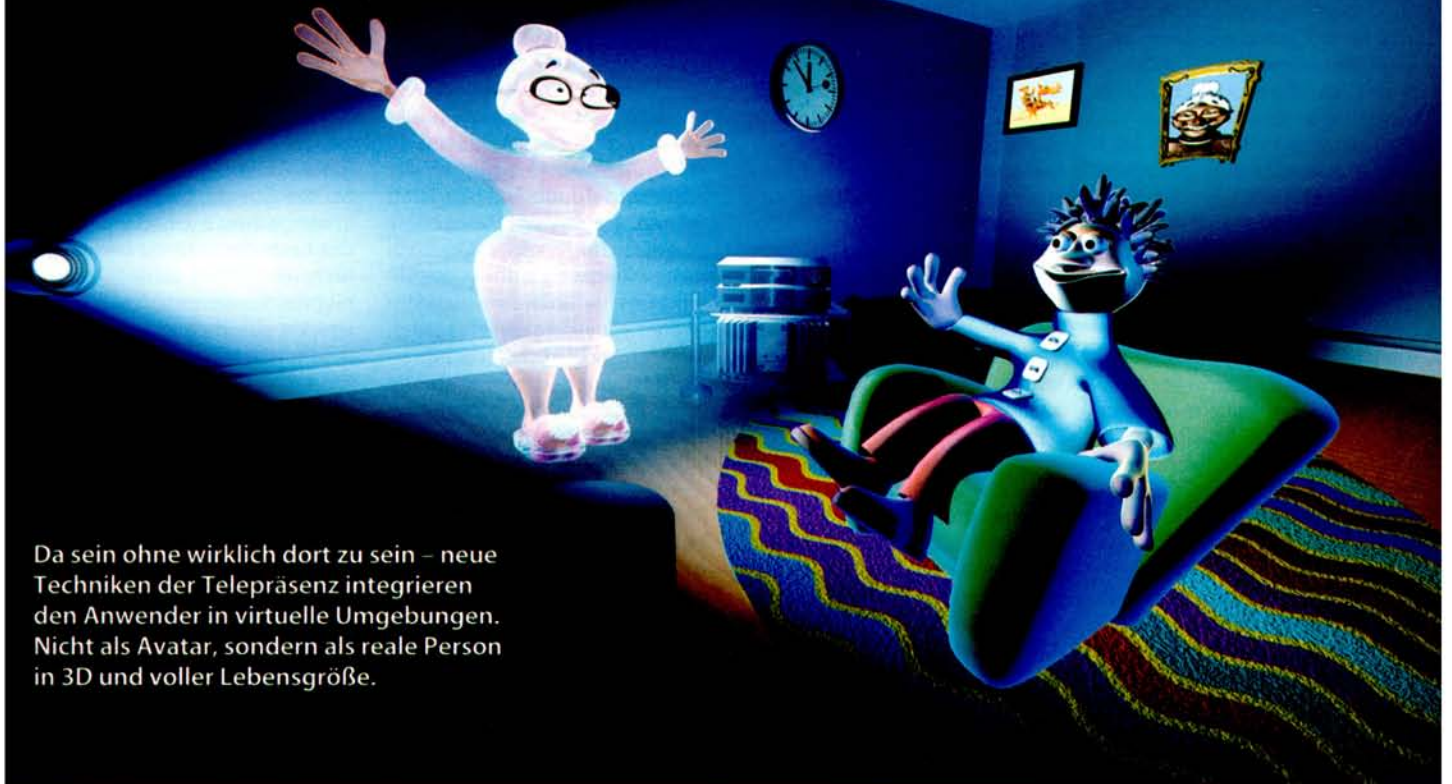


Tom Sperlich

Holodeck light

Auf dem Weg zur 3D-Telepräsenz



Da sein ohne wirklich dort zu sein – neue Techniken der Telepräsenz integrieren den Anwender in virtuelle Umgebungen. Nicht als Avatar, sondern als reale Person in 3D und voller Lebensgröße.

