

The Office of the Future

A Unified Approach to Image–Based Modeling and Spatially Immersive Displays

Von

Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake,
Lev Stesin und Henry Fuchs

University of North Carolina at Chapel Hill and the NFS Science and Technology Center for Computer Graphics and
Scientific Visualization

Vortrag von Markus Lupert

Inhaltsverzeichnis

1 Einführung.....	3
2 Elementare Komponenten.....	3
2.1 Dynamisches Image–Based Modeling.....	3
2.2 Rendering.....	4
2.3 Spatially Immersive Displays.....	4
3 Methoden.....	4
3.1 Dynamische Image–Based Modellierung.....	4
3.1.1 Tiefenextraktion.....	4
3.1.2 Farbe und Reflectance.....	5
3.2 Nicht wahrnehmbares, strukturiertes Licht.....	6
3.3 Geometrische Registrierung.....	6
3.4 Rendering und Display.....	6
3.4.1 Herausforderungen.....	7
3.4.1.1 Geschwindigkeit.....	7
3.4.1.2 Parallelisierung.....	8
3.4.1.3 Latenz.....	8
3.5 Erzeugen von Übergangsfunktionen.....	8
3.6 Vereinfachung der Tiefe von Daten.....	9
3.7 Tracking.....	9
4 Momentane Implementation.....	10
5 Zukünftige Arbeit.....	11
6 Zusammenfassung und Schlussfolgerung.....	11
7 Referenzen.....	12

1 Einführung

Telekollaborationssysteme haben sich in den letzten Jahren bedeutend verbessert. Es gibt eine Vielzahl verschiedener Systeme, um eine virtuelle Umgebung zu implementieren, wie zum Beispiel das Head-Mounted Display (HMD). Der Benutzer setzt sich ein HMD auf und verschwindet von der realen Welt. Er taucht in eine virtuelle Umgebung »innerhalb des Displays« ein, in der er zum Beispiel virtuelle Objekte zusammen mit 3D Repräsentationen seiner Mitarbeiter sehen kann. Aber solche Systeme haben leider mehrere Nachteile. Am deutlichsten sind die Ergonomieprobleme wie Grösse, Gewicht, Mobilität, begrenzte Auflösung und begrenztes Blickfeld.

Eine attraktivere Alternative bei der der Anwender kein Display aufsetzen muss, ist ein spatially immersive Display (SID). Ein SID ist ein Display, das den Betrachter physikalisch mit einem Panoramabild umgibt. SID's sind typischerweise in Raumgrösse, dadurch haben sie Platz für mehrere Betrachter. Die wahrscheinlich bekanntesten Beispiele für Mehrzweck SID's sind das Cave Automated Virtual Environment CAVE™, Alternate Realities' VisionDome. Aber alle haben Nachteile, die man gerne mit dem Office of the Future System beheben möchte. Allen gemeinsam ist, dass die viel Platz erfordern, zum Beispiel in einem Labor, in dem genug Raum für die Kameras und Projektoren ist. Ihr System aber soll in einem bereits existierenden Büro leicht integriert werden können, indem man die benötigten Kameras und Projektoren an der Decke installiert. (Siehe Abbildung 1)



Abbildung 1: Wie das Office of the Future einmal aussehen sollte



Das einzige System, das ähnlich dem Office of the Future ist, ist der Teleport. Das Teleportsystem benutzt wandgrosse Displays. Jedes zeigt eine synthetische Szene, die mit Videobildern der entfernten Teilnehmer vermischt wird. Es ist in einem sorgfältig entworfenen Raum eingebettet, um in den gerenderten Raum zu passen. Das Ziel des virtuellen Raumes ist es, als Erweiterung des realen Raumes zu erscheinen. Das lokale und entfernte Büro scheinen entlang einer gemeinsamen Verbindung physikalisch miteinander verbunden zu sein. Aber anders als der Teleport, welcher 2D Video des entfernten Teilnehmers über ein virtuell angrenzendes Büro legt, wollen sie image-based 3D Rekonstruktionen des gesamten Büros.

Auf Grund des dynamischen image-based Modells kann nun beinahe alles im Büro als SID dienen.

2 Elementare Komponenten

Die Idee des Office of the Future vereint mehrere fundamentale Bereiche der Informatik.

2.1 Dynamisches Image-Based Modeling

Eine Hauptkomponente des Systems ist ein Modul, welches in Echtzeit image-based Modelle der Büroumgebung erfasst, einschliesslich der bestimmten Projektionsflächen. Über die Bestimmung der Tiefe einer Szene wurde schon viel geschrieben. Allgemeinere Ansätze beinhalten die Tiefe von Bewegung, Stereo, Focus und Defocus. Verglichen mit Echtzeit sind aber viele der traditionellen Techniken wegen ihrer Computer- und Bandbreitenanforderungen schwierig. Es gibt aber schon einige Verfahren, die dieses Problem behandeln.

