

BodyTalk - Gestenbasierte Mensch-Computer-Interaktion zur Steuerung eines multimedialen Präsentationssystems

Christian Leubner, Jens Deponte, Sven Schröter, Helge Baier
Universität Dortmund, Fachbereich Informatik VII (Graphische Systeme)

Zusammenfassung

Es wird ein computerbasiertes System zur Steuerung multimedialer Präsentationen durch Armgesten vorgestellt. Das System ist in der Lage, zuvor angelernte Armbewegungen (Gesten) des Benutzers zu erkennen und zur Interaktion mit einem Präsentationssystem einzusetzen. Die Gestenerkennung erfolgt über Hidden-Markow-Modelle (HMM), zur Modellierung einer Präsentation wird ein spezielles Interaktionsparadigma verwendet. Die Erweiterbarkeit des Systems und das Ansteuern und Einbinden existierender, weit verbreiteter Präsentationssoftware auf unterschiedlichen Plattformen ist wesentlicher Bestandteil des Konzepts. Eine übersichtliche intuitive Benutzungsoberfläche für Systemeinstellungen sowie ein Kontrollbildschirm auf Anwenderseite sorgen für einen reibungslosen Ablauf der Präsentation.

1 Einführung

Multimediafähige Rechner zusammen mit der Videoprojektion werden zunehmend das zentrale Präsentationshilfsmittel für Vortragsveranstaltungen. Dies reicht von Notebook-basierten Präsentationen hin zur gleichzeitigen Präsentation in örtlich entfernten, vernetzten Hörsälen (Deponte 1997). Solche Vorträge erfordern die Interaktion mit dem Präsentationssystem, im einfachsten Fall etwa zum Wechseln auf die nächste Vortragsseite. Hierfür kann die Rechnertastatur verwendet werden, die aber häufig an den Aufstellungsort des Benutzers gebunden ist. Fernbedienungen, wie sie etwa bei Videoprojektoren mitgeliefert werden, können ortsungebunden eingesetzt werden. Sie haben jedoch den Nachteil geringer Flexibilität bzgl. der Anpassung an unterschiedliche Vortragsmodi und es fällt etwas schwer, sich in diesem Fall an die Tastenbelegungen zu gewöhnen.

Im folgenden wird das System BodyTalk zur Interaktion mit rechnerbasierten Präsentationssystemen vorgestellt. BodyTalk löst die Ortsgebundenheit der Tastatur in gewissem Umfang auf, behält jedoch die Flexibilität der tastaturbasierten Bedienung in gewissem Umfang bei. Der Benutzer wird über eine Videokamera erfasst. Durch Ausführung spezieller Armbewegungen, sogenannter Bewegungsgesten, kann er diesen zugeordnete Kommandos an das Präsentationssystem auslösen. Die möglichen Kommandos werden dem Vortragenden auf einem großflächigen Monitor mit grob strukturierten Fenstern angezeigt, die ein Erkennen auch über eine größere Distanz erlauben.

Die Verwendung von Videokameras zur Interaktion mit technischen Systemen findet zunehmend Interesse. Es gibt eine Vielzahl von Szenarien bezüglich eingesetzter Körperteile und Anwendungen. Eine recht umfassende Übersicht zum Einsatz der Hand ist bei (Kohler, 1998, 1999) zu finden. Der Vorteil von Videokameras gegenüber anderen Sensoren liegt darin, dass am Benutzer keine Sensoren angebracht werden müssen, wie dies etwa bei elektromagnetischen Verfahren der Fall ist. Das Computersehen stellt eine Möglichkeit, auch in Kombination mit anderen, zum Ersatz von Tastatur und Maus in Interaktionssituationen dar, wo diese ungünstig sind (Cooperstock 1997).