Wer in der Entwicklung an vorderster Front arbeitet, muss gelegentlich eine Erklärung vorausschicken. So auch bei „blue-c“, dem Telepräsenz-Portal [1], das auf dem Kongress IPT-EGVE (Immersion Projection Technology – Eurographics Virtual Environments) 2003 in Zürich erstmals einem kleinen Fachpublikum vorgestellt wurde. Professor Markus Gross, Leiter des Computer Graphics Laboratory (CGL) der ETH Zürich und Leiter der ersten blue-c-Projektphase, warnte sein Publikum: Man solle sich dreidimensionale Videokommunikation so vorstellen „wie damals in den ersten Tagen des Fernsehens, als auch dieses noch in den Kinderschuhen steckte“. Die Experten auf dem internationalen VR-Kongress quittierten es mit verständigem Kopfnicken.

Der so genannte Teleimmersionsraum blue-c (<http://blue-c.ethz.ch>), aufgebaut im Rechenzentrum der ETH Zürich, sieht einer Cave sehr ähnlich. Eine Cave (CAVE Automated Virtual Environment) ist eine meist etwa drei Meter mal drei Meter mal drei Meter große Umgebung zur Erzeugung eines virtuellen Raumeindrucks, die mit Mehrseiten-Stereoprojektion arbeitet.

blue-c besitzt vergleichbare Maße, auf drei Wände wird projiziert, man tritt allein oder mit Partner in den halboffenen Raum. Doch während Caves wie ein Kino nur 3D-Szenen abspielen, funktioniert blue-c wie ein Videokonferenzsystem in voller Lebensgröße und 3D. Bei einem Besuch im Prototypen der Telekommunikationseinrichtung kann man bereits heute mit einem oder zwei entfernten Ge-

genüßern in Kontakt treten. Der Interaktionspartner befindet sich dabei in einem ähnlichen System, in einem weiteren ETH-Gebäude auf dem Höggerberg, am anderen Ende der Stadt – verbunden über Glasfaser. Das System erfasst das 3D-Bild des Users und überträgt es in Echtzeit zur Gegenstelle; gleichzeitig stellt es die in der Gegenstelle agierende Person in Echtzeit stereoskopisch dar.

An beiden Orten sind jeweils 16 FireWire-Kameras mit VGA-Auflösung montiert. Fünf davon sind oben an den vier Ecken des Raumes, eine an der Decke befestigt. Elf Kameras befinden sich außen rings um blue-c beziehungsweise die Gegenstelle. Die Wände von blue-c bestehen aus aktiven Flüssigkristall-Glas-Panels. Sie schalten, wie die sechs Projektoren für die stereoskopi-

sche Rückprojektion, die Brillen sowie eine spezielle Beleuchtung aus 10 000 weißen LEDs im Takt von 60 Hz exakt aufeinander abgestimmt zwischen einem transparenten und einem undurchsichtigen Zustand hin und her.

Nicht wenig Hard- und Software musste eigens von den Wissenschaftlern für das blue-c-Projekt entwickelt oder in bislang einmaligen Kombinationen neu zusammengefügt werden, etwa für die Synchronisation der Projektion oder die automatische Selbstkalibrierung der Kameras. Insgesamt rund 20 ETH-Wissenschaftler waren in den vergangenen drei Jahren in das derzeit größte eigenfinanzierte Polyprojekt der Eidgenössischen Technischen Hochschule involviert.

Die augenfälligste neue Technik im blue-c-Projekt ist die 3D-Darstellung der menschl-

chen Kommunikationspartner. Sie setzt nicht auf gewöhnliches Video oder auf modellbasierte Techniken, die den Gesprächspartner im Brustbildformat darstellen oder einen Video-Avatar erzeugen. Die Zürcher Forscher verwendeten für blue-c eine bisher nur wenig eingesetzte Technik der Computergrafik, die so genannten „Point Samples“, um die sich frei im Raum bewegend Personen darzustellen.

