Virtuelles Konstruieren mit Gestik und Sprache

Bernhard Jung, Stefan Kopp, Marc Erich Latoschik, Timo Sowa, Ipke Wachsmuth

Im Bielefelder Labor für Künstliche Intelligenz und Virtuelle Realität liegt der Forschungsschwerpunkt auf der Integration von gestischer und sprachlicher Kommunikation in einem Szenario des Virtuellen Konstruierens. Dabei werden hochaufgelöste räumliche Visualisierungen CAD-basierter Bauteilmodelle in realistischer Größe auf einer Großbildleinwand präsentiert und über Eingabegeräte der Virtuellen Realität (Datenhandschuhe, Positionssensoren, Spracherkennungssystem) zu komplexen Aggregaten zusammengebaut. Wissensbasierte Techniken kommen dabei einerseits bei der Montagesimulation mit den computergraphischen Bauteilmodellen und andererseits bei der Auswertung der sprachlich-gestischen Eingaben zum Einsatz.

1 Einleitung

„Virtuelle Realität" bezeichnet ein neuartiges Kommunikationsmedium, das die unmittelbare Wechselwirkung des Menschen mit rechnergenerierten Darstellungen erlaubt. Es beinhaltet wesentlich die Abkehr von üblichen Bildschirm-Displays und damit verbundenen WIMP-Techniken (windows, icons, menus, and pointing devices) in der Mensch-Maschine-Interaktion und ersetzt diese durch betrachterabhängige, dreidimensionale Ein- und Ausgabeverfahren, die neben visuellen Präsentationen auch Schall und taktile Reize einbeziehen [1], [2]. Leitziel der Forschungsarbeiten im Bielefelder Labor für Künstliche Intelligenz und Virtuelle Realität ist die Entwicklung intuitiver Mensch-Maschine-Schnittstellen zu und mit Systemen der virtuellen Realität. Die konkreten Forschungsarbeiten betreffen dabei einerseits die verbesserte Manipulierbarkeit der virtuellen Umgebung durch wissensbasierte Modellierungen sowie andererseits den Einbezug der natürlichen Modalitäten Gestik und Sprache in die Mensch-Maschine-Interaktion.

Um eine intuitive, der menschlichen Erfahrungswelt entsprechende Manipulierbarkeit virtueller Umgebungen zu gewährleisten, ist in vielen Fällen die Rekonstruktion bestimmter physikalischer Objekteigenschaften notwendig. Soll zum Beispiel in einem Innenarchitektur-Szenario die Position eines an der Wand hängenden Regals verändert werden, dann sollten die Freiheitsgrade bei der Objektbewegung derart eingeschränkt werden, daß während der gesamten Interaktion - zumindest jedoch als Endergebnis - Kontakt zwischen Regal und Wand besteht. Im von uns betrachteten Szenario des *Virtuellen Konstruierens* geht es um die Manipulation CAD-basierter Bauteilmodelle, die in der virtuellen Realität zu komplexen Aggregaten zusammengesetzt werden. Dazu werden die Geometriemodelle um wissensbasierte Modellierungen ihrer Verbindungsstellen angereichert und Freiheitsgrade bei Objektmanipulationen so eingeschränkt, daß ein Verbinden der virtuellen Bauteile weitgehend in Übereinstimmung mit realen Montagen vollzogen werden kann. Anwendungspotential dieser Arbeiten liegt in der komfortablen Erstellung sog. *virtueller Prototypen*, die wichtige Eigenschaften eines Produkts abbilden und die damit - ohne Materialverbrauch - eine Überprüfung von Entwürfen in frühen Phasen des Entwicklungszyklus ermöglichen [6].

Als Erprobungsdomäne werden die CAD-basierten Bauteilmodelle in realistischer Größe auf einer Großbildleinwand präsentiert und über Eingabegeräte der Virtuellen Realität (Datenhandschuhe, Positionssensoren, Spracherkennungssystem) zusammengesetzt (siehe Abb. 1). Im Gegensatz zu Ansätzen wie [1] und [4] werden dabei als Fokus unserer Arbeiten nicht zeichensprachliche Gesten, sondern natürliche, koverbale Gesten untersucht.

Abb. 1: Virtuelles Konstruieren mit Gestik und Sprache an einer interaktiven Wand.

Abb. 2: In der Desktop-Version des Virtuellen Konstrukteurs können dreidimensional visualisierte Bauteile u.a. mittels sprachlicher Anweisungen zusammengebaut werden.

Die nachstehend im Überblick beschriebenen Arbeiten bauen auf einem wissensbasierten Simulationssystem auf, dem *Virtuellen Konstrukteur,* mit dem die interaktive Montage komplexer Aggregate aus CAD-basierten Grundbausteinen auf einer bildschirmpräsentierten virtuellen Werkbank möglich ist [6]. In Abschnitt 2 wird ein allgemeiner Ansatz zur wissensbasierten Montagesimulation vorgestellt, der den Zusammenbau verschiedenartiger Bauteiltypen in der virtuellen Umgebung ermöglicht. Benutzereingaben erfolgen in der Desktop-Version des Virtuellen Konstrukteurs durch direkte, mausbasierte Manipulationen bzw. mittels einfacher sprachlicher Instruktionen (Abb. 2). Die in Abschnitt 3 dargestellten Arbeiten zielen auf eine gestische und sprachliche Steuerung solcher Montagesimulationen an einer interaktiven Wand mit 2.5m x 3m Großprojektion ab.

2 Virtuelles Konstruieren

Im Projekt CODY (DFG, SFB 360) wurde mit dem Virtuellen *Konstrukteur* ein wissensbasiertes System entwickelt, das eine interaktive Montagesimulation mit dreidimensional modellierten Grundbausteinen ermöglicht. Die Zielvorstellung ist, daß alle im Realen aus einer gegebenen Menge von Grundbauteilen physikalisch konstruierbaren Aggregate auch in der virtuellen Umgebung herstellbar sind. Neben den standardmäßig von interaktiven Graphiksystemen bereitgestellten Interaktionsmöglichkeiten wie Bewegen von Objekten und Navigation zu beliebigen Betrachterstandpunkten sind im Virtuellen Konstrukteur insbesondere folgende Montage-bezogene Manipulationen in Echtzeit simulierbar:

- paßgenaues Fügen von Bauteilen und Aggregaten
- Trennen von Bauteilen und Aggregaten
- Modifikation entstandener Aggregate durch Relativbewegung (Rotation und Translation) von Bestandteilen gemäß Verbindungsart-spezifischen Freiheitsgraden.