Das BodyTalk-System repräsentiert eine Möglichkeit der computersehensbasierten Interaktion, deren Funktionsfähigkeit anhand eines Prototypen nachgewiesen wurde. Darüber hinaus wurden in BodyTalk weitere Konzepte integriert, die für ein Interaktionssystem zur Vortragspräsentation unabhängig von der computersehensbasierten Vorgehensweise interessant und relevant sind:

- Bodytalk ist ein Interaktionssystem, das unabhängig vom Präsentationssystem ist. Im vorgestellten Prototyp wird dies an KPresenter (Stadlbauer 1999) demonstriert. Das Anbinden anderer Fremdprodukte ist vorgesehen.
- Die Navigation durch die Präsentation ist unabhängig vom Präsentationssystem und wird von BodyTalk überlagert.
- Zur Navigation bietet BodyTalk einen Präsentationsbaum mit roten Fäden an, mit denen über die Gesteneingabe interagiert wird.

In Kapitel 2 wird der Aufbau und die Funktionsweise des BodyTalk-Systems erläutert. Insbesondere wird das Konzept des Präsentationsbaums vorgestellt. In Kapitel 3 wird auf die Softwarestruktur eingegangen. Dabei werden auch ausgewählte technische Aspekte skizziert, insbesondere der Lösungsansatz für die Computersehenskomponente. Kapitel 4 beschreibt Erfahrungen mit der Anwendung des BodyTalk-Systems.

2 Aufbau und Funktionsweise des BodyTalk-Systems

2.1 Hardware-Konfiguration

Zum Aufbau des BodyTalk-Systems werden zwei Rechner benötigt, die mittels einer Netzwerkverbindung miteinander verbunden sind: ein Interaktionsrechner mit Framegrabberkarte und ein Präsentationsrechner. Weiterhin sind ein Videoprojektor zur Darstellung der Präsentation, eine Videokamera zur Erfassung des Vortragenden sowie ein Display zur Visualisierung des Präsentationszustandes erforderlich. Da das BodyTalk-System zur Zeit noch keine kontinuierliche Gestenerkennung leistet, wird zudem ein Fußschalter eingesetzt, dessen Betätigung den Anfang und das Ende einer Steuerungsgeste markiert. Abbildung 1 zeigt den schematischen Aufbau der BodyTalk-Hardware.

2.2 Funktionsweise

Der Ablauf einer Präsentation mit BodyTalk erfolgt prinzipiell ähnlich wie eine rein computergestützte Vorführung. Lediglich die Interaktion des Vortragenden mit dem Präsentationscomputer wird durch das BodyTalk-System übernommen. Auf dem Kontrollbildschirm des Benutzers wird der Zustand der Präsentationssteuerung und die aktuelle Position innerhalb des Vortrags angezeigt. Durch Ausführen von Armgesten kann der Benutzer durch seinen Vortrag navigieren, d.h. zwischen den Folien wechseln. Um eine größtmögliche Kompatibilität mit bestehender Präsentationssoftware zu gewährleisten, verwendet BodyTalk ein abstraktes Präsentationsmodell, das den strukturellen Aufbau und den zeitlichen Ablauf des Vortrages erfasst. Die Struktur des Vortrages wird durch einen *Präsentationsbaum* abgebildet, der die Gliederung und die Bestandteile der Folien erfasst.

Der Präsentationsbaum ordnet die Präsentationselemente hierarchisch an und ermöglicht so die effiziente Auswahl eines Elements. Ein Vortrag enthält eine Menge von Folien, der einzelne Textelemente so zugeordnet sind, dass sich eine Baumstruktur ergibt. An den Knoten befinden sich die Objekte der Präsentation, die sich in folgende Klassen gliedern:

