Multimodale Interaktion in Virtueller Realität am Beispiel der virtuellen Konstruktion

M. E. Latoschik

Systeme der Virtuellen Realität (VR) stellen neue Herausforderungen an die Art und Weise der Systembedienung. Die Übertragung herkömmlicher 2D orientierter WIMP–Schnittstellen (Windows, Icon, Menue, Pointer) erweist sich hier häufig als ungeeignet. In diesem Zusammanhang verfolgt die vorgestellte Arbeit den Ansatz, VR–Interaktionen über natürliche menschliche Kommunikationsmöglichkeiten mit Gestik und Sprache zu realisieren. Zu diesem Ziel wurden Kernkomponenten entwickelt, welche in den Echtzeit–getriebenen Programmfluss von VR Systemen eingebettet sind und Funktionen für die Gestikerkennung und –analyse sowie die multimodale Integration und Auswertung bereitstellen. Diese Komponenten werden in einer Reihe von aktuellen Projekten etwa im Kontext der Virtuellen Konstruktion, der Deixisforschung oder zur Interaktion mit MAX, einem antropomorphen Gegenüber in virtuellen Umgebungen eingesetzt.

Einleitung

Der Gestikbegriff beschreibt ein breites Spektrum menschlicher Bewegungen mit kommunikativer Relevanz. Neben den deiktischen Gesten sind im koverbalen Gebrauch insbesondere die kinemimischen oder mimetischen und die ikonischen oder spatiographischen Gesten wichtige Informationsträger. Sie können relative Objektpositionen und –lagen sowie Objektbewegungen und –formen kommunizieren und eignen sich insbesondere für VR bezogene Interaktionen wie 3D Objektmanipulationen. Abbildung 1 zeigt einen Ausschnitt aus einer kinemimischen Interaktion. Nach der Auswahl eines Objekts („Nimm (Zeigegeste) dieses Rad") wird es anschließend kontinuierlich manipuliert („Dreh es (Rotation der Hand) so herum") wobei das Ergebnis schritthaltend inspiziert und modifiziert werden kann.

Der Aufbau von komplexen VR Systemen erfolgt i.d.R. mit Hilfe modularer Komponenten etwa für graphische Objektbeschreibungen, zur Einbindung VR–typischer Ein– und Ausgabegeräte, aber auch für die Interaktions– oder die Simulationsmodellierung. Szenengraph–basierte Ansätze stellen eine Möglichkeit bereit, diese Komponenten in einer hierarchischen Struktur zu organisieren und binden die einzelnen Graphobjekte über sogenannte Anwendungs– und Rendertraversierungen in die geschlossene Renderpipeline des VR Systems ein. Zusätzliche Verbindungen (Routen) zwischen einzelnen Objekten auf Attributbasis (Felder) erlauben den bildgenerierungssynchronen Informationsfluss zwischen Objekten. Diese Modellierung liefert den technischen Rahmen für die im Folgenden beschriebenen Module zur Auswertung multimodaler Interaktionen in der VR.

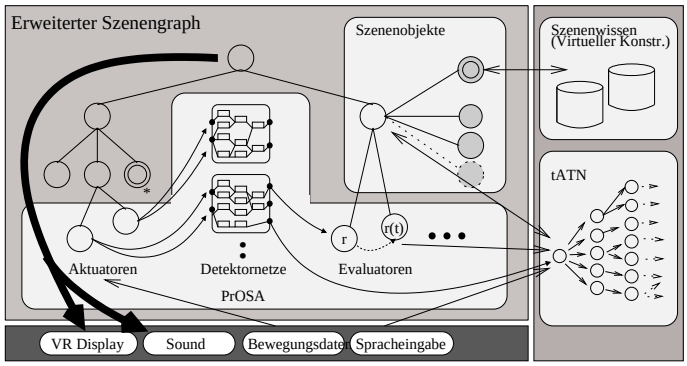

Abbildung 2Verknüpfte Gestenverarbeitungskomponenten

Gestenverarbeitung mit PrOSA

Abbildung 2 illustriert die prinzipielle Struktur einer um multimodale Interaktionen erweiterten VR Anwendung. Das PrOSA–Framework (Patterns on Sequences of Attributes) stellt die Gestenverarbeitungskomponenten zur Verfügung. Grundlage des Datenaustauschs für die Gestenverarbeitung sind Attributsequenzen (AS). Im Gegensatz zu den bisherigen Feldkonzepten kapseln AS eine *beliebige* Anzahl *zeitgestempelter* Attributwerte (je nach Verarbeitungsstufe etwa 3D–Positionen, Fingerbeugungen, etc.). Am Anfang der Gestenverarbeitung stehen **Aktuatoren** als virtuelle Repräsentanten der für die Gestik interessanten Extremitäten, etwa die Position der Hände oder der Fingerspitzen. Aktuatoren routen über AS diese Informationen in **Detektornetze**. Diese bestehen aus einzelnen verbundenen Berechnungsknoten, welche einzelnd betrachtet einfache mathematische Operationen kapseln. Im Verbund führen sie komplexe Berechnungen durch, etwa die Analyse von Handformen oder Trajektorien. Detektornetze routen ihre Ergebnisse entweder

Abbildung 1Kontinuierliche Objekmanipulation mittels kinemimischer Geste

an **Evaluatoren** oder stellen sie direkt als Abfragefunktionen (**Gestenprädikate**) zur Verfügung. Erstere sortieren etwa die umgebene Szene relativ zu bestimmten Körperreferenzsystemen, um diese Information für eine Rückschau bei der Analyse deiktischer Äußerungen in **Spacemaps** bereitzustellen. Letztere werden direkt zur Integration der Gestik mit der Sprache herangezogen.