Grundlage der Montagesimulation ist ein allgemeiner, wissensbasierter Ansatz zur Modellierung der Bauteilverbindungsmöglichkeiten, der im folgenden näher beschrieben wird.

2.1 Verbindungsports und Verbindungsarten

Die Simulation von Montagehandlungen beim Virtuellen Konstruieren beruht auf der expliziten Modellierung verschiedener Arten von Verbindungsstellen (*Ports*) der virtuellen Bauteile, z.B. dem Schaft einer Schraube oder dem Gewinde einer Mutter, sowie verschiedener Typen von Verbindungsbeziehungen zwischen den Verbindungsports, z.B. Schrauben oder Stekken. Auf Grundlage verschiedener Baukastensysteme, die von Baufix-Holzbauteilen bis hin zu industriellen, CAD-basierten Grundbausteinen reichen, wurde dazu eine Wissensbasis mit taxonomisch angeordneten Porttypen entwickelt, die sich bezüglich ihrer mechanischen und geometrischen Eigenschaften sowie der daraus induzierten Freiheitsgrade bei Portverbindungen unterscheiden. Top-level Konzepte der Wissensbasis sind *ExtrusionPort* (spezialisiert in *GeberPort* und *NehmerPort*), mit dem z.B. Schraubverbindungen modelliert werden, *PlanePort*, mit dem Objektverbindungen an koplanaren Flächen modelliert werden, sowie PointPort, mit dem Punkt-artige Verbindungen ohne translatorische Freiheitsgrade modelliert werden. Neben den Verbindungsports der Bauteile werden im Virtuellen Konstrukteur verschiedene Verbindungsarten modelliert, wie sie auch in der Konstruktionslehre z.B. nach physikalischen Schlußkräften oder verbleibenden Freiheitsgraden unterschieden werden [7], [12]. So kann in der Montagesimulation z.B. weiter zwischen Stecken (Translation und Rotation ungekoppelt), Schrauben (Translation und Rotation gekoppelt) oder Schweißen (keine Freiheitsgrade) unterschieden werden. Die verschiedenen Verbindungen werden dabei als Relationen zwischen Verbindungsports aufgefaßt und ebenfalls in einer Hierarchie angeordnet. Abbildung 4 zeigt die bisher modellierte, anwendungsspezifisch erweiterbare Taxonomie von Verbindungsports.

Montagerelevante Eigenschaften der virtuellen Bauteile werden mit Hilfe der oben eingeführten Beschreibungsmittel in einer Wissensbasis modelliert. Für jedes Bauteil werden Typ, Kapazität, Geometrie und relative Lage der Verbindungsports angegeben. Durch den Typ des Verbindungsports sind dabei automatisch die möglichen Verbindungsarten festgelegt, die entlang der Porttyp-Taxonomie (siehe Abb. 4) vererbt werden. Sog. „Hotspots" definieren präferierte Objektpositionierungen

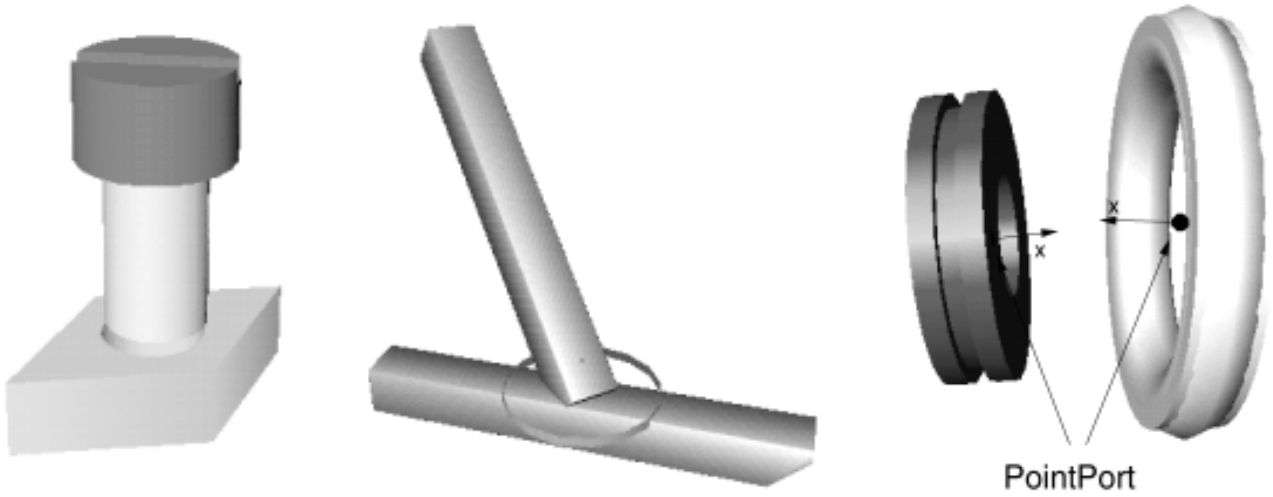


Abb. 3: Verbindungen in der virtuellen Montage beruhen auf verschiedenen Arten von Verbindungsports der virtuellen Bauteile: ExtrusionPorts (links), PlanePorts (Mitte) und PointPorts (rechts). Bei derartigen Verbindungen bestehen jeweils ein Freiheitsgrad bzgl. Rotation sowie bis zu zwei Freiheitsgrade bzgl. Translation.