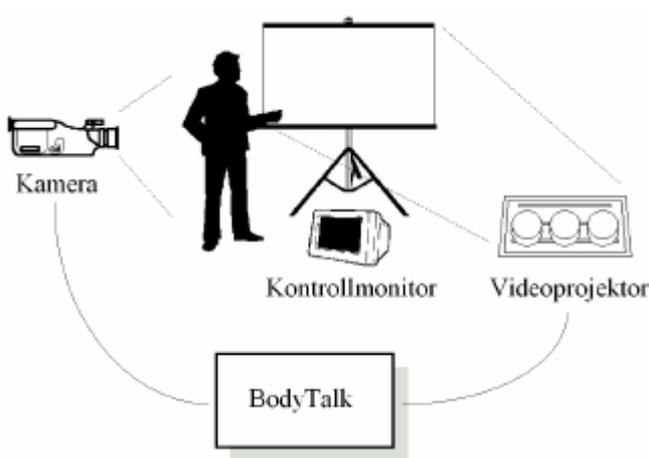


Abb. 1: Schematischer Aufbau der Interaktionshardware des BodyTalk-Systems

1. *Folien* beinhalten untergeordnete, grafische Folienelemente, die auf einer Seite der Präsentation angezeigt werden sollen.
2. *Foliengruppen* dienen dazu, Folien zu Gruppen zusammenzufassen, um die Navigation in großen Mengen zu erleichtern. Foliengruppen können beliebig viele Elemente der beiden Typen Foliengruppen und Folien beinhalten.
3. *Folienelementgruppen* dienen dazu, Folienelemente zu Gruppen zusammenzufassen, um somit die Navigation in großen Mengen zu erleichtern. Folienelementgruppen können beliebig viele Elemente der Typen Folienelementgruppen und Folienelemente beinhalten.
4. *Folienelemente* stellen die Blätter des Baumes dar. Folienelemente repräsentieren Bestandteile eines Vortrages, also z.B. Videos, Bilder und Texte. Folienelemente sind Folien untergeordnet. Sie enthalten Informationen über:
 - Typ (z.B. HTML, Gif, Text, Postscript, ...)
 - eindeutige Bezeichnung zum Vergleichen und Ansteuern
 - Referenz auf das Element der angesteuerten Präsentation
 - Aktivierungsaktion, die bei Betreten des Objekts im Roten-Faden-Modus ausgeführt werden soll
 - Deaktivierungsaktion, die bei Verlassen des Objekts, im Roten-Faden-Modus, ausgeführt werden soll.

Ein Vortrag besteht oft aus Sequenzen, in der die zeitliche Reihenfolge der Vortragspunkte vom Vortragenden vorgeben ist. Eine solche zusammenhängende Sequenz wird in BodyTalk als *Roter Faden* bezeichnet. Da eine solche Struktur häufig vorkommt und den normalen, geplanten Verlauf einer Präsentation widerspiegelt, ist es für den Vortragenden vorteilhaft, wenn das Präsentationssystem solche Strukturen unterstützt.

In Abbildung 2 ist ein Präsentationsbaum mit einem Roten Faden, angedeutet durch die durchgezogene dickere Linie, dargestellt. Das Beispiel zeigt die baumartige Gliederung der Präsentation. Sowohl die „Einleitung“ als auch die „Kapitel“ sind Foliengruppen, die wiederum aus mehreren Folien bestehen. Der Inhalt der einzelnen Folien wiederum besteht aus Folienelementen und Folienelementgruppen, die hier nicht mehr dargestellt sind.

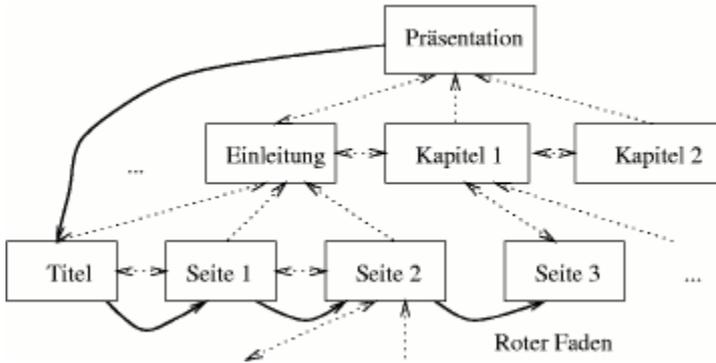


Abb. 2: Präsentationsbaum mit Rotem Faden, angedeutet durch die durchgezogenen Linien