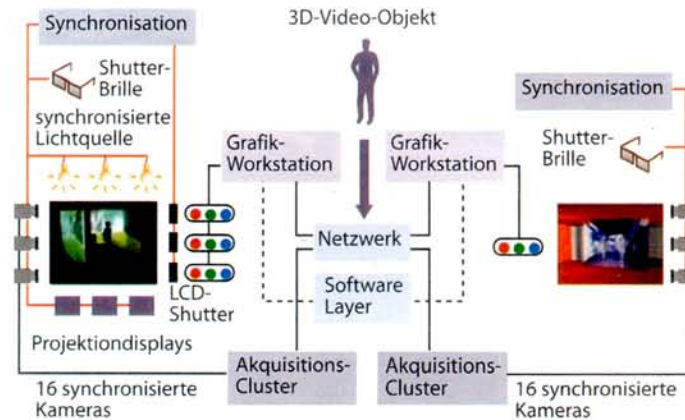
Auf den Punkt

Bereits 1985 wurde von Marc Levoy und Turner Whitted von der Universität North Carolina der Punkt als Primitive zur Darstellung von dreidimensionalen Objekten in der Computergrafik vorgeschlagen [2]. Es dauerte jedoch bis Mitte der Neunzigerjahre, bis einige Forscher die zugrunde liegenden Konzepte auch praktisch umsetzten.

Bis dahin galt die virtuelle Dreiecks-konnektivität als Standard: Man stellt eine 3D-Szenarie dar, indem man zunächst die Punkte eines Gitternetzes angibt, und dann eine Liste der wechselseitigen Verbindungen dieser Punkte nachschiebt, sodass die Flächen der virtuellen Szenarie von Dreiecken ausgefüllt werden. Was für statische Szenarien kein Problem ist, kann bei schnell wechselnden, dynamischen Bildern schwierig werden, wenn sich die Konnektivität grundlegend ändert – etwa wenn ein Loch auftaucht, oder ein völlig neues Objekt eingeführt wird. Dann muss das gesamte Gitternetz neu übertragen und aufgebaut werden – ein Prozess, der Bandbreite und Zeit kostet.

Punktwolken, argumentieren nun die Schweizer, sind viel pflegeleichter, denn man braucht keine Konnektivität zu konstruieren und zu unterhalten. Aber das Rendering von Oberflächen aus Punktwolken stellt Computergrafiker vor neue Probleme: Die 3D-Punkte sind ja diskrete Samples einer Oberfläche. Das Ziel beim Rendering ist, diese Oberfläche wieder kontinuierlich zu rekonstruieren. Gelöst wird das Aufgabe im Wesentlichen, indem man die Punkte als Scheibchen (Splats) darstellt [3].

Die notwendige Größe der Scheibchen kann man leicht aus der Dichte der Samples auf der Oberfläche abschätzen. Mit Hilfe von Mitteln wie beispiels-



Schema der Hard- und Software-Komponenten der blue-c Architektur. Die roten Linien zeigen Synchronisationen beziehungsweise die Verbindung mit der Gegenstelle an.

weise Gaussfiltern werden die Scheibchen, deren Ränder transparent sind, dann auf der Oberfläche überblendet. Dieses Alpha-Blending sorgt dafür, dass sich an den Rändern der Scheibchen Mischfarben ergeben. Mit Hilfe von Antialiasing, das auf dem anisotropischen EWA-Texturfilterung von Paul Heckbert beruht, konnten die Schweizer bereits 2001 beeindruckende Resultate vorweisen.

Um die Punktsamples zu „akquirieren“, rechnet zunächst jeder Kamera-Knoten den statischen Hintergrund in Echtzeit aus den 2D-Video-Bildern seiner ihm zugeordneten Kamera heraus. Dafür wird eine statistische Methode aus der Computer-Bildverarbeitung für den Vergleich zwischen Vordergrund – also den Personen – und Hintergrund eingesetzt, die ermittelt, ob die in einen Vektor einsortierten Farben der Nachbapixel eines Punktes sich mathematisch hinreichend vom Farbvektor des statischen Hintergrundes unterscheiden. Aus den segmentierten Bildern entstehen erste grobe Silhouetten, die anschließend, je nach erwünschter Qualität und Geschwindigkeit, adaptiv verfeinert werden können.