bei der Simulation von Fügeschritten. Dadurch kann z.B. bei Verbindungen einer Schraube mit einer Mutter unterschieden werden, ob als Standardsystemverhalten die Schraube vollständig oder nur bis zu einem vordefinierten Ansatzpunkt in die Mutter geführt werden soll (nachfolgende Modifikationen sind in jedem Fall möglich). Die explizite Modellierung einer Schraube, die als einzigen Port einen Schaft besitzt, ist im folgenden aufgeführt.

```
part: L_SCH_SCHRAUBE
((type:        Schraubschaft)
(name:         schaft)
(position:     (0,-16,0))
(orientation:  (0,0,90))
(connect.:     stecken, schrauben)
(capacity:     (16,-16))
(hotspots:     (((position: (16,0,0)) (orientation: (0,1,0)))))
(geometry:     ((type:     cylinder)
               (radius:    7))))
```

2.2 Erweiterte Freiheitsmatrizen

Zur Spezifikation von Port- und Verbindungsart-bedingten Bewegungseinschränkungen verbundener Objekte wurde das Beschreibungsmittel der *erweiterten Freiheitsmatrizen* entwickelt [8]. Diese basieren auf den logischen Freiheitsmatrizen, die in der Konstruktionslehre als digitale Darstellungen der Berührungen zweier Körper in diskreten Punkten definiert sind [12]. Während die logische 3x4 Freiheitsmatrix zu diesem Zweck für jeden der 12 Freiheitssinne[1] einen binären Eintrag enthält („0" für Bewegungssperrung, „1" sonst), ist es für die virtuelle Simulation eines realistischen Konstruktionsprozesses notwendig, den Bewegungsfreiraum eines Objektes quantitativ, mittels numerischer Werte zu beschreiben. Dies wird durch den Übergang zu erweiterten Freiheitsmatrizen erreicht, die pro Freiheitssinn eine beliebige positive Zahl oder die Bezeichner BLOCKED (für vollständige Bewegungssperrung) und FREE (für unbeschränkte Bewegungsfreiheit) enthalten können. Um die Bewegungscharakteristika verschiedenster Objektverbindungen adäquat beschreiben zu können, z.B. die Gewindesteigungen bei Schraubverbindungen, lassen sich zusätzlich Kopplungen zwischen je zwei Freiheitssinnen angeben. Eine Kopplung ist dabei durch die Indizes der beiden Freiheitssinne in der Matrix und einer numerischen Angabe des Kopplungsverhältnisses gegeben. Eine erweiterte Freiheitsmatrix ist also ein 2-Tupel aus einer 3x4-Matrix und einer Kopplungsliste (siehe auch Abb. 6). Erweiterte Freiheitsmatrizen sind für jeden Porttyp, z.B. Schraubschaft, und für jede Verbindungsart, z.B. stecken, modelliert und definieren somit typische Bewegungsmerkmale möglicher Objektverbindungen. Des weiteren werden die erweiterten Freiheitsmatrizen auch, wie unten dargestellt, zur Beschreibung der verbleibenden möglichen Relativbewegungen von Bauteilen verwendet, die in der virtuellen Umgebung verbunden wurden.

2.3 Verbindbarkeit von Bauteilen

Bei der Simulation von Fügehandlungen in der virtuellen Umgebung wird das Wissen über die Verbindungsports und die Verbindungsarten zur Herstellung und zur weiteren Modifikation von Objektverbindungen ausgenutzt. Zur Überprüfung der prinzipiellen Verbindbarkeit zweier Bauteile werden dabei für jeweils zwei Ports der zu verbindenden Objekte folgende Bedingungen getestet:

1. Gleichartigkeit: Beide Port-Objekte müssen Instanzen desselben generischen Top-level Konzepts sein, d.h. beide sind entweder *PlanePorts*, *PointPorts* oder *ExtrusionPorts*. Im Fall von ExtrusionPorts muß ein Port ein *GeberPort*, der andere ein *NehmerPort* sein. Es kann somit z.B. keine Verbindungsfläche (*PlanePort*) mit einem Loch (*ExtrusionPort*) verbunden werden.
2. Geeignete Positionierung: Beide Ports müssen korrekt positioniert und ausgerichtet sein. Diese Bedingung betrifft die Port-Koordinatensysteme und ist symmetrisch. Im einzelnen ist diese Bedingung für die verschiedenen Port-Typen wie folgt definiert:
   - Zwei ExtrusionPorts P_1 und P_2 mit den Portachsen $\vec{a_1}$ bzw $\vec{a_2}$ sind korrekt zueinander positioniert, wenn die Portachsen kollinear zueinander sind: $\vec{a_1} \| \vec{a_2}$ und $\vec{a_1} \| \vec{p_2^1}$, wobei $\vec{p_2^1}$ die Position von P_2 bzgl. P_1 ist.
   - Für zwei PlanePorts P_1 und P_2 müssen die Portachsen $\vec{a_1}$ bzw. $\vec{a_2}$, die als die Normalen der Portebenen definiert sind, antiparallel und die Portebenen darüber hinaus koplanar sein, d.h. $\vec{a_1} \uparrow\downarrow \vec{a_2}$ und $\vec{a_1} \perp \vec{p_2^1}$.
   - Für zwei PointPorts P1 und P2 müssen die Portachsen $\vec{a_1}$ bzw. $\vec{a_2}$ parallel sein und die Positionen der beiden punktförmigen Ports übereinstimmen: $\vec{a_1} \| \vec{a_2}$ und $\vec{p_2^1} = \vec{0}$.
3. Ausreichende Kapazitäten: Die Kapazitäten der beiden Ports müssen sich in legaler Weise überschneiden und jeweils genügend Freiraum für den Überlappungsbereich bieten. Das heißt, bei keiner der beiden Portkapazitäten darf ein bereits belegter Bereich geschnitten werden.
4. Passende Geometrien: Die beiden Ports müssen eine gleichartige Form aufweisen, z.B. kann ein zylinderförmiger Schaft nur in zylinderförmige Bohrungen gesteckt werden. Darüber hinaus dienen weitere geometrische Eigenschaften, wie z.B. der Durchmesser bei Zylindergeometrien, zur Zurückweisung von Verbindungen, in denen z.B. eine vergleichweise dünne Schraube in eine Bohrung großen Durchmessers geführt werden soll.
5. Kompatible Verbindbarkeiten: Für jede Portart sind mögliche Verbindungsarten durch eine Liste von formalen Namen, die zu entsprechenden Verbindungsrelationen korrespondieren, spezifiziert. Die Listen der beiden Ports müssen mindestens einen formalen Verbindungsnamen gemein haben, der den Typ der zu erstellenden Verbindungsrelation festlegt. Die Verbindbarkeiten eines Ports werden typischerweise über dessen Einordnung in die Port-Taxonomie per Vererbung inferiert, so daß diese i.a. nicht in der Portbeschreibung des Bauteils modelliert werden müssen.