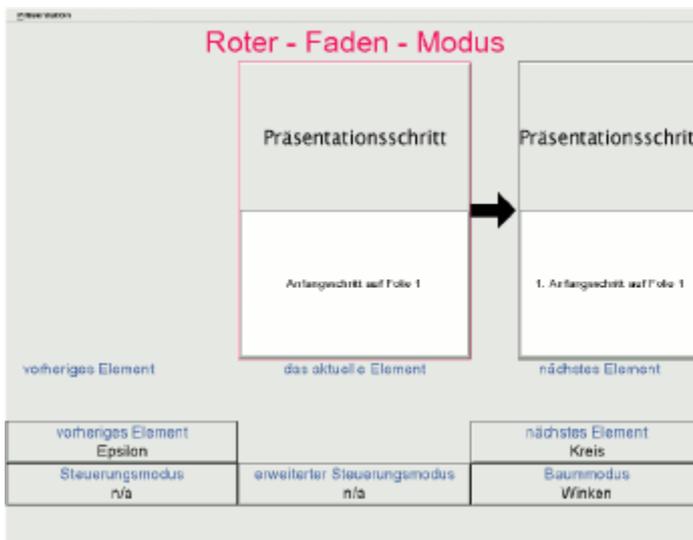


Abb. 3: Benutzungsoberfläche des Rote-Faden-Modus, die auf dem Kontrollmonitor dargestellt wird

Die Steuerung von BodyTalk kann wahlweise in einem „Rote-Faden Modus“ oder in einem „Baum-Modus“ erfolgen. Beim Start des Systems wird der Rote-Faden-Modus verwendet. Abhängig vom Modus wird ein entsprechendes Fenster auf dem Bildschirm dargestellt. Abbildung 3 zeigt das Fenster des Rote-Faden-Modus. Im oberen Teil der gleichartig aufgebauten Fenster wird der aktuelle Kontext im Vortrag dargestellt, dem die lokalen Navigationsmöglichkeiten zu entnehmen sind. Im unteren Teil werden die möglichen Anweisungen angeboten und die zugehörige auslösende Geste (s.u.) aufgeführt.

Die Informationen auf dem Kontrollbildschirm sind leicht und schnell zu erfassen, so dass sich zum einen der Vortragende ganz auf seine Präsentation konzentrieren kann und zum anderen der Bildschirm auch in größerer Entfernung zum Benutzer aufgestellt werden kann. BodyTalk bietet die Möglichkeit mehr als einen Roten Faden zu verwalten. Im Baum-Modus wird die Navigation innerhalb der Baumstruktur unterstützt, indem Vater-, Sohn- und Geschwisterknoten innerhalb des Baumes angezeigt werden.

Die Interaktion zwischen Benutzer und Computer erfolgt über die Ausführung von Bewegungsgesten, die durch das System zuvor angelernt wurden. Im Rahmen der praktischen Erfahrungen mit BodyTalk hat sich das Gestenset, das Abbildung 4 entnommen werden kann, als gute Wahl erwiesen. Die Kurven beschreiben die Armbewegung. So wird bei Geste 1 und 2 der Arm so bewegt, dass sich Hand auf einem Kreis bewegt, einmal im Gegenuhrzeigersinn, einmal entgegen. Bei Geste 3 und 4 wird der Arm so bewegt, dass die Hand ein Epsilon beschreibt, einmal von unten nach oben, einmal von oben nach unten. Grundsätzlich sind beliebige andere Gesten möglich.

