

# Photometrisch konsistente Radiositysimulation und Bildwiedergabe in Virtual und Augmented Reality Anwendungen



Vom Fachbereich Informatik  
der Technischen Universität Darmstadt  
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs (Dr.-Ing.)

von

Dipl.-Inform. **Wolfram Kresse**

Referenten der Arbeit:

Prof. Dr.-Ing. José L. Encarnação

Prof. Dr. Hans-Peter Seidel

Tag der Einreichung:

15. September 2003

Tag der mündlichen Prüfung:

27. Oktober 2003



# Danksagung

Die vorliegende Arbeit entstand während meiner Zeit als wissenschaftlicher Mitarbeiter im Fraunhofer-Institut für graphische Datenverarbeitung.

Mein spezieller Dank gilt Professor Dr.-Ing. Encarnaçao für die Förderung des Themas, sowie die Betreuung meiner Arbeit während der vergangenen Jahre. Herrn Professor Seidel danke ich sehr für die Übernahme des Korreferats.

Ein super-besonderer Dank geht an Professor Stefan Müller, der mich schon in meinem Studium an die Thematik Lichtsimulation, der ich bis heute treu geblieben bin, heranführte. Sein Enthusiasmus und sein Engagement nicht nur im Bereich seines damaligen Lieblingsthemas „Licht“ sind einfach unvergleichlich. Ohne ihn wäre diese Arbeit so nicht möglich gewesen.

Besonders bedanken möchte ich mich auch bei Didier Stricker und Christian Knöpfle, für die Unterstützung und Genehmigung von Freiräumen, die für die Erstellung dieser Arbeit sehr wichtig waren.

Vielen Dank insbesondere auch meinem Bürokollegen Dirk Reiners für Diskussionen zur Thematik, gute Ideen und gemeinsam durchhackte Nächte...

Ebenso ein dickes Dankeschön an meine Studenten Robert Rapljenovic und Markus Seeger, die immer mehr als vollen Einsatz leisteten, wenn mal wieder eine Deadline anstand, sowie Christoph Groth und Gundolf von Bachhaus, die in kürzester Zeit unheimlich viel geschafft haben. Danke an Frank Schöffel für lange Zusammenarbeit im Thema Licht, und Thorsten Grosch, für seine tatkräftige Unterstützung als Hiwi.

Ein großer Sonder-Dank latürnich allen A4-Kollegen, -Ex-Kollegen, -Studis, -Hiwis, usw. Ohne das in dieser Abteilung bestehende außerordentliche Arbeitsklima würde das Arbeiten trotz fesselnder Thematik nur noch vermindert Spaß machen. Vielen Dank auch an Andreas Fröhlich für das detaillierte Korrekturlesen dieser Arbeit.

Ich bedanke mich beim Universum, dafür, daß es da ist.

Wolfram Kresse

Mörfelden, im Oktober 2003



<b>1</b>	<b>Einleitung</b>	<b>9</b>
1.1	Motivation	9
1.2	Überblick der Arbeit	10
1.3	Zusammenfassung der wichtigsten Ergebnisse	11
<b>2</b>	<b>Ein effizientes Radiosity-System</b>	<b>13</b>
2.1	Benutzergruppen und -anforderungen	13
2.1.1	Benutzergruppen	13
2.1.2	Anforderungen	15
2.1.3	Szenentypen	18
2.2	Mängel existierender Radiositysysteme	19
2.2.1	Optische Mängel	20
2.2.2	Physikalische Genauigkeit	23
2.2.3	Physikalische Wiedergabequalität	24
2.2.4	Unnötige Unterteilungen	25
2.2.5	Ungenauere Sichtbarkeitsberechnung	25
2.2.6	Szenenvorverarbeitung	26
2.2.7	Bedienbarkeit	30
2.2.8	Zusammenfassung	32
2.3	Photometrische und colorimetrische Konsistenz	34
2.3.1	Physiologische Grundlagen	34
2.3.2	Interne Konsistenz	36
2.3.3	Konsistenz in den Eingabedaten	40
2.3.4	Konsistenz in den Ausgabedaten	41
2.4	Kunst- und Tageslicht	42
2.4.1	Flächenlichtquellen	43
2.4.2	Punktförmige Lichtquellen	43
2.4.3	Goniometrische Lichtquellen	45
2.4.4	Diffuses Tageslicht	45
2.5	Fehlerschranken	46
2.5.1	Evolution des Refinementkriteriums	47
2.5.2	Fehlerarten	48
2.5.3	Basisfunktionen	50
2.5.4	Implikationen durch einen linearen Empfänger	53
2.6	Refinement	57
2.6.1	Energiebasierte Tools	60

2.6.2	Geometrische Tools	66
2.6.3	Sichtbarkeitstools	69
2.6.4	Refinementkriterium	69
2.7	Visibilitätsbetrachtung	71
2.7.1	Zweistufigkeit	72
2.7.2	Sichtbarkeitsklassifikation	73
2.7.3	Ermittlung des sichtbaren Anteils	76
<b>3</b>	<b>Techniken zur beleuchtungstechnisch konsistenten Wiedergabe</b>	<b>79</b>
3.1	Lichttechnische Eigenschaften von immersiven Displaysystemen	81
3.1.1	Physikalische Eigenschaften	81
3.1.2	Stereotechnologien	83
3.1.3	Projektorttechnologien	88
3.1.4	Leinwandmaterialien	92
3.1.5	Exemplarische Setups	95
3.2	Mängel existierender Displaysysteme	102
3.2.1	Testumgebung	102
3.2.2	Leuchtdichtevertelung	104
3.2.3	Leuchtdichtebereich und Farbraum	109
3.2.4	Interreflexion	112
3.2.5	Projektorspezifische Problematiken	117
3.3	Displaykalibrierung	122
3.3.1	Erforderliche Messungen und Kalibrierungsschritte	125
3.3.2	Bestimmung des gemeinsamen Gamuts	127
3.3.3	Ermittlung der maximalen Leuchtdichten	129
3.3.4	Kompensation nichtlinearen Verhaltens	129
3.3.5	Ersetzen der Gammakurve	130
3.3.6	Die Farbe „Schwarz“	131
3.3.7	Konsistente Eingangsdaten	133
3.3.8	Beispielszenarien Digitale CAVE und HEyeWall	134
3.3.9	Ergebnisse	138
3.4	Automatische Kalibrierung von Displaysystemen	144
3.4.1	Vorgehensweise	145
3.4.2	Kamerakalibrierung	145
3.4.3	Displaykalibrierung	152
3.4.4	Konsistente Darstellung	160

3.4.5	Ergebnisse .....	163
<b>4</b>	<b>Integration einer AR-Bildsynthesekomponente .....</b>	<b>169</b>
4.1	Mängel existierender AR-Systeme .....	169
4.2	Rekonstruktion der Licht- und Materialeigenschaften .....	171
4.2.1	Reflexionskoeffizient .....	171
4.2.2	Lichtposition .....	172
4.2.3	Abstrahlcharakteristik .....	173
4.2.4	Verwendung einer Lightprobe .....	180
4.3	Konsistente Erweiterung von Fotos .....	182
4.3.1	Datenstrukturen .....	183
4.3.2	Vorgehensweise .....	184
4.3.3	Direktes Licht .....	187
4.3.4	Sichtbarkeitsbestimmung .....	188
4.3.5	Bewegung virtueller Objekte .....	189
4.3.6	Aktualisierung indirekten Lichts .....	190
4.3.7	Darstellung .....	192
<b>5</b>	<b>Konkrete Anwendungsbeispiele .....</b>	<b>195</b>
5.1	Lichtsimulation im Dom von Siena .....	195
5.2	Nachtsituation im Fahrzeuginnenraum .....	200
5.3	Konsistente Erweiterung von Fotos von Wohnräumen .....	202
5.4	HEyeWall .....	204
5.5	Weitere Beispiele .....	206
<b>6</b>	<b>Zusammenfassung und Ausblick .....</b>	<b>211</b>
6.1	Zusammenfassung .....	211
6.1.1	Das Radiositysystem .....	211
6.1.2	Die Bildwiedergabe .....	212
6.1.3	Die Verbindung mit realen Szenen .....	215
6.1.4	Das Gesamtkonzept .....	215
6.2	Ausblick .....	216
<b>7</b>	<b>Literaturverzeichnis .....</b>	<b>219</b>
7.1	Eigene Veröffentlichungen .....	230





# 1 Einleitung

## 1.1 Motivation

Im Laufe der letzten Jahre hat Echtzeitgrafik durch günstige Grafikkarten mit der Rechenleistung von Grafik-Workstations Einzug in viele Haushalte gehalten, sodass die prinzipiellen Grundlagen zur 3D-Visualisierung beliebiger Szenarien jedem zur Verfügung stehen.

Nachdem von der Softwareseite her Verfahren zur Darstellung von *Virtuellen Realitäten* heute durchaus etabliert sind – man denke nur an immer ausgefeiltere Computerspiele, oder den immer stärker wachsenden Anteil an Special Effects in Kinofilmen bis hin zu vollständig computergenerierten Filmen – beginnen nun mehr und mehr Leute darüber nachzudenken, was man mit der zur Verfügung stehenden Technik denn nun eigentlich alles anfangen kann, und insbesondere wie die Darstellungsqualität weiter verbessert werden kann.

Neue Projektionstechnologien und Stichworte wie LCD, DLP und zirkuläre Polarisation drängen auf den Markt. Die Anwendbarkeit immersiver Projektionssysteme für konsistente, quantitative Aussagen und Entscheidungen, sei es im Design-, Konstruktions- oder reinen Visualisierungsbereich, war jedoch bis jetzt unklar.

Zum anderen entwickelt sich das Gebiet der Augmented Reality mehr und mehr, und die Frage nach Möglichkeiten zur Erhöhung des Realismus wird aufgeworfen.

Zusammenfassend läßt sich sagen, daß die Verfahren zur globalen Beleuchtungssimulation durch eine gewisse *Awareness* allmählich in den breiten Markt vordringen. Auf der anderen Seite haben sich für den industriellen, täglichen Einsatz der Virtuellen Realitäten verschiedene immersive Projektionstechnologien entwickelt. Und zum Dritten ist man auf dem Gebiet der *Augmented Reality* immer mehr danach bestrebt, reale Szenen nicht nur überhaupt mit virtuellen Objekten zu erweitern, sondern dies auch in einer Form zu tun, daß sich diese Objekte nahtlos in die reale Szene einfügen.

Der Bedarf an konsistenter Simulation und Darstellung ist vorhanden und wächst. Für den produktiven Einsatz von computergraphischen Verfahren z.B. im Bereich Design-Review ist jedoch eine hohe visuelle Qualität allein schließlich nicht mehr ausreichend, und daher eine durchgehend quantitativ verlässliche Pipeline von Ausgangsdaten über Simulation bis hin zur Darstellung unabdingbar.

Eine relevante Beispielanwendung hierfür ist z.B. die Evaluierung von Reflexionen auf rückbeleuchteten Instrumenten im Fahrzeuginnenraum, oder die Spiegelung ebendieser Instrumente in Front- und Seitenscheibe, ob sie zu hell sind, den Fahrer stören, ablenken, oder die Sichtbarkeit z.B. auf den Außenspiegel dadurch eingeschränkt wird.

Um Technologien wie Computergraphik, Lichtsimulation und immersive Projektionen im Designprozeß einsetzen zu können, müssen die präsentierten Bilder verlässlich die „reale Welt“ widerspiegeln. Es muß garantiert werden können, daß das was zu sehen ist die exakte Helligkeit und Farbe besitzt, die das entsprechende reale Äquivalent auszeichnet.

Durch genaue Analyse der verschiedenen Aspekte, die eine solche Darstellung beeinflussen, wird mit der vorliegenden Arbeit ein photometrisch konsistentes Framework geschaffen, mit dem der nutzbringende Einsatz globaler Simulationssysteme im produktiven Betrieb ermöglicht werden kann.

Diese einzelnen Aspekte setzen sich zusammen aus dem Lichtsimulationssystem an sich, den Eigenschaften von Projektionsdisplays, sowie der Vereinigung von realen und virtuellen Welten in der Augmented Reality.

## 1.2 Überblick der Arbeit

Im Folgenden werden die erforderlichen Einzelkomponenten für ein Framework besprochen, das die Methoden zur Beleuchtungssimulation mit der computertechnischen Erweiterung von Fotografien kombiniert und dabei eine vollständige Konsistenz-Pipeline von der Eingabe über die Berechnung bis hin zur Darstellung sicherstellt.

Aufbauend auf einer photometrisch und colorimetrisch konsistenten Berechnungsbasis wird in **Kapitel 2** ein Radiosity-System mit logisch sinnvollen Patchunterteilungskriterien, sowie einer effizienten und genauen Sichtbarkeitsbetrachtung beschrieben.

Um eine photometrisch konsistent berechnete Szene korrekt darstellen zu können, sind eine Analyse der möglichen Darstellungstechniken, sowie Methoden zur Beseitigung ihrer Unzulänglichkeiten erforderlich, was in **Kapitel 3** durchgeführt wird. Eine Displaykalibrierung sowohl mehrerer Displays aufeinander, als auch der konsistenten Kalibrierung insgesamt wird hier erarbeitet. Durch ein automatisiertes Kalibrierungsverfahren wird die einfache Ermittlung der für konsistente Bildwiedergabe relevanten Parameter sichergestellt.

In **Kapitel 4** wird das Konzept eines photometrisch kalibrierten Systems auf Augmented Reality Anwendungen ausgedehnt. Die Voraussetzungen für kalib-

rierte Erweiterung realer Szenen werden klargestellt und die Verbindung zwischen realen Fotos und dem vorgestellten Radiositysystem geschaffen, wodurch konsistente Darstellung auch für Realszenen oder mit virtuellen Objekten erweiterte Szenen möglich wird.

**Kapitel 5** beschäftigt sich mit konkreten Anwendungsbeispielen des vorgestellten Systems aus dem computergraphischen Umfeld. Nach einem Beispiel für eine Lichtsimulation in einer komplexen Architekturszene wird die Anwendung des beschriebenen Verfahrens auf die photometrisch konsistente Repräsentation der Beleuchtungssimulation eines Fahrzeuginnenraums in einer Nachtsituation vorgestellt. Im weiteren dient die farblich korrekte Erweiterung von Wohnraumaufnahmen als nutzbringendes Beispiel der Kombination von Radiositysimulation und AR-Bildsynthese in einem photometrisch konsistenten Framework.

Abschließend dient **Kapitel 6** als globale Zusammenfassung der in dieser Arbeit vorgestellten Verfahren und Ergebnisse, sowie als Ausgangspunkt für weiterführende Forschungen.

## 1.3 Zusammenfassung der wichtigsten Ergebnisse

In der vorliegenden Arbeit wird ein Framework zur photometrischen und colorimetrischen Konsistenz entwickelt, das kalibrierte Simulation und Bildwiedergabe derart ermöglicht, daß die Korrektheit der gesamten Prozeßkette Eingabedaten—Simulation—Ausgabedaten—Darstellung quantitativ sichergestellt und durch direkte Messungen an der Leinwand überprüft werden kann.

Durch die Sicherstellung photometrischer und colorimetrischer Konsistenz an jeder Stelle der Simulationspipeline und insbesondere bei immersiven Projektionsdisplays als Ausgabemedium wird so ein globales Konsistenz-Framework geschaffen, das es dem Benutzer ermöglicht, vollsynthetische oder durch virtuelle Objekte erweiterte, reale Szenarien mit ihren tatsächlichen Leuchtdichten und Farben zu simulieren und quantitativ verläßlich darzustellen.

Das Radiositysystem wird entsprechend der relevanten Gesichtspunkte erweitert und auf die Anforderungen unterschiedlicher Benutzergruppen angepaßt, wodurch die Effizienz steigt. Die Simulationsqualität wird verbessert, und die quantitative Aussagekraft der Simulation garantiert.

Typische Artefakte und Problematiken existierender Radiositysysteme werden diskutiert und Methoden zu deren Beseitigung vorgestellt. Tageslicht und künstliche Lichtquellen werden nahtlos in das konsistente, hierarchische Radiositysystem integriert.

Durch Algorithmen zur Berechnung verlässlicher Formfaktorschranken für alle Lichtquellentypen wird effizientes Refinement und letztendlich ein anwendbares Radiositysystem realisiert.

Um korrekt simulierte Daten entsprechend wiederzugeben, werden die für konsistente Darstellung relevanten Parameter ermittelt, sowie die Nachteile untersucht, die eine photometrisch konsistente Darstellung auf verschiedensten Projektionssystemen einschränken oder verhindern.

Unterschiedliche Ausgabetechnologien werden entsprechend dieser Parameter analysiert und bewertet. Es werden Algorithmen entwickelt, diese Displays zu kalibrieren, um so ihre photometrische und colorimetrische Konsistenz zu gewährleisten.

Insbesondere wird eine Methode entwickelt und vorgestellt, durch ein automatisches Kalibrierungsverfahren mit Hilfe einer digitalen Kamera eine gleichförmige Ausleuchtung von Projektionsflächen vorzunehmen, sowie die gleichzeitige Vermessung mehrerer Ausgabedispays zu ermöglichen.

Neue Methoden zur Rekonstruktion von Lichtquelleninformationen aus Realbildern werden entwickelt, existierende Verfahren zur Erweiterung von Fotos mit dem Radiosityverfahren beschleunigt.

Durch Verwendung von *high dynamic range images* können photometrische Informationen aus Realbildern gewonnen werden, was die Anwendung des Ansatzes zur photometrischen und colorimetrischen Konsistenz auch für Augmented Reality Szenarien ermöglicht.

## 2 Ein effizientes Radiosity-System

Die Grundlage für heutige Finite-Element-Radiositysysteme bildet, abgesehen von den frühen Erkenntnissen aus der Wärmestrahlungsphysik, die erste algorithmische Beschreibung von 1984 [GTGB84].

Nach Experimenten mit verschiedenen Berechnungsmethoden setzte sich nach und nach das ursprünglich 1990 entstandene Verfahren des Hierarchischen Radiositys durch [HSA91], das wenige Jahre später durch die Clustering-Methode zur Behandlung komplexer Szenen erweitert wurde [S94b][SAG94].

Doch die allgemeine Akzeptanz bei den Anwendern ließ immer ein wenig zu wünschen übrig. Nur wenige Softwaresysteme unterstützen die Radiositymethode, und auch das erst seit kurzem.

### 2.1 Benutzergruppen und -anforderungen

Wer ein System zur globalen Beleuchtungssimulation verwendet, tut dies aus ganz unterschiedlichen Gründen – im Vordergrund steht sicherlich die Visualisierung einer Lichtverteilung, jedoch kann auch insbesondere die Überprüfung physikalischer Beleuchtungsverhältnisse im Vordergrund stehen. Man möchte „eben mal“ das 3D-Modell eines möblierten Zimmers ausleuchten, oder den Licht- und Schattenverlauf in einem komplexen Fahrzeuginnenraum bei bestimmten Sonnenstand simulieren.

Je nach Anwendertyp ergeben sich hieraus sich teils gravierend unterscheidende Anforderungen, was z.B. Berechnungsgeschwindigkeit, visuelle Qualität oder physikalische Verlässlichkeit anbelangt.

Von verschiedenen Seiten betrachtet gibt es also unterschiedliche **Benutzergruppen**. Jede dieser Benutzergruppen hat explizite **Anforderungen**, auf die ein Radiosity-System eingehen muß, um vom Benutzer akzeptiert und gleichzeitig effizient zu sein. Schließlich können die mit einem Radiosity-System bearbeiteten **Szenentypen** klassifiziert werden, um je nach Typ unterschiedliche Berechnungsstrategien zu entwickeln.

#### 2.1.1 Benutzergruppen

Wie sehen nun diese unterschiedlichen Benutzergruppen aus?

Um diese Frage zu beantworten, ist es hilfreich, sich die generellen Möglichkeiten eines Lichtsimulationssystems nochmals ins Gedächtnis zu rufen:

Ein Radiosity-System kann die physikalischen Beleuchtungseigenschaften einer gegebenen Szene berechnen und entweder als Standbild, in einer Animationssequenz oder mit Hilfe von Echtzeit-Renderingverfahren darstellen. Die Berechnung kann hierbei grob und dafür schnell erfolgen, sie kann jedoch für eine höhere visuelle Qualität auch verfeinert werden. Es können spezielle Lichtquellen zur Anwendung kommen, der Einfluß von Tageslicht kann untersucht werden. Einfache Visualisierungsverbesserungen von Architekturszenen sind genauso denkbar wie exakte ergonomische Designuntersuchungen. In reale Fotos können, passen zur vorhandenen Beleuchtung, Objekte eingefügt werden.

Insgesamt lassen sich fünf unterschiedliche Benutzergruppen aufstellen, die aus teils sehr unterschiedlichen Motivationen heraus und mit unterschiedlichen Ergebnisvorstellungen ein Radiositysystem anwenden, und damit sehr unterschiedliche Anforderungen an ein solches System stellen.

### **VR-/Echtzeitanwender**

Für Virtual Reality Präsentationen wird Lichtsimulation verwendet, um den Realismus der Darstellung zu erhöhen, indem verschiedene Beleuchtungssituationen und vor allem Schatteninformationen in das 3D-Modell integriert werden, welches in Echtzeit dargestellt werden muß. Im Allgemeinen ist es ausreichend, die Beleuchtung als Präprozeß vorab zu berechnen, aber es sind genauso Situationen denkbar, in denen es gewünscht ist, virtuelle Objekte interaktiv zu bewegen oder zu verändern, was ein schnelles Update einer bestehenden Radiosity-Simulation erforderlich macht.

### **Lichttechniker**

Der Lichttechniker ist an physikalisch exakten, realitätsnahen Effekten und Eigenschaften von realen, vermessenen Lichtquellen interessiert, um die beste Anzahl, Art und Verteilung dieser Lichtquellen in einem gegebenen Szenario bestimmen zu können. Des weiteren möchte er bestehende Lampen und Lichtquellen im Design oder bezüglich ihrer physikalischen Eigenschaften wie Intensität und Abstrahlcharakteristik verändern.

### **Animationsspezialist**

Für computergenerierte Animationssequenzen aus einer Vielzahl einzelner Bilder ist eine optimale visuelle Qualität unabdingbar. Da jedoch die gesamte Animation im Vorfeld geplant und spezifiziert ist, ist eine voll interaktive Lichtsimulation nicht erforderlich. Auf der anderen Seite möchte man sich jedoch die zeitliche Kohärenz benachbarter Einzelbilder zunutze machen, um nicht für je-

des dieser Einzelbilder eine vollständige Radiositylösung von Grund auf berechnen zu müssen.

### **CAD-Anwender**

Lichtdesigner und Architekten sowohl im Außen- als auch im Innenbereich verwenden eine Lichtsimulation, um die beste Lampenwahl und die Platzierung von Wänden, Gebäuden und anderen Strukturen zu finden, und so eine genügende und angemessene Beleuchtung einer beliebigen 3D-Szene zu gewährleisten. Sie wollen einen Eindruck über die Wirkung ihres Modells bekommen und die verwendeten Komponenten ggf. interaktiv verändern.

### **AR-Anwender**

Im Bereich von Augmented Reality ist eine Lichtsimulation wichtig, um virtuelle Objekte nahtlos in eine reale Szene zu integrieren. Hierbei ist es nicht nur wichtig, Effekte wie Schatten und indirekte Beleuchtung durch virtuelle Objekte zu simulieren, sondern insbesondere die vorhandene, reale Beleuchtungssituation zu rekonstruieren, um die virtuellen Objekte unter den gleichen Bedingungen auszuleuchten.

Man kann sich sicherlich noch weitere Benutzergruppen überlegen, etwa der Bastler an seinem PC, der sich mal eben mit Radiosity ausgeleuchtete Ansichten seines Hauses oder seiner Umgebung schaffen will, und insbesondere auch solche Anwender, die bisher noch nicht darüber nachgedacht haben, welche Vorteile sie von einer Lichtsimulation in ihrem Bereich erlangen könnten.

Jedoch lassen sich all diese zusätzlichen Gruppen mehr oder weniger in eine der genannten fünf Kategorien einordnen, deren jede unterschiedliche Ziele und Prioritäten an ein Ergebnis eines Lichtsimulationssystems hat.

## **2.1.2 Anforderungen**

Der nächste Schritt ist nun, diese unterschiedlichen Ziele und Anforderungen der oben beschriebenen Benutzergruppen zu definieren und zu priorisieren:

Ein **VR-/Echtzeitanwender** ist hauptsächlich an einem Ergebnis interessiert, das nicht zu komplex sein darf, um es noch in Echtzeit darstellen zu können. Er benötigt keine präzisen, physikalischen Werte der verwendeten Lichtquellen, sondern ist mit einem Ergebnis zufrieden, das möglichst wenig visuelle Artefakte enthält. Was die Repräsentation der Lichtverteilung auf einzelnen Flächen angeht, kann er großen Nutzen aus einem linearen Verlauf ziehen, da dieser durch Hardware-Gouraudshading [G71] schnell dargestellt werden kann. Gegebenen-

falls ist eine gewisse Interaktion mit Szenenobjekten und -materialien erforderlich.

Der **Lichttechniker** legt keinen besonderen Wert auf „schöne Bilder“, er benötigt vielmehr präzise und verlässliche Simulationsergebnisse von direkten und indirekten Beleuchtungssituationen durch verschiedene, physikalische Lichtquellen. Diese Ergebnisse können durchaus als Falschfarben oder gar numerisch visualisiert werden. Für ihn sind daher quantitative, reproduzierbare Daten auf Basis seiner realen, vermessenen Lichtquellen wichtig.

**Animationssequenzen** erfordern hohe, visuelle Korrektheit und tolerieren keinerlei Artefakte, weder statisch noch zeitlich variierend (sog. *temporal aliasing*). Da der Ablauf der gesamten Animationssequenz vor der Berechnung bekannt ist, ist keine Interaktion erforderlich.



Der **CAD-Anwender** kann als Mischung dieser drei Typen angesehen werden: Während er nicht auf ein vollkommen perfektes, visuelles Ergebnis aus ist, so ist doch eine gewisse Korrektheit und Qualität erforderlich; Interaktion mit der Szene und schnelle Updates sind besonders wichtig.

Für den **AR-Anwender** schließlich ist ein anderer Aspekt der Radiositysimulation wichtig: die Rekonstruktion vorhandener Beleuchtungseigenschaften.

Man beachte den Unterschied zwischen *physikalischer* und *visueller* Korrektheit: Eine physikalisch exakte Simulation muß nicht zwangsläufig auch visuell hochqualitativ sein – sind nur Beleuchtungswerte erforderlich, ist es unnötig, zusätzlichen Berechnungsaufwand in die Beseitigung von visuellen Artefakten zu investieren. Auf der anderen Seite muß ein visuell ansprechendes Ergebnis nicht notwendigerweise physikalisch korrekt sein, um glaubwürdig zu wirken – im Extremfall kann es sogar erwünscht sein, „unerlaubte“ Beleuchtungssituationen zu realisieren, um den gewünschten optischen Effekt zu erzeugen, z.B. in der Unterhaltungsindustrie [A98].

Die unterschiedliche Relevanz der verschiedenen Anforderungen in Bezug zu den genannten Benutzergruppen sind in folgender Tabelle zusammengefaßt:



Skala: sehr wichtig  wünschenswert  irrelevant 









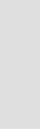




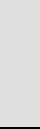
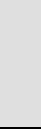








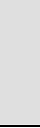
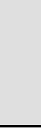








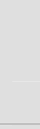
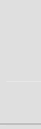



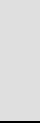
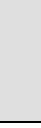
	VR	Lichttechnik	Animation	CAD	AR
physikalische Korrektheit					
visuelle Korrektheit					
visuelle Artefakte					
niedriger Speicherbedarf					
geringe Komplexität					
Interaktion	 *				
Dynamik	 *				
Lineare Darstellung			 **		

Tabelle 1: Relevanz verschiedener Anforderungen je Benutzergruppe.

\* Bei statischen Szenen ist Interaktion nicht erforderlich: die vorberechneten Radiositywerte ändern sich nie. Wird die Szene jedoch verändert (Stühle bewegt, Materialien geändert, ...), so sind Interaktion und schnelles Radiosityupdate essentiell.

\*\* Eine mindestens lineare Repräsentation ist erforderlich – falls nichts besseres verfügbar ist.

Anhand dieser Aufstellung ist ersichtlich, daß es nicht wünschenswert sein kann, einen einzigen, hochkomplexen Algorithmus zu entwerfen, der diese unterschiedlichsten Anforderungen gleichzeitig betrachten und bearbeiten kann. Statt dessen ist es sinnvoll, einzelne, spezialisierte Algorithmen zu konstruieren, die gezielt auf die expliziten Anforderungen einzelner Benutzergruppen angepaßt sind, wie beispielsweise ein pixelbasierter *final gathering*-Algorithmus, um qualitativ hochwertige Bilder zu erzeugen.

Auf der anderen Seite ist die Kombination und Integration dieser Algorithmen in einem einzigen, globalen Simulationssystem erstrebenswert, um die dadurch bereitgestellten Möglichkeiten zu erweitern: Die physikalische Lichtverteilung kann durch eine halbtransparente Falschfarbenebene in Echtzeit dargestellt werden. Lichttechniker könnten an einer tatsächlichen Visualisierung bestimmter Lichtquellen interessiert sein, auch wenn sie bisher an 2D-Diagramme und Tabellen gewöhnt waren. Animationssequenzen beleuchteter Szenen können in einer Echtzeit-Vorschau betrachtet werden, um z.B. Kamerapfade und Objektpositionen zu optimieren. CAD-Anwender sind gewohnt, ihre Szene in einem einzigen System zu erstellen, zu bearbeiten und darzustellen. Und im AR-Bereich schließlich ist eine möglichst unauffällige Integration von korrekt ausgeleuchteten Objekten in das Visualisierungssystem wichtig.

Die ganze Aufgabe wird mehr und mehr durch immer leistungsfähigere Rechner und Systeme erleichtert, die es inzwischen auch dem Heimanwender ermöglichen, komplexe 3D-Modelle zu bearbeiten.

### 2.1.3 Szenentypen

Zusätzlich zu den unterschiedlichen Benutzergruppen und ihren Anforderungen wirkt sich auch die Beschaffenheit einer vorliegenden Szene entscheidend auf die optimale Funktionsweise und damit die Anwendbarkeit eines Radiosity-Simulationssystems aus. Szenen können sich, je nach Anwendungsgebiet, teilweise gravierend in ihrer prinzipiellen Beschaffenheit voneinander unterscheiden, und so spezialisierte Algorithmen erforderlich machen – bzw. begünstigen.

#### Architekturszenen

Architekturszenen bestehen üblicherweise aus relativ großen, ebenen Flächen (Wände, Fußböden, Möbel, ...), die in Schattenbereichen eine hohe Unterteilung erforderlich machen.

#### Fahrzeuginnenräume

Typisches Merkmal von Fahrzeuginnenräumen sind insbesondere teilweise stark gekrümmte, sich kaum überlappende Flächen, die üblicherweise bereits als feintessellierte Polygonmenge vorliegt. Eine zusätzliche Unterteilung ist selten erforderlich, jedoch muß besonderen Wert auf die Beseitigung von sichtbaren Übergängen in den Schattierungen benachbarter Flächen gelegt werden.

### **Innenraumszenen**

Dieser Szenentyp zeichnet sich durch eine Vielzahl Lichtquellen mit unterschiedlichsten Abstrahlcharakteristiken aus, Objekte sind meist dicht aneinandergedrängt und verschachtelt.

### **Außenszenen**

Hier sind einzeln stehende Objekte, die meist innerhalb einer großen Ebene verteilt sind, hervorstechend. Tageslichtsimulation ist wesentlicher Bestandteil.

### **Unterhaltungsindustrie**

Jeder vorstellbare Szenentyp.

Es ist ebenfalls denkbar, daß jede Benutzergruppe ihren „typischen“ Szenentyp besitzt, aber im Prinzip sind Benutzer und Szenentyp frei kombinierbar.

Ein Radiosity-System muß flexibel genug sein, all diese Szenentypen effizient behandeln zu können.

## **2.2 Mängel existierender Radiositysysteme**

Lichtsimulationssysteme auf Basis des Radiosity-Verfahrens werden allmählich von mehr und mehr kommerziellen und professionellen Animationssystemen unterstützt. Warum die „Markteinführung“ so lange dauerte – man bedenke, daß seit der Entwicklung des ersten Radiosity-Algorithmus’ nunmehr 19 Jahre vergangen sind – hat vielschichtige Gründe.

In den ersten Jahren war das Radiosityverfahren sicherlich reines Forschungsgebiet und für den Normalanwender kaum sinnvoll nutzbar. Nachdem sich aber ersteinmal grundsätzliche Algorithmen etabliert hatten, wie insbesondere z.B. das Progressive Refinement [CCWG88], begannen erste Softwarepakete darauf aufzubauen, während sich die Forschung nun mit der Verbesserung dieser oder der Entwicklung neuer Algorithmen befassen konnte.

Trotzdem sind auch heute noch Special Effects oder gar computeranimierte Filme zu sehen, die keine globale Beleuchtungssimulation verwenden, da die verfügbaren Radiositysysteme immer noch starke Einschränkungen haben.

Worin die einzelnen Problematiken von Radiositysystemen bestehen, soll in diesem Abschnitt genauer analysiert werden. Neben der Analyse werden hierbei insbesondere auch Lösungsmöglichkeiten vorgeschlagen, diese Probleme zu beseitigen, oder doch zumindest abzumildern.

Ein Teil der hier vorgestellten Eigenschaften von Radiositysystemen wurden im Rahmen des Esprit Forschungsprojektes LTR 24944 ARCADE [A00] zusammengestellt.

### 2.2.1 Optische Mängel

Die Klasse der optischen Mängel umfasst insbesondere alle sichtbaren Artefakte, d.h. alle Unstetigkeiten, die den Realitätseindruck stören.

#### Schatten- und Lichtlöcher

In allen Fällen, in denen ein verdeckendes Objekt einen Empfänger berührt, entsteht eine D0-Unstetigkeit, ein Sprung in der Radiosityfunktion. Solange diese Berührungskanten nicht explizit im Radiositymesh repräsentiert sind und zusätzliche Informationen enthalten (wie z.B. zwei unterschiedliche Radiositywerte an jedem Punkt der Berührungskanten), werden Schatten unter dem verdeckten Objekt und unverschattete Bereiche „herauslaufen“, oder umgekehrt Licht in eigentlich verschattete Zonen hinüberstrahlen, da die Samplingwerte auf den Empfängerpatches linear interpoliert werden.

Identifikation und explizite Unterteilung entlang solcher Berührungskanten verhindert Schatten- und Lichtlöcher. In Frage kommende Kanten können in einem Präprozess vor der eigentlichen Lichtsimulation ermittelt werden.

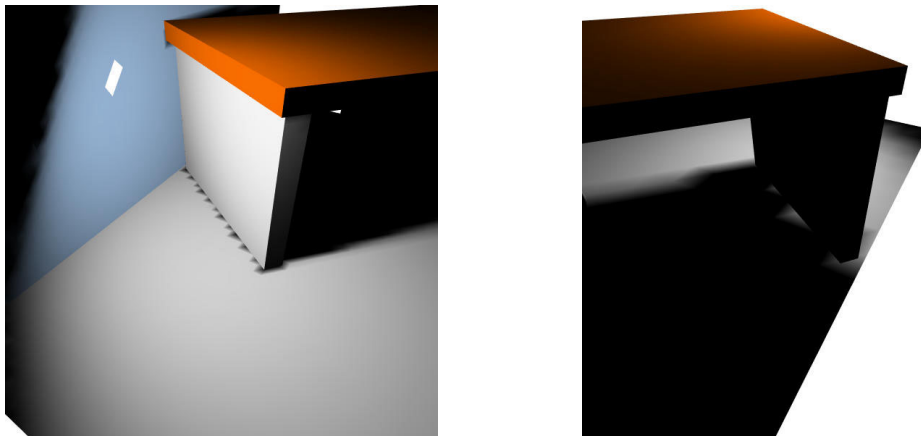


Abb. 1: Schattenlöcher (links) und Lichtlöcher (rechts).

#### Gezackte Schattenkanten

Gezackte Schattenkanten sind ein typisches Problem aller symmetrischen Unterteilungsverfahren und werden, je nach Orientierung, manchmal durch Gouraudshading verstärkt, manchmal reduziert. Sie entstehen überall dort, wo eine Schattenkante nicht zu einem Empfänger ausgerichtet (d.h. parallel zu einer sei-

ner Mesh-Kanten) ist. Das Problem wird bei der Verwendung von Eckpunktradiositäten durch den fehlenden „Verwischungseffekt“ der zweiten linearen Interpolation von Patchmittelpunkten zu dargestellten Eckpunkten verstärkt.

Der Effekt kann durch sehr hohe, ansonsten jedoch überflüssige Unterteilung entlang entsprechender Schattenkanten vermindert, nicht jedoch vollständig beseitigt werden. Optimalerweise sollte der Empfänger an solchen Stellen asymmetrisch unterteilt werden, allerdings sind entsprechende vollständige *discontinuity meshes* [H92][LTG92] zu rechenintensiv und numerisch zu instabil, um für ein anwendbares Radiositysystem in Frage zu kommen. Soft shadow textures [SS98] können eine Lösung für einen Teil der Fälle darstellen.

### Mach-Bänder

Gouraud-Schattierung ist anfällig und bekannt dafür, Mach-Bänder [M1865], also physiologisch verstärkte Unstetigkeiten ersten Grades, zu erzeugen (siehe z.B. auch [G95]).

Mach-Bänder müssen erkannt werden (z.B. mit Hilfe des Radiosity-Gradienten) und können nur durch nichtlineare Interpolation, was die Verwendung von Echtzeitrendering-Hardware ausschließen würde, oder durch eine höhere Unterteilung beseitigt werden.

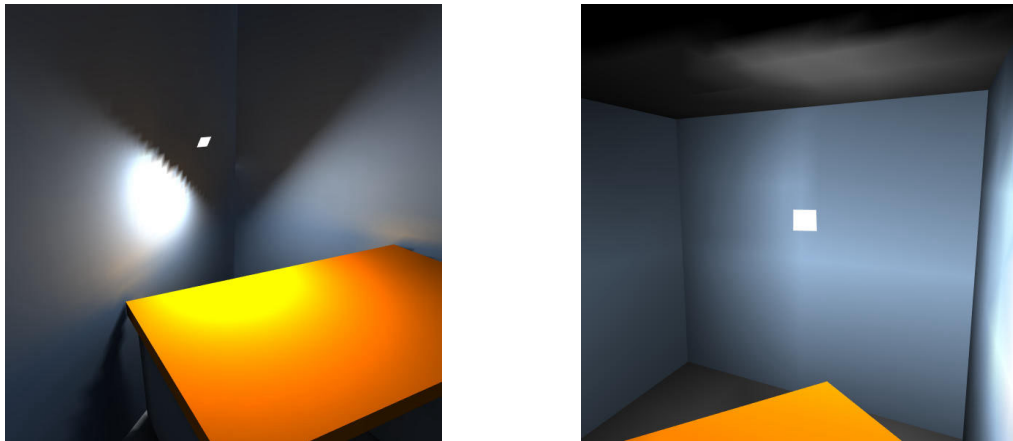


Abb. 2 : Gezackte Schattenkanten (links) und Machbänder (rechts).

### Nähte

Nähte entstehen überall dort, wo Eckpunkte entlang von Berührungskanten benachbarter Polygone nicht aufeinander fallen und keine topologische Information über Polygongrenzen hinweg vorhanden ist. Sie werden durch die Verwendung von Mittelpunktradiositäten erheblich verstärkt.

Eckpunkte von aneinandergrenzenden, koplanaren Flächen müssen gemeinsam von allen beteiligten Flächen verwendet werden, gegebenenfalls muß dies durch sog. *zip-locking* in einem Präprozeß sichergestellt werden.

Bei nicht zusammenfallenden Berührungspunkten muß in einem Postprozeß eine Verankerung der entsprechenden Kanten durchgeführt werden.

### T-Vertices

Ähnlich den Nähten weisen T-Vertices auf fehlende Topologieinformationen innerhalb eines Polygons hin und entstehen überall dort, wo sich die Unterteilungsebene benachbarter Patches unterscheiden.

Sie können wirkungsvoll durch ein *restricted mesh*, d.h. einem balancierten Mesh, das anschließend verankert wird, beseitigt werden.

Da die durch Balancierung und Verankerung erzeugten Patches für die eigentliche Radiosity-Simulation irrelevant sind, bietet es sich an, sie nur temporär und direkt vor der Darstellung zu berechnen.

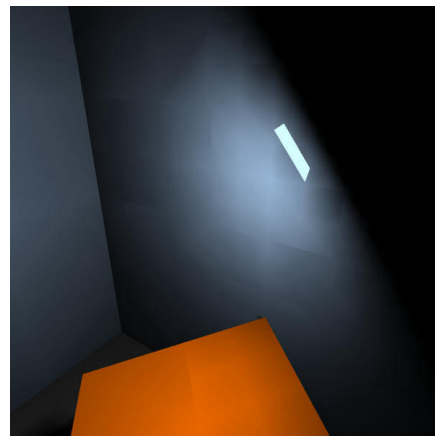
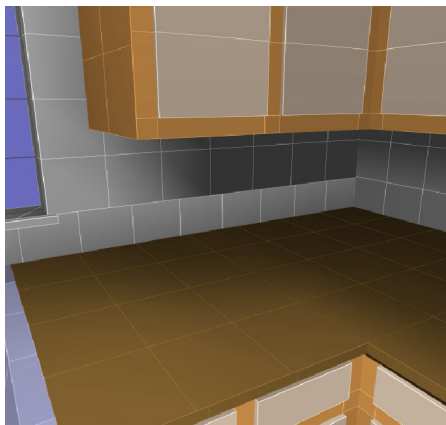


Abb. 3 : Nähte (links) und T-Vertices (rechts).

### Fehlende Schattenkanten

Sicherlich die schwerwiegendsten Artefakte sind vollständig fehlende Schatten oder Schattenbereiche, die durch unzuverlässige Sichtbarkeitserkennung, wie z.B. bei der Verwendung von *ray casting*, verursacht werden.

Die logische Alternative hierzu ist, eine verlässliche Sichtbarkeitsklassifizierung zur Verfügung zu stellen, und diese zusätzlichen Informationen während dem Refinement zu nutzen. Dies wird in Kapitel 2.7 näher erläutert.

### *temporal aliasing*

Zeitliches Aliasing, d.h. sichtbare Sprünge durch plötzlich wechselnde Unterteilungstiefen oder auftauchende Schatten, ist das wesentliche Artefakt, das bei animierten Sequenzen auftreten kann.

Sichtbare Wechsel der Unterteilungstiefen deuten auf unzureichende Unterteilung oder ungenügende Behandlung von Mach-Bändern und/oder T-Vertices hin. Plötzlich auftauchende oder verschwindende Schatten sind Fehler in der Sichtbarkeitserkennung. Die entsprechenden Vorgehensweisen zur Beseitigung dieser Problematiken sind oben beschrieben. Zusätzlich hilfreich können Informationen über die beweglichen Objekte, sowie die von diesen Objekten verschatteten Empfänger sein.

## 2.2.2 Physikalische Genauigkeit

Bei der physikalischen Genauigkeit geht es eher um die Exaktheit des Simulationsergebnisses, als um ein ansprechendes, optisch einwandfreies Bild.

### Genauigkeit der Formfaktorberechnung

Die Frage der Formfaktorberechnung war lange Zeit wichtiger Gegenstand jeder Radiositysimulation. Das grundlegende Problem bestand darin, daß der Energieaustausch zwischen zwei Flächen durch den sogenannten Formfaktor beschrieben wird, ein Doppelintegral, das zwar seit mehreren hundert Jahren bekannt, aber erst vor kurzem analytisch gelöst wurde [SH93]. Die analytische Berechnung ist jedoch sehr aufwendig, weshalb es verschiedenste numerische und analytische Verfahren gibt, Formfaktorvereinfachungen oder -abschätzungen zu berechnen (vgl. z.B. [M96]).

Die beste Lösung der Problematik ist nun, den Formfaktor immer von einer Senderfläche zu einem expliziten Empfängerpunkt (dem Mittelpunkt bei Mittelpunktradiosities, einem Eckpunkt bei Eckpunktradiosities) zu berechnen. Hierzu kann dann eine exakte, analytische Formel, die sog. *Prisma-Formel* [HS67][M96], bei dem die erste Formfaktorvereinfachung mit Hilfe des Satzes von Stokes [BS85] in ein Konturintegral umgewandelt wird, eingesetzt werden.

Bei der Berechnung des Formfaktors für Spezialfälle in der geometrischen Konfiguration zweier Patches müssen besondere Maßnahmen getroffen werden, um mathematische Korrektheit zu garantieren. Zum Beispiel müssen sog. *support plane splits* korrekt behandelt werden, um sichtbare Schattenkanten oder gar negative Formfaktoren zu vermeiden.

## Formfaktorsingularitäten

Für verschiedene geometrische Situationen existieren bestimmte, degenerierte Fälle, in der die Anwendung der analytischen Fläche-zu-Punkt-Formel aufgrund von Singularitäten zu keiner Lösung führt, da z.B. einzelne Abschnitte oder gar die Eingangsvektoren des Konturintegrals 0 werden. Das Problem liegt z.B. dann vor, wenn der Empfänger den Sender (oder genauer, die Senderebene) berührt, oder der Empfängerpunkt auf der Umrandung des Senders liegt, und stellt sich häufig, wenn Eckpunktradiositäten verwendet werden. Für diese Fälle liefert selbst das Nusselt-Analogon [N28][MBG86] keine brauchbaren Werte mehr, da die Projektion des Senders auf eine eindimensionale Linie ohne Fläche reduziert wird. Für solche Fälle ist der korrekte Formfaktor jedoch keinesfalls 0 (vgl. [K97]), sondern müßte eigentlich durch eine Grenzwertbetrachtung bestimmt werden.

Für einzelne Spezialfälle können hierfür analytische Formeln gefunden werden (z.B. rechter Winkel zwischen Sender und Empfänger, usw.), jedoch existieren für die Vielzahl der unterschiedlichen Fälle erfahrungsgemäß keine allgemeingültigen, mathematischen Lösungen, daher ist die praktikabelste Vorgehensweise, den Empfängerpunkt in solchen Fällen für die Berechnung um ein  $\epsilon$  aus der Senderebene herauszubewegen.

## Support plane split auf der Empfängerseite

Ein Split des Empfängers durch die erweiterte Senderebene erzeugt eine D1-Unstetigkeit, die – wenn das Refinementkriterium es für erforderlich hält – durch entsprechende Unterteilungen abgefangen werden muß, im optimalen Fall asymmetrisch entlang des Splits.

## Support plane split auf der Senderseite

Tritt ein support plane split auf der Senderseite auf, so ist eine physikalische Unterteilung des Patches nicht erforderlich, da keine sichtbare Schattenkante entstehen kann. Jedoch darf der Formfaktor nur für den vom Empfänger aus gesehen im vorderen Halbraum befindlichen Teil berechnet werden.

### 2.2.3 Physikalische Wiedergabequalität

Neben der visuellen Qualität ist, insbesondere für den Lichttechniker, eine Verlässlichkeit bzgl. der photometrischen Werte, sowie der reproduzierten Farben erforderlich. Insbesondere die Verwendung der RGB-Farbkonvention macht eine konsistente Simulation schwierig, da die Wirkung von RGB-Farben von dem jeweils verwendeten Ausgabemedium abhängt und nicht eindeutig definiert ist.



Die Thematik der photometrischen Konsistenz des Radiosity-Systems wird in Kapitel 2.3 näher besprochen.

### 2.2.4 Unnötige Unterteilungen

Klassische Radiosityverfahren verwenden fast ausschließlich das einfache Unterteilungskriterium des Originalalgorithmus für hierarchisches Radiosity: das *BF*-Kriterium [HSA91]. Der grundsätzliche Nachteil hierbei liegt darin, daß eine Interaktion unterteilt wird, sobald die übertragene Gesamtenergie über einer vorgegebenen Schranke liegt. Aber eine solche Unterteilung ist oft nicht notwendig, und es gibt viele Situationen, bei denen durch eine Unterteilung weder die mathematische, noch die visuelle Qualität merklich gesteigert wird, weshalb andere, intelligentere Refinementkriterien eingesetzt werden müssen.

Ein zweiter Fall, bei dem zur Vermeidung unnötiger Komplexität Unterteilungen eingespart werden können liegt vor, wenn man die Darstellungsmethode in die Simulation mit einbezieht. Weiß man z.B., daß Gouraud-Shading zur Darstellung verwendet wird, kann man daraus ableiten, daß Unterteilungen überall dort überflüssig sind, wo die Radiosityfunktion auf dem Empfänger einen linearen Verlauf besitzt.

Eine genauere Diskussion dieser Thematik findet sich in den Kapiteln 2.5 und 2.6.

Unnötige Unterteilung tritt auch überall dort auf, wo eine Interaktion als partiell verdeckt erkannt wurde, obwohl keine Verdeckung (oder gravierender: eine unerkannte vollständige Verdeckung) vorliegt.

### 2.2.5 Ungenaue Sichtbarkeitsberechnung

Ein großes Manko klassischer Radiositysysteme liegt in einer fehlerhaften, unzuverlässigen Sichtbarkeitsberechnung, wodurch viele der in Kapitel 2.2.1 beschriebenen Artefakte verursacht werden.

Verwendet man einfache Methoden zur Sichtbarkeitsbestimmung, wie z.B. Ray Casting, so unterliegt man der Gefahr, Objekte (und damit möglicherweise sichtbare Schatten) zu verpassen, wodurch Fehler in der Simulation entstehen. Auch bei bestimmten Sonderfällen (z.B. *support plane splits*) sind verlässliche Sichtbarkeitsberechnungen erforderlich.

## 2.2.6 Szenenvorverarbeitung

Geometrie, für die eine Lichtsimulation durchgeführt werden soll, muß bestimmten Kriterien genügen, die im Allgemeinen etwas einschränkender sind als entsprechende Anforderungen an CAD-Systeme, wo solche Szenen üblicherweise entstehen.

Während mit entsprechendem Aufwand im Radiosity-System jeder Sonderfall von Geometrieentartung direkt behandeln werden könnte, ist es dagegen viel nutzbringender, wenn entsprechende Problemfälle in Vorverarbeitungsschritten in Standardobjektstrukturen verwandelt werden, mit denen das System effizient umgehen kann. Diese Schritte sollten, wenn irgendwie möglich, automatisch erfolgen, um den Benutzer nicht mit einer Vielzahl von interaktiv durchzuführenden, aufwendigen Optimierungsaufgaben zu belasten, wie z.B. bei einem Modell mit inkonsistenten Normalen die Flächennormalen von jedem einzelnen Polygon mit falscher Normalenrichtung umzudrehen.

### Triangulierung

Der normale Radiosityalgorithmus kann nur konvexe, planare Polygone mit entweder drei oder vier Ecken bearbeiten. Alle anderen Varianten müssen vorher trianguliert werden.

Bei bereits triangulierten Modellen besteht die Möglichkeit, daß die vorliegenden Dreiecke ungünstig geformt sind (schlechte aspect-ratio, *fans*). Hier sollte ein weiterer Vorverarbeitungsschritt zwischengeschaltet werden, der koplanare, verbundene Dreiecke vereinigt und sinnvoll (z.B. durch *constrained delaunay triangulation*) retrianguliert, um später visuelle Artefakte zu vermeiden.

Hat man die Möglichkeit, die Triangulierung zu beeinflussen (was normalerweise selten möglich ist, da das Ausgangsmodell oft bereits in triangulierter Form vorliegt), so ist eine Delaunay-Triangulierung sinnvoll, oder zumindest die Angabe einer minimalen Kantenlänge. Ansonsten entstehen lange, dünne, und manchmal entartete Dreiecke, oder „triangle fans“, die sichtbare, kaum zu reduzierende Artefakte verursachen.

### Degenerierte Geometrien

Problematische Geometrien, also z.B. Polygone mit Löchern oder Selbstdurchdringungen, müssen vor Beginn der Radiositysimulation bereinigt werden. Sich überlappende oder gar identische Geometrien müssen beseitigt werden. Sich durchdringende Geometrien sollten vermieden werden, da sie zu Problemen

(und damit letztendlich zu Artefakten) bei der Sichtbarkeitsbestimmung führen können.

Um Selbstverschattungen von schlecht tessellierten, „verknöteten“ Objekten zu vermeiden, bietet sich ein benutzerdefinierbarer Epsilon-Parameter in Welt-/Modellkoordinateneinheiten an, der festlegt, bis zu welchem Abstand von Empfänger/Sender *keine* Sichtbarkeitsüberprüfung stattfindet. Dieser Parameter kann auch dazu dienen, die Problematiken von sich überlappenden Geometrien zu reduzieren.

## Nachbarschaft

Verschiedene Stufen von Nachbarschaftsbeziehungen und -informationen sind relevant: auf niederem Level zwischen einzelnen, während der Simulation erzeugten Patches, sowie zwischen einzelnen, sich berührenden Polygonen.

Letzteres ist vor allem bei CAD-Szenen, die aus Freiformflächen entstanden sind relevant, da es sonst nicht möglich ist, Radiositywerte zwischen verschiedenen Polygonen zu interpolieren, wodurch sichtbare Kanten entstehen. Solche benachbarten Polygone (Polygone mit gemeinsamen Kanten) müssen z.B. in einem Präprozeß bestimmt werden, um dann vor der Darstellung Radiosity direkt zu interpolieren, oder aufgrund unterschiedlichem Unterteilungslevel solcher Polygone sie durch zip-locking zu verankern.

Interpoliert wird hierbei zwischen Punkten, die dieselben Eckpunktnormalen besitzen. Fehlt diese Information, ist ein Grenzwinkel (zwischen benachbarten Polygonflächennormalen) erforderlich, damit über gewollt „scharfe“ Kanten hinweg nicht interpoliert wird. Es darf dabei natürlich generell nur zwischen Polygonen interpoliert werden, die dasselbe Material besitzen.

## *Stitching*

Bei bestimmten, bereits triangulierten Modellen kann es passieren, daß eigentlich benachbarte Meshes von unterschiedlichen Freiformflächen nicht zusammenhängen und so Löcher in der triangulierten Geometrie entstehen. Um eine Interpolation zwischen solchen Polygonen zu ermöglichen (und dadurch sichtbare Kanten zu vermeiden), müssen diese Löcher durch sog. *stitching* gestopft werden. I.A. besitzt bereits der Triangulierer eine solche Option (ggf. im Zusammenhang mit „Skins“), sodass dies bereits beim Export aus dem CAD-System behandelt werden kann.

## Normalenkonsistenz

Übliche Radiosityverfahren sind auf eine konsistente Ausrichtung der Flächennormalen angewiesen, da für einen Lichtaustausch definierte „Vorder-“ und „Rück-“seiten benötigt werden. Diese muß im Normalfall durch geeignete manuelle oder halbautomatische Vorverarbeitung sichergestellt werden.

Die für den Radiosityprozeß relevante Flächennormale ergibt sich normalerweise implizit aus dem Umlaufsinn der Polygoneckpunkte. Nur selten wird diese Eigenschaft jedoch von CAD-Systemen bei der Ausgabe von triangulierten Polygonmodellen beachtet.

Statt nun den Benutzer interaktiv das gesamte Modell nach falschen Normalen durchforsten zu lassen ist es möglich, statt der Flächennormalen die Eckpunktnormalen zu verwenden. Tessellierte Modelle sind üblicherweise mit Eckpunktnormalen versehen, um ein interpoliertes Gouraudshading zu ermöglichen. Bei manchen Tessellierern scheint es üblich zu sein, daß die Eckpunktnormalen exakte Informationen der Freiformfläche beinhalten, während die 3D-Punkte der tessellierten Polygone innerhalb des Sehnenfehlers unkorrelierte Abweichungen besitzen, sodass die Flächennormalen dieser Polygone einen weit höheren Fehler besitzen, und manchmal sogar einfach in die entgegengesetzte Richtung weisen.

Die Verwendung von Eckpunktnormalen im Radiosityverfahren ergänzt sich auch sehr gut mit der Berechnung von Eckpunktradiositäten.

Nachteil dieser Methode ist, daß mit Hilfe von  $u/v$ -Koordinaten bei jedem Zugriff auf eine Patchnormale die Eckpunktnormalen interpoliert oder bei den Eckpunkten gespeichert werden müssen, während eine Flächennormale für sämtliche Patches eines Polygons gleich ist. Dies bedeutet einen höheren Aufwand. Der Benefit ist sicherlich jedoch eine genauere Darstellung, da selbst bei grober Tessellierung die Normale über ein Polygon nicht als konstant angenommen, sondern jeweils korrekt interpoliert wird, sodass trotz Flächenrepräsentation die die Beleuchtung beeinflussende Objektkrümmung erhalten bleibt.

Sollte doch mal eine interaktive Normalenkorrektur erforderlich sein, weil z.B. das Eingangsmodell keine Eckpunktnormalen zur Verfügung stellt, so gibt es verschiedene Vorgehensweisen:

- Der User kann einzelne Polygone umdrehen. Sehr mühsam.
- Polygone, die eine gemeinsame Kante haben, werden in die gleiche Richtung gedreht. Das vom User selektierte Initialpolygon kann hierbei zuerst dem Betrachter zugewandt werden. Problematisch ist diese Methode bei nicht eindeutigen Topologien (z.B. falls eine Kante zu mehr als zwei Polygonen gehört).

- Alle sichtbaren Polygone werden durch einen Raytracing-Sweep dem Betrachter zugewandt. Problematisch ist, daß leicht Polygone verpaßt werden und als einzelne „Flecken“ übrig bleiben, oder daß durch Löcher in der Geometrie Polygone fälschlicherweise umgedreht werden.
- Eine vollautomatische Methode ist, durch Heuristiken „innen“ und „außen“ von Objekten zu entscheiden und daraufhin Polygone zu drehen [MF97].
- Sind die Informationen der Eckpunktnormalen korrekt, die Punktreihenfolgen aber falsch, ist ggf. eine Funktion nötig, die die Flächennormalen entsprechend der allgemeinen Richtung der Eckpunktnormalen anpaßt.

### Freiformflächen

Speziell Karosseriemodelle bestehen hauptsächlich aus nichtplanaren Freiformflächen, die zu Beginn trianguliert wurden. Qualitativ wichtig ist die Möglichkeit, nachträglich aus dem NURBS- oder sonstigen Freiform-Modell Punkte zu generieren. Damit ist es möglich, bei einer Radiosityunterteilung exakte Patches zu erzeugen, indem durch Unterteilung entstandene Punkte auf der Freiformfläche liegen, anstatt die vorhandenen, groben Polygone planar weiter zu unterteilen. Durch Interpolation von Eckpunktnormalen werden jedoch die im letzteren Falle entstehenden (visuellen) Artefakte bereits stark abgemildert, was eine entsprechende Behandlung von Freiformflächen nur bei hoher geforderter physikalischer Korrektheit erforderlich macht.

### Objekthierarchie

Neben der Unterteilung ab Polygonlevel (die „untere“ Hierarchie der Patches) müssen für hierarchisches Radiosity mit Clustering Polygone zu Clustern zusammengefaßt werden („obere“ Hierarchie). Während sich die Struktur der Patchhierarchie während des Radiosityverfahrens von selbst ergibt und durch das verwendete Meshing-Verfahren festgelegt ist, ist der Aufbau der Clusterhierarchie eine eigene (Präprozeß-)Aufgabe. Hierbei ist unbedingt zu beachten, daß die Struktur der Clusterhierarchie ganz wesentlich auf Geschwindigkeit und Darstellungsqualität der Radiosityberechnung Einfluß nimmt.

Im einfachsten Fall wird hierzu die vorhandene Modellierungshierarchie übernommen, und zusätzlich entsprechend der Methode von Goldsmith&Salmon [GS87] Zwischenknoten eingefügt. Die so entstandene Hierarchie ist aber normalerweise unzureichend, da die Modellierungshierarchie selten über eine für ein Radiositysystem geeignete Struktur verfügt (wenn z.B. alle Türgriffe eines Gebäudes in einem Objekt vereint sind).

In solchen Fällen muß in einem Präprozeß eine geeignete Hierarchie erstellt werden. Es bietet sich an, dies objektweise zu tun, d.h. nicht allein durch Bounding-Boxes bestimmt, sondern anhand von zusammengehörenden Polygongruppen (d.h. Polygone mit gemeinsamen Kanten/Punkten) [HDSD99]. Noch vorteilhafter und sehr effizient ist die Verwendung von *face clustering* [WHG99][GWH01], bei dem zusammenhängende Polygone unter Berücksichtigung ihrer Normalenausrichtung (*cone of normals*) gruppiert werden.

Hat man auf diese Weise Gruppen gebildet, können diese durch weitere Goldsmith&Salmon-Gruppierung nach unten bis auf Polygonebene, sowie nach oben bis zur Szenenwurzel zu einer Hierarchie ausgebaut werden.