Zusammengefaßt bedeutet dies: Damit zwei beliebige Ports miteinander verbindbar sind, müssen ihre Typen kompatibel sein, sie müssen geeignet positioniert sein, ihre Kapazitäten müssen eine Verbindung zulassen, ihre Geometrien müssen, falls spezifiziert, von gleicher Form sein, und sie müssen die gleiche Art von Verbindungbeziehung eingehen können. Da die Bedingungen für das paarweise Verbinden von Objekten im Rahmen einer allgemeinen Taxonomie von Verbindungsports formuliert sind, decken sie eine große Klasse von Bauteilen ab,

[1] Ein *Freiheitssinn* ist ein gerichteter Freiheitsgrad. Es gibt also zu jedem Freiheitsgrad genau zwei entgegengesetzt gerichtete Freiheitssinne.

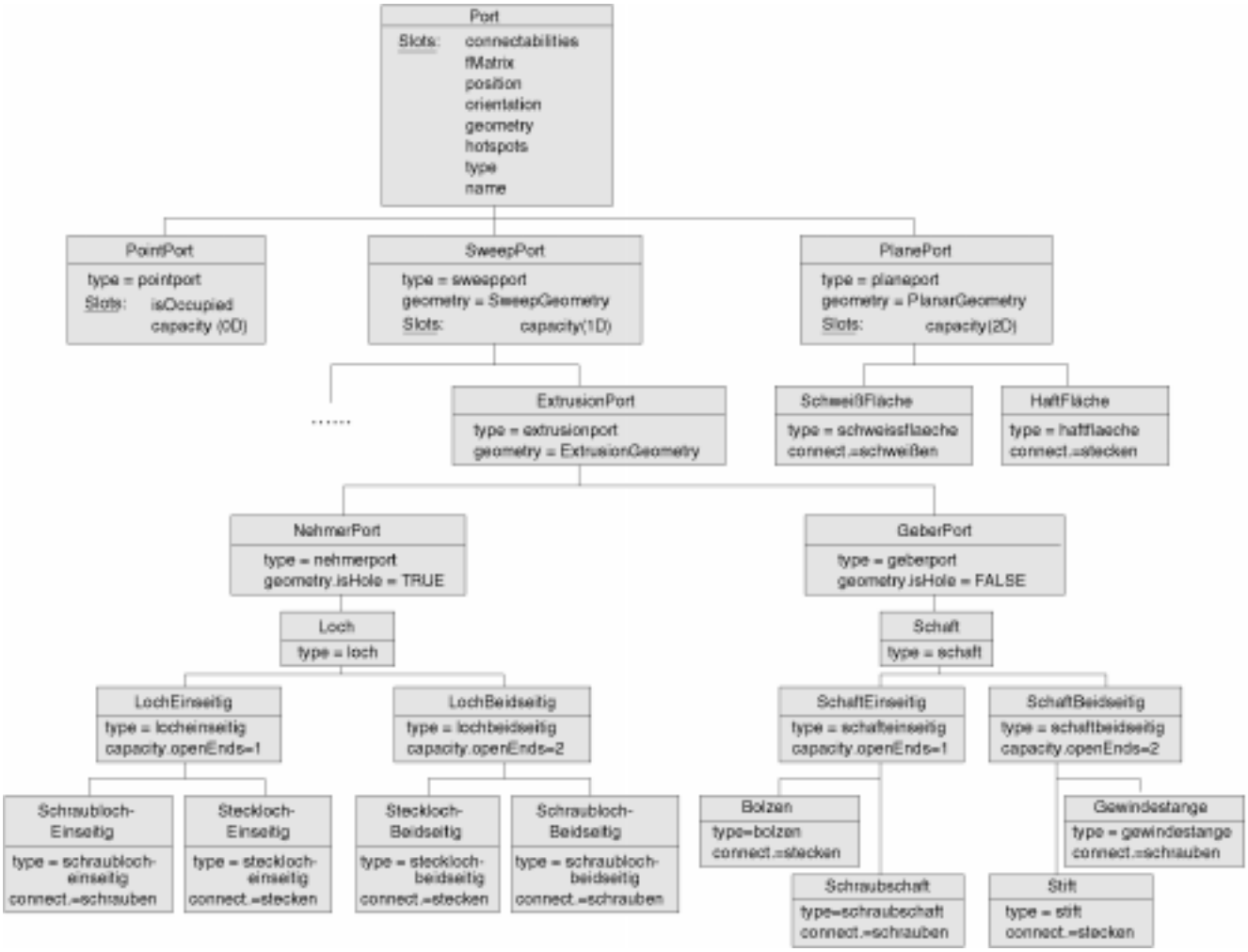


Abb.4: Taxonomie der Verbindungsports.

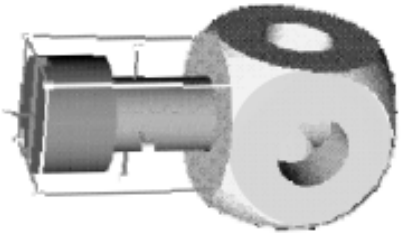
wie z.B. die Grundbausteine eines Kleinfahrzeugs (Abb. 1) oder die Baufix-Teile des SFB 360-Szenarios (Abb. 2). Insbesondere stellen die genannten Bedingungen eine Verallgemeinerung der im Kontext der robotischen Montageplanung formulierten Kontaktbedingungen nach [10] dar.