Bisher wurde BodyTalk aus Sicht des präsentierenden Anwenders vorgestellt. Daneben bietet BodyTalk eine Reihe von Funktionen zur Systemadministration an, die sich in die Konfigurationen der Gesten- und der Präsentationskomponente aufteilt. Alle Konfigurationsaufgaben können komfortabel über eine dafür angebotene grafische Benutzeroberfläche von BodyTalk bewältigt werden. Innerhalb der Gestenkomponente kann das Trainieren von Gesten, die Auswahl und Bestimmung eines Gestensets sowie die Videoeinstellungen erfolgen.



Abb. 4: Beispielhaftes Gestenset zur Navigation. Die Kurven deuten die weiträumig auszuführenden Handbewegungen an.

Bei der Präsentationskomponente sind folgende Einstellungen konfigurierbar:

- *Importieren einer Präsentation:* Für die Steuerung einer Präsentation mit BodyTalk ist die Aufbereitung der Präsentationsdaten entsprechend der spezifischen Präsentationssoftware erforderlich.
- *Nachbearbeiten einer Präsentation:* ermöglicht die Änderung der Präsentationsstruktur, die durch das Importieren erzeugt wurde.
- *Laden/Speichern einer Präsentation:* Dateioperationen im BodyTalk Format.
- *Profilverwaltung:* Für unterschiedliche Benutzer können jeweils spezifische Profile angelegt werden, die beispielsweise Angaben über das zu verwendende Gestenset enthalten.
- *Gestenverwaltung:* Unter diesem Punkt sind Funktionen zur Verwaltung der Gesten zu finden, wie zum Beispiel das Antrainieren neuer Gesten oder die Zuweisung einer Bedeutung zu einer Geste.
- *Speichern von Profildaten.*

3 Software

Die BodyTalk-Software setzt sich im wesentlichen aus einer Gestenerkennungs- und einer Präsentationskomponente zusammen, deren Inhalte im folgenden näher erläutert werden.

3.1 Software-Struktur

Die Gestenerkennungskomponente besteht aus zwei Packages, wobei eines in C++ und eines in IDL (Sun 1997) erstellt wurde, das sowohl in Java als auch in C++ umgesetzt werden kann. Das C++-Package enthält im wesentlichen die Klassen zur Bildanalyse und zur Gestenerkennung,

wohingegen im IDL-Package die Klassen für die Schnittstelle zwischen der Gestenerkennung und der Präsentationskomponente enthalten sind.

Die Präsentationskomponente ist vollständig in Java implementiert worden und setzt sich aus mehreren Packages zusammen:

- *Package „Action“*: legt die Aktionen fest, die anhand sogenannter Profile den Gestenobjekten zugeordnet werden können.
- *Package „Basics“*: Verwaltung von Listen und der Objektspeicherung.
- *Package „Editor“*: umfasst die Klassen für die Benutzeroberfläche.
- *Package „Gesture“*: enthält die Klassen zur Verwaltung von Gesten und ihrer Bedeutung sowie die Beschreibung der Schnittstelle zur Gestenerkennungskomponente.
- *Package „Import“*: leistet die Konvertierung von Präsentationen eines Fremdpräsentationssystems in das BodyTalk eigene Format.
- *Package „Presentation“*: stellt Klassen für einzelne Präsentationselemente und den Präsentationskoordinator zur Verfügung.
- *Package „KPresenter“*: bildet das Importmodul für KPresenter (Stadlbauer 1999) als Beispiel für die konkrete Ansteuerung eines Präsentationssystems.
- *Package „Profile“*: Verwaltung der Benutzerprofile.

3.2 Kommunikation zwischen den Komponenten

Aus Abbildung 5 können sowohl die Komponenten des BodyTalk-Systems sowie weitere kommunizierende Komponenten innerhalb des Präsentationssystems entnommen werden. Durchgehende Linien symbolisieren eine direkte Kommunikation, gepunktete Linien markieren Vermittlungsfunktionen. Die Pfeile zeigen jeweils in die Richtungen, in die Informationen ausgetauscht werden.