Multimodale Integration

Die rechte Seite in Abbildung 2 illustriert die Integration der Informationen aus Spracherkennung und Gestenverarbeitung mittels eines modifizierten *Übergangsnetzwerks*, eines **tATNs** (temporal augmented transition network). Ein tATN verfügt an jedem Zustand (State) über die zeitliche Information des zum Erreichen dieses States führenden sprachgestischen Perzepts und erlaubt die gleichzeitige parallele Aktivierung mehrerer Teilzweige. Die einzelnen Constraints an den tATN–Kanten prüfen etwa auf die Erfüllung der beschriebenen, durch PrOSA ermittelten Gestenprädikate, klassifizieren die hereinkommenden Worte, testen ob gewisse Anwendungszustände erreicht sind (ob etwa ein Objekt ausgewählt wurde) und ermöglichen die temporäre Verknüpfung der Constraintfunktionen auf Grund deren Gültigkeitsdauer. So können parallele Eingaben mit Hilfe des tATN Formalismus verarbeitet werden. Die Integration erfolgt durch ein kombiniertes syntaktisch–semantisches Parsing, wobei mit den tATN–Zustandsänderungen Funktionen für die eigentliche Auswertungsfunktionalität verknüpft werden. In Abbildung 3 wird dies an einem Teil–tATN zur Verarbeitung von Objektbeschreibungen verdeutlicht. Der obere Zweig testet allein auf das Vorkommen eines Personalpronomens. Der mittlere und der untere Zweig parsen definite Nominalphrasen ohne und mit begleitender Zeigegeste. Dabei kombiniert der Übergang OD1–>OD41 die Gesten– und

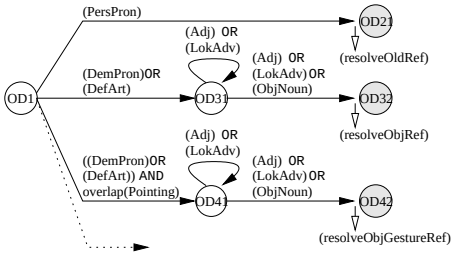

Abbildung 3Einfaches Teil–tTAN zur Verarbeitung von Objektbeschreibungen

Spracheingaben an Hand des temporalen *Overlap* Prädikats. Es vergleicht den Äußerungszeitpunkt des Demonstrativpronomens oder definiten Artikels mit Zeiträumen in welchen eine Zeigegeste vorliegt. Solche Intervalle sind auf PrOSA Seite parametrisierbar (im aktuellen System nach Experimentalergebnissen ein Intervall 250msec vor und 100msec nach der Referenzzeit). Am Ende der jeweiligen Zweige werden Funktionen angestoßen (resolve*), welche mit Hilfe der bereits beschriebenen Spacemaps die Referenzanalyse durchführen.

Interaktionsumsetzung

Neben den Komponenten für die Gestenverarbeitung und die multimodale Integration wurden Konzepte für die Umsetzung der eigentlichen Interaktionen entwickelt. Dies beinhaltet insbesondere die Bereitstellung eng in die Simulationsschleife gekoppelter Module für kontinuierliche Interaktionen wie in Abbildung 1. Die von einem Benutzer distant ausgeführte Gestik ist von ihrer Natur aus nicht in präzise Objektmanipulationen umzusetzen. Wird etwa eine gewünschte Objektrotation mit Hilfe der oberen Extremitäten ausgeführt, so entspricht die Trajektorie der Extremität nur grob der intendierten Objektbewegung. Um hier eine schritthaltende Umsetzung der unpräzisen Bewegungen in präzise Manipulationen zu gewährleisten, filtern sogenannte **Motion–Modifikatoren** die von den Aktuatoren bereitgestellten Bewegungsdaten im Hinblick auf eine „homogene" Trajektorie. Von diesen Motion–Modifikatoren werden die gefilterten Bewegungen an Manipulatoren weitergegeben, die die gewünschte Veränderungen auf Zielobjekte anwenden. Im Kern werden die Motion–Modifikatoren ebenso wie die Gestenerkenner aus PrOSA Berechnungsknoten zusammengeschaltet, da die auszuführenden mathematischen Operationen für beide Aufgaben identisch sind.

Zusammenfassung

In der hier beschriebenen Arbeit wurden Methoden zur Exploration multimodaler Interaktion für VR–Anwendungen entwickelt. Die beschriebenen Komponenten für die Gestenverarbeitung und die multimodale Integration sowie Auswertung zeigen ihre besondere Stärke in der einfachen Konfigurierbarkeit. Alle Module werden über ein Skriptinterface vor oder während der laufenden VR Simulation in den Echtzeit–getriebenen Ablauf eingebracht und über die AS miteinander verknüpft. Etwaige Modifikationen können online erfolgen und inspiziert werden. Die Stärke dieses Ansatzes wird momentan von unterschiedlichen Projekten im Bereich der Mensch–Maschine–Interaktion genutzt. Aktuelle Arbeiten am hier beschriebenen Grundsystem entwickeln neue Detektornetze für ikonische Gesten und erlauben die deklarative tATN Spezifikation über MIML (multimodal integration markup language), einen neu entwickelten XML–Dialekt.

Kontaktadresse:

Dr. Marc Erich Latoschik
AG WBS, Technische Fakultät
Universität Bielefeld
Postfach 10 01 31
D–33501 Bielefeld
Fon: (0521) 106 2919
marcl@techfak.uni–bielefeld.de