2.4 Erzeugung und Modifikation von Verbindungen in der virtuellen Montage

Der Virtuelle Konstrukteur unterstützt die 3D-Simulation verschiedener montagebezogener Operationen wie Fügen und Trennen von Bauteilen sowie Korrekturbewegungen von Bestandteilen schon zusammengebauter Teilaggregate. Dabei wird einerseits das oben beschriebene Modellwissen über die Verbindungsmöglichkeiten der Bauteile ausgenutzt, indem die Lage der virtuellen Bauteile bei der Montagesimulation derart verändert wird, daß die oben dargestellten Bedingungen für die Bauteilverbindungen in der Zielposition des transformierten Objekts erfüllt sind. Im Gesamtablauf der Montagesimulation werden die Port-bezogenen Bedingungen für das paarweise Verbinden von Bauteilen andererseits durch die globale Anforderung nach Kollisionsfreiheit zwischen allen virtuellen Bauteilen ergänzt. Da die Bedingungen für das paarweise Verbinden von Bauteilen zwar Kollisionen zwischen den beiden beteiligten Objekten, nicht aber mit dritten Objekten ausschließen, werden initiale Lagevorschläge auf globale Kollisionsfreiheit überprüft, sowie gegebenfalls durch Kollisionsvermeidungsstrategien korrigiert, die eingegangene Verbindungen entlang portspezifischer Bewegungsmöglichkeiten modifizieren (siehe Abb. 5). D.h. die Simulation von Montagehandlungen ist ein multimodaler Schlußfolgerungsprozeß, der sowohl konzeptuelles Modellwissen wie auch exaktes geometrisches Wissen über die virtuellen Bauteile ausnutzt.

Nach dem Verbinden zweier Bauteile in der virtuellen Umgebung wird eine erweiterte Freiheitsmatrix ermittelt, welche die genauen Grenzen beschreibt, in denen Relativ-Bewegungen beider Verbindungspartner erlaubt sind. Die verbleibenden Bewegungsmöglichkeiten verbundener Objekte hängen dabei von der Art ihrer Verbindung, der Art der verbundenen Ports und deren Belegungszuständen ab. Die gesuchte erweiterte Freiheitsmatrix wird also aus den Freiheitsmatrizen der Verbindungsrelation und Port-Objekte sowie deren aktuellen Kapazitäten ermittelt. Zu diesem Zweck sind formale Operatoren definiert, die in Anlehnung an die ursprünglichen logischen Freiheitsmatrizen [12] disjunktive und konjunktive Verknüpfungen zwischen zwei erweiterten Freiheitsmatrizen bilden [8]. Beispielsweise werden zwei Freiheitsmatrizen konjunktiv verknüpft, indem die Schnittmenge der Bewegungsfreiräume gebildet wird, d.h. jeder Eintrag der Ergebnismatrix ist gleich dem kleineren der beiden Einträge der Ausgangsmatrizen; die Kopplungslisten beider Matrizen werden vereinigt. Die

Abb.5: Beim Verbinden virtueller Bauteile wird zunächst auf Grundlage der aktuell verfügbaren Portkapazitäten ein Lagevorschlag gemacht, der in Bezug auf die beiden zu verbindenden Objekte, im Beispiel die obere Schraube und die untere Leiste, zwar lokal kollisionsfrei ist, wobei jedoch noch Kollisionen mit weiteren beteiligten Objekten auftreten können (Bild Mitte). Kollisionsvermeidungsstrategien, wie das schrittweise Herausziehen in entgegengesetzter Verbindungsrichtung, garantieren einen kollisionsfreien Endzustand (Bild rechts).



Erweiterte Freiheitsmatrix Schraube-1:

$$\begin{pmatrix} 3.3696 & 18.6304 & \text{FREE} & \text{FREE} \\ \text{BLOCKED} & \text{BLOCKED} & \text{BLOCKED} & \text{BLOCKED} \\ \text{BLOCKED} & \text{BLOCKED} & \text{BLOCKED} & \text{BLOCKED} \end{pmatrix}$$

[(0,0,2,0, 0.8),(1,0,3,0, 0.8)]

Abb.6: Beschreibung von Relativbewegungen durch erweiterte Freiheitsmatrizen: Die Schraube (Gewinde nicht dargestellt) ist 3.4 Einheiten in den Würfel gedreht und kann um weitere 18.6 Einheiten eingeführt werden. Eine Rotation um die durch den Schaft gegebene Achse ist in beiden Drehsinnen möglich. Alle anderen Relativbewegungen sind ausgeschlossen. Beim Hinein- und Herausschrauben besteht im gezeigten Beispiel ein Kopplungsverhältnis von 0.8 zwischen Translation und Rotation, was die Steigung des Gewindes modelliert.

konkrekte Freiheitsmatrix für einen zu manipulierenden Port P_m, der mit einem anderen raumfesten Port P_f verbunden ist, berechnet sich durch konjunktive Verknüpfung der zugehörigen erweiterten Freiheitsmatrix der Verbindungsart mit den beiden – in ein gemeinsames Bezugssystem transformierten – Freiheitsmatrizen der Port-Objekte, die jeweils die durch ihre Porttypen zugelassenen Freiheitssinne gemäß der aktuellen Kapazitäten begrenzen: $M_f = M_{Verbindung} \wedge M_{P_m} \wedge M_{P_f}$. M_f spezifiziert somit die erlaubten Relativ-Bewegungen, die dem manipulierten Port in seiner direkten Verbindung zu dem festen Port verbleiben. Rechnet man die einzelnen Einträge in Objektkoordinaten um, so kann man die Benutzermanipulationen des Szenenobjekts, bei ortsfestem Verbindungsgegenstück, unmittelbar einschränken. Zusätzlich lassen sich Bewegungen simulieren, die durch bestehende Kopplungen erzwungen werden. Abbildung 6 zeigt die erweiterte Freiheitsmatrix, welche die Randbedingungen für das weitere Hinein- oder Herausdrehen einer Schraube beschreibt, die mit einem Gewindeblock durch eine Schraubverbindung verbunden ist. Indem die erweiterten Freiheitsmatrizen die relativen Bewegungsmöglichkeiten paarweise verbundener Objekte beschreiben, stellen sie die Grundlage für die weitere Modifikation der entstandenen Aggregate in nachfolgenden Interaktionschritten dar. Zusammengefaßt bieten Portverbindungen mit erweiterten Freiheitsmatrizen einen allgemeinen Ansatz zur Beschreibung von Objektverbindungen, wie er z.B. in [5] gefordert wurde.