Zur Realisierung der Kommunikation zwischen der Gestenerkennungs- und der Präsentationskomponente wird CORBA (OMG 2000) verwendet. Der in C++ implementierte „GRMServer“, der die videobasierte Gestenerkennung bereitstellt, meldet sein Objekt GRM bei dem CORBA Naming Service an, der als separater Serverprozess läuft. Die mit „IOR“ gekennzeichneten Komponenten in Abbildung 5 deuten darauf hin, dass an diesen Stellen lediglich Objektreferenzen des Naming Service ausgetauscht werden.

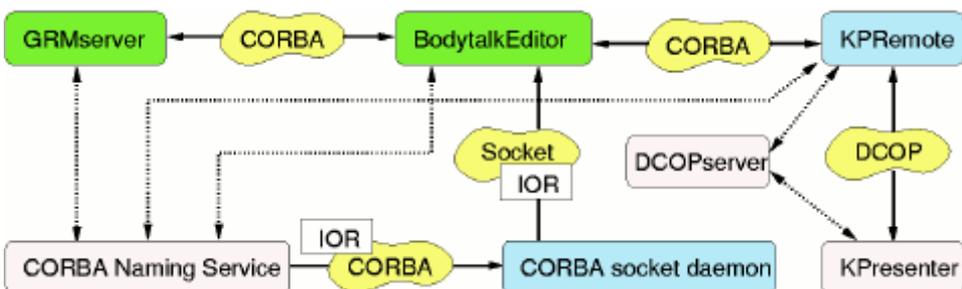


Abb. 5: Kommunikationsstruktur in BodyTalk

Die angesteuerte Präsentationssoftware ist in diesem Beispiel das Programm „KPresenter“ (Stadlbauer 1999), dessen Importmodul „KPRremote“ die Verbindung zu BodyTalk herstellt. Die Verbindung des Importmoduls zu KPresenter wird über das *Desktop COmmunication Protocol*

(DCOP) der KDE Oberfläche realisiert, das weniger Ressourcen benötigt und weniger leistungsfähig als CORBA ist, sich für diese Anwendung aber als hinreichend erwiesen hat.

3.3 Einbindung von Fremdpräsentationssystemen

Durch seine offene Schnittstelle ermöglicht BodyTalk die Anbindung beliebiger Präsentationssysteme, wie Microsoft Powerpoint oder StarPresenter (Sun 2000). Möglich wird diese Kompatibilität durch das in Abschnitt 2.2 erläuterte abstrakte Präsentationsmodell. Die Kommunikation zwischen den BodyTalk Komponenten und einem externen Präsentationsprogramm erfolgt über ein Importmodul, das für die spezifische Präsentationssoftware entwickelt werden muss. Beispielfähig realisiert wurde die Anbindung des Programms „KPresenter“, das Teil des KOffice Projekts (KDE 1999) ist. Wie Abbildung 4 entnommen werden kann, erfolgt die Kommunikation ebenfalls über CORBA.

3.4 Bildverarbeitung und Gestenerkennung

Eine Möglichkeit, Bewegungsabläufe zu klassifizieren, stellen Trajektorien dar, die einen Bewegungspfad, dessen Punkte eindeutig bzgl. ihrer Reihenfolge und Position festgelegt sind (z.B. Liste von Punkten), beschreiben. Nach Anwendung von Merkmalsextraktions- und Klassifikationsverfahren, werden die einzelnen Gesten erkannt. Die meisten Merkmalsextraktionen basieren auf geometrischen oder numerischen Algorithmen zur Auswertung der Punktfolgen, wie z.B. Polygonzugberechnung (Lipscomb 1991), Kreisgitterextraktion (Müller 1998) und Momente (Hu 1962, Wood 1995). Welche der aufgeführten Verfahren ausgewählt werden, hängt von den zu erkennenden Gesten (z.B. Größe der Geste) und von dem auf die Merkmale angewandten Klassifikationsverfahren ab. So werden die Momentmerkmale mit Hilfe von Kreuzkorrelationsklassifikatoren am besten klassifiziert. Für die beiden anderen Verfahren liefern Raumklassifikatoren wie der Abstandsclassifikator und der Trennebenenklassifikator (Politt 1993) die besten Ergebnisse.