Im nächsten Stadium der 3D-Video-Pipeline wird ein geometrisches Modell, die so genannte visuelle Hülle der anwesenden Person erzeugt. Zur Generierung dieser visuellen Hülle und somit zur Erzeugung der Tiefeninformation für jedes Pixel werden die Silhouetten miteinander geschnitten. Für jede Kameraansicht wird ein Silhouettenkonus erzeugt, der vom Kameraz-

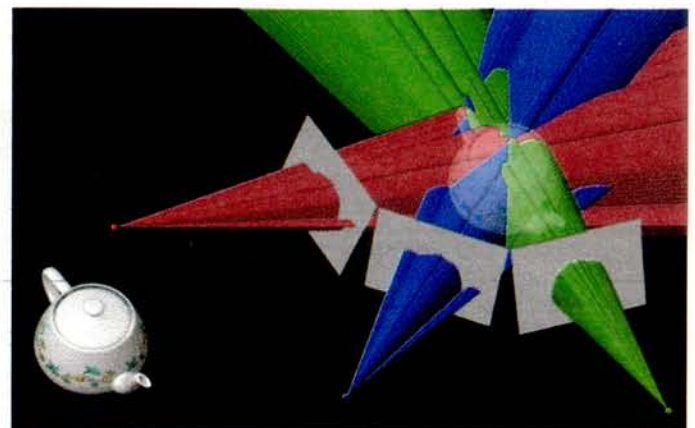
trum durch Projektionen der Silhouettenpunkte im dreidimensionalen Raum entsteht. Das Schnittvolumen der Silhouetten aller Kameras nennt man die visuelle Hülle. Aus diesem Volumen kann dann durch Rückprojektion in die einzelnen Kameras ein Tiefenwert für alle Pixel erzeugt werden. Die visuelle Hülle wird dabei umso realistischer, je mehr Kameras für die Bildung der Schnittmenge zugezogen werden. Demgegenüber erhöht sich aber auch der Berechnungsaufwand, je mehr Kameras involviert sind.

Zur Optimierung dieses Prozesses wählt die Software der blue-c jeweils die relevantesten Kameras zur Schnittberechnung aus. Über die Tiefenkarten kann dann, mittels Projektion der Pixel mit Tiefenwert in den dreidimensionalen Raum, eine punkt-

basierte 3D-Geometrie des akquirierten Objektes erzeugt werden. Dies entspricht einem Punkt-Sampling der visuellen Hülle in den Ursprungspixeln der Kameras. Zum Zusammenführen der Geometrien aus den verschiedenen Kameras ist eine Farbabgleichung zwischen allen Kameras nötig.

Aus dem Input aller Kameras werden die Punkte zu einem „3D-Video-Fragment-Stream“ zusammengefügt und an den Rendering-Knoten der Gegenstelle übertragen. Die Punktdarstellung eignet sich sehr gut für inkrementelle 3D-Video-Streams: Punkte, die neu aufgetaucht sind, werden beim Rendering am Ort des Gegenübers eingefügt, nicht mehr sichtbare Punkte einfach gelöscht. Ein Update-Operator modifiziert Punkte, die immer noch gültig sind, aber veränderte Attribute wie Flächennormale oder Farbe haben. Dank dieser Methode muss nicht in jedem Bild die gesamte 3D-Geometrie neu berechnet werden. Zudem zeigt sich die Methode besonders tolerant bei eventuell aufgetretenen Netzwerkstörungen. Sollten einzelne Datenpakete nicht ankommen, interpoliert die Point-Sample-Rendering-Architektur der ETH.

An Feinheiten arbeiten die Wissenschaftler allerdings noch: So reagiert das virtuelle Gegenüber in blue-c zwar quasi in Echtzeit, wer aber beispielsweise schnell den Arm hebt, dem hängen an der Gegenstelle von den Achseln „virtuelle Fetzen“ herunter, die erst später gelöscht werden. Vor allem bei schnellen und



Visual Hull: Der 3D-Umriss der Teekanne wird ermittelt, indem man das Objekt durch Kegel einschließt, die an verschiedenen Beobachtungspunkten münden. Je mehr Kegel man einfügt, desto vollständiger wird die dreidimensionale Gestalt des Objektes erfasst.

größeren Veränderungen der Bildkontur hat die Punktmethode noch Probleme. Generell ist „vorerst leider keine Spitzenqualität der 3D-Video-bilder zu erwarten“, bedauern die Wissenschaftler.