2.2 Rendering

Ihre Vision des Office of the Future erfordert die Fähigkeit Bilder zu erzeugen, die zu einem sich bewegenden head-tracked Betrachter korrekt erscheinen, wenn sie auf die Projektionsfläche projiziert werden. Dies gilt auch für ein System wie CAVE™. Aber hier ist die Situation schwieriger, da sie in der Lage sein wollen, auf allgemeine Oberflächen zu projizieren, während das CAVE™ System auf planare Oberflächen »zurechtgeschnitten« ist.

Sie müssen bestimmen, wie sie die Bilder im voraus verzerren, sodass sie vom Blickpunkt des Betrachters aus als korrekt erscheinen, wenn sie vom Blickpunkt des Projektors aus projiziert werden.

2.3 Spatially Immersive Displays

Das bekannteste spatially immersive Display ist wahrscheinlich das CAVE™. Es behandelt weder Intensitätsübergänge, noch kennt es eine Methode, die Geometrie der Umgebung zu erfassen. Dies ist auch verständlich, da dies kein beabsichtigtes Ziel des Systems ist.

Militärsimulations-/Flugsimulatorindustrie beschränken sich gewöhnlich (aber nicht immer) auf die Übereinstimmung der Naht der Displays. Eine andere Technik der Flugsimulatorindustrie ist die Platzierung eines hochauflösenden Displays in die Mitte der Sicht des Users und die Projektion eines niederauflösenden Bildes auf den umliegenden Bildschirm. Obwohl dies effektiv ist, kann es nicht repositioniert werden und eine Naht kann entstehen, wo das hochauflösende Bild auf das niederauflösende trifft.

Ein Kuppeldisplay wie der VisionDome ist ein weiteres Beispiel für ein SID. Solche Systeme sind oft auf einen hochauflösenden Projektor beschränkt und verwenden selten Mechanismen, um die Tiefe oder die Projektionsoberflächeninformationen einer Szene zu erfassen.

Ein letzter, wichtiger Punkt über all diese Systeme ist, dass sie sich auf spezielle Geometrieconfigurationen verlassen und keine allgemeine Lösung präsentieren, da sie typischerweise eine nicht veränderbare Umgebung mit einheitlichen, idealen Projektionsflächen hatten. Also hatten sie auch Kontrolle über jede Ausgabe des Displaysystems, von der Beleuchtung bis hin zur genauen Kalibrierung der Projektoren. Im Office of the Future wird jedoch eine allgemeine Lösung des Projizierens auf willkürliche Projektionsflächen mit Echtzeit, automatischen Kalibrierungsverfahren vorgezogen.

3 Methoden

Die benutzten Kameras und Projektoren können in einem Capture oder Display Modus betrieben werden. Im Capture Modus können Kameras und Projektoren zusammen genutzt werden, um Tiefe und Reflectance Informationen von den bestimmte Displayflächen pro Pixel zu erhalten. Im Display Modus können die image-based Modelle zum Rendern und zur Projektion von geometrisch und photometrisch korrekte Bilder auf die möglicherweise unregelmässigen Oberflächen gebraucht werden.

3.1 Dynamische Image-Based Modellierung

Image-based Modeling ist ein schwieriges Problem, das die Computer Vision schon seit mehreren Jahren beschäftigt.

3.1.1 Tiefenextraktion

Die Tiefenextraktionsmethode sollte in grossen Arbeitsräumen funktionieren, welche Hochfrequenztexturen, als auch keine Texturen aufweisen. Um Projektionsoberflächen zu modellieren, braucht man eine hohe Genauigkeit, damit die Projektionen korrekt sind. Um dynamisch ändernde Szenen zu modellieren sind höhere Updateraten erforderlich, um die Bewegung mit möglicherweise geringerer Auflösung und Genauigkeit darzustellen. Diese Methode sollte nicht aufdringlich sein, um die Arbeitsatmosphäre nicht zu stören. Damit sind also z.B. Laser ausgeschlossen.

Das System benutzt momentan eine Videokamera zusammen mit einem Projektor, obwohl auch mehrere Kameras mit einem Projektor arbeiten könnten. Das Problem der Übereinstimmung wurde mit der Projektion von binär kodierten, vertikalen Balken [1] gelöst. Die Kamera blickt auf ein Set von n aufeinanderfolgenden Bildern, kriert mit adaptiven Schwellwerten auf Pixelbasis binäre Bilder, und erzeugt einen n-Bit Code für jedes Pixel, der mit dem Code der vertikalen Balken, der bei diesem Pixel dargestellt wird, übereinstimmt. Dies erlaubt zwischen 2^n projizierten, vertikalen Balken zu unterscheiden. Eine vorberechnete Triangulations Lookup Table, basierend auf Kalibrierungsdaten, ermöglicht trilineare Interpolation, um die Schnittpunkte von Pseudo-Kamera Strahlen und projizierten, vertikalen Ebenen zu berechnen. Die 3D Koordinaten von den Oberflächenpunkten, dargestellt bei jedem Pixel, werden später mit den Farbinformationen gebraucht, um das image-based Modell zu vervollständigen. Die vertikalen Balken können wiederholt projiziert werden, um die Tiefe für dynamische Szenen zu berechnen. Dieselbe Methode kann benutzt werden für das Scannen von Wänden, Menschen oder sich bewegende Objekte. (Siehe Abbildung 2)