Wichtig ist hierbei, daß im Originalmodell vorhandene Transformationsmatrizen entsprechend beachtet und durchgereicht werden. Ist dies nicht möglich, müssen in einem Vorverarbeitungsschritt sämtliche Transformationsmatrizen rausgerechnet werden (sog. *flushing*).

## Meshing

Statt eines normalen Quadtree-Meshings hat sich ein erweiterter Quadtree als am sinnvollsten erwiesen, da hierdurch insbesondere das Auftreten von T-Vertices unterbunden wird. Drei- und Vierecke werden hierzu in jeweils 4 regelmäßige Kinder unterteilt. Ist jedoch die Aspect-Ratio des Eingangspatches zu schlecht ( $> \sqrt{2}$  oder  $< \frac{1}{\sqrt{2}}$ ), so werden statt vier Kindern nur zwei erzeugt. Diese Methode wird auch für Dreiecke angewendet, da ein überwiegender Teil der zu bearbeitenden Geometrien aus triangulierten Modellen besteht. Hierzu wird ein schmales Dreieck zunächst in der Mitte der Längsachse in ein kongruentes Dreieck und ein Viereck unterteilt. Mit diesen wird wie gewohnt weiterverfahren.

Der Algorithmus um Nachbarschaften in einem solchen Quadtree zu finden ist jedoch komplizierter als im einfachen Fall und muß entsprechend angepaßt werden, damit die Generierung eines *restricted meshes* zur Visualisierung möglich ist.

### 2.2.7 Bedienbarkeit

#### Parametervielfalt

Der Benutzer eines Radiositysystems ist generell mit der Vielzahl von Parametern überfordert, die die verschiedensten Aspekte des Verhaltens eines Radiositysystems beeinflussen, und die teilweise gegenseitig voneinander abhängen.

Sehr viele der in einem Radiositysystem relevanten Parameter lassen sich durch Automatismen auf vernünftige Werte voreinstellen, um einen unbedarften Benutzer nicht zu überfordern. Möchte ein Benutzer explizit Standardparameter anpassen, kann er dies z.B. in einer „Experten-Variante“ des Benutzerinterfaces tun.

Hier sollen nur ein paar ausgewählte Parameter exemplarisch als Beispiele dienen. Eine genauere Diskussion findet sich z.B. in [K97]:

- **Level**  
Der maximale Unterteilungslevel eines Polygons wird festgelegt. Üblicherweise kann dieser Wert auf das Maximum, das die Datenstrukturen verkraften gesetzt werden, da sich das Refinement-Kriterium um die notwendige Genauigkeit kümmert. Es kann jedoch sinnvoll sein, den Benutzer einen kleineren Wert wählen zu lassen (z.B. weil sich Flächen sehr dicht an Lichtquellen befinden, oder die Szene aus sowohl sehr großen, als auch sehr kleinen Polygonen besteht).
- **Area**  
Die Fläche, bis zu der ein Patch maximal unterteilt werden darf. Üblicherweise kann dieser Wert auf 0 gesetzt werden. Ansonsten gelten die gleichen Richtlinien wie für den Level.
- **Refinement-Epsilon**  
Dieser Parameter ist der einzige, der vom Benutzer eingestellt werden muß. Er ist ein Maß dafür, ob eine Szene eher fein mit hoher Qualität berechnet wird, oder grob, aber dafür schnell. Er ist, was die Skalierung betrifft, von der Szenendimension unabhängig, hängt ansonsten jedoch stark von Geometrie-eigenschaften, sowie Lichtquellen-Eigenschaften wie Platzierung und Intensität ab.

### **Berechnungszeit vs. Feedback vs. Qualität**

Statt einer Vielzahl von nicht-intuitiven Parametern (z.B. „Refinement-Epsilon“) ist es für ein benutzerfreundliches System wichtig, selbsterklärende Parameter und Zusammenhänge zu verwenden. So ist zum Beispiel offensichtlich, daß im Allgemeinen eine reziproke Beziehung zwischen Berechnungszeit und Qualität besteht. Indem man die von Benutzer steuerbaren Einstellungen auf so einfache Begriffe reduziert, erhöht man die Verwendbarkeit – und damit die Akzeptanz – eines Radiositysystems.

Auch z.B. die Frage des zeitlichen Feedbacks spielt in diese Konstellation mit ein: Möchte man einfach nur eine schnelle Berechnung, ohne Wert auf sichtbare Zwischenergebnisse zu legen, oder möchte man nach jeder Iteration einen Eindruck der Szene bekommen? Ist man einfach an einer fertigen Lösung interessiert, oder ist ein ständiges Feedback, z.B. für den interaktiven VR-User, zwin-

gend erforderlich? Das Radiositysystem muß in der Lage sein, auf solche Anforderungen eingehen zu können.

### **Unterstützte Lichtquellen**

Radiositysysteme können oft nur mit einem einzigen Lichtquellentyp umgehen: Lambert'schen Flächenstrahlern – Polygonen, denen ein Emissionswert zugewiesen wurde.

Um für Benutzer wie Lichttechniker und Designer interessant zu sein, ist es jedoch unabdingbar, daß das Radiositysystem weitere, komplexere Lichtquellen verarbeiten kann, insbesondere Lichtquellen mit Lichtstärkeverteilungskurven, und Tageslicht.

Die besondere Handhabung und Integration dieser Lichtquellen in das Simulationsverfahren wird in Kapitel 2.4 beschrieben.

### **2.2.8 Zusammenfassung**

In der folgenden Tabelle werden die vorgestellten zusammen mit weiteren Aspekten entsprechend ihrer Relevanz der definierten Benutzergruppen bewertet.



Dabei bedeuten:



V – VR-/Echtzeitanwender

L – Lichttechniker

A – Animationen

C – CAD-Anwender

M – Augmented/Mixed Reality Anwender

Skala: sehr wichtig     wünschenswert     irrelevant 
































































































Thematik	Relevanz				
	V	L	A	C	M
Artefakt: Schatten- und Lichtlöcher					
Artefakt: Gezackte Schattenkanten					
Artefakt: Mach-Bänder					
Artefakt: Nähte					
Artefakt: T-Vertices					
Artefakt: Fehlende Schattenkanten					
Artefakt: temporal aliasing					
Visibility: Support plane split auf der Senderseite					
Formfaktorberechnung (Genauigkeit)					
Formfaktorsingularitäten					
Refinement: Physikalische Wiedergabequalität)					
Refinement: Unnötige Unterteilungen					
Visibility: Ungenaue Sichtbarkeitsberechnung					
Degenerierte Geometrien					
Normalenkonsistenz					
Freiformflächen					
Speicheranforderungen					
Parametervielfalt					
Unterstützte Lichtquellen					

Tabelle 2 : Relevanz einzelner Aspekte des Radiositysystems für die Benutzergruppen.

## 2.3 Photometrische und colorimetrische Konsistenz

Um verlässliche, reproduzierbare Aussagen über wahrgenommene Helligkeiten, Farbtreue und Realismus eines ganz oder teilweise simulierten Szenarios treffen zu können, muß an allen Stellen der Simulationspipeline photometrische und colorimetrische Konsistenz sichergestellt werden.

Nach einer kurzen Einführung in die Wahrnehmungsfähigkeit des menschlichen Sehapparates werden die wesentlichen Stationen hierzu untersucht und analysiert, welche Schritte vorgenommen werden müssen, um diese Konsistenz zu gewährleisten.

Zunächst muß sicherlich das Radiositysystem in der Lage sein, mit kalibrierten photometrischen Daten umzugehen. Diese kalibrierten Daten müssen jedoch auch von Anfang an in die Simulation einfließen, d.h. es gelten bereits für die Eingangsdaten bestimmte Voraussetzungen. Und schließlich, als nicht weniger wichtiger Schritt, muß das jeweilige Display kalibriert werden und das Radiositysystem auf das Display kalibrierte Daten liefern.

Nur bei lückenloser Sicherstellung all dieser einzelnen Kalibrierungsschritte ist eine photometrisch und colorimetrisch konsistente Simulation und Bildwiedergabe möglich.

### 2.3.1 Physiologische Grundlagen

Das menschliche Auge ist nicht für alle Wellenlängen des sichtbaren Lichts gleich empfindlich (vgl. z.B. [H89][G95]). Die **spektrale Empfindlichkeitskurve des Auges** (*luminous efficiency function* oder  $V(\lambda)$ -Kurve) beschreibt die Empfindlichkeit des Auges für Licht unterschiedlicher Wellenlängen und wird unter anderem dazu verwendet, zwischen physikalischen Strahlungsgrößen wie der Strahldichte (*radiance*, Einheit  $\text{W}/\text{m}^2 \cdot \text{sr}^{-1}$ ), und wahrnehmungsbasierten Größen wie Leuchtdichte (*luminance*, Einheit  $\text{cd}/\text{m}^2$ ) umzurechnen.

Durch die Adaptionfähigkeit des Auges ändert sich dessen Empfindlichkeit in Abhängigkeit der mittleren Beleuchtungsverhältnisse. Bei sehr niedriger Helligkeit, dem Nachtsehen (*scotopic illumination*,  $L < 10^{-5} \text{ cd}/\text{m}^2$ ), liefern Stäbchen die einzige Helligkeitsinformation, welche monochromatisch ist. Bei höheren Helligkeiten, also z.B. Tageslicht oder auch Raumbeleuchtung (*photopic illumination*,  $L > 10 \text{ cd}/\text{m}^2$ ) sind die Stäbchen übersättigt und liefern keine Informationen. Statt dessen sind die Zäpfchen aktiv, die aufgrund ihrer drei unterschiedlichen Typen erst Farbsehen ermöglichen (z.B. [WS82][SB94]).

In dem dazwischenliegenden Bereich, dem Dämmerungssehen (*mesopic illumination*), wird die Farbwahrnehmung zum scotopischen Bereich hin immer schwächer. Aufgrund der biophysikalischen Eigenschaften der Zäpfchen und Stäbchen verschiebt sich der Bereich maximaler Empfindlichkeit in Abhängigkeit von der Helligkeit. Die sich daraus ergebenden drei Varianten der spektralen Empfindlichkeitskurve sind in Abb. 4 angegeben.

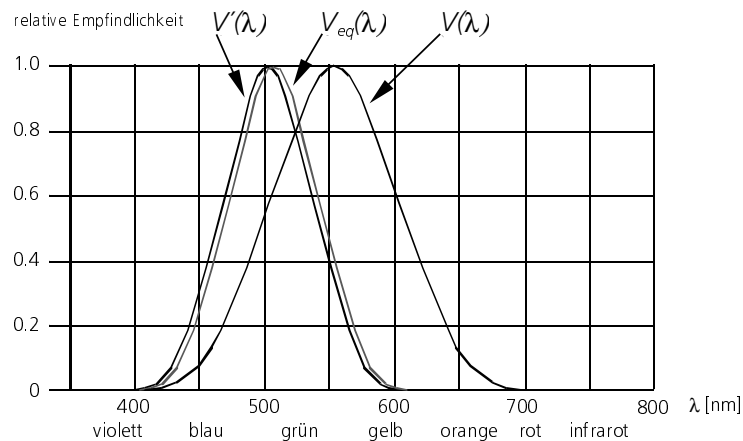


Abb. 4 : Normalisierte spektrale Empfindlichkeitskurven.  $V(\lambda)$ : photopisches Sehen ( $L > 10 \text{ cd/m}^2$ , Tageslicht).  $V_{eq}(\lambda)$ : mesopisches Sehen (Dämmerung).  $V'(\lambda)$ : scotopisches Sehen ( $L < 10^{-5} \text{ cd/m}^2$ , Nachtsicht).

Das Auge, bzw. die anschließende Bildverarbeitung im Gehirn nehmen weniger Helligkeiten direkt wahr, als vielmehr Helligkeitsdifferenzen: Was das Auge als „Weiß“ empfindet, ergibt sich ausschließlich aus dem Zusammenhang. So wird es eine weiße Wand solange als „Weiß“ akzeptieren, bis z.B. von einem Projektor ein Bild darauf geworfen wird. Jetzt ist plötzlich dasselbe „Weiß“ (nämlich die unbeleuchteten Bildteile) ein „Schwarz“, obwohl sich die Absoluthelligkeit nicht verändert hat – stattdessen wird nun der vom Projektor beleuchtete Bereich als „Weiß“ definiert, und alles andere in Relation dazu.

Auch die Farbwahrnehmung basiert bei konstanter Beleuchtung auf den *Reflektivitäten*, und nicht den tatsächlichen physikalischen Farben: ein weißes und ein blaues Objekt wird unter Tageslichtbedingungen dem Auge genauso weiß und blau erscheinen wie unter Glühlampenbeleuchtung – obwohl sich im letzteren Fall die physikalischen Farben tatsächlich von Weiß nach Tieforange, und von Blau nach Dunkelgrau verändert haben. Das Objekt mit der höchsten Reflektivität (hier das „weiße“) dient dem Auge hierbei als Referenzpunkt, und alle anderen Farbeindrücke werden auf diese Referenz bezogen, und nicht auf die tatsächlichen physikalischen Gegebenheiten. Überschreitet ein Objekt bei gleicher Beleuchtung in Teilen seines Spektrums die Reflektivität dieses Referenzobjekts, so wird es als unnatürlich, fluoreszierend wahrgenommen.

Diese Kompensation der Helligkeits- und Farbwahrnehmung im Gehirn ist allerdings nicht perfekt – so wird man z.B. sicher nicht denselben Bildeindruck von einem Raum haben, wenn er einmal von einer Kerze, und einmal von einer 1000 Watt Xenonlampe beleuchtet wird. Eine photometrisch und colorimetrisch korrekte Darstellung simulierter oder anderer kalibrierter Daten ist daher für hohe visuelle Qualität trotzdem unentbehrlich.

### 2.3.2 Interne Konsistenz

Sinn der photometrischen Konsistenz ist die korrekte Darstellung der Ergebnisse der Radiositysimulation auf einem beliebigen Display.

Hat die Simulation für ein Patch eine Leuchtdichte  $L$  ergeben, so soll die auf dem Display dargestellte Farbe genau diese Helligkeit haben (sofern das Display diese Leuchtdichte darstellen kann).

Aus diesem Grund müssen vor der Simulation für alle Emitter (Flächen- und Punktlichtquellen) photometrisch konsistente Farben gesetzt werden, die aus der vom Benutzer vorgegebene Lichtfarbe (RGB) und der Leuchtdichte bzw. dem Lichtstrom berechnet werden. Auch die Eingabe eines Spektrums ist möglich.

Weiterhin wird eine Kalibrierungsmatrix benötigt, die aus den Primärvalenzen des Displays ermittelt wird.

Es ist festzuhalten, daß eine einzelne RGB-Farbe keinerlei photometrische Aussagekraft besitzt. Erst durch Kalibrierung eines bestimmten RGB-Raums auf ein zugehöriges Ausgabemedium kann zwischen RGB-Farben und photometrisch sinnvollen Einheiten wie XYZ-Farben umgerechnet werden.

Intern wird in einem beliebigen, aber physikalisch definierten und quantifizierbaren Farbraum gerechnet (z.B. im RGB Farbraum eines genormten Monitors (NTSC)), für die Darstellung wird in den Farbraum des Displays konvertiert (NTSC-RGB $\rightarrow$ XYZ $\rightarrow$ Display-RGB).

Da Radiosity und Leuchtdichte über die konstante Beziehung  $B = \pi L$  im Prinzip gleichwertig sind, sind diese Einheiten austauschbar. Die Verwendung der Leuchtdichte als interne Einheit im Radiositysystem vereinfacht jedoch die Umrechnung für lichttechnische Anwendungen.

Bei der Umrechnung zwischen unterschiedlichen Farbräumen kann es auftreten, daß manche Farben in einigen dieser Farbräume nicht darstellbar sind, was sich durch negative Einträge in den Umrechnungs- und Kalibrierungsmatrizen bemerkbar macht. Dies ist mathematisch kein prinzipielles Problem – die Rech-

nung bleibt weiterhin korrekt. Jedoch muß das Radiositysystem darauf „vorbereitet“ sein, d.h. es muß mit negativen Farbwerten umgehen können.

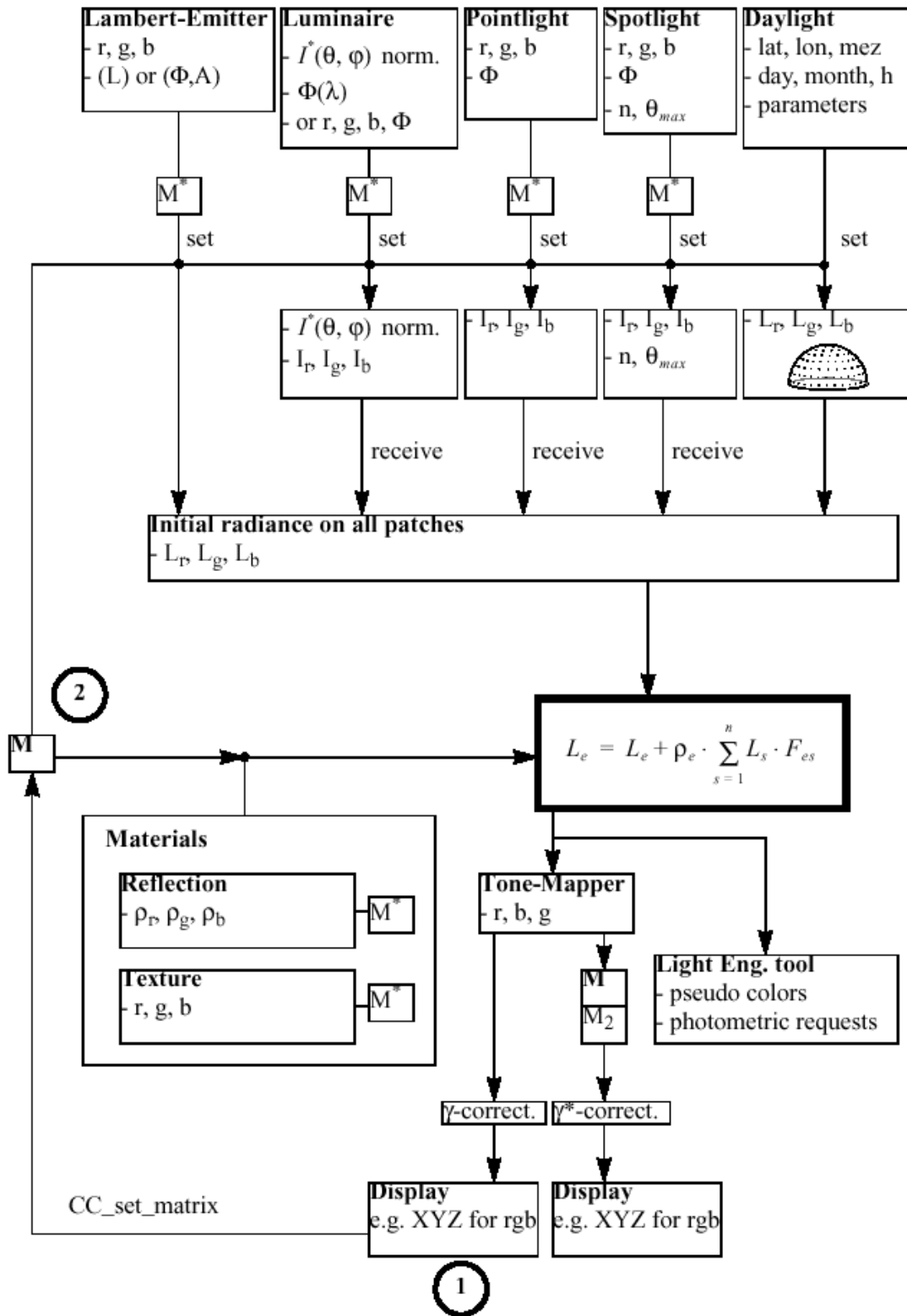


Abb. 5: Framework zur photometrischen und colorimetrischen Konsistenz

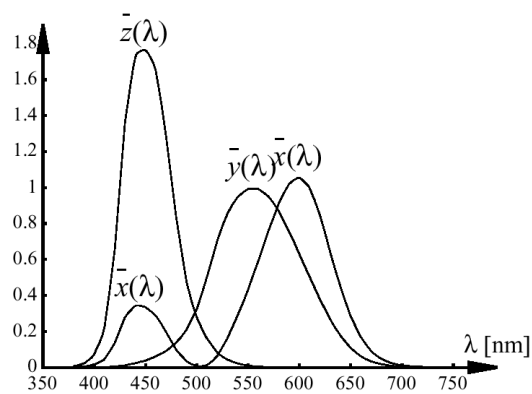


Abb. 6 : CIE Normspektralkurven

Im Folgenden dient der Faktor  $K_m = 683 \text{ lm/W}$  zur Umrechnung zwischen strahlungsphysikalischen und lichttechnischen Größen (vgl. [M96]).

### XYZ und xyL

Farbbeschreibungen mit photometrischer und colorimetrischer Aussagekraft stellt das CIE XYZ-System zur Verfügung. Eine Variante davon ist das xyL-System, bei dem Farb- und Helligkeitsinformationen getrennt sind und direkt abgelesen werden können (Farbort  $xy$ , Leuchtdichte  $L$ ).

Konvertierung zwischen xyL und XYZ:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$L = K_m \cdot Y$$

bzw.

$$Y = \frac{L}{K_m}$$

$$X = \frac{x}{y} \cdot Y$$

$$Z = \frac{1 - x - y}{y} \cdot Y$$

## Spektrum

Ein gegebenes Spektrum kann in eine XYZ Farbe konvertiert werden:

$$X = \int_{380}^{780} L(\lambda) \cdot \bar{x}(\lambda) d\lambda$$

$$Y = \int_{380}^{780} L(\lambda) \cdot \bar{y}(\lambda) d\lambda$$

$$Z = \int_{380}^{780} L(\lambda) \cdot \bar{z}(\lambda) d\lambda$$

Dabei ist  $L(\lambda)$  die spektrale Strahldichte und  $\bar{x}$ ,  $\bar{y}$ ,  $\bar{z}$  die CIEXYZ Kurven.

Als Näherung für  $n$  äquidistante Werte  $\lambda_i$  ( $i=0..n-1$ ) kann verwendet werden:

$$f(\lambda) = L(\lambda) \cdot \bar{x}(\lambda)$$

$$X = \frac{\lambda_{\max} - \lambda_{\min}}{n-1} \left( \frac{1}{2} L(\lambda_{\min}) \bar{x}(\lambda_{\min}) + \sum_{i=1}^{n-2} L(\lambda_i) \cdot \bar{x}(\lambda_i) + \frac{1}{2} L(\lambda_{\max}) \bar{x}(\lambda_{\max}) \right)$$

Für nicht-äquidistante Abstände gilt:

$$X = \frac{1}{2} \sum_{i=0}^{n-2} (f(\lambda_{i+1}) + f(\lambda_i)) \cdot (\lambda_{i+1} - \lambda_i)$$

(Y und Z berechnen sich jeweils analog).

## Farbtemperatur

Für eine Farbtemperatur kann zunächst mit der Planck-Formel [WS82][G99] ein Spektrum aufgebaut werden, das dann mit der obigen Formel in eine XYZ Farbe konvertiert wird. Über die Planck-Formel zur Beschreibung der Schwarzkörperstrahlung lässt sich die spektrale Flußdichte in  $W \cdot m^{-3}$  entweder in Abhängigkeit von der Frequenz  $\nu$  oder der Wellenlänge  $\lambda$  ausdrücken:

$$M_{e\nu} = \frac{c}{4} \cdot \frac{8\pi h \nu^3}{c^3 (e^{\frac{h\nu}{kT}} - 1)} = \frac{2\pi h \nu^3}{c^2 (e^{\frac{h\nu}{kT}} - 1)}$$

$$M_{e\lambda} = \frac{c}{4} \cdot \frac{8\pi h c}{\lambda^5 (e^{\frac{hc}{\lambda kT}} - 1)} = \frac{2\pi h c^2}{\lambda^5 (e^{\frac{hc}{\lambda kT}} - 1)}$$

Um ein Wellenlängenspektrum zu erzeugen, wird die Variante  $M_{e\lambda}$  verwendet.

### 2.3.3 Konsistenz in den Eingabedaten

Damit eine durchgehende Konsistenz gewährleistet ist, müssen insbesondere bereits für die Eingangsdaten konsistente Werte vorliegen. Dies gilt sowohl für die Farb- und Helligkeitsbeschreibungen von Lichtquellen, als auch für die Reflexionseigenschaften von Materialien – eine Angabe in einem unkalibrierten RGB-Farbraum genügt hier nicht.

Eine besondere Problematik stellt sich hier bei Texturen, die meist als Fotografien mit unkalibrierten Kameras, oder gar eingescannten Materialproben unter unbekanntem Beleuchtungsbedingungen als RGB-Bilder vorliegen. Konsistente Texturen müssen im optimalen Fall als *high dynamic range images* einer kalibrierten Kamera unter genau beschriebenen Beleuchtungsbedingungen (d.h. als *Reflexionsspektrum*) gegeben sein.

Eine konsistente Farbe kann auf unterschiedliche Weisen angegeben werden, wobei zwischen Emissionswerten (Lichtquellen) und Reflexionswerten (Materialien) unterschieden werden muß:

- RGB-Werte für einen Standard Monitor (NTSC)
- RGB-Werte mit zugehöriger Kalibrierungsmatrix  $M^*$
- Spektrum (Reflexion und Emission)
- Farbtemperatur (nur Emission)
- Wellenlänge und Sättigung (nur Emission)
- Farbbort  $xy$  (Reflexion und Emission)
- RAL Nummer (nur Reflexion)
- ...

Zusätzlich muß jedoch im Falle von Flächenlichtquellen die Leuchtdichte, im Falle von punktförmigen Lichtquellen der Lichtstrom gegeben sein.



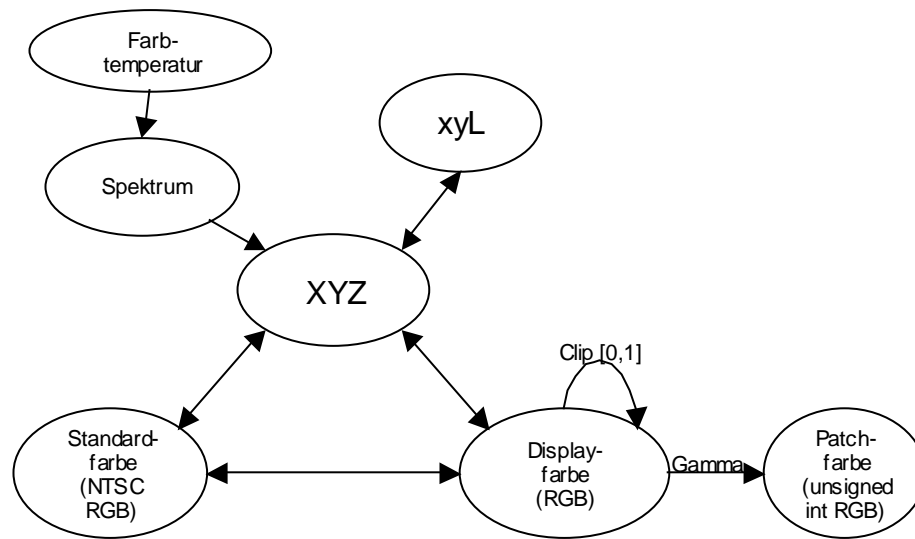


Abb. 7 : Konvertierung zwischen Farbsystemen.

### 2.3.4 Konsistenz in den Ausgabedaten

Für eine korrekte Farbdarstellung werden die Primärvalenzen des aktuellen Displays (xyL für Rot, Grün, Blau) benötigt. Intern wird in einem genormten Farbraum (z.B. eines genormten Monitors, hier NTSC) gerechnet, für die Darstellung wird in den Farbraum des aktuellen Displays konvertiert.

Ist die Farbe außerhalb des darstellbaren Bereichs, so muß entweder an [0,1] geclippt werden oder die ähnlichste Farbe dargestellt werden (Gamut Mapping).

Nach der Konvertierung müssen die Farben eine Gammakorrektur durchlaufen, d.h. aus (R,G,B) wird  $(R^{\gamma_r}, G^{\gamma_g}, B^{\gamma_b})$ .

Das Display muß über seine Primärvalenzen im xyL-Farbraum, sowie der Gammakurven dem System bekannt gemacht werden. Die Gammakurven für R, G und B sind bei realen Displays oft leicht unterschiedlich (d.h. unterschiedliche Grauwerte besitzen leicht unterschiedliche Farborte), und nur selten lassen sich die Kurven durch einen einfachen Exponenten beschreiben. Der Einsatz von drei *look-up tables* ist hier sinnvoll.

Insbesondere für Helligkeiten, die sich außerhalb des vom Display darstellbaren Bereichs befinden, aber auch für nicht darstellbare Farben muß das System z.B. mit Hilfe von Fehlfarben diejenigen Bildbereiche markieren können, aus denen keine photometrisch verlässlichen Rückschlüsse mehr gezogen werden können.

Zur Displaykalibrierung, Gleichförmigkeit über die Leinwand, sowie Problematiken im Zusammenhang mit Texturen siehe insbesondere Kapitel 3.2.5.

Sollen die Daten außerhalb des Radiositysystems visualisiert werden, ist es wichtig, daß sie in einem displayunabhängigen Format gespeichert werden. Das Visualisierungssystem muß dann diese norm-kalibrierten Daten, sowie Kalibrierungsdaten des entsprechenden Displays in displayspezifische Simulationsdaten konvertieren.

## 2.4 Kunst- und Tageslicht

Klassische Radiositysysteme kennen nur Lambert'sche Flächenlichtquellen.

Insbesondere für die Benutzergruppe des Lichttechnikers, aber auch für eine Augmented Reality Umgebung, sind jedoch Lichtquellen mit physikalischen Leuchtenbeschreibungen unerlässlich.

Eine detaillierte Beschreibung, punktförmige Lichtquellen in ein Radiositysystem zu integrieren, findet sich in [M96]. Für die Integration in ein hierarchisches Radiositysystem ist jedoch ebenfalls eine saubere, konsistente Integration möglich.

Lichtquellen können keine Energie empfangen und besitzen somit auch keine Empfängerlinks. Entsprechende Datenstrukturen sind also prinzipiell nicht erforderlich, was die Speicherkomplexität verringert.

Für jede Lichtquelle wird ein Skalierungsfaktor  $s$  zur Umrechnung von kalibrierter Originalfarbe und Lichtstrom in die photometrisch konsistente Farbe  $(L_R, L_G, L_B) = s \cdot (R, G, B)$  benötigt. Dieser Faktor wird dabei so gewählt, daß die Rückrechnung in lichttechnische Größen vereinfacht wird. Man beachte, daß  $(L_R, L_G, L_B)$  keine direkte Aussagekraft besitzt – es repräsentiert weder spektrale Samples der Radiancefunktion, noch kann es in der Einheit  $\text{W}/\text{m}^2 \cdot \text{sr}^{-1}$  ausgedrückt werden. Es beschreibt jedoch eine kalibrierte Farbe und enthält gleichzeitig einen zugehörigen photometrischen Wert.

Die auf einem Patch empfangene Radiosity  $L_e$  kann dann über die unten jeweils angegebenen Formeln ermittelt werden.

Die Elemente  $m_{ij}$  beziehen sich auf die in 2.3.2 ermittelte Farbkonvertierungsmatrix.

### Zurückrechnen lichttechnischer Größen

Für ein gegebenes Patch können aus der photometrisch konsistenten Farbe  $(L_R, L_G, L_B)$  folgende Größen ermittelt werden:

- Leuchtdichte (*luminance*)

$$L = K_m \cdot (L_R \cdot m_{21} + L_G \cdot m_{22} + L_B \cdot m_{23})$$

- Spezifische Ausstrahlung (*radiosity*)

$$B = \pi \cdot L$$

- Lichtstrom (*luminous flux*)

$$\Phi = L \cdot A \cdot \pi$$

- Beleuchtungsstärke (*illuminance*)

$$E = \frac{\pi \cdot L}{\rho}$$

### 2.4.1 Flächenlichtquellen

Flächige Lambertstrahler [L1760] sind die Standardlichtquellen klassischer Radiositysysteme. Ihre photometrischen Beziehungen sind:

$$L = \frac{\Phi}{A \cdot \pi}, \text{ wenn } L \text{ nicht gegeben ist.}$$

$$s = \frac{L}{K_m (m_{21}R + m_{22}G + m_{23}B)}$$

$$L_e = \rho \cdot F_{es} \cdot L_s$$

Ein Spezialfall einer Flächenlichtquelle besteht z.B. aus einer texturierten Lichtquelle, bei der die Pixelwerte über die Fläche variierende Emissionswerte beschreiben.

### 2.4.2 Punktförmige Lichtquellen

Diese Lichtquellen besitzen keine räumliche Ausdehnung.

#### Punktlichtquelle

Diese Lichtquelle ist durch gleichförmige Lichtausbreitung in jede Richtung beschrieben.

$$s = \frac{\Phi}{4\pi K_m (m_{21}R + m_{22}G + m_{23}B)}$$

$$L_e = \rho \cdot \frac{\cos\theta_e}{\pi \cdot d^2} \cdot L_s$$

## Spotlight

Eine Erweiterung der Punktlichtquelle. Das Spotlight besitzt eine vorgegebene Hauptausbreitungsrichtung, einen maximalen Öffnungswinkel und eine  $\cos^n$ -Verteilungsfunktion.

$$s = \frac{\Phi \cdot (n+1)}{2\pi(1 - \cos^{n+1}\theta_{\max})K_m (m_{21}R + m_{22}G + m_{23}B)}$$

$$L_e = \rho \cdot \frac{\cos^n\theta_s \cdot \cos\theta_e}{\pi \cdot d^2} \cdot L_s, 0 < \theta_s < \theta_{\max}$$

## LVK

LVKs sind Lichtquellen mit gegebener Abstrahlcharakteristik („LichtStärkeVerteilungskurve“). Sie werden in Kapitel 2.4.3 eingehender beschrieben.

## Sonne

Das direkte Sonnenlicht ist, im Gegensatz zum diffusen Tageslicht (vgl. Kapitel 2.4.4), im Prinzip eine Punktlichtquelle in genügender Entfernung. Sinnvoller ist die Definition als gerichtete Lichtquelle, da hierdurch weitere Vereinfachungen im Kontext der Formfaktorberechnung ermöglicht werden.

Ein erweitertes Verfahren kann diese Lichtquelle auch als Raumwinkel mit 30° Durchmesser annehmen, um realistische Schatten zu erzeugen. Dies ist jedoch im Allgemeinen nur für hochqualitative Einzelbilder (z.B. mit Hilfe des *final gathering*) sinnvoll, da die feinen Unterschiede zwischen den echten Schlagschatten einer Punktlichtquelle und der verhältnismäßig kleine Halbschattenzonen generierenden Raumwinkel-Lichtquelle durch das (tiefenbegrenzte) Meshing ohnehin verwischt werden.

Die jeweilige Richtung des Sonnenlichts wird aus den geographischen Informationen ermittelt, die ebenfalls in Kapitel 2.4.4 beschrieben sind.

### 2.4.3 Goniometrische Lichtquellen

Die Abstrahlcharakteristik dieser Lichtquelle ist als LVK gegeben, die als Datei vorliegt und auf einen Gesamtlichtstrom von 1000 lm normiert ist. Bei LVK-Lichtquellen handelt es sich üblicherweise um reale, mit einem Goniometer vermessene Lichtquellen.

Es ist zu beachten, daß die vermessenen LVKs üblicherweise nur ab einem gewissen Abstand von der Lichtquelle, im sog. Fernfeld gültig sind und im Nahfeld, also unterhalb des *photometrischen Grenzabstandes* [SC00], aufgrund der Lichtquellengeometrie der realen Leuchte nur bedingt eingesetzt werden können.

$$s = \frac{\Phi}{K_m 1000(m_{21}R + m_{22}G + m_{23}B)}$$

$$L_e = \rho \cdot \frac{\cos \theta_e}{\pi \cdot d^2} \cdot L_s(\theta, \varphi)$$

#### Punktförmige LVKs

Diese gebräuchlichste Variante der goniometrischen Lichtquelle besitzt keine Ausdehnung und ist daher punktförmig.

#### Linien- und Flächen-LVKs

Linienförmige oder flächige Lichtquellen mit LVK-Beschreibungen besitzen kein direktes physikalisches Äquivalent, sondern dienen als Erweiterung der punktförmigen LVKs und werden eingesetzt um z.B. die Nahfeld-Einschränkungen etwas abzumildern.

Hierbei besitzt die Lichtquelle eine linienförmige oder flächige Geometrie, wobei für jede Stelle der Geometrie dieselbe LVK gilt. Umgekehrt läßt sich so also für einen Raumwinkel die LVK sampeln, wodurch realistischere, weiche Schatten erzeugt werden.

### 2.4.4 Diffuses Tageslicht

Das diffuse Tageslicht aus der gesamten Himmelshalbkugel, das in klassischen Radiositysystemen – falls überhaupt unterstützt – als Geometrie gegeben ist [MKGS95][DSSD97], jedoch sinnvoller durch Raumwinkel und Pseudogeometrien definiert wird.

Die Lichtverteilung ergibt sich durch Himmelsmodelle oder Datenfiles aus Vermessungen und ist von geographischem Ort und Tageszeit, sowie z.B. der Bedeckungssituation abhängig.

Diese Lichtquelle erfordert die Festlegung einer Nord- und einer Zenitrichtung im Weltkoordinatensystem.

## Skylights

Dies ist eine rechnerisch sehr effiziente Variante der internen Repräsentation der diffusen Tageslichtquelle und kann für Innenraumausleuchtungen eingesetzt werden.

Hierbei werden die Szenenbereiche, durch die normalerweise der freie Himmel sichtbar ist (also im Allgemeinen die Fenster) durch flächige goniometrische Lichtquellen mit speziellen Eigenschaften ersetzt.

Die generelle Idee bei dieser Lichtquelle ist, daß in einem Raum mit beispielsweise nur einem Fenster keine Sichtbarkeitsbestimmung mit der gesamten Himmelshalbkugel durchgeführt werden muß, sondern direkt die (vom Benutzer als „Fenster“ vordefinierte) Lichtquelle befragt wird. Die „Abstrahlcharakteristik“ ergibt sich wie beim normalen diffusen Tageslicht entsprechend dem Himmelsmodell und der Richtung zum Empfänger.

## 2.5 Fehlerschranken

Die Fehler in einer Radiosityberechnung z.B. zwischen einem bestimmten Sender und einem Empfänger abzuschätzen und verlässliche obere und untere Schranken anzugeben sind die Grundlagen, die es einem ermöglichen, die Radiositysimulation entsprechend zu steuern. Besonders für die verbreiteten Flächenlichtquellen existieren entsprechende Methoden, solche Schranken zu bestimmen [LSG94][HS95]. Jedoch fehlen sie für andere Lichtquellen, bzw. für ein Gesamtkonzept mit verschiedenen Lichtquellentypen und Clustern.

Es ist wichtig zu erkennen, daß es unterschiedliche Arten von Fehlern gibt, die unterschiedlich behandelt werden müssen. Die tiefere Bedeutung dieser „Fehler“ ist stark anwendungsabhängig, da sie direkt an die Benutzeranforderungen und auch die Szenentypen gekoppelt ist. Es muß also anhand der Anforderungen des Benutzers festgelegt werden, welche Arten von Fehlerabschätzungen zur Anwendung kommen und das Refinementkriterium beeinflussen sollen.

Diese Entscheidungen sollten natürlich für den Benutzer eines Radiositysystems vollkommen transparent sein. Das System muß Fehlerabschätzungen automa-

tisch vornehmen und so die Refinementstrategie auf Basis weniger benutzerspezifischer Einstellungen weitgehendst automatisiert ablaufen lassen, was besonders wichtig ist für die Anwendbarkeit eines Radiositysystems im produktiven Einsatz, sowie für Benutzer, die nicht mit den Details des Radiosityprozesses vertraut sind.

### 2.5.1 Evolution des Refinementkriteriums

Das erste generelle Refinementkriterium für hierarchisches Radiosity war das *BF*-Kriterium [HSA91]. Hierbei wurde eine gegebene Sender/Empfänger-Konfiguration immer dann unterteilt, wenn die Beleuchtungsstärke (nämlich eben der Faktor zwischen eingehender Radiosity  $B$  und dem Formfaktor  $F$ ) über einer benutzerdefinierten Schranke lag.

Die Gründe für ein solches Kriterium lagen darin, daß z.B. der Fläche-zu-Fläche-Formfaktor durch einen differentiellen Punkt-zu-Scheibe-Formfaktor abgeschätzt wurde, was eine sehr grobe Beschreibung des echten Formfaktors bedeutete, sodass der Fehler zwischen echtem und abgeschätzten Formfaktor mit der Größe des Formfaktors zunahm. Zusätzlich wurde durch die Integration der Senderradiosity eine Unterteilung von dunklen oder gar unbeleuchteten Regionen vermieden.

Bereits diesem *BF*-Kriterium lag also eine (zugegebenermaßen sehr grobe) Fehlerabschätzung des Formfaktors – und damit der Energieübertragung – zugrunde.

Sobald genauere, analytische Formfaktorberechnungsverfahren zum Einsatz kamen, konnte man sich darauf konzentrieren, die *tatsächlichen* Gründe, die eine Unterteilung erforderlich machten, zu untersuchen. Zum Beispiel bedeutet nur weil die transportierte Energiemenge über einem bestimmten Wert liegt noch nicht, daß eine Unterteilung wirklich notwendig ist: Eine auf eine Wand gerichtete helle Lichtquelle wird hohe Intensitätswerte erzeugen. Ein reines *BF*-Kriterium würde diese Region sehr fein unterteilen, obwohl die tatsächliche Radiosityfunktion verhältnismäßig glatt ist und bereits durch ein grobes Mesh hinreichen genau repräsentiert werden würde. Des weiteren reagiert das menschliche Auge auf Helligkeitsschwankungen in hellen Bereichen deutlich unempfindlicher als auf solche in dunklen Bereichen [S98]. Je heller also eine Region ist, desto größer können auch die Absolut-Differenzen werden, bevor das Auge einen Unterschied wahrnimmt.

Ausgehend von der *constant radiosity assumption* erkennen Lischinski et al. [LSG94], daß man, anstatt nur den Absolutwert der transportierten Energie (al-

so die resultierende Intensität) für das Refinementkriterium zu verwenden, besser versuchen sollte, die Annahmen für stückweise konstante Funktionen direkt zu integrieren: Schranken für die Radiosityfunktion werden entwickelt um zu entscheiden, ob die Funktion konstant ist, oder ob sie zu stark von einer Konstanzannahme abweicht, in welchem Fall das Refinementkriterium eine Unterteilung forcieren wird.

Solche Ansätze können erweitert werden, indem man lineare Funktionen zur Repräsentation der Radiosity auf dem Empfänger verwendet. Dies ist genau dann wichtig, wenn bei einem Radiositysystem auf Basis finiter Elemente Hardware-Gouraud-Shading zur Ergebniswiedergabe verwendet wird, da Gouraud-Shading nichts weiter ist als eine vereinfachte Form der linearen Interpolation einer Funktion – in dem Fall der Radiosityfunktion. Wird die Radiosityfunktion auf Basis stückweise linearer Funktionen rekonstruiert, so ist eine Unterteilung nur dann erforderlich, wenn die Radiosityfunktion auf einem Patch nicht als linear angenommen werden kann. Ein Ansatz hierzu ist, Lischinskis Idee der Beschränkung der Radiosityfunktion auf lineare Schranken zu erweitern [PB95].

Auf der anderen Seite läßt sich Lischinskis Ansatz konstanter Schranken mit einem Gradienten-basierten Kriterium kombinieren [K97][S98]. Eckpunktradiositäten können verwendet werden, um zusätzliche Fehlerquellen bei der Extrapolation von Mittelpunktradiositäten auf Patcheckpunkte auszuschließen.

Refinementkriterien haben eine Evolution erfahren seit ihrem ersten Einsatz in frühen hierarchischen Radiositysystemen. Ein aus heutiger Sicht sinnvolles Kriterium wird immer auf Low-Level-Informationen des Energietransfers (insbesondere auf Clusterebene), sowie der voraussichtlichen Wahrnehmung durch den menschlichen Sehapparat basieren.

Leider verwenden auch heute noch viele kommerzielle Radiositysysteme ein einfaches *BF*-Kriterium.

## 2.5.2 Fehlerarten

Innerhalb einer Radiositysimulation gibt es fundamental verschiedene Fehlerarten, die den Zustand der Simulation beschreiben. Meist nimmt man diese Fehler bewußt in Kauf, um dadurch die Simulation in vernünftiger Zeit (oder überhaupt) durchführen zu können. Nur wenn die Größen dieser Fehler bekannt sind und angegeben werden können, besitzt das Ergebnis – im Rahmen dieser Fehler-schranken – immer noch eine gültige Aussagekraft.

Zunächst gibt es den **Fehler in der Energieübertragung** zwischen einem sendenden und einem empfangenden Patch. Die Menge der übertragenen Energie wird hierbei über den Formfaktor bestimmt. Da es jedoch zu aufwendig (und bis



vor ein paar Jahren sogar unmöglich [SH93]) wäre, für jede Energieinteraktion einen vollständigen Doppelintegral-Formfaktor zu berechnen, werden in klassischen Systemen meist Abschätzungen oder die erste Formfaktorvereinfachung verwendet, wobei das kleinere Element als infinitesimal angenommen wird. Auch an anderen Stellen können Fehler durch bestimmte Annahmen in die Simulation einfließen, wie z.B. die Radiosityverteilung auf den Patches, die als konstant angenommen wird, oder durch Raycasting inkorrekte Sichtbarkeitserkennung für partiell sichtbare Szenenbereiche. Insbesondere für eine Fläche-Fläche-Sichtbarkeit reicht es nicht aus, einen prozentualen Wert für die Sichtbarkeit anzugeben, da es von der Position auf dem Empfänger abhängt, wieviel und welche Bereiche des Senders sichtbar sind. Gibt man statt dessen einen für den ganzen Empfänger gültigen Bereich der partiellen Sichtbarkeit mit Hilfe von Fehlerschranken an, weiß man, daß man im Rahmen dieser Schranken korrekt ist.

Für Formfaktor, Senderradiosity-Verteilung und Sichtbarkeit können obere und untere Schranken ermittelt werden. Die Kombination dieser einzelnen Schranken liefert Fehlerschranken für die Energieübertragung, durch die man in der Lage ist, die Größe des Gesamtfehlers in der Simulation zu bestimmen.

Alleinige Kenntnis der Energieübertragung liefert jedoch nicht genug Informationen über das *wahrgenommene* Erscheinungsbild des Ergebnisses: Ein gewisser mathematischer Fehler an gewissen Stellen der Radiositysimulation bedeutet nicht gleichzeitig, daß der menschliche Sehapparat diesen Fehler auch, z.B. in Form eines Artefakts, wahrnehmen wird oder kann. Auf der anderen Seite ist es denkbar daß, obwohl der mathematische Fehler in manchen Bereichen klein sein mag, das menschliche Auge merkt, daß „etwas nicht stimmt“, daß Artefakte existieren, die nicht hinreichend durch Energieübertragungsschranken beschrieben werden.

Daraus läßt sich die zweite Art von Fehler ableiten, der **Fehler bezüglich der visuellen Qualität**.

Was der Benutzer letztendlich von einer ausgeleuchteten Szene wahrnimmt, ist die Beleuchtungsverteilung auf den Objekten, also die Radiosityfunktion auf den Empfängern.

Während mit Hilfe der Energieübertragungsschranken verlässliche *Werte* für die Empfängerradiosity ermittelt werden können, läßt sich durch Schranken für die Radiosityfunktion auf dem Empfänger die *Form* dieser Funktion verlässlich beschreiben. Der absolute Wert der Radiosityfunktion definiert sich über die Energiemenge, die übertragen wird, während sich die visuelle Qualität der Lichtverteilung über die Form der Radiosityfunktion, und wie gut sie während der Simulation angenähert wurde, definiert.

Der Unterschied zwischen diesen Auswertungsmethoden spiegelt sich in den unterschiedlichen Benutzerszenarien wider: Ein Benutzer, der hauptsächlich an physikalischer Korrektheit interessiert ist, wie z.B. ein Lichttechniker, legt nicht so viel Wert auf ein perfektes, optisches Endergebnis. Ein Benutzer, der jedoch eine Animationssequenz oder ein hochauflösendes Einzelbild berechnen will, benötigt eher eine hohe visuelle Qualität als 100%ig exakte Candelawerte: solange das Ergebnis „gut aussieht“, ist es nicht so wichtig, ob die berechnete Energie jetzt 5% zu hoch oder zu niedrig ist. Je nach Benutzergruppe kann der Radiosityalgorithmus seine Berechnungsstrategien so den expliziten Anforderungen der jeweiligen Gruppe anpassen.

Vom mathematischen Standpunkt aus betrachtet wird der Wert einer Funktion, ihre „Größe“, über die Funktion selbst beschrieben, während ihre Form über ihre Ableitung repräsentiert wird. Insbesondere kann die Annahme konstanter Sender- und Empfängerelemente nicht die notwendigen Informationen bereitstellen, um die Form der Radiosityfunktion zufriedenstellend zu rekonstruieren.

### 2.5.3 Basisfunktionen

Klassische Radiositysysteme basieren auf der *constant radiosity assumption*. Hierbei wird angenommen, daß die Radiosityverteilung auf allen sendenden und empfangenen Elementen konstant ist, was bestimmte vereinfachende Annahmen beim Refinementkriterium und der generellen Radiositysimulation gestattet.

Doch die Annahme durchgehend konstanter Radiosity ist stellenweise zu restriktiv, und es gehen Vorteile verloren, die man sich durch andere, z.B. lineare Annahmen zunutze machen könnte.

Da Cluster ohnehin eine approximierte Repräsentation für den Energietransfer sind und die Form der Radiosityfunktionen auf einem solchen Element irrelevant ist, kann die *constant radiosity assumption* durchaus für Cluster angewandt werden.

Für Patch-Elemente kann jedoch zwischen folgenden Situationen unterschieden werden:

#### **Konstanter Sender**

Ein konstanter Sender bedeutet, daß die Radiosityverteilung über den gesamten Sender hinweg als konstant angenommen wird. Sobald diese Annahme nicht mehr eingehalten werden kann, würde das Refinementkriterium eine Unterteilung des Senders forcieren. Ein Sender mit konstanter Radiosity vereinfacht bestimmte Berechnungen im Zusammenhang mit dem Sender –insbesondere die Formfaktorberechnung.

## Linearer Sender

Lineare Basisfunktionen auf dem Sender bilden eine Erweiterung zu konstanten Sendern und erfordern entsprechend angepasste Formfaktor- und Gradientenalgorithmen. Analytische Berechnungsmethoden hierfür sind bekannt [HS95], aber deutlich aufwendiger als für den konstanten Fall.

Ein anderer Punkt sind nicht-konstante Sender, wie z.B. texturierte Sender, oder Lichtquellen mit LVK-Beschreibungen, die auch punktförmig sein können und dafür eine richtungsabhängige Komponente beinhalten.

## Konstanter Empfänger

Dies ist die einfachste Variante, eine Radiosityfunktion auf dem Empfänger darzustellen. Ein einziger Radiositywert wird berechnet und für das gesamte Patch als gültig angenommen. Wieder liegt es am Refinementkriterium, z.B. aufgrund von (konstanten) Fehlerschranken nötigenfalls eine Unterteilung herbeizuführen.

Würde man den konstanten Radiositywert direkt darstellen, würde dies in einem *flatshading* der einzelnen Patches resultieren. Um dieses Artefakt zu vermeiden, extrapoliert man den einen Radiositywert, ggf. zusammen mit den Werten benachbarter Patches, auf die Patcheckpunkte und verwendet schnelles Gouraud-Shading für die Darstellung. Die eigentlich nicht-konstante Radiosityfunktion wird also durch einen konstanten Wert pro Patch angenähert, der dann wieder linear interpoliert wird, damit das Ergebnis ansprechender aussieht. Diese Vorgehensweise scheint etwas widersprüchlich, was in Kapitel 2.5.4 noch einmal aufgegriffen wird. Jedoch beinhaltet sie den Vorteil vereinfachter Radiosityberechnung.

## Linearer Empfänger

Bei linearen Basisfunktionen auf dem Empfänger wird der Verlauf der Radiosityfunktion als linear angenommen. Anstatt dies über eine kompliziertere Formfaktorberechnung zu bewerkstelligen (die in der Tat sogar unverändert bleibt, wenn der Formfaktor von der Senderfläche zu einem Empfängerpunkt berechnet wird), wird die Radiosityfunktion nicht an einem einzelnen Empfängerpunkt berechnet, sondern an mehreren (d.h., an den Eckpunkten statt nur am Mittelpunkt). Der Vorteil ist, daß die so erhaltenen Radiositywerte direkt und insbesondere ohne zusätzliche Interpolation mit Gouraud-Shading dargestellt werden können, was die visuelle Qualität steigert.

Refinementkriterien für eine *linear radiosity assumption* sind auf zusätzliche, lineare Fehlerschranken [PB95] angewiesen, sowie ggf. einem Gradienten [HS95], um vernünftige Ergebnisse zu liefern [K97][S98].

Die Verwendung von linearen Basisfunktionen auf der Empfängerseite verhindert Unterteilungen dort, wo die Radiosityfunktion zwar nicht-konstant, aber linear steigt oder abfällt, was die Meshkomplexität verringert.

Im Prinzip ist die Verwendung von konstanten oder linearen Basisfunktionen für Sender oder Empfänger unabhängig voneinander, sodass jede Kombination möglich ist.

### *higher order basis functions*

Es ist denkbar, quadratische oder Basisfunktionen noch höherer Ordnung auf Sender- und/oder Empfängerseite einzusetzen, um z.B. die Meshkomplexität noch weiter zu reduzieren. Es sprechen jedoch mehrere wichtige Faktoren gegen eine solche Vorgehensweise:

- Der Berechnungs- und insbesondere Speicheraufwand steigt rapide mit jedem zusätzlichen Grad der Funktion an [WH97], sodass es letztendlich teurer wird als ein komplexeres Mesh basierend auf niedrigeren Basisfunktionen. Versuche, Superlineare Basisfunktionen für Radiosity einzusetzen ergaben Probleme *Gibbs ringing artefacts* zu vermeiden [Z93].
- Die Berechnung von verlässlichen Formfaktor- und anderen Fehlerschranken für höherwertige Basisfunktionen ist extrem aufwendig oder nicht zufriedenstellend lösbar, die Komplexität des Refinementkriteriums steigt gleichfalls mit der Ordnung der Funktionen an.
- Lineare Funktionen sind sehr handlich und erlauben die Behandlung der meisten Probleme, die sich durch eingeschränkte Funktionsrepräsentation ergeben.
- Für schnelle Darstellung (also insbesondere für die Benutzergruppe der VR-User) wird hardwareunterstütztes Gouraud-Shading verwendet, das auf linearer Interpolation von Farbwerten basiert. Durch Funktionen höherer Ordnung zu Verfügung gestellte Informationen würden einfach ignoriert.
- Und schließlich ist die menschliche Wahrnehmung nicht in der Lage (oder zumindest diesbezüglich sehr unempfindlich), Sprünge (d.h. Diskontinuitäten) in der zweiten oder in höheren Ableitungen von Intensitätsfunktionen wahrzunehmen – die tatsächliche Verbesserung der wahrgenommenen Qualität ist daher eher eingeschränkt.

### 2.5.4 Implikationen durch einen linearen Empfänger

Für konstante Sender und konstante Empfänger sind die Radiosityalgorithmen und Berechnungsstrategien wohlbekannt und verbreitet. Dieses Kapitel wird sich mit den Änderungen befassen, die sich für die Verwendung von linearen Empfängern ergeben.

#### Mittelpunkt- und Eckpunktradiosities

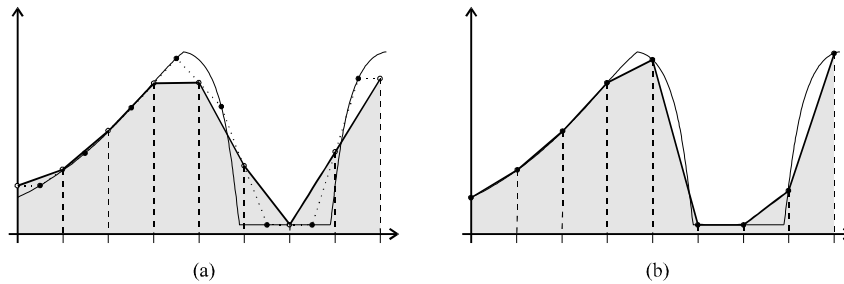
In Abhängigkeit von der geforderten visuellen Qualität (oder der geforderten Berechnungsgeschwindigkeit) wird entweder ein Mittelpunkt- oder ein Eckpunktradiosityalgorithmus gewählt. Radiosities an Mittelpunkten einzusammeln – wie beim ursprünglichen hierarchischen Radiosityalgorithmus – ist schneller, aber weniger exakt, während der Berechnungsaufwand (und die Speicheranforderungen) bei einem Einsammeln auf Patcheckpunkten höher sind, jedoch die visuelle und physikalische Qualität zunimmt.

Die Entscheidung zwischen Mittelpunkt- und Eckpunktverfahren kann pro Objekt (automatisch) entschieden werden, wobei das Radiositysystem natürlich auf beide Varianten – Fläche-zu-Fläche und Fläche-zu-Punkt – angepaßte Verfahren zur Abschätzung von Fehlerschranken bereitstellen muß.

Die übliche Vorgehensweise eine Finite-Element-Radiositysimulation durchzuführen besteht darin, die Radiosity auf den Mittelpunkten jedes Empfängerpatches zu bestimmen und sie für das gesamte Patch gelten zu lassen (sie also als konstant anzunehmen).

Die Algorithmen für Mittelpunktradiosities sind wohlbekannt, der Formfaktor (oder eine entsprechende Abschätzung davon) muß von (Sender-)Fläche zu (Empfänger-)Fläche berechnet werden, das Gleiche gilt für die Sichtbarkeit.

Vor der Darstellung werden die Mittelpunktwerte auf die Patcheckpunkte extrapoliert, um Gouraud-Shading zu ermöglichen. Dies bedeutet, daß die ursprünglich berechneten Werte zweimalig interpoliert werden, sodass die letztendlich dargestellte Radiosityfunktion stark von der tatsächlichen abweicht (Abb. 8a).



**Abb. 8:** Die horizontale Achse beschreibt die Ausdehnung des Empfängers, der in 8 Patches unterteilt ist. Auf der vertikalen Achse sind die Radiositywerte aufgetragen. (a) Die Verwendung von Mittelpunktradiositäten bedeutet zwei lineare Interpolationen. (b) Eckpunktradiositäten können durch Gouraud-Shading direkt dargestellt werden. Man beachte, daß sich das Ergebnis viel besser an die tatsächliche Radiosityfunktion anschmiegt. Dargestellt ist die tatsächliche Radiosityfunktion (dünne Linien), die berechneten Punkte (schwarze Punkte), und die resultierenden interpolierten Werte (dicke Linien). Gestrichelte senkrechte Linien markieren die Eckpunkte des Meshs. Die gepunktete Linie in (a) beschreibt die lineare Interpolation von den Mittelpunkten (schwarze Punkte) zu den Eckpunkten (schwarze Kreise).

Die Verwendung von Eckpunktradiositäten hingegen bietet mehrere Vorteile. Zunächst führt die Berechnung der Radiosityfunktion direkt an den Eckpunkten zu einer genaueren Repräsentation, da nur ein einziger Interpolationsschritt, nämlich eben das Gouraud-Shading, erforderlich ist [WEH89].

Des Weiteren wird der Einfluß von sichtbaren Nähten an den Berührungskanten von Polygonen erheblich reduziert, da die Radiositywerte einer solchen Begrenzung nicht mehr von jedem Polygon einzeln auf die Eckpunkte extrapoliert werden, sondern direkt entlang derselben Begrenzung berechnet werden, wodurch z.B. auch Polygonecken immer exakte Werte enthalten (Abb. 8b).

Insbesondere für ein Eckpunktradiosity-Verfahren wird die notwendige Unterteilung zwischen einer Sichtbarkeitsklassifizierung und der Ermittlung des sichtbaren Anteils (vgl. Kapitel 2.7) deutlich:

Während die Unterteilung und der Refinementalgorithmus auf Links zwischen Clustern und/oder Patches (den *node links*) basiert, erfolgt der eigentliche Energietransfer zwischen einem Sender und den Empfängereckpunkten (mit Hilfe von *vertex links*) [K97][DDP99] – es sei denn der Empfänger ist ein Cluster.

Die *node links* erfordern keinen genauen Formfaktor (und damit auch keinen exakten Sichtbarkeitsanteil), sondern nur Energieschranken. Sie enthalten jedoch Informationen zur Sichtbarkeitsklassifizierung von Sender und Empfänger. Dies ist die Linkstruktur, die während des Refinementsschritts im Radiosityalgorithmus ausgewertet wird.