3 Interaktion mit Gestik und Sprache

Bei der Steuerung der Montagesimulation vor einer Großbildleinwand kommen Techniken der sprachlich-gestischen Interaktion zum Einsatz. Dabei werden drei Gestentypen ausgewertet, die signifikant zur Bestimmung der auszuführenden Anweisung beitragen. *Deiktische Gesten* („nimm <Zeigegeste> dieses Teil") spezifizieren ein Objekt oder einen Ort der virtuellen Umgebung, *mimetische Gesten* („drehe es <kreisender Zeigefinger> so herum") qualifizieren die Ausführung einer Aktion, und *ikonische Gesten* („das so <Andeutung eines Zylinders durch die Handform> geformte Objekt...") werden zur Eigenschaftsbeschreibung von Objekten verwendet. Alle Gestentypen sind koverbal, d.h. sie werden immer im Zusammenhang mit einer sprachlichen Äußerung vollführt.

Die signaltechnische Erfassung der Körperbewegung erfolgt mittels elektromagnetischer Sensoren (Flock of Birds) sowie zweier Datenhandschuhe (CyberGlove).[2] Ein Datenserver empfängt die sensorischen Werte und stellt sie weiteren Datenverarbeitungsmodulenüber eine Netzwerkverbindung zur Verfügung. Die verwendete Spracherkennung, ein an der Universität Bielefeld entwickelter Forschungsprototyp [3], arbeitet benutzerunabhängig und inkrementell und erkennt auch kontinuierlich gesprochene Sprache. Die Zusammenführung von Sprache und Gestik, die multimodale Integration, ist wesentlich durch eine zeitliche Koordination beider Modalitäten bestimmt. Dabei ist das Korrespondenzproblem [14] zu lösen, d.h. die semantisch-pragmatische Zuordnung von sprachlichen zu gestischen Äußerungssegmenten, wobei als Hauptkriterium deren zeitliche Nähe ausschlaggebend ist. Perzepte des gestischen und sprachlichen Kanals werden dazu zeitgestempelt und in einem Kurzzeitspeicher abgelegt, auf den der Integrationsprozeß zugreifen kann.

Die Gestenerkennung basiert auf der Detektion definitorischer Merkmale, die sowohl die Form als auch die zeitliche Expression einer Geste betreffen. Als Formmerkmale werden etwa Fingerstellung, Handorientierung und -position betrachtet; expressive Elemente, die generell auf das Vorliegen einer

[2] Langfristig ist auch eine kamerabasierte Erkennung geplant.

bedeutungstragenden Geste hinweisen, sind z.B. Ruhepunkte, hohe Beschleunigungen, Symmetrien und Abweichungen von Ruhestellungen, etwa der Handspannung oder der Positionen der Hände. Die Detektion der Form- und Expressionsmerkmale wird durch spezialisierte Erkenner geleistet, die auf den Sensordaten der Datenhandschuhe und Positionssensoren aufsetzen. Zum Beispiel müssen für eine Zeigegeste folgende Merkmale vorliegen: Der Zeigefinger ist gestreckt, die anderen Finger sind eingerollt und die Hand ist momentan in Ruhe. Liegen alle diese Merkmale vor, dann werden sie über eine geeignete Gewichtung zu einem Evidenzwert für das Vorliegen einer Zeigegeste zusammengefaßt. Dieser Evidenzwert wird dann mit einem Schwellwert verglichen, durch dessen Einstellung die Erkennung je nach Anforderung sensitiv oder träge gestaltet werden kann. Als ein Basisansatz zur Formbeschreibung wird das Gesten-Notationssystem *HamNoSys* [11] zugrundegelegt, das vollständige Gesten auf der Basis von atomaren Formelementen kompositionell (symbolisch) beschreibt. Die Erkennung läßt sich mit einer zeitlichen Integration der Formelemente bewerkstelligen [13]. Mit den beschriebenen Verfahren wurden so Erkenner für Basisinteraktionen (Pointing, Grasp, GraspRelease, Rotation, Translation) in einer Virtuellen Umgebung realisiert.

Eine reine Formerkennung gestischer Eingaben ist jedoch für die semantisch-pragmatische Auswertung unzureichend, sie dient rein der Klassifikation, d.h. der Detektion des Vorliegens einer Geste an sich. Die anschließende Interpretation der Geste erfolgt grundsätzlich im Kontext einer sprachlichen Äußerung und vor dem Hintergrund des Anwendungsszenarios der Virtuellen Konstruktion. Da alle betrachteten Gesten koverbal sind, kann die Gestenerkennung sensitiv eingestellt und trotzdem eine robuste Interaktionsauswertung erreicht werden. So reagiert das System durch niedrige Schwellwerte in den Gestenerkennern bereits auf geringste Anzeichen einer Benutzergeste. Auf diese Weise werden sehr viele Gesten-Hypothesen generiert, von denen aber nur wenige in geeigneter zeitlicher Beziehung zur sprachlichen Äußerung stehen und daher weiter interpretiert werden. Kontextinformationen über die virtuelle Umgebung fließen direkt in die Analyse der Gesten ein und ermöglichen deren situationsbezogene Auswertung. In Bezug auf die Benutzerdeixis bedeutet das etwa: Durch die Einbettung des Anwenders in das Referenzsystem der virtuellen Umgebung ist ein deiktischer Bezug zu den ihn umgebenden Objekten definiert. Das gestische Bedeuten von Objekten schränkt die für eine weitere Desambiguierung in Frage kommenden Objekte durch die Menge der in Zeige- und Blickrichtung [3] liegenden Objekte weiter ein. Die Interpretation mimetischer Gesten wird dagegen durch die mögliche Lage- und Positionsänderung der dargestellten Bauteile beschränkt. Unverbundene Teile lassen prinzipiell alle manipulativen Freiheitsgrade zu; nach einem Verbindungsschluß werden diese Freiheitsgrade gemäß der resultierenden Freiheitsmatrizen eingeschränkt.