Die Wahl der Kamera und des Projektors, um aktiv 3D Umgebungen zu scannen, ist erforderlich, um genug Anhalts-



Abbildung 2: Einige Resultate von Tiefenextraktion mittels binär kodiertem, strukturierten Licht

punkte für eine Übereinstimmung bereitzustellen, da die Projektionsflächen keine Texturen haben können. Der Gebrauch des selben Projektors zum Scannen und Darstellen erlaubt eine Vereinigung der beiden Aufgaben. Die einzige Zusatzkomponente ist dabei eine Kamera. Geschwindigkeit gegenüber Genauigkeit Trade-offs führten sie zu zwei Typen von Kamera-Projektor-Paaren. Relativ statische Displayflächen, wie Wände und Möbel in Büros, werden von den nach aussen schauenden Kamera-Projektor Paaren genauer und langsamer modelliert, als Menschen und sich bewegende Objekte, welche von nach innen schauenden Kamera-Projektor Paaren gescannt werden.

Die Schwierigkeit bei der Benutzung von zwei separaten Geräten für die Tiefenextraktion, bildet die Kalibrierung. Zuerst müssen die intrinsischen und extrinsischen Parameter der Kamera und des Projektors bestimmt werden, d.h. die Parameter der Optik und die Position und Orientierung des Gerätes. Die Kombination von beiden ergibt das Verhältnis zwischen Kamera und Projektor. Um das Verhältnis zwischen zwei Kamera-Projektor Paaren zu finden, wird zuerst die Transformation zwischen den beiden Kameras bestimmt. Nun wendet man die gleiche Methode wie oben an. Danach sind die zwei Paare hinsichtlich der Arbeitsvolumen kalibriert. Die Prozedur ist einfacher, wenn die Frustums der beiden Kameras zu einem beträchtlichen Teil überlappen.

Die Entdeckung von Veränderungen in der Szenengeometrie durch Unterscheiden der Kamerabilder ist nicht gewährt, wenn den Displayflächen Texturen fehlen. Dagegen aber können Veränderungen in einer projizierten, zufälligen Textur dargestellt werden. Die zufällige Textur selbst wird vom menschlichen Auge nicht wahrgenommen werden, wie in Abschnitt 3.2 beschrieben. Entdeckte Veränderungen in der Szenengeometrie über eine Zeitperiode können das Resultat von tatsächlichen Veränderungen in der Oberfläche oder Abweichung im kalibrierten System sein.

Der Gebrauch von Kameras ermöglicht es, das System periodisch selbst zu kalibrieren, um die Fehler zu kompensieren, die auf Grund von Umweltfaktoren, wie Temperatur oder Vibrationen im Setup entstehen.

3.1.2 Farbe und Reflectance

Der Projektor wird in zwei Modi gebraucht, Szenenextraktion und -darstellung. Um Farbinformation über die Oberfläche zu erhalten, wird der Projektor als helle Lichtquelle zusammen mit einer synchronisierten Kamera gebraucht.

Momentan beleuchten sie die Szenen mit einem schwarzen Muster, dem ein weisses folgt. Dann beobachten sie das resultierende dunkle und helle Bild von einem bestimmten Blickpunkt, um die Reflectancefunktion pro Pixel zu schätzen. Die Reflectancefunktion dient hauptsächlich als Schwellwert für die Bilder von projizierten, binär-kodierten, strukturierten Lichtmuster, weil sie annehmen, dass sich die Kameraantwort linear zur Intensität verhält. Die Kurven der Kameraantwort können durch Beleuchtung der Szene mit verschiedenen Intensitätslevel geschätzt werden. Um das image-based Modell zu vervollständigen, können die Oberflächen in der Szene von mehreren Blickpunkten aus gesammelt werden, um eine bidirektionale Reflectanceverteilungsfunktion (BRDF) zu schätzen.

Die Kamera wird sowohl für Tiefenextraktion, als auch für Farbextraktion pro Pixel gebraucht. Da die zwei Prozeduren die selbe optische Achse teilen, ist keine Abweichung vorhanden. Entsprechend wird das gleiche Projektor-Kamera Paar für die Projektion von strukturierten Lichtmuster, für Tiefenextraktion und für die Darstellung auf tiefenextrahierte Oberflächen gebraucht. Dieses Verfahren eliminiert Probleme, die auf Grund von Abweichungen oder ungenaue Ausrichtungen entstehen.