Stephan Würmlin ist dennoch zufrieden: „Mit anderen computergrafischen Methoden gerenderte Personen sehen eigentlich immer sehr artifizuell aus, wie die bekannten Avatare. Das allerdings hat von meinen 3D-Repräsentationen noch niemand gesagt. Vielleicht, dass sie aussehen wie bei einem schlechten Fernsehempfang, aber nicht, dass sie künstlich aussehen.“

Auch das Thema Netzwerkverbindung ist bei blue-c nicht trivial. Das komplexe „shared & distributed environment“ setzt auf dedizierte Kommunikations-Middleware, die von Edouard Lamoray auf der Basis von CORBA entwickelt wurde. Die einzelnen lokalen Knoten von blue-c sind über ein gewöhnliches Ethernet (100 MBit/s) miteinander vernetzt, für die Verbindung zwischen zwei Sites wird momentan Gigabit-Ethernet benutzt. Benötigt wird derzeit eine Mindestbandbreite von drei bis fünf MBit/s.

An der Gegenstelle angekommen, wird die lokale 3D-Repräsentation von dem 3D-Video-Stream upgedated und von einem lokalen, punktbasierten sogenannten, „elliptischen Splat-



Rendering-Verfahren“ dargestellt. Dies wird von einem eigenen Rendering-Knoten ausgeführt – am ETH-Standort Hänggerberg ist dies ein Linux-PC mit Nvidia QuadroFX Grafikkarte. Der Rendering-Knoten sendet Feedback-Informationen, etwa die gegenwärtige Blickrichtung (viewpoint) zurück an den 3D-Video-Knoten der Gegenstelle. Der 3D-Video-Knoten wiederum kontrolliert entsprechend den Akquisitions-Prozess durch den Kamera-Knoten.

Die wahrgenommene Bewegung kann als gerade noch fließend bezeichnet werden, denn der Kamera-Knoten kann derzeit lediglich circa zehn Bilder pro Sekunde berechnen. Auch eine kurze Latenz ist festzustellen, obwohl für das lokale Rendering im ETH-Rechenzentrum eine Silicon Graphics (SGI) Onyx

3200 mit acht Prozessoren und zwei Grafik-Pipes eingesetzt wird. Im nächsten Schritt wollen die Wissenschaftler diese Aufgabe an einen Linux-PC-Cluster übertragen.

Auch an Kompressionstechniken zur weiteren Reduktion der benötigten Bandbreite wird gearbeitet. Aus diesem Grund haben die ETH-Wissenschaftler an mehreren der vierteljährlichen Meetings des MPEG-Standardisierungsgremiums (ISO/IEC JTC 1/SC 29/WG 11) teilgenommen. Die dynamische Repräsentation von 3D-Video-Objekten mit frei wählbarem Viewpoint ist nämlich bei MPEG-4 ein noch nicht definitiv abgeschlossenes Thema, und die Technik der Punktgrafik findet im MPEG-4-Gremium einige Beachtung. Nur an den bisher standardmäßig benötigten 3-MBit/s-Verbindungen stießen sich die MPEG-Experten, denn für eine normale Nutzung kommt diese Basis-Bandbreite nicht in Frage. Mittelfristig müsste man eher auf 256 kBit/s benötigte Bandbreite herunterskalieren können, forderten sie.

Die ETH-Forscher sind bislang bei 500 – 1000 kBit/s Minimalbandbreite, wenngleich freilich auch die Qualität darunter leidet.

Allgemein sei es aber gerade ein günstiger Zeitpunkt, bei MPEG noch Einfluss zu nehmen, meint Stephan Würmlin: „Für MPEG fängt jetzt die heiße Phase an, in der Standards im Bereich 3D-Video beraten werden.“ Würmlin hofft denn auch, im MPEG-Gremium mit der blue-c-typischen Point-Sampling-Methode eine Alternativ-Repräsentation zu ebenfalls diskutierten oder bereits verwendeten Techniken (basierend auf Meshgrafik und bereits fertigen Modellen) etablieren zu können.