Die Umsetzung multimodaler Anweisungen kann auf zwei unterschiedliche Arten erfolgen [9]. Sind einerseits alle Parameter einer Manipulation durch die Eingaben und den Kontext bestimmt, so wird sie in einem *diskreten* Ausführungsschritt durchgeführt. Alle gestischen Eingaben mit deiktischem/ikonischem Typ liefern die Informationen für solche diskreten Interaktionen. Dagegen bieten virtuelle Umgebungen andererseits den Vorteil des explorativen Vorgehens; die Resultate der Operationen lassen sich in Echtzeit während der Ausführung verfolgen. Mimetische Gesten, die gewünschte Objektmanipulationen vormachen, werden als derartige kontinuierliche Interaktionen umgesetzt (vgl. Abb. 7). Da bei den kontinuierlichen Interaktionen eine zeitlich ausgedehnte Bindung der sensorischen Eingabedaten an Bewegungen der virtuellen Objekte notwendig ist, wird eine Szenengraphstruktur der Interface-Komponente zugrundegelegt, in der spezielle Knotenklassen eingeführt werden [9]: Aktuatoren dienen als virtuelle Repräsentanten der Benutzerextremitäten, sie binden während einer kontinuierlichen Aktion an Motion-Modifikatoren, welche die übermittelten Daten auf die beschriebene Form hin analysieren. Die so erhaltenen Formprimitive binden dann an Manipulatoren, welche wiederum die virtuellen Objekte modifizieren (Abb. 8).

[3] Die Blickrichtung wird über einen in die Stereobrille integrierten Positionssensor erfaßt.

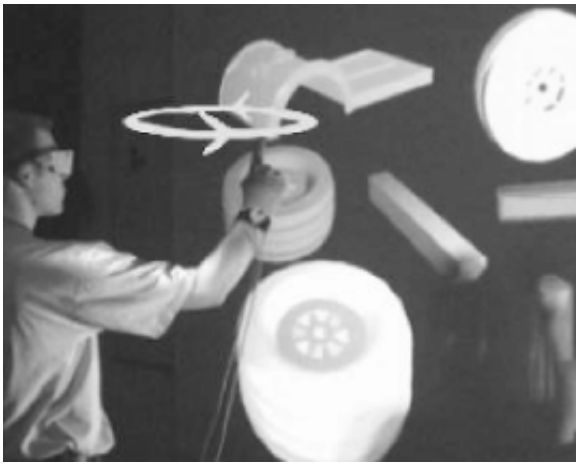


Abb. 7: Mimetische Geste zur Beschreibung einer Objektrotation.

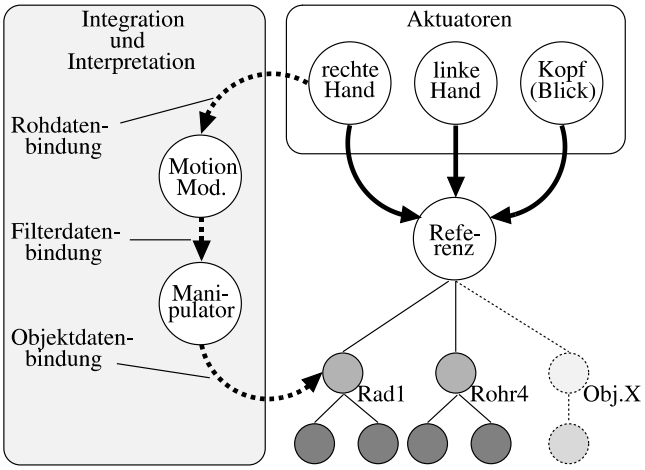


Abb. 8: Auswertung gestischer Benutzereingaben über temporäre Bindungen während der Interaktion.

4 Resümee

In diesem Beitrag wurden aktuelle Forschungsarbeiten im Bielefelder Labor für Künstliche Intelligenz und Virtuelle Realität dargestellt. Der Fokus lag dabei einerseits auf wissensbasierten Techniken für das Virtuelle Konstruieren, durch welche eine realitätsnahe, komfortable Manipulierbarkeit der virtuellen Umgebung erreicht wird. Andererseits wurden Techniken für die Ver-

arbeitung sprachlich-gestischer Eingaben bei der Mensch-Maschine-Interaktion an Großbild-Displays betrachtet. Die beiden Forschungsrichtungen ergänzen einander, da durch die Ausnutzung von Domänenwissen Vagheiten in Benutzereingaben und Ungenauigkeiten der VR-Eingabegeräte ausgeglichen werden können. Eine Verfeinerung der temporalen Zuordnung sprachlicher und gestischer Äußerungen ist derzeit in Arbeit. Möglicher Ausgangspunkt ist die Nutzung rhythmischer Muster, wie sie z.B. in der Betonung der Sprache und der Ausprägung der Gestik festgestellt werden; eine erste technische Studie liegt dazu vor[17]. Weitere Forschungsergebnisse betreffen die dynamische Konzeptualisierung der virtuellen Montageszene, bei der das System durch Abgleich von Struktur- und Formmerkmalen die entstehenden Baugruppen erkennt und dadurch u.a. dem Benutzer deren funktionsbezogene Benennung in sprachlichen Instruktionen ermöglicht [18]. Demonstrationen der beschrieben Arbeiten finden sich im WWW auf der Labor-Seite unter

www.TechFak.Uni-Bielefeld.DE/techfak/ags/wbski/

Hinweis

Die hier beschriebenen Arbeiten wurden durch die DFG im Sonderforschungsbereich 360 und vom Land NRW im Forschungsverbund „Die virtuelle Wissensfabrik" unterstützt.