3.2 Nicht wahrnehmbares, strukturiertes Licht

Einem gewöhnlichen Beobachter soll nicht mehr als weisses Licht erscheinen, und nicht eine Folge von aufblitzenden binären Mustern. Sie benutzen dabei die Methode von nicht wahrnehmbaren, strukturiertem Licht. Die Muster werden in einer schnellen Serie von weissen Lichtprojektionen versteckt. Abbildung 3 stellt eine Sequenz von Muster auf die Zeit projiziert dar. Ein Ansatz mit binär strukturiertem Licht benützt n Muster, um 2^n projizierte vertikale Balken aufzulösen. Abbildung 3 stellt dar, wie ein gegebenes Muster und das Komplement vom visuellen System so integriert werden, dass die Sequenz als Projektion eines ebenen Feldes oder weissen Lichtes erscheint. Der gleiche Ansatz kann verwendet werden, um nicht wahrnehmbares, strukturiertes Licht zusammen mit Video- oder Graphikbildern zu projizieren. Wobei er die nicht wahrnehmbare, automatische Kalibrierung von bestimmten Projektionsflächen erleichtert. Sie benutzen einen digitalen Lichtprojektor, um das Muster und sein Komplement zu projizieren. Eine synchronisierte Kamera kann das strukturierte Licht messen, in dem sie das Licht während der Musterprojektion integriert. (Beispiele siehe Abbildung 5 und Abbildung 4)

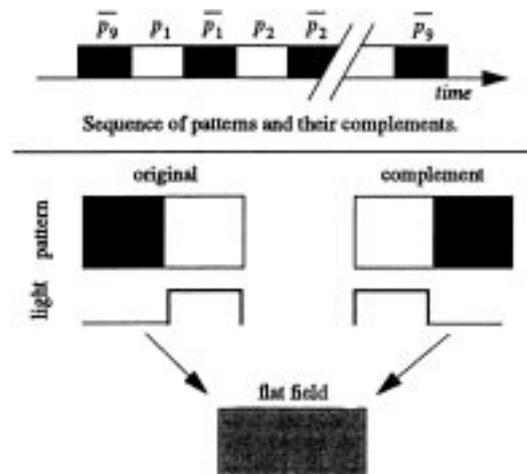


Abbildung 3: Das Muster und sein Komplement werden über die Zeit integriert, das Resultat ist ein ebenes Feld oder »weisses« Licht

3.3 Geometrische Registrierung

Die Tiefe von Daten, die von verschiedenen Kameras anliegen, müssen zusammen gezogen werden, um die Geometrie der Projektionsoberfläche zu vervollständigen. Dies wird benötigt, um korrekte Bilder von einem einzelnen Projektor für Multisurfaces zu erzeugen und für den Übergang von Intensitäten, wenn die Projektionen von mehreren Projektoren überlappen. Die Tiefen von Bildern werden von verschiedenen Kameras genommen und sind in verschiedenen Koordinatensystemen. Um die extrahierten Flächen zusammenzufügen, werden übereinstimmende Punkte von überlappten Regionen benutzt. Die Übereinstimmungspunkte werden generiert, indem man den binär-kodierte, strukturierte Licht Ansatz für Zeilen und Spalten der Projektorpixel benutzt. Der binäre Code von einem dargestellten Punkt identifiziert eindeutig die übereinstimmenden Projektorpixel. Paare von Pixel in zwei Kameras, die den binären Code teilen, werden benutzt, um die Transformation zwischen der Tiefe der Datenmenge zu berechnen. Die Transformation zwischen den zwei Kameras kann auch während der Kalibrierungsphase vorberechnet werden, aber dies ist gewöhnlich zu ungenau, um zwei Mengen der Tiefendaten zu registrieren. Für Übergangszwecke wiederum wird die geometrische Übereinstimmung zwischen den Pixel zweier unterschiedlicher Projektoren durch Beobachtung der Überlappungsregion der Projektion von einer einzelnen Kamera aus hergestellt. Sie nehmen an, dass bei jedem Paar von Bildern eine wesentliche Überlappung auftritt (ungefähr 1/3 seiner Gesamtfläche).

3.4 Rendering und Display

Ein Ziel ist es, Bilder zu generieren, die einem Beobachter korrekt erscheinen, wenn sie auf eine (möglicherweise unregelmässige) Projektionsfläche projiziert werden. Da der Beobachter sich im Büro umher bewegen kann, benutzen sie momentan magnetisches Head-Tracking, um die Position des Betrachters zu bestimmen. Die Eingabeparameter für den Algorithmus sind ein Modell der Oberflächen, die intrinsischen und extrinsischen Parameter der Projektoren, die Position des Betrachters und das »gewünschte Bild«, welches man sehen will. Das gewünschte Bild wird typischerweise das Resultat eines konventionellen 3D-Renderings sein.

Der Algorithmus kann mit jedem herkömmlichen Typ von Oberflächenrepräsentation (z.B. NURBS) arbeiten, so lange wie das Modell der Projektionsfläche der realen Welt genau ist. Ebenso kann das Rendering durch vielen verschiedenen Methoden (z.B. Ray-Tracing) erfolgen. Die momentane Implementation benutzt projektive Texturen mit OpenGL Primitiven, um eine Beschleunigung der Hardware zu erreichen.