Videokonferenzen mit Augenkontakt

Auch in Deutschland gibt es Entwicklungen für teleimmersive Videokonferenzen, die ebenfalls bereits MPEG-4 einsetzen, wie etwa das PC-basierte Projekt „Immersive Meeting Point“ (im.point) des Berliner Heinrich-Hertz-Institut – jetzt Fraunhofer Institut für Nachrichtentechnik. Der im.point ermöglicht im Unterschied zu bisherigen Videokonferenz-Technologien, Augenkontakt sowie Blickrichtung naturgetreu zu reproduzieren. Das System läuft auf einem handelsüblichen monoskopischen Display. Zur Erzeugung des Augenkontakts werden, erläutert Ralf Tanger vom Fraunhofer HHI, zwei Kameras benötigt, die an der Seite sowie oberhalb des Displays platziert werden. Mit Techniken aus der 3D-Video-signalverarbeitung wird eine Form von Tiefeninformationen, so genannte Disparitäten, berechnet. Mit Hilfe der Disparitäten kann dann über eine View-Synthese, also die Erzeugung einer Zentralperspektive durch Zusammenführen mehrerer Kameraaufnahmen, eine künstliche Kameraansicht erzeugt werden, die einer exakt an der Kopfposition des Gesprächspartners befindlichen Kamera entspricht.

Die Nutzer des Systems sitzen an einem halben, runden Tisch, dessen andere Hälfte im Videobild zu sehen ist, welches auf einem 61-Zoll-Plasmadisplay läuft. Auch bei dieser so ge-



Rund 10 000 weiße LEDs, die mit den Panels und den Projektoren synchronisiert werden müssen, leuchten die Szenerie aus.

blue-c in Aktion: Videokonferenz mit einem in 3D dargestellten Gegenüber



nannten „Shared Virtual Table“-Lösung werden die Personen zunächst vom Hintergrund segmentiert, dann erfolgen Disparitätsschätzung und die Berechnung der virtuellen Kameraansicht an den Kopfpositionen der Gesprächspartner. Über verschiedene MPEG-4-Funktionen wird der Videostream kodiert und übertragen. Die vorab entworfene virtuelle Szene liegt jeweils lokal auf dem PC der Gegenstelle, sodass nur noch die Videos an der entsprechenden Stelle im virtuellen Raum dargestellt werden müssen.

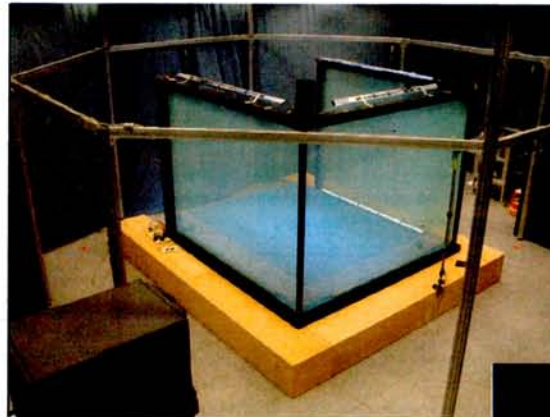
Einen vergleichbaren Ansatz wie das Heinrich-Hertz-Institut entwickelte vor etwa drei Jahren auch das amerikanische Projekt für 3D-Telepräsenz, die so genannte „National Tele-Immersion Initiative“ (NTII) (www.advanced.org/teleimmersion.html). VR-Pionier Jaron Lanier, der einst den Begriff Virtual Reality prägte,

Services in Armonk, New York, wo Laniers Arbeitsbereich war. Obwohl eine parallele Verbindung nur zwischen zwei Orten befriedigend verlief und stets nur der Oberkörper der Teilnehmer als 3D-Videobild zu sehen war, wurde das Projekt als grundlegender Erfolg verkauft. „Eine ultimative Konvergenz zwi-

Bislang wurde der umgebende Raum bei der Daten-Akquisition eigentlich gar nicht in Betracht gezogen, der anfängliche Fokus lag auf der Geometrie der dynamischen Elemente, ihrer Akquisition, Segmentierung und dem Point-Rendering. In der nächsten Phase soll nun auch ein kompletter Raum, mögli-

doch, so die Überlegungen, als dynamische 3D-Point-Sample-Repräsentation in Echtzeit gerendert werden.