Auf der anderen Seite beschreiben *vertex links* die Energiemenge, die auf expliziten Punkten des Empfängers ankommt. Hier sind ein Formfaktor und qualitative Sichtbarkeitsinformationen erforderlich – beide jedoch nur zwischen einer Fläche und einem Punkt, was die Komplexität dieser Algorithmen erheblich verrin-

gert. Diese Linkstruktur wird während den Gathering- und PushPull-Schritten des Radiosityalgorithmus befragt, um den Energieaustausch durchzuführen.

Ob Eckpunkt- oder Mittelpunktradiositities verwendet werden sollen, sollte objektbasiert entschieden werden, denn nicht in allen Fällen ist die Verwendung der genaueren (und teureren) Eckpunktradiositymethode erforderlich oder erwünscht. Dies hängt sowohl von den Benutzeranforderungen, als auch von der Stellung der Wichtigkeit einer bestimmten Patch-Patch-Interaktion innerhalb der Simulation ab.

### Formfaktor und Gradient

Für einen Algorithmus auf Basis von Eckpunktradiositities werden keine Fläche-zu-Fläche-Formfaktoren benötigt. Der Formfaktor von einer (als konstant angenommenen) Senderfläche zu einem bestimmten Punkt auf dem Empfänger kann analytisch – und damit genau – über die Prisma-Formel bestimmt werden [HS67][BRW89], ohne auf das rechenintensive Doppelintegral oder fehlerbehaftete Approximationen angewiesen zu sein. Es ist jedoch zu beachten, daß entartete Fälle entstehen können, für die eine direkte Formfaktorberechnung nicht oder nur bedingt möglich ist (vgl. 2.2.2).

Um den Verlauf der Radiosityfunktion auf dem Empfänger beschreiben zu können, sind Informationen in Form eines Gradienten erforderlich. Dieser kann beispielsweise durch Differenzbildung der vorhandenen Formfaktoren abgeschätzt werden – was jedoch selbst bei einer korrekten Sichtbarkeitsklassifizierung zu deutlichen Artefakten führen kann, da diese Abschätzungen beliebig falsch sein können. Stattdessen ist es möglich, den exakten Radiositygradienten für Empfängerpunkte zu berechnen. Der Aufwand für dessen Berechnung ist nur geringfügig größer, wenn Gradient und Formfaktor gleichzeitig berechnet werden [HS95].

Für unverdeckte Situationen liefern Formfaktor und Gradient bereits sehr viele Informationen über Größe und Form der tatsächlichen Radiosityfunktion, die vom Refinementkriterium ausgewertet werden können.

Für partiell verdeckte Situationen läßt sich ebenfalls der exakte Radiositygradient bestimmen [A94][PC96], jedoch ist dessen Gültigkeitsbereich aufgrund von möglichen Diskontinuitäten in der Sichtbarkeitsfunktion auf dem Empfänger eingeschränkt: Selbst wenn aufgrund der Gradienten die Radiosityfunktion zwischen zwei Punkten auf dem Empfänger glatt zu verlaufen scheint, können eine oder gar mehrere Schattenkanten dazwischenliegen, wodurch das Verhalten der Radiosityfunktion unvorhersagbar wird. Nur ein aufwendiges *discontinuity meshing* [H92][LTG92] oder die zusätzlichen Informationen durch ein *visibility*

*skeleton* [DDP97] können hier eine verlässliche Beschränkung der Radiosityfunktion garantieren.

Die Vorteile, die aus der Verwendung der Gradienteninformation gezogen werden können sind jedoch stellenweise so bedeutend, daß sie selbst ohne vorhandenes *discontinuity mesh* nützlich sein können, da die Gefahr Schattenkanten zu verpassen vergleichsweise gering ist [K97]. Dies liegt einfach darin begründet, daß Gradienten innerhalb einer partiellen Situation in unterschiedliche Richtungen (nämlich immer von der Schattenregion weg) weisen. Eine Fehlinterpretation der Gradienteninformation kann also nur vorkommen, wenn für zwei beliebige Punkte sowohl der tatsächliche Wert der Radiosityfunktion, als auch der mit Hilfe des Gradienten vom zweiten Punkt extrapolierte Wert nahezu gleich sind.

Dieses Risiko kann durch zusätzliche Überlegungen weiter eingeschränkt werden. Ergibt z.B. die Sichtbarkeitsbestimmung, daß alle Eckpunkte vollständig sichtbar oder vollständig verdeckt sind, die Sichtbarkeitsklassifizierung jedoch klarlegt, daß eine partielle Verdeckung vorliegt, so ist offensichtlich, daß sich mindestens zwei Schattenkanten innerhalb des Patches befinden müssen, und die Gradienteninformation sollte nicht verwendet werden.

Diese Erweiterung ist insbesondere dann wichtig, wenn z.B. ein kleines Objekt auf einem ansonsten leeren Tisch liegt, oder bei einem Fußboden gerade die vier Ecken von verschiedenen Objekten blockiert werden. Weitere Einschränkungen sind denkbar [K97].

Außer einer verbesserten Abschätzung des Verlaufs der Radiosityfunktion liegen die Vorteile eines Gradienten in der sehr guten Eliminierung von Mach-Bändern (vgl. 2.2.1), und der Vermeidung unnötiger Unterteilungen aufgrund des nachfolgenden Gouraud-Shadings. Er kann prinzipiell sogar dazu verwendet werden, Formfaktor-Extrema zu finden [A96].

## Gathering und PushPull

Die Gathering- und PushPull-Schritte des hierarchischen Radiosityverfahrens müssen für die Verwendung von Eckpunktradiositäten angepaßt werden, da in diesem Fall geometrisch gleiche Eckpunkte von verschiedenen Hierarchiestufen gemeinsam genutzt werden, und so eigentlich identische Eckpunkte unterschiedliche Radiositäten von unterschiedlich unterteilten Sendern aufgrund der an den Vertices variierenden Linkstruktur erhalten. Genauere Beschreibungen der erforderlichen Anpassungen, sowie der während der Simulation nötigen, zusätzlichen Datenstrukturen, wie z.B. *vertex links* und die Speicherung von *shooting* und *gathering* Radiositäten finden sich in [K97], sowie in [DDP99].



Bei Eckpunktradiosities muß beachtet werden, daß ein als konstant angenommener Sender, auch wenn er nicht unterteilt ist, an jedem Eckpunkt Radiositywerte besitzt, die im Allgemeinen nicht identisch sind. Diese werden zur Berechnung oberer und unterer Schranken der Senderradiosity herangezogen. Als zu versendende Radiosity dient der Mittelwert aus diesen Werten, sodass der Sender nun durch eine konstante Radiosity plus Fehlerschranken definiert ist.

Der beim Patch gespeicherte Gathering-Radiositywert wird nur für den Push-Schritt benötigt. Im Pull-Schritt kann dieser Speicherbereich daher für das Weiterreichen zusätzlicher Informationen, insbesondere der Radiosityextrema und des Residuums (vgl. 2.6.1), wiederverwendet werden. Des Weiteren werden beim Push bei der Verwendung von Vertex-Radiosities die entsprechenden Werte linear interpoliert an die Kinder weitergereicht. Die angepaßte, kombinierte Gather-/PushPullroutine ist hier im Pseudocode angegeben:

```
GatherPushPull:
  Gathering:
    Element ist Lichtquelle?
      Setze Bmax, Bmin, Beps. Fertig.
    B = B + ankommende Push-Radiosity
    Element besitzt Emission und ist Rootpatch?
      B=B+E
    Für alle Links
      Berechne Linkradiosity. Bei Clustern:
        reiche gerichtete Radiosity an alle Kinder
  PushPull:
    Element ist Blatt?
      Beps=0, reiche Radiosity nach oben (Pull)
      Fertig.
    Für alle Kinder
      GatherPushPull-Rekursion
      Berechne Pull-Radiosity flächengewichtet
    Berechne Residuum Beps
```

## 2.6 Refinement

Das „Herz“ eines Radiositysystems bildet das Refinementkriterium, das dafür Sorge trägt, daß vorhandene Patches fein genug unterteilt werden, um eine gewünschte visuelle oder photometrische Qualität zu garantieren. Es trägt in klassischen Radiositysystemen oft den etwas mystischen Beinamen „Orakel“, denn lange Zeit war man sich nicht einig darüber, was die genaue Aufgabe des Refinementkriteriums war (oder besser: wie es diese am besten zu lösen hatte).

Die grundlegende Aufgabe des Refinements kann nun jedoch auf Basis der Beobachtungen und Feststellungen in obigen Kapiteln klar umrissen werden: Die Szene muß genau so unterteilt werden, daß

- das entstehende Mesh fein genug ist, daß z.B. der Simulationsfehler unterhalb einer vorgegebenen Schranke liegt, und
- das entstehende Mesh nicht zu fein ist, indem z.B. Unterteilungen vermieden werden, die keine nennenswerte Genauigkeitssteigerung erbringen.

### Applikationsabhängigkeit

Dies beschreibt jedoch nur die algorithmische Sicht. Aus Anwendersicht ist wichtig, daß das Refinementkriterium sich derart an dessen spezielle Anforderungen anpaßt, daß die für ihn bestmöglichen Ergebnisse geliefert werden.

Wie beeinflussen nun die unterschiedlichen Anforderungen der verschiedenen Benutzergruppen (vgl. Tabelle 1) den Aufbau und das Verhalten eines Algorithmus zur Berechnung von Fehlerschranken? Was die gewünschte (visuelle) Qualität angeht, so interessiert den Lichttechniker offensichtlich nur die mathematische und physikalische Korrektheit. Das optimale System für solch einen Benutzer wird daher aus der einfachsten Konfiguration bestehen: *constant radiosity assumption* für sowohl Sender als auch Empfänger, sowie ein reines (Energietransfer-)fehlerbasiertes Unterteilungssystem scheinen angebracht.

Die VR-Gruppe wird definitiv hardwareunterstütztes Gouraud-Shading verwenden. Es erscheint daher logisch, die Repräsentation der Radiosityfunktion auf dem Empfänger auf lineare Basisfunktionen zu erweitern.

Eine Funktion über Gouraud-Shading wiederzugeben bedeutet lineare Interpolation von (Eckpunkt-)Farben. Solange die Funktion selbst keine Unstetigkeiten besitzt, wird die generelle Repräsentation zufriedenstellend sein. Aufgrund der Natur stückweise linearer Interpolation werden jedoch Unstetigkeiten in der ersten Ableitung erzeugt, welche als Mach-Bänder sichtbar werden [G95]. Als Nebeneffekt einer Empfängerunterteilung (während die Radiosityfunktion durch lineare Segmente angenähert wird) werden sich die Unstetigkeiten der ersten Ableitung verringern, und Mach-Bänder verschwinden. Um dieses Artefakt also zu beseitigen ist eine lineare Repräsentation auf Empfängerseite ausreichend, sofern die Unterteilung fein genug ist, daß die Unstetigkeiten klein gehalten werden. Angemessene Schranken oder Abschätzungen, die Deutlichkeit von Machbändern zu quantifizieren sind jedoch erforderlich. Ein Mesh, das unter diesen Gesichtspunkten unterteilt wurde kann mit Gouraud-Shading ohne (wahrnehmbaren) Qualitätsverlust dargestellt werden.

Was die Senderseite betrifft wird deutlich, daß eine lineare Repräsentation hier das visuelle Eergebnis nicht beeinflusst, da die Senderradiosities nicht direkt im Ergebnis repräsentiert sind und damit keine sichtbaren Artefakte hervorrufen können; sie sind nur für die Energieübertragung wichtig. Daraus läßt sich schließen, daß lineare Sender nicht die (vom menschlichen Auge wahrgenommene) visuelle Qualität steigern und es also ausreichend ist, für Sender weiterhin die *constant radiosity assumption* zu verwenden – abgesehen natürlich für nicht-konstante Lichtquellen, wie texturierte Flächen oder LVKs.

Für den VR-Anwender sind somit also konstanter Sender und linearer Empfänger die optimale Wahl.

Ein weiterer Schritt hin zu hoher visueller Qualität kann für den Animations-Anwender vorgenommen werden. Er ist nicht auf Hardware-Gouraud-Shading beschränkt und könnte so von quadratischer oder höherer Interpolation profitieren. Wie jedoch in Kapitel 2.5.3 bereits diskutiert wurde, ist der menschliche Sehapparat gegenüber Unstetigkeiten der zweiten oder gar höheren Ableitungen von Intensitätsfunktionen eher unempfindlich. Solange das Mesh ausreichend unterteilt ist, daß Mach-Bänder verschwinden, wird die wahrgenommenen Qualität einer stetigen Funktion nicht mit höherer Interpolation zunehmen.

Somit kann sogar der Benutzer mit hochqualitativen Ansprüchen durch die konstante Sender/lineare Empfänger-Kombination voll zufriedengestellt werden. Gegebenenfalls kann natürlich das Linearitäts- $\epsilon$  auf Kosten eines feineren Meshes kleiner gewählt werden, um alle noch wahrnehmbaren Machbänder zu entfernen.

Die unterschiedlichen Anforderungen der Benutzergruppen können in ein einziges System auf Basis eines einfachen Mechanismus (z.B. einem „visuelle Qualität“-Schieberegler) zur Kontrolle des die visuelle Wahrnehmung betreffenden Teils des Refinementkriteriums integriert werden, während der Kernalgorithmus weiterhin das Energietransferkriterium verwendet um die mathematische Korrektheit des Ergebnisses zuzusichern.

Wiederum in Abhängigkeit von den Benutzeranforderungen, aber z.B. auch je nach zur Verfügung stehender Rechenzeit, kann (automatisch) zwischen Mittelpunkt- und Eckpunktradiosityalgorithmus gewählt werden. Die Unterscheidung zwischen diesen beiden Möglichkeiten ist wichtig, da sie direkt die Anwendbarkeit der berechneten Werte ebenso beeinflussen, wie die Annahmen oder Methoden zur Fehlerabschätzung. Z.B. kann der Radiositygradient nur sinnvoll mit Eckpunktradiosities eingesetzt werden.

Insgesamt muß sich das Radiositysystem also an die Benutzeranforderungen anpassen können.

## Eine Toolsammlung

Das eigentliche Refinementkriterium greift zur Realisierung dieser Anpassungsfähigkeit auf verschiedene Low-Level-Informationen einer gegebenen Sender-Empfänger-Konstellation zurück, die über eine entsprechende Tool-Sammlung zur Verfügung gestellt werden.

Diese Tools können in verschiedene Gruppierungen klassifiziert werden. Die Gruppe der **energiebasierten** Tools beinhaltet Werkzeuge zur Thematik des reinen Lichtenergieaustauschs zwischen dem zu betrachtenden Sender und Empfänger. Die **geometrischen** Tools befassen sich mit der gegenseitigen räumlichen Lage der beiden Objekte, und die **Sichtbarkeitstools** schließlich liefern Informationen über zwischen Sender und Empfänger liegenden Objekten.

Die einzelnen Tools müssen nichts über die Gesamtsituation wissen, sondern dienen nur dazu, möglichst schnell sehr grundlegende, einfache Fragestellungen zu beantworten und dem Refinementkriterium genügend Information zur Verfügung zu stellen, damit es seine Arbeit machen kann. Das Ergebnis solcher Tools kann z.B. in einer einfachen ja/nein-Antwort bestehen, in welchem Fall die Information bei dem entsprechenden Link als Flag gespeichert wird. Oder es kann sich um eine komplexe Information wie z.B. eine komplette Schattenliste verdeckender Objekte handeln. Sodann muß aufgrund der Benutzeranforderungen entschieden werden, ob diese Information gespeichert werden soll (und damit die Speicherkomplexität erhöht), ob sie *on-the-fly* Neuberechnet wird (und damit den Berechnungsaufwand erhöht), oder ob sie durch einfachere Annahmen ersetzt wird (und damit ggf. die Genauigkeit reduziert).

Die eigentliche Aufgabe des Refinementkriteriums ist es dann schließlich, die Frage zu beantworten, wann – oder grundlegender: **warum** – eine Sender-Empfänger-Situation unterteilt werden muß. Es ist dafür zuständig, die Einzelinformationen in die richtige Reihenfolge zu bringen, und effizient zu entscheiden, ob eins der beiden Objekte unterteilt werden muß, und wenn ja, welches.

Welche Tools das Refinementkriterium letztendlich auswertet, hängt zum einen sicherlich von der jeweiligen Situation zwischen Sender und Empfänger ab. Zum anderen sind es jedoch insbesondere die Benutzeranforderungen und Szenentypen, die dieser Entscheidung zugrunde liegen.

### 2.6.1 Energiebasierte Tools

Die erste Gruppe von Tools befaßt sich mit der Fragestellung, wieviel Energie zwischen zwei bestimmten Hierarchieelementen ausgetauscht wird.

Hierbei ist prinzipiell zu unterscheiden, ob der Empfänger ein Cluster oder ein Patch ist. Bei einem Cluster ist nur der *Betrag* der Energie relevant – je höher die Energie, desto feiner die Unterteilung. Bei einem Patch ist hingegen die *Energieverteilung* auf dem Empfänger interessant, d.h. der relative Verlauf der Radiosityfunktion. Die alleinige Tatsache einer hohen übertragenen Energiemenge bedeutet hier noch nicht die Notwendigkeit feiner Unterteilung.

Diese Unterscheidung ergibt sich jedoch implizit durch die Antworten der entsprechenden Tools und der Formulierung des eigentlichen Refinement-Kriteriums, wie sich später zeigen wird.

Die übertragene Energie ist beschrieben durch Senderradiosity mal Formfaktor ( $BF$ ), und da die Hierarchieelemente nicht punktförmig sind, sind obere und untere Schranken zu bestimmen. Die Genauigkeit dieser Schranken bestimmt hierbei direkt die Qualität (d.h. das Verhältnis zwischen physikalischer Korrektheit und Meshkomplexität) des Endergebnisses.

### Radiosityextrema

Die oberen und unteren Schranken der (Sender-)Radiosity sind bereits in den Radiosity-Datenstrukturen enthalten und durch den maximalen bzw. minimalen Senderradiositywert eines jeden Patchbaumes definiert. Durch einfaches Traversieren der Bäume und min/max-Bildung werden für jeden Level diese Extrema bestimmt. Dieser Schritt kann, wie in Kapitel 2.5.4 beschrieben, in den normalen PushPull-Schritt des Radiosityverfahrens integriert werden.

Falls es sich bei dem Sender um eine nicht-Lambert'sche Lichtquelle handelt, ergibt sich die Berechnung aus dem Kontext der Lichtquelle: Für Punktlichtquellen, Spotlights ohne Helligkeitsabfall und Sonnenlicht gilt  $\Delta B = 0$ , und  $B_{\max}$  und  $B_{\min}$  sind bekannt. Letzteres gilt auch für Spotlights mit Helligkeitsabfall. Für LVKs und Tageslicht sind die Radiositywerte innerhalb eines bekannten Raumwinkels gegeben, und die Extrema können z.B. durch Sampling ermittelt werden.

Bei einer groben Radiositylösung z.B. durch ein großes Refinement- $\varepsilon$  kann es natürlich vorkommen, daß ein Patch nicht unterteilt wurde, obwohl es einen nicht-konstanten Radiosityverlauf besitzt. In diesem Fall scheint es, als würde daraus eine falsche Schlußfolgerung für die Radiosityextrema resultieren, sobald ein solches Patch als Sender herangezogen wird.

Für die direkt nachfolgende Iteration trifft das zu – die Simulation kann den Fehler des nicht unterteilten Patches natürlich nicht ausgleichen. Wird jedoch  $\varepsilon$  im Laufe der Simulation verkleinert und so das Patch nachträglich doch unterteilt, werden automatisch während dem nächsten Gathering- und PushPull-

Schritt die Radiosityextrema aktualisiert und spiegeln jetzt den nicht-konstanten Radiosityverlauf wider. Es ergibt sich so eine positive Rückkopplungsschleife: Je genauer die Radiosityberechnung, desto genauer die Radiosityschränken, was sich wieder auf die Berechnung auswirkt.

## Residuum

Die Differenz des Radiositywertes eines Elements zwischen zwei Iterationen. Das Residuum ist ein Maß der Konvergenz und dient dazu festzustellen, ob weitere Iterationen berechnet werden müssen. Es wird ebenso wie die Radiosityextrema während des PushPull-Schritts bestimmt.

## Formfaktormaxima

Formfaktorextremwerte beschreiben die Veränderung des Formfaktors an verschiedenen Punkten auf dem Empfänger. Eine korrekte Bestimmung des Formfaktors ist entscheidend für die Funktionstüchtigkeit und Effektivität des Refinementkriteriums.

Zu jeder Kombination zweier Hierarchieelemente ist eine Funktion erforderlich, die eine verlässliche Abschätzung oder wenn möglich eine exakte Beschreibung des maximalen Formfaktors auf dem Empfängerelement bestimmt. Diese werden wie folgt ermittelt:

Empfänger ist ein Patch, Sender ist:

- **Punktlicht:** Das analytisch exakte Formfaktormaximum läßt sich berechnen, indem zunächst der der Lichtquelle nächstgelegene Punkt der Empfängerfläche gefunden wird. Befindet dieser sich innerhalb des Patches, ist das Maximum gefunden. Ansonsten befindet es sich am nächstgelegenen Empfängereckpunkt, bzw. dem nächstgelegenen Punkt eine Patchkante. Das Maximum ist der Formfaktor an diesem Punkt.
- **Spotlicht:** Das analytisch exakte Maximum liegt auf der Schnittlinie zwischen der Patchebene und der Ebene, die von Spotrichtung und Patchnormalen aufgespannt wird.
- **LVK:** Durch die generelle Unvorhersagbarkeit von LVKs muß über das Patch gesampelt werden. Ein konservativer Ansatz lautet: Punktlichtmaximum mal maximales LVK-Sample im Patch-Raumwinkel.
- **Tages-, Sonnenlicht:** Das analytisch exakte Maximum ist identisch zum Formfaktor. Diese Lichtquellen sind unendlich weit entfernt, der Formfaktor ist daher konstant auf dem gesamten Empfänger.

- Patch: Es ist sichergestellt, daß nur ein Maximum existiert [DF93]. Für die Annahme eines infinitesimalen Senders gibt es eine analytische Lösung zur Bestimmung des Maximums auf der Patchebene. Hierbei ergibt sich die Lage des Maximums bei:

$$\vec{P}_{\max} = \vec{M}_S + \frac{\overrightarrow{M_E M_S} \cdot \vec{n}_E}{1 - \cos^2 \varphi} ((a - \cos \varphi) \vec{n}_S + (a \cos \varphi - 1) \vec{n}_E), \text{ mit}$$

$$a = \frac{1}{3} \left( \cos \varphi + \sqrt{\cos^2 \varphi + 3} \right).$$

$$FF_{\max} = FF(\vec{M}_S, \vec{P}_{\max})$$

Dabei sind  $\vec{M}_S$  bzw.  $\vec{M}_E$  die Mittelpunkte des Senders, resp. Empfängers, und  $\varphi$  der Winkel zwischen den beiden Normalen. Das Maximum muß analog zur Punktlichtquelle auf das Empfängerpatch geclippt werden. Dieses Maximum ist für flächige Sender jedoch nur näherungsweise korrekt. Die Formel sollte dann für Sendermittel- und -eckpunkte mehrfach angewendet werden.

- Cluster: Das analytisch exakte Maximum ergibt sich analog zur Punktlichtquelle durch Bestimmung des nächstgelegenen Punktes auf der Empfängerenebene, Clippen gegen den Empfänger und Berechnung des Formfaktors. Das Maximum ist gleich 1, falls sich der Empfänger oder Teile des Empfängers innerhalb des Senders befinden.

Empfänger ist ein Cluster, Sender ist:

- Punktlicht: Das analytisch exakte Maximum befindet sich an dem der Punktlichtquelle nächstgelegenen Punkt.
- Spotlight: Eine konservative Lösung ist, den Formfaktor für den nächstgelegenen Punkt und den kleinsten Winkel zu bestimmen. Das Finden einer analytischen Lösung steht noch aus.
- LVK: Durch die generelle Unvorhersagbarkeit von LVKs muß über den Cluster (genauer: über die eine bis maximal drei der LVK zugewandten Seiten des Clusters) gesampelt werden.
- Tages-, Sonnenlicht: Das analytisch exakte Maximum ist identisch zum Formfaktor. Diese Lichtquellen sind unendlich weit entfernt, der Formfaktor ist daher auf dem gesamten Empfänger konstant.

- Patch: Für die Annahme eines infinitesimalen Senders gibt es unter der Verwendung eines kugelförmigen Clusters eine analytisch exakte Lösung. Der Formfaktor für eine solche Konfiguration berechnet sich zu:

$$FF(\vec{M}_S, \vec{M}_E) = \frac{\cos \theta_s}{\pi \cdot \vec{d} \cdot \vec{d}} = \frac{\vec{d} \cdot \vec{n}_s}{\pi (d^2)^{3/2}}, \quad \vec{d} = \overrightarrow{M_E M_S}$$

Dieses Maximum ist für flächige Sender jedoch nur angenähert korrekt. Die Formel sollte dann für Sendermittel- und -eckpunkte mehrfach angewendet werden.

- Cluster: Das analytisch exakte Maximum kann über den Abstand der Cluster-Boundingboxes bestimmt werden.

Ist der Empfänger ein Cluster, so gilt außerdem, daß das Maximum gleich 1 ist, falls sich der Sender oder Teile des Senders innerhalb des Empfängers befinden. Bei Tages- und Sonnenlicht kann dies nicht passieren.

### Formfaktorminima

Analog zu den Formfaktor-Maxima gibt es Funktionen zur Bestimmung der Minima. Für sie gilt im Prinzip das gleiche wie für die Maxima. Die entsprechenden Varianten ergeben sich wie folgt:

Empfänger ist ein Patch, Sender ist:

- Punktlicht: Das analytisch exakte Minimum befindet sich an dem am weitesten von der Lichtquellen entfernten Patcheckpunkt.
- Spotlight: Das analytisch exakte Minimum befindet sich an einem der Empfängereckpunkte. Das Minimum ist 0, falls mindestens ein Eckpunkt außerhalb des Spot-Öffnungswinkels liegt. Ansonsten wird in einer *brute force* Methode das Minimum aller Eckpunktformfaktoren bestimmt.
- LVK: Durch die generelle Unvorhersagbarkeit von LVKs muß über das Patch gesampelt werden. Ein konservativer Ansatz lautet: Punktlichtminimum mal minimales LVK-Sample in Patch-Raumwinkel. Aus Effizienzgründen sollten zum Sampling Minimum und Maximum gemeinsam bestimmt werden.
- Tages-, Sonnenlicht: Das analytisch exakte Minimum ist identisch zum Formfaktor. Diese Lichtquellen sind unendlich weit entfernt, der Formfaktor ist daher auf dem gesamten Empfänger konstant.



- Patch: Das analytisch exakte Minimum befindet sich an einem der Empfängerereckpunkte [D94]. Die *brute force* Methode ist, das Minimum aller Eckpunktformfaktoren zu bestimmen.
- Cluster: Das analytisch exakte Minimum befindet sich an dem am weitesten von der Lichtquelle entfernten Patchcheckpunkt.

Ist der Empfänger ein Cluster, muß immer vom *worst case* (d.h., der Cluster enthält eine abgewandte oder verdeckte Fläche, die kein Licht empfangen kann) ausgegangen werden. Das Minimum ist daher immer gleich 0 zu setzen. Im Wesentlichen ist dies der Umstand, der bewirkt, daß sich das Refinementkriterium für diesen Fall automatisch auf ein energiebasiertes Kriterium reduziert.

### Formfaktoren

Der eigentliche Formfaktor wird nicht mehr für die Frage, wann zu unterteilen ist benötigt, sondern ist nun nur noch für den Energieaustausch entlang bereits etablierter Links erforderlich.

Für eine Patch-Patch-Situation kann mit der sog. Prisma-Formel für jeden beliebigen Empfängerpunkt der Formfaktor analytisch exakt ermittelt werden. Für punktförmige Lichtquellen sind spezielle, sog. Receive-Funktionen erforderlich, die die jeweiligen Ausbreitungscharakteristiken in Betracht ziehen. Für Tageslicht und Sonnenlicht vereinfachen sich die Formeln entsprechend (kein Winkel mit der Sendernormalen, kein Konzept eines Abstandes, ...).

Da es sich bei der Prisma-Formel um ein Konturintegral handelt, kann der exakte Formfaktor prinzipiell auch für eine partiell verdeckte Situation berechnet werden, vorausgesetzt es wurde eine entsprechend geclippte Kontur bestimmt [PC96]. Da das Berechnen einer solchen Kontur jedoch numerische Genauigkeitsprobleme mit sich bringt, beschränkt man sich oft darauf, im partiellen Fall den unverdeckten Formfaktor mit einem prozentualen Sichtbarkeitsterm zu multiplizieren. Hierdurch kann der ermittelte Formfaktor für einen bestimmten Punkt geringfügig von seinem tatsächlichen Wert abweichen – er bewegt sich jedoch immer innerhalb der ermittelten Fehlerschranken.

Für Cluster ist keine besondere Formfaktor-Formel erforderlich. Vielmehr wird im Falle eines Senderclusters die Punkt-zu-Punkt-Formfaktorvereinfachung bis auf den Senderwinkel ausgewertet. Das Ergebnis wird mit der rekursiv ermittelten Summe der richtungsabhängigen Senderradiosity aller im Cluster enthaltenen Polygone gewichtet. Wichtig ist hierbei, daß sich dieser Formfaktor im Prinzip ändert, wenn sich die Radiosityverteilung innerhalb des Clusters ändert.

Diese Formfaktoren müssen also **nach jeder Iteration neu berechnet** werden. Ggf. wird nur die prozentuale Sichtbarkeit im Link gespeichert.

Bei einem Empfängercluster muß die im Gathering-Schritt bei jedem Link erhaltene Energie sofort richtungsabhängig an alle im Cluster enthaltenen Polygone heruntergereicht werden.

Der sog. *self factor* eines Clusters läßt sich z.B. durch

$$FF_{self} = 1 - e^{-\frac{\text{Clusterfläche} \cdot \text{Clusterdiagonale}}{4 \cdot \text{Clustervolumen}}}$$

abschätzen.

## Gradient

Der Gradient des Formfaktors kann für Patch-Patch-Situationen mit geringem Mehraufwand bestimmt werden [HS95]. Für nichtflächige Lichtquellen wie z.B. LVKs ist seine Bestimmung schwierig. Die Berechnung ist, wiederum mit Hilfe des Konturintegrals, auch bei partiell verdeckten Situationen möglich, wozu eine *depth clipped contour* erforderlich ist [A94][PC96]. Die Aussagekraft eines solchen Gradienten ist jedoch aufgrund möglicher Diskontinuitäten in der Sichtbarkeitsfunktion eingeschränkt [K97]. Eine vollständige Lösung dieser Problematik kann nur mit einem *discontinuity mesh* [H92][LTG92] erfolgen.

Der Formfaktorgradient an einem bestimmten Empfängerpunkt kann herangezogen werden, um die Linearität der Radiosityfunktion auf dem Empfänger zu überprüfen.

## Wahrnehmungsbasierte Schranken

Eine Sonderklasse der energiebasierten Tools sind solche, die auf wahrnehmungsbasierten und physiologischen Aspekten des (nicht-linearen) menschlichen Sehapparates basieren. Sie sind zum Teil blickpunktabhängig (z.B. [MPT97]) und müssen insbesondere die Quantisierungstechnik zur Umwandlung von Radiositywerten in darstellbare, und damit vom Ausgabemedium abhängige, Farben in ihren wahrnehmungsbasierten Kontext integrieren (siehe speziell [GH97]).

### 2.6.2 Geometrische Tools

Diese Gruppe von Tools bestimmt geometrische Eigenschaften von Sender und Empfänger – und nur von diesen, d.h. ohne möglicherweise dazwischenliegende

Objekte in Betracht zu ziehen. Sie dienen dazu festzustellen, ob die Objekte prinzipiell in der Lage sind, Energie auszutauschen.

### Halbraumtest

Dieser Test stellt fest, ob zwei Objekte prinzipiell Energie austauschen können, oder ob sie z.B. voneinander abgewandt sind.

Dieser Test ist weiterhin nur für Patches relevant, da nur für sie die Beschränkung gilt, daß sie nur mit dem vorderen Halbraum interagieren (Der Sonderfall des Spotlights wird bereits durch das FF-Minimum abgefangen).

Ist eines der Objekte ein Patch, müssen alle Eckpunkte des zweiten Objektes gegen die Patchebene getestet werden. Nur wenn mindestens ein Eckpunkt im vorderen Halbraum liegt, ist ein Energieaustausch prinzipiell möglich.

Aus Effizienzgründen ist ein Boundingbox-Test vorangestellt:

- Ist der am weitesten „hinten“ liegende Boundingbox-Punkt im vorderen Halbraum, liegen sämtliche Objekte in dieser Boundingbox ebenfalls im vorderen Halbraum.
- Ist der am weitesten „vorne“ liegende Boundingbox-Punkt im hinteren Halbraum, liegen sämtliche Objekte in dieser Boundingbox ebenfalls im hinteren Halbraum.

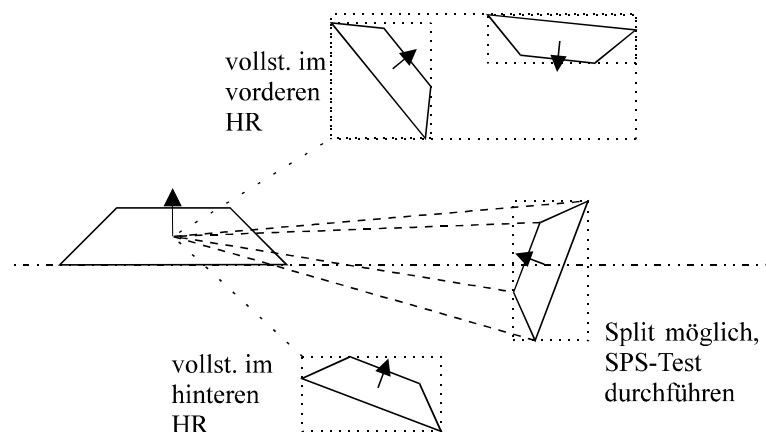


Abb. 9 : Halbraum und *support plane* Test: Bezüglich der einzelnen Fläche braucht der *support plane split* Test nur gegen Objekte in der geschnittenen Boundingbox durchgeführt zu werden. Objekte in der im vorderen Halbraum liegenden Boundingbox können mit der Fläche interagieren, während die Objekte der unteren Boundingbox abgewandt sind.

Die Patch- bzw. Polygonebene wird an verschiedenen Stellen im Radiosityalgorithmus verwendet, daher sollte die Ebenengleichung  $\vec{x} \cdot \vec{n} - d = 0$  in der Polygon-Datenstruktur gespeichert werden. Da die Normale hier üblicherweise sowieso

bereits vorhanden ist, beschränkt sich die zusätzliche Datenmenge auf den Fließkommawert  $d$ .

Ein erweiterter (und entsprechend teurerer) Algorithmus zieht die Varianz der Eckpunktnormalen auf einem Polygon mit in Betracht, um das Konzept der Eckpunktnormalen für Radiosity entsprechend für den Halbraumtest zu erweitern: Auch wenn die Flächennormale in den hinteren Halbraum zeigt, ist es dennoch möglich, daß interpolierte Normalen Teile des Polygons sichtbar erscheinen lassen.

### *support plane split* Test

Dieser Test ist ebenfalls nur für Patches relevant.

Durchdringt eine (verlängerte) Patchebene das zweite Objekt, so muß dies im Radiosity-Verfahren berücksichtigt werden, um Fehler zu vermeiden:

- Wird der Sender geschnitten, so darf der Formfaktor (und somit auch die Formfaktorextrema) nur für den im vorderen Halbraum befindlichen Senderteil bestimmt werden. Eine physikalische Unterteilung ist nicht erforderlich, da keine sichtbaren Artefakte entstehen können.
- Wird der Empfänger geschnitten, entsteht eine Schattenkante (genauer: eine DI-Unstetigkeit), und es ist Aufgabe des Refinement-Kriteriums zu entscheiden, ob der Empfänger unterteilt werden muß.

Der Halbraumtest ist im Prinzip ein Sonderfall dieses Tests; die beiden Verfahren werden daher kombiniert.

Wird eines der Objekte unterteilt, werden gewonnene Informationen weitergereicht:

- Nur wenn ein *support plane split* vorlag, bzw. das Objekt ein Cluster war, muß dieser Test für die Kinder wiederholt werden. Um diese Information transparent zu halten, wird das Flag „verursacht *support plane split*“ für Cluster immer gesetzt.
- Wurde einmal „vollständig im vorderen/hinteren Halbraum“ erkannt, gilt dies für alle weiteren Unterteilungen.
- Ein Sender-Split ändert seine Eigenschaften wie z.B. die Lage nicht, wenn der Empfänger unterteilt wird, und umgekehrt.

Hieraus folgt, daß der eigentliche Test nur für *root*-Patches durchgeführt werden muß – für bereits unterteilte Patches können wesentliche Informationen vorangegangener Tests wiederverwendet werden.

Weitere Kriterien zur Effizienzsteigerungen dieser Tests finden sich in [K97].

### 2.6.3 Sichtbarkeitstools

Der Einfluß dritter Objekte wird über eine Sichtbarkeitsbestimmung ermittelt. Diese wird aufgrund ihrer Komplexität und Wichtigkeit im Radiosityverfahren in einem gesonderten Kapitel (Kapitel 2.7) detailliert beschrieben.

### 2.6.4 Refinementkriterium

Dadurch, daß die in den vorigen Abschnitten beschriebenen Tools zu jedem Elementtyp passende Informationen liefern, kann das Refinementkriterium vollkommen allgemein gehalten werden. Dabei ist die entscheidende Fragestellung die, ob ein Link überhaupt unterteilt werden muß, und falls ja, ob entweder der Sender oder der Empfänger unterteilt werden muß.

Das weithin bekannte Kriterium  $BF$  hat in einem effizienten Radiositysystem nichts zu suchen, da es den Verlauf der Radiosityfunktion auf dem Empfänger (d.h. das, was letztendlich sichtbar ist) nicht wiedergeben kann und daher einen ungleich höheren Unterteilungslevel erforderlich macht, um die gleiche sichtbare Qualität zu erreichen, wie ein gut durchdachtes Kriterium.

Eine Unterteilung ist genau dann nötig, wenn der Energieaustausch, der von dem Link beschrieben wird, nicht als konstant angenommen werden kann, d.h. der Wert  $\Delta E = \rho_E A_E \cdot (F_{\max} B_{\max} - F_{\min} B_{\min})$  (vgl. [LSG94][M96]) größer einer gegebenen Schranke  $\varepsilon$  ist. Eine Unterteilung hängt somit von der relativen Energiedifferenz ab, nicht vom Absolutwert.

Die Skalierung mit der Empfängerfläche ist sinnvoll zum einen, um  $\varepsilon$  möglichst szenenunabhängig (bzgl. Dimension) zu machen, zum andern, um die Unterteilung von großen Empfängern zu begünstigen.

Nachdem so entschieden wurde, daß unterteilt werden muß, muß die Frage beantwortet werden, welches der beiden Elemente unterteilt werden muß. Abweichungen auf dem Sender sind nur bzgl. der Radiosity relevant, während bei Abweichungen auf dem Empfänger der Formfaktor entscheidend ist. Der Sender wird also unterteilt, wenn:

$$F_{\max} \Delta B > \Delta F B_{\max}, \text{ was sich umformen läßt zu: } F_{\max} B_{\min} < F_{\min} B_{\max}.$$

Ansonsten wird der Empfänger unterteilt.

Da für Cluster  $F_{\min}$  immer mit 0 abgeschätzt werden muß, reduziert sich das Refinementkriterium auf Cluster-Cluster-Ebene automatisch zu einem reinen Energieübertragungskriterium:  $\Delta E = \rho_E A_E F_{\max} B_{\max}$ . Der zweite Schritt des Krite-

riums würde sich jedoch zu  $F_{\max} B_{\min} < 0$  reduzieren – der Sender würde also nie unterteilt werden. Für den Fall, daß beide Elemente Cluster sind, muß also die Frage, welches der Elemente unterteilt werden muß, durch „Unterteilung des Senders, falls  $A_S > A_E$ , sonst Unterteilung des Empfängers“ beantwortet werden.

Zu beachten ist, daß im Falle partieller Sichtbarkeit (was durch die Sichtbarkeitstools festgestellt wird) der Formfaktor kleiner sein kann als im unverdeckten Fall. Dementsprechend muß also  $F_{\min}$  mit dem noch sichtbaren Anteil gewichtet werden. Dies ist ein konservativer Ansatz, denn für eine exakte Lösung müßten die Formfaktorextrema genau für die sichtbare Teile bestimmt werden, denn z.B. könnte auch  $F_{\max}$  kleiner werden, wenn nämlich die Position des tatsächlichen Maximums verschattet ist. Die Berechnung der Extrema für komplizierte partielle Situationen ist jedoch sehr aufwendig, und nicht immer analytisch lösbar, weshalb der vorgeschlagene Ansatz die bessere Wahl ist.

Speziell wenn mehrere Lichtquellen zusammenwirken und einzeln betrachtet Unterteilungen erforderlich machen würden, ist es leicht möglich, daß sich als Ergebnis eine verhältnismäßig homogen beleuchtete Fläche ergibt. Spezielle globale Refinementmethoden, deren Unterteilungsentscheidung nicht nur von einem einzelnen Sender und Empfänger abhängig ist, können für diese Situationen die Meshkomplexität reduzieren [SSS00b].

## Reihenfolge

Die Effizienz des Refinements hängt unter anderem davon ab, in welcher Reihenfolge welche der beschriebenen Tools aufgerufen werden. Die maximal effizienteste Reihenfolge ist letztendlich szenenabhängig, und kann normalerweise nur im Nachhinein festgestellt werden. Die Erfahrung zeigt jedoch, daß die hier beschriebene Abfolge die besten Ergebnisse liefert.

- „ist unterteilbar“ überprüfen.  
Vorgegebene Werte, die gegebenenfalls vom Benutzer modifiziert werden können (maximaler Level, minimale Patch-Fläche) auswerten. An dieser Stelle sollte auch überprüft werden, ob der Reflexionskoeffizient des Empfängers 0 ist.
- Empfänger-Split überprüfen.  
Beinhaltet auch die Halbraum-Sichtbarkeit. Der Test ist nur nötig, wenn der Sender ein Patch ist und das Flag gesetzt ist, daß der Test erfolgen muß, oder bei einem unterteilten Self-Link. Die Lage des Splits wird ggf. bestimmt, um Sichtbarkeitstests auf den vorderen (Sender-)Halbraum zu beschränken.
- Sender-Split überprüfen.  
Beinhaltet auch Halbraum-Sichtbarkeit. Der Test ist nur nötig, wenn der Empfänger ein Patch ist und das Flag gesetzt ist daß der Test erfolgen muß,

oder bei einem unterteilten Self-Link. Die Lage des Splits wird ggf. bestimmt, um Sichtbarkeitstests auf den vorderen (Empfänger-)Halbraum zu beschränken und den Formfaktor bzw. die Extrema korrekt berechnen zu können.

- Formfaktor-Minimum und -Maximum bestimmen.  
Abbruch, falls Maximum gleich 0 (kein Energieaustausch).
- Sichtbarkeits-Klassifizierung vornehmen.  
Bestimmung erfolgt nur, wenn die Klassifizierung vorher unbekannt war oder „partiell“ lautete.
- Prozentuale Sichtbarkeit bestimmen.  
Bestimmung erfolgt nur, wenn die Klassifizierung „partiell“ lautet.  $F_{\min}$  wird mit diesem Wert gewichtet.
- Refinement-Kriterium auswerten.
- Formfaktor für den Energieaustausch berechnen.  
Nur falls ein Link erzeugt wird.

Wird an irgendeiner Stelle entschieden, weder zu unterteilen, noch einen Link anzulegen, ist es dennoch erforderlich, die gewonnene Information dann in einem *passiven* Link zu speichern, wenn im späteren Verlauf dynamisches Radiosity zur Anwendung kommen soll [DS97].

## 2.7 Visibilitätsbetrachtung

Mit der aufwendigste Teil (und damit auch mit die kritischste Stelle hinsichtlich der Berechnungsgeschwindigkeit [HSD94]) in einem Radiosity-Verfahren – und der empfindlichste in Bezug auf sichtbare Artefakte, und somit der Ausgabequalität – ist die Sichtbarkeitsbestimmung, d.h. der oder die Tests um festzustellen, ob ein Objekt (ggf. teilweise) im Schatten liegt oder nicht.

Warum das so ist wird klar, wenn man sich die möglichen Fehlentscheidungen einer Sichtbarkeitsbestimmung genauer ansieht:

- Wird eine in Wirklichkeit vollständig verdeckte oder vollständig sichtbare Situation als partiell verdeckt klassifiziert, wird das Refinementkriterium durch eine feinere Unterteilung versuchen, diese – in diesem Fall nicht existente – Sichtbarkeitsänderung zu finden. Hierbei entsteht jedoch kein Fehler, weder in Form einer sichtbaren Störung, noch in der mathematischen Korrektheit. Statt dessen wird durch die unnötige Unterteilung ein erhöhter Rechenaufwand verursacht, und aus dem dadurch komplexeren Mesh folgt eine langsame Renderingzeit.

- Bleibt eine in Wirklichkeit partiell verdeckte Situation unerkannt, indem sie als vollständig verdeckt oder vollständig sichtbar klassifiziert wird, wird eine Schattenkante verpaßt und sichtbare Artefakte entstehen.

Während der erste Fall, als Überabschätzung der partiell verdeckten Situation, sicherlich unerwünscht ist – aus der Statistik ist dieser Fall auch als **Fehler 1. Art** bekannt – so ist es jedoch weit wichtiger, den zweiten Fall zu vermeiden, da hierdurch meß- oder sichtbare Fehler in der Radiositysimulation entstehen (**Fehler 2. Art**).

Die Aufgabe des Sichtbarkeitstools muß es demnach sein, ein robustes und verlässliches Klassifizierungsschema zur Verfügung zu stellen, das in mehrere Gruppen unterteilt ist, um dem Refinementkriterium zu ermöglichen, jeweils die optimale Auswahl an Tools in Abhängigkeit von den Benutzeranforderungen oder den zu Verfügung stehenden Ressourcen wie Berechnungsaufwand oder Speicherbedarf zu wählen.

Konkret bedeutet dies, daß das Modul zur Sichtbarkeitsbestimmung überprüft, ob zwischen zwei Elementen, zwischen denen Strahlungsenergie ausgetauscht werden soll, eine komplette oder teilweise Verdeckung vorliegt. Da beide Elemente räumlich ausgedehnt sein können, ist diese Prüfung notwendig z.B. für eine genaue Berechnung des Formfaktors oder als Unterteilungskriterium im hierarchischen Radiosity-Prozeß. Für teilweise Verdeckungen wird des weiteren der prozentuale Anteil der Verdeckung bestimmt.

### 2.7.1 Zweistufigkeit

Die Unterteilung der Sichtbarkeitsbestimmung in zwei unabhängige Abschnitte, nämlich die **Sichtbarkeitsklassifikation** und die **Ermittlung des sichtbaren Anteils**, ist ein essentieller Schritt, um das Verhalten des Algorithmus auf beste Performance in den entsprechenden Abschnitten eines Radiositysystems hin zu optimieren. Ein experimentelles Verfahren einer solchen Unterteilung wird in [K97] diskutiert.

Nachdem also bei der Klassifizierung der Sichtbarkeit eines Links die generelle Entscheidung erfolgt, ob keine, eine teilweise oder eine vollständige Verdeckung vorliegt (dies ist also ein **qualitativer Test**), dient der zweite Teil zur Ermittlung des Anteils der Verdeckung. Dieser **quantitative Test** ist nur für die partiell verdeckten Fälle notwendig, für die vollständig verdeckten Fälle kann ohne weitere Überprüfung sofort der Wert 0 geliefert werden, für die unverdeckten Fälle der Wert 1.

Bei der Sichtbarkeitsprüfung müssen alle Fälle verschiedener Hierarchieelemente berücksichtigt und unter Umständen unterschiedlich behandelt werden.



## 2.7.2 Sichtbarkeitsklassifikation

Verschiedene Ansätze zur Sichtbarkeitsklassifikation sind denkbar.

[TH93] beschreibt beispielsweise eine Strategie unter Verwendung von Plucker-Koordinaten. Neuere Verfahren wie das *visibility skeleton* [DDP97][DDP99] erlauben Möglichkeiten zur schnellen Sichtbarkeitsklassifikation, jedoch gestatten ihre extremen Speicheranforderungen und fehlende numerische Robustheit keinen Einsatz einer solchen Methode in einem anwendbaren Radiositysystem.

Die hier beschriebene Sichtbarkeitsklassifikation besteht aus mehreren Tools, die sich zu einem vollständigen Verfahren ergänzen.

Die in Kapitel 2.6.2 beschriebenen geometrischen Tools bilden hierbei eine Vorstufe, während sich die hier vorgestellten Tools ausschließlich auf den Einfluß dritter Objekte beschränken.

### Ray Casting

Die einfachste und vermutlich schnellste Variante ist ein Ray Casting Schritt, der mit einer bestimmten (z.B. vom abgedeckten Raumwinkel abhängigen) Anzahl zufällig über Sender und Empfänger verteilten Strahlen [WEH89] verdeckende Objekte zu finden versucht. Der große Nachteil bei dieser Methode ist jedoch, daß die Wahrscheinlichkeit hoch ist, daß Objekte (und damit Schattenkanten) verpaßt werden. Insbesondere können „vollständig sichtbar“ und „vollständig verdeckt“ nicht verlässlich mit dieser Methode bestimmt werden.

Ray Casting ist daher zur Sichtbarkeitsklassifikation **ungeeignet**.

### Shaft Culling

Zur Sichtbarkeitsklassifikation für einen Link zwischen zwei Hierarchie-Elementen wird zunächst ein Shaft nach Haines&Wallace konstruiert [HW91], so er nicht bereits vorliegt, da er z.B. bei einem bestehenden Link gespeichert wurde. Das Shaft Culling von Haines&Wallace liefert verlässliche Ergebnisse und wird daher statt einem Ray Casting verwendet. Hierbei werden verschiedene Klassifizierungsflags eingesetzt:

- **Unknown:** Noch keine Sichtbarkeit ermittelt
- **Visible:** Definitiv vollständig sichtbar
- **Invisible:** Definitiv vollständig verdeckt
- **Partial:** Definitiv partiell verdeckt

- **Complex:** Wahrscheinlich partiell, aber ggf. überabgeschätzt.

Mit jedem Link in der Hierarchie korrespondiert somit genau ein Shaft. Shafts können bei den Links gespeichert werden, damit sie z.B. für dynamisches Radiosity mit der *line space hierarchy* [DS97] zur Verfügung stehen. Es ist jedoch vorgesehen, daß die Shafts eventuell zwecks Speicherplatz-Reduzierung auch gelöscht werden können. Gegen die Ebenen des Shafts werden die Boundingboxen der potentiell verdeckenden Hierarchie-Elemente entsprechend der *overlap open* Strategie unter Verwendung der Cluster-Hierarchie getestet. Das Ergebnis ist eine Liste von Hierarchie-Elementen, die potentiell den Shaft schneiden, die *Kandidatenliste*.

Die Kandidatenliste wiederum sollte nur temporär im Speicher gehalten werden, solange sie noch benötigt wird (z.B. bis die Sichtbarkeit eines Links genügend genau bestimmt wurde und der Anteil der Verdeckung bestimmt ist). Sie sollte danach gelöscht werden um Speicher zu sparen.

### Ebenentests

Durch verschiedene zusätzliche Tests der (potentiellen) Verdecker gegen Sender- und Empfängerebenen (sofern es sich um Patches handelt), sowie Tests mit der Verdeckerebene läßt sich die durch reines Shaft Culling stark überabgeschätzte Menge der Verdecker durch Ausschluß einiger nicht in Frage kommender Objekte weiter verkleinern. Bei diesen Tests handelt es sich um schnelle Ebenentests, wie sie auch für die oben beschriebenen geometrischen Tools angewandt werden – nun jedoch zwischen Sender bzw. Empfänger und einem potentiellen Verdecker [K97].

Mit diesen Tests wird unter anderem auch festgestellt, ob die Verdeckerebene eine *separating plane* von Sender und Empfänger darstellt, was Grundvoraussetzung für eine totale Verdeckung ist.

### Erweiterte Tests

Wenn die Kandidatenliste nicht leer ist, werden nun alle Kandidaten genauer getestet. Dieser genauere Test kann anstelle des Shafts auch gegen die – paßgenaue und damit exakte – konvexe Hülle der beiden Hierarchie-Elemente durchgeführt werden. Da die Aufstellung der konvexen Hülle zusätzlichen Aufwand bedeutet, kann für diesen Test zunächst auch der bereits ermittelte Shaft verwendet werden. Bei diesem Test werden nicht die Boundingboxes, sondern die Polygone selbst getestet, ob sie ganz oder teilweise im Shaft liegen.

Da die konvexe Hülle nur für sehr simple, eindeutig definierte Konfigurationen von genau zwei drei- oder viereckigen, planaren Flächen bestimmt werden muß,

kann ihre Berechnung sehr effizient erfolgen, ohne auf teure Standardalgorithmen zur Berechnung konvexer Hüllen angewiesen zu sein, indem man sich an den Kanten der beteiligten Polygone „entlanghangelt“ und so die konvexe Hülle sukzessive aufbaut. Durch Empfänger und Sender selbst sind bereits zwei der Grenzflächen sowie sämtliche Punkte gegeben, die die konvexe Hülle ausmachen. Die fehlenden Flächen bilden schlicht eine Sender und Empfänger verbindende „Mantelfläche“, die aus maximal acht Einzelflächen bestehen kann.

Der Vorteil einer konvexen Hülle ist sicherlich, daß der Fall der partiellen Sichtbarkeit **exakt** bestimmt werden kann, wodurch das Refinementkriterium in die Lage versetzt wird, auf Basis dieser Information eine optimierte Strategie für maximale visuelle Qualität anzuwenden.

### Totale Verdeckung

Ähnlich der Überabschätzung partieller gegenüber unverdeckter Situationen wird das Nichterkennen einer vollständigen Verdeckung keine sichtbaren Artefakte erzeugen. Es wird jedoch eine unnötige Unterteilung verursachen, wodurch Szenenkomplexität und Berechnungsaufwand vergrößert werden. Es liegt daher im Interesse einer effizienten Radiositysimulation, so viele Fälle von totaler Verdeckung zu erkennen wie möglich.

Cluster, die sich noch in der Kandidatenliste befinden, müssen sich aufgrund der *overlap open* Strategie vollständig innerhalb der konvexen Hülle befinden und können daher keine totale Verdeckung verursachen. Sie brauchen also in den folgenden Tests nicht beachtet zu werden.

Durch sukzessives Ausschließen einer totalen Verdeckung, bzw. sicheres Erkennen partiell verdeckter Situationen läßt sich schließlich der Fall der totalen Verdeckung durch ein einzelnes Polygon verhältnismäßig einfach ermitteln:

- Befindet sich ein Verdeckereckpunkt im Innern der konvexen Hülle, wird eine partielle Verdeckung verursacht. Dieses Objekt kann keine totale Verdeckung hervorrufen.
- Schneidet eine Verdeckerkante die konvexen Hülle, wird eine partielle Verdeckung verursacht. Dieses Objekt kann keine totale Verdeckung hervorrufen.
- Eine Ebenendurchdringung (partielle Verdeckung) oder daß sich Sender oder Empfänger im selben Halbraum des Verdeckers befinden (totale Sichtbarkeit) wurde bereits in den oben beschriebenen „Ebenentests“ ausgeschlossen.

Der Verdeckter befindet sich nun entweder vollständig außerhalb der konvexen Hülle, oder zerschneidet sie in zwei völlig getrennte Gebiete. Welches hiervon der Fall ist, kann nun durch einen einfachen Strahltest ermittelt werden.

### Komplexe totale Verdeckung

Außer durch ein einzelnes Polygon verursacht zu werden, existieren zwei weitere Fälle, wie eine vollständig verdeckte Situation entstehen kann.

Ob eine vollständige Verdeckung von mehreren, miteinander verbundenen Polygonen verursacht wird, kann durch Methoden des *occlusion culling* [CT96][CT97] erkannt werden.

Eine Verdeckung von mehreren, voneinander unabhängigen Polygonen ist sehr schwer sicher zu erkennen. Der Vorschlag ist deshalb, diesen Fall außer Acht zu lassen und darauf zu bauen, daß ggf. weitere Unterteilungen schnell zu einem der beschriebenen Fälle führen.

Kommt statt der beschriebenen Boundingbox-Hierarchie eine *space subdivision* Strategie zum Einsatz, können auch HOM (*hierarchical occlusion maps*) [ZMHH97] weiterhelfen.

### Visibility Bounds

Da sich für partiell verdeckte Umgebungen die Sichtbarkeitsfunktion auf dem Empfängerpatch ändern und damit Brüche oder gar Sprünge im Verlauf der Radiosityfunktion hervorrufen kann, reicht eine explizite Berechnung des sichtbaren Anteils (vgl. Kapitel 2.7.3) und deren Gültigkeitsannahme für ein ganzes Patch nicht aus, um dem Refinementkriterium eine verlässliche Fehlereingrenzung zu ermöglichen.

Daher ist es sinnvoll, entsprechend den in Kapitel 2.5 diskutierten Fehlerschranken solche Schranken ebenfalls für die partielle Sichtbarkeit zu bestimmen [GH96][HS97].

### 2.7.3 Ermittlung des sichtbaren Anteils

Da es bei der Ermittlung der prozentualen Sichtbarkeit einer bereits als partiell verdeckt klassifizierten Situation nicht auf die 100%ige Verlässlichkeit des ermittelten Wertes ankommt, ist eine Vielzahl von Sichtbarkeitsverfahren einsetzbar, wie z.B. *hemi cube*-Verfahren [CG85], *back projection* [DF94][SG94] oder die Berechnung der *anti-penumbra* [T92].

Am effizientesten und naheliegendsten ist jedoch ein einfacher Ray Casting-Ansatz. Nur für den Fall, daß das Refinement auf komplexere Sichtbarkeitsinformationen, z.B. zur Gradientenberechnung im partiell sichtbaren Fall, angewiesen ist, sind komplexere Projektions- und Clippingverfahren erforderlich, da Ray Casting keine Informationen über Größe oder Richtung des Gradienten liefern kann.

### Ray Casting

Ein einfacher, schneller und naheliegender Ansatz zur prozentualen Sichtbarkeitsabschätzung ist das bereits erwähnte Ray Casting. Da der ermittelte sichtbare Anteil nur zur Energieübertragung herangezogen wird, nicht jedoch zur Entscheidung dient, ob eine Situation partiell verdeckt ist oder nicht, ist in diesem Fall, im Gegensatz zur Sichtbarkeits*klassifikation*, ein Ray Casting Verfahren ausreichend.

Zur Effizienzsteigerung ist eine Schnittpunktroutinen-Beschleunigung wie z.B. Kay&Kajiya [KK86] mit kombinierter Goldsmith&Salmon-Hierarchie [GS87] unerlässlich.

### Projektion und Clipping

Durch Projektion der Verdeckter und Clippen dieser Projektionen gegen die Senderkontur [LN97] läßt sich der exakte Bereich des von einem bestimmten Empfängerpunkt aus noch sichtbaren Senders ermitteln. Die Prismamethode zur Formfaktorberechnung kann selbst für eine solch geclippte Kontur den korrekten Formfaktor bestimmen.

Durch Modifikation des Projektionsalgorithmus ist es sogar möglich, Tiefeninformationen zu erhalten, wodurch eine *depth clipped contour* [A94] entsteht. Für eine derartige Kontur ist sogar der analytisch exakte Formfaktorgradient ermittelbar [PC96].

Leider sind die bekannten Clipping-Verfahren numerisch sehr instabil, weshalb diese Methode nur begrenzt einsetzbar ist, auch wenn die resultierende visuelle Qualität des Ergebnisses hoch ist [K97].



### 3 Techniken zur beleuchtungstechnisch konsistenten Wiedergabe

Großformatige Stereoprojektionen haben sich heutzutage als ein wichtiges Werkzeug in der Industrie z.B. in der Produktentwicklungsphase etabliert. Der Wunsch der produzierenden Industrie ist hierbei eine Reduktion in der Anzahl teurer, realer Prototypen und damit eine signifikante Verringerung der Zeitspanne zwischen Entwicklung und Markteinführung.

In vielen Bereichen sind mit immersiven Displaytechnologien hier bereits vielversprechende Ergebnisse erzielt worden: Visualisierung von Strömungs- oder Verformungssimulationen, virtuelle Konstruktion und Begutachtung von Präßwerkzeugen und Karosserieteilen, Einbau-/Montagesimulation, und so fort.