Sie beschreiben einen Ansatz für Rendering und Displaying der Bilder von 3D Szenen auf möglicherweise unregelmässige Oberflächen in zwei Schritten.



Abbildung 5: Der Text kann nur mit einer synchronisierten Kamera gesehen werden.



Abbildung 4: a) normales Bild b) binäres Bild, das nicht wahrgenommen werden soll c) beide Bilder kombiniert d) Resultat

Im ersten Schritt wird die vom Betrachter »gewünschte View« durch Rendering der 3D Szene vom Blickpunkt des Betrachters aus berechnet. Die gewünschte View wird als Texture-Map gespeichert. Zuerst wird die Farbe der gewünschten View vom Blickpunkt des Betrachters aus berechnet.

$$I1(r,g,b) = E*[G]$$

Dann wird die Tiefe des Projektionsflächenmodell, auf das der Betrachter schaut, berechnet und zu I1 hinzugefügt.

$$I1(z) = E*[D]$$

Nun muss I1 vom Projektionssystem E des Betrachters zum Projektionssystem P des Projektors transformiert werden. Dies geschieht mit einem 3D Warp mittels projektiver Texturen.

Im zweiten Schritt wird die Textur vom Blickpunkt des Betrachters auf das polygonale Modell der Displayfläche projiziert. Die Projektionsoberfläche (mit der gewünschten Bildtextur darauf abgebildet) wird dann vom Blickpunkt des Projektors gerendert.

$$I2(r,g,b,z) = P*E^{-1}*I1(r,g,b,z)$$

Das resultierende Bild, wenn vom Projektor dargestellt, wird das gewünschte Bild für den Betrachter erzeugen. Wenn sich der Benutzer bewegt, ändert sich das gewünschte Bild und es wird von der neuen Position des Benutzer her projiziert.

$$\begin{aligned} I3(r,g,b) &= E*[Projektionsoberfläche] \\ &= E*P^{-1}*I2(r,g,b,z) \\ &= E*P^{-1}*P*E^{-1}*I1(r,g,b,z) \\ &= \text{gewünschtes Bild} \end{aligned}$$

(E: Projektionsmatrix des Betrachters, P: Projektionsmatrix des Projektors, G: Graphisches Modell, der gewünschten View, D: Projektionsflächenmodell)

Es wird angenommen, dass die Projektoren keine radialen Verzerrungen haben und deswegen könne die Projektoren mit einer Lochkamera modelliert werden. Wenn aber die Projektoren radiale Verzerrungen aufweisen, muss das gerenderte Bild vorverzerrt werden, bevor es zum Framebuffer des Projektor geschickt wird. Diese Vorverzerrung kann erfolgen, indem man einen nicht-linearen 3D Warp der Projektionsflächengeometrie. (Siehe Abbildung 6)

3.4.1 Herausforderungen

3.4.1.1 Geschwindigkeit

Die in zwei Schritten erfolgte Renderingmethode besteht im ersten Schritt aus einem normalen 3D Rendering. Der zweiten Schritt mapped das gewünschte Bild auf ein Projektionsflächenmodell. Die zusätzlichen Kosten des Algo-

rhythmus entstehen durch das Übertragen des Framebuffers vom ersten Schritt in das Texturememory und durch das Rendern des Projektionsflächenmodells mit dem Texturemapping, das im zweiten Schritt angewendet wird. Somit ist es entscheidend die Projektionsflächengeometrie zu vereinfachen.

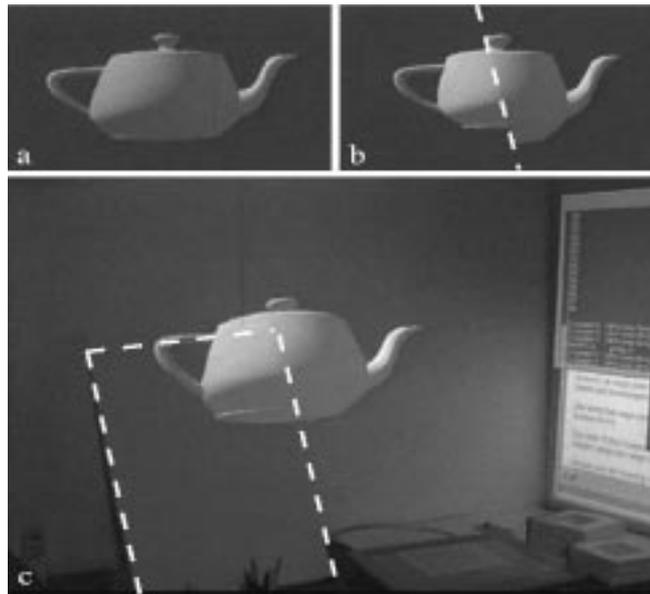


Abbildung 6: a) normaler Teekrug b) verzerrte Projektion (wegen zweiter, ebener Projektionsfläche) c) korrigierte Projektion

3.4.1.2 Parallelisierung

Wenn mehr als ein Projektor gebraucht wird, kann jeder Projektor eine eigene Displayengine besitzen. Das Rendering des zweiten Schrittes kann parallel erfolgen, und jede Displayengine muss nur die Teile des Projektionsflächenmodells laden, die vom Projektor aus sichtbar sind.