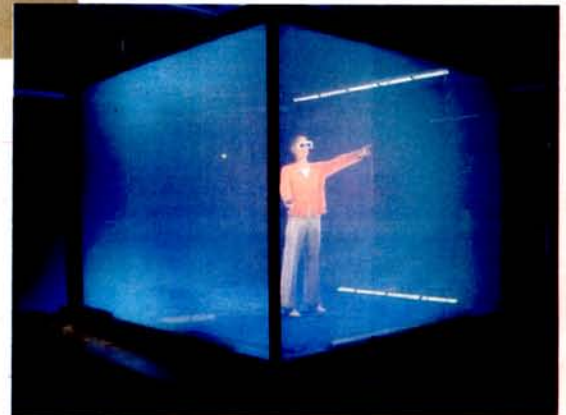
Diese Vorstellungen durchmischter, belebter 2D- und 3D-Environments bringen freilich auch neue Herausforderungen für die Computergrafikmethoden mit sich. Wann blue-c tatsächlich den Kinderschuhen entwachsen und der Bildqualität heutigen Fernsehens vergleichbar sein wird, darauf lassen sich



Die Wände des Teleimmersionsraumes werden von drei überdimensionalen LC-Panels gebildet, die mit 60 Hz zwischen durchsichtig und undurchsichtig hin- und hergeschaltet werden.



Infoticles: Den Einträgen von Datenbanken sind virtuelle physikalische Eigenschaften zugeordnet – durch die Immersion in ein 3D-Environment können Anwender die Muster bestimmter Daten entdecken.



war der leitende Wissenschaftler des Projekts. Auch beim NTII waren allerdings die Bandbreitenbedürfnisse am oberen Ende angesiedelt – kein Wunder, was das mit Spitznamen Tele-Cubicle titulierte Projekt als Demonstrationsobjekt für das US-Langzeitprojekt Internet 2 gedacht. Ohne Kompression wurden die Szenen mit 15 bis 20 MBit/s über das Abilene-Netzwerk – ein 10-Gbit/s-Backbone für die am Internet-2-Projekt beteiligten 220 Universitäten und Unternehmen – hin- und hergeschickt.

Eines der jüngsten publizierten Ergebnisse war eine (Teil-)Verbindung zwischen gleich drei Orten: der University of North Carolina (UNC), der University of Pennsylvania (Penn) und zwischen den Labs des Projekt-sponsors Advanced Network &

schen realer Welt und der Computergrafik“ habe begonnen, hieß es. Heute sei die NTII in einer interessanten Phase, so Jaron Lanier zu c't, denn in praktischen Tests müsse nun herausgefunden werden, wie „durchkonstruiert die Designs sein müssen, bevor sie in der realen Welt etwas in Gang setzen.“

Phase zwei

Bei blue-c hat Anfang 2004 die zweite Phase begonnen, nachdem im Sommer 2003 eine internationale, interdisziplinäre Jury die Güte und Fortschritte des Projekts mit erhobenem Daumen beurteilt hatte. Nun sollen, wie es Markus Gross bereits in der ersten Projektphase vorschwebte, auch größere Räume als bisher abgedeckt werden können.

cherweise gar in Stadiongröße, dargestellt werden können, nebst aller sich darin befindlichen dynamischen Elemente in ihrer Relation zum Raum. Es gehe jetzt weniger um die Immersion in einen Raum, sondern darum, einen Raum frei navigierbar zu machen, etwa in Form von interaktivem Fernsehen, erläutert Kai Strehlke.

Beim Beispiel Stadion könnte dies bedeuten, eine virtuelle Kamera für das Stadionrund mit Tribünen anzubieten, deren Blickpunkt sich interaktiv frei steuern lässt. Während das Stadion bereits als Gittergrafik vormodelliert würde, könnten die Zuschauer auf der Tribüne als eine gängige 2D-Videotextur mit dynamischen Eigenschaften dargestellt werden. Die Akteure auf dem Stadionrasen, die Sportler auf der Aschenbahn sollten je-

die ETH-Forscher natürlich nicht festlegen. Doch Abenteuer im Holodeck erleben konnte ja auch erst die dritte Generation der Enterprise-Besatzung: Archer und Kirk mussten noch ohne auskommen. (wst)

Literatur

- [1] Papiere zu blue-c online unter: http://graphics.ethz.ch/Downloads/Publications/Papers/2003/gro03/p_Gro03.pdf und http://graphics.ethz.ch/Downloads/Publications/Papers/2004/wue04/p_Wue04.pdf
- [2] <http://graphics.stanford.edu/papers/points>, The Use of Points as a Display Primitive; zwei grundlegende Aufsätze
- [3] Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, Markus Gross, Surface Splatting, Proceedings of SIGGRAPH 2001, online unter: http://graphics.ethz.ch/Downloads/Publications/Papers/2001/p_Zwi01a.pdf