Aber es gibt Anwendungsgebiete, die höhere und auch andere Anforderungen an die Bildqualität haben, als einfach nur „nette Bilder“ darzustellen. Ergonomiestudien sind auf konsistente Darstellung von reproduzierbaren Farben und Helligkeiten angewiesen, um verwertbare Ergebnisse zu liefern. Designentscheidungen werden basierend auf dem visuellen Eindruck getroffen, den ein bestimmtes Design hervorruft, weshalb die auf den unterschiedlichen Darstellungsmedien präsentierten Bilder und VR-Umgebungen so nahe wie möglich an der Realität liegen müssen, um von Designern akzeptiert zu werden. Die für fundierte Entscheidungen benötigte Qualität kann von einfachen, abstrakten Displays nicht geliefert werden.

Durch die Radiositysimulation ist ein Werkzeug geschaffen worden, den visuellen Realismus computergenerierter Szenen drastisch zu verbessern. Doch was nützt ein „schönes buntes Bild“, wenn keinerlei Aussagen über Verlässlichkeit und Korrektheit der präsentierten Farben oder Helligkeiten im Vergleich zum realen Objekt getroffen werden können, oder das Display z.B. gar nicht in der Lage ist, die geforderten Leuchtdichten darzustellen?

Um Designern und Entwicklern bei Entscheidungen helfen zu können, ob z.B. eine Reflexion in der Tat störend ist oder nur durch das Display übertrieben dargestellt ist, ist eine Sicherstellung der 1:1 Wiedergabe von Farben und Helligkeiten zwingend erforderlich.

Das vorige Kapitel beschäftigte sich unter anderem mit der Sicherstellung der photometrischen und colorimetrischen Konsistenz innerhalb der Radiosity-Simulationspipeline. Hier werden nun mögliche Ausgabegeräte näher betrachtet und Methoden zur verlässlichen Kalibrierung vorgestellt, wodurch das Ergebnis

einer Radiositysimulation photometrisch reproduzierbar präsentiert werden kann.

Eine Vielzahl unterschiedlicher Displaytechnologien stehen heute für VR und sonstige Anwendungen zur Verfügung, angefangen beim *head mounted display* bis hin zu einer voll-immersiven Umgebung in einer CAVE [CSD93]. Abgesehen von Unterschieden in Auflösung, Blickfeldgröße, Bildschirmdurchmesser und mehreren anderen Charakteristiken [BRB98] ist eine Haupteigenschaft, die diese verschiedenen Ausgabemedien unterscheidet, ihre maximale Helligkeit (genauer: Leuchtdichte), und damit auch der maximal zur Verfügung stehende Kontrast. Neben Einflüssen wie Szenendetail und Realismus innerhalb einer VR-Szene tragen diese Eigenschaften entscheidend zum visuellen Realismus eines virtuellen Szenarios bei.

Nur wenn sich die Leuchtdichtenbereiche einer realen Szene und der darstellbare Bereich eines immersiven Displays decken, kann dieses Display für quantitative Untersuchungen einer solchen Szene herangezogen werden. Ist das Display nicht in der Lage, die geforderten Leuchtdichten darzustellen, nützt selbstverständlich auch eine korrekte Simulation wenig (man kann sich in diesen Fällen zwar mit Tonemapping-Verfahren auf den ersten Blick helfen – der tatsächlich wahrgenommene Eindruck unterscheidet sich dann jedoch wieder von der Realität).

In vielen der für die fünfseitige CAVE des Fraunhofer-IGD entwickelten Anwendungen [U98] wird deutlich, daß Farbsättigung und Kontrast deutlich schlechter sind im Vergleich zu einem Monitor oder sogar einer einfachen Großbildprojektion. Um dem entgegenzuwirken, werden oft extreme Farben innerhalb der VR-Szene verwendet, um beim Betrachten der 3D-Szene noch etwas Farbbrillanz zu bekommen.

Statt die Wirkungen ohne genaue Kenntnis der Ursachen beseitigen zu wollen, befaßt sich dieses Kapitel damit, die genauen Ursachen von Phänomenen wie Kontrastverlust, reduzierter Farbwahrnehmung usw. zu untersuchen, und Methoden vorzuschlagen, wie diese Eigenschaften effektiv verbessert werden können.

Dazu werden zunächst unterschiedlichste Ausgabemedien speziell auf ihre lichttechnischen und optischen Eigenschaften hin untersucht. Sodann werden die Ursachen für diejenigen Faktoren ermittelt, welche die visuelle Qualität und lichttechnische Verwendbarkeit einschränken, sowie Lösungsmöglichkeiten vorgestellt, die Qualität immersiver Displaysysteme zu verbessern und sie so photometrisch und colorimetrisch zu kalibrieren. Des Weiteren wird ein automatisiertes Verfahren zum Helligkeitsausgleich und zur Farbkalibrierung vorgestellt.



## 3.1 Lichttechnische Eigenschaften von immersiven Displaysystemen

Während hier nur die grundlegenden Eigenschaften, Prinzipien und Aspekte der einzelnen Komponenten eines Displaysystems vorgestellt und erläutert werden [KR02], befaßt sich Kapitel 3.2 mit einer genaueren Analyse der photometrischen und farblichen Eigenschaften der einzelnen Technologien.

### 3.1.1 Physikalische Eigenschaften

In diesem Abschnitt werden kurz die grundlegenden physikalischen Eigenschaften von Projektionssystemen beschrieben, die direkt die Bildqualität beeinflussen. Es wird analysiert, warum diese Eigenschaften so wichtig sind, und von welchen Teilen des Systems (Projektor, Leinwand, Stereotechnologie, ...) sie beeinflusst werden können.

#### Leuchtdichte

Die Leuchtdichte eines Displays bestimmt dessen maximal darstellbare Helligkeit. Während der Gesamtlichtstrom z.B. eines Projektors meist in ANSI-Lumen gegeben ist, wird die Leuchtdichte in Candela pro Quadratmeter ( $\text{cd}/\text{m}^2$ ) gemessen und beschreibt die tatsächlich von Auge wahrgenommene Helligkeit, die eine Oberfläche abgibt.

Für Lichtdesigner ist die quantitative Leuchtdichte eines simulierten Objektes wichtig, um Ergonomieentscheidungen treffen zu können. Nur wenn das Display in der Lage ist, den Leuchtdichtebereich des simulierten, realen Objektes wiederzugeben, sind solche Entscheidungen auf Basis von rein virtuellen Modellen möglich.

Die maximale Leuchtdichte hängt direkt von der Lampenleistung und der Bildfläche ab und wird von Faktoren wie Filtern, Umspiegelung, Leinwandreflexion und Stereobrillenabsorption verringert.

#### Farbraum

Der Farbraum, oder *color gamut* beschreibt die darstellbaren Farben und hängt allein von den verwendeten Phosphoren, bzw. Lampenspektrum und Farbfiltern der Primärfarben des Projektors ab. Für CRT-Röhrenprojektoren sind diese Primärfarben wie bei Monitoren üblicherweise relativ ähnlich, und bei den meisten Displaygeräten ist der Farbraum eine vergleichsweise kleine Teilmenge der exis-

tierenden Farben, wie sie im CIE *chromaticity diagram* auf Basis der reinen Wellenlängen definiert sind.

Auf einem Displaysystem, das nicht farbkalibriert ist, lassen sich noch nicht einmal so einfache „Farben“ wie z.B. Weiß konsistent darstellen.

## Highlights

Es existieren generell zwei Arten von Highlights, mit unterschiedlichen Ursachen und Eigenschaften.

Der erste Typ ist ein blickpunktunabhängiges, vollkommen diffuses Highlight nahe der Displaymitte, dessen Ursache einfach darin begründet liegt, daß unterschiedliche Punkte auf dem Schirm unterschiedliche Abstände zur Projektionslinse, sowie unterschiedliche Winkel zur optischen Achse besitzen. Z.B. ergibt sich bei einem Projektionsabstand der einfachen Bildschirmdiagonale ein Streckenverhältnis von Projektionseckpunkt zu Mittelpunkt von  $r = \frac{\sqrt{5}}{2} : 1$ , was zusammen mit der sich daraus ergebenden Winkelabweichung von  $\varphi = \arccos(\frac{1}{r}) = 26.6^\circ$  zu einer Lichtminderung der Ecken auf  $\frac{\cos \varphi}{r^2} = \frac{1}{r^3} = 72\%$  gegenüber dem Mittelpunkt führt. Hinzu kommen noch gleichartige Verluste im Strahlengang des Projektors.

Highlights des zweiten Typs sind spiegelnder Natur – die sog. *Glanzlichter*, welche blickpunktabhängig sind. Sie sind eine reine Eigenschaft des Leinwandmaterials und daher unabhängig von der verwendeten Projektionstechnologie. Allerdings ergeben sich insbesondere bei CRT-Projektoren durch die baulich getrennten Röhren der Primärfarben oft entsprechend eingefärbte, räumlich versetzte Highlights. Diese Form der Highlights ist aufgrund ihrer Blickpunktabhängigkeit schwer in den Griff zu bekommen.

Insbesondere kann es je nach verwendetem Leinwandtyp auch dazu kommen, daß die Winkelabhängigkeit – und damit die Größe des Highlights – zusätzlich von der Wellenlänge abhängt, wodurch die Highlights Farbsäume bekommen können.

Wird mehr als ein Projektor pro Projektionsfläche verwendet (sog. *tiled display*), ergibt sich der zusätzliche Nachteil einer verstärkten Leuchtdichtedifferenz an den Berührungskanten, in Abhängigkeit von der Beobachterposition. Eine sorgfältige Wahl des korrekten Leinwandmaterials ist daher von entscheidender Bedeutung für ein photometrisch konsistentes Display.

## Interreflexion

Interreflexionen zwischen nicht koplanaren Leinwandflächen sind ein bekanntes Phänomen von CAVEs und reduzieren sowohl Kontrast, als auch Farbsättigung. Des weiteren wird der Schwarzpunkt des Displays erhöht – teilweise so stark, daß ein „schwarzes“ Bild, das indirekt von 4 hellen, benachbarten Leinwänden angestrahlt wird eine fast genauso hohe Leuchtdichte besitzt, wie ein einzelnes weißes Bild (vgl. Kapitel 3.1.5). Die genaueren Einflüsse zur Thematik Interreflexion werden in Kapitel 3.2.4 näher untersucht.

Das Problem der Interreflexion kann ausschließlich durch spezielle Leinwandmaterialien, wie z.B. „schwarze Leinwände“ verringert werden und ist insbesondere **unabhängig von der Größe der Projektionsflächen**.

## Kontrast

Das Verhältnis eines vollständig weißen zu einem vollständig schwarzen Bild ist der Bildkontrast. Je besser das „reine Schwarz“ eines Projektors ist, desto höher ist der Kontrast, und desto besser ist die Bildqualität. Umgebungslicht, Interreflexion von anderen Leinwänden und schlechte Qualität der Projektoroptik reduzieren den Kontrast, wobei die beiden zuerst genannten Faktoren von der Reflektivität der Leinwand abhängen.

### 3.1.2 Stereotechnologien

Die Hauptaufgabe von stereoskopischen Systemen ist die Projektion eines Stereobildpaares auf eine Leinwand, wobei jedoch nur jeweils eins der Bilder pro Auge wahrgenommen werden darf. Hierfür gibt es zwei prinzipiell unterschiedliche Ansätze: entweder durch den Einsatz von zwei Projektoren und der Verwendung von entsprechenden Filtern auf Basis verschiedener Prinzipien sowohl auf Projektorseite als auch bei den dann passiven Stereobrillen, oder indem in schneller Folge abwechselnd von einem Projektor zwei unterschiedliche Bilder projiziert werden und synchron dazu die aktiven Brillengläser hell und dunkel schalten.

Abgesehen von frühen, bedingt erfolgreichen Experimenten mit rot-/grün-Filtern waren für Großbildprojektionen lange Zeit lineare Polarisierung und Shuttering die einzig möglichen Alternativen. In den letzten Jahren sind jedoch zwei wichtige neue Technologien hinzugekommen.

Im folgenden werden die verschiedenen Ansätze mit ihren spezifischen Vor- und Nachteilen näher betrachtet.

## Lineare Polarisisation

Durch Aufbringen von linearen Polarisationsfiltern vor den Projektionslinsen, sowie passend dazu orientierten Filtern als Stereobrille geschieht die Kanaltrennung bei dieser wohl am weitesten verbreitetsten Stereotechnik, die z.B. auch in kommerziellen IMAX-Kinos angewandt wird.. Ihr Hauptvorteil liegt in der Verwendung von sehr günstigen Stereobrillen, wobei allerdings zwei Projektoren erforderlich sind, um ein Bild zu erzeugen.

Durch die Winkelabhängigkeit geht jedoch der Stereoeindruck beim Neigen des Kopfes verloren, wodurch diese Technologie für CAVEs und andere nicht koplare Multiscreen-Projektionen nicht verwendbar ist.

Des weiteren erfordert die Erhaltung der Polarisisation spezielle Leinwandmaterialien, was speziell bei Aufprojektionen, aber auch bei Kachelprojektionen zu erheblichen Problemen führt, wie im nächsten Abschnitt gezeigt wird.

## Zirkulare Polarisisation

Als Erweiterung der linearen Polarisisationstechnologie erlaubt die in den letzten Jahren entwickelte zirkulare Stereoprojektion das Neigen des Kopfes, da zirkular polarisiertes Licht von Drehungen um die optische Achse unbeeinträchtigt bleibt. Dadurch ist es möglich geworden, passive Stereotechniken erstmals auch in CAVEs einzusetzen.

Wie bei der linearen Polarisisation ist jedoch auch diese Technologie auf spezielle Leinwandmaterialien angewiesen.

## Aktives Shuttering

Das zweite der weit verbreiteten Stereoprinzipien beruht auf der Verwendung von aktiven LCD-Brillengläsern, die durch abwechselndes Hell- und Dunkel-schalten von je einem Auge synchron zu einem entsprechend wechselndem Videosignal eine Kanaltrennung erzeugen. Solche Stereobrillen sind verhältnismäßig teuer, der Vorteil ist jedoch, daß bei dieser Technik nur ein einzelner Projektor erforderlich ist.

Der Gesamtlichtverlust liegt interessanterweise in derselben Größenordnung wie bei den Polarisisationstechniken. Während dort das Licht, von zwei Projektoren kommend, über zwei Stufen von Polarisationsfiltern zum Auge gelangt, ist es hier jedoch nur ein einziger Projektor, der das Licht durch die Shutterbrille auf beide Augen verteilen muß. Bei gleicher Projektorleistung ergibt also die aktive Shuttertechnik ein dunkleres Bild.

Die Funktion des Shutter-Stereo ist vom Leinwandmaterial unabhängig, und es ist daher die einzige Stereotechnologie, die auch bei handelsüblichen (CRT-)Monitoren anwendbar ist. Als problematisch können sich jedoch langsame Phosphore erweisen, die bei Umschaltfrequenzen von 100Hz und mehr durch zu langes Nachleuchten dazu führen, daß ein Übersprechen des für ein Auge bestimmten Bildes in das andere auftritt.

### **Infitec**

Ein weiteres, neueres passives Stereosystem ist das von Daimler-Chrysler entwickelte Infitec-Verfahren [J02]. Hier erfolgt die Trennung der beiden Stereokanäle nicht mit Hilfe von unterschiedlich polarisiertem Licht, sondern durch die Verwendung von schmalbandigen Spektralfiltern. Diese Interferenzfilter (INTERferenz-FILTER-TECHnologie) besitzen pro Auge leicht unterschiedliche, getrennte Durchlaßbereiche für jede der drei Primärvalenzen.

Die Vorteile liegen zum einen in einer Rotationsunabhängigkeit entlang der optischen Achse. Des weiteren werden insbesondere keine speziellen Anforderungen an das Leinwandmaterial gestellt. Und schließlich ist durch die hohe Effektivität von Interferenzfiltern die Kanaltrennung ungleich besser gegenüber polarisierenden Verfahren (wozu auch das Shuttering zählt).

Der Gesamtarbeindruck wird bei diesem Verfahren jedoch aufgrund der sich unterscheidenden Spektralbereiche für das linke Auge ins rötliche, für das rechte ins grünliche hinein verfälscht, sodass dieses Stereoprinzip für farblich konsistente Bildwiedergabe ohne zusätzliche Ausgleichstechniken unbrauchbar ist. Das Meß- und Kalibrierungsverfahren, das colorimetrisch konsistente Darstellung auch mit Infitec ermöglicht, wird in Kapitel 3.2.5 vorgestellt.

Sowohl bei CAVE-ähnlichen Umgebungen, als insbesondere auch bei Kachelprojektionen wird der wesentliche Vorteil von Infitec deutlich, da die Eigenschaften der Leinwand – neben dem Projektor der wichtigste Faktor für eine hohe Bildqualität – frei gewählt werden können und nur so der Einsatz von z.B. volldiffusen Leinwänden ermöglicht wird (vgl. Kapitel 3.3.8).

Der reine Lichtverlust aufgrund projektorseitigen Filtern und Stereobrillen ist vergleichbar mit dem anderer Stereoverfahren. Der erforderliche Farbausgleich reduziert jedoch die Gesamtlichtausbeute nochmals um einen wesentlichen Faktor.

Die nutzbare Gesamthelligkeit hängt hierbei jedoch sehr stark von dem Lampenspektrum ab: heutzutage in den meisten digitalen Projektoren eingesetzte Quecksilber-Hochdrucklampen erbringen z.B. im langwelligen Rotbereich nur eine sehr geringe Lichtausbeute. Da hier das rote Spektralband des rechten Au-

ges angesiedelt ist, wird die somit für einen vollständigen Farbausgleich verfügbare Rotintensität extrem reduziert, womit auch die Gesamtlichtausbeute entsprechend verringert wird.

Interferenzfilter erlauben aufgrund ihres physikalischen Prinzips nur verhältnismäßig geringe Abweichungen von der optischen Achse, bevor sich ihr Verhalten umkehrt und die eigentlich zu transmittierenden Spektralbereiche reflektiert werden, und umgekehrt. Bei heute verfügbare Stereobrillen treten hierbei in den Außenbereichen der Filter – trotz bereits leicht gewölbter Gläser – Farbverfälschungen und sogar Kanalwechsel auf. Dieses Problem kann jedoch durch ein Redesign der Stereobrillen und Verwendung von stärker gekrümmten Filtern beseitigt werden.

Es muß angemerkt werden, daß aufgrund von Abweichungen individueller Beobachter gegenüber dem CIE-Standardbeobachter (Abb. 6) ein für jeden Betrachter gültiger Farbausgleich ausgeschlossen ist – bei Infitec noch mehr als bei einfachem Farbausgleich. Durch die schmalbandige Spektralselektion der Filter ist es möglich, daß Farben von denen sowohl das Messgerät (welches auf dem Standardbeobachter basiert), als auch ein normalsichtiger Beobachter sagt sie seien identisch, von farbfehlsichtigen Personen als deutlich unterschiedlich wahrgenommen werden. Es kann sogar Situationen geben, in denen ein Normalsichtiger das von dem einen Projektor dargestellte Bild als „grünlich“, und das des anderen als „rötlich“ beschreibt – während der Farbfehlsichtige (im konkreten Fall aufgrund von Protanomalie) das genaue Gegenteil sieht!

## Weitere Techniken

Es existieren weitere Verfahren zur Stereodarstellung, die jedoch meist von geringer Bedeutung sind, da sie entweder die Farbwahrnehmung zerstören oder zumindest stark einschränken, oder oft sehr gewöhnungsbedürftig sind.

Die einfachste (und älteste: [W1838]) Variante der Stereodarstellung geschieht einfach auf der Nebeneinanderdarstellung zweier Bilder, wobei das Auge, bzw. der **Konvergenzmechanismus**, selbst dafür sorgen muß, je nach Anordnung der beiden Bilder entweder durch Überkreuzen des Blicks (Schielen) oder entsprechendes "Hindurchsehen" die Bilder zur Deckung zu bringen. Es ist möglich, die hierbei erforderliche unnatürliche Betrachtungsweise durch ein Linsensystem („Stereoskop“) zu kompensieren, das sich jedoch in festem Bezug zur Bildebene befinden muß.

Ein weiteres, jedoch ungleich neueres Verfahren, das sich Konvergenzänderungen zunutze macht und dabei mit nur einem Bild auskommt sind die **SIRDS**, oder *single image random dot stereograms* [J71]. Hierbei wird per Computer zeilenweise mit einem Randpunkt begonnen und für das rechte Auge ein Schnittpunkt auf

dem Objekt bestimmt. Für diesen Punkt wird für das linke Auge der Schnittpunkt mit der Bildebene berechnet, für den wiederum für das rechte Auge ein Schnittpunkt auf dem Objekt bestimmt wird, und so fort (Abb. 10). Konvergiert nun das Auge die so entstandene Punkte außerhalb der Bildebene, so entsteht ein dreidimensionales Bild.

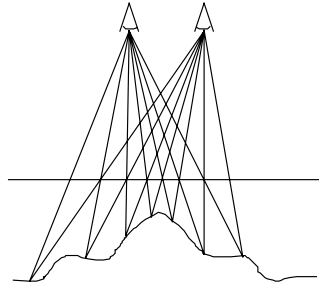


Abb. 10 : Konstruktion eines SIRDS

Die in verschiedenen Varianten vorkommenden **anaglyphischen Stereoverfahren**, wie z.B. Rot-Grün (1922), Rot-Cyan oder AmberBlue, können vielleicht als ferne Vorläufer des Infitec-Verfahrens gesehen werden, da sie letztendlich auch auf Kanaltrennung durch disjunkte Spektralbänder basieren. Im Gegensatz zu den bei Infitec verwendeten drei jeweils nur leicht verschobenen Einzelbändern gibt es hier jedoch nur jeweils ein einziges, sehr breites Band.

Der Vorteil, den anaglyphische Stereoverfahren jedoch bieten, ist die Verwendung nur eines einzigen Ausgabemediums für beide Bilder gleichzeitig, wodurch eine Verwendung dieser Verfahren in Bereichen wie z.B. Fernsehen oder Film möglich wurde. Des weiteren eignen sich diese Technologien insbesondere auch für Printmedien. Insbesondere AmberBlue [C02] hat den Vorteil, daß man auch ohne Brille einen relativ guten, farblich sinnvollen 2D-Bildeindruck hat.

Eine andere Methode, die ebenfalls in Fernsehmedien testweise zum Einsatz kam nutzt das Prinzip zeitlicher Wahrnehmungsunterschiede aufgrund unterschiedlicher Helligkeiten, den Pulfrich-Effekt [P22]. Die zugehörige Stereobrille besitzt zwei farblich zwar leicht unterschiedliche Filter, deren Haupteigenschaft jedoch einfach die unterschiedliche Gesamttransmission ist. Durch die so induzierte **verzögerte Wahrnehmung** der beiden Stereokanäle erscheinen sich von links nach rechts bewegende Objekte näher zum Betrachter als sich von rechts nach links bewegende, da das gefilterte Bild später im Gehirn ankommt als das ungefilterte. Stillstehende Objekte liegen in der Bildebene. Diese Prinzip funktioniert daher nur bei Bewegungsbildern und nur bei speziellen Kamerafahrten.

„Wiggle-Bilder“ [G03b] stellen eine weitere, experimentelle Stereotechnik dar. Hier werden die Stereobilder zeitlich alternierend ohne jegliche Filter an dersel-

ben Position dargestellt. Bei einer bestimmten Frequenz (ca. 1.5-3 Hz) entsteht hierdurch ein erstaunlich guter Stereoeindruck. Ist die Frequenz zu hoch, verwischen die Bilder zu einem, ist sie zu niedrig verliert sich der Stereoeindruck.

Heutige LCD-Monitore generieren per Definition polarisiertes Licht. Es gibt experimentelle Ansätze, durch Aufbringen von einfachem Cellophanpapier durch dessen physikalische Eigenschaften die Polarisation zu drehen, um so zwei senkrecht aufeinanderstehende Polarisationsrichtungen zu erzeugen. Mit normalen Stereobrillen auf Basis linearer Polarisation kann nun ein Stereoeindruck erzeugt werden. Hierdurch entstehen jedoch nur zwei nebeneinanderliegende Stereobilder, die mit dem Konvergenzverfahren zur Deckung gebracht werden müssen. Versuche, durch streifenweises Aufbringen von Cellophanfolie ein Interlace-Stereobild zu erzeugen gibt es noch nicht.

### 3.1.3 Projektortechnologien

Für lange Zeit waren CRT- oder Röhrenprojektoren die einzige Alternative, wenn ein Projektionsdisplay zu realisieren war. Gerade in den letzten Jahren gab es jedoch entscheidende, neue Technologien im Bereich Projektortechnik, wodurch die Frage nach der für immersive Projektionssysteme am geeignetsten Technologie plötzlich an Wichtigkeit gewann.

Die einzelnen Technologien werden hier vorgestellt und in Bezug auf ihre für immersive Displaysetups relevanten Eigenschaften hin untersucht.

#### CRT

CRT(*cathode ray tube*)-Projektoren sind die klassischen Arbeitspferde der bildgebenden Ausgabemedien. Es ist die älteste, und doch bis vor ein, zwei Jahren immer noch am weitesten verbreitete Projektionsdisplaytechnologie. Das Bild wird, ähnlich dem Monitor oder Fernseher, durch Elektronenstrahlen erzeugt, die auf entsprechend phosphorisierte Teile der Elektronenröhren treffen. Dabei ist es, nicht zuletzt für eine höhere Auflösung und Bildschärfe üblich, pro Primärfarbe eine eigene, monochromatische Röhre einzusetzen.

Die CRT-Technik hat, nicht zuletzt aufgrund ihrer Ausgereiftheit, mehrere Vorteile. Da das Bild durch wechselndes An- und Abschalten des Elektronenstrahls erzeugt wird, ist der erreichbare Kontrast extrem hoch. Des Weiteren sind CRT Bildpunkte über ihre gesamte Fläche beleuchtet, was Röhrenprojektoren zu einem ansprechenden, gleichmäßig ausgeleuchteten Bild verhilft. Da der Elektronenstrahl elektronisch kontrolliert werden kann, erlauben CRTs beinahe grenzenlose Flexibilität in Bezug auf Bildentzerrung, und damit Projektorpositionierung und Leinwandausrichtung bzw. -form. Die Bildauflösung kann frei gewählt



werden und ist unabhängig von einem vordefinierten Pixelraster, wodurch die maximale Auflösung sehr hoch sein kann und prinzipiell nur von der Qualität und Leistungsfähigkeit der Ablenkelektronik abhängt. Bildwiederholraten sind hoch genug, um aktives Shutter-Stereo zu gestatten, und das projizierte Licht ist unpolarisiert, was passive Stereotechniken durch Vorsatz von Polarisationsfiltern ermöglicht.

Pixel derselben Bildzeile werden kontinuierlich erzeugt, wodurch Aliasing-Artefakte stark reduziert werden. Dies wird häufig als Vorteil gegenüber anderen Technologien gesehen, bei denen oft deutlich quadratische, räumlich getrennte Einzelpixel zu erkennen sind.

Der Hauptnachteil eines CRT ist seine geringe Lichtausbeute. Röhrenprojektoren sind mit einer typischen Maximalhelligkeit von etwa 250-300 ANSI-Lumen vergleichsweise dunkel und erfordern dadurch meist abgedunkelte Räume, wodurch die Einsatzgebiete stark eingeschränkt werden. Der Vorteil einer feinen und flexiblen Justierbarkeit ist gleichzeitig auch ein Nachteil, da die Justierung sehr aufwendig ist und sich im Laufe der Zeit verstellen kann, weshalb sie in regelmäßigen Abständen nachgeregelt werden muß.

Insgesamt gesehen erzeugen CRTs hochwertige, einheitliche Bilder, aber sind sehr lichtschwach (etwa ein bis eineinhalb Größenordnungen) im Vergleich mit anderen Technologien.

## LCD

LCD(*liquid crystal display*)-Projektoren erlangen mehr und mehr Bedeutung, da sie kompakt gebaut und vergleichsweise günstig hergestellt werden können. Um das projizierte Bild zu erzeugen, durchleuchtet eine helle Lichtquelle ein LCD-Display, wobei das Licht von dunkel geschalteten Bereichen absorbiert wird. Hochwertige Projektoren verwenden drei monochrome, mit Farbfiltern versehene LCD-Einheiten zur Farbwiedergabe, deren Strahlengänge vor der Projektion rekombiniert werden. Da diese Projektoren im Vergleich zu CRTs externe Lichtquellen verwenden, können sie sehr leuchtstark sein. Typische Helligkeiten von heutigen LCD-Projektoren liegen im Bereich von 4000-7000 ANSI-Lumen, erste High-End-Varianten mit 10000 ANSI-Lumen sind erhältlich.

Das Hauptproblem bei LCD-Projektoren ist der geringe Kontrast. Lichtabschwächung in den LCD-Einheiten geschieht auf Basis verschränkter Polarisationsfilter, die kein perfektes Schwarz erzielen können – ein nicht zu vernachlässigender Anteil des Lichtes kann selbst die geschlossenen Filter durchdringen. Das Ergebnis ist eher ein verwaschenes Grau als ein wirkliches Schwarz. Diese physikalische Eigenart macht sich besonders bei gekachelten Displays aus mehreren

angrenzenden Projektoren bemerkbar, da sich das „Schwarz“ benachbarter Projektoren überlagert und damit aufaddiert – ein Effekt, der sich elektronisch nicht verringern läßt, wodurch sichtbare, graue Ränder entstehen.

Einzelne Pixel sind, aufgrund Leiterbahnen zur Ansteuerung, als solche erkennbar und von schwarzen Stegen umgeben, was sich vor allem bei großen Projektionen mit stark vergrößerten Bildpunkten störend bemerkbar macht.

Ein weiteres Problem ist Geschwindigkeit. LCDs sind verhältnismäßig träge; Bildpunkte können nicht schnell genug zwischen hell und dunkel wechseln, so dass dieser Projektorentyp für aktives Shuttering ungeeignet ist. Da das von LCDs projizierte Licht per Definition bereits linear polarisiert ist, ist es schwierig sie für passiven Stereobetrieb einzusetzen, bei dem die Polarisationsrichtung zweier überlagerter Projektoren um exakt 90 Grad gedreht sein muß.

Als Fazit muß man sagen, daß LCD-Projektoren zwar lichtstark, aber durch das ungenügende Schwarz sehr kontrastarm sind. Einsatz im Stereobetrieb ist nur eingeschränkt möglich.

## D-ILA

D-ILA(*direct drive image light amplifier*)-Projektoren stellen eine Weiterentwicklung der LCD-Technik dar. Als Basis dient auch hier eine LCD-Einheit, jedoch ist diese auf einer spiegelnden Chip-Oberfläche aufgebracht, wodurch reflektiertes Licht zweimal durch die LCD-Einheit fallen muß. Dadurch wird das Problem des ungenügenden Schwarzlevels etwas abgeschwächt, jedoch nicht eliminiert. Durch die Chip-basierte Technologie (LCoS, *liquid crystal on silicon*) sind etablierte Fertigungstechniken, hohe Baudichte und damit entsprechend große Auflösungen möglich. Da die einzelnen Pixel direkt von hinten über die Siliziumschicht angesteuert werden, ist es möglich, die Stege zwischen den Pixeln relativ begrenzt zu halten.

Die Stereoverwendbarkeit bleibt unverbessert, da es sich immer noch um eine LCD-Technik handelt.

## DLP

DLP(*digital light processing*)-Projektoren verwenden ebenfalls einen reflexiven Ansatz, jedoch kommen hier mechanische Spiegel in Form von *digital mirror devices* (DMD) zum Einsatz. Das Grundprinzip sind hier mikroskopische Spiegel auf Siliziumchips, die einzeln digital angesteuert werden können.

Da diese Spiegel binäre Schalter darstellen, müssen Helligkeitsabstufungen durch unterschiedliche Zeitintervalle simuliert werden. Durch die Farbneutrali-

tät sind zur Farbdarstellung entweder drei DMD-Chips erforderlich, oder es kommt ein Farbrad zum Einsatz, sodass nacheinander die Farben Rot, Grün und Blau dargestellt werden. Solche Einchip-DLPs sind für Shutteringtechniken ungeeignet. Ein bescheidener Vorteil ist, daß Konvergenzprobleme ausgeschlossen sind – natürlich nur sofern die Optik keine chromatische Aberration aufweist.

Qualitativ gesehen befinden sich DLPs zwischen D-ILA und LCD. Die Ränder um einzelne Pixel sind stärker ausgeprägt, aber das Schwarz ist aufgrund von Restlichtstreuung kaum besser als bei LCDs. Sie können deutlich schneller schalten als die polarisationsbasierten Systeme, wodurch kürzlich die ersten für aktives Shuttering geeigneten DLP-Projektoren auf dem Markt erschienen. Da DMDs das Licht nicht (merklich) polarisieren, ist dieser Projektortyp daher für alle Stereotechnologien geeignet.

Die Lichtabschwächung der DMDs ist sehr gering, weshalb DLP-Projektoren die leuchtstärksten Projektortypen darstellen und Helligkeiten bis etwa 20000 ANSI-Lumen erreicht werden.

Bei Einchip-DLPs mit Farbrad bedient man sich manchmal eines Tricks, um zusätzliche Helligkeit zu gewinnen. In das Farbrad wird ein viertes, filterloses Feld eingefügt, das bei hellen Farben zusätzliches weißes Licht durchlässt. Die Farbwiedergabe ist hierdurch jedoch in Abhängigkeit von allen drei Farbkanälen nichtlinear, daher lässt sich dieser Effekt nicht durch eine einfache 3x3-Farbmatrix ausgleichen [S01c].

Interessant ist zu erwähnen, daß Einchip-DLPs für das Infitec-Verfahren optimiert werden können, indem die Farbfilter des 3er-Farbrades durch entsprechende Interferenzfilter mit je einem der Primärvalenz-Spektralbänder der Infitecfilter ersetzt werden, was einen deutlichen Helligkeitsgewinn zur Folge hat.

## **Laser**

Eine sich gerade entwickelnde, neue Technologie stellen echte Laserprojektoren dar. Hier kommen drei Laser in den Primärfarben, sowie rotierende Prismenspiegel zur Bilderzeugung zum Einsatz.

Die offensichtlichen Vorteile dieses Prinzips sind ein maximal vergrößerter Farbraum, der jegliche andere Displaytechnologie in den Schatten stellt und bis dato unerreichte Farbbrillanz ermöglicht; außerdem unlimitierte Schärfentiefe, und damit die Möglichkeit zur Projektion auf beliebig geformte oder gar bewegliche oder halbtransparente Oberflächen. Hinzu kommt eine extrem hohe Farbstabilität und -konsistenz durch die Verwendung von Lasern mit durch physikalische Eigenschaften festgelegten Wellenlängen. Von der Lasereinheit zum Projekti-

onskopf wird das Licht der drei Laser in einer einzigen Glasfaser übertragen, wodurch keine Konvergenzprobleme auftreten können.

Der stark vergrößerte Farbraum erlaubt die nahezu exakte Reproduktion sogar von so extremen Lichtquellen wie LEDs mit festen Wellenlängen z.B. in Fahrzeugarmaturen. Wie beim Röhrenprojektor ist prinzipiell perfektes Schwarz, und damit sehr hoher Kontrast, durch Strahlabschaltung in dunklen Bereichen erzielbar.

Die maximale Leuchtdichte von Laserprojektoren ist, insbesondere aufgrund der neuartigen Technologie speziell des blauen Lasers, zur Zeit noch niedriger als man vermutlich erwarten würde, liegt jedoch trotzdem bereits im Bereich leistungsstarker digitaler Projektoren (vgl. Kapitel 3.1.5).

Bei der Erstellung dieser Arbeit waren Laserprojektoren nur für monoskopische Displays verfügbar, jedoch gibt es keine grundsätzlichen Einschränkungen, die einen Einsatz für Shuttering in Frage stellen würden. Da Laserlicht von Natur aus polarisiert ist, wird sich der Einsatz von passiven Stereotechnologien wohl ähnlich schwierig wie beim LCD-Projektor gestalten – es sei denn man kann die Polarisationsrichtung kontrollieren, wodurch dann sogar vorgeschaltete Polarisationsfilter entfallen würden. Für die Infitec-Stereotechnologie ist ein Laserprojektor optimal geeignet, da hier die stark lichtabschwächenden Interferenzfilter entfallen können. Voraussetzung ist, daß für die beiden Stereobilder Laser mit unterschiedlichen Wellenlängen für jede Primärfarbe eingesetzt werden – andernfalls ist Infitec hier nicht anwendbar.

Laser-Projektoren haben erst kürzlich die prototypische Produktphase verlassen und sind daher noch extrem teuer und unhandlich. Ihre immensen Vorteile lassen aber auf eine fortschreitende Verbesserung dieser Technologie und weite Verbreitung hoffen.

### 3.1.4 Leinwandmaterialien

Die Wichtigkeit von Qualität und optischen Eigenschaften des Leinwandmaterials wird oft vernachlässigt. Ein möglichst diffus reflektierender Projektionschirm ist notwendig, um eine gleichmäßige Ausleuchtung der gesamten Schirmfläche zu ermöglichen und blickpunktabhängige Highlights und sichtbare Helligkeitssprünge bei gekachelten Projektionen zu eliminieren. Auf der anderen Seite zerstört eine rein diffuse Leinwand die Polarisation von Licht, was sie für auf Polarisation basierende Stereotechnologien unbrauchbar macht.

Ein hoher Reflexionskoeffizient (d.h., eine weiße Leinwand) reduziert den Kontrast, insbesondere durch Interreflexion von nicht koplanaren Nachbarleinwänden und Streuung von Umgebungslicht.

Eine Auswahl von unterschiedlichen Leinwandtechniken soll hier bezüglich ihrer Aspekte zur konsistenten Bildwiedergabe diskutiert werden.

### Aufprojektion und Polarisation

Ein besonderes Problem stellt sich, wenn Polarisationstechnik und Aufprojektion kombiniert werden sollen. Erhaltung der Polarisationsrichtung bedeutet insbesondere in diesem Fall einen hohen Anteil an spiegelnder Reflexion, weshalb meist silbrig beschichtete Materialien für ein solches System verwendet werden. Dadurch entstehen jedoch auch gravierende, blickpunktabhängige Highlights, sodass sich Aufprojektion und Polarisationstechnik im Prinzip gegenseitig ausschließen und für photometrisch konsistente Bildwiedergabe unbrauchbar sind.

### Fresnellinsenprojektionen

Als Fresnellinsen geschliffene Glasmattscheiben dienen bei Rückprojektionen dazu, die Helligkeit auf die vorgegebene Position eines potentiellen Betrachters zu bündeln, um dort eine höhere Lichtausbeute (*gain*) zu erlangen. Dies ist jedoch auch gleichbedeutend mit einer besonderen Art von blickpunktabhängigem Highlight: Bewegt man sich aus dieser optimalen Betrachterposition heraus, verliert das dargestellte Bild für die neue Blickrichtung einen erheblichen Teil seiner Leuchtdichte.

Dieser Leinwandtyp ist für photometrisch konsistente Darstellung, und insbesondere für aus mehreren Projektionen zusammengesetzte Displays ungeeignet.

### Starre Rückprojektionen

Heutige Rückprojektionssysteme bestehen meist aus Mattscheiben aus Glas. Das starre Material erlaubt große Projektionsflächen bei gleichzeitiger Erhaltung der Polarisation. Solche Mattscheiben besitzen jedoch immer ein blickpunktabhängiges Highlight, das nicht vollkommen beseitigt werden kann.

Die Intensität des Highlights wird üblicherweise als *gain factor* bezeichnet. Während unter Umständen ein hoher *gain factor* wünschenswert sein kann, um z.B. die Maximalintensität für einen begrenzten Bildwinkel zu steigern, so erfordert jedoch photometrische Konsistenz insbesondere für unterschiedliche Betrachterpositionen einen Faktor optimalerweise von 1.0.

## Flexible Rückprojektionen

Mit flexiblen Materialien ist es möglich, nahezu 100% diffuse Leinwände herzustellen. Solche Leinwände sind jedoch für diejenigen Stereotechniken ungeeignet, welche auf gerichtete Lichtpolarisation angewiesen sind.

Des Weiteren sind flexible Schirme für große Projektionen aufgrund ihrer steigenden Empfindlichkeit gegenüber Luftbewegungen ungeeignet. Dieses Problem kann jedoch durch mit demselben Material beschichtete Klarglasscheiben behoben werden, wobei jedoch darauf geachtet werden muß, daß der verwendete Kleber die optischen Eigenschaften (wie insbesondere die Polarisation) nicht beeinträchtigt.

## Kombination starr und flexibel

Bei auf Klarglasscheiben aufgetragenen Leinwandmaterialien entsteht durch den verbindenden Klebstoff eine Beugungs- oder Reflexionsschicht, durch die ein gewisser Prozentsatz des einfallenden Lichtes unter festem Winkel in die Scheibe gestreut, und von deren Rückseite auf die Leinwand reflektiert wird. Speziell bei Leinwänden mit einem Gainfaktor in der Nähe von eins sind diese Reflexionen als deutliche Streifen in festen Abständen zu kontrastreichen Objekten erkennbar.

Um dieses Phänomen zu unterbinden und trotzdem die Vorteile einer starren Scheibe nutzen zu können, ist es möglich die Leinwand über die Scheibe zu spannen, ohne beides zu verkleben. Zur Sicherheit kann auch ein einige Millimeter dickes Luftpolster zwischen Scheibe und Leinwand erzeugt werden. Diese Technik wurde bei der HEyeWall (Kapitel 3.3.8) erfolgreich angewendet.

## Schwarze Leinwände

Eine sehr wichtige Verbesserung gegenüber Standard-Leinwandmaterialien sind Leinwände aus möglichst dunklem Material. Die Reflexion von auftreffendem Streulicht wird dadurch nahezu vollständig abgeschaltet, wobei jedoch die an sich diffusen Eigenschaften der Leinwand nicht beeinträchtigt werden. Diese Vorgehensweise ist der einzig mögliche Ansatz, Interreflexionsprobleme und damit Kontrastminderung in einer CAVE wirkungsvoll zu beseitigen. Selbst für planare Projektionssysteme bieten schwarze Leinwände einen höheren Kontrast und geringere Anfälligkeit für Streulicht aus der Umgebung.

Dies geht zwar auf Kosten der Lichtdurchlässigkeit, kann jedoch durch die Verwendung von leistungsstärkeren Projektoren leicht ausgeglichen werden kann.

Bis heute scheint jedoch die Herstellung großflächiger schwarzer Leinwände mit gleichmäßig verteilter „Schwärze“ problematisch zu sein, was sich durch „Wölkchen“, abwechselnd helleren und dunkleren Bereichen auf der Leinwand bemerkbar macht. Diese Eigenschaft könnte, speziell für Designer, den Vorteil der stark verbesserten Kontrasteigenschaften zunichte machen.

### 3.1.5 Exemplarische Setups

Mehrere, unterschiedliche Setups auf Basis der beschriebenen Technologien wurden untersucht. Für Leuchtdichte- und Farbmessungen wurde ein Chromameter (vgl. Kapitel 3.2.1) verwendet.

Soweit nicht anders vermerkt, handelt es sich bei den beschriebenen Setups um Rückprojektionssysteme.

Die hier angegebenen Meßwerte sollen nur einen ungefähren Eindruck der verschiedenen Technologien vermitteln und sind nur bedingt vergleichbar, da aufgrund der Gegebenheiten leider bei jedem Setup andere Versuchsbedingungen vorherrschten.

Man beachte, daß insbesondere die angegebenen maximalen Leuchtdichten nicht direkt die tatsächliche Leistungsfähigkeit des beschriebenen Projektortyps widerspiegeln, da sowohl die individuellen Einstellungen der Projektoren bezüglich Kontrast und Helligkeit variierten und nicht notwendigerweise auf 100% eingestellt waren, als auch die Projektionsflächen der einzelnen Szenarien unterschiedliche Abmessungen besaßen. Des Weiteren sind Kontrastwerte teilweise durch ungenügend abgeschottetes Restlicht reduziert.

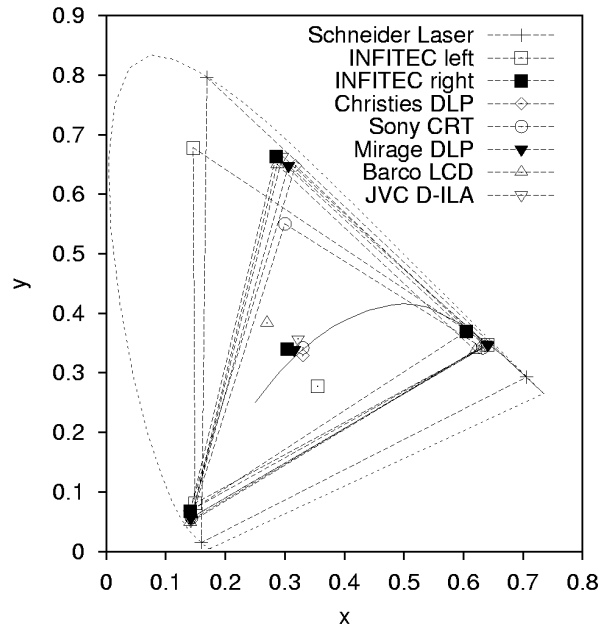


Abb. II: Farbräume und Weißpunkte der untersuchten Setups. Man beachte, daß sämtliche Primärvalenzen abgesehen vom Laserprojektor sowie dem Grünwert des linken Infitec-Filters trotz unterschiedlicher Technologien sehr nahe beieinanderliegen.

### Sony CRT, lineare Polarisation, Aufprojektion

Hierbei handelt es sich um die Werte für das gleiche Setup, das in Kapitel 3.2.1 beschrieben ist. Es dient hier als Vergleichssetup.

Maximales Weiß ohne Brille:	24 cd/m <sup>2</sup>
CIE-xy rot:	0.632 0.342
CIE-xy grün:	0.300 0.550
CIE-xy blue:	0.143 0.057
Kontrastverhältnis:	ca. 1000:1
Durchlässigkeit Brille:	72%
Crosstalk:	ca. 1:100
Leuchtdichte bei 45° Winkel:	40%
Leuchtdichte bei 75° Winkel:	10%

Der Wert für das maximale Weiß gilt nur, wenn man direkt in das deutlich ausgeprägte blickpunktabhängige Highlight schaut. Die mittlere Leuchtdichteverteilung über den gesamten Schirm gemessen lag im Bereich von 4.5 cd/m<sup>2</sup>. Die silbrige Leinwand besitzt eine starke Winkelabhängigkeit. Der Schwarzlevel befand sich innerhalb der Meßgenauigkeit des Meßgerätes.

### Electrohome CRT, Shuttering

Dies sind die Werte des fünfseitigen CAVE-Setups aus Kapitel 3.2.1.



Maximales Weiß ohne Brille:	2.5 cd/m <sup>2</sup>
CIE-xy rot:	0.597 0.345
CIE-xy grün:	0.337 0.560
CIE-xy blue:	0.148 0.067
Kontrastverhältnis:	ca. 1000:1
Durchlässigkeit Brille:	13%
Crosstalk (Kanaltrennung links/rechts):	ca. 1:100
Leuchtdichte bei 45° Winkel:	98%
Leuchtdichte bei 75° Winkel:	80%

Die Reflektivität der verwendeten flexiblen Leinwände betrug etwa 47% (vgl. Kapitel 3.2.4), wodurch deutliche Interreflexionseffekte auftreten:

Frontseite Weiß, andere Seiten Schwarz:	2.5 cd/m <sup>2</sup>
Frontseite Schwarz, andere Seiten Schwarz:	0.01 cd/m <sup>2</sup>
Frontseite Schwarz, andere Seiten Weiß:	1.5 cd/m <sup>2</sup>

Das daraus resultierende Kontrastverhältnis kann aus dem Verhältnis des letzten und ersten Wertes ermittelt werden, und lag in diesem Fall nur noch bei 1:1.7(!).

Der starke Lichtverlust der Shutterbrillen in Kombination mit den von Natur aus lichtschwachen Röhrenprojektoren führen letztendlich zu mesopischen Sichtbedingungen.

### TAN DLP, Infitec

#### Links:

Maximales Weiß ohne Brille:	8.8 cd/m <sup>2</sup>
CIE-xy rot:	0.641 0.347
CIE-xy grün:	0.146 0.678
CIE-xy blue:	0.149 0.081
Kontrastverhältnis:	60:1
Durchlässigkeit Brille:	69%
Crosstalk:	1:120
Leuchtdichte bei 45° Winkel:	84%
Leuchtdichte bei 75° Winkel:	63%

#### Rechts:

Maximales Weiß ohne Brille:	13.1 cd/m <sup>2</sup>
CIE-xy rot:	0.605 0.370
CIE-xy grün:	0.285 0.663
CIE-xy blue:	0.141 0.068
Kontrastverhältnis:	40:1

Durchlässigkeit Brille:	65%
Crosstalk:	1:170
Leuchtdichte bei 45° Winkel:	84%
Leuchtdichte bei 75° Winkel:	63%

Aufgrund der Spektralselektion unterscheiden sich die Farborte der Primärvalenzen zwischen rechts und links erheblich. Obwohl die individuellen R/G/B-Intensitäten seitens der Projektoren bereits extrem verändert wurden, um diese farblichen Verschiebungen aufgrund der Spektralfilter auszugleichen, war es nicht möglich, die Weißpunkte beider Projektoren aufeinander abzugleichen (0.355/0.277 vs. 0.304/0.340). Für eine farbkonsistente Darstellung muß also ein komplexeres Kalibrierungsverfahren durchgeführt werden (Kapitel 3.2.5).

Die Durchlässigkeit der optischen Stereobrillen in Kombination mit den projektorseitigen Filtern ist mit anderen passiven Techniken vergleichbar (wenn auch für beide Augen unterschiedlich).

Bei der Leinwand handelte es sich um eine Glasmattscheibe. Da die Umgebung während der Messungen nicht vollständig abgedunkelt werden konnte, sind die Meßwerte für Kontrast und Kanaltrennung unzuverlässig.

### Eiki LCD, Infitec

Dies sind die Werte des in Kapitel 3.3.8 vorgestellten CAVE-Setups.

Maximales Weiß ohne Brille (vorjustiert):	27 cd/m <sup>2</sup>
CIE-xy rot (l/r):	0.646 0.352 , 0.656 0.338
CIE-xy grün (l/r):	0.170 0.696 , 0.329 0.665
CIE-xy blue (l/r):	0.140 0.055 , 0.136 0.067
Kontrastverhältnis:	ca. 100:1
Durchlässigkeit Brille (l/r):	70% , 75%
Crosstalk (links/rechts):	ca. 1:70 – ca. 1:250
Leuchtdichte bei 45° Winkel:	29%
Leuchtdichte bei 75° Winkel:	12%

Folgende Interreflexionseffekte wurden in diesem fünfseitigen CAVE-Setup gemessen:

Frontseite Weiß, andere Seiten Schwarz:	23.7 cd/m <sup>2</sup>
Frontseite Schwarz, andere Seiten Schwarz:	0.23 cd/m <sup>2</sup>
Frontseite Schwarz, andere Seiten Weiß:	1.47 cd/m <sup>2</sup>

Hieraus ergibt sich ein Kontrastverhältnis von 1:19, sowie eine Reflektivität der Leinwände von 7%. Dies stellt jedoch nur einen verhältnismäßig unzuverlässigen

Schätzwert dar, da die Leinwand nicht mehr als Lambertstrahler angenommen werden kann.

Durch Anwendung des in Kapitel 3.2.5 entwickelten Farbausgleichsalgorithmus wird die nutzbare Gesamthelligkeit zwangsläufig weiter reduziert (vgl. Messwerte in Kapitel 3.3.9).

### Christie LCD, Infitec

Dies sind die Werte des in Kapitel 3.3.8 vorgestellten HEyeWall-Setups.

Maximales Weiß ohne Brille (vorjustiert):	97.6 cd/m <sup>2</sup>
CIE-xy rot (l/r):	0.655 0.342 , 0.659 0.336
CIE-xy grün (l/r):	0.104 0.735 , 0.327 0.671
CIE-xy blue (l/r):	0.140 0.039 , 0.137 0.065
Kontrastverhältnis (l/r):	ca. 700:1 – ca. 900:1
Durchlässigkeit Brille (l/r):	70% , 75%
Crosstalk (links/rechts):	ca. 1:160 – ca. 1:2000
Leuchtdichte bei 45° Winkel:	91%
Leuchtdichte bei 75° Winkel:	83%

Wie beim obigen CAVE-Setup reduziert sich die nutzbare Gesamthelligkeit durch den Farbausgleich entsprechend.

### Barco LCD, zirkulare Polarisation

Maximales Weiß ohne Brille:	125 cd/m <sup>2</sup>
CIE-xy rot:	0.614 0.337
CIE-xy grün:	0.290 0.645
CIE-xy blue:	0.143 0.049
Kontrastverhältnis:	90:1
Durchlässigkeit Brille:	63%
Crosstalk:	1:75-1:200
Leuchtdichte bei 45° Winkel:	20%
Leuchtdichte bei 75° Winkel:	7%

Die Durchlässigkeit der zirkularen Polarisationsbrillen ist mit der linearer Brillen vergleichbar. Interessanterweise war das Kanalübersprechen nicht vollkommen unabhängig gegenüber einer Rotation der Stereobrillen um die optische Achse, was von der Kompatibilität von auf Projektorseite eingesetzten Filtern und den Brillen abhängt.

Die in diesem Fall verwendete Mattscheibe war für Einzelpersonen-Betrachtung optimiert und verwendete eine Fresnellinsentechnik, wodurch das durchgelasse-

ne Licht hochgradig blickpunktabhängige Eigenschaften besaß. Der Vorteil lag in einem fast völlig schwarzen Erscheinungsbild bezüglich diffuser Reflexionen.

### JVC D-ILA, mono

Maximales Weiß:	105 cd/m <sup>2</sup>
CIE-xy rot:	0.637 0.348
CIE-xy grün:	0.296 0.664
CIE-xy blue:	0.142 0.052
Kontrastverhältnis:	100:1
Leuchtdichte bei 45° Winkel:	33%
Leuchtdichte bei 75° Winkel:	22%

Auch hier besaß die verwendete Mattscheibe stark blickpunktabhängige Eigenschaften.

### Mirage DLP, Shuttering

Maximales Weiß ohne Brille:	63 cd/m <sup>2</sup>
CIE-xy rot:	0.641 0.347
CIE-xy grün:	0.306 0.648
CIE-xy blue:	0.142 0.056
Kontrastverhältnis:	1:150
Durchlässigkeit Brille:	13%
Crosstalk:	ca. 100:1
Leuchtdichte bei 45° Winkel:	84%
Leuchtdichte bei 75° Winkel:	63%

Die geringen Lichtdurchlässigkeiten sind typisch für Shutterbrillen (vgl. Kapitel 3.2.3). Die verwendete Leinwand war mit der des Infitec-Setups identisch.

### Christies DLP, zirkulare Polarisation, schwarze Leinwände

Maximales Weiß ohne Brille:	57 cd/m <sup>2</sup>
CIE-xy rot:	0.636 0.345
CIE-xy grün:	0.318 0.648
CIE-xy blue:	0.140 0.068
Kontrastverhältnis:	110:1
Durchlässigkeit Brille:	13%
Crosstalk:	ca. 1:100
Leuchtdichte bei 45° Winkel:	84%
Leuchtdichte bei 75° Winkel:	63%

Man beachte, daß der Schwarzlevel (und damit das direkte Kontrastverhältnis) durch den Einsatz schwarzer Leinwände nicht verbessert wird, da sowohl schwarzes als auch weißes Licht um denselben Faktor reduziert werden. Der entscheidende Vorteil der schwarzen Leinwände liegt jedoch in der Reduktion von Streulicht und wird offensichtlich, wenn man sich das Lichtverhalten auf der Frontseite eines CAVE-Szenarios anschaut:

Frontseite Weiß, anderen Schwarz:	61 cd/m <sup>2</sup>
Frontseite Schwarz, anderen Schwarz:	0.53 cd/m <sup>2</sup>
Frontseite Schwarz, anderen Weiß:	0.83 cd/m <sup>2</sup>

Das Kontrastverhältnis lag hiermit bei 1:75, im Vergleich zu den oben genannten 1:1.7(!) für weiße Leinwände. Aus der in Kapitel 3.2.4 hergeleiteten Formel ergibt sich in diesem vierseitigen CAVE-Setup eine Reflektivität der Leinwände von etwa 2%.

### Schneider Laser, mono, Aufprojektion

Maximales Weiß:	23 cd/m <sup>2</sup>
CIE-xy rot:	0.706 0.293
CIE-xy grün:	0.169 0.796
CIE-xy blue:	0.160 0.015
Kontrastverhältnis:	250:1

Die CIE-xy Farbkoordinaten wurden aus dem Datenblatt entnommen, da das Meßgerät offenbar mit reinen Spektralquellen überfordert war.

Man beachte, daß die Projektionsfläche (eine weiß gestrichene Wand) etwa 9x6m maß. Für gebräuchlichere Leinwandgrößen ergibt sich eine entsprechend höhere Maximalintensität.

Die Helligkeitsspezifikation des Projektors in ANSI-Lumen ist nicht bekannt. laut Datenblatt beträgt die Lichtleistung 7 Watt. Damit lässt sich die Obergrenze des möglichen Lichtstroms mit  $683 \frac{\text{lm}}{\text{W}} \cdot 7\text{W} = 4781\text{lm}$  angeben. Je nach Spektralverteilung liegt der tatsächliche Wert entsprechend niedriger.

Eine wie hier verwendete mattweiße Wand lässt sich jedoch grob als Lambertstrahler mit einem Reflexionskoeffizienten von ca. 0.8 abschätzen. Über den Zusammenhang von Lichtstrom, Leuchtdichte und Fläche (vgl. Kapitel 2.4):

$$\Phi = L \cdot A \cdot \pi$$

läßt sich so der *ankommende* Lichtstrom des Projektors mit ca. 5000 Lumen am hellsten Punkt überschlagen.

Der relativ hohe Schwarzlevel liegt in der Tatsache begründet, daß der Raum nicht vollständig abgedunkelt werden konnte und gibt daher nicht das tatsächliche Kontrastverhältnis wieder, das deutlich höher liegt. Laut Datenblatt liegt der erreichbare Kontrast bei  $>5000:1$ .

## 3.2 Mängel existierender Displaysysteme

Nachdem im vorigen Abschnitt die zur Verfügung stehenden und teilweise sehr unterschiedlichen Technologien vorgestellt sowie ihre grundlegenden, für photometrisch konsistente Bildwiedergabe relevanten Eigenschaften diskutiert wurden, beinhaltet dieser Abschnitt eine genaue Analyse dieser relevanten Eigenschaften, ihre qualitativen Auswirkungen, quantitative Untersuchungen, sowie Lösungsvorschläge, wie sich negative Eigenschaften verringern oder sogar ausgleichen lassen [SKUM99]. Hierzu wurden konkret zwei der in Kapitel 3.1.5 vorgestellten Aufbauten mit photometrischen und colorimetrischen Meßmethoden analysiert und entsprechend der vorgefundenen Effekte aufgeschlüsselt. Insbesondere werden Leuchtdichteumfang, Farbwiedergabe, Helligkeitsverteilung über der gesamten Bildfläche, sowie Interreflexionseffekte genauer betrachtet.

### 3.2.1 Testumgebung

Als Meßinstrument wurde ein Minolta Chroma-Meter CS-100A (Abb. 12) verwendet. Dieses *tristimulus colorimeter* besitzt eine spektrale Farbempfindlichkeit, die dem des menschlichen Sehens sehr gut angepaßt ist. Die Meßwerte wurden im standardisierten CIE Yxy Format vorgenommen [CIE31]. Das Gerät ist in der Lage, Leuchtdichten in einem Bereich von  $0.01 \text{ cd/m}^2$  bis  $299000 \text{ cd/m}^2$  bei einem maximalen Fehler von  $\pm 2\%$ , sowie xy Farbwerte mit Fehlern von  $\pm 0.004$  zu bestimmen. Bei den Farbwerten ist angegeben, daß die gemessenen Werte nur oberhalb vom  $12 \text{ cd/m}^2$  entsprechend der angegebenen Fehlertoleranz verlässlich sind. Verschiedene Meßreihen haben jedoch ergeben, daß der Farbfehler bis hinunter zu etwa  $1 \text{ cd/m}^2$  nicht wesentlich steigt, was für die hier teilweise im mesopischen Bereich liegenden Messungen wichtig ist.



Abb. 12 : Das Minolta CS-100A Colorimeter.

### Großbildstereoprojektion aus vier Projektoren

Beim ersten Setup handelte es sich um eine Breitbild-Aufprojektion mit linearer Polarisation als passivem Stereoverfahren. Die von  $2 \times 2$  Projektoren nebeneinander erzeugten Bilder wurden durch ein elektronisches Edgeblending überlagert

und ergaben so ein projiziertes Bild von insgesamt 5x2 Metern. Die vier SONY VPH-1292QMG Röhrenprojektoren wurden gespeist von einer Dualpipe SGI Onyx2 InfiniteReality2 mit 4 MIPS R10000 CPUs und einer Auflösung von 2400x1200 Bildpunkten. Die äußeren beiden Projektoren erzeugten die beiden Teilbilder für das linke Auge, die inneren beiden für das rechte (Abb. 13a). Der Betrachter wurde nicht getrackt.