Wenn der erste Schritt auch parallelisiert wird, müssen alle die übereinstimmenden Graphikpipelines Zugriff auf das synthetische Modell haben, das gleichzeitig dargestellt wird. Dies kann zu einem Problem werden, wenn sich das Modell dynamisch verändert und die Graphikpipeline das Modell während jeder Darstellungsiteration lesen muss. Andere Parallelisierungsaufgaben schliessen dieses Problem mit ein, wenn das Modell der Projektionsfläche dynamisch ist. Dies gilt auch für Netzwerkaufgaben, wenn die Displayengines sich auf verschiedenen Computer befinden.

3.4.1.3 Latenz

Neben der Tracking- und Renderinglatenz der zwei-Schritt Methode für das Zeichnen von Modellen existiert noch eine zweite. Für grosse Modelle kann die Rendering-Zeit von den verschiedenen Projektoren abweichen, so dass sich eine Verzögerung zwischen den Projektoren während des Rendering ergibt. Wenn alle Projektoren von einer einzigen Maschine betrieben werden, werden das Errichten von Viewports innerhalb eines einzigen Renderingprogramms für jeden Projektor und das synchrone Aktualisieren des Framebuffers dieses Problem beseitigen.

3.5 Erzeugen von Übergangsfunktionen

Sie benutzen »Alpha-Werte«, um die Übergangsfunktion zu beschreiben. Wenn man eine Art von deckenden Multi-projektor SID entwirft, wird man mit zwei Ansätzen konfrontiert, um die Übergänge zwischen den Bildern zu behandeln. Man kann entweder ein System entwerfen, bei dem die Bilder nicht überlappen, aber so angepasst werden können, dass sie sich knapp berühren, d.h. »nahtlos«, oder man lässt zu, dass sich die projizierten Bilder überlappen. Der zweite Ansatz braucht typischerweise Funktionen, um die Übergänge zwischen den Projektoren zu glätten.

Die Entwickler des CAVE™ erprobten die erste Möglichkeit, in dem sie das System auf eine wohldefinierte, relativ einfache Bildschirmanordnung beschränken, bei der keine Projektorüberlappung entsteht. Sie aber wollen Bilder auch auf willkürlichen, möglicherweise unregelmässigen Projektionsflächen im Büro projizieren, d.h. dass sie den ersten Ansatz nicht gebrauchen können, da sie keine Kontrolle über die Oberflächen haben. Ferner stellen sie sich ein flexibles Setup vor, bei dem mehrere Projektoren für die Projektion auf der selben Stelle gebraucht werden können, um eine höhere Auflösung oder erhöhtes Licht zu erreichen.

Sie implementierten eine Gewichtsfunktion, in dem sie jedem Pixel in jedem Projektor Alpha-Werte zwischen 0 und 1 zuweisen, und damit gewährleisten, dass jeder beleuchtete Weltpunkt, der mit einem einzelnen Kamerapixel übereinstimmt, eine Alphasumme gleich 1 hat. Dies setzt voraus, dass die Projektoren eine ähnliche Intensitätsantwort