3.4.1 Hidden-Markov-Modelle in der Gestenerkennung

Ein neuerer Ansatz nutzt zur Modellierung von Gesten *Hidden-Markov-Modelle* (HMM), die unter anderem in der Spracherkennung erfolgreich eingesetzt werden. HMM ordnen Beobachtungsfolgen Folgen von Zuständen mit Übergangswahrscheinlichkeiten zu, die von außen nicht wahrnehmbar sind. Zum Erkennen einer Geste wird aus der Trajektorie eine Beobachtungsfolge bestimmt. Für jedes HMM, das eine Geste repräsentiert, wird die Wahrscheinlichkeit bestimmt, dass es diese Beobachtungsfolge erzeugt. Die Geste mit dem HMM größter Wahrscheinlichkeit wird als „erkannt“ klassifiziert.

Auch in der bildbasierten Gestenerkennung sind HMM in verschiedenen Anwendungen zum Einsatz gekommen. Beispielsweise wurde in (Rigoll 1997) ein videobasiertes System zum Erkennen von Personengesten vorgestellt, das in der Lage ist, zwischen 24 verschiedenen Gesten zu unterscheiden. Eine Erweiterung macht es möglich, kontinuierlich und positions-unabhängig Gesten zu erkennen (Rigoll 1998).

In BodyTalk wird ebenfalls nach diesem Ansatz verfahren. Die Modellierung einer Bewegungsgeste mittels eines HMM erfolgt durch ein *Training*, währenddessen die Geste mehrfach ausgeführt wird. Aus den Videosequenzen werden Beobachtungsfolgen berechnet, die zur Ermittlung der Parameter einer HMM genutzt werden. Zum Erkennen einer Geste wird aus der Videosequenz einer Geste wieder eine Beobachtungsfolge errechnet. Für jedes HMM, das eine Geste repräsentiert, wird die Wahrscheinlichkeit bestimmt, dass es diese Beobachtungsfolge erzeugt.

Die aus der Handbewegung gewonnene Trajektorie wird durch eine Folge von eindimensionalen Werten repräsentiert. Diese werden mit Hilfe einer *Quantisierung* ermittelt, die den Bildbereich der Kamera in verschiedene Bereiche (*Cluster*) einteilt, die mit einer eindeutigen Nummer

versehen werden. Die Folge der zweidimensionalen Positionen wird auf die Folge der zugehörigen Clusternummern abgebildet, die sich aus der Reihenfolge des Durchlaufs der Bildbereiche ergibt.

3.4.2 Empirische Parameterevaluierung

Einige Parameter, beispielsweise die Clusteranzahl bei der Quantisierung, konnten nicht im Vorfeld allgemeingültig angegeben werden, sondern mussten durch ständige Versuche ermittelt werden. Auch über die Anzahl der Trainingssequenzen für die Gestenerlernung konnten im Vorfeld keine Annahmen gemacht werden. Idealerweise werden diese Parameter so gewählt, dass die Erkennungswahrscheinlichkeit maximal wird. Die für die Merkmalsvektoren ermittelten Wahrscheinlichkeiten der HMM sind ein qualitatives Maß für die Erkennung. Ein Optimieren der Parameter auf möglichst große Wahrscheinlichkeiten beim HMM führt aber nicht notwendigerweise zu einer optimalen Erkennungsrate. Es erweist sich als sinnvoller, das Verhältnis zwischen der Wahrscheinlichkeit der ausgeführten Geste zu den Wahrscheinlichkeiten der *falschen* Gesten zu optimieren.