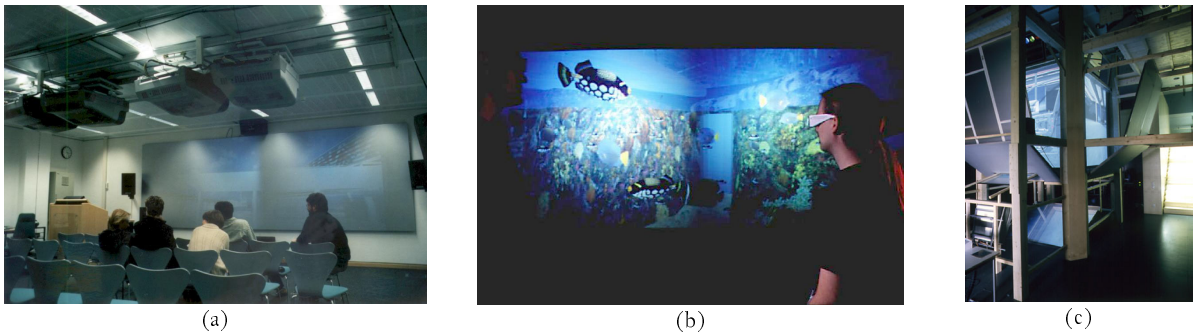


Abb. 13 : (a) Die Breitbild-Stereoprojektion mit vier Projektoren. (b) Ein Foto der Projektion in Betrieb. (c) Die fünfseitige CAVE-Installation.

### Fünfsseitige CAVE

Das CAVE-Setup (Abb. 13c) besteht aus fünf Leinwänden, die drei Seiten, sowie Decke und Boden eines Würfels mit einer Kantenlänge von 2.4 Metern bilden. Um das verwendete elektromagnetische Tracking nicht zu beeinträchtigen, besteht die Konstruktion weitestgehend aus Holz. Der Rückprojektions-Fußboden ist eine Spezialkonstruktion bestehend aus einer 2.4m x 2.4m x 30mm großen *Paraglas*-Platte, einem speziellen Acrylglas, und hält einem Gewicht von wenigstens 10 Personen stand, ohne dabei nennenswert den Strahlengang zu beeinflussen, Licht zu absorbieren – oder durchzubrechen. Die übrigen Leinwände bestanden im Originalaufbau bis 2002 aus dem Leinwandkunststoffmaterial Geriets Opera milchig matt [G03], das an den Stoßstellen verschweißt war, um nahtlose Übergänge zwischen den Seiten zu ermöglichen. Der Boden war mit dem gleichen Material beschichtet.

Betrieben wurde die Anlage mit fünf modifizierten Electrohome Marquee 7500 mit einer Auflösung von 1024x1024 Bildpunkten und einer Bildwiederholfrequenz von 2x57Hz im Shutterbetrieb. Die Bilderzeugung übernahm eine Triplepipe SGI ONYX InfiniteReality mit 6 CPUs. Die Shutterbrillen waren vom Typ CrystalEyes von StereoGraphics. Es wurden zusätzliche Messungen mit NuVision 60GX Brillen durchgeführt, die jedoch außer einer leicht verbesserten Rotdurchlässigkeit und einem kleineren Blickfeld keine nennenswerten Unterschiede aufwiesen.

Inzwischen wurden beide Displaysysteme durch neue Generationen ersetzt, eine digitale CAVE und eine gekachelte Stereoprojektion. Diese Systeme werden in Kapitel 3.3.8 näher beschrieben.

### 3.2.2 Leuchtdichteverteilung

Eine uneinheitliche Verteilung der Leuchtdichte über die Bildfläche kann zwar relativ gut durch die automatische Bildverarbeitung im Gehirn „weggedacht“ werden, jedoch wird man die Darstellungsqualität bei einheitlicher Verteilung als ungleich besser wahrnehmen. Wie stark diese Abweichungen sind und wie sie beseitigt werden können, soll in diesem Abschnitt näher untersucht werden.

#### Helligkeitsschwankungen

Eine der Hauptvoraussetzungen, ein Projektionssetup als photometrisch konsistentes Ausgabemedium zu verwenden, ist die gleichmäßige, konstante Leuchtdichte über der gesamten Bildfläche und für jede Blickrichtung. Ohne diese Konstanz wäre die wahrgenommene Helligkeit eines Objektes abhängig von dessen Position auf der Leinwand, bzw. abhängig von der Betrachterposition, wodurch eine verlässliche Darstellung unmöglich würde.

Diese Eigenschaften hängen, wie in Kapitel 3.1 besprochen, zum einen von der Güte des Projektors ab, was die Gleichmäßigkeit der Verteilung angeht, zum anderen von den physikalischen Eigenschaften des Leinwandmaterials.

Die Leuchtdichte des diffusen Highlights, das durch die einfache Tatsache des geringsten Projektionsabstandes und einem rechten Winkel in der Nähe der Projektionsmitte entsteht, fällt entsprechend einer Gauß-ähnlichen Funktion senkrecht zur Projektionsrichtung zu den Ecken hin ab. Hochwertige CRT-Projektoren sind mit entsprechenden Einstellungsmöglichkeiten ausgerüstet, um diesem Highlight entgegenzuwirken. Bei anderen Projektortypen ist dies jedoch generell nicht der Fall.

Um den Helligkeitsverlauf auf einer der symmetrisch aufgebauten CAVE-Leinwände zu analysieren, wurde in Intervallen von 24cm (d.h. einem 10x10-Raster) die Leuchtdichte vermessen. Um hierbei mögliche Interreflexionseffekte auszuschließen, erfolgte die Messung bei abgedunkeltem Raum direkt auf der Leinwand. Das resultierende Isoliniendiagramm ist in Abb. 14 zu sehen. Die Leuchtdichtewerte ändern sich fließend von  $0.8 \text{ cd/m}^2$  in den unteren Ecken, bis hin zu  $2.5 \text{ cd/m}^2$  an der hellsten Stelle.

Der Grund, daß sich der hellste Punkt nicht genau im Zentrum der Leinwand befindet, liegt an dem Aufbau von Projektor und den beiden Umlenkspiegeln re-



lativ zur Leinwand (vgl. Abb. 13c); die optische Achse steht nicht senkrecht auf der Leinwandebene.

Aufgrund der hochgradigen Winkelabhängigkeit der Stereo-Aufprojektion, welche im nächsten Abschnitt noch genauer analysiert wird, und der Tatsache, daß das Meßinstrument einen Schatten auf den zu messenden Punkt werfen würde, mußte für diesen Aufbau ein anderer Ansatz gewählt werden. Hier wurde das Meßgerät an der optimalen Betrachterposition, nämlich zentral vor der Leinwand in einem Abstand von 3m und einer mittleren Augenhöhe eines sitzenden Betrachters von 1.2m, positioniert. Nur die äußeren beiden Projektoren waren während der Messung aktiv, was der Leuchtdichtevermessung für ein einzelnes Auge (das linke) entspricht.

Die Leuchtdichtewerte für die Aufprojektion (Abb. 13a) variieren stark zwischen  $0.4$  und  $24 \text{ cd/m}^2$ , also fast zwei Größenordnungen gegenüber einer halben Größenordnung bei einer rein diffusen Leinwand wie in der CAVE. Das Diagramm zeigt außerdem, daß die Projektoren zur Zeit der Messung nicht optimal eingestellt waren; der linke Projektor ist etwa 20% heller als der rechte. Abb. 13b zeigt ein Foto der Leinwand in Betrieb. Während die außerordentlichen Helligkeitsunterschiede für das bloße Auge nicht so extrem wahrgenommen werden, weil der menschliche Sehapparat und die Bildverarbeitung im Gehirn hier bereits beträchtliche Ausgleichsarbeit leisten, ist der Effekt der abfallenden Leuchtdichte und des ausgeprägten Highlights in einer Photographie sehr deutlich sichtbar.

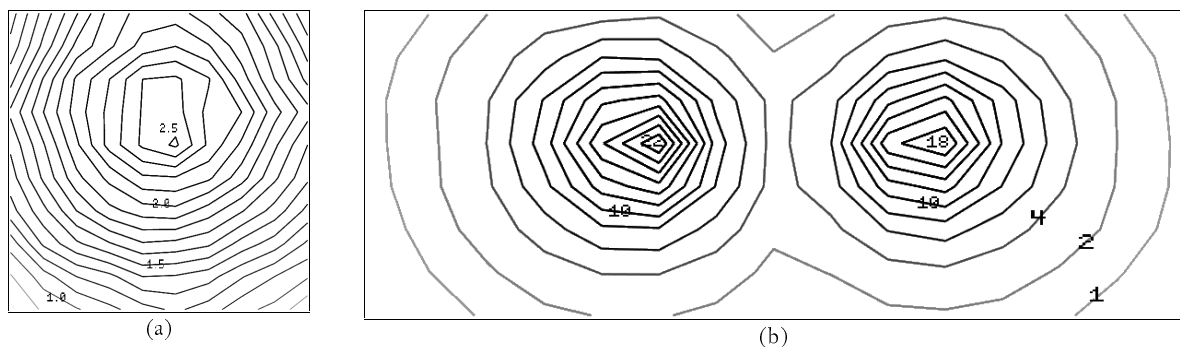


Abb. 14 : (a) Isoliniendiagramm der CAVE-Frontseite. Leuchtdichten im Bereich von  $0.8$  bis  $2.5 \text{ cd/m}^2$ , Linienabstand  $0.1 \text{ cd/m}^2$ . (b) Isoliniendiagramm der Aufprojektion für einen zentrierten Beobachter in  $3\text{m}$  Abstand. Leuchtdichten im Bereich von  $0.4$  bis  $24 \text{ cd/m}^2$ , Linienabstand  $2 \text{ cd/m}^2$ .

### Winkelabhängigkeit

Ein weiterer wichtiger Gesichtspunkt bei der Thematik gleichmäßiger Ausleuchtung sind winkelabhängige Reflexions- bzw. Transmissionseigenschaften des Leinwandmaterials.

Für CAVE-Umgebungen ist selbst für photometrisch nicht konsistente Setups eine möglichst nur diffus durchscheinende Leinwand Grundvoraussetzung, da sonst sofort winkel- und damit auch blickpunktabhängige Highlights entstehen, die den Gesamteindruck stören oder gar zerstören würden.

Die in Kapitel 3.1.4 beschriebenen Fresnellinsenscheiben mit ihren nicht-uniformen Charakteristiken werden oftmals für Stereoprojektionen mit nur einer Leinwand verwendet. Solange sich der oder die Beobachter in der Nähe des optimalen Bereichs befinden, erhöhen diese Scheiben die Lichtausbeute. Sind die Betrachter jedoch sehr nahe an der Leinwand, oder ist die projizierte Fläche sehr groß, werden solche Scheiben unbrauchbar.

Um eine gleichmäßige Ausleuchtung der projizierten Fläche zu erhalten, sind diffuse Leinwände ein wesentlicher Gesichtspunkt.

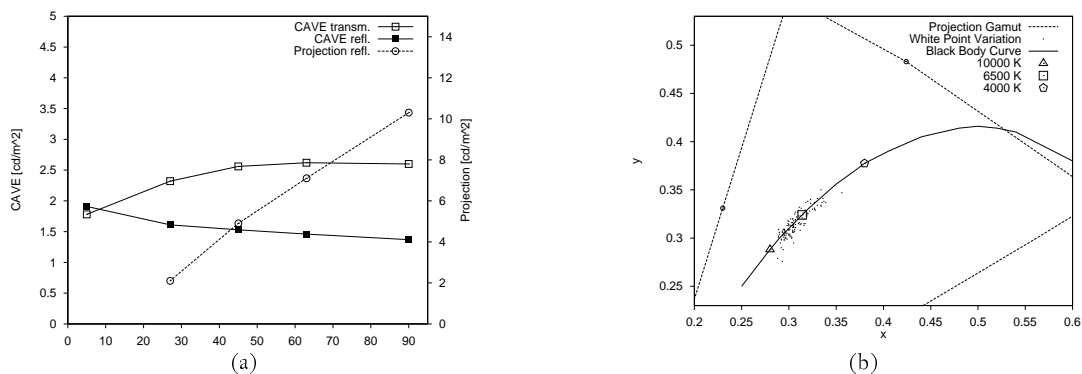


Abb. 15: (a) Winkelabhängigkeit von CAVE-Leinwänden (transmissive und reflektiv) und Aufprojektion. (b) Variationen im Weißpunkt der Aufprojektion über die Leinwandfläche aufgrund hoher Winkelabhängigkeit und räumlich getrennten CRT-Röhren. Zum Vergleich sind einige Standard-Farbttemperaturen gegeben.

Die Winkelabhängigkeit wurde an einem festen Punkt auf der Leinwand durch Messungen aus unterschiedlichen Richtungen bestimmt.

Während die resultierende Kurve in der CAVE nahezu konstant verläuft und nur sehr geringfügig für kleine Winkel abnimmt (Abb. 15a, Kurve „CAVE transm.“) und damit als nahezu ideal diffus eingeordnet werden kann, ergibt sich für die Breitbildprojektion (Kurve "Projection refl.") eine für Winkel  $< 90^\circ$  stark abfallende Kurve.

Die Transmissionsfunktion der CAVE-Leinwände wurde ermittelt, indem die einzig beleuchtete Leinwand vermessen wurde, während die anderen abgeschaltet waren. Die Reflexionskurve entstand auf ähnliche Weise in der umgekehrten Situation. Fehler durch tertiäre indirekte Reflexionen (Licht, das von der gemessenen Leinwand auf eine Seitenwand fällt und zurückreflektiert wird), die das Meßergebnis beeinträchtigen könnten, lagen bei  $< 2\%$ , wie man mit den in Kapitel 3.2.4 ermittelten Gleichungen nachvollziehen kann.

## Positionsabhängige Highlights

Die gemessene Winkelabhängigkeit ist direkt verantwortlich für die Entstehung von Phong-ähnlichen Highlights an den imaginären Positionen der gespiegelten Projektorobjektive im Verhältnis zum Beobachter. Darüber hinaus ergeben sich durch baulich getrennte Farbröhren in den drei Primärfarben Farbverschiebungen und nicht unbeträchtliche Schwankungen im Weißpunkt über die gesamte Leinwandfläche. Durch die Ausrichtung der Projektoren sind die vier Highlights an ihrem linken Rand bläulich verfärbt, an ihrem rechten rötlich. Dies erzeugt zusätzliche Farbverzerrungen in der Nähe der Kante des elektronischen Edge-Blendings, was zwar durch Umbau der Röhrenanordnung in den beiden rechten Projektoren vermieden würde, aber den Effekt hätte, daß nunmehr der zentrale Bereich der Projektion rötlich erscheinen würde, während die äußeren Ränder bläulich wären.

Um die positionsabhängigen Highlights einschließlich des Einflusses des Edge-Blending-Moduls zu vermessen, wurden sechs unterschiedliche Betrachterpositionen gewählt: Zentral vor der Leinwand, sowie um je 1.5m nach rechts und links versetzt, das Ganze jeweils im Abstand von 3 bzw. 6 Metern. Von jeder dieser Positionen wurde aus der gemittelten Augenhöhe von 1.2m einer sitzenden Person die Projektion entlang der mittleren Projektionshöhe bei 1.65m vermessen.

Die Highlights verlagern sich für eine nicht zentral sitzende Person zu der jeweiligen Seite hin. Durch das Edge-Blending ergibt sich hierbei jedoch für einen dichter an der Leinwand platzierten Beobachter eine außerordentlich auffällige Hell-/Dunkelkante, sodass die visuelle Qualität für Betrachter außerhalb der Mittellinie stark leidet (Abb. 16).

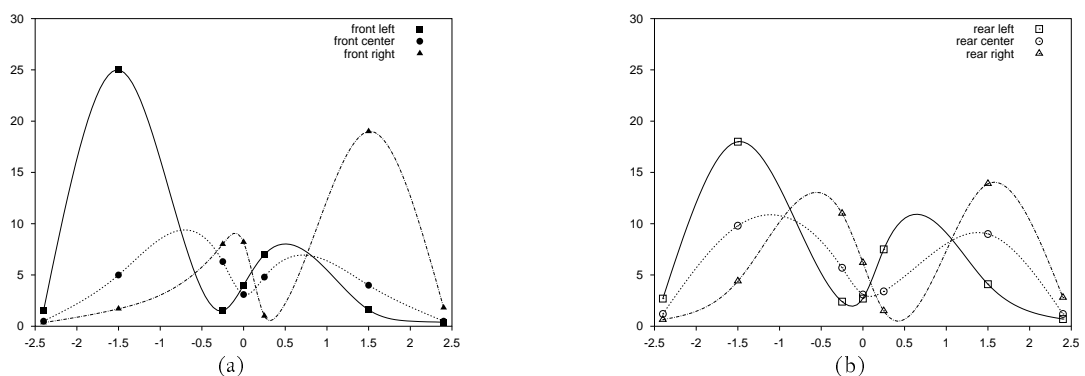


Abb. 16: Blickpunktabhängige Leuchtdichtevertelung für zentrierte, sowie 1.5m nach links/rechts versetzte Betrachterpositionen. (a) bei 3m Abstand, (b) bei 6m Abstand. Man beachte, daß die gesampelten Punkte der Anschaulichkeit halber durch Splines verbunden wurden, was nicht notwendigerweise dem tatsächlichen Verlauf entspricht.

Insbesondere für Kachelprojektionen (*tiled displays*) sind voll-diffuse Leinwände ein absolutes Muß, da ansonsten Hell-/Dunkelkanten an jeder Kachelgrenze auftreten, deren Intensität und Lage sich nicht nur für das rechte und linke Auge, sondern insbesondere auch für einen sich bewegenden Betrachter ändern, wodurch der Eindruck eines einzigen, hochauflösenden Bildes sehr effizient zerstört wird (vgl. Kapitel 3.3.8).

### Einfluß von Streulicht

Weitere Messungen umfaßten den Einfluß von Stör- und Streulicht, wie beispielsweise ein eingeschalteter Monitor im selben Raum (Abb. 18k), oder gar ein Notbeleuchtungsschild. Auch wenn diese Einflüsse in der CAVE-Umgebung nahe der Genauigkeitsgrenze des Meßgerätes im Bereich von 0.02-0.05 cd/m<sup>2</sup> lagen, wird doch deutlich, daß sie einen nicht zu vernachlässigenden Einfluß auf von Natur aus mesopische Szenarien haben können, wie z.B. einem Fahrzeuginnenraum in einer nur von Armaturenleuchten erhellenen Nachtsituation.

### Bewertung

Zur Beseitigung von festen, diffusen Highlights kann man z.B. Projektoren einsetzen, die sich über ihre Fläche genau justieren lassen. Eine Softwarelösung hierzu ist eine transparente Textur, die aus der inversen Leuchtdichteverteilung berechnet und während dem Rendering über die Szene gelegt wird (vgl. Kapitel 3.4). Der offensichtliche Nachteil hierbei ist, daß die maximal darstellbare Leuchtdichte auf einen Wert reduziert wird, der der niedrigsten tatsächlichen Leuchtdichte auf der Leinwand (also typischerweise einer der Ecken) entspricht. Betreibt man die Projektoren jedoch mit maximaler Leistung und nutzt die transparente Textur nur dazu, die stärksten Highlights zu beseitigen, verbessert sich dieser Wert.

Der Ansatz der transparenten Textur könnte auch zur Beseitigung blickpunktabhängiger Highlights dienen, jedoch käme die verbesserte Bildqualität nur einem Benutzer zugute, der sich in direkter Nähe der Position befindet, von der aus die Textur berechnet wurde. Ist der Hauptgrund für die Entstehung dieser Art der Highlights jedoch der, daß die Leinwand z.B. die Polarisation erhalten muß, sollte man in Erwägung ziehen, eine andere Stereotechnik einzusetzen.

Die Farbverschiebungen von blickpunktabhängigen Highlights aufgrund von physikalisch getrennten Farbröhren lassen sich nur vermeiden, wenn man eine Projektortechnologie verwendet, die mit nur einer einzigen, kombinierten Lichtaustrittsöffnung auskommt.

Eine generelle Methode, die Stärke von blickpunktabhängigen Highlights etwas einzuschränken besteht darin, konkave oder zylindrische Projektionsschirme zu

verwenden. Aufgrund der fehlenden Hardwareunterstützung für die Bildgenerierung auf gewölbten Flächen würde dies jedoch die Berechnungszeiten für VR-Applikationen drastisch erhöhen, es sei denn, man nimmt geometrische Verzerrungen bis zu einem gewissen Grad in Kauf. Eine andere Variante ist, für gekrümmte Leinwände ein zweistufiges Renderingverfahren [RCWS98] einzusetzen.

Um den Einfluß von Fremdlichtquellen zu minimieren, ist die naheliegendste Lösung vermutlich auch die praktikabelste: Man sollte sie abschalten, oder den gesamten Projektionsraum durch Wände oder Vorhänge abschirmen. Schwarze Leinwände würden ebenfalls verhindern, daß Streulicht von den Leinwänden reflektiert wird.

### 3.2.3 Leuchtdichtebereich und Farbraum

In diesem Abschnitt sollen insbesondere diejenigen Gründe näher untersucht werden, die für ein reduziertes Farbempfinden speziell in CAVE-Umgebungen verantwortlich sein könnten.

#### Leuchtdichtebereich

Betrachtet man die maximal darstellbaren Leuchtdichten verschiedener Projektionssysteme, so fällt auf, daß sie z.B. im Vergleich zu einem Standardmonitor sehr viel dunkler sind, je nach Konfiguration um bis zu zweieinhalb Größenordnungen: Während ein typischer CRT-Monitor etwa um die  $100 \text{ cd/m}^2$  aufweist, liegt die höchste Leuchtdichte der Stereoleinwand je Auge bereits nur bei  $24 \text{ cd/m}^2$ , während in der CAVE bei einem maximal weißen Bild am hellsten Punkt nur noch  $2.5 \text{ cd/m}^2$  gemessen werden.

Hierbei bleiben des weiteren noch die Stereobrillen unberücksichtigt: Für die Großbildprojektion, welche lineare Polarisation als Stereoverfahren verwendet, führen zwei Lagen von Polarisationsfiltern zu weiterer Lichtabschwächung. Zunächst müssen die Objektive mit linear polarisierenden Filtern versehen werden. Diese besitzen eine Transmission von 36%. Da sie normalerweise fest an den Projektoren montiert sind, beinhaltet der oben angegebene Maximalwert diese Abschwächung bereits.

Um nun pro Auge jeweils nur den gewünschten Kanal wahrzunehmen, müssen Stereobrillen verwendet werden. Diese bestehen aus einem zweiten Satz linearer Polarisationsfilter. Da das einfallende Licht bereits polarisiert ist, wird der größte Teil des Lichtes durchgelassen, sodass die Transmission dieser Filter bei immerhin 72% liegt.

Ausgehend von den genannten  $24 \text{ cd/m}^2$  läßt sich also rückrechnen, daß die eigentliche Projektorleistung an der hellsten Stelle der Leinwand bei  $24/0.36=67 \text{ cd/m}^2$  liegt. Hinzukommt, daß aufgrund des passiven Verfahrens zwei Projektoren zwei überlagerte Bilder liefern. Beide Projektoren ohne Filter und auf Weiß gestellt würden demnach zusammen in einer Leuchtdichte von  $133 \text{ cd/m}^2$  resultieren. Was am Ende pro Auge hinter der Brille davon noch wahrgenommen wird, sind  $24*0.72=17.3 \text{ cd/m}^2$ . Die effektive Transmission beläuft sich damit also auf  $36\%*72\%/2=13\%$ .

Es muß dazugesagt werden, daß der Maximalwert von  $24 \text{ cd/m}^2$  aufgrund der stark spiegelnden Eigenschaften der Leinwand nur auf einen äußerst beschränkten Bereich zutrifft. Mittelt man die Gesamtleuchtdichte eines projizierten weißen Bildes, ergibt sich ein Wert von nur  $4.5 \text{ cd/m}^2$ , bzw. von  $3.2 \text{ cd/m}^2$  hinter der Brille.

Bei der im aktiven Stereomodus betriebenen CAVE verhält es sich etwas anders. Hier liefert nur *ein* Projektor die oben genannte Leuchtdichte von  $2.5 \text{ cd/m}^2$  (gemessen ohne Interreflexionseinfluß). Vor den Projektoren befinden sich keine Filter. Die Transmission inaktiver Shutterbrillen liegt jedoch bereits bei nur 31%. Arbeiten sie synchron zum Videosignal (im Meßaufbau  $2x57 \text{ Hz}$ ), sinkt die Gesamttransmission ebenfalls auf 13%, womit von den  $2.5 \text{ cd/m}^2$  beim Auge noch  $0.33 \text{ cd/m}^2$  ankommen. Beachtet man auch hier die Gesamtverteilung der Leuchtdichte über die Leinwand, ergeben sich Mittelwerte von 1.7 bzw.  $0.22 \text{ cd/m}^2$ .

	Shutterbrille			Polarisationsbrille	
	transparent	geschlossen	geschüttert	durchlässig	90° verdreht
Transmission	~ 31 %	< 1 %	~ 13 %	~ 72 %	< 1 %

Tabelle 3 : Durchlässigkeit von Stereobrillen.

	CAVE		Breitbildprojektion		CRT Monitor
	ohne Brille	mit Shutterbrille	ohne Brille	mit Polarisationsbrille	
Maximale Leuchtdichte	$2.5 \text{ cd/m}^2$	$0.33 \text{ cd/m}^2$	$24 \text{ cd/m}^2$	$17.3 \text{ cd/m}^2$	~ 100 $\text{cd/m}^2$
Mittlere Leuchtdichte	$1.7 \text{ cd/m}^2$	$0.22 \text{ cd/m}^2$	$4.5 \text{ cd/m}^2$	$3.2 \text{ cd/m}^2$	~ 95 $\text{cd/m}^2$

Tabelle 4 : Maximale und mittlere Leuchtdichten für ein 100% weißes Bild.

Damit befindet man sich spätestens im CAVE-Setup weit im mesopischen Sehbereich, dem Dämmerungssehen, in dem die Farbwahrnehmung bereits stark eingeschränkt ist. Dies ist einer der Gründe, warum der Farbeindruck in einem CAVE-Szenario mangelhaft ist. In der Tat wirken die Farben auf der Stereoleinwand brillanter, deren mittlere Leuchtdichte um mehr als eine ganze Größenordnung über der der CAVE – und damit bereits im photopischen Bereich, dem Tagsehen – liegt.

Eine Folgerung aus dieser Beobachtung ist, daß die CAVE besonders zur Darstellung von Szenen geeignet ist, die eine generell niedrige Leuchtdichte in der „realen Welt“ besitzen, d.h. bei denen ebenfalls mesopische Sichtbedingungen vorherrschen. Würde eine solche Szene z.B. auf einem Monitor dargestellt werden, würden die Farben als übertrieben intensiv empfunden werden.

Bei vielen, neu aufgebauten, immersiven Installationen kommen seit kurzem mehr und mehr leistungsstarke DLP-Projektoren zum Einsatz, die den darstellbaren Helligkeitsbereich nach oben hin stark vergrößern. Durch die Problematiken der Schwarz-Darstellung ergeben sich jedoch entsprechende Genauigkeitsprobleme im Bereich geringer Leuchtdichten, sodass z.B. eine Nachtsituation im Fahrzeug mit einer solchen Installation meist nicht mehr befriedigend dargestellt werden kann.

### Farbraumunterschiede

Neben der geringen Helligkeit könnte jedoch auch ein eingeschränkter Farbraum – sei es durch schlechte Projektoren oder Eigenschaften der Stereobrillen – für die geringe Farbsättigung in der CAVE verantwortlich sein. Daher wurden mit den colorimetrischen Meßeinrichtungen des Minolta-Meßgerätes die sechs Primär- und Sekundärfarben, sowie Weiß, mit und ohne Stereobrillen in einem definierten Farbraum vermessen.

Die resultierenden Chromazitätswerte sind in Abb. 17 für die untersuchten Displaygeräte zusammengestellt. In der CAVE war hierbei nur die vermessene Leinwand aktiviert, um Interreflexionen zu unterdrücken.

Überraschenderweise liegen alle gemessenen Werte sehr dicht beieinander und weichen nicht wesentlich von den Gamut-Werten eines Standard-CRT-Monitors ab. Die einzigen Werte, die signifikant abweichen betreffen die CAVE-Messungen durch die Shutterbrille. Im CIE-Chromazitätsdiagramm sollten Farbgamuts eigentlich die Form von Dreiecken annehmen, wobei Komplementärfarben auf am Weißpunkt gespiegelten Punkten der Dreieckskanten liegen müßten. Dies trifft auch auf die einzelnen Meßreihen zu – abgesehen von dem „CAVE with shutter glasses“-Diagramm. Es liegt daher nahe anzunehmen, daß diese

Meßreihe aufgrund der sehr geringen Leuchtdichten und der Empfindlichkeitspezifikation des Meßgerätes ungünstige Werte lieferte. Dies bestätigt sich in Abb. 17b, wo sowohl Shutterbrillen als auch die CAVE-Projektion einzeln betrachtet aufgrund der höheren Leuchtdichten in einem Monitor-Setup ein sehr viel besseres Verhalten zeigen.

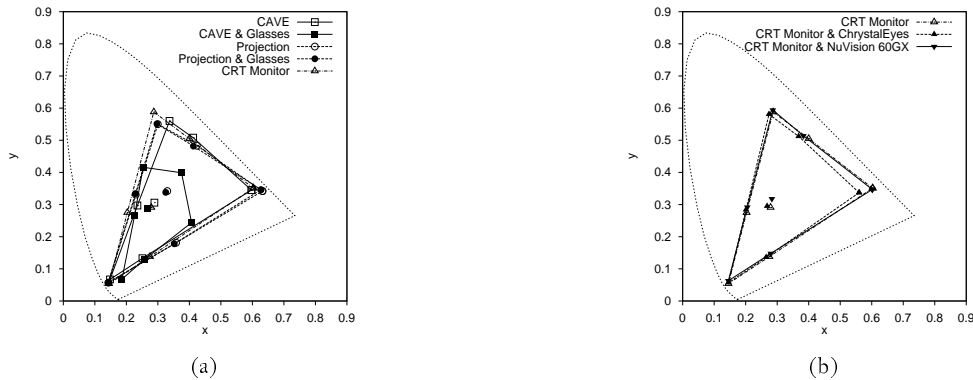


Abb. 17 : (a) Farbgamuts und Weißpunkte für CAVE und Aufprojektion, gemessen mit und ohne Stereobrillen. (b) Zum Vergleich die entsprechenden Werte für einen CRT-Monitor, sowie zwei unterschiedlichen Shutterbrillentypen.

## Bewertung

Die Ergebnisse legen nahe, daß die reduzierte Farbwahrnehmung in der CAVE nicht auf einen eingeschränkten Farbraum zurückgeführt werden kann.

Der Weißpunkt der Frontprojektion ist gegenüber dem eines Monitors verschoben. Man muß jedoch bedenken, daß dies eine reine Einstellungsache der RGB-Einzelintensitäten ist, nämlich eben genau der Farbtemperatur. Solange sich die Farbwerte für die drei Primärfarben zweier Displays decken, sind auch ihre Farbfähigkeiten vergleichbar.

Polarisationsbrillen sind nahezu farbneutral und ändern den zur Verfügung stehenden Farbraum nicht. Für die CAVE-Shutterbrillen-Situation können aufgrund der fehlerhaften Meßreihe keine verlässlichen Aussagen getroffen werden, jedoch ergaben die Monitor-Shutter-Tests, daß die Rot-Durchlässigkeit der CrystalEyes-Brillen erheblich reduziert war (55%), wodurch auch alle verwandten Farben wie Gelb, Violett, Orange und Braun beeinflusst waren. Die Durchlässigkeitswerte der NuVision-Brillen sind in diesem Bereich gleichförmiger.

### 3.2.4 Interreflexion

Im vorigen Abschnitt wurde festgestellt, daß der Bereich darstellbarer Farben in einer CAVE im Prinzip durch das komplette Projektionssetup hindurch unbeeinträchtigt bleibt.



Ein Teilaspekt sind sicherlich die Bedingungen des mesopischen Sehens – jedoch bleibt die Frage bestehen: „Warum erscheinen Farben in einer CAVE dumpf, mehr grau als farbig, und warum machen dargestellte Bilder einen verwaschenen, nebligen Eindruck?“. Schaut man sich die physikalischen Verhältnisse in einem CAVE-Würfel näher an, wird deutlich, daß der Grund hierfür fast ausschließlich in Interreflexionseffekten zwischen den Leinwänden zu suchen ist.

Dieser Effekt kann manchmal so stark sein, daß eine „schwarze“ Wand, ausschließlich indirekt von vier hellen Wänden beleuchtet, dieselbe Leuchtdichte besitzt, wie eine „weiße“ Wand – auch wenn dies aufgrund der menschlichen Wahrnehmung nicht so empfunden wird (es sei hierzu auf Kapitel 2.3.1 verwiesen). Der sichtbare Effekt ist letztendlich wiederum die Kontrastreduktion.

Ohne eine genaue Untersuchung gelangt man leicht zu der Überzeugung, daß diese Interreflexionseffekte durch geringe Leuchtstärken vermieden werden können – weshalb viele CAVEs oft nur mit 50% Projektorleistung oder weniger betrieben werden. Dieser Eindruck ist jedoch irreführend – der prozentuale Anteil der Interreflexion ändert sich nicht mit variierender Gesamtleuchtdichte! Einzuräumen ist allerdings, daß die Möglichkeit besteht, daß bei niedriger Gesamtleuchtdichte diese Effekte weniger auffällig erscheinen könnten.

Die zweite, häufige Vermutung ist, daß mit zunehmendem CAVE-Durchmesser die Interreflexionseffekte eigentlich abnehmen müßten. Doch auch dies bestätigt sich, nach einer strahlungstechnischen Analyse der geometrischen Beziehungen, als ebenfalls nicht zutreffend, wie im Folgenden gezeigt wird.

### Theoretische Überlegungen

Ideale CAVE-Leinwände sollten folgende Eigenschaften besitzen:

- Total diffuse Transmission ohne Absorption und Farbverschiebung
- Keine Reflexion

In der Realität ist der diffuse Reflexionskoeffizient von üblichen (und daher weißen) Leinwänden jedoch sehr viel größer als null. Um den tatsächlichen Reflexionskoeffizienten einer Leinwand zu messen, ist neben dem Leuchtdichtemeßgerät Spezialequipment in Form eines Beleuchtungsmeßgerätes erforderlich, welches leider nicht zur Verfügung stand. Die Radiositygleichung liefert jedoch eine ausreichende Grundlage, um auch ohne dieses Equipment den Koeffizienten zu bestimmen. Geht man von einer uniformen, konstanten Beleuchtungsstärke  $E$  durch die Darstellung eines vollweißen Bildes auf allen Seiten und einem für alle Leinwände einheitlichen Reflexionskoeffizienten  $\rho$  aus, kann der endgültige Radiositywert  $B_c$  durch folgende iterative Gleichung berechnet werden:

$$B_e^t = E + \rho \cdot \sum_{s \neq e} B_s^{t-1} \cdot F_{es}.$$

Der Formfaktor  $F_{es}$  in einer würfelförmigen CAVE ist  $F = 0.2$ , innerhalb einer Toleranz von 0.3%, je nachdem ob es sich um gegenüberliegende oder rechtwinklig zueinander stehende Seiten handelt. Dadurch kann die iterative Radiositygleichung in die folgende geometrische Reihe umgewandelt werden, die für jede würfelförmige CAVE mit den beschriebenen Eigenschaften und  $n$  Seiten Gültigkeit besitzt:

$$B_e = B_e^\infty = E \cdot \frac{1}{1 - (n-1)\rho F}$$

### CAVE-Umgebung

Um die direkte Beleuchtungsstärke  $E$  einer der Leinwände zu messen, erfolgte die Messung direkt auf der Leinwand, wodurch alle indirekten Effekte ausgeschlossen wurden. Zur indirekten Beleuchtungsmessung wurden die Wände in einem groben 3x3-Raster in vier Meßreihen mit wechselnden Beleuchtungssituationen vermessen, indem ein weißes (bzw. schwarzes) Bild auf die zu vermessende Wand (bzw. die vier übrigen Wände) projiziert wurde. Setzt man die so ermittelten Meßwerte  $E = 2.5 \text{ cd/m}^2$  und  $B_e = 4.0 \text{ cd/m}^2$  in die obige Gleichung ein und formt diese nach dem Reflexionskoeffizienten um, ergibt sich:

$$\rho = \frac{B_e - E}{0.8 \cdot B_e} \approx 47\%$$

Ein Reflexionskoeffizient  $\rho$  von 47% erzeugt ganz offensichtlich einen signifikanten Anteil an indirekter Beleuchtung in einer CAVE-Umgebung. Um nun den tatsächlichen Einfluß dieses  $\rho$  auf das visuelle Erscheinungsbild einer virtuellen Umgebung zu untersuchen, wurde ein 3D-Modell des CAVE-Würfels mit dem in Kapitel 2 vorgestellten Radiosityverfahren simuliert. Indem nun reale Snapshots einer virtuellen CAVE-Umgebung als Texturen auf dieses 3D-Modell gelegt und die Texturpixelwerte als Emissionswerte verwendet wurden, konnte mit Hilfe einer solchen Szene das tatsächliche Verhalten einer CAVE-Umgebung simuliert werden. Der besondere Vorteil des 3D-Modells und der Radiositysimulation war hierbei, nun jedes beliebige  $\rho$  als Materialeigenschaft der Leinwände angeben zu können.

Die Radiosityberechnung war hier in nahezu interaktiven Frameraten möglich, da die Szene aus nur fünf Polygonen bestand, und die einmal berechnete Linkstruktur bei neuen Materialwerten oder anderen Emissionstexturen einfach nur nochmals traversiert werden mußte.

Die berechneten Bilder kommen erstaunlich nahe an das tatsächliche Erscheinungsbild der realen CAVE heran (vgl. Abb. 18c und d), und es wird offensichtlich, daß der Hauptgrund für den schlechten Kontrast und die sich daraus ergebende geringe Farbsättigung in den Interreflexionseffekten zwischen benachbarten Leinwänden zu finden ist.



Abb. 18 : Radiositysimulationen für unterschiedliche Testszenen. Dunhuang-Szenario: (a) Optimaler Fall: Uniforme Leuchtdichteverteilung, keine Interreflexion. (b) Die tatsächliche Leuchtdichteverteilung, unter Beachtung von Abb. 14, jedoch noch ohne Interreflexionseffekte. (c) Der Reflexionskoeffizient wird auf 0.5 festgelegt, wodurch Interreflexion erzeugt wird. (d) Fotografie der realen Situation. VW Sharan Testszene: (e) Optimaler Fall. (f) Tatsächliches Erscheinungsbild. Schachbrett Testszene: (g) Optimaler Fall. (h) Tatsächliches Erscheinungsbild. BMW Testszene: (i) Tatsächliches Erscheinungsbild. (k) Streulicht durch einen in 3m Abstand vom Eingang platzierten, simulierten Monitor mit  $100 \text{ cd/m}^2$ .

## Bewertung

Welches ist nun die beste Vorgehensweise, um dieses doch erhebliche Manko einer CAVE-Umgebung zu verringern, oder gar auszugleichen?

Es wurde bereits festgestellt, daß dunkle Szenarien für eine lichtschwache CAVE am besten geeignet sind. Was geschieht nun, wenn der generelle Beleuchtungslevel z.B. durch den Einsatz leistungsstärkerer Projektoren angehoben wird? Die höheren Leuchtdichten werden die allgemeinen Beobachtungsbedingungen verbessern, indem photopisches Sehen ermöglicht wird. Jedoch werden proportional zur Gesamthelligkeit auch die sekundären Beleuchtungseffekte, also die Interreflexionen, zunehmen. Während das Verhältnis von direktem zu indirektem Licht unverändert bleibt, wird der verminderte Kontrast letztendlich durch das Nichtvorhandensein von Schwarz hervorgerufen: Dunkle Regionen der projizierten Szene werden heller, je mehr Licht von den hellen Bereichen zurückgestrahlt wird, wodurch letztendlich bei erhöhtem Reflexionskoeffizient auch die Farben ausbleichen.

In der Tat ergaben sowohl Messungen als auch Radiositysimulation, daß eine „schwarze“ CAVE-Seite von vier weißen Seiten beleuchtet fast dieselbe Leuchtdichte besitzt, wie eine weiße CAVE-Seite. Das Auge nimmt dieses Phänomen jedoch nicht direkt wahr: da die weißen Leinwände immer noch heller sind als die indirekt angestrahlte schwarze Seite, wird diese immer noch als „Schwarz“ interpretiert – jedoch mit entsprechend herabgesetztem Kontrast.

Unterschiedliche Konfigurationen, wie ein hypothetisches, optimales CAVE-Leinwandmaterial beschaffen sein müßte, um Kontrastreduktion durch Interreflexion zu vermeiden sind denkbar:

- Eine Leinwand, die Licht aus einer bestimmten Richtung (d.h., vom Projektor kommend) diffus in die CAVE streut, diffus *einfallendes* Licht jedoch absorbiert.
- Eine Leinwand, die von hinten einfallendes Licht diffus nach vorne streut, und von vorne einfallendes Licht nach hinten.
- Eine Leinwand, die 100% des vom Projektor einfallenden Lichts diffus in die CAVE streut, die aber kein von vorne einstrahlendes Licht reflektiert.

Die in Kapitel 3.1.4 vorgestellten „schwarzen Leinwände“ kommen dem dritten Typ am nächsten, indem sie zwar nur einen bestimmten Anteil des vom Projektor kommenden Lichts streuen, aber dafür kaum noch Licht reflektieren.

### 3.2.5 Projektorspezifische Problematiken

Im Zuge des CAVE-Umbaus am Fraunhofer-IGD wurden mehrere digitale Projektortypen bezüglich ihrer optischen und sonstigen Eigenschaften untersucht.

Dabei zeigten sich bei jeder Untersuchung neue Artefakte, die vorher nicht aufgefallen sind, oder die man in dieser Stärke überhaupt nicht erwartet hätte.

In Zusammenarbeit mit der Projektorfirma konnten einige dieser Probleme beseitigt oder zumindest verbessert werden, teilweise durch Austausch von LCD-Panelblöcken oder ganzer Projektoren. Es bleibt jedoch zusammenfassend zu sagen, daß bei digitalen Projektoren von den Herstellern viele oft vermeidbare Qualitätseinbußen in Kauf genommen werden, was eine bedenkliche Entwicklung darstellt, da hiervon auch High-End-Projektoren betroffen sind.

#### Geometrische Entzerrung

Das Problem der extrem limitierten geometrischen Entzerrung bei jeder Art von digitalem Projektor ist weitläufig bekannt und soll hier nur der Vollständigkeit halber erwähnt werden.

Während sich CRT-Projektoren in jeder vorstellbaren Art und Weise sogar für einzelne Bildbereiche geometrisch feinjustieren lassen, so ist man bei digitalen Projektoren auf einen Lensshift beschränkt – und auch diesen gibt es je nach Projektortyp manchmal nur vertikal, oder überhaupt nicht.

Viele Projektoren stellen immerhin ein digitales *keystoning*, also eine Trapezentzerrung zur Verfügung – es verbietet sich jedoch von selbst, diese einzusetzen, da hierdurch nicht nur Auflösung verloren geht, sondern meist heftige Aliasing-Artefakte und Moiré-Effekte erzeugt werden.

Steht die optische Achse nicht senkrecht auf der Bildmitte, oder ist die Bildebene gar nicht-planar, hat man bei digitalen Projektoren schlechte Karten, es sei denn man nimmt einen aufwendigen Softwareausgleich mit begleitendem Aliasing in Kauf.

#### Pixelstrukturen

Speziell bei LCD-Projektoren, aber auch bei D-ILAs und DLPs, ist das Verhältnis zwischen Pixelbreite und Abstand zum nächsten Pixel ein wichtiger Faktor, der die visuelle Qualität des projizierten Bildes (hier speziell die Erkennbarkeit von Pixelstrukturen) beeinflusst.

Bei LCD-Projektoren muß jedes Pixel in einem Panel einzeln von außen angesprochen werden können. Die hierfür erforderlichen Leiterbahnen verlaufen daher neben den Pixeln und vergrößern so den Bereich, der nicht den Pixeln selbst zugeordnet wird, was sich durch horizontale und vertikale schwarze Streifen zwischen den Pixelzeilen und -spalten bemerkbar macht.

Dabei ist die Breite dieser Zone offenbar völlig unabhängig von logisch erklärbar- re Gründen wie z.B. Projektorhersteller oder Panelauflösung: Bei drei Projektoren desselben Herstellers mit unterschiedlichen Auflösungen waren in der 1024x768er Variante die Streifen minimal (<10%) und sogar besser als bei einigen DLP-Projektoren, die des 1600x1200 Projektors erwartungsgemäß etwas schlechter (ca.20%) – die des Projektors mit der 1280x1024er Auflösung jedoch dermaßen miserabel (>50% des Pixelabstandes!), daß man bei kleinem Betrachtungsab- stand nur noch mit Mühe überhaupt ein zusammenhängendes Bild erkennen konnte.

Gegen diesen Effekt gibt es keine Gegenmaßnahme, außer der Wahl eines ande- ren Projektors.

### Farbkonvergenz

Da LCD-Panels monochromatisch sind, geschieht die Erzeugung eines farbigen Bildes durch Strahlauftrennung mit Hilfe dichroitischer Spiegel, und Weiterlei- tung durch drei getrennte Panels für die drei Primärfarben. Über einen Prismen- block werden die drei Strahlengänge wieder zusammengeführt.

Aufgrund geometrischer Ungenauigkeiten zwischen Panelposition relativ zum Pixelblock, sowie der Position des Pixelblocks an sich können hierbei Konver- genzprobleme in allen drei Primärfarben auftreten. Leider treten diese Probleme auch in der Praxis auf, und zwar nicht nur bei Low-Cost-Projektoren, sondern insbesondere auch bei den High-End-Varianten. Sich über die Bildfläche (linear) verändernde Konvergenzabweichungen von über zwei Pixeln(!) sind hierbei kei- ne Seltenheit.

Um Produktionskosten zu sparen, wurden Feinjustagemöglichkeiten der Panel- positionen meist wegrationalisiert, wodurch das Problem überhaupt nur durch Eingriffe mit dafür eigentlich ungeeignet erscheinenden Werkzeugen verbessert werden kann, indem z.B. mit einer Zange die Panels in ihren Fassungen regel- recht zurechtgebogen werden.

### Fokus

Ähnlich verhält es sich mit der Fokussierung der einzelnen Primärfarben: Nur zu oft sind unterschiedliche Fokussierebenen der einzelnen Primärfarben feststell-

bar, die noch dazu gegen die optische Achse individuell horizontal und/oder vertikal geneigt sein können.

Während sich Farbkonvergenz durch Verschiebungen in der Bildebene beschreiben lässt, liegen Abweichungen im Fokus in Verschiebungen der Panels entlang des Strahlengangs begründet. Meist ergibt sich eine Kombination aus beidem, wenn der Panelblock nicht optimal sitzt und leicht verdreht ist.

Wie bei der Konvergenz kann hier, sofern der Projektor keine Einstellmöglichkeiten vorsieht, wieder nur die Zange helfen (vgl. „optische Aufhängung“).

### Shading

Neben dem durch optisch-physikalische Eigenschaften bestimmten diffusen Highlight (vgl. Kapitel 3.1.1) kann es aufgrund elektronischer Abweichungen innerhalb der Panels zu statischen Helligkeits- und damit auch Farbschwankungen über die Bildebene kommen, die noch dazu bei unterschiedlicher Pixelintensität variieren können (siehe z.B. auch [SB99]).

Dies lässt sich zumindest bei manchen Projektoren durch (proprietäre) Einstellmöglichkeiten auf einem verhältnismäßig feinem Raster über die Panelflächen aufwendig (aber immerhin überhaupt) justieren.

Ist eine solche Justage nicht möglich, kann die in Kapitel 3.4.3 vorgestellte Ausgleichstextur eingesetzt werden.

### Optische Verzeichnung

Je nach Brennweite und Qualität der Projektionsoptik erzeugen die meisten Projektoren ein verhältnismäßig rechteckiges Bild. Geometrische Abweichungen von mehreren Pixeln insbesondere in den Ecken sind jedoch nicht ungewöhnlich. Diese Abweichungen beschränken sich im Allgemeinen jedoch nicht nur auf eine radiale Verzerrung, sondern es können auch wellenförmige oder andere, asymmetrische Abweichungen auftreten.

Ein weiterer Effekt ist eine Verzeichnung der Pixel, die über eine normale Unschärfe hinausgeht, Strahlenkränze bei einzelnen Primärfarben, „verschmieren“ der Pixelstrukturen entlang einer oder mehrerer Achsen, und so fort. Diese Effekte können durch Materialunreinheiten innerhalb der Optik, oder durch unsauber verarbeitete oder justierte Panels und Panelblöcke verursacht werden.

Hier helfen nur Austausch entweder der Optik, oder des gesamten Panelblocks – oder beides.

## Optische Aufhängung

Die Optiken fast aller untersuchten Projektoren waren sehr berührungsempfindlich. Die Aufhängung hatte teilweise soviel Spiel, daß eine leichte Berührung zu einem Pixelversatz von 5-10 Pixeln im projizierten Bild führte. Hinzu kommt durch den dadurch veränderten Strahlengang eine entsprechende Verkippung der Fokusebene.

Ursachen sind meist im Lensshift-System, aber auch in der mechanischen Verarbeitung der Objektivs selbst zu suchen.

Problematisch wird dieses Verhalten z.B. bei senkrechter Montage, da nun die Optik innerhalb ihres Spiels völlig frei beweglich ist und nicht von der Schwerkraft in eine stabile Lage gezwungen wird. Ähnlich kritisch ist die Deckenmontage: da die Optiken normalerweise für Standprojektionen justiert sind, fällt das Objektiv bei einer Deckenmontage in seine entgegengesetzte Maximallage, wodurch insbesondere die Fokusebenen erheblich verkippeln können.

Eine Möglichkeit, die Wirkungen etwas zu reduzieren ist das Unterlegen von dünnen Abstandshaltern in die Objektivhalterung, wodurch zwar das Spiel nicht reduziert wird, was aber die Fokusebene wieder senkrecht zur optischen Achse ausrichten helfen kann.

Eine professionelle Gegenmaßnahme, die noch dazu zwei Fliegen mit einer Klappe schlägt, ist eine feinjustierbare, mechanische Arretierung der Objektivs in beiden Achsen. So wird nicht nur die Stoßempfindlichkeit beseitigt, sondern es wird außerdem noch die Möglichkeit geschaffen, die Lage der Fokusebenen genauestens zu justieren.

## Staub

Staub- und Schmutzpartikel sind ein inhärentes Problem von LCD- und DILA-Projektoren. Aufgrund der großen Schärfentiefe digitaler Projektoren (es ist nicht ungewöhnlich, daß das Bild in einer räumlichen Tiefe von mehreren Metern erkennbar ist, während z.B. bei CRT-Projektoren das Bild bei variierendem Projektionsabstand sehr schnell sehr unscharf wird) und der Tatsache, daß insbesondere die LCD-Panels frei und luftumströmt im Projektor angebracht sind, können und werden sich sehr schnell Staubpartikel auf dem Panel ansammeln, welche aufgrund ihrer polarisationsverändernden Wirkung bei schwarzem Bild als vergleichsweise hell leuchtende, unscharfe, farbige Flecken erkennbar sind.

Gute Feinstaubfilterung der Luft im Projektionsraum – insbesondere bereits in der Aufbauphase – und ein wahrscheinlich nicht vermeidbares Ausbauen der



Projektoren und Ausblasen mit Preßluft in bestimmten Wartungsintervallen scheinen die einzig möglichen Maßnahmen gegen dieses Manko zu sein.

### Polarisierte Reflexionen

Ein interessanter Effekt konnte nach dem Einbau und Inbetriebnahme der Frontwall-Projektoren der IGD-CAVE (mit installierten Infitec-Filtern, bei denen der eine ein rötliches, der andere ein grünliches Bild erzeugt) festgestellt werden: Aufgrund der spekularen Reflexionseigenschaften der verwendeten Leinwände reflektiert das Bild der Frontprojektion stark an den Seitenwänden, sodass speziell von einem Betrachterstandpunkt einige Meter vor der CAVE die unbeleuchteten Seitenwände hell erscheinen. Diese Reflexionen waren jedoch *farbig*: die Reflexionen in den Leinwänden rechts und links waren deutlich grün, die in Decke und Boden deutlich rot.

Naheliegende Vermutungen wie eine Richtungsabhängigkeit von Infitec oder auch eine Richtungsabhängigkeit oder sich unterschiedlich verhaltende Vorder- und Rückseiten der Projektionsleinwände konnten relativ leicht widerlegt werden.

Stattdessen war die Ursache in dem von den LCD-Projektoren erzeugten, von Natur aus linear polarisierten Licht zu suchen. Reflexionen an dem Leinwandmaterial sind per Definition senkrecht zur Einfallsebene polarisiert, und zwar maximal bei einem Einfallswinkel, der dem Brewster-Winkel entspricht. Da sich aufgrund des Prismenblocks die Polarisationsrichtungen der einzelnen Farbkanäle unterscheiden (und zwar ist meist das grüne Licht um 90° gegenüber den anderen beiden Farben gedreht), verlässt bei dem verwendeten Projektor das grüne Licht den Projektor waagrecht polarisiert, während Rot und Blau annähernd senkrecht polarisiert sind.

Die einfache Lösung für das Problem war die Verwendung von  $\lambda/4$ -Filtern, die vor die Objektive montiert wurden. Diese optisch nahezu volltransparenten Filter verursachen eine Phasenverschiebung und Drehung einer der Polarisationsachsen, wodurch sie aus linear polarisiertem Licht zirkular polarisiertes Licht machen. Nun sind die Reflexionen zwar noch da – aber sie sind farblos.

### Bewertung

Sicherlich ist es so, daß viele der hier beschriebenen Problematiken bei „normalen“ Applikationen (Besprechungsräume, Großbildprojektionen) aufgrund des großen Abstandes der Zuschauer zur Leinwand nicht so offensichtlich zu Tage treten. Jedoch gerade in einem CAVE-Szenario, bei dem sich der Betrachter bei sehr großer Projektionsfläche außergewöhnlich dicht an der Leinwand befindet,

sind insbesondere Faktoren wie Pixelstrukturen, Farbkonvergenz und Fokus von extremer Wichtigkeit.

Es ist ratsam, sich – speziell beim Aufbau einer CAVE – unterschiedliche Projektoren genau anzusehen, bevor man sich für einen entscheidet. Steht der Typ einmal fest, ist es ratsam, aus mehreren Projektoren die Besten auszuwählen, und diese ggf. noch entsprechend ihrer farblichen Eigenschaften oder optischen Mängel für bestimmte Leinwände oder Positionen vorzusehen (z.B. die Leuchtstärken und Schärfsten nach vorne, auf gleichmäßige optische Verzerrung an den Leinwandübergängen achten, bei Verzeichnungen oder Unschärfen in der linken Hälfte an die linke Wand, bei schlechter Rotintensität nicht als Infitec-„Grün“-Projektor auswählen, usw.). Dies wurde für die IGD-CAVE und die HEyeWall (vgl. Kapitel 3.3.8) durchgeführt.

### 3.3 Displaykalibrierung

Großflächige Projektionstechnologien und immersive Displays werden immer verbreiteter und kommen heute insbesondere in vielen unterschiedlichen VR-Anwendungsbereichen zum Einsatz. Viele verschiedene Faktoren beeinflussen jedoch die Wiedergabequalität solcher Geräte, die stark von den verwendeten Projektoren und sonstigen Komponenten abhängt (vgl. Kapitel 3.1, und insbesondere 3.2.5). Farbraum und darstellbare Leuchtdichten sind oft eingeschränkt, und je nach Konfiguration (Mono/Stereo, Auflösung, ...) verhalten sich diese Medien unterschiedlich.

Aufgrund dieser Charakteristiken wirkt eine auf verschiedenen Displays dargestellte Szene jeweils sehr unterschiedlich. Selbst wenn alle Farben und Helligkeiten im darstellbaren Bereich des jeweiligen Displays liegen, so weichen die Darstellungen aufgrund unterschiedlicher interner Einstellungen voneinander ab, was insbesondere bei Kachelprojektionen durch regelrecht bunte Kachelsegmente den Gesamtbildeindruck zerstört. Diese Abweichungen sind selbst dann signifikant und mit bloßem Auge erkennbar, wenn Projektoren des gleichen Typs oder gar mit fortlaufenden Seriennummern zum Einsatz kommen.

In vielen Bereichen, in denen sich der Benutzer auf das was er sieht verlassen muß um Entscheidungen treffen zu können, sind korrekte Farb- und Helligkeitswiedergabe unabdingbar, insbesondere z.B. bei Design- oder ergonomischen Studien. Dies macht Methoden erforderlich, die (innerhalb einer vorgegebenen Fehlertoleranz) photometrische und colorimetrische Konsistenz für die Darstellung auf immersiven Displays sicherstellen können, sodass Farben und Leuchtdichten direkt auf der Leinwand mit einem Meßgerät überprüft werden können.

Dabei ist es zweitrangig, ob es sich bei den verwendeten Ausgabegeräten um Monitore oder analoge oder digitale Projektoren handelt – bei korrekter Kalibrierung lassen sich auf all diesen Varianten konsistente Daten 1:1 wiedergeben.

Es ist verhältnismäßig einfach zwei oder mehr Displays bezüglich eines gemeinsamen Weißpunktes durch Intensitätsregulierung der Farbkanäle abzugleichen. Die meisten Displays weisen jedoch ebenfalls erhebliche Farb(ort)unterschiede bei der Darstellung von Reinfarben auf. Abweichungen in den Primärvalenzen bedeuten unterschiedliche Farbräume (Gamuts), was letztendlich bedeutet, daß ein Displays bestimmte Farben darstellen kann, zu denen ein anderes nicht in der Lage ist, unabhängig davon, wie R, G und B gemischt werden. Ein drastisches Beispiel solcher sich unterscheidender Farbräume stellt das Infitec-Verfahren dar (vgl. Abb. 11). Als Resultat der spektralen Selektion ergibt die Darstellung eines beliebigen RGB-Wertes für beide Augen deutlich unterschiedliche Farb- und Helligkeitswerte (vgl. Abb. 26, linke Spalte).

Um zu garantieren, daß jeder Eingangs-RGB-Wert auf jedem der beteiligten Displays identisch aussieht, müssen die Farbräume aller Displays bestimmt, das Gammaverhalten vor und nach der Farbkorrektur ausgeglichen, und alle Eingangswerte unter Beachtung von Schwarzwerten mit Intensitäten ungleich 0 entsprechend transformiert werden [KRR03].

Der Nachteil jedes Farbraum- und Helligkeitskalibrierenden Ansatzes ist eine Einschränkung (d.h. Verkleinerung) des dem Display eigenen Farbraums, was eine geringere Farbsättigung zur Folge haben kann. Insbesondere definiert das dunkelste Display die darstellbare Maximalhelligkeit *aller* Displays.

Der wesentliche Vorteil ist völlige Farbgleichheit zwischen allen entsprechend kalibrierten Displays, unabhängig von ankommenden Eingangswerten (vgl. Abb. 26, rechte Spalte). Insbesondere können auch kalibriert vorliegende Eingangsdaten photometrisch und colorimetrisch konsistent wiedergegeben werden.

### Vorangehende Arbeiten

Das Problem unterschiedlicher Farbdarstellung auf Ausgabegeräten blieb lange Zeit von Entwicklern wie Anwendern von VR-Systemen unbeachtet. Erst kürzlich, mit dem Einsatz von VR-Applikationen in Bereichen, in denen eine korrekte Farbwiedergabe unentbehrlich ist, wurden entsprechende Studien von Charakteristiken immersiver Projektionsdisplays durchgeführt [BRB98][SKUM99]. Bisher wurden zur Verfügung stehende Helligkeits- und Farbmeßgeräte nicht für eine Kalibrierungsmethode für VR-Systeme mit konsistenten Simulationsdaten eingesetzt.

Was die colorimetrische Kalibrierung von Displays anbelangt, so gibt es entsprechende Vorarbeiten im Bereich von Monitoren (z.B. [V99]), jedoch behandeln diese nur Teilaspekte einer vollständigen Kalibrierung relevanter Aspekte immersiver Projektionsdisplays, betrachten z.B. nicht die Helligkeitsvarianzen auf der Leinwandfläche oder unterschiedliche Farbräume, und haben insbesondere die Einschränkung, daß sie nur für einzelne Meßpunkte Kalibrierungsdaten liefern können.