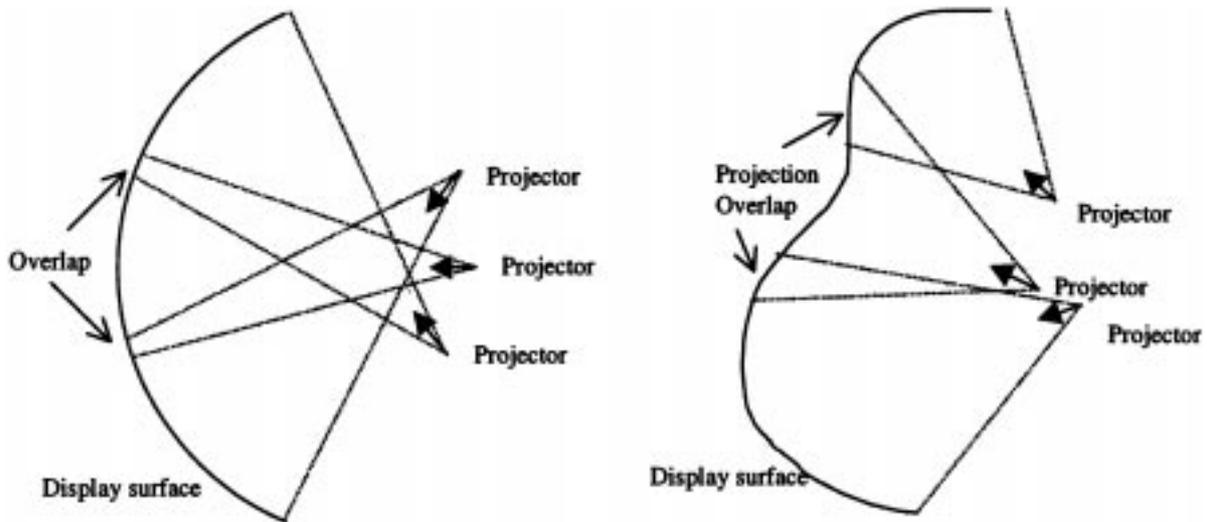


Abbildung 7: Traditionelles Setup

Flexibles Setup

haben. Es gibt dabei zwei Fälle: ein Punkt, der von der Projektion eines einzelnen Projektors stammt und ein Punkt, der von mehreren Projektoren stammt. Im ersten Fall ist die Lösung einfach. Dagegen machen sie im zweiten Fall, in dem die Überlappung auftritt, die Alpha-Werte zu einer Funktion der Distanz vom Anfang bis zum Ende der überlappenden Region. Es gilt dabei die Beschränkung, dass die Alpha-Werte von Punkten verschiedener Projektoren, die mit dem gleichen Punkt im Raum übereinstimmen, eine Summe von 1 aufweisen müssen. Um verschiedenen Gewichte den Projektorpixel zuzuweisen, kreierte sie ein Alphabild für jeden Projektor. Dieses Bild enthält $(1 - \text{gewünschtes_alpha})$ bei jedem Pixel. Dieses Alphabild wird zuletzt gerendert. In ihrer OpenGL Implementation wird dies durch den Gebrauch von transparenten Texturen erreicht. Eine Kamera im geschlossenen Loop-System lässt es zu ein Bild auch dann photometrisch zu korrigieren, wenn die Projektoren verschiedene Helligkeit haben oder wenn die Projektionsfläche nicht einheitliche Reflexionseigenschaften besitzt.

3.6 Vereinfachung der Tiefe von Daten

Aus dem dynamischen image-based Modeling eines ganzen Büros wird eine enorm grossen Datenmengen resultieren. Die Mehrheit der Daten aber wird zeitlich und räumlich sehr stark zueinander in Beziehung stehen. Tatsächlich wird der grösste Teil des Büros in einer kurzen Zeitperiode kaum verändert, dies gilt besonders für die Bürowände und viele andere bestimmte Projektionsflächen. Es ist daher sinnvoll, die Daten zu vereinfachen, damit das System nicht eine solche Datenmenge behandeln muss. Die Vereinfachungsmethode muss sorgfältig sein, um nicht in Regionen Vereinfachungen vorzunehmen, die sich schnell verändern oder starke Krümmung besitzen, da dort Informationen verloren gehen könnten.

Unglücklicherweise erlauben die dynamische Natur und das Rauschen im System nicht die Benutzung der gebräuchlichen Vereinfachungsalgorithmen. Die Methode, die sie momentan anwenden, ist kein genau definierter mathematischer Ansatz, lieferte aber qualitativ angenehme Resultate. Sie wenden zuerst einen Krümmungsansatz auf den Datensatz an, indem sie eine Tangentenmethode [2] gebrauchen. Zu jedem Datenpunkt wird eine Tangentialfläche erzeugt. Man muss dann nur noch das Zentrum und die Normale der Tangentialebene speichern. Das Zentrum und die Normale werden so berechnet, dass die Tangentialebene die beste anliegende Fläche gemäss dem Least Square Algorithmus ist. Diese Tangentialebenen dienen dann als lokale Approximationen der Oberfläche. Dann benutzen sie einen Euklidischen Distanzansatz für die verbleibenden Punkte. Sie wählten diese Reihenfolge der Algorithmen, weil die Krümmungsmethode gewöhnlich erfolgreicher ist im Beseitigen von Punkten als die zweite Methode. Dieser Ansatz produziert Eliminationsraten von 80% bis 90% ohne irgendeinen sichtbaren Informationsverlust. Das kommt daher, da die meisten Objekte in der Büroumgebung lokal planar sind.

3.7 Tracking

Obwohl das Office of the Future sicher und möglicherweise effektiv in einem 2D Modus benutzt werden könnte, wollen sie sich seinen zusätzlichen Gebrauch als 3D Visualisationsumgebung überlegen. Sie brauchen die Fähigkeit, die Köpfe der Betrachter zu verfolgen, um perspektivisch korrekte Bilder zu rendern. Interessanterweise braucht man für monoskopisches Betrachten nicht die Orientierung des Auges, weil das Displaybild eindeutig durch die Position des Auges bestimmt wird. Für stereoskopisches Betrachten muss man fähig sein, entweder ein Auge und die Orientierung des Kopfes, oder beide Augen nur mit ihrer Position zu tracken. Das System involviert das Projizieren von synthetischen Bildern auf reale Oberflächen, für welche man annimmt, dass das extrahierte Oberflächenmodell korrekt im Weltenraum ist. Jeder kleine Fehler beim Trackerreading (nach der Transformation) im Weltenraum



Abbildung 9: a) Kameras und Projektoren sind an der Decke installiert b) Schaumstoffplatten, die als SID benutzt werden

5 Zukünftige Arbeit

Sie planen Akquisition von Szenen zu integrieren und so darzustellen, dass diese Akquisition nicht wahrnehmbar ist, oder wenigstens nicht störend. Dies wird einige Kombinationen von Lichtkontrolle und Kameras einschließen, möglicherweise auch wide-field-of-view high-resolution Cluster (Siehe Abbildung 10 und Abbildung 11). Sie erforschen die Optionen zwischen strikter Kontrolle des ganzen Lichtes in der Umgebung (wie in diesem Artikel angesprochen) und schwächerer oder keiner Kontrolle des Lichtes, bei der sie aber mehrere Kameras verwenden.

