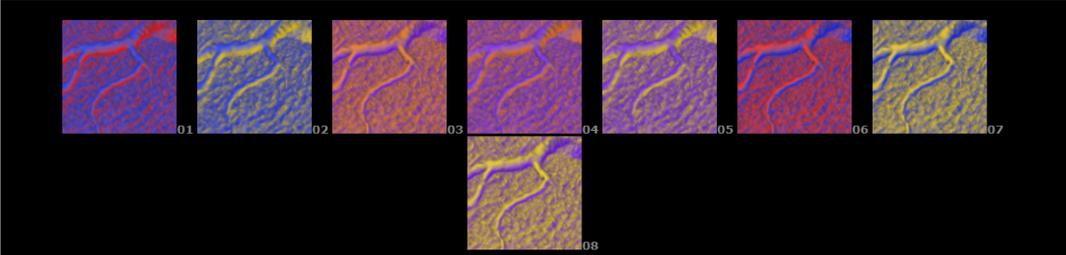
Im Rahmen meiner Bachelor-Arbeit versuche ich optimale Beleuchtungs- und Materialparameter für medizinische Volumenvisualisierungen zu ermitteln. Dieses wird durch Benutzertests geschehen, die mit statistischen Mitteln ausgewertet werden. Um die zahlreichen Konfigurationen an Beleuchtungssituationen auf eine handhabbare Zahl zu reduzieren, wird dieser "Vortest" durchgeführt, für den kein medizinisches Fachwissen benötigt wird.

Es werden keine persönlichen Informationen erhoben, aufgrund der simplen technischen Umsetzung dieses Fragebogens wird jedoch eine Email über Ihr Email-Programm an den Autor mit den Antworten versandt. Dafür gibt es am unteren Ende der Website einen Submit-Button.

Am besten führen Sie die Umfrage im Vollbildmodus (Taste F11 zum an-/ausschalten) durch, sodass sie alle Bilder eines Sets möglichst ohne Scrollen vollständig sehen können.

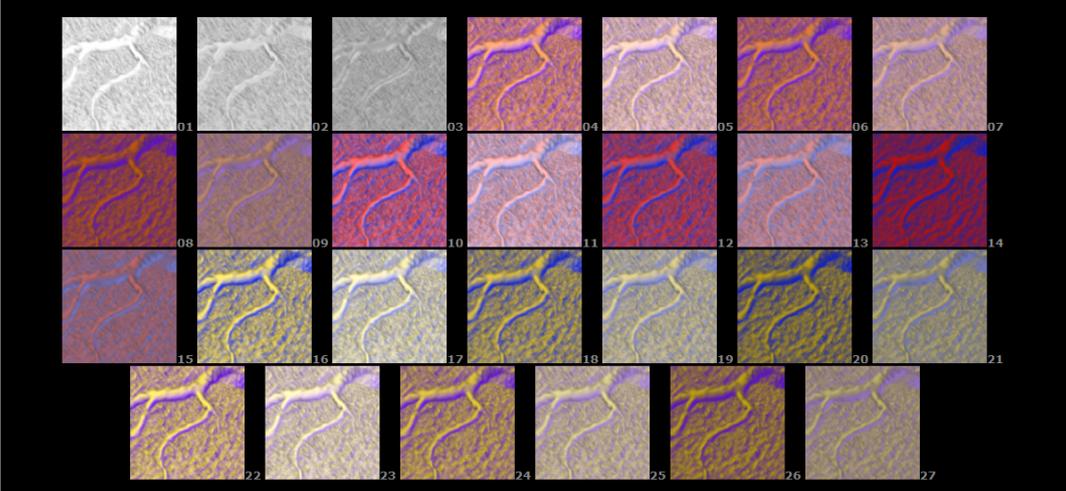
Versuchen Sie die Fragen immer losgelöst von Ihren bereits gegebenen Antworten bei den vorherigen Bilder-Sets zu beantworten. Die Beantwortung der Fragen erfolgt durch eine Rangfolge der Bilder, auf die die Fragestellung am meisten zutrifft. Bei der Frage "Welche der Bilder zeigen die meisten Details?" ist also das Bild, welches Sie hinter 1. angeben, dasjenige, welches unter den aufgeführten Bildern die meisten Details zeigt; dann folgt hinter 2. das Bild mit der zweit-besten Sichtbarkeit an Details usw. ...

Tipp: Auf den ersten Blick mögen einige Bilder gleich aussehen. Achten Sie z.B. für Fragen der Detailsichtbarkeit unter anderem auf die Details in den Schattenbereichen oder die Details der Glanz- und hellen Bereiche.



In welchen der oberen 8 Bilder sind die meisten Details erkennbar? 1. 2. 3. 4. 5.

Welche der oberen 8 Bilder haben die beste räumliche Wirkung? 1. 2. 3. 4. 5.



In welchen der oberen 24 Bilder sind die meisten Details erkennbar? 1. 2. 3. 4. 5.

Welche der oberen 24 Bilder haben die beste räumliche Wirkung? 1. 2. 3. 4. 5.