Standards wie das *ICC Farbprofil* [I97] wurden entwickelt, um im Bereich von Printmedien und Monitoren ein Austauschformat zur Verfügung zu stellen, das Informationen über Display- und Kalibrierungseigenschaften enthält. Ob jedoch die einzelnen Hersteller von Hardware (Displays, Drucker) oder Software (Kamera-, Grafikkarten- und Druckertreiber, Bildbearbeitungssoftware, ...) immer alle Möglichkeiten (und insbesondere Erfordernisse) dieses Formats ausschöpfen, ist fraglich. Des weiteren bleiben in dieser Spezifikation viele Fragen, wie z.B. uneinheitliche Leuchtdichteverteilung und der Farbraumabgleich mehrerer Displays aufeinander entweder ganz oder teilweise unbeantwortet.

Stone [S01b] erkennt viele der wichtigen Problematiken im Zusammenhang mit der Kalibrierung von digitalen Projektoren und der Konfiguration von Kachelprojektionen, insbesondere die Wichtigkeit von möglichst diffusen Leinwänden (vgl. Kapitel 3.3.8). Unter anderem werden Punkte wie die Korrektur des Schwarzwertes und das Finden eines gemeinsamen Farbgamuts angesprochen. Die Problematik des *white boosts* von DLP-Projektoren wird in [S01c] behandelt.

Majumder et al. [MHTW00] entwickelten einen Algorithmus, der ein paar gute Ansätze zur Kalibrierung von Multiprojektordisplays aufweist. Ihre Methode kann jedoch nicht mit sich unterscheidenden Farbgamuts umgehen und ist unmodifiziert daher nur für solche Setups anwendbar, die reine Helligkeits- und Weißpunktabweichungen aufweisen, nicht jedoch Unterschiede in den Farborten ihrer Primärvalenzen. Des weiteren wird die Problematik von unsauberem Schwarz nicht zufriedenstellend behandelt, und den dargestellten wird ein lineares Gammaverhalten von 1.0 aufgezwungen, was zum einen in einer unnatürlichen Darstellung resultiert, zum anderen Quantisierungseffekte aufgrund des auf 3x8 Bit pro Pixel limitierten Framebuffers erzeugt.

In dem Ansatz von Wallace und Chen [WC03], der zeitgleich mit [KRK03] auf dem Immersive Projection Technologies Workshop 2003 vorgestellt wurde, wird ein vielversprechender Farbgamut-Matchingalgorithmus vorgestellt, der dem hier vorgestellten Verfahren in Teilen vergleichbar ist. Es bleiben jedoch ebenfalls essentielle Problematiken wie ein Schwarzlevel ungleich null sowie unsaubere Gammfunktionen unbehandelt. Sie vermessen jeden Projektor mit über 2000 Samplewerten, was besonders bei Kachelprojektionen einen extremen

Aufwand erfordert. Das Ergebnis ist ein 3D-Lookup-Table, welcher über Pixelshader ausgewertet werden kann. Der Vorteil dieser Methode gegenüber der hier verwendeten einfachen Matrix ist, daß spezielle Verhaltensweisen wie beispielsweise der *white boost*, der bei bestimmten DLP-Arten hinzugeschaltet werden kann, ausgleichbar ist. [S01c] schlägt hierfür statt einer numerischen Vermessung des vollständigen Farbwürfels einen algorithmischen Ansatz und die Verwendung einer erweiterten Farbmatrix vor.

### 3.3.1 Erforderliche Messungen und Kalibrierungsschritte

#### Messwerte

Für jedes Display und jede Farbkomponente müssen Leuchtdichte ( $\text{cd/m}^2$ ) und Farbort (CIE- $xy$ ) bei mehreren Intensitäten gemessen werden. Durch die Ansteuerbarkeit des verwendeten Farbmeßgerätes über PC können so verhältnismäßig schnell viele Messwerte in kleinen Intervallen erzeugt werden.

Die R-, G- und B-Intensitäten bei voller Intensität ergeben hierbei die maximalen Leuchtdichten und Farborte der einzelnen Farbkanäle, woraus die Transformationsmatrix des Displays abgeleitet werden kann. Die einzelnen vollständigen Kurven repräsentieren das Gammaverhalten des Displays.

Leuchtdichte und Farbort müssen hierbei ebenfalls für Schwarz aufgenommen werden. Da dieses Schwarz als letzter Schritt in der photometrischen Pipeline vom Projektor zu einer dargestellten Farbe immer hinzuaddiert wird, muß es von allen Messwerte zu Beginn subtrahiert werden. Geschieht dies nicht, wird die Farbinformation insbesondere von Farben niedriger Intensität verfälscht (vgl. Abb. 19 und Abb. 20). Des weiteren würden kalibrierte Eingangswerte um den entsprechenden Anteil inkorrekt wiedergegeben.

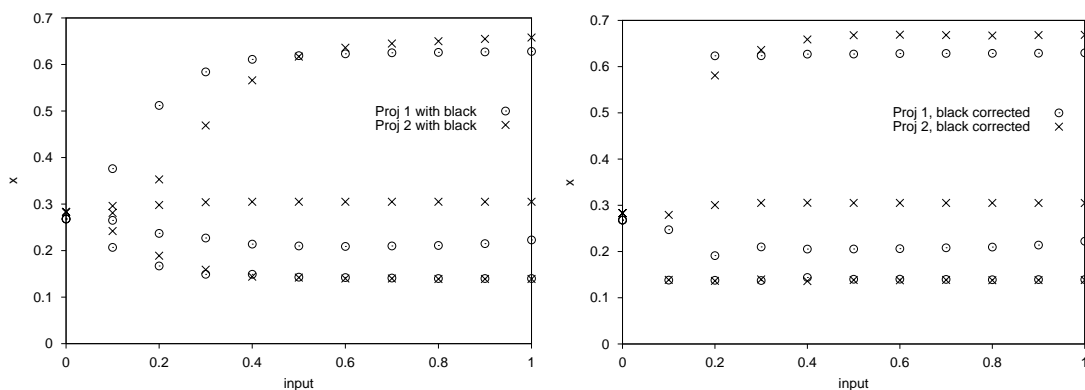


Abb. 19 :  $x$  Farbkoordinaten für rot, grün und blau unkorrigiert (links) und korrigiert (rechts). Man beachte daß sich im rechten Diagramm die Nichtlinearitäten von Projektor 2 oben links ( $x$  Farbkoordinate der grünen Primärvalenz) sowie Projektor 1 unten links ( $x$  Farbkoordinate der roten Primärvalenz) durch die extrem niedrigen Intensitäten dieser Farben begründen (vgl. Tabelle 6).

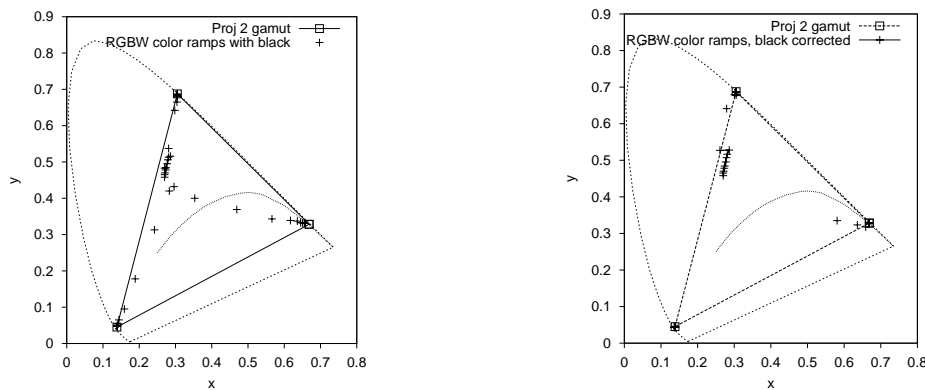


Abb. 20: R-, G-, B- und Graustufen-Farbwerte ohne (links) und mit (rechts) Schwarzkorrektur. Die Punkte der Graukurve verbleiben verzerrt, da aufgrund ungleicher Gammakurven der Farbkanäle die Grauwerte in der Tat unterschiedliche Farbwerte aufweisen.

Man beachte jedoch, daß sich die Farbwerte aus Abb. 19 und Abb. 20 durch rechnerische Subtraktion des Schwarzlevels ergeben – der tatsächliche Gamut darstellbarer Farben verkleinert sich aufgrund des Schwarzanteils mit sinkender Intensität zwangsläufig immer weiter, denn wie bei der Problematik des reinen Schwarzes kann ein vorhandener Schwarzlevel nicht durch Software verbessert werden. Es *kann* jedoch sichergestellt werden, daß die durch den Schwarzlevel verursachten Abweichungen in die Kalibrierung mit eingehen und so eine konsistente Darstellung garantiert wird.

### Kalibrierungsschritte

Damit das VR-System eine Szene konsistent darstellen kann, muß es Kalibrierungsinformationen zu Verfügung gestellt bekommen. Diese Kalibrierungen müssen daher im Voraus entsprechend den Charakteristiken des Displays vorgenommen werden.

Die Displaykalibrierung befaßt sich somit mit den in Kapitel 3.2 als am relevantesten für die konsistente Bildwiedergabe identifizierten Faktoren:

*Gleichmäßigkeit:* Wie verändern sich Leuchtdichte und Weißpunkt über die Leinwandfläche?

*Colorimetrie:* Was ist der Farbraum des Displays?

*Photometrie:* Was ist die maximal darstellbare Leuchtdichte, und wie weit weicht ein Leuchtdichteverlauf von einer linearen Funktion ab (d.h., was ist die Gammafunktion)?

Aus den einzelnen Messungen und Kalibrierungsschritten werden folgende Informationen gewonnen:

- Eine transparente Ausgleichstextur, um eine nicht-uniforme Leuchtdichtenverteilung inklusive Farbverschiebungen zu beseitigen.
- Eine Konvertierungsmatrix von CIE XYZ-Farben in den Farbraum und Leuchtdichtebereich des Displays.
- Drei Gammakurven des Displays, die das nichtlineare Intensitätsverhalten der einzelnen Farbkanäle beschreiben.

Diese Daten werden anschließend vom VR-System verwendet, um eine konsistente Darstellung zu ermöglichen.

Die Bestimmung der Ausgleichstextur, sowie eine automatisierte Kalibrierung auf Basis aller drei Punkte wird in Kapitel 3.4 diskutiert.

## Übersicht

Zusammenfassend sind also drei Punkte erforderlich, um eine konsistente Wiedergabe sicherzustellen:

1. Eine Szenenbeschreibung mit korrekten Farb- und Helligkeitsinformationen, welche dadurch insbesondere unabhängig vom verwendeten Display ist. Oder die Verwendung von unkalibrierten Bildern mit dem Ziel, sie auf mehreren Displays identisch darzustellen.
2. Eine Möglichkeit, die Charakteristiken von Display zu messen und entsprechende Kalibrierungsdaten zur Verfügung zu stellen.
3. Eine Gruppe von Tools, basierend auf der Kenntnis der Displaycharakteristiken, die in das VR-System integriert werden und so eine konsistente Bildwiedergabe für verschiedene Ausgabemedien ermöglichen.

Da Kapitel 2 den ersten Punkt bereits erfüllt, wird sich dieses Kapitel mit dem zweiten Punkt befassen, während der letzte Punkt in Kapitel 3.4.4 behandelt wird.

### 3.3.2 Bestimmung des gemeinsamen Gamuts

Der erste Schritt besteht nun darin, den Bereich darstellbarer Farben zu bestimmen, der von *allen* beteiligten Displays wiedergegeben werden kann. Die individuellen Intensitätsfähigkeiten sind für diesen Schritt irrelevant.

Nachdem ein beliebiges der beteiligten Displays als Startpunkt ausgewählt wurde, werden die RGB-Primärvalenzen des nächsten Displays in dessen Farbraum

gemappt. Treten hierbei negative Werte auf bedeutet dies, daß die entsprechende Primärfarbe nicht vom ersten Display dargestellt werden kann.

Ist nur eine der Farbkomponenten negativ, muß sie im CIE-xy Farbraum in Richtung der durch die Farbkomponente definierten Primärvalenz des zweiten Displays geclippt werden. Hierbei kann ein Sonderfall auftreten, bei der die Verbindungslinie zwischen Original- und Zielprimärfarbe den Originalfarbraum nicht schneidet. In diesem Fall liegt jedoch die Originalprimärfarbe *innerhalb* des zweiten Gamuts, und stellt daher die neue, gemeinsame Primärfarbe dar.

Bei zwei negativen Komponenten muß der Test mit vertauschten Rollen wiederholt werden, indem nun die Primärfarben des ersten Displays in den Farbraum des zweiten gemappt werden. Dadurch reduziert sich aufgrund der geometrischen Beziehungen die Maximalzahl von negativen Werten auf eins.

Sind auf diese Weise die neuen Primärfarben gefunden, die von beiden Displays dargestellt werden können, wird dieser neue Farbraum als Ausgangspunkt herangezogen und mit dem nächsten Display fortgefahren.

Eine andere Variante mit gleichem Ergebnis ist, durch direktes 2D-Dreiecksclipping aller RGB-Primärvalenztripel im CIE-xy-Raum das größte eingeschriebene Dreieck der Schnittmenge zu finden. Dies ist der gemeinsame Farbraum aller Displays.

Für den so gefundenen Farbraum ist garantiert, daß alle darin enthaltenen Farben auf allen beteiligten Displays dargestellt werden können (Abb. 21).

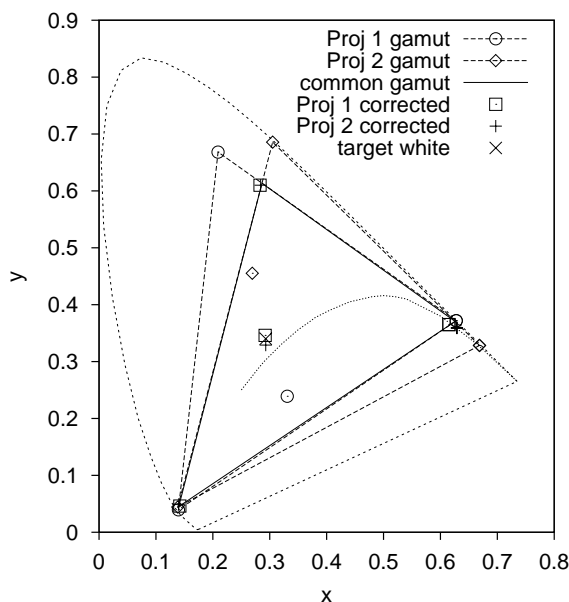


Abb. 21: Verifikation der Bestimmung des gemeinsamen Gamuts für zwei Infitec-Projektoren



### 3.3.3 Ermittlung der maximalen Leuchtdichten

Nach der Bestimmung des gemeinsamen Farbraums müssen nun für jedes Display die maximalen Intensitäten gefunden werden, mit der diese neuen Primärvalenzen dargestellt werden können.

Da die Summe dieser drei Maximalprimärvalenzen den neuen gemeinsamen Weißpunkt aller Displays definiert, ist es sinnvoll, von den maximal möglichen Intensitäten abzuweichen. Auf diese Weise kann die Lage des Weißpunktes frei definiert werden: durch Vorgabe eines neuen Weißpunktes, z.B. dem natürlichen Weißwert eines der Displays, oder einem Standardweißpunkt wie z.B.  $D_{65}$ , und entsprechender Skalierung der neuen Primärfarben kann der neue Farbraum – und damit alle beteiligten Displays – bezüglich diesem neuen Weißpunkt kalibriert werden. Insgesamt muß natürlich darauf geachtet werden, daß keine Farbkomponente der einzelnen Displays größer als eins werden darf.

Für jedes Display wird also der vorgegebene Weißpunkt in den Original-Displayfarbraum transformiert und dessen Helligkeit so skaliert, daß die für alle Displays größte Farbkomponente genau eins ist. Hierbei muß jedoch auf korrekte Integration des Schwarzwertes geachtet werden.

Aus den resultierenden gemeinsamen Maximalprimärvalenzen und den Originalwerten jedes Displays lassen sich so RGB-RGB-Konvertierungsmatrizen berechnen, die beliebige, ankommende RGB-Werte zwischen 0 und 1 in den gemeinsamen Farbraum umrechnen.

Das Ergebnis ist, das ein Eingangswert von z.B.  $RGB=(1,0,0)$  für jedes Display in neuen RGB-Werten resultiert, wobei das „reine“ Rot, das allen Displays sowohl bezüglich Farbort als Intensität gemeinsam ist, nun für jedes Display aus einer Kombination seiner einzelnen Primärvalenzen zusammengesetzt ist.

Um die so transformierten RGB-Werte darstellen zu können, muß jedoch zunächst das nichtlineare Verhalten der Displays korrigiert werden.

### 3.3.4 Kompensation nichtlinearen Verhaltens

Jedes Ausgabegerät weist eine eigene Kennkurve beim Umwandeln von Pixelwerten in Leuchtdichten auf. Diese Kennkurve hängt von unterschiedlichsten Faktoren ab, wie z.B. Displaytechnologie, Projektortyp, und insbesondere Einstellungen wie dem *blacklevel* (landläufig als „Helligkeit“ bekannt), und der *intensity* (dem „Kontrast“) (vgl. auch Kapitel 3.4.3, Abschnitt „Photometrie“).

Aufgrund des nichtlinearen Verhaltens des menschlichen Sehapparates, welcher Intensitätsänderungen in dunklen Bereichen empfindlicher ist gegenüber solchen

in hellen Bereichen, kommt oftmals eine exponentielle Funktion – der Gamma-wert – bei bildbezogenen Ein- und Ausgabegeräten zur Anwendung. So befindet sich z.B. bezogen auf die volle Intensität bei einem Pixelwert von 255 der entsprechende Pixelwert für 10% dieser Intensität nicht bei 25.5, sondern bei einem Pixelwert von etwa 100, wenn der sehr gebräuchliche Gammaexponent von 2.4, welcher bis auf geringe Abweichungen üblichen CRT-Monitoren und auch TFT-Displays zugrunde liegt, zum Einsatz kommt.

Die Annahme eines festen Exponenten ist jedoch nur selten exakt. Insbesondere z.B. bei digitalen Projektoren kommt meist eine S-förmige Kurve zur Anwendung.

Farbumrechnungen und Schwarzwertkorrektur müssen in einem linearen Raum durchgeführt werden. Daher müssen die Kennkurven des Displays (eine pro Farbkanal) vermessen und jede linear berechnete Farbtransformation zunächst mit den Inversen dieser Kennkurven transformiert werden, bevor das Ergebnis an das Ausgabedisplay geschickt werden kann. Das Display wendet dann auf diese neuen Werte seine Kennkurven an, wodurch dann korrekte (und lineare) Leuchtdichten dargestellt werden. Die Vermessung muß hierbei für jeden Farbkanal getrennt vorgenommen werden. Die Kurven sollten bei einem guten Display zwar für alle Kanäle identisch sein (denn nur dann ist ein konstanter Farbbort für z.B. alle Grauintensitäten garantiert), dies ist jedoch im Allgemeinen nicht der Fall, und es können teilweise signifikante Abweichungen in den einzelnen Farbkanälen auftreten.

### 3.3.5 Ersetzen der Gammakurve

Eine Darstellung von Bildern mit unkalibrierten Eingangsdaten an dieser Stelle würde diese jedoch verzerren: sie würden mit einem Gammawert von 1.0 gemappt werden, nämlich eben einer linearen Kurve, wodurch dunkle Bildbereiche viel zu hell dargestellt würden, denn unkalibrierte Bilder sind üblicherweise so gemappt, daß sie auf einem unkalibrierten Display mit besagtem Gamma von ca. 2.4 „normal“ aussehen.

Daher müssen alle Eingangswerte vor der Farbtransformation mit einer Gamma-korrektur in den linearen Raum überführt werden. Der naheliegende Gedanke wäre hier, die ausgemessenen Kennkurven für diese Prozedur heranzuziehen: Die Farbkonvertierung zwischen Eingangswerten und Display würde dann zuerst die Displaykennkurven verwenden, die Farbtransformation durchführen, und anschließend den Nichtlinearitätsausgleich durch Anwendung der inversen Kurven negieren.

Diese Vorgehensweise ist jedoch unnötig und würde nur die Unstetigkeiten und möglichen Unterschiede zwischen den einzelnen Kurven der Farbkanäle auf Neu in die Darstellungspipeline einbringen.

In der Tat gibt es keinen wirklichen Grund, die Displaykennkurven für diesen Schritt zu verwenden. Es ist statt dessen einfach nur eine beliebige, ideale Funktion erforderlich, welche die (vor-gammakorrigierten) Eingangswerte in den linearen Raum transformiert. Diese Funktion kann z.B. die gemittelte Kurve der vermessenen Displaykennkurven sein, oder es kann die Kennkurve einer digitalen Kamera verwendet werden, wenn z.B. Bilder dieser Kamera dargestellt werden sollen.

Aber es ist noch praktischer, einfach eine gleichmäßige Funktion wie z.B. den Gammaexponent 2.4 zu verwenden. Letztendlich ist diese Funktion nur dazu da, die (im Allgemeinen unbekannt) Gammatransformation, die bereits in den Eingangsdaten enthalten ist, rückgängig zu machen.

Nachdem auf diese Weise nun die Eingangswerte in den linearen Raum transformiert wurden, kann die eigentliche Farbtransformation durchgeführt werden.

### 3.3.6 Die Farbe „Schwarz“

Wie in Kapitel 3.1 festgestellt, leiden digitale Projektoren insbesondere an einem hohen Schwarzlevel. Es ist natürlich nicht möglich, das Schwarz an sich zu verbessern – jedoch muß es in einem Farbkonsistenzalgorithmus gesondert und korrekt behandelt werden, sonst geht die colorimetrische Konsistenz verloren.

In jedem Fall muß „Schwarz“ als Farbe behandelt werden, um Konsistenz zu gewährleisten: Projektorschwarz mag im Vergleich zu den anderen Farben dunkel sein, aber es enthält trotzdem noch Farbinformation und wird so Farben niedriger Intensität beeinflussen und verfälschen (Abb. 19 und Abb. 20).

Da es sich bei Schwarz um einen additiven Term handelt, muß es aus den Farb-raumberechnungen herausgerechnet werden, da es sonst mitskaliert würde. Stattdessen ist eine Sonderbehandlung erforderlich.

Es gibt prinzipiell drei Varianten, wie mit einem Schwarzlevel ungleich null umgegangen werden kann. Welche Methode für eine jeweilige Anwendung am Sinnvollsten ist, kann nur vom Benutzer selbst entschieden werden, da keine der möglichen Alternativen alle drei Vorteile visuelle Qualität, photometrische Korrektheit und maximalen Kontrast vereinbaren kann:

### a) Schwarz ignorieren

Wird der Schwarzlevel außer Acht gelassen, so ist weder photometrisch oder colorimetrisch konsistente Wiedergabe, noch die konsistente Darstellung bei mehreren Displays möglich. Der induzierte Fehler wird hierbei immer größer, je dunkler die darzustellende Farbe ist, da das Ausgabegerät seinen Schwarzlevel zu den berechneten Farbwerten hinzuaddiert.

### b) Nicht darstellbare Werte clippen

Das vermessene Schwarz kann in den Display-Farbraum umgerechnet und so dann von den berechneten Werten nach der Farbkonvertierung (aber noch im linearen Raum) abgezogen werden. Entstehen hierbei negative Komponenten, bedeutet dies daß die entsprechenden Farbkanäle unterhalb des Schwarzlevels liegen müssten, was nicht möglich ist. Diese Werte werden daher an 0 geclippt.

Alle nicht geclippten Farben können jedoch photometrisch exakt und **ohne jeglichen Einfluss durch schlechtes Schwarz** dargestellt werden, wodurch insbesondere der Farbkontrast verbessert wird. In der Tat wird also das Display-schwarz als wirksamer Bestandteil der darzustellenden Farbe genutzt, indem die Soll-Farbe minus dem Schwarzlevel an das Display gesendet wird.

Dunklere Farben werden auf das Displayschwarz abgebildet, wodurch für den Betrachter eine Art „Nebeleffekt“ entsteht, da erst ab einer bestimmten Grundhelligkeit bzw. -farbe Details sichtbar werden.

Je nach Stärke des Schwarzlevels kann dies jedoch bedeuten, daß ein erheblicher Teil an Detailinformation in dunklen Bildbereichen verloren geht. Insbesondere können sich diese Clippinggrenzen bei benachbarten Displays unterscheiden, was zwar oberhalb des schlechtesten Schwarzlevels photometrisch konsistent sein mag, aber vom menschlichen Auge als störend empfunden wird.

Ein Vorteil dieser Methode, der jedoch entscheidend sein kann ist, daß abgesehen von den ganz dunklen Bildbereichen der Rest des Bildes in einer Farbbrillanz dargestellt wird, als wenn perfektes Schwarz möglich wäre. Insbesondere erhält diese Alternative die maximale Bilddynamik, was bei der physikalisch korrekten Visualisierung wichtiger sein kann als optimale visuelle Qualität.

Das Ergebnis ist ein **individuelles Schwarz** für jedes beteiligte Display.

### c) Das schlechteste Schwarz verwenden

Die dritte Methode ist, das „kleinste gemeinsame Vielfache“ der einzelnen Display-Schwarz zu finden, indem die Schwarzwerte nacheinander in die Displayfarbräume transformiert und an 0 geclippt werden (Vorsicht ist hierbei jedoch

geboten, wenn ein Schwarz außerhalb des Gamuts seines eigenen Displays liegt, was nicht ausgeschlossen ist).

Mit dieser Vorgehensweise ist es möglich, sowohl photometrische Konsistenz zu garantieren, als auch visuelle Qualität (d.h. keine Sprünge durch geclippte dunkle Farben) zu wahren.

Entscheidender Nachteil ist jedoch ein nicht zu vernachlässigender Kontrastverlust – insbesondere wenn sich die Schwarztöne farblich stark unterscheiden, da hier die Gesamthelligkeit des gemeinsamen Schwarzes unnatürlich weit angehoben werden muß. Bei reinem Abgleich von Displays aufeinander (d.h. ohne kalibrierte Eingangsdaten) wird außerdem der Farbkontrast reduziert, da dunkle Farben zum Farbart des gemeinsamen Schwarzes konvergieren.

Diese Variante ist allerdings als einzige in der Lage, ein gleichmäßiges „schwarzes“ Bild (oder genauer: für jede beliebige Farbkombination ein gleichmäßiges Bild) darzustellen, was insbesondere bei gekachelten Projektionen von Wichtigkeit sein kann.

Das Ergebnis ist also ein **gemeinsames Schwarz** für jedes beteiligte Display.

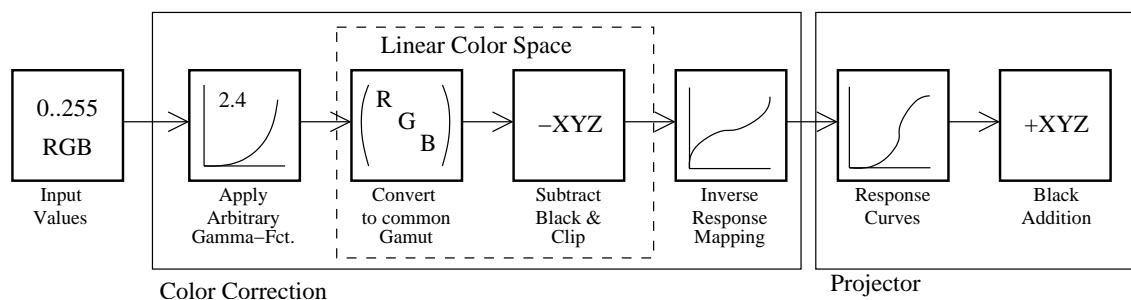


Abb. 22: Die vollständige Verarbeitungspipeline zur konsistenten Darstellung unkalibrierter Eingangsdaten auf mehreren Displays.

### 3.3.7 Konsistente Eingangsdaten

Für die konsistente Darstellung bereits kalibrierter Eingangsdaten, wie z.B. Ergebnissen einer Radiositysimulation entsprechend Kapitel 2, Bildern in einem definierten Farbraum, high dynamic range images, etc. vereinfacht sich der in Abb. 22 für unkalibrierte Daten dargestellte Ablauf entsprechend. Die Gamma-korrektur zu Beginn, sowie insbesondere die Bestimmung eines gemeinsamen Farbraums sind nicht mehr erforderlich, und das Schwarz (gemeinsam oder individuell) kann direkt von den Eingangsdaten subtrahiert werden.

Im Gegensatz zur Transformation von Eingangswerten zwischen 0 und 1 ist es nun jedoch möglich, daß manche Farben nicht mehr dargestellt werden können, da sie außerhalb des Displayfarbraumes liegen oder zu hell sind.

Treten Farben mit zu großen Leuchtdichten auf, können Tonemapping-Verfahren eingesetzt werden [TR93][S94][WRP97][SSS00], wobei jedoch photometrische Konsistenz insgesamt verloren geht. Um diese Konsistenz zumindest für die darstellbaren Bereiche zu wahren, müssen nicht darstellbare Bildbereiche z.B. durch Fehlfarben kenntlich gemacht werden.

Für Bildbereiche außerhalb des Displayfarbraums ist colorimetrisch konsistente Darstellung nicht möglich. Solche Farben können entweder ebenso kenntlich gemacht werden, oder werden auf eine darstellbare Farbe geclippt. Im einfachsten Fall erfolgt dieses Clipping im  $xy$ -Farbraum durch Verschiebung in Richtung des Display-Weißpunktes, oder des Energiegleichgewichts-Weißpunktes  $E$  ( $x=y=0.33$ ). Der menschliche Sehapparat wird hierbei jedoch Farbverschiebungen wahrnehmen, die durch geeignetere Clippingmethoden vermieden werden können [A10][H89][ML97].

### 3.3.8 Beispielszenarien Digitale CAVE und HEyeWall

Als Beispielsetup für Infitec und entsprechenden Farbausgleich mehrerer Displays dienen die 2003 am Fraunhofer-IGD installierten immersiven Stereodisplaysysteme einer 5-seitigen CAVE mit digitalen Projektoren, sowie einer 6 zu 4 gekachelten Projektion aus 48 Projektoren. Diese werden betrieben von einem (10+1)- bzw. (48+1)-PC-Linux-Cluster. Als farbkonsistenzfähige Renderingsoftware wird in beiden Fällen das VR-System Avalon als gemeinsame Entwicklung von Fraunhofer-IGD und ZGDV e.V. Darmstadt verwendet, dem der ebenfalls am Fraunhofer-IGD entwickelte Open-Source-Szenengraph OpenSG zugrunde liegt.

#### Digitale CAVE

Die neue Version der Fraunhofer-IGD-CAVE basiert auf dem gleichen Grundgerüst wie das in Kapitel 3.1.5 beschriebene Setup. Ein Holzrahmen, um elektromagnetisches Tracking zu ermöglichen. Eine verhältnismäßig dünne Paraglas-Platte. Um eine bessere Helligkeit zu erhalten, kommen nun 10 digitale Projektoren zum Einsatz, die im passiven Stereobetrieb die 5 CAVE-Leinwände ausleuchten. Da das projizierte Bild digitaler Projektoren nicht nennenswert geometrisch entzerrt werden kann, wurden die Folien Spiegel durch hochplanare, bedampfte Glas Spiegel ersetzt. Da ohne Shuttering keine



Abb. 23 : Die digitale IGD-CAVE

besonderen Anforderungen mehr an die Grafikkarte gestellt wurden, kann die neue CAVE durch einen einfachen Beowulf-Cluster mit nVidia GeForce 4 betrieben werden.

Zirkulare Polarisation kam aufgrund der schlechten Kanaltrennung und insbesondere da der Boden die Polarisation nicht erhält nicht in Frage, womit Infitec als einzige verwendbare Stereotechnik verblieb. Durch die in diesem Kapitel vorgestellten Kalibrierungsmethoden ist jedoch selbst ein solch extrem farbverfälschendes Verfahren wie Infitec zu photometrisch und colorimetrisch konsistenter Darstellung in der Lage, wie noch gezeigt werden wird.

Als Projektoren kamen nach intensiven Untersuchungen (vgl. Kapitel 3.2.5) 10 LCD EIKI LC-UXT1 zum Einsatz, deren Hauptvorteile in einer sehr guten Helligkeit, sowie gutem Pixel-zu-Leiterbahn-Verhältnis liegen.

Die Auflösung der Projektoren liegt bei 1600x1200, von der aufgrund der quadratischen Projektionsflächen 1200x1200 genutzt werden. Durch Montage der Projektorpaare nebeneinander und Nutzung der jeweils inneren 1200 Pixel steht die optische Achse senkrecht und zentriert auf den Projektionsflächen. Hierdurch werden geometrische Verzerrungen ausgeschlossen, die bei digitalen Projektoren nur schwer oder gar nicht ausgeglichen werden können.

Aus Auflösung und Projektionsmaßen ergibt sich eine Breite pro Pixel von 2mm. Da sich des weiteren in einem CAVE-Szenario der Betrachter stets sehr dicht an einer der Leinwände befindet (meistens weniger als einen Meter), ergeben sich sehr hohe Anforderungen an die Abbildungsqualität der Pixel – insbesondere bezüglich der Farbkonvergenz und Schärfe in allen Bildbereichen – die bei „normalen“ Projektionssetups kaum ins Gewicht fallen.

Neben den Projektoren der wichtigste Faktor für ein qualitativ hochwertiges Display sind die Leinwandeigenschaften. Die alten Leinwände waren zwar sehr gut, was ihre diffuse Charakteristik betraf, jedoch erzeugte die hohe Reflektivität einen nicht tolerierbaren Anteil an Interreflexionen, wodurch der Kontrast in der CAVE entscheidend herabgesetzt wurde (vgl. Kapitel 3.2.4). Das Wunschmaterial „Schwarze Leinwand“ existiert jedoch noch nicht in zufriedenstellender Qualität, und insbesondere in den erforderlichen Größen von mindestens 2.4x2.4 Metern. Bis heute gibt es nur einen einzigen Hersteller, der schwarze Leinwände mit hinreichend diffusen Eigenschaften liefern kann. Jedoch wird deren visuelle Qualität durch uneinheitliche Transmissionseigenschaften („Wölkcheneffekt“) stark herabgesetzt, sodass schließlich ein Kompromiß zwischen Reflektivität und Diffusität getroffen werden mußte. Die Stewart FilmScreen 150 als „beste der schlechten Alternativen“ kam daher zum Einsatz, mit Klarglasscheiben als

Trägermaterial. Ein wirklich zufriedenstellendes Leinwandmaterial für eine CAVE-ähnliche Umgebung gibt es weltweit bis heute nicht.

## HEyeWall

Das Ziel der HEyeWall als Kachelprojektion war die Konstruktion einer hochqualitativen, hochauflösenden Stereoprojektion auf Basis von bezahlbaren Standardkomponenten. Während in letzter Zeit mehrere Kachelprojektionen mit Digitalprojektoren entstanden sind [W94][CPSD97][HJS00][FL00][WC03], so ist außer der HEyeWall bis heute noch keine von ihnen stereofähig. Etwa zeitgleich mit der HEyeWall wurde eine 4x2-Stereoprojektion in den Niederlanden errichtet [GSSB03], die jedoch aufgrund der verwendeten Polarisierungstechnik neben Kanalübersprechen insbesondere mit starken richtungsabhängigen Highlights zu kämpfen hat.

Das zentrale Problem einer solchen Projektion ist die Vermeidung von sichtbaren Kachelübergängen, wodurch der Eindruck eines einzigen, einheitlichen Gesamtbildes zerstört werden kann. Der einfache Ansatz, die Projektionen überlappen zu lassen und eigens aufwendige Algorithmen zu entwickeln, diese zu justieren [RWCL98][RBYC99][S99][CCFH00] oder gar photometrisch anzugleichen [MS02], scheitern spätestens dann, sobald ein schwarzes oder hinreichend dunkles Bild dargestellt wird: das Schwarz überlappender Bereiche summiert sich auf und kann somit an Kreuzungspunkten viermal so hell sein wie normal. Sind dann noch die Projektoren arbiträr ausgerichtet, entsteht ein bunter Flickenteppich aus unterschiedlich breiten und hellen Streifen und Flecken.

Einige der angesprochenen Problematiken im Zusammenhang mit Kachelprojektionen werden z.B. in [S01b] analysiert.

Weitere Nachteile eines solchen beliebig ausgerichteten Setups ergeben sich durch die nicht-achsparallele Ausrichtung der Pixelraster, wodurch deutliche Aliasing-Artefakte bei harten horizontalen oder vertikalen Gradienten, welche z.B. bei einer bildschirmfüllenden Desktoparbeitsfläche zwingend vorkommen, entstehen.

Stattdessen wurde für die HEyeWall ein hochpräzises mechanisches Setup entwickelt, das durch paarweise Gruppierung nicht nur Vorjustage erlaubt und Ausfallzeiten minimiert, sondern das außerdem durch Metallblenden Bereiche außerhalb der darstellbaren Bildflächen (die bei digitalen Projektoren einen nicht zu vernachlässigenden Lichtanteil abgeben) abschattet und so pixelgenau die Kachelübergänge justierbar sind, wobei nebenbei noch optische Verzerrungen an den Projektionsrändern verdeckt werden. Durch zusätzliche Einstellmöglichkeiten wurden die in Kapitel 3.2.5 diskutierten Fokalebeneproblematiken kontrollier- und sogar justierbar.



Neben dem mechanischen Setup ist auch bei der HEyeWall die Leinwand der entscheidende Faktor. Hier jedoch sind die Anforderungen verglichen mit einer CAVE gänzlich verschieden: Interreflexionseffekte sind nicht relevant, da alle Projektionsflächen koplanar sind. Stattdessen gewinnt der Faktor *Diffusität* an Relevanz, ja entpuppt sich sogar als „Killerkriterium“: Ist die Leinwand nicht vollkommen diffus, entstehen blickpunktabhängige Highlights, die an jeder Kachelgrenze als deutlich sichtbare Helligkeitssprünge sichtbar werden würden (vgl. Abb. 16). Eine Leinwand mit nahezu optimaler Lambert-Charakteristik konnte in einer Variante der alten IGD-CAVE-Leinwand gefunden werden, der Gerriets Opera hellblau [G03]. Dieser Variante wurde aufgrund besserer Farbneutralität in der Transmission der Vorzug gegenüber der Opera milchig matt gegeben.

Durch volldiffuse Eigenschaften verbietet sich ein Stereoverfahren nach dem Polarisationsprinzip, womit auch hier das Infitec-Verfahren die einzig logische Alternative darstellte.

Die 24 Kacheln in einer 6x4 Matrix ergeben mit den 48 verwendeten Christie Vivid LX-41 LCD-Projektoren mit einer Auflösung von 1024x768 eine durch die Metallblenden leicht reduzierte Gesamtauflösung von 6104x3064 Pixeln.

Bei einer Projektionsfläche von 5x2.5m ergibt dies Pixelgrößen von 0.67 mm<sup>2</sup>. Hier erklärt sich auch die Namensgebung – denn bei einem Betrachterabstand von ca. 3 Metern (z.B. [G99]) übersteigt die HEyeWall das Auflösungsvermögen des menschlichen Auges.

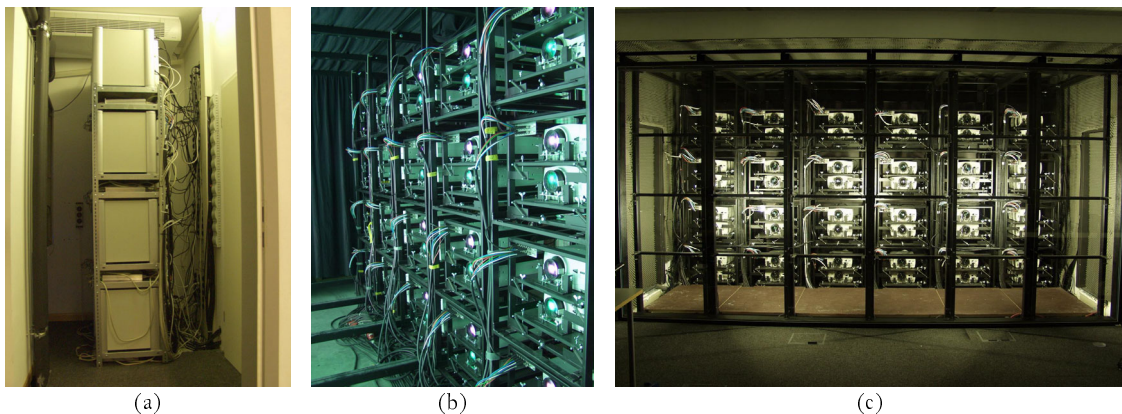


Abb. 24 : HEyeWall. (a) die 24 linken PCs. (b) Nahansicht der Projektormodule. (c) Blick von vorn, ohne Leinwand. Der Metallrahmen dient zur Abschattung von Streulicht auf Nachbarkacheln.

### 3.3.9 Ergebnisse

#### Infitec

Als Testszenario wurden zwei der (zu dem Zeitpunkt noch nicht in die CAVE eingebauten) Eiki LC-UXT1 mit Infitec-Filtern verwendet. Die Aufprojektions-Leinwand bestand aus weiß lackierten Holzplatten, wodurch sich die in Abb. 26 auffallenden, sehr starken spekularen Highlights erklären.

Tabelle 6 zeigt die gemessenen Farbwerte für die drei Primärvalenzen plus einiger Graustufen, den ermittelten gemeinsamen Farbraum, das vorgegebene Zielweiß (welches dem entsprechend helligkeitsreduzierten Weiß der Projektoren ohne Filter entspricht), sowie die Vergleichswerte nach durchgeführter Farbkorrektur. In Tabelle 5 sind zur besseren Fehlerabschätzung die gemittelten XYZ-Distanzen jeder Testfarbe zum gemeinsamen Gamut aufgelistet.

Farbe	Projektor 1		Projektor 2	
	unkalibriert	kalibriert	unkalibriert	kalibriert
R	1.018	0.692	91.802	0.320
G	131.297	1.366	6.341	0.641
B	54.627	1.073	113.168	1.143
W	179.462	1.906	175.645	1.480
50%	41.242	0.422	34.715	1.070
20%	1.538	0.114	1.650	0.094

Tabelle 5: Gemittelte XYZ-Absolutdifferenzen zwischen den Testprojektoren und dem gemeinsamen Gamut.

Farbe	Projektor 1			Projektor 2			gemeinsamer Farbgamut			Projektor 1 korrigiert			Projektor 2 korrigiert		
	L	x	y	L	x	y	L	x	y	L	x	y	L	x	y
R	13.41	.668	.328	118.72	.627	.371	16.26	.625	.370	16.50	.629	.359	16.20	.615	.365
G	327.51	.305	.685	67.82	.209	.668	57.45	.286	.613	54.80	.284	.610	58.30	.283	.610
B	11.31	.138	.044	17.02	.139	.039	4.14	.140	.045	4.64	.140	.049	4.05	.142	.046
W	351.51	.269	.458	202.72	.357	.268	77.86	.293	.340	76.00	.293	.329	77.70	.292	.346
50%	82.81	.276	.495	39.22	.357	.265	14.75	.293	.340	13.90	.293	.330	13.50	.292	.336
20%	4.95	.286	.527	2.18	.307	.223	1.63	.293	.340	1.47	.283	.320	1.63	.285	.360
0%	0.49	.283	.420	0.28	.268	.184									

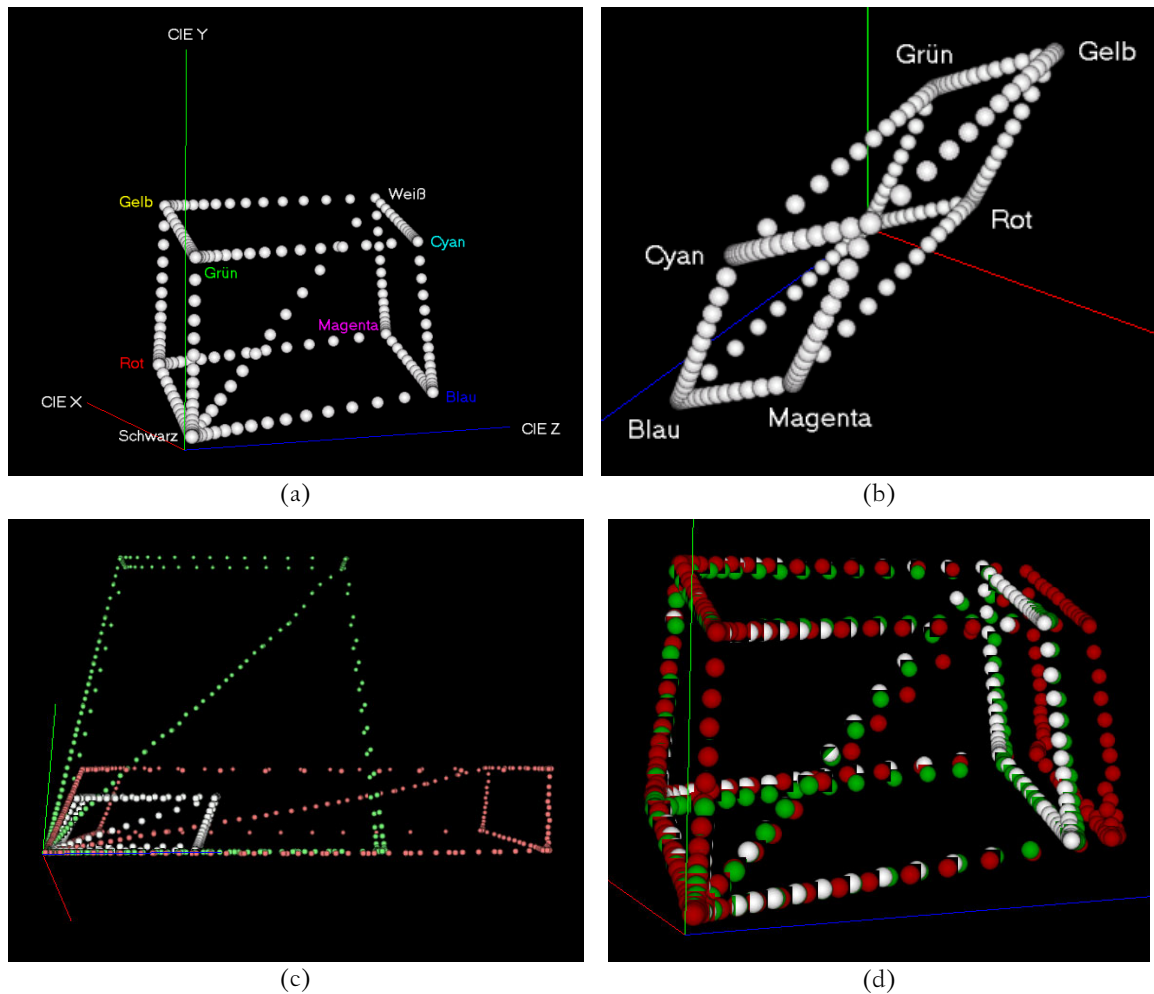
Tabelle 6: Gemessene Primärfarben und mehrere Graustufen für die unkalibrierten Projektoren, der gemeinsame Farbgamut, sowie Kontrollergebnisse nach der Farbkalibrierung.

In Abb. 25 sind Snapshots einer interaktiven Visualisierung der vermessenen Farbwürfel (genauer: CIE-XYZ-Farbspäte) zu sehen. Abb. 25a und b dienen hier als Legende: Es sind die Koordinatenachsen, sowie die Primär- und Sekundärfarben zu sehen, außerdem der Grauverlauf von Schwarz nach Weiß. Man beachte, daß sich das Schwarz aufgrund eines von 0 verschiedenen Schwarzlevels nicht im Ursprung befindet. Abb. 25b zeigt die Sicht entlang der Grauachse: die Farben sind nun entsprechend ihres Farbtons (*hue*) angeordnet.

In Abb. 25c sind die gemessenen Farbwerte der beiden Projektoren zu sehen, grün und rot eingefärbt entsprechend der Filterfarben. Man beachte die unre-

gelmäßig verlaufende Graulinie: durch sich unterscheidende Gammakurven pro Farbkanal ändert das Grau mit steigender Intensität ständig seinen  $xy$ -Farbort.

Der weiße Spat ist der gemeinsame Farbgamut bezüglich des vorgegebenen Weißpunktes. Im Prinzip besteht er aus der Schnittmenge der beiden Originalspäte, weiter reduziert um sich dem vorgegebenen Weißpunkt anzugleichen.



**Abb. 25:** Echtzeitvisualisierung des XYZ-Farbwürfels. (a) Schematische Darstellung mit Farbnamen als Orientierungshilfe. Man beachte, daß aufgrund des Schwarzlevels Schwarz nicht im Ursprung liegt. (b) Blick entlang der Grauchse. (c) zwei vermessene Infitecprojektoren plus der ermittelte gemeinsame Gamut. (d) Kontrollmessung nach dem Farbausgleich. Aufgrund eines Datenfehlers ist das Blau des roten Projektors zu intensiv.

Die Kontrollmessung nach dem Farbausgleich ist in Abb. 25d zu abgebildet (aufgrund eines Datenfehlers ist das Blau des roten Projektors etwas zu hell, weshalb die entsprechende Fläche des Spats verschoben ist). Man beachte, wie genau die Messwerte dem vorgegebenen Farbgamut entsprechen; insbesondere der Grauverlauf besteht nun aus einer geraden Linie. Insbesondere fällt auf, daß die Linien der drei Farbwürfel zueinander vollkommen parallel sind. Dies ist ein Indiz dafür, daß in der Tat der Farbraumabgleich erfolgreich war: Die Richtung einer vom Ursprung ausgehenden Gerade im XYZ-Raum entspricht einem konkreten

Punkt im  $xy$ -Farbdiagramm. Parallele Linien bedeuten gleichen Farbart, Linien unterschiedlicher Richtung (z.B. die drei Grünlinien in Abb. 25c, linker Bildrand) bedeuten unterschiedliche Farborte.



Abb. 26 : Vergleich zwischen unkorrigierten (links) und korrigierten (rechts) Bildpaaren. Deutliche Unterschiede in der rechten Spalte werden verursacht durch Abweichungen in der Spektralantwort der Kamera, sowie insbesondere durch Ausdrucken auf einem nicht kalibrierten Drucker.

Die Fotografien der Ergebnisse wurden mit einer Olympus E-20p Digitalkamera durchgeführt. Die visuellen und quantitativen Abweichungen der farbkalibrierten Bilder waren in der Realität sehr viel geringer, was die Vermutung nahe legt, daß die spektrale Empfindlichkeit der Kamera nicht sehr genau mit der des menschlichen Auges, bzw. des CIE-Standardbeobachters übereinstimmt und insbesondere Schwierigkeiten mit den extremen Spektralverteilungen aufgrund

der Infitec-Filter hat (man vergleiche z.B. die Grünwerte in der rechten Spalte des Farbverlaufs-Bildes in Abb. 26. Insbesondere durch das Ausdrucken verstärken sich diese nochmals extrem, da der Drucker nicht auf die Kameraeigenschaften kalibriert ist).

Durch den in den Bildern sichtbaren, in unkalibriertem „Weiß“ dargestellten Mauscursor bekommt man einen ungefähren Eindruck davon, wie die Farbkorrektur die Maximalhelligkeiten beeinflusst.

Obwohl ein Schwarzlevel aufgrund seiner vergleichsweise geringen Helligkeit zunächst unwichtig scheint, so ist seine Behandlung als *Farbwert*, sowie die Einbeziehung in den gesamten Kalibrierungsprozeß essentiell für eine konsistente und kalibrierte Farbwiedergabe bis hinunter zu sehr leuchtschwachen Pixeln.

Selbst für extreme Farbraumdifferenzen, wie z.B. aufgrund von speziellen Filtern, kann eine gute Farbkonsistenz erzielt werden. Standardprojektoren, mit üblicherweise sehr viel kleineren Unterschieden in ihren Gamuts, verhalten sich durch die Kalibrierung dementsprechend besser. Der limitierende Faktor ist hier meist die Empfindlichkeit des Meßgerätes, insbesondere was die Farbgenauigkeit bei geringen Leuchtdichten ( $<1\text{cd/m}^2$ ) anbelangt.

## HEyeWall

Das gleiche Verfahren wurde für die 48 HEyeWall-Projektoren eingesetzt. Zunächst wurden die Projektoren entsprechend Kapitel 3.3.1 vermessen und der gemeinsame Gamut mit den Maximalintensitäten bezüglich eines vorgegebenen Weißpunktes von 8500K (CIE-*xy*: 0.291, 0.300) bestimmt. Abb. 27 zeigt die entsprechend nutzbaren Bereiche der Primärvalenzen von allen Projektoren. Es ist erkennbar, daß das Rot von Projektor Nr. 26 die Gesamthelligkeit limitiert: 100% dieses Farbkanals müssen aktiviert werden, um bei allen Projektoren dasselbe Weiß mit gleicher Intensität darzustellen.

Es fällt auf, daß die Projektoren extrem starke Schwankungen nicht nur in ihren Helligkeiten aufgrund unterschiedlicher Lampenbetriebsstunden aufweisen (bei reinen Helligkeitsunterschieden müssten die Kurven für R, G und B bis auf einen Skalierungsfaktor identisch sein), sondern auch deutliche Unterschiede in den Einzelintensitäten besitzen. Die Schwankungen insgesamt sind beträchtlich: Von den Grünkanälen der Projektoren 0-23 kommen je nach Projektor zwischen 58% und 95% zum Einsatz.

Aufgrund der schmalbandigen Filter dürfen bei den „rechten“ Projektoren (Infitec-Filter für das rechte Auge, deutlicher Grünstich) vom Grünkanal nur ca. 10% der maximal darstellbaren Helligkeit genutzt werden. Da dies einen deutlichen Verlust der verfügbaren Farbauflösung bedeuten würde (von 0-255 auf ca. 0-100

bei unverändertem Projektorgamma), wurde beschlossen, die Maximalintensitäten bereits im Service-Menü der Projektoren zu beschränken (im normalen Benutzermenü sind solche Feineinstellungen selten möglich).

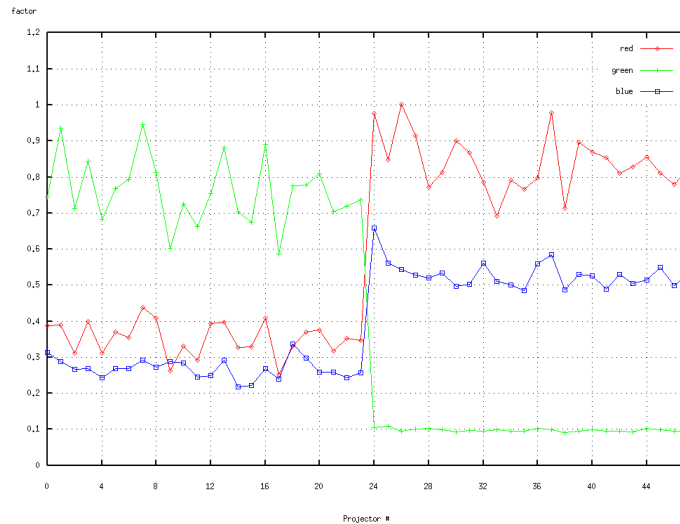


Abb. 27 : Ergebnis der maximal verwendbaren Intensitäten nach dem Ausgleich auf ein vorgegebenes Weiß. Projektoren 0–23 verwenden den „roten“ Infitec-Filter, Projektoren 24–47 den „grünen“.

Die Einstellung erfolgte pro Projektor derart, daß um das gemeinsame maximale Weiß darzustellen jeder seiner Farbkanäle mit ca. 80-90% Eingangssignal betrieben werden muß. Die 100% wurden deshalb nicht ausgeschöpft, um dem anschließenden Softwareausgleich bei mit der Zeit schwächer werdenden Lampen noch etwas Justagespielraum zu geben.

Gleichzeitig mit der Vorjustage wurde bei Projektor Nr. 26 aufgrund des schlechten Rotwertes die Lampe ausgetauscht. Nach dieser Vorjustage wurde nochmals gemessen, das Ergebnis ist in Abb. 28 zu sehen.

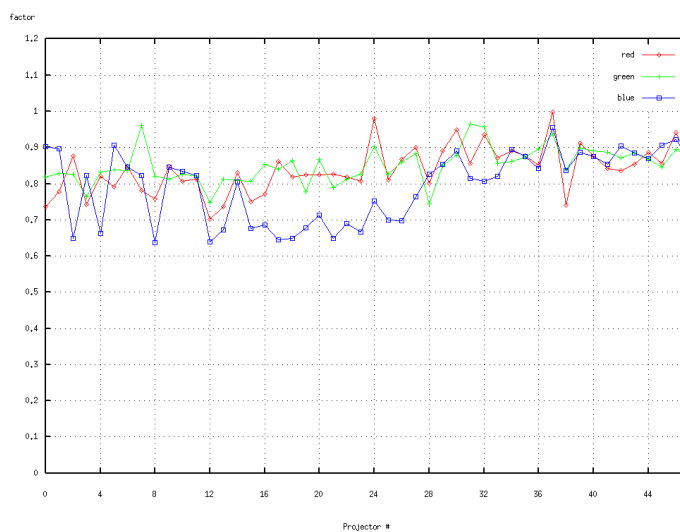
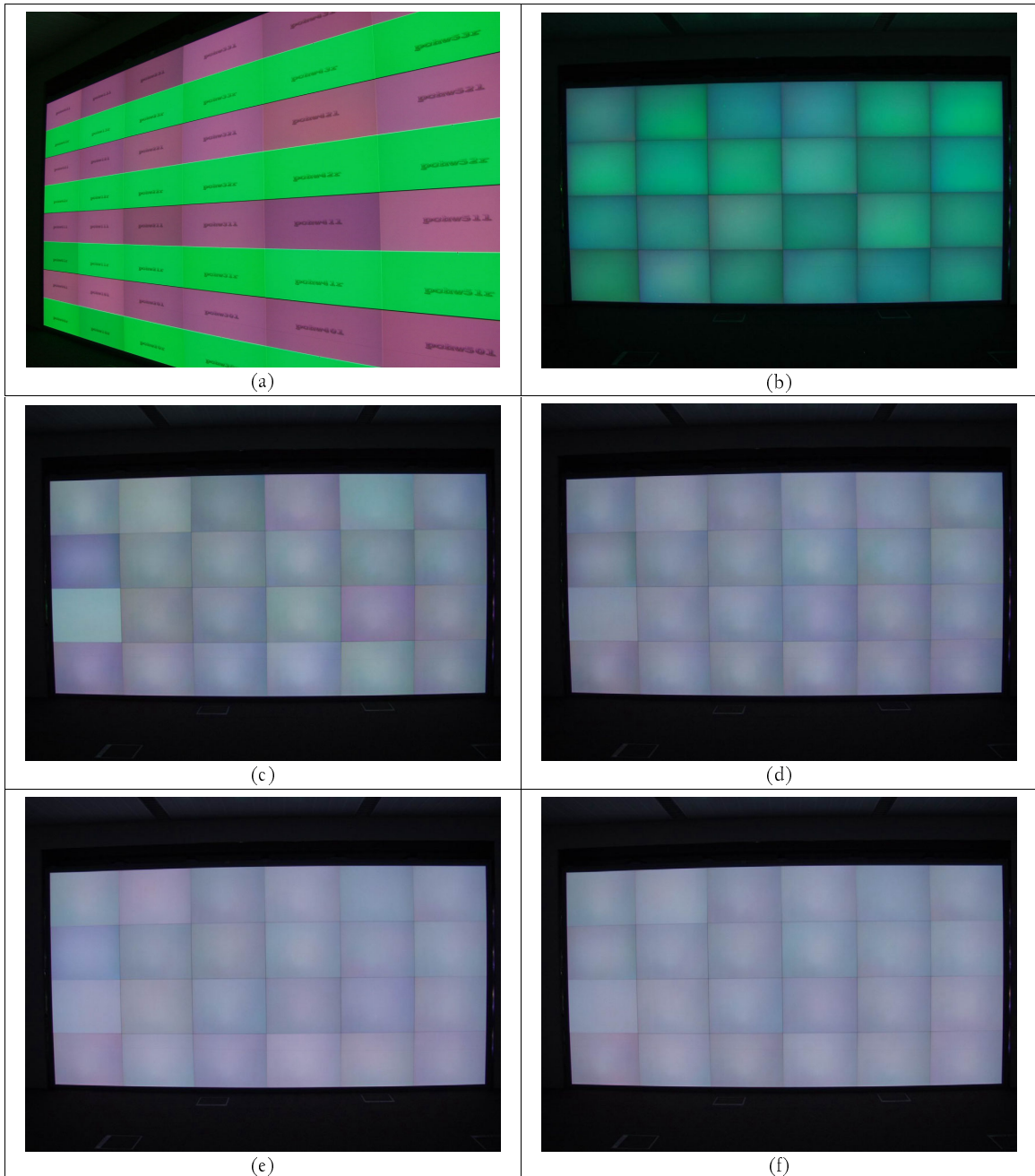


Abb. 28 : Durch Verringerung der darstellbaren Helligkeiten pro Farbkanal im Service-Menü befinden sich nun alle Kurven einheitlich bei ca. 80-90%.