Kontaktadresse:

Bernhard Jung:
Universität Bielefeld, Technische Fakultät, AG WBS (KI)
Postfach 10 01 31, D-33501 Bielefeld
Tel.: +49 521 106 2923
Email: jung@TechFak.Uni-Bielefeld.dE

Literatur

[1] K.Böhm, W. Hübner, and K.Väänänen. GIVEN:Gesture driven interactions in virtual environments:A toolkit approach to 3D interactions. Proc. Conf. Interface to Real and Virtual Worlds, 1992.

[2] F. Dai and M. Göbel. Virtual Prototyping - An approach using VR-techniques. In Proc. of the 14th ASME Int. Computers in Engineering Conference, Minneapolis, ASME, 1994.

[3] G.A. Fink, C. Schillo, F. Kummert, and G. Sagerer. Incremental speech recognition for multimodal interfaces. In Proceedings 24th Annual Conference of the IEEE Industrial Electronics Society, 2012-2017, 1998.

[4] A.G. Hauptmann and P. McAvinney. Gestures with speech for graphic manipulation. Int. Journal of Man Machine Studies, 38:231-249, 1993.

[5] B.W. Henson and N.P. Juster. Information requirements for the support of assembly mating conditions. In Proceedings of DETC'97, ASME, 1997.

[6] B. Jung, M.Hoffhenke, and I. Wachsmuth. Virtual assembly with construction kits. In 1998 ASME Design for Manufacturing Symposium – Virtual Design and Prototyping, 1998.

[7] R. Koller. Entwicklung einer Systematik für Verbindungen – ein Beitrag zur Konstruktionsmethodik. Konstruktion, 36(5):173-180, 1984.

[8] S. Kopp. Ein wissensbasierter Ansatz zur Modellierung von Verbindungen für die virtuelle Montage. Diplomarbeit, Technische Fakultät, Universität Bielefeld, März 1998.

[9] M.E. Latoschik, B. Jung, and I. Wachsmuth. Multimodale Interaktion mit einem System zur Virtuellen Konstruktion. In K. Beiersdörfer, G. Engels & W. Schäfer (Hrsg.): Informatik '99, 29. Jahrestagung der Gesellschaft für Informatik, Paderborn, Springer, 88-97, 1999.

[10] R.J. Popplestone, A.P. Ambler, and I.M. Bellos. An interpreter for a language for describing assemblies. Artificial Intelligence, 14(1):79-107, August 1980.

[11] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide. International Studies on Sign Language and Communication of the Deaf, vol.5, Signum Press, Hamburg, 1989.

[12] K. Roth. Konstruieren mit Konstruktionskatalogen, Band I, 2. Auflage, Springer, 1994.

[13] T. Sowa, M. Fröhlich and M.E. Latoschik. Temporal Symbolic Integration Applied to a Multimodal System Using Gestures and Speech. In A. Braffort et al. (Eds.): Gesture-based Communication in Human-Computer Interaction – Proceedings GW'99, Springer (LNAI 1739), 1999, 291-302.

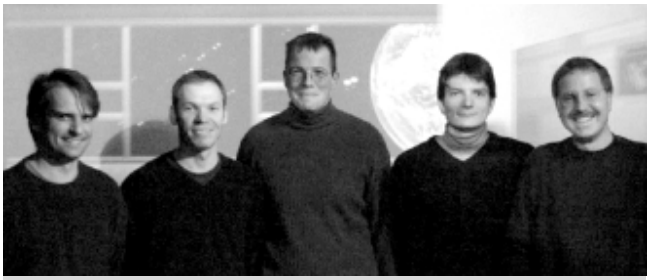
[14] R.K. Srihari. Computational models for integrating linguistic and visual information: A survey. Artificial Intelligence Review, 8:349-369, 1994.

[15] A. van Dam. Post-WIMP User Interfaces. In: Communications of the ACM, Vol.40, No.2, 1997.

[16] I. Wachsmuth. Virtuelle Realität, KI, 1/1998, p. 34.

[17] I. Wachsmuth. Kommunikative Rhythmen in Gestik und Sprache. Kognitionswissenschaft, 8(4), Januar 2000, 151-159.

[18] I. Wachsmuth and B. Jung. Dynamic conceptualization in a mechanical-object assembly environment. Artificial Intelligence Review, 10(3-4):345-368, 1996.



Die Autoren von links nach rechts:

Bernhard Jung studierte Informatik und Computerlinguistik an der Universität Stuttgart und an der University of Missouri, St. Louis, USA. Seit 1993 arbeitet er an der Universität Bielefeld, wo er 1996 promovierte und seit 1997 als Wissenschaftlicher Assistent beschäftigt ist. Gegenwärtige Forschungsarbeiten liegen der Entwicklung wissensbasierter Methoden für interaktive virtuelle Umgebungen.

Stefan Kopp studierte Naturwissenschaftliche Informatik mit dem zweiten Hauptfach Physik an der Universität Bielefeld und ist dort seit 1998 wissenschaftlicher Mitarbeiter. Seine Forschungsinteressen umfassen multimodale Systeme, virtuelle anthropomorphe Agenten, Computergraphik/-animation und Virtual Reality.

Marc Erich Latoschik studierte Naturwissenschaftliche Informatik mit dem zweiten Hauptfach Biologie an der Universität Bielefeld und ist dort seit 1996 als wissenschaftlicher Mitarbeiter beschäftigt. Er arbeitet an der Entwicklung multimodaler Steuerungssysteme für die Virtuelle Realität am Beispiel der Virtuellen Konstruktion.

Timo Sowa studierte an der Universität Bielefeld Naturwissenschaftliche Informatik mit dem zweiten Hauptfach Biologie und ist dort seit 1998 wissenschaftlicher Mitarbeiter. Seine Interessenschwerpunkte liegen im Bereich der intelligenten, multimodalen Mensch-Maschine Kommunikation.

Ipke Wachsmuth studierte Mathematik und Informatik an der TU Hannover, wo er 1980 promovierte. Nach Lehr- und Forschungstätigkeiten an der Universität Osnabrück, der Northern Illinois University und bei IBM Deutschland habilitierte er sich 1989 an der Universität Osnabrück; im gleichen Jahr wurde er auf die Professur für Wissensbasierte Systeme (Künstliche Intelligenz) an die Universität Bielefeld berufen.