Abbildung A.5: Bildersetz 1 und 2 des Fragebogens

Anhang B

Tabellen

B.1 Rankings aus dem Vorversuch

Tabelle B.1: Ranking der besten Rendering-Bedingungen für Bilderset 1

Rang	Punktzahl	Cool-Warm Färbung	Cool-Warm-Richtung
1	175	Gelb-Blau	Warm-to-Cool
2	140	Gelb-Blau	Cool-to-Warm
3	120	Gelb-Lila	Warm-to-Cool
4	87	Gelb-Lila	Cool-to-Warm

Tabelle B.2: Ranking der besten Rendering-Bedingungen für Bilderset 2

Rang	Punktzahl	Cool-Warm Färbung	Lichtstärke	Sättigung
1	108	Gelb-Blau	Voll	Voll
2	98	Gelb-Blau	Halb	Voll
3	94	Gelb-Blau	Aus	Voll
4	75	Gelb-Lila	Voll	Voll
5	57	Gelb-Lila	Halb	Voll
6	51	Orange-Lila	Aus	Voll

Tabelle B.3: Ranking der besten Rendering-Bedingungen für Bilderset 3
(zur Bedeutung der Buchstaben der Lichter siehe Seite 44)

Rang	Punktzahl	Glossiness	Lichtstärke	Lichter
1	71	Voll	Halb	F
2	66	Halb	Voll	F
3	65	Voll	Voll	F
4	56	Voll	Voll	F + H
5	48	Voll	Halb	F + H
6	47	Halb	Voll	F + H

Literatur

- [AL11] ANSORGE, U.; LEDER, H.: *Wahrnehmung und Aufmerksamkeit*. Springer, 2011
- [BD06] BORTZ, J.; DÖRING, N.: *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler*. 4., überarb. Aufl. Springer, 2006
- [BG07] BRUCKNER, S.; GRÖLLER, M.E.: Style transfer functions for illustrative volume rendering. In: *Computer Graphics Forum* Bd. 26 Wiley Online Library, 2007, S. 715–724
- [Boy03] BOYCE, P.R.: *Human factors in lighting*. 2nd Edition. Taylor & Francis, 2003
- [CF07] CANIARD, F.; FLEMING, R.W.: Distortion in 3D shape estimation with changes in illumination. In: *Proceedings of the 4th symposium on Applied perception in graphics and visualization* ACM, 2007, S. 99–105
- [EHK⁺06] ENGEL, K.; HADWIGER, M.; KNISS, J.M.; REZK-SALAMA, C.; WEISKOPF, D.: *Real-time volume graphics*. AK Peters, Ltd., 2006
- [FTA04] FLEMING, R.W.; TORRALBA, A.; ADELSON, E.H.: Specular reflections and the perception of shape. In: *Journal of Vision* 4 (2004), Nr. 9
- [GGSC98] GOOCH, A.; GOOCH, B.; SHIRLEY, P.; COHEN, E.: A non-photorealistic lighting model for automatic technical illustration. In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* ACM, 1998, S. 447–452
- [Hor74] HORN, B.K.P.: Determining lightness from an image. In: *Computer Graphics and Image Processing* 3 (1974), Nr. 4, S. 277–299
- [KHSI04] KIM, S.; HAGH-SHENAS, H.; INTERRANTE, V.: Conveying shape with texture: experimental investigations of texture’s effects on shape categorization judgments. In: *IEEE Transactions on Visualization and Computer Graphics* 10 (2004), Nr. 4, S. 471–483
- [KK91] KNILL, D.C.; KERSTEN, D.: Apparent surface curvature affects lightness perception. In: *Nature* 351 (1991), Nr. 6323, S. 228–230

- [Lei] *Leitfaden für kontrollierte Experimente.* – online unter: <http://www.cs.uni-paderborn.de/fachgebiete/ag-domik/hard-software/onlinehilfe-der-ag-domik/kontrollierteexperimente.html>
- [LOPR97] LEHMANN, T.; OBERSCHELP, W.; PELIKAN, E.; REPGES, R.: *Bildverarbeitung für die Medizin.* Springer, 1997
- [Mül08] MÜLLER, J.: *Handbuch der Lichttechnik - Das Kompendium für den Praktiker.* 4., aktual. und erw. Aufl. PPVMedien, 2008
- [Ram88] RAMACHANDRAN, V.S.: Perception of shape from shading. In: *Nature* 331 (1988)
- [RDRS10] ROPINSKI, T.; DÖRING, C.; REZK-SALAMA, C.: Advanced Volume Illumination with Unconstrained Light Source Positioning. In: *IEEE Computer Graphics and Applications, IEEE* (2010), S. 29–41
- [RSH06] ROPINSKI, T.; STEINICKE, F.; HINRICHS, K.: Visually supporting depth perception in angiography imaging. In: *Smart Graphics* Springer, 2006, S. 93–104
- [SM99] SCHUMANN, H.; MÜLLER, W.: *Visualisierung: Grundlagen und allgemeine Methoden.* Springer, 1999
- [Spa06] SPADA, H.: *Lehrbuch Allgemeine Psychologie.* 3., vollst. überarb. und erw. Aufl. Huber, 2006
- [Ste10] STEFFEN, F.: *Applications and Advantages of Style Transfer Functions for Heart Data Visualization,* Universität Paderborn, Master-Arbeit, 2010
- [TM83] TODD, J.T.; MINGOLLA, E.: Perception of surface curvature and direction of illumination from patterns of shading. In: *Journal of Experimental Psychology: Human Perception and Performance* 9 (1983), Nr. 4, S. 583
- [War04] WARE, C.: *Information visualization: perception for design.* Morgan Kaufmann, 2004
- [WB08] WEIGLE, C.; BANKS, D.: A comparison of the perceptual benefits of linear perspective and physically-based illumination for display of dense 3D streamtubes. In: *IEEE Transactions on Visualization and Computer Graphics* (2008), S. 1723–1730
- [WFG92] WANGER, L.C.; FERWERDA, J.A.; GREENBERG, D.P.: Perceiving spatial relationships in computer-generated images. In: *IEEE Computer Graphics and Applications* (1992), S. 44–51