Nun kann von allen Farbkanälen fast die gesamte Bittiefe genutzt werden. Da die Projektoren über DVI angeschlossen sind und sie intern mit 10 Bit rechnen, gewinnt man trotz der Helligkeitsreduzierung an Farbtiefe, wodurch Quantisierungseffekte bei der Darstellung vermieden werden.



**Abb. 29:** Farbkalibrierung HEyeWall, ohne Shadingausgleich. (a) reines Infitec-„Weiß“, ohne Vorabgleich der Projektoren. Die Bilder der zwei Projektoren pro Kachel wurden hier bei halber Höhe übereinander dargestellt. (b) Schwarz, unkalibriert (lange Belichtung). (c) 50% Grau, unkalibriert. (d) 50% Grau, farbkalibriert. Man beachte, daß in der linken Spalte beim zweiten Projektor von unten bereits ein projektorinterner Shadingabgleich vorgenommen wurde: Der Helligkeitsverlauf über die Fläche ist hier viel gleichmäßiger. (e) Weiß, unkalibriert. (f) Weiß, farbkalibriert. Man beachte daß zwischen Ausmessen und Aufnehmen der Fotos gut ein Monat lag, wodurch sich einige Lampen bereits leicht verändert hatten.

### 3.4 Automatische Kalibrierung von Displaysystemen

Projektions- und anderen Displaysysteme zur Visualisierung kalibrierter Daten können technologisch sehr vielfältig sein. Ein System zur Kalibrierung solcher Ausgabegeräte muß daher flexibel genug sein, um den unterschiedlichen Anforderungen solcher Systeme gerecht zu werden und das VR-System ohne große Anforderungen an den Benutzer an einzelne Displays anzupassen.

Heute in digitalen Projektoren verwendete Lampen verändern ihre Intensität – und teilweise auch ihr Spektrum – im Laufe ihrer Lebenszeit. Die gravierendsten Helligkeitsverluste treten hierbei üblicherweise während der ersten 100-200 Betriebsstunden auf. Eine Neukalibrierung in bestimmten Zeitintervallen ist somit notwendig.

Das Kalibrierungsverfahren an sich muß daher möglichst automatisch ablaufen, wodurch die Anwendbarkeit steigt und der Zeitaufwand zur Kalibrierung reduziert wird [KSM00].

Um eine Vielzahl an Messpunkten gleichzeitig aufnehmen und verarbeiten zu können, wird eine digitale Kamera als Meßgerät eingesetzt. Um dies zu ermöglichen, müssen zunächst deren interne Charakteristiken ermittelt werden (Kamerakalibrierung).

#### Vorangehende Arbeiten

Rykowski stellte eine Methode zur Verwendung von CCD Kameras für Quaitätsuntersuchungen von Beleuchtungssimulationssystemen vor [R94]. Der Vorteil dieser Methode liegt in der Geschwindigkeitssteigerung gegenüber konventionellen Meßmethoden mit Photodetektoren, da Messungen sehr vieler Positionen in einem Schritt durchgeführt werden können.

Um eine digitale Kamera für Messungen einsetzen zu können, ist jedoch eine Kenntnis der internen Farbverarbeitung dieser Kamera erforderlich. Üblicherweise handelt es sich bei solchen Informationen um proprietäre Daten, die von den Herstellern ungern herausgegeben werden [VFTB97]. Da das genaue Abbildungsverhalten von Farben und Intensitäten weiterhin für jede Kamera individuelle Abweichungen aufgrund von Herstellungsvarianzen und unterschiedlichen CCD-Empfindlichkeiten aufweisen kann und außerdem von Parametern wie Zoom oder Blende anhängig ist, müssen allgemeingültige Methoden verwendet werden, um diese kameraspezifischen Abbildungseigenschaften zu rekonstruieren [DM97].

Majumder et al. [MS02] stellten ein Verfahren vor, das auf Basis einer automatisch generierten „Luminance Attenuation Map“ einen Helligkeitsausgleich für



Multiprojektordisplays mit Hilfe einer digitalen Kamera berechnet. Das Verfahren gleich jedoch nur Helligkeiten aus, und auch dies nur für mittlere bis helle Bildbereiche, da das Projektorschwarz ignoriert wird. Insbesondere handelt es sich nur um einen Leuchtdichteausgleich, wodurch die photometrischen Displayeigenschaften verbessert werden, die colorimetrischen jedoch unverändert bleiben.

### 3.4.1 Vorgehensweise

Um eine Kamera als Meßgerät einzusetzen, müssen zunächst deren Charakteristiken ermittelt werden. Danach kann jedes zu verwendende Display kalibriert werden. Das VR-System verwendet schließlich die ermittelten Informationen zur und während der Darstellung. Dieser Ablauf ist in Abb. 30 dargestellt.

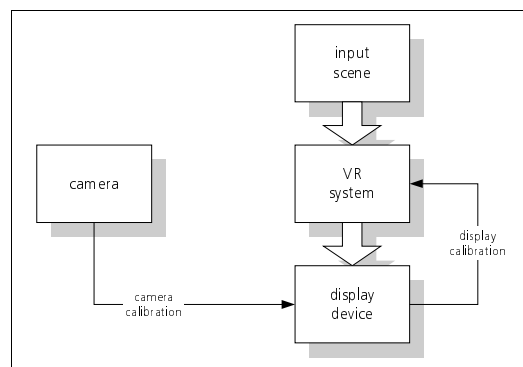


Abb. 30 : Schaubild des Kalibrierungsprozesses.

### 3.4.2 Kamerakalibrierung

Zur Displaykalibrierung wurde eine handelsübliche digitale Low-Cost-Kamera mit programmierbarem Software-Interface verwendet, eine Kodak DC265. CCD(charge coupled device arrays)-Chips an sich besitzen eine lineare Zuordnung zwischen eingehenden Beleuchtungsstärken und CCD-Ausgabeintensitäten, jedoch werden kameraintern üblicherweise nicht-lineare Mappings bei der Transformation in die RGB-Farben des resultierenden Bildes dazwischengeschaltet, um die Bildwiedergabe zu beeinflussen.

Daher muß zunächst die Kamerakennkurve bestimmt werden. Die Transferfunktion der Kamera muß bekannt sein, um aus den Pixelwerten Rückschlüsse auf die Leuchtdichten ziehen zu können. Diese Transferfunktion ist für jede Kamera eindeutig und fest, daher muß der Kalibrierungsschritt pro Kamera nur einmalig durchgeführt werden. Da die Kennkurve jedoch von vielen Parametern wie Blende, Zoom usw. abhängt, muß die Displaykalibrierung mit denselben Einstellungen vorgenommen werden, mit denen die Bestimmung der Kennkurve erfolgte.

## Kamerakennkurve

Mit Hilfe der von Debevec und Malik [DM97] beschriebenen Methode zur Erzeugung von *high dynamic range images* aus Fotos läßt sich die Kamerakennkurve  $f$  als Funktion der (logarithmischen) Belichtung  $X$  zu einem Pixelwert (0..255) aus einer Belichtungsreihe derselben Szene und unter denselben Bedingungen und Einstellungen der Kamera bestimmen. Die Belichtung  $X$  ist definiert als Produkt der Beleuchtungsstärke  $E$  und der Belichtungszeit  $\Delta t$ . Hierbei ist nur das Produkt  $E\Delta t$  relevant – eine Halbierung der Belichtungszeit bei gleichzeitiger Verdopplung der Beleuchtungsstärke läßt  $X$  – und damit den resultierenden Pixelwert – unverändert. Die Kennkurve läßt sich schließlich aus den unterschiedlichen Pixelwerten der gleichen Pixel in Bildern unterschiedlicher Belichtungszeit  $\Delta t$  mit Hilfe eines *least square* Optimierungsalgorithmus rückrechnen [DM97].

Zur Rekonstruktion wurde eine Serie von zehn Fotos eines Testbildes mit Hilfe der Scriptsteuerung der DC265 erstellt (Abb. 31). Das Testbild wurde hierbei so gewählt, daß die Fotos eine große Variation von unterschiedlichen RGB Pixelkombinationen enthielten, was den Vorteil hat, daß sich der Rekonstruktionsalgorithmus stabiler verhält – obwohl prinzipiell jede beliebige Szene mit nicht-konstanter Leuchtdichte zur Rekonstruktion verwendet werden kann.

Ist  $f$  schließlich bekannt, läßt sich die relative Beleuchtungsstärke  $E$  für jedes Pixel von beliebigen Bildern dieser Kamera berechnen, sofern die Belichtungszeit bekannt ist.

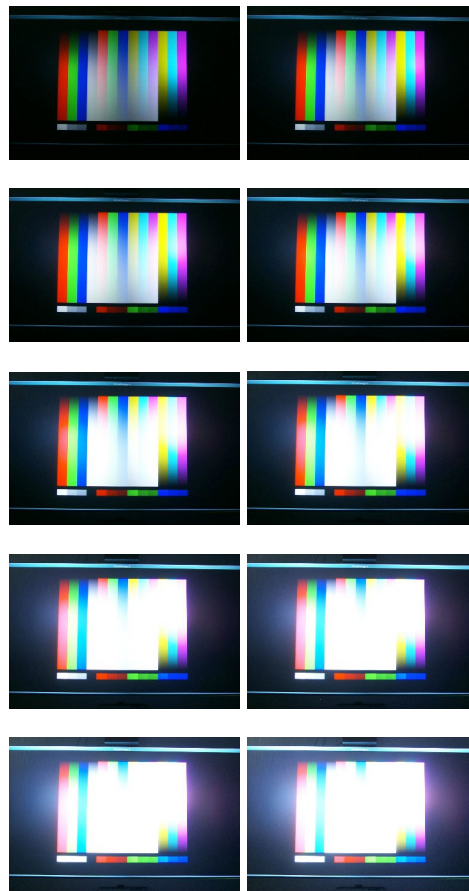


Abb. 31: Belichtungsserie zur Rekonstruktion der Kamerakennkurve  $f$ . Die Bilder wurden mit Belichtungen von 0.5s, 1.0s, 1.5s, 2.0s, 3.0s, 4.0s, 6.0s, 8.0s, 12.0s, bzw. 16.0s aufgenommen.

### Skalierungsfaktoren

Bei den so ermittelten Beleuchtungsstärken handelt es sich jedoch noch nicht um Absolutwerte. Um direkt die Leuchtdichte aus Pixelhelligkeit und Belichtungszeit bestimmen zu können, muß eine Lichtquelle mit bekannter Leuchtdichte als Referenzwert fotografiert werden, wobei es sich jedoch als problematisch erweisen kann, falls die spektrale Empfindlichkeit der Kamera nicht genau genug dem menschlichen Sehapparat vergleichbar ist. Pro Farbkanal wird also aus Leuchtdichte  $L$  und Pixelwert  $p$  ein Skalierungswert  $c$  ermittelt:

$$\ln(L \cdot \Delta t \cdot c) = f^{-1}(p), \text{ und damit}$$

$$c = \frac{e^{f^{-1}(p)}}{L \cdot \Delta t}$$

Während die inverse Kennkurve  $f^{-1}(p)$  zwar monoton ist, so ist es jedoch möglich, daß verschiedene Leuchtdichten auf gleiche Pixelwerte abgebildet werden, insbesondere natürlich bei über- und unterbelichteten Bereichen.

Diese Pixel können nicht zur Rekonstruktion photometrischer Information herangezogen werden und müssen daher entweder ignoriert, oder aus Fotos mit anderen Belichtungszeiten gewonnen werden.

Mit Hilfe der beschriebenen Methode wurden so für die Kodak DC265 folgende Werte ermittelt:

	Rot	Grün	Blau
Pixelwert $p$ ( $\Delta t = 1s$ )	140	111	192
$f^{-1}(p)$	1.838	1.733	1.899
gemessene Leuchtdichte $L$	$0.5 \text{ cd/m}^2$	$0.86 \text{ cd/m}^2$	$0.21 \text{ cd/m}^2$
Skalierung $c$	2.33	0.84	10.28

**Tabelle 7 :** Pixelwerte, gemessene Leuchtdichten und resultierende Skalierungsfaktoren für rot, grün und blau.

Ausgehend von auf Pixelwert 128 und  $\ln(L \cdot \Delta t) = 0$  normierten Kurven (Abb. 32a) ergeben sich durch die Skalierungsfaktoren seitlich zueinander verschobene Kurven (Abb. 32b), die in Abb. 32c nochmals nicht-logarithmisch aufgetragen sind.

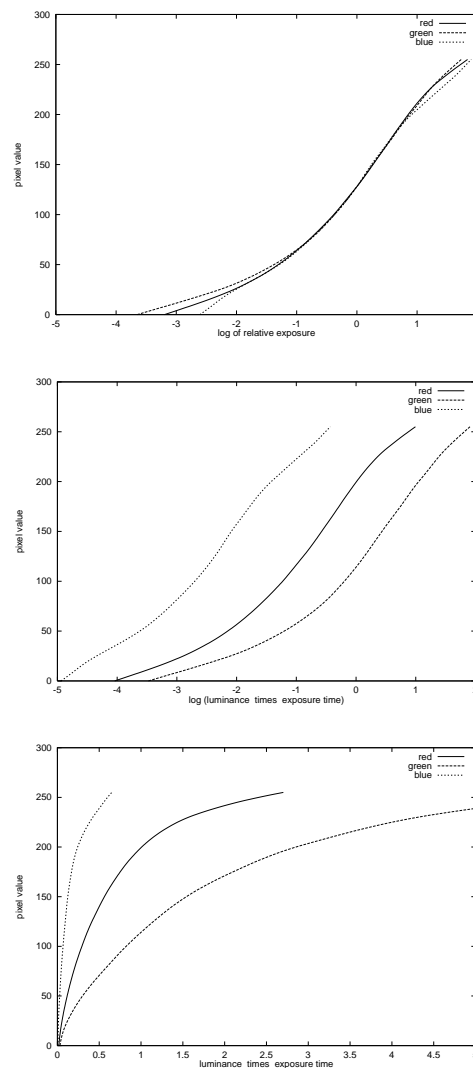


Abb. 32: Kamerakennkurven für rot, grün und blau, aus der Belichtungsserie rekonstruiert. Oben:  $\log(\text{relative Belichtung})$  zu Pixelwert, normiert auf  $f(0)=128$ . Mitte: Nach der Integration des Skalierungsfaktors. Unten: Leuchtdichte mal Belichtung zu Pixelwert.

## Gammawerte

Die so ermittelten Skalierungswerte liefern jedoch für sich betrachtet keine verlässlichen Werte, da z.B. ein „rotes“ Pixel auch einen bestimmten Anteil Grün und/oder Blau ungleich 0 besitzen wird, wodurch sich die gemessene Leuchtdichte entsprechend auf die drei Farbsensoren verteilt, und die tatsächliche „Rot“-Leuchtdichte geringer ausfällt. Dieses Problem lässt sich umgehen, wenn ein Weg gefunden wird, unabhängig von den Skalierungsfaktoren zu werden.

Und in der Tat wird die Berechnung dieser Faktoren überflüssig, wenn die relativen Kennkurven als „Kamera-Gammafunktionen“ betrachtet werden, um also Pixel-RGB von  $[0..255]$  in lineare RGB-Intensität  $[0..255]$  umzuwandeln. Die Absolut-Skalierung wird in diesem Fall statt dessen in die Farbmatrix der Kame-

ra mit aufgenommen (siehe folgender Abschnitt) und muß somit nicht mehr explizit errechnet werden. Es ist zu beachten, daß in diesem Fall die ermittelte Farbmatrix immer noch die Belichtungszeit  $\Delta t$  als Faktor enthält.

Des Weiteren darf diese Umrechnung nicht in Integer erfolgen, da aufgrund der nicht-linearen Funktion durchaus unterschiedliche Pixel niedriger Intensitäten auf dieselben Pixelwerte abgebildet werden können.

## Farbmatrix

Mit einer so vermessenen Kamera lassen sich nun also Pixelwerte in lineare Intensitäten umwandeln. Der Wertebereich ist jedoch auf 255 normiert. Um nun sowohl die tatsächliche Leuchtdichte, als auch die CIE  $xy$ -Farbe eines gegebenen Pixels ermitteln zu können, ist eine kameraspezifische Farbkalibrierungsmatrix erforderlich.

Zur Berechnung dieser Matrix müssen drei Farben mit bekannten XYZ-Farbkoordinaten fotografiert und mit Hilfe der Kennkurven in lineare Intensitäten umgewandelt werden. Durch diese Matrix werden Intensitäts-RGBs in kalibrierte (und mit  $\Delta t$  skalierte) XYZ-Farben umgewandelt.

Die berechnete Farbmatrix ist pro Kamera eindeutig, jedoch hängt die Qualität der Matrix von den verwendeten Kalibrierungsfarben ab.

Liegen die drei Farben nahe beieinander, bilden also im  $xy$ -Farbraum nur ein kleines Dreieck, vergrößert sich entsprechend der Einfluß von Meßfehlern. Handelt es sich bei den Farben jedoch um Extremwerte (z.B. bei Verwendung der Primärvalenzen eines Monitors, oder gar LEDs mit festgelegten Wellenlängen), so besteht die Gefahr, daß diese Farben außerhalb des Farbgamuts der Kamera liegen und von dieser gar nicht korrekt dargestellt werden können.

Eine Methode, solche Referenzfarben zur Verfügung zu stellen wird in Kapitel 4.2.4 vorgestellt.

## Aufnahme

Die kürzeste, vom Benutzer einstellbare Belichtungszeit der DC265 liegt bei 0.5 Sekunden. Durch diesen verhältnismäßig großen Wert werden alle Leuchtdichten jenseits von  $15 \text{ cd/m}^2$  überbelichtet. Auf die Blende der Kamera hat der Benutzer überhaupt keinen Einfluß. Da sich für unterschiedliche Blenden und Brennweiten jedoch die Kennkurven ändern, müssen die Kameraparameter, die in den digitalen Bildformaten mitgespeichert sind, dahingehend überprüft werden, ob die Kamera vielleicht von sich aus die Einstellungen geändert hat.

## Über-/Unterbelichtung

Vorsicht ist geboten, wenn die Fotos über- oder unterbelichtete Pixel (d.h. Pixelwerte nahe bei 255 bzw. 0) enthalten. Aus einem Pixelwert von 255 kann nicht geschlossen werden, ob die tatsächliche Leuchtdichte nahe bei diesem Wert oder sehr viel höher liegt. Um eine Sicherheitsschranke für CCD-Fluktuationen mit einzuschließen, sollten Pixelwerte nahe bei 255 (z.B. größer als 250) ignoriert werden.

Das gleiche gilt für Werte nahe bei 0. Hier macht sich das typische „CCD-Rauschen“ besonders stark bemerkbar: Unterbelichtete Pixel sind fast nie vollständig schwarz. Je nach Kamera, CCD-Qualität, interner Firmware oder auch Faktoren wie z.B. Umgebungstemperatur besitzen „schwarze“ Pixel im Allgemeinen RGB-Werte ungleich null.

Um sich hiergegen abzusichern, wurden für die Messungen Werte ignoriert, bei denen alle drei Farbkomponenten (vor der Kennkurventransformation) unter 15 lagen. Zusätzlich wird der Benutzer über den Anteil Über- und Unterbelichteter Pixel informiert, um so eine Rückmeldung über die Verlässlichkeit der vorliegenden Fotos zu erhalten.

Ein gewisser Anteil des CCD-Rauschens kann auch direkt ausgeglichen werden, da es statisch ist. Werden also vollständig schwarze Bilder aufgenommen (*clean plate image*), so sind oft immer dieselben Pixel um denselben Wert verfälscht, so dass sich diese Differenzen einfach aus einem vorliegenden Bild herausrechnen lassen.

## Optische Entzerrung

Qualitativ sehr hochwertige Kameraoptiken verhalten sich meist asphärisch, d.h. sie gleichen Verzerrungen im Bild durch ihr Linsensystem aus. Bei vielen Kameras wird das Bild jedoch radial verzerrt, sodass sich insbesondere in der Nähe der Bildränder Linien in Kurven verwandeln. Dieser Effekt verstärkt sich mit abnehmender Brennweite

Diese *radial distortion* muß zunächst durch entsprechende Bildtransformationen [DF01][P02] herausgerechnet werden, um eine eindeutige Zuordnung von aufgenommenen Pixeln zu Kamerapixeln zu ermöglichen.

## Vignettierung

Vergleichbar zu anderen optischen Systemen (siehe Kapitel 3.1.1, Abschnitt „Highlights“), ergibt sich aufgrund der unterschiedlichen Distanzen und Winkel zur Bildebene der Kamera eine Lichtabschwächung zu den Ecken hin. Da diese

Abschwächung signifikant sein kann, muß sie durch Aufnahme einer möglichst gleichförmig ausgeleuchteten Fläche (z.B. mehrere gemittelte Aufnahmen eines blauen Himmels aus unterschiedlichen Richtungen, Aufnahme eines Testbildes unter gleichzeitiger Kontrolle der Leuchtdichten einzelner Bereiche mit einem Leuchtdichtemeßgerät, oder unter kontrollierten Bedingungen mit Hilfe einer Ulbricht-Kugel [U20]) und Bestimmung der entsprechenden Leuchtdichteabschwächung pro Pixel ausgeglichen werden.

### Pixelwertmittelung

Einzelne CCD-Bildpunkte – insbesondere in dunklen Bildbereichen – sind üblicherweise nicht sehr repräsentativ und verlässlich bezüglich ihrer Helligkeit, aufgrund des (dynamischen) CCD-Rauschens, oder auch aufgrund von nicht abschaltbarer Bildkomprimierung, wie z.B. bei der DC265.

Daher sollte immer eine ganze Gruppe ähnlicher Pixel gemittelt werden, um verlässlichere photometrische und colorimetrische Informationen ableiten zu können.

### 3.4.3 Displaykalibrierung

Nach diesen – pro Kamera und für feste Blende und Zoom einmalig ausgeführten – Kalibrierungsschritten ist eine solche Kamera nun als schnelles, leicht zu verwendendes und vergleichsweise günstiges Farb-, Leuchtdichte- und auch Beleuchtungsmeßgerät mit akzeptabler Fehlertoleranz einsetzbar. Durch eine Feedback-Schleife läßt sich die Kamera nun zur automatischen Kalibrierung von Monitoren, Leinwänden oder auch CAVEs einsetzen.

Abb. 33 zeigt den hierzu verwendeten Aufbau: Die digitale Kamera wird an ein Laptop angeschlossen, das zum einen die Kamera ansteuert und die Bilder herunterlädt, zum anderen die Ergebnisse an eine über LAN angeschlossene SGI (oder eine andere OpenGL-fähige Workstation) überträgt, die die neuen Informationen wieder auf dem Display darstellt, das kalibriert werden soll.

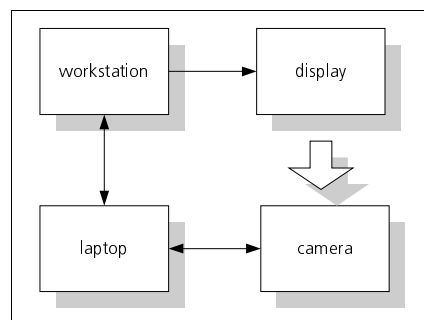


Abb. 33 : Schema des Kalibrierungsaufbaus.



Wie in Kapitel 3.4.1 diskutiert, müssen verschiedene Teilaspekte behandelt werden, um ein Display vollständig zu kalibrieren:

*Gleichmäßigkeit:* Leuchtdichte und Weißpunkt müssen an jeder Stelle des Displays gleich sein, damit gleiche Pixelwerte reproduzierbare Leuchtdichten liefern.

*Colorimetrie:* Simulierte Farbwerte müssen auf der Leinwand in der korrekten CIE-Farbe wiedergegeben werden.

*Photometrie:* Die dargestellten Leuchtdichtewerte müssen den simulierten Werten entsprechen.

### Gleichmäßigkeit

Die allermeisten Projektionsdisplays besitzen aufgrund physikalischer Gegebenheiten keine konstante Leuchtdichteverteilung über die Leinwand (vgl. Kapitel 3.1.1, Abschnitt „Highlights“). Der erste Schritt, um eine photometrisch und colorimetrisch konsistente Bildwiedergabe zu ermöglichen ist daher, diese konstante Verteilung – optimalerweise für jeden beliebigen Blickpunkt – sicherzustellen. Letzteres ist jedoch leider nur bei rein diffusen Leinwänden möglich, da sich die Lichtintensität ansonsten immer in Abhängigkeit vom Betrachterstandpunkt ändert. Aber selbst für Leinwände mit blickpunktabhängigen Highlights garantiert die vorgestellte Methode Gleichmäßigkeit zumindest für einen (z.B. den bevorzugten) Blickpunkt.

Die Grundidee liegt hierbei darin, eine transparente Textur zu berechnen, die Bildbereiche eines einheitlichen Bildes abdunkelt, welche im Verhältnis zu dunkleren Regionen (im Allgemeinen den Ecken) zu hell sind. Um zu vermeiden, daß in mühsamen, manuellen Meßreihen jeder Punkt der Leinwand mit einem Photometer gemessen und daraus eine invertierte Textur erzeugt werden muß, ist es wichtig, eine computerunterstützte, oder gar vollautomatisch arbeitende Lösung auf Basis einer selbststabilisierenden Feedback-Schleife zu finden, die ungewollte Highlights analysiert und eine transparente Textur erzeugt, um diese zu beseitigen.

Einmal berechnet, wird die invertierte Textur bei jedem gerenderten Frame statisch über den gesamten Bildbereich gelegt. Dadurch entstehen keine merklichen Performance-Verluste, es sei denn die Grafikkarte besitzt eine niedrige *pixel fill rate*.

Der offensichtliche Nachteil bei dieser Vorgehensweise ist sicherlich, daß Bildbereiche nicht heller gemacht werden können als vor dem Ausgleich. Um also eine vollständig gleichmäßige Ausleuchtung zu gewährleisten, muß jede Bildregion bis auf die niedrigste sichtbare Leuchtdichte bei Darstellung eines weißen

Bildes abgedunkelt werden. Für stark spiegelnde Leinwände bedeutet dies einen außerordentlich hohen Intensitätsverlust. In diesen Fällen sollte die auszugleichende Fläche eingeschränkt werden, um zumindest die stärksten Highlights zu beseitigen, jedoch werden in diesem Fall die Ecken merklich dunkler sein.

### *Pixelkorrelation*

Damit der Kalibrierungsalgorithmus feststellen kann, welche Pixel der Textur (und damit der Leinwand) mit welchen Kamerapixeln korrespondieren, muß zunächst ein Korrelationsschritt durchgeführt werden. Hierzu werden an definierten Stellen Farbmarkierungen projiziert, ein Foto aus der zu kalibrierenden Beobachterposition davon gemacht und die Markierungen in dem Foto unter Beachtung radialer Entzerrung durch Computervision-Methoden wiedergefunden. Dieser Vorgang läuft automatisiert ab.

Mindestens vier solcher Markierungen sind für eine eindeutige Rekonstruktion der Kameraposition erforderlich, mehr Punkte erhöhen jedoch die Genauigkeit.

Es ist sicherlich weder erforderlich, noch ratsam, eine Intensitätsverteilung pixelgenau für die volle Auflösung von 2400x1024 Pixeln bestimmen zu wollen. Die maximale Auflösung der DC265 beträgt 1536x1024 bei JPEG-komprimierten Bilddaten. Und insbesondere treten die Helligkeitsunterschiede über die Fläche nicht als individuelle Pixelfehler entsprechend dem CCD-Rauschen auf, sondern als niederfrequente Abweichungen im Helligkeits- und Farbraum. Der Fehler individueller Pixel muß somit durch die Mittelung mehrerer benachbarter Bildpunkte verringert werden.

Für das in Kapitel 3.2.1 beschriebene Setup einer Stereoleinwand war eine Texturauflösung von 64x32 Bildpunkten ausreichend. Niedrigere Texturauflösungen sind für einen Leuchtdichteausgleich unzureichend, und durch Gouraudshading können Mach-Bänder (vgl. Kapitel 2.2.1) entstehen.

### *Feedback-Schleife*

Der erste Schritt zur eigentlichen Texturerzeugung besteht darin, aus der nun kalibrierten Kameraposition ein vollweißes Bild aufzunehmen. Hier könnte der Algorithmus nun die Leinwanddimensionen erkennen und die niedrigste vorgefundene Pixelhelligkeit als zu adaptierenden Grenzwert annehmen. Es ist jedoch aus den diskutierten Gründen nutzbringender, den Benutzer einen Threshold-Wert wählen zu lassen, bis zu dem der Helligkeitsausgleich erfolgen soll – dunklere Pixel bleiben somit unverändert.

Nun wird die erste Iteration der Textur berechnet, an die Grafikworkstation übertragen und zur erneuten Aufnahme dargestellt. Dieser Prozeß wiederholt sich, bis eine gleichmäßige Lichtverteilung ermittelt ist.

Um Konvergenz zu garantieren und Oszillationen zu vermeiden, wird mit einem 50%-Graubild (8 Bit) begonnen. In der ersten Iteration wird ein Pixel um 25% Absolutelligkeit abgedunkelt oder erhellt. Das Intervall wird halbiert und die Prozedur beginnt von neuem, bis das niederwertigste Bit bearbeitet wurde.

Für passive Stereosetups muß dieser Prozeß für jedes Auge wiederholt werden, da durch die unterschiedliche geometrische Lage der Projektoren Highlights an unterschiedlichen Stellen entstehen können.

### *Erweiterung auf Farbe*

Für bestimmte Projektionen ist eine Graustufentextur unzureichend: Speziell bei CRT-Projektoren besteht die Möglichkeit, daß die Highlights einen farbigen Effekt verursachen, da die drei Primärfarbröhren räumlich getrennt sind. Optik oder interne Komponenten wie z.B. LCD-Panels können Farbabweichungen über die Fläche erzeugen. Die Leinwand kann wellenlängenabhängige Reflexionseigenschaften besitzen. Bei einem Multi-Projektorsystem können die einzelnen Projektoren unterschiedliche Farbeigenschaften und Weißpunkte aufweisen, oder verschiedene Bildflächen können aus unterschiedlichen Materialien bestehen (z.B. Rückprojektions-Seitenwand und Aufprojektions-Fußboden in einer Vier-Seiten-CAVE), usw.

Da für die Kamera jedoch eine Farbmatrix berechnet wurde, ist man nun in der Lage, für jedes Pixel (bzw. jeden Pixelflächenmittelwert) eine definierte CIE XYZ-Farbe zu bestimmen und so die RGB-Werte der Textur auf jede beliebige XYZ-Farbe und Leuchtdichte – sofern es Displayfarbgamut und maximale Projektorleuchtdichte zulassen – zu eichen.

Zusätzlich zu dem oben beschriebenen Threshold wird also die mittlere Gesamtfarbe des weißen Bildes ermittelt und nach CIE  $xy$  normalisiert. Dies ist der tatsächliche Weißpunkt des Displays. Die Farbinformation des gewählten Threshold-Wertes hingegen sollte jedoch ignoriert werden, da sie durch Farbverschiebungen des Displays verfälscht sein kann. Stattdessen liefert die Skalierung der gemittelten  $xy$ -Farbe mit der Threshold-Leuchtdichte die XYZ-Farbe, auf die jedes Pixel im zu kalibrierenden Bildbereich eingestellt wird.

Alternativ ist es ebenfalls denkbar, einen beliebigen anderen Weißpunkt (z.B. den Tageslicht-Normwert  $D_{65}$ ) vorzugeben und die Ausgleichstextur für eine Kalibrierung zu diesem Weißpunkt hin einzusetzen.

Da alle drei Farbkanäle zusammen erforderlich sind, um eine XYZ-Farbe zu definieren, muß die Berechnung der Ausgleichstextur nicht mehrfach durchgeführt werden (nämlich pro Farbkanal einmal), sondern kann einfach auf eine farbige Textur erweitert werden.

### *Gammakorrektur*

Da die ermittelte Textur Pixelwerte im Display-Gammaraum enthält, und nicht in einem linearen Gammaraum, muß sie durch Transformation mit den inversen gemessenen Gammakurven in den linearen Raum transformiert werden. Erst jetzt kann sie über die Szene gelegt werden.

### **Colorimetrie**

Der Farbgamut eines Displays hängt insbesondere von den verwendeten Röhren bzw. Farbfiltern ab und kann für die allermeisten Projektoren und Displays durch eine einfache RGB/XYZ-Konvertierungsmatrix  $M_{Display-RGB \rightarrow XYZ}$  beschrieben werden.

Diese Matrix kann ebenfalls mit Hilfe des beschriebenen Setups automatisch bestimmt werden: Durch Darstellung von maximalem Rot, Grün, bzw. Blau unter Anwendung der Ausgleichstextur können die XYZ-Werte der Primärvalenzen des Displays ermittelt und so die 3x3-Matrix aufgestellt werden.

Voraussetzung hierfür ist allerdings, daß der Farbraum der verwendeten Kamera größer als der des Displays ist, da sonst das Farbmeßgerät „Kamera“ fehlerhafte Werte liefert.

Man beachte daß selbst nach der Farbkalibrierung Farben außerhalb des Display-Gamuts (d.h. mit einer oder mehreren negativen RGB-Komponenten) per Definition weiterhin nicht darstellbar sind und geclippt werden müssen, wodurch an diesen Stellen natürlich eine colorimetrische Inkonsistenz entsteht. Das Renderingsystem „weiß“ jedoch, für welche Bildbereiche dies gilt, und kann so den Benutzer entsprechend (z.B. durch Fehlfarbdarstellung) darauf aufmerksam machen.

### **Photometrie**

Die Darstellungspipeline von Pixelwerten über Ansteuerströme bis zur dargestellten Leuchtdichte verwendet im Allgemeinen ein nicht-lineares Mapping, oft aufgrund von Nichtlinearitäten in der verwendeten Projektionshardware (CRT-Gamma) oder auch bewußt, um das Darstellungsverhalten nichtkalibrierter Systeme dem menschlichen Sehempfinden anzugleichen.

Ist das nichtlineare Verhalten des Ausgabegeräts bekannt, so kann durch entsprechend reziproke Anpassung der darzustellenden Eckpunktfarben ein Ausgleich erfolgen. Dieser Ausgleich ist erforderlich, um die kalibriert simulierten, in einem linearen Raum berechneten Farb- und Intensitätswerte auf einem solchen Display darzustellen, ohne die Konsistenz zu verlieren.

### *Gamma als Exponentialkurve*

Für den einfachsten Fall wird angenommen, daß die Nichtlinearität durch eine Exponentialfunktion folgender Form beschrieben werden kann (Gammakurve):

$$I_{out} = k \cdot I_{in}^{\gamma} + b,$$

wobei  $I_{in}$  die Eingangsintensität [0..1] entsprechend den RGB-Pixelwerten [0..255], sowie  $I_{out}$  die Ausgangsintensität repräsentiert.  $k$  ist ein pro Farbkanal unterschiedlicher Skalierungsfaktor, die zusammen den eingestellten Weißpunkt beschreiben, und insgesamt die Einstellung des „Kontrast“-Reglers wiedergibt.  $b$  ist der *blacklevel* des Displays, hängt stark von Projektortechnologie und Hintergrundhelligkeit ab, und wird durch den Regler mit dem etwas unglücklich gewählten Namen „Helligkeit“ eingestellt (zur Begriffsproblematik siehe z.B. auch [P97]).

Die korrekte Einstellung dieses *blacklevels* vor Beginn der Kalibrierung ist maßgeblich zur Substitution der Gammakurve durch einen einzelnen Gammaexponenten. Andernfalls kann photometrische Konsistenz für niedrige Leuchtdichten nicht sichergestellt werden: Ist der *blacklevel* zu niedrig, resultieren alle Eingangswerte unterhalb von  $I_{in_{min}}$  in  $I_{out} = 0$ . Ist er zu hoch, ist kein echtes „Schwarz“ darstellbar, und es gilt immer  $I_{out} > 0$ . Im optimalen Fall muß der *blacklevel* also so eingestellt werden, daß  $b = 0$  gilt.

Der gesuchte Exponent ist  $\gamma$ . Es wird angenommen, daß er für alle drei Farbkanäle gleich ist – wäre dem nicht so, würden dargestellte Graustufen wahrnehmbare Farbunterschiede aufweisen.

Analog zur Ausgleichstextur und Farbkalibrierung wird der Gammawert durch Aufnahme unterschiedlicher Grauwerte und Anwendung eines *best fit* Algorithmus bestimmt. Zur Kontrolle werden hier die drei Farbkanäle getrennt betrachtet, wobei die daraus resultierenden Gammakurven jedoch im Idealfall identisch sind.

In Abb. 34 ist das Verhältnis von Eingangsintensität zu Leuchtdichte dargestellt.

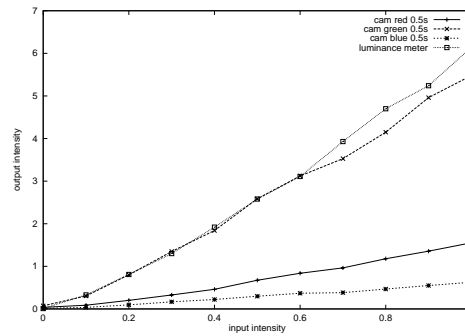


Abb. 34 : Eingangintensitäten zu Displayleuchtdichte. Dargestellt sind die über die digitale Kamera gemessenen Werte für Rot, Grün und Blau bei einer Belichtungszeit von 0,5s, sowie zum Vergleich die Ergebnisse eines Leuchtdichtemeßgerätes.

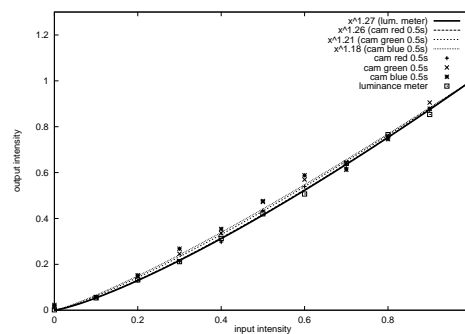


Abb. 35 : Eingangintensitäten zu normalisierter Displayleuchtdichte. Hier wird das nichtlineare Mapping in Form der ebenfalls dargestellten exponentiellen Gammakurven sichtbar. Zum Vergleich die Ergebnisse eines Leuchtdichtemeßgerätes.

Die normalisierten Graphen (Abb. 35) zeigen, daß die Nichtlinearitäten pro Farbkanal ähnlich sind. Die jeweils verwendeten Exponentialfunktionen sind in die Graphen integriert. Zum Vergleich wurde die Gammakurve ebenfalls mit einem dedizierten Leuchtdichtemeßgerät bestimmt. Diese kommt denen der automatischen Kalibrierung per digitaler Kamera sehr nahe.

		Gamma
Kameramessung $\Delta t = 0.5s$	Rot	1.26
	Grün	1.21
	Blau	1.18
Kameramessung $\Delta t = 1.0s$	Rot	1.29
	Grün	1.19
	Blau	1.14
Leuchtdichtemeßgerät		1.27

Tabelle 8 : Durch Kameramessung bzw. Leuchtdichtemeßgerät ermittelte Gammawerte. Die verhältnismäßig niedrigen Gammawerte erklären sich durch die Tatsache, daß die bildgebende SGI-Workstation einen voreingestellten Systemgamma von 1.7 besaß.

Durch Anwendung des ermittelten Gammawertes werden Eingangsintensitäten nun linear dargestellt, d.h. ein nochmals gemessener Gammawert sollte idealerweise 1.0 liefern.

Manche Systeme haben bereits einen soft- oder hardwareseitigen Gammawert voreingestellt. Der berechnete Gammawert  $\gamma_{corr}$  beschreibt einen Korrekturwert, der **zusätzlich** zu etwaigen Systemgammawerten  $\gamma_{system}$  angewandt werden muß. Mit  $\gamma_{total} = \gamma_{corr} \cdot \gamma_{system}$  erhält man einen Gammawert, der unabhängig von eingestellten (und veränderbaren) Systemgammawerten ist. Durch Speicherung dieses Wertes kann der je nach Systemgamma erforderliche Korrekturwert immer wieder ermittelt werden, auch wenn sich der Systemgamma geändert haben sollte. Jedoch muß in diesem Fall die transparente Ausgleichstextur mit  $\gamma_{adj} = \gamma_{system\_old} / \gamma_{system\_new}$  gewichtet werden, um für das neue Systemgamma gültig zu sein.

### *Gamma als Lookup-Table*

Für eine exaktere Berechnung – und auch um fehlerhafte *blacklevel* korrigieren zu können – müssen statt eines einzelnen Gammawertes je Farbkanal Gammafunktionen ausgemessen werden, die dann über eine *look-up table* (LUT) und lineare Interpolation auf die darzustellende Szene angewendet werden.

Der zusätzliche Vorteil dieser Vorgehensweise ist, daß hiermit auch Mappingkurven abgedeckt werden, die deutlich von einer Exponentialkurve abweichen, wie dies insbesondere bei fast allen digitalen Projektoren der Fall ist, deren Kurven üblicherweise S-förmig sind.

Die automatische Bestimmung dieser LUTs geschieht analog zu der Exponentialkurven-Variante, jedoch werden die vermessenen Werte direkt in die LUTs eingetragen, anstatt einen Exponenten daraus zu ermitteln.

### **In welcher Reihenfolge?**

Der erste Schritt besteht in der Aufnahme der Messwerte für Farbmatrix und Gammakurven. Nun kann die Ausgleichstextur bestimmt und mit den inversen Gammakurven transformiert werden. Da die transparente Farbtexur die Darstellungsfähigkeiten des Displays beeinflusst (insbesondere bezüglich der maximal darstellbaren Intensität), muß anschließend die Farbmatrix des Displays entsprechend den Textureigenschaften (Maximalhelligkeiten, voreingestellter Weißpunkt) angepasst werden.

### 3.4.4 Konsistente Darstellung

Nachdem die drei Displaycharakteristiken Ausgleichstextur, Farbraum und Gammakorrektur ermittelt wurden, muß dafür Sorge getragen werden, daß das Renderingsystem diese Informationen bei der Darstellung einer kalibriert berechneten Szene korrekt anwendet, um photometrisch verlässliche Werte auf der Leinwand zu erhalten. Um korrekte Farbwiedergabe auf unterschiedlichen Projektionsmedien (oder bei unterschiedlichen Einstellungen) zu ermöglichen, werden die pro Display ermittelten Werte für Textur, Farbmatrix und Gammawert zusammen mit einer eindeutigen Identifizierung des Displays szenenunabhängig in einer Datenbank gespeichert.

Im VR-System kann dann angegeben werden, für welches Display eine kalibrierte Ausgabe erfolgen soll, wobei die jeweils erforderlichen Korrekturwerte aus der Datenbank entnommen werden.

Dadurch kann, ohne VR-System-Konfiguration oder gar Szenenbeschreibung manuell anpassen zu müssen, schnell zwischen kalibrierter Darstellung auf unterschiedlichen Displays umgeschaltet werden. Wie die einzelnen Kalibrierungsinformationen hierbei im VR-System zur Anwendung kommen wird im Folgenden beschrieben.

#### Transparente Textur

Die berechnete Ausgleichstextur wird einfach als statisches, texturiertes Fullscreen-Polygon dargestellt. Für passive Stereotechnologien sind hierbei je Auge unterschiedliche Texturen zu verwenden.

Um den gewünschten Abschwächungseffekt und Farbkalibrierung zu ermöglichen, kann die übliche OpenGL *alpha blending function* zur Darstellung von transparenten Objekten nicht verwendet werden. Statt dessen ist eine Funktion erforderlich, die die RGB-Farben im Bildspeicher mit den RGB-Farben der Textur skaliert. Dies wird durch Verwendung der folgenden OpenGL-Kommandos bewerkstelligt:

```
glBlendFunc(GL_ZERO, GL_SRC_COLOR);  
glTexEnvf(GL_TEXTURE_ENV, GL_TEXTURE_ENV_MODE, GL_MODULATE);
```

Als Ergebnis wird, unabhängig von der dargestellten Szene, ein Punkt mit gegebenem RGB-Wert an jeder Stelle der Leinwand dieselbe (jedoch weiterhin unkalibrierte) Intensität und Farbe erhalten, wobei für Leinwände mit blickpunktabhängigen Eigenschaften dies nur für die Position, aus der die Ausgleichstextur berechnet wurde, zutrifft.



## Display-Farbkonvertierung

In der Szenendatei liegen die Ergebnisse der Lichtsimulation als standardisierte NTSC-RGB-Farben vor. Dabei werden explizit Werte außerhalb des Intervalls [0..1] zugelassen, sodass jede beliebige Farbe und Helligkeit beschrieben werden kann.

### *Vorab-Konvertierung*

Die naheliegendste Vorgehensweise wäre, die Konvertierung dieser Standardfarben in den Display-Farbraum beim Laden der Szene in das VR-System vorzunehmen:

Dies wird durch eine Umrechnung zunächst in den XYZ-Farbraum bewerkstelligt, wofür eine standardisierte Konvertierungsmatrix  $M_{NTSC-RGB \rightarrow XYZ}$  vorliegt. Sodann werden diese Farben in den RGB-Raum des Displays durch Anwendung der Inversen der oben ermittelten Konvertierungsmatrix  $M_{Display-RGB \rightarrow XYZ}^{-1}$  umgerechnet. Diese Matrizen können sodann durch eine einzige Matrix  $M_{NTSC-RGB \rightarrow Display-RGB}$  ersetzt werden.

Durch diese Matrix transformierte RGB-Werte können auf dem entsprechenden Display konsistent wiedergegeben werden, sofern jede Farbkomponente des Display-RGB-Wertes im Intervall [0..1] liegt.

Ein Wert größer eins bedeutet, daß das Display nicht in der Lage ist, die entsprechende Leuchtdichte (oder genauer: die vorgegebene Farbe mit der entsprechenden Leuchtdichte) darzustellen.

Sind ein oder zwei Werte negativ (drei negative Werte sind undefiniert und können nicht auftreten), liegt die Farbe außerhalb des darstellbaren Bereichs des Displays und können entsprechend Kapitel 3.3.7 behandelt werden.

Für alle darstellbaren Werte ergeben sich so nach der Umrechnung Eckpunktfarben, die sowohl für Farbraum, als auch für maximal darstellbare Leuchtdichten auf das jeweilige Display kalibriert sind. Solange jedoch der Systemgamma nicht dem gemessenen Gammawert entspricht, wird die Intensitätsantwort des Displays für Grauwerte falsch sein, und da sich die Gammakorrektur auf jeden Farbkanal auswirkt, gilt das Gleiche für die Farben.

### *On-the-Fly am Ende der Grafikpipeline*

Es ist jedoch nicht ausreichend, diese Werte in einem Vorprozeß als neue Eckpunktfarben zu berechnen. Insbesondere bei der Verwendung von Texturen, aber auch bei der zwar nicht photometrisch konsistenten, jedoch über mehrere Displays kalibrierten Ausgabe unkalibrierter Eingangswerte muß eine Methode ge-

funden, die entsprechenden Transformationen als letzte Operation vor der Darstellung, also insbesondere nach GL-Lighting und Texturierung, am Ende der Grafikpipeline durchzuführen.

Heute verfügbare PC-Grafikkarten besitzen entsprechende Möglichkeiten die Pixel-Pipeline zu programmieren und so zu beeinflussen.

Die erforderlichen Transformationen werden daher mit nVidia's RegisterCombinern realisiert. Hierbei werden drei generelle Combiner, sowie der abschließende Combiner verwendet, wodurch genug Ressourcen für die Nutzung weiterer Grafikkarteneffekte freibleiben. Es wird allerdings die `NV_register_combiners2` Extension aufgrund der Menge erforderlicher Konstanten benötigt, und erfordert somit mindestens eine GeForce3 oder besser.

Bei ATI-Grafikarten können die gleichen Transformationen mit Hilfe der ATI-spezifischen Erweiterungen oder den OpenGL 1.4 Fragment Programs durchgeführt werden.

Statt die Transformationen für jedes berechnete Pixel durchzuführen, kann dasselbe Resultat erzielt werden, indem das Bild normal in eine Textur gerendert wird, und diese Textur dann mit Hilfe der Fragment Programs bzw. Register Combiners bildschirmfüllend dargestellt wird. Welche Variante hier besser ist, hängt von der Effizienz des *direct-to-texture renderings* ab, die in älteren Grafikkartentreibern nur mäßig unterstützt wird.

### Gamma-Ausgleich

Die vermessenen, invertierten Gammakurven können dem Grafikkartentreiber zur Veränderung seines Darstellungsverhaltens übergeben werden. Dadurch werden Rechenzeit und Ressourcen der entsprechenden Grafikkartenfeatures freigegeben, da der komplette Part der Umwandlung vom linearen System in das Displayverhalten von diesem dedizierten Grafikkartenfeature übernommen werden kann.

Der Schwarzlevel – als additiver Term – kann hierzu in die Kurven mit integriert werden, wobei es jedoch von der verwendeten Methode abhängt, ob der Term einfach subtrahiert wird, oder der Differenzwert zu einem berechneten gemeinsamen Schwarz verwendet wird (vgl. Kapitel 3.3.6).

Der entscheidende Nachteil dieser Methode insgesamt ist jedoch, daß sich die im Framebuffer eingetragenen Pixelwerte immernoch in einem linearen Raum befinden, wodurch sich die pro Farbe verfügbaren Helligkeitsstufen auf 256 linear verteilte Schritte reduzieren. Aufgrund der logarithmischen Helligkeitsempfind-

lichkeit des menschlichen Auges sind durch diesen Quantisierungseffekt jedoch die unteren Stufen als deutliche Helligkeitssprünge sichtbar.

Auch eine Anpassung des Displaygammas auf 1.0 (sofern überhaupt möglich) ändert diese Situation nicht: die dunkelste Helligkeitsstufe neben Schwarz umfasst so immer noch  $1/256$  der Maximalhelligkeit des Displays – ein Wert, der viel zu hoch ist und bei einem normalen Display mit einem Gamma von 2.4 etwa ein Zweimillionstel beträgt – was vom menschlichen Auge bereits unterschieden werden kann.

Für eine gute visuelle Qualität auch in dunklen Bereichen ist daher die Kombination der Gammakorrektur mit der Farbmatrix erforderlich, damit die Transformation in den Displayraum noch in der Grafikkpipeline geschieht, wo Fließkommazahlen vorliegen, und nur das Ergebnis in dem 3x8 Bit Framebuffer quantisiert wird. Hierbei müssen auch die Schwarzanpassungen von der Grafikkpipeline übernommen werden.

### 3.4.5 Ergebnisse

Das vorgestellte Autokalibrierungsverfahren wurde auf das in Kapitel 3.2.1 beschriebene Setup der Breitbild-Stereoprojektion angewendet.

Für mehrere Beispielszenen wurde eine entsprechende Radiositysimulation durchgeführt, von denen eine ein virtuelles Modell des VR-Labs mit der CAVE am Fraunhofer-IGD darstellt (Abb. 36). Für einige Beispielpositionen sind hierbei die zu erwartenden Leuchtdichtewerte angegeben. In der Testszene wurden die eingesetzten Lichtquellen so weit gedimmt, daß die sich ergebenden Leuchtdichtewerte auf der gegebenen Leinwand darstellbar sind. Man beachte, daß der dargestellte Snapshot noch nicht auf den Farbraum des Displays kalibriert wurde.



Abb. 36: Das Ergebnis der Radiositysimulation.

Die Vermessung des Helligkeitsausgleichs resultierte in der in Abb. 37 dargestellten Ausgleichstextur – hier für das rechte Auge, für das die mittleren beiden

Projektoren zuständig waren. Die Textur für das linke Auge weicht entsprechend ab.

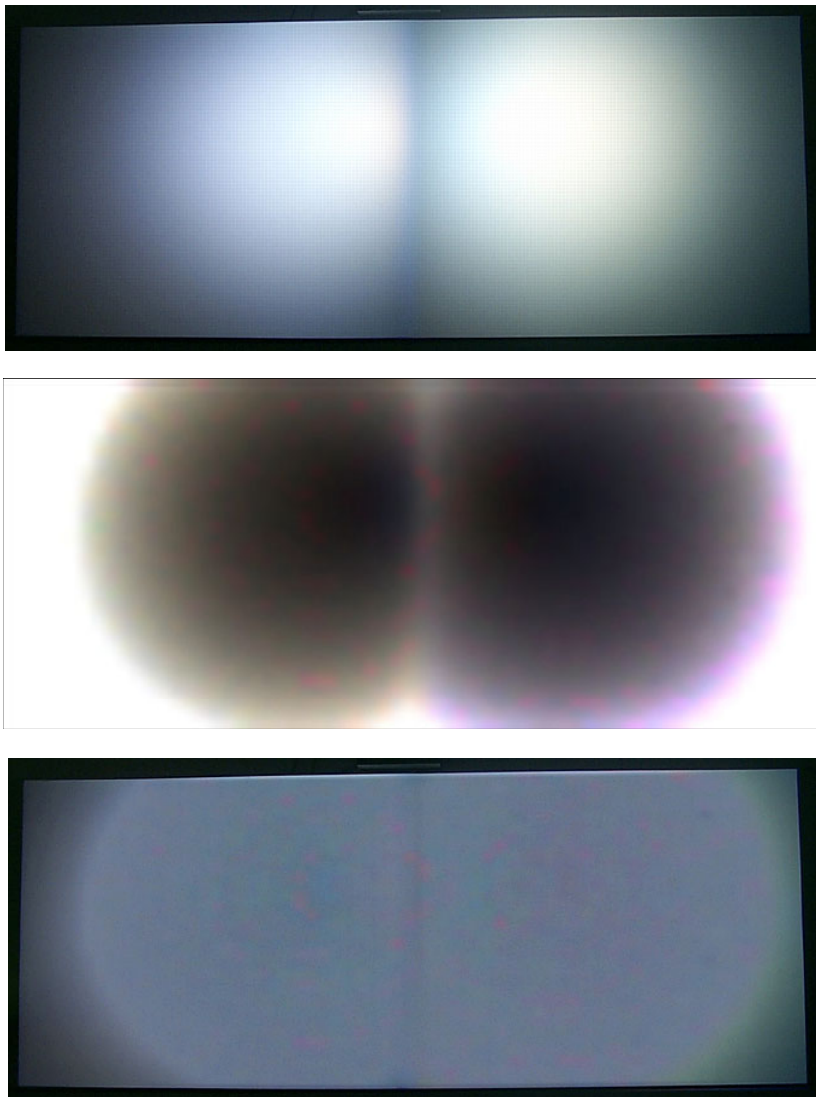


Abb. 37 : Die ausgehend von einem Foto eines weißen Bildes (oben) errechnete Ausgleichstextur für das rechte Auge (Mitte), sowie ihre Anwendung auf das weiße Bild (unten). Man beachte im oberen Bild die Farbränder blau(links) und rot(rechts).

Die Auswirkungen der Ausgleichstextur lassen sich direkt aus Abb. 37 ablesen: Highlights werden beseitigt, Farbverschiebungen verschwinden, und die Leuchtdichte ist auf weiten Bereichen der Leinwand gleichmäßig und farbig einheitlich. Jedoch wird die Gesamtleuchtdichte aufgrund der spiegelnden Eigenschaften der vorhandenen Leinwand stark reduziert. So wurde das Bild mit aktivierter Ausgleichstextur beispielsweise mit doppelter Belichtungszeit aufgenommen, um den Effekt deutlich zu machen.

Durch Anwendung der passenden Farbkonvertierungsmatrix und Gammakorrektur zusammen mit der Ausgleichstextur ergibt sich die in Abb. 38 dargestellte

Situation. Zum Vergleich ist bei diesen Snapshots die Szene einmal mit und einmal ohne Ausgleichstextur dargestellt. Es ist deutlich sichtbar, wie die Textur die zentralen Bereiche abdunkelt.

Abb. 39 zeigt nun schließlich Fotos der Leinwand während der Darstellung dieser beiden Szenarios. Man beachte insbesondere die Farbverfälschungen auf dem Tisch im Vordergrund, sowie das starke Highlight an der rechten Wand. Beide Effekte sind in dem Foto mit aktivierter Ausgleichstextur verschwunden. Auch für diese Fotos wurde die Belichtungszeit bei aktiver Textur verdoppelt.

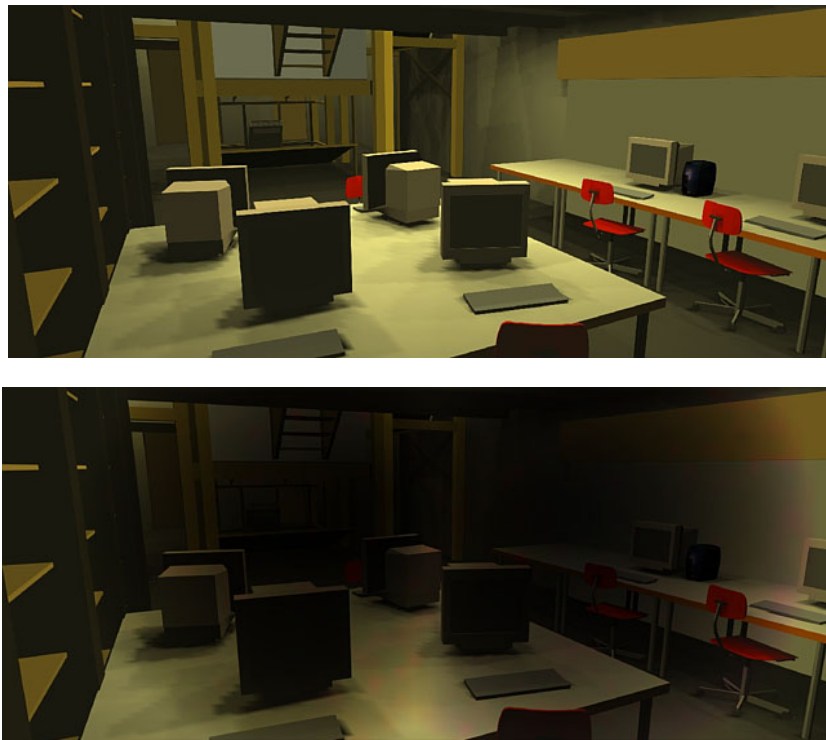


Abb. 38 : Snapshots der von dem VR-System dargestellten Szene, mit (unten) und ohne (oben) Ausgleichstextur. Der extreme Farbeindruck kommt dadurch zustande, daß die Bilder im Farbraum des Displays gerendert wurden.



Abb. 39 : Fotos der Leinwand mit projizierter, farbkorrigierter Szene, mit (unten) und ohne (oben) Ausgleichstextur.

Abschließend wurde die photometrische und colorimetrische Displayqualität durch Kontrollmessungen mit dem Minolta Colorimeter CS-100A für die in Abb. 36 dargestellten Bildpunkte überprüft.

Die Fehler in der Leuchtdichte lagen überall unter 10%, während Abweichungen in *xy chromaticity* meist bei  $\pm 0.008$ , stellenweise bei  $\pm 0.015$  lagen. Das Datenblatt des CS-100A – um ein Vielfaches teurer als die digitale Kamera DC265 – gibt einen Leuchtdichtefehler von 2%, und einen Farbfehler von  $\pm 0.004$  an. Es besagt jedoch auch, daß Farbmessungen für geringe Leuchtdichten nicht verlässlich sind, wodurch der Fehler des CS-100A für das Beispielsetup aufgrund der vorherrschenden Helligkeiten von ca.  $1-10 \text{ cd/m}^2$  vermutlich größer als  $\pm 0.004$  ist.

Es sei angemerkt, daß die Genauigkeit der digitalen Kamera als Farbmeßgerät nicht durch geringe Lichtintensitäten beeinträchtigt wird – es muß nur entsprechend die Belichtungszeit erhöht werden. Statt dessen ist der limitierende Faktor hier die Qualität und Genauigkeit des verwendeten CCD-Chips bei Langzeitaufnahmen, da hierbei das CCD-Rauschen an Intensität zunimmt.

Der Fehler kann weiter reduziert werden, wenn die Farbmatrix der Kamera genauer bestimmt wird: Durch die eigenartigen Belichtungszeiteinschränkungen von 0.5s liegt die maximal meßbare (Weiß-)Leuchtdichte der DC265 bei etwa  $15 \text{ cd/m}^2$ , was eine maximale Leuchtdichte von etwa  $2 \text{ cd/m}^2$  für hauptsächlich blaue Pixel bedeutet. Da das verwendete CS-100A jedoch bei diesen Helligkeiten

keine verlässlichen  $xy$ -Farben mehr liefern kann, waren die ermittelten XYZ-Werte der Kalibrierungsfarben möglicherweise nicht genau genug.





## 4 Integration einer AR-Bildsynthesekomponente

Augmented Reality ist auf dem besten Wege, sich als Standardmethode der Computer-Graphik zu etablieren, wie es seinerzeit bei der Virtual Reality der Fall war.

Die Idee dabei ist, die „echte“ Realität zu *augmentieren*, also mit virtuellen Objekten zu erweitern, oder reale Gegebenheiten zu verändern.