Als Teil der Szenenakquisition kann man Veränderungen der Projektionsfläche entdecken und das Rendering entsprechend anpassen. Momentan kann dies mit nicht-interaktiven Raten gemacht werden. Schlussendlich wollen sie auch Methoden erforschen, um Oberflächenänderungen zu entdecken, für Ziele wie Tracking und Gestik.

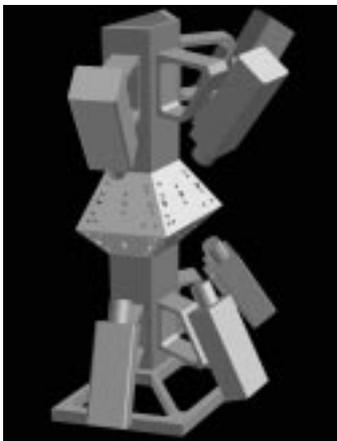


Abbildung 10: WFOV-Kamera



Abbildung 11: Mit WFOV-Kamera aufgenommenes Bild

Sie wollen auch durch Auswerten der image-based Rendering Methoden die Erzeugung von Bildern mit genaueren Übergängen verbessern, um ein Zielbild aus mehreren Referenzbildern zu konstruieren. Zusätzlich wollen sie Unstimmigkeiten in der Oberflächenreflexion dynamisch korrigieren, und die Information während der Laufzeit benutzen, um das gerenderte Bild anzupassen.

6 Zusammenfassung und Schlussfolgerung

Sie haben erste Resultate gezeigt, für ein neues semi-immersiv Display in einer büroähnlichen Umgebung. Dieses verbindet Akquisition und Darstellung. Sie haben Techniken entwickelt, um die Geometrie von unregelmässigen Oberflächen zu aquirieren und dann das Rendering zu modifizieren, um die Projektion auf diese unregelmässigen Oberflächen zu ermöglichen, so dass diese für einen Betrachter korrekt aussehen. Sie haben eine Methode beschrie-

ben, die strukturiertes Licht in eine Szene einführt, das nicht wahrnehmbar für die Teilnehmer, aber messbar für eine synchronisierte Kamera ist. Diese Techniken können für andere Displayumgebungen verwendet werden, welche mehrere Projektoren gebrauchen oder die komplexe Displaygeometrien involvieren.

Ein Gebiet, das nicht von der globalen Entwicklung beeinflusst wurde, ist der Computerbildschirm oder der Bereich, auf welchem die Resultate der Arbeit repräsentiert werden. Ihr System treibt diese Entwicklung voran, welche ermöglicht, dass jedes Objekt, oder eine Menge solcher Objekte, die irgendwo liegen, als Projektionsfläche gebraucht werden kann. Von jetzt an muss man die Informationen nicht mehr auf einen relativ kleinen Bildschirm pressen, sondern man hat so viel Platz wie zur Verfügung steht. Er wird nur durch die Grösse des Raums begrenzt wird. Alles kann eine Projektionsfläche sein – eine Wand oder ein Tisch. Sie kann auch überall sein – sei es in einem Büro oder in einem Konferenzraum. Natürlich birgt das System viele Herausforderungen in sich, aber diese können durch die steigende Geschwindigkeit der Graphikhardware und der Computer überwunden werden.

7 Referenzen

- [1] Depiero, F.W., und Trivedi, M.M., »3D Computer Vision using Structured Light: Design, Calibration and Implementation Issues«, *Advances in Computers*(43), Academic Press, pp.243–278, 1996
- [2] Hoppe, Hugues, Tony DeRose, Tom Duchamp, John McDonald, Werner Stuetzle, »Surface Reconstruction from Unorganized Points«, *SIGGRAPH 92 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison–Wesley*, pp.71–76, 1992
- [3] Ramesh Raskar, Matt Cutts, Greg Welch, Wolfgang Stürzlinger, »Efficient Image Generation for Multiprojector and Multisurface Displays«, University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98–016, 1998
- [4] Ramesh Raskar, Greg Welch, Henry Fuchs, »Spatially Augmented Reality«, University of North Carolina at Chapel Hill, Dept of Computer Science, 1998
- [5] Ramesh Raskar, Greg Welch, Henry Fuchs, »Seamless Projection Overlaps Using Image Warping and Intensity Blending«, University of North Carolina at Chapel Hill, Dept of Computer Science, 1998