Informationen über die reale Szene können entweder in Form von Fotografien vorliegen, oder über einen Videostrom z.B. von einer mit einer Kamera versehenen See-Through-Datenbrille geliefert werden, auf der dann die virtuellen Objekte eingeblendet werden und sich so mit der Realität überlagern.

Ähnlich wie bei Virtual Reality konzentrierte man sich zunächst darauf, eine solche Erweiterung *überhaupt* zu erzeugen. Nachdem Augmented Reality nun langsam den Kinderschuhen entwachsen ist, beginnt man auch hier darüber nachzudenken, wie die visuelle Qualität – und damit der Realismus – von Augmented Reality Umgebungen gesteigert werden kann.

Dabei merkt man sehr schnell, daß ein wesentlicher Gesichtspunkt bei dem nahtlosen Einbinden von simulierten Objekten eine Kenntnis der Beleuchtungssituation der realen Szene ist. Neben der reinen Geometrie und verschiedenen Kameraparametern müssen also die realen Beleuchtungsverhältnisse (d.h. Lichtquellen) ebenso rekonstruiert werden, wie die Objektbeschaffenheiten wie Reflexionseigenschaften der Materialien und Texturen.

Werden diese Informationen in ein Radiositysystem eingebracht, lassen sich die virtuellen Objekte passend zu den in der Vorlage vorherrschenden Beleuchtungsverhältnissen simulieren und in die Szene einpassen.

Das Konzept der photometrischen Konsistenz läßt sich dabei auch auf Augmented Reality Szenarien ausdehnen – Voraussetzung dafür ist neben einem konsistent arbeitenden Simulationssystem, die photometrischen und colorimetrischen Informationen aus der realen Szene zu gewinnen [GCSV03].

### 4.1 Mängel existierender AR-Systeme

Bei klassischen Augmented Reality Systemen werden reale Aufnahmen – Video oder Einzelbilder – mit virtuellen Objekten erweitert. Hierzu müssen zunächst

Kameraposition, sowie Geometrie der realen Szene rekonstruiert werden. Diese kann z.B. unter Verwendung von mehreren, sich überschneidenden Fotos und Ausnutzung der Epipolargeometrie unter Verwendung z.B. der Fundamentalmatrix [LF96] geschehen. Andere Methoden können aufgrund bestimmter Annahmen, wie z.B. Orthogonalität von Raumbegrenzungen usw., diese Informationen auf Basis nur eines Fotos rekonstruieren [V98][GMMB00].

Aktuelle Methoden, sowohl Kameraparameter, als auch 3D-Szenengeometrie einfach und interaktiv aus einem einzigen Foto zu rekonstruieren finden sich z.B. in [BS03].

Die Lichtverhältnisse der realen Szene werden jedoch, wenn überhaupt, nur grob abgeschätzt, und gehen daher kaum in die Visualisierung mit ein. Speziell werden die virtuellen Objekte oft mit einfachem Hardwarerendering in die Szene eingeblendet.

In jüngerer Zeit wurde versucht, aus den realen Fotos die vorherrschenden Beleuchtungsverhältnisse zu rekonstruieren [FGR93][DRB97]. Nur wenige Verfahren können hierbei jedoch indirekte Lichteinflüsse korrekt behandeln [LFDW99]. Auf Basis der rekonstruierten Geometrie- und Beleuchtungsinformationen der Realszene werden die virtuellen Objekte mit Hilfe eines Radiosityverfahrens entsprechend der realen Beleuchtung simuliert.

Um photometrisch konsistent Fotos mit Augmented Reality und Radiosity erweitern zu können, ist insbesondere die photometrische und colorimetrische Rekonstruktion der Originalszene erforderlich. Dazu muß zunächst die Kamera-kennkurve der Kamera mit der das Bild aufgenommen wurde rekonstruiert werden. Dies geschieht auf Basis von *high dynamic range (HDR) images* [DM97] (vgl. Kapitel 3.4.2). Sodann kann eine Beziehung zwischen Belichtung, Pixelhelligkeit und vorhandener Leuchtdichte hergestellt werden.

Zur Rekonstruktion von Materialeigenschaften müssen die Reflexionskoeffizienten der realen Objekte ermittelt werden. Dies geschieht oft einfach auf Basis von Heuristiken [FGR93][DRB97] unter Zuhilfenahme eines ambienten Terms [CCWG88]. In [LFDW99] wird die Szene mit einer bekannten Lichtquelle beleuchtet, was zu einer exakten Bestimmung der direkten Beleuchtungsstärke pro Pixel führt. Die indirekte Beleuchtungsstärke wird allerdings ebenfalls durch einen ambienten Term abgeschätzt, was Ungenauigkeiten verursacht. Zur Materialienrekonstruktion gibt es Ansätze z.B. auf Basis mehrere HDR-Bilder [YDMH99] oder unter Voraussetzung bekannter Lichtquellen [BG01].

Literatur zur direkten Rekonstruktion von Lichtquellen und ihrer Eigenschaften ist bisher kaum vorhanden. Meist werden Position und Helligkeit als bekannt

vorausgesetzt, oder es wird versucht, die Beleuchtungsverhältnisse durch mehrere virtuelle Lichtquellen in einem Optimierungsverfahren anzunähern [GM00].

Eine gute Methode, die zwar nicht die Position und Abstrahlcharakteristik rekonstruieren kann, stattdessen jedoch für einen Bezugspunkt in der Szene die dort ankommende Beleuchtung aus jeder Raumrichtung, verwendet eine sogenannte *light probe* – eine verspiegelte Kugel, die sich im Foto befindet [D98].

Um die erweiterte Szene interaktiv verändern zu können, sind dynamische Radiosityverfahren, z.B. auf Basis der *line space hierarchy* [DS97][S01], unabdingbar.

## 4.2 Rekonstruktion der Licht- und Materialeigenschaften

Liegen außer dem Ausgangsfoto keine Informationen über Materialeigenschaften oder Eigenschaften der Lichtquellen vor, so ist deren Rekonstruktion ein schwieriges, teilweise sogar unlösbares Problem. Kapitel 4.2.3 beschreibt die Problematik der Mehrdeutigkeiten, die hierbei entsteht. Einfacher ist die Rekonstruktion einer Eigenschaft, wenn die andere bekannt ist; hierfür werden in diesem Abschnitt Verfahren vorgestellt (vgl. [G01]).

### 4.2.1 Reflexionskoeffizient

Die Radiosity  $B_E$  eines Patches  $E$  berechnet sich immer als Produkt von Beleuchtungsstärke  $E_E$  und Reflexionskoeffizient  $\rho_E$ :

$$B_E = \rho_E \cdot E_E,$$

$$E_E = E_{dir} + E_{ind} = \sum_{i=1}^L \frac{I_i(\theta, \varphi) \cdot \cos(\theta_{Ei})}{d_i^2} + \sum_{S=1}^n B_S F_{ES}$$

Die Beleuchtungsstärke ist in einen direkten und einen indirekten Teil aufgeteilt. Der direkte Anteil beschreibt die  $L$  Lichtquellen, der indirekte Anteil die Beleuchtungsstärke der  $n$  Patches der gesamten Szene.

Ist die Lichtstärke  $I_i(\theta, \varphi)$  der Lampen bekannt, so kann nach  $\rho_E$  aufgelöst werden:

$$\rho_E = \frac{B_E}{\sum_{i=1}^L \frac{I_i(\theta, \varphi) \cdot \cos(\theta_{Ei})}{d_i^2} + \sum_{S=1}^n B_S F_{ES}}$$

Betrachtet man für ein Polygon, bestehend aus  $m$  Patches, den Reflexionskoeffizienten als konstant, so kann man mit folgender Fehlerminimierung einen Reflexionskoeffizienten für das Polygon bestimmen:

$$D = \sum_{j=1}^m \Delta B_{E_j}^2 = \sum_{j=1}^m (B_{E_j} - \rho_E E_{E_j})^2$$

Zur Minimierung der Abweichung  $D$  wird die Nullstelle der ersten Ableitung gesucht, also:

$$\frac{\partial D}{\partial \rho_E} = \sum_{j=1}^m -2 \cdot (B_{E_j} - \rho_E E_{E_j}) \cdot E_{E_j} = -2 \cdot \left( \sum_{j=1}^m B_{E_j} E_{E_j} - \sum_{j=1}^m E_{E_j}^2 \right) = 0$$

Daraus folgt für den Reflexionskoeffizienten des Polygons:

$$\rho_E = \frac{\sum_{j=1}^m B_{E_j} E_{E_j}}{\sum_{j=1}^m E_{E_j}^2}$$

### 4.2.2 Lichtposition

Ist die Lichtquelle auf den Fotos nicht sichtbar, so muß sie aus Merkmalen des Fotos rekonstruiert werden. Für die folgenden Überlegungen wird davon ausgegangen, daß nur ein punktförmiger Emittter die Szene beleuchtet und ein 3D Modell der Szene vorliegt.

Das einfachste Merkmal zur Rekonstruktion der Lichtposition sind die auf dem Foto sichtbaren Schatten. Ist beispielsweise ein Eckpunkt eines Schattens und der zugehörige verursachende Eckpunkt der Geometrie sichtbar, so wird durch diese beiden Punkte eine Gerade definiert, die durch die Lichtquelle verläuft. Sind mehrere solcher Schattengeraden erkennbar, so beschreibt deren Schnittpunkt die Position der Lichtquelle.

Sind die Eckpunkte des Schatten bzw. die der verursachenden Geometrie auf dem Foto nicht sichtbar, sondern nur die Schattenkante und die verursachende Kante der Geometrie, so ist durch diese beiden Kanten eine Schattenebene definiert, welche die Lichtquelle enthält. Für mehrere solcher Schattenebenen ist der gemeinsame Schnittpunkt die Position der Lichtquelle.

In der Praxis gibt es allerdings Probleme bei der Anwendung dieses Verfahrens:

- Lichtquellen sind nie punktförmig, sondern besitzen eine Ausdehnung was zu unscharfen Schattenkanten führt.

- Auch bei relativ scharfen Schattenkanten entstehen durch die begrenzte Auflösung des Bildes Ungenauigkeiten

Diese Ungenauigkeiten haben zur Folge, daß sich die Schattengeraden (und Ebenen) nicht in einem Punkt schneiden, sondern in der Nähe der Lichtposition aneinander vorbeilaufen. Als Lichtposition wird dann der Punkt berechnet, für den die Summe seiner quadratischen Abstände zu allen Geraden und Ebenen minimal wird (z.B. [G01]).

### 4.2.3 Abstrahlcharakteristik

Ist die Position der Lichtquelle bekannt, so stellt sich als nächstes die Frage, wie hell die Lampe in verschiedene Richtungen strahlt und welche Lichtfarbe sie besitzt.

Die korrekte Einstellung dieser Werte ist für die Erweiterung von Fotos durch virtuelle Objekte extrem wichtig, da durch eine falsche Helligkeit der Lampe oder eine nicht passende Lichtfarbe die photometrische Konsistenz verlorengeht.

Ist die Lampe auf einem der Fotos sichtbar, so hat man bei vorhandenen Kamerakurven bereits einen Anhaltspunkt für deren Leuchtdichte. Voraussetzung ist allerdings eine sehr kurze Belichtungszeit, die bei vielen Digitalkameras nicht möglich ist.

Die Leuchtdichten innerhalb der Szene (Wände oder Boden) sind dagegen um Größenordnungen kleiner und im Normalfall kann aus deren Pixelfarbe  $(R, G, B)$  die Leuchtdichte  $(L_R, L_G, L_B)$  mit Hilfe der invertierten Kamerakurve [DM97] bestimmt werden.

Mit der bekannten Leuchtdichte eines Patches läßt sich allerdings noch nicht auf die Lichtstärke einer unbekanntenen Lampe schließen, wie folgendes einfaches Beispiel zeigt:

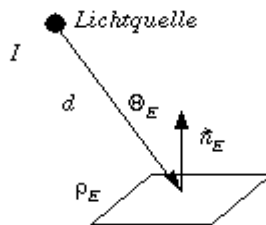


Abb. 40 : Zur Berechnung der Lichtstärke.

Ein Patch  $E$  befindet sich im Abstand  $d$  von der Lampe mit unbekannter Lichtstärke  $I$ .

Die Radiosity auf dem Patch beträgt (vgl. 2.4.2):

$$B_E = \rho \cdot \frac{I \cos \theta_e}{d^2} \cdot \Omega_0,$$

wobei der Einheitsraumwinkel  $\Omega_0$  zur Korrektur der Einheiten dient [SC00].

Da sowohl  $\rho_E$  als auch  $I$  unbekannt sind, läßt sich die Lichtstärke nicht eindeutig bestimmen: Verdoppelt sich die Lichtstärke und halbiert sich die Reflektivität, so entsteht wieder die gleiche Patch-Radiosity:

$$B_E = \frac{1}{2} \cdot \rho \cdot \frac{2I \cos \theta_e}{d^2} \cdot \Omega_0$$

Für eine gegebene Radiosity-Simulation kann das folgendes bedeuten:

Jedes Patch liefert eine Gleichung und zwei Unbekannte, was dazu führt, daß die Bestimmung von  $I$  und  $\rho$  für jedes Patch mehrdeutig ist, d.h. es gibt unendlich viele Lösungen für das Radiosity-Gleichungssystem.

### Nur eine Lichtquelle vorhanden

Für die Beleuchtung mit einer Lichtquelle lautet die Radiosity-Gleichung:

$$B_E = \rho_E \cdot \left( \frac{I(\theta, \varphi) \cdot \cos(\theta_E)}{d^2} \cdot \Omega_0 + \sum_{s=1}^n B_S F_{ES} \right),$$

wobei der Summenterm die indirekte Beleuchtungsstärke darstellt.

Ist der Reflexionskoeffizient bekannt, so kann man nach der unbekanntem Lichtstärke auflösen:

$$I(\theta, \varphi) = \left( \frac{B_E}{\rho_E} - \sum_{s=1}^n B_S F_{ES} \right) \cdot \frac{d^2}{\cos(\theta_E) \cdot \Omega_0},$$

womit die Lichtstärke in Richtung aller Patches ermittelt werden kann, die auf den Fotos sichtbar sind und nicht durch dritte Objekte im Schatten der Lichtquelle liegen oder von der Lichtquelle abgewandt sind.

### Vorgabe eines Reflexionskoeffizienten

Normalerweise sind weder die Lichtstärke der Lampe noch die Reflexionskoeffizienten bekannt. Der im folgenden beschriebene Algorithmus bestimmt die LVK der Lampe und die Reflexionskoeffizienten aller sichtbaren Polygone mit der Vorgabe eines Reflexionskoeffizienten durch den Benutzer.

Gibt der Benutzer einen Reflexionskoeffizienten  $\rho_{Ref}$  vor, so können für alle Patches mit diesem Material die Lichtstärken mit obiger Gleichung bestimmt werden. Da sich die Lichtfarbe der Lichtquelle im Normalfall für verschiedene Richtungen nicht ändert, wird für jedes Patch ein Skalierungsfaktor, bezogen auf eine Referenzlichtstärke, berechnet.

$$I_n^C = k_n \cdot I_{ref}^C$$

mit

$$C \in \{R, G, B\}$$

$I_{ref}$  wird dabei am Anfang für die drei Farbkomponenten für ein Patch mit Material  $\rho_{ref}$  berechnet. Alle anderen rekonstruierten Lichtstärken beziehen sich auf  $I_{ref}$ . Das heißt, für jedes Patch muß durch die bekannte Lichtfarbe nur der Skalierungsfaktor  $k$  berechnet werden, anstatt von drei Werten für die Lichtstärke. Sind alle Lichtstärken für Patches mit Material  $\rho_{ref}$  berechnet, so wird als nächstes der kleinste Winkel  $\alpha_{min}$  zwischen der Richtung einer bereits bestimmten Lichtstärke  $I_1$  und der Richtung einer noch nicht bestimmten Lichtstärke  $I_2$  berechnet. Unter der Annahme, daß sich die Lichtstärke mit dem Winkel nur langsam ändert, wird als Näherung  $I_2 = I_1$  gesetzt. Mit der vorhandenen Lichtstärke wird der noch unbekannte Reflexionskoeffizient  $\rho$  entsprechend Kapitel 4.2.1 bestimmt und damit alle Lichtstärken für Patches mit Material  $\rho$ . Diese Vorgehensweise wird wiederholt, bis für alle von der Lichtquelle aus sichtbaren Patches die Lichtstärke und der Reflexionskoeffizient bestimmt wurden.

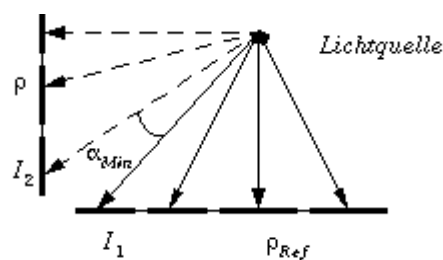


Abb. 41: Automatische Ermittlung von Lichtstärkeverteilungskurven,

Zur Optimierung der Werte  $k_1 \dots k_n$  und  $\rho^{\{R,G,B\}}$  wird der Levenberg-Marquadt-Algorithmus [PTVF92] angewendet, um mögliche Fehler durch die Näherung auszugleichen. Aus den Mittelwerten der das Patch  $i$  überdeckenden Pixel läßt sich die Radiosity  $B_i^{\{R,G,B\}}$  des Patches bestimmen. Andererseits berechnet sich

die Radiosity  $B_i^C$ , von Patch  $i$  mit den Näherungswerten für  $k_1 \dots k_n$  und  $\rho^{\{R,G,B\}}$  als

$$B_i^C = \rho^C \cdot \left( \frac{k_i \cdot I_{\text{Ref}}^C \cdot \cos(\theta_i)}{d_i^2} + \sum_{j=1}^n B_j F_{ij} \right)$$

Der Unterschied  $\Delta B_i = B_i - B_i^C$  sollte minimal werden, d.h. gesucht ist das Minimum der Funktion

$$D = \sum_{i=1}^n (\Delta B_i^{\{R,G,B\}})^2$$

Da ein Direktansatz über die Nullstellen der partiellen Ableitungen von  $F$  hier zu keinem Erfolg führt und das Modell nicht-linear von den Parametern abhängt, wird der Levenberg-Marquadt-Algorithmus als nicht-lineares Optimierungsverfahren eingesetzt. Voraussetzung für die Konvergenz ist ein guter Näherungswert, der mit dem Trick  $I_2 = I_1$  ermittelt wird.

In Abb. 43 sieht man die Anwendung des Algorithmus: Ausgehend von mehreren Fotos (hier nur eines abgebildet) werden die LVK der Lichtquelle und die Reflexionskoeffizienten der Materialien bestimmt. Vorgegeben wurde dabei der Reflexionskoeffizient der Wand als (0.6, 0.6, 0.6). Die daraus resultierende Radiositysimulation sieht dem Foto ähnlich. Man sieht eine Erweiterung des Fotos aus Abb. 42 basierend auf dieser Radiositysimulation.



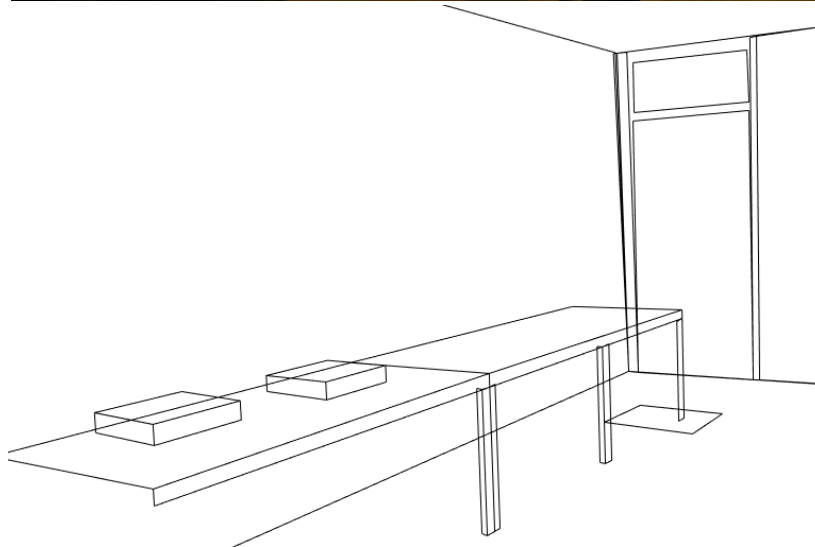


Abb. 42 : Von oben nach unten: Das Originalfoto, die rekonstruierte Geometrie, Überlagerung beider Bilder.

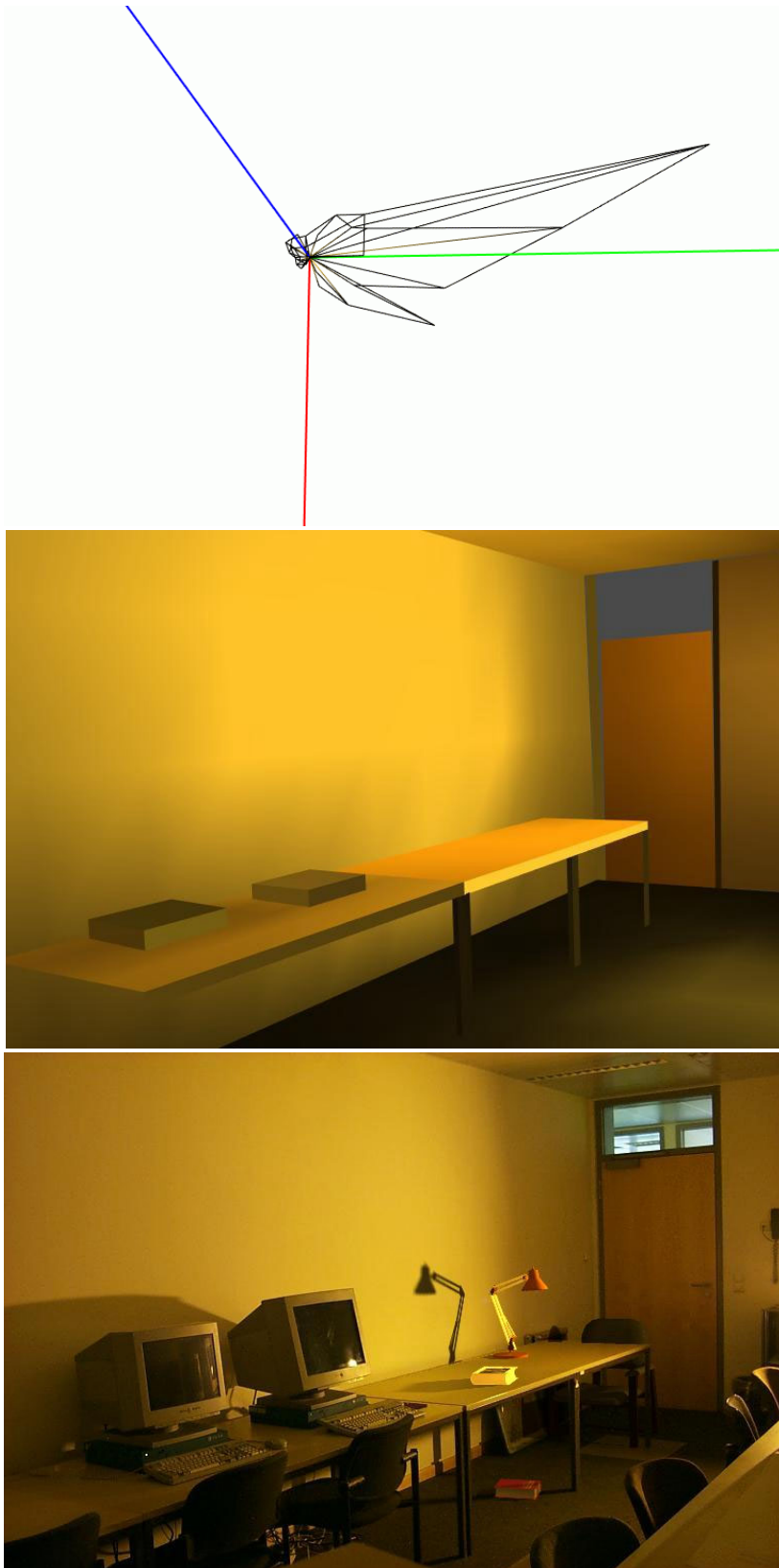


Abb. 43 : Von oben nach unten: Die rekonstruierte LVK. Das Ergebnis der 1.Radiositysimulation. Das fertige, mit Büchern und Lampe erweiterte Bild.

## Heuristiken für fehlende Geometrien

Bei der Rekonstruktion von LVK und Reflexionskoeffizienten wird die indirekte Beleuchtungsstärke benötigt. Liegt von der Szene nur der auf den Fotos sichtbare Teil vor, so ist die Berechnung der indirekten Beleuchtungsstärke nur von diesem Teil möglich.

Für ein gegebenes Patch kann dessen Radiosity mit der durchschnittlichen Pixel-Radiosity der das Patch überdeckenden Pixel gesetzt werden, d.h. bei  $k$  Pixeln für ein Patch

$$B_{xy} = \pi T^{-1}(p_{xy})$$

$$B_{Patch} = \frac{1}{k} \sum_{i=1}^k \pi T^{-1}(p_i)$$

Hat man für die Szene eine Patchauflösung festgelegt, so kann man jedem Patch den Radiositywert  $B_{Patch}$  zuweisen und die indirekte Beleuchtungsstärke für einen gegebenen Empfängerpunkt berechnen.

Sind nicht alle Patches der Szene auf den Fotos sichtbar, so können für den fehlenden Szenenteil Heuristiken angewendet werden. Man kann beispielsweise den indirekten Anteil in zwei Teile aufspalten, einen sichtbaren und einen unsichtbaren.

Ist die 3D Szene zu einem Anteil rekonstruiert worden, so geht man davon aus, daß der nicht-rekonstruierte Teil genauso zur indirekten Beleuchtungsstärke beiträgt und schätzt ihn entsprechend ab. Natürlich kennt man den genauen Wert des sichtbaren Anteils nicht, weshalb dieser ebenfalls abgeschätzt werden muß.

Ist die nicht sichtbare Geometrie ergänzt worden, so kann man auch allen nicht sichtbaren Patches den durchschnittlichen Radiositywert  $\bar{B}$  aller Pixel zuweisen.

$$\bar{B} = \frac{1}{N} \cdot \sum_{i=1}^N \pi T^{-1}(p_i)$$

Hierbei ist  $N$  die Gesamtzahl aller Pixel der Fotos.

Das heißt, man geht hier davon aus, daß der nicht sichtbare Teil der Szene genauso hell ist wie der sichtbare.

Wird der nicht sichtbare Teil einfach vernachlässigt, so hat dies Auswirkungen auf die Rekonstruktion des direkten Lichts: Durch die Unterschätzung des indirekten Lichts wird der direkte Anteil überschätzt, d.h. die Lampe wird heller.

#### 4.2.4 Verwendung einer Lightprobe

Ein anderer Ansatz ist die Verwendung einer Lightprobe [D98] zur Lichtrekonstruktion.

Die Vorteile einer Lightprobe liegen darin, daß die genauen Richtungen und Raumwinkel mehrerer, komplexer Lichtquellen direkt aus den Pixelpositionen ihrer Reflexionen abgelesen werden können. Des weiteren spiegelt sich in der Lightprobe die gesamte Szene – insbesondere auch Gebiete, die in dem zu erweiternden Bild nicht sichtbar sind. Diese Informationen können während der Radiositysimulation genutzt werden.



Abb. 44 : Eine Lightprobe.

Rekonstruktion von Tageslicht wird ermöglicht, da beliebig geformte und Lichtquellen bearbeitet werden können. Die Integration von direktem Sonnenlicht erweist sich jedoch als problematisch, da insbesondere eine Lightprobe nur an ihrer genauen Position in der Szene exakt ist und so leicht Schatten von Fensterrahmen o.ä. verpaßt werden können. Des weiteren ist die Erzeugung nicht überbelichteter Lightprobefotos für diesen Fall kaum durchführbar.

Die aus einer Lightprobe gewonnenen Informationen sind natürlich nur für die Position der Lightprobe in der Szene korrekt – insbesondere die Rekonstruktion einer LVK ist auf Basis einer einzelnen Lightprobe nicht möglich.

#### Wahl der Fotos

Es ist sinnvoll die Restriktion „aus einem Bild“ etwas abzumildern. Sinnvoller ist die Verwendung von insgesamt drei Bildern:

1. Das eigentliche zu erweiternde Bild, ohne Lightprobe („*clean plate*“).
2. Ein zweites Bild mit identischer Kameraposition und -einstellungen, diesmal jedoch mit Lightprobe. Dieses Bild dient zur Bestimmung der Lightprobe-position in Relation zur Szene.
3. Eine Nahaufnahme der Lightprobe mit kurzer Belichtungszeit. Durch die dadurch höhere Auflösung kann die Lightprobe genauere Beleuchtungsinformationen liefern.

Sind die Kamerakennkurven bekannt (Kapitel 3.4.2), lassen sich aus den aufgenommenen Bildern photometrische und colorimetrische Lichtinformationen ablesen. Hierzu sind nicht notwendigerweise komplette Belichtungsserien und daraus rekonstruierte HDRIs erforderlich. Vielmehr liefert auch ein einzelnes Foto mit bekannter Belichtungszeit für alle nicht über- oder unterbelichteten Bildbereiche bereits exakte Werte. Dies ist im Allgemeinen ausreichend.

Voraussetzung für die Rekonstruktion korrekter Werte für die Lichtquellen ist in diesem Fall jedoch, daß keine der Lichtquellen in der Nahaufnahme der Lightprobe überbelichtete Bereiche verursachen, weshalb die Belichtungszeit dieses dritten Bildes entsprechend kurz gewählt werden muß.

### Das Calibrate-O-Meter

Stellt man statt einer einfachen Lightprobe ein erweitertes, spezielles Gerät (Abb. 45) in das Foto, können mehrere Zusatzinformationen auf einmal gewonnen werden:

- Der weiße Rahmen um die Lightprobe besitzt eine bekannte Dimension. Hierdurch wird ein Skalierungsfaktor zur maßstabsgerechten geometrischen Rekonstruktion vorgegeben.
- Der (mattweiße) Rahmen besitzt einen bekannten (da vorher bestimmten) Reflexionskoeffizient, der als Ausgangspunkt für die in diesem Kapitel vorgestellten Materialrekonstruktionsalgorithmen dient.
- Durch die feste Zuordnung von Lightprobe-Position zum Rahmen läßt sich deren genaue 3D-Position in der Szene (bei bekanntem Lightprobe-radius) rekonstruieren.
- Über das Verhältnis von Rahmenfarbe und der Farbe/Intensität der Reflexion des Rahmens in der Lightprobe lassen sich Reflektivität und Farbe der Lightprobe bestimmen.
- Schließlich dienen die von hinten beleuchteten Farbfilter, deren XYZ-Farbe bestimmt wurde, als drei bekannte, weit auseinanderliegende Referenzfarben, woraus sich die Farbmatrix der Kamera selbst entsprechend Kapitel 3.4.2 rekonstruieren läßt.

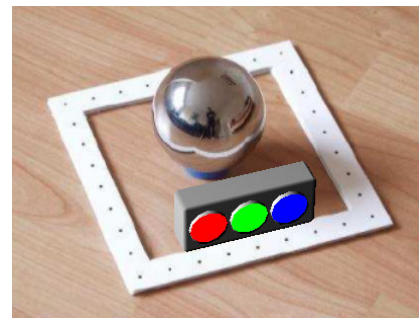


Abb. 45: Das Calibrate-O-Meter: Licht-, Materialien- und Kamerafarbmatrixrekonstruktion in einem.

- Und natürlich sämtliche Informationen, welche durch die integrierte Lightprobe selbst gewonnen werden können.

### Interaktive Echtzeitvisualisierung

Zur weiteren Visualisierung können die globalen Beleuchtungsverhältnisse über ein Irradiance Volume [GSHG98] in einem räumlichen Gitter bestimmt werden. Eingefügte Objekte werden mit Techniken aktueller Grafikhardware, wie Echtzeitgenerierung von Environment-Maps, sowie Shadow-Maps [GCHH03] diffus, spekulär, sowie unter Einfluss von Schatteneffekten visualisiert (siehe insbesondere [GCSV03], sowie [A01]). Der Nachteil dieser Methode von einem quantitativen Gesichtspunkt ist, daß insbesondere indirekte Lichteffekte verursacht durch die virtuellen Objekte nicht wiedergegeben werden können. Die interaktive photorealistische Darstellung ist jedoch visuell sehr überzeugend.

## 4.3 Konsistente Erweiterung von Fotos

Sind alle für eine konsistente Radiositysimulation notwendigen Parameter bestimmt, so kann mit der Erweiterung des Fotos begonnen werden.

Ein Foto kann durch mehrere Möglichkeiten erweitert werden:

- Änderung von Materialien
- Einfügen zusätzlicher und Änderung real vorkommender Lichtquellen
- Einfügen von virtuellen Objekten

Der zu Grunde liegende Algorithmus basiert auf [FGR93] und legt sich wie folgt dar:

1. Radiositysimulation ohne Erweiterungen und Speichern der Patch Radiosities als  $L_{old}$
2. Radiositysimulation mit Erweiterungen und Speichern der Patch Radiosities als  $L_{new}$
3. Fallunterscheidung pro Pixel:
  - Reales Objekt: Skalierung der Pixelfarben des Fotos, abhängig von  $L_{old}$  und  $L_{new}$ .
  - Virtuelles Objekt: Darstellung von  $L_{new}$

### 4.3.1 Datenstrukturen

#### PixeldataArray

In Anlehnung an [LFDW99] werden pro Pixel folgende Informationen gespeichert:

- Schnittpunkt mit der Szene
- Zeiger auf getroffenes Polygon (NULL wenn kein Objekt)
- Liste mit Formfaktor, Sichtbarkeit und Pixelhaft für jeden Emitter
- direkte und indirekte Radiosity aus erster Simulation

Diese Datenstruktur wird als *PixelData* bezeichnet, bzw. als *PixeldataArray* für das gesamte Bild.

#### Pixelhaft

Zur Bestimmung einer möglichen Änderung im direkten Licht für ein Pixel werden Shafts verwendet [HW91], die jedoch, aufgrund der punktförmigen Pixel, an einem Ende keine Boundingbox besitzen. Diese Shafts werden hier als Pixelshafts bezeichnet, sie beschreiben das Volumen zwischen dem 3D Punkt eines Pixels und der Boundingbox der Lichtquelle. Für jedes Pixel kann mit Hilfe des Pixelshafts in guter Näherung festgestellt werden, ob durch das virtuelle Objekt das direkte Licht für das Pixel verändert wird. Nur bei einer Überlagerung des Pixelshafts mit der Boundingbox des virtuellen Objekts liegt das Pixel möglicherweise im Schatten des virtuellen Objekts. Somit erspart man sich für viele Pixel die Neuberechnung des direkten Lichts. Ragt das virtuelle Objekt in den Pixelhaft hinein, so wird die Sichtbarkeit neu berechnet und der gespeicherte unverdeckte Formfaktor damit gewichtet.

#### Optimierung durch Hierarchie

Als Verbesserung der einfachen Pixelshafts lassen diese sich durch eine Hierarchie von Pixelshafts darstellen [G01]. Da benachbarte Pixel meist ähnliche Schnittpunkte besitzen, bietet sich eine Zusammenfassung benachbarter Pixelshafts an (z.B. 4 x 4 Pixel Blöcke). Die Boundingbox für einen Pixelblock ergibt sich dabei aus dem minimalen und maximalen Wert aller Schnittpunkte entlang aller Achsen.

Diese Pixelblöcke können wiederum zusammengefaßt werden, sodass eine Hierarchie von Pixelshafts entsteht. Die Wurzel des Baum beschreibt den Shaft für

das gesamte Bild. Zur Lokalisierung der Pixel, für die das direkte Licht verändert wurde, kann der Baum von der Wurzel aus traversiert werden: Überlagert das virtuelle Objekt den Shaft, so werden alle Kinder-Pixelshafts untersucht. Ist keine Überlagerung vorhanden, so braucht für den gesamten Pixelblock kein direktes Licht neu berechnet werden. Abb. 46 zeigt eine Visualisierung des hierarchischen Shaft Tests: Jeder eingefärbte Block beschreibt eine Region im Bild, die durch einen Shaft-Test von der Neuberechnung des direkten Lichts ausgeschlossen wurde.



Abb. 46 : Visualisierung der Pixelshaft-Hierarchie.

### 4.3.2 Vorgehensweise

Die beschriebenen Möglichkeiten, Fotos zu erweitern, lassen sich wie folgt mit aufsteigender Komplexität gruppieren:

1. Nur Materialänderungen
2. Lichtänderungen (und evtl. Materialänderungen)
3. Virtuelle Objekte (und evtl. Lichtänderungen oder Materialänderungen)

#### Materialänderungen

Werden nur Materialien geändert (und keine Lichtquellen oder virtuellen Objekte eingefügt), so wird auf den Aufbau des PixeldataArray verzichtet. Da keine virtuellen Objekte vorhanden sind, werden auch keine Pixelshafts benötigt. Auch die Trennung von direktem und indirektem Licht ist nicht notwendig. Der Algorithmus sieht dann folgendermaßen aus:



```

1. Radiositysimulation
Speichere  $L_{old}$  bei jedem Patch
Aktiviere geänderte Materialien
2. Radiositysimulation
Speichere  $L_{new}$  bei jedem Patch
Berechne Korrekturterm für jedes Patch
Berechne Korrekturterm für jeden Patchcheckpunkt
durch Extrapolation
Für alle Pixel
    Berechne Schnittpunkt mit Szene
    Interpoliere Korrekturterm aus Patchcheckpunkten
    Skalieren Pixel-Leuchtdichte mit Korrekturterm
    und ermittle neue Pixelfarbe

```

Der Korrekturterm für das differentielle Rendering ist dabei entweder  $L_{new} - L_{old}$  oder  $L_{new} / L_{old}$  (vgl. Kapitel 4.3.7).

## Lichtänderungen

Werden Lichteigenschaften verändert, so wird das PixeldataArray aufgebaut und zwischen direktem und indirektem Licht unterschieden. Da keine virtuellen Objekte vorhanden sind, werden keine Pixelshafts benötigt. Durch die Änderung der Lichtquellen muß das direkte Licht für alle Pixel neu berechnet werden. Der Algorithmus ändert sich wie folgt:

```

1. Radiositysimulation
Entferne direktes Licht
Erzeuge PixeldataArray ohne Pixelshafts
Für alle Pixel
    Berechne direktes Licht  $L_{old}^{dir}$  und speichere es in PixelData
    Interpoliere indirektes Licht  $L_{old}^{ind}$  aus EckpunktRadiosities und
    speichere es in PixelData
Aktiviere geänderte Lichtquellen
Füge neue Lichtquellen ein
Aktiviere geänderte Materialien
2. Radiositysimulation
Entferne direktes Licht
Für alle Pixel
    Erweitere Emitterinformationen (keine Pixelshafts)
    Berechne direktes Licht  $L_{new}^{dir}$ 
    Interpoliere indirektes Licht  $L_{new}^{ind}$ 

$$L_{new} = L_{new}^{dir} + L_{new}^{ind}$$


```

$$L_{old} = L_{old}^{dir} + L_{old}^{ind}$$

Berechne KorrekturTerm und neue Pixelfarbe

Zur Entfernung des direkten Lichts, sowie der Neuberechnung pro Pixel vgl. Kapitel 4.3.3. Die Erweiterung der Emitterinformationen besteht aus der Berechnung von Formfaktor und Sichtbarkeit für alle Einträge im PixeldataArray. Bei modifizierten Lichtquellen werden die vorhandenen Werte aktualisiert, bei neu eingefügten Lichtquellen werden die Arrays mit Formfaktoren und Sichtbarkeit erweitert.

## Virtuelle Objekte

Bei eingefügten virtuellen Objekten werden alle Datenstrukturen angelegt. Im PixeldataArray werden neben Formfaktor und Sichtbarkeit Pixelshafts zur Erkennung von Änderungen des direkten Lichts gespeichert. Direktes und indirektes Licht aus der ersten Radiositysimulation werden getrennt pro Pixel gespeichert. Für alle Pixel eines virtuellen Objekts wird das direkte Licht neu berechnet. Bei einem realen Objekt wird das direkte Licht für ein Pixel nur dann neu berechnet, wenn ein virtuelles Objekt den Pixelshaft überlagert. Bei Änderungen an den Lichtquellen müssen bei der Erweiterung der Emitter Informationen Formfaktor, Sichtbarkeit und Pixelshaft aktualisiert werden (Direktes Licht wird in diesem Fall für alle Pixel neu berechnet). Die Aktualisierung des indirekten Lichts wird in Kapitel 4.3.6 beschrieben. Zur besseren Übersicht wurde im folgenden Pseudocode auf die vorgestellte Hierarchie von Pixelshafts verzichtet.

```

Radiositysimulation
Entferne direktes Licht
Erzeuge PixeldataArray (mit Pixelshafts)
Für alle Pixel
    Berechne direktes Licht  $L_{old}^{dir}$  und speichere es in PixelData
    Interpoliere indirektes Licht  $L_{old}^{ind}$  aus EckpunktRadiosities
    und speichere es in PixelData
Aktiviere geänderte Lichtquellen
Aktiviere geänderte Materialien
Für alle Pixel
    Erweitere Emitterinformationen
    Füge virtuelle Objekte ein
    Berechne neues indirektes Licht
    Für alle Pixel
        Falls virtuelles Pixel oder reales Pixel und Pixelshaft
        von virtuellem Objekt betroffen
            Berechne direktes Licht  $L_{new}^{dir}$ 
        sonst
            Setze  $L_{new}^{dir} = L_{old}^{dir}$ 

```

Interpoliere indirektes Licht  $L_{new}^{ind}$

$$L_{new} = L_{new}^{dir} + L_{new}^{ind}$$

$$L_{old} = L_{old}^{dir} + L_{old}^{ind}$$

Berechne KorrekturTerm und neue Pixelfarbe

### 4.3.3 Direktes Licht

#### Entfernen direkten Lichts

Die Entfernung des direkten Lichtanteils gestaltet sich in einer konvergierten hierarchischen Radiositysimulation sehr einfach.

Da die Linkstruktur eine vollständige Beschreibung der Energieverteilung innerhalb der Szene beschreibt, wird hierzu einfach ein zusätzlicher Gathering-Schritt durchgeführt, bei dem Links von Lichtquellen ignoriert werden. Übrig bleibt die rein von indirektem Licht verursachte Energiesituation.

#### Neuberechnung durch *final gathering*

Da es sich bei dem gewünschten „Ausgabemedium“ um ein Standbild handelt – nämlich gerade das erweiterte Foto – und eine Echtzeitvisualisierung mit veränderbarem Kamerastandpunkt nicht notwendig ist, bietet es sich an, die in Kapitel 2.1.2 vorgeschlagene, pixelbasierte Qualitätsoptimierung des *final gathering* (z.B. [SSS01]) anzuwenden. Dabei wird wie beim Raytracing für jedes Pixel der Schnittpunkt des Strahls Kamera→Pixel mit der Szenengeometrie bestimmt. Aus der Radiositysimulation ist der indirekte Anteil für die Helligkeit des Pixels bekannt. Zur besseren Darstellungsqualität wird mit Gouraud-Shading ein aus den Eckpunkt-Radiosities interpolierter Wert für den indirekten Anteil verwendet. Der direkte Anteil wird mit den in den nächsten Abschnitten vorgestellten Verfahren neu berechnet. Die "Pixel-Radiosity"  $B_{Pixel}$  berechnet sich dann als Summe aus direktem und indirektem Anteil:

$$B_{Pixel} = B^{dir} + B^{ind} .$$

Für Flächenlichtquellen lautet die Radiositygleichung des direkten Anteils

$$B^{dir} = \rho_E \cdot \sum_{S=1}^L E_S F_{ES}$$

Da aufgrund des punktförmigen Pixels über den Sehstrahl nur auf einem punktförmigen Bereich des sichtbaren Objektes der direkte Lichtanteil gesucht ist,

kann als Formfaktor  $F_{ES}$  direkt die Prisma-Formel herangezogen werden. Zur weiteren Beschleunigung der Berechnung wird die Prisma-Formel durch die zweite Formfaktorvereinfachung ersetzt, wenn die folgende Faustformel der Lichttechnik erfüllt ist:

*„Der Abstand zwischen Sender und Empfänger muß mindestens zehnmals größer sein als die größte Querdimension des Senders.“*

Für punktförmige Lichtquellen kann der direkte Lichtanteil entsprechend der in Kapitel 2.4 angegebenen Formeln ermittelt werden.

#### 4.3.4 Sichtbarkeitsbestimmung

Für Punktlichtquellen genügt normalerweise ein Strahltest Empfängerpunkt→Lichtposition zur Sichtbarkeitsbestimmung. Die daraus resultierenden Schatten sind allerdings unrealistisch aussehende Schlagschatten, wie sie z.B. von Raytracern produziert werden. Um diesen Effekt zu vermeiden, werden auch bei Punktlichtquellen mehrere Strahlen für die Sichtbarkeitsberechnung verwendet, die um die Lichtposition verteilt werden.

Die Form der Lichtquelle wird dabei als (nicht notwendigerweise achsenparalleler) Quader angenommen, wobei dessen Größe ggf. vom Benutzer so eingestellt wird, daß sie der Form der realen Lichtquelle möglichst nahe kommt. Zur Berechnung der Sampling-Positionen wird der Quader rekursiv jeweils entlang der längsten Achse halbiert, bis eine maximale Anzahl an Quadern erreicht ist. Durch diese Art der Unterteilung wird gewährleistet, daß die entstehenden Gitterzellen möglichst würfelförmig sind.

Innerhalb der Gitterzellen werden jetzt  $m$  Zufallspunkte berechnet und in einem Array abgelegt. Das Refinementkriterium bestimmt sodann einen Level  $L$ , auf dem die Sichtbarkeit bestimmt werden soll, d.h. es werden  $2^L$  Samples verwendet. Für jede Gitterzelle wird einer der vorberechneten Punkte zufällig ausgewählt.

Der Grund für diese Vorgehensweise liegt darin, daß auf diese Weise für unterschiedliche Pixel und unterschiedliche Genauigkeitslevel sichergestellt ist, daß immer dieselben geometrischen Punkte zur Sichtbarkeitsbestimmung herangezogen werden, wodurch sichtbare Artefakte, die als *pixel noise* in Erscheinung treten würden, erfolgreich vermieden werden.

```
for i = 0..2^L-1
    k=Random(m*2^(Lmax-L))
```

$$\text{point\_index} = i * m * 2^{(L_{\max} - L)} + k$$

Für das Level  $L$  ist die Intervallbreite  $m \cdot 2^{L_{\max} - L}$ , sodass mit  $k$  einer der Punkte aus dem Intervall  $i$  gewählt wird. Der Index  $i$  durchläuft alle Quader des Levels  $L$ .

Durch die Vereinfachung der ausgedehnten Lichtquelle als punktförmige Lichtquelle ist eine exakte Berechnung des Halbschattenbereichs nicht möglich, das oben beschriebene Verfahren liefert aber gute Näherungslösungen.

### 4.3.5 Bewegung virtueller Objekte

Für eine intuitive Bewegung virtueller Objekte in einer Szene erfolgt die Positionierung des Objekte, ausgehend von einer per Mausklick ausgewählten 2D-Position, automatisch in der 3D-Szene.

Dazu wird für die Gerade Kamera→Pixel der Schnittpunkt mit der Szene bestimmt. Das virtuelle Objekt wird zunächst mit seinem Mittelpunkt an diesen Punkt verschoben. Da es jetzt nur zur Hälfte sichtbar ist, muß es jedoch noch zum Betrachter hin verschoben werden. Ist die Flächennormale des Polygons eine der drei Hauptachsen, so kann man das virtuelle Objekt einfach um die Hälfte der Länge seiner Boundingbox entlang der Flächennormalen verschieben. Abb. 47 zeigt, daß somit das virtuelle Objekt mit seiner Boundingbox genau an die getroffene Fläche angrenzt.

Ist die Richtung der Flächennormale ungleich einer Achsenrichtung, so kann man für den Verschiebungsabstand eine Näherung anwenden.

Für alle Eckpunkte der Boundingbox des virtuellen Objekts wird der Vektor  $\vec{v}_i$  (BBoxMittelpunkt→BBoxEckpunkt) auf den Normalenvektor  $\vec{n}$  der getroffenen Fläche projiziert ( $d_i = \vec{v}_i \cdot \vec{n}$ ). Der größte der  $d_i$  Werte wird als Verschiebung entlang des Normalenvektors verwendet. Die Boundingbox grenzt dann mit einem Eckpunkt an die getroffene Fläche an (Abb. 47).

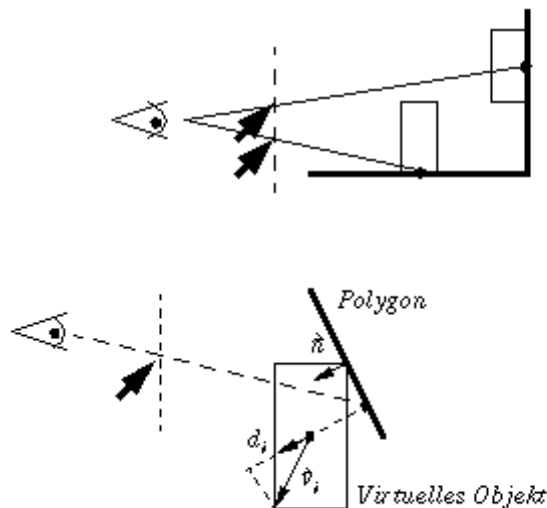


Abb. 47: Plazierungshilfen beim Einfügen virtueller Objekte.

### 4.3.6 Aktualisierung indirekten Lichts

Bewegt der Benutzer ein virtuelles Objekt, so muß die Radiosity-Lösung aktualisiert werden. Da das direkte Licht pro Pixel ohnehin neu berechnet wird, bezieht sich die Aktualisierung nur auf das indirekte Licht. Eine komplette Neuberechnung dauert in der Regel für interaktive Anwendungen zu lange, daher wird hier die *line space hierarchy* [DS97] eingesetzt. Ausgehend von einer konvergierten Lösung werden dabei schnell die von der Bewegung des virtuellen Objekts betroffenen Links identifiziert und deren Empfänger-Radiosity korrigiert. Anschließend wird das direkte Licht entfernt und mit dem Ergebnis ein erweitertes Bild generiert.

#### Vorberechnung und Interpolation

Für indirektes Licht lassen sich vor allem zwei Eigenschaften beobachten:

- Wird ein Objekt bewegt, so ändert sich das indirekte Licht in der Szene meistens langsam und stetig, selten sprunghaft.
- Durch die Bewegung eines Objekts sind die sichtbaren Änderungen der Szene bezüglich des indirekten Lichts meistens auf einen kleinen Teil der Geometrie beschränkt, viele Objekte bleiben visuell unverändert.

Basierend auf diesen Beobachtungen kann das indirekte Licht auch mit folgendem Verfahren angenähert werden: Das virtuelle Objekt wird zunächst an verschiedene Positionen bewegt und die Radiosity-Lösung mit der *line space hierarchy* aktualisiert. Nach Entfernen des direkten Lichts wird die neue Radiosity-Verteilung mit einer Referenzlösung verglichen und deutlich sichtbare Änderun-

gen werden gespeichert. Bewegt der Benutzer das virtuelle Objekt später an eine Position, so wird das indirekte Licht aus den vorberechneten Werten interpoliert.

```

Festlegung einer ROI und Schrittweiten
Bewege virtuelles Objekt in die Mitte der ROI
Berechne konvergierte Radiosity Lösung und entferne direktes Licht

Speichere indirekte Radiosity-Werte als Referenzwerte  $B_{ind}^{ref}$ 
Füge direktes Licht ein
Für alle Positionen in der ROI
    Bewege virtuelles Objekt an Position
    Aktualisiere Radiosity-Lösung mit line space hierarchy
    Entferne direktes Licht
    Für alle Patches
        Falls  $|T(B_{ind}) - T(B_{ind}^{ref})| > \epsilon$ 
            Speichere  $B_{ind}$  und Patch in IndirectInfoArray
        Füge direktes Licht ein

```

Eine *region of interest* (ROI) ist dabei ein Quader, der den für ein virtuelles Objekt interessanten Bereich darstellt. Dieser Bereich muß vom Benutzer festgelegt werden, z.B. ein Bereich an einer Wand für ein Regal oder ein Bereich am Fußboden für einen Tisch. Bei der Verwendung einer Lightprobe also beispielsweise die Position der Lightprobe im Foto.

Zum Speichern der vorberechneten indirekten Radiositywerte werden in der Datenstruktur IndirectInfo die vorgefundene indirekte Radiosity, sowie ein Zeiger auf das zugehörige Patch festgehalten.

Für jede Position in der Region of Interest wird in der Vorberechnungsphase ein Array aus dieser Datenstruktur angelegt (IndirectInfoArray). In einem Eintrag steht die indirekte Radiosity und ein Zeiger auf das entsprechende Patch. Befindet sich das virtuelle Objekt in der Vorberechnungsphase an einer Position in der Region of Interest, so wird für jedes Patch überprüft, ob die Farbe der indirekten Radiosity eine deutliche Abweichung von der Referenzlösung besitzt. Ist dies der Fall, so wird im IndirectInfoArray ein Eintrag mit der indirekten Radiosity und einem Zeiger auf das Patch angelegt.

Bewegt der Benutzer das virtuelle Objekt, so werden zunächst die acht umliegenden Positionen der Region of Interest ermittelt. Die indirekte Radiosity eines Patches wird nun als interpolierter Mittelwert aus den acht Radiositywerten bestimmt, die das Patch hatte, als das virtuelle Objekt in der Vorbereitung an diesen Positionen war. Für alle acht Positionen werden die dort gespeicherten IndirectInfoArrays durchlaufen. Für jedes dort eingetragene Patch wird ein Ar-

ray mit acht Einträgen (IndirectValues) zur trilinearen Interpolation beim Patch angelegt. Da die indirekte Radiosity bei geringer Abweichung nicht gespeichert wurde, werden zunächst alle acht Einträge mit dem Referenzwert  $B_{ind}^{ref}$  für dieses Patch initialisiert. Befindet sich im IndirectInfoArray an Position  $i$  ein Eintrag mit einer indirekten Radiosity  $B_{ind}$ , so wird im IndirectValues Array des Patches an der Position  $i$  die indirekte Radiosity  $B_{ind}$  eingetragen. Zur Aktualisierung der indirekten Radiosity werden nur die Patches mit angelegten IndirectValues herangezogen und mit trilinearer Interpolation ein Mittelwert bestimmt.

### 4.3.7 Darstellung

#### Differentielles Rendering

Die grundlegende Idee beim Erweitern von Fotos durch virtuelle Objekte ist, die vorhandenen Pixelfarben zu modifizieren, anstatt komplett neue Farben zu berechnen [FGR93][D98].

$p_{xy}$  ist die Pixelfarbe eines realen Objekts. Mit  $L_{xy}$  wird die Leuchtdichte des Pixels beschrieben, also

$$L_{xy} = T^{-1}(p_{xy})$$

wobei  $T$  die Kamerakurve zur Umrechnung von Leuchtdichten in Pixelfarben ist.

$L_{old}$  ist der Radiositywert aus der Simulation ohne virtuelle Objekte und  $L_{new}$  ist der Radiositywert aus der Simulation mit hinzugefügten virtuellen Objekten. Bei einer guten Simulation werden  $L_{xy}$  und  $L_{old}$  ähnliche Werte annehmen.

Für die Änderung der Pixelfarbe gibt es zwei mögliche Strategien:

1. Die Differenz der Helligkeiten soll erhalten bleiben

$$L_{xy} = T^{-1}(p_{xy})$$

$$\Delta L = L_{new} - L_{old}$$

$$L_{xy}' = L_{xy} + \Delta L$$

$$p_{xy}' = T(L_{xy}')$$

2. Die relative Helligkeitsänderung soll erhalten bleiben



$$L_{xy} = T^{-1}(p_{xy})$$

$$s = \frac{L_{new}}{L_{old}}$$

$$L_{xy}' = L_{xy} \cdot s$$

$$p_{xy}' = T(L_{xy}')$$

Für eine perfekte Simulation mit  $L_{xy} = L_{old}$  liefern beide Varianten als neue Pixel-farbe

$$p_{xy}' = T(L_{new})$$

Bei Abweichungen zwischen  $L_{xy}$  und  $L_{old}$  unterscheiden sich die Ergebnisse. Insbesondere wird Variante 1 undefiniert, wenn der neue Radiositywert negativ wird, hier sollte an 0 geclippt werden. Bei Variante 2 ist dies der Fall, wenn  $L_{old} = 0$  ist. Auch dunkle Pixel mit kleinen Leuchtdichten sind hier problematisch, ein gar komplett schwarzes Pixel wird hiermit auch nicht heller skaliert.

### Anti-Aliasing

Fotos realer Szenen enthalten durch die Kameraoptik ein natürliches Anti-Aliasing. Werden hier ohne weitere Bearbeitung virtuelle Objekte eingeblendet, entstehen Aliasing-Treppenstufen an den Pixelkanten, und trotz übereinstimmenden Beleuchtungsverhältnissen zwischen realen und virtuellen Szenenbereichen geht hierdurch ein Großteil des Realismus verloren.

Abhilfe schaffen hier adaptierte Standardverfahren zur Aliasing-Vermeidung [FDFH90]. Für jedes Subpixel wird der Schnittpunkt mit der Szene ermittelt. Trifft jeder Strahl ein reales Objekt so wird das Pixel nicht verändert. Wird bei mindestens einem Subpixel ein virtuelles Objekt getroffen, so wird die Farbe des Pixels neu berechnet. Für jedes Subpixel  $ij$  wird zwischen realem und virtuellem Objekt unterschieden: Wird ein virtuelles Objekt getroffen, so wird die Leuchtdichte  $L_{ij}$  des virtuellen Objekts an dieser Stelle bestimmt. Wird ein reales Objekt getroffen, so wird mit differentielltem Rendering aus der vorhandenen Pixel-Leuchtdichte die korrigierte Leuchtdichte berechnet. Dazu müssen für den Schnittpunkt des Subpixels die Werte  $L_{old}$  und  $L_{new}$  bestimmt werden.

Die neue Leuchtdichte des Pixels wird dann mit

$$L_{xy} = \frac{1}{N^2} \cdot \sum_{i=1}^N \sum_{j=1}^N L_{ij}$$

bestimmt. Das differentielle Rendering-Verfahren wird hier praktisch mit einer höheren Auflösung angewandt, was zu einer realistischeren Darstellung führt. Der einzige Nachteil ist der hohe Zeitbedarf.

## 5 Konkrete Anwendungsbeispiele

### 5.1 Lichtsimulation im Dom von Siena

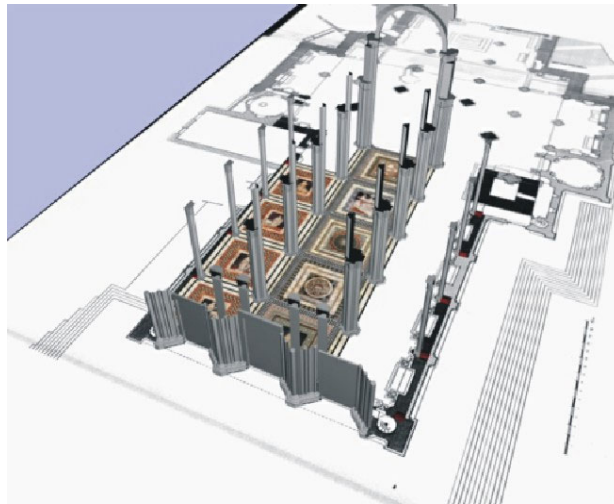


Abb. 48 : Generierung des Dommmodells auf Basis von Bauplänen und Skizzen

Das komplexe Modell des Doms von Siena [BFKK01] stellte hohe Ansprüche an die Leistungsfähigkeit und Speichereffizienz des Radiositysystems. Ausgehend von einem Polygonalen Modell aus 150000 texturierten Polygonen wurden 500 Spot- und Punktlichtquellen in der Szene platziert, um eine realistische Wiedergabe der in der Realität vorhandenen Beleuchtungsverhältnisse im Dom zu ermöglichen. Weitere 300 Polygone bildeten außerhalb des eigentlichen Doms eine Repräsentation der Lichtverteilung des diffusen Himmels um Tageslicht zu simulieren. Das Tageslicht konnte in den Dom durch die oberen Fenster und die zwei großen Rundfenster eindringen und leuchtete so den Dom zusammen mit den Kunstlichtquellen aus. Für alle Lichtquellen wurden passende Farbtemperaturen und Helligkeiten vorgegeben.



Abb. 49 : Blick in die Kathedrale. Links Standard-OpenGL-Shading, rechts das Ergebnis der Radiositysimulation.



Abb. 50 : Untexturierte und Wireframe-Darstellung der *Libreria piccolomini*. Man beachte die Adaption der Schatten von Stuhl und Statue auf dem Fußboden.