Um die optimalen Werte für die Parameter experimentell zu ermitteln, wurde ein Gestenset mit den fünf Gesten aus Abbildung 4 angelegt, zu denen jeweils 40 Sequenzen aufgezeichnet wurden. Anschließend wurden daraus mehrere Gestensets erstellt, in dem jeweils fünf, zehn, 20 und 30 Sequenzen kopiert wurden. Da die beiden Parameter nicht unabhängig voneinander sind, wurden sämtliche Kombinationsmöglichkeiten der verschiedenen Gestensets und der Anzahl der Cluster (5, 10, 15, 20, 25, 30, 35 und 40) getestet. Als Eingabe dienten dazu 20 Testsequenzen, wobei jeweils vier Sequenzen eine Geste zeigten.

Es hat sich herausgestellt, dass 40 Cluster bei fünf Sequenzen zu viel zu sein scheinen, so dass die HMM nicht mehr ausreichend trainiert werden können. Zum anderen scheinen 40 Trainingssequenzen auszureichen, da bei dieser Anzahl fast kein Unterschied bei einer Veränderung der Cluster-Anzahl erkannt werden kann. Bei wenigen Clustern und wenigen Testsequenzen sind die Werte zwar gut, aber sehr schwankend.

4 Praktische Erfahrung und Bewertung

Bei verschiedenen Einsätzen hat sich BodyTalk als brauchbares und zuverlässiges System zur Präsentationsunterstützung erwiesen. In einer empirischen Untersuchung mit fünf Versuchspersonen wurde die Erkennungsleistung und die Erlernbarkeit von BodyTalk evaluiert. Dabei wurde das Gestenset aus Abbildung 4 verwendet. Die Funktionsweise des Systems konnte allen Versuchspersonen in maximal drei Minuten erläutert werden. Zur Bewertung interindividueller Unterschiede in Bezug auf die Erkennungsrate wurden die Probanden zunächst aufgefordert, mit einem bereits angelernten Gestenset zu arbeiten. Jede Geste wurde von jeder Person fünf bis sechs Mal durchgeführt und ihre Erkennung gewertet. Die daraus resultierende Erkennungsrate der jeweiligen Versuchsperson schwankte deutlich zwischen 36% und 100%. Bezogen auf alle ausgeführten Gesten aller Personen lag die Rate der richtig erkannten Gesten bei 68%. Aufgrund der starken individuellen Schwankung der Erkennungsrate selbst bei dieser eher kleinen Anzahl an Versuchspersonen kann geschlossen werden, dass das Anlernen der Gesten des Vortragenden dringend empfehlenswert ist. Zum Vergleich wurde daher für jede Versuchsperson das Gestenset individuell neu angelernt, indem alle fünf Gesten fünf Mal durchgeführt wurden. Pro Versuchsperson hat dies zwischen vier und sechs Minuten in Anspruch genommen. Daraufhin wurde die Aufgabe wiederholt, alle Gesten fünf bis sechs Mal durchzuführen. Als Resultat konnte eine erhebliche Verbesserung der Erkennungsrate festgestellt werden, die selbst im schlechtesten Fall bei einer Person noch 93% betrug. Bezogen auf alle ausgeführten Gesten aller Versuchspersonen wurde eine Erkennung von 97% erreicht. Abschließend wurden die Versuchspersonen gebeten, eine Schulnote für BodyTalk zu vergeben und ihre Eindrücke zu äußern. BodyTalk wurde durch-

weg von allen Probanden als „gut“ beurteilt. Negativ bemerkt wurde, dass der Fußschalter störe, die zu verwendenden Gesten zu groß und das häufige Ausführen der Gesten anstrengend sei.

Aufgrund der Eigenschaften der Hidden-Markov-Modelle wird bei Betätigung des Fußschalters in jedem Fall eine Geste erkannt, auch wenn keine der definierten Gesten ausgeführt wurde. Durch die hohe Stabilität bei der Gestenerkennung treten Fehlerkennungen, die unerwünschte Aktionen in der Präsentationssoftware auslösen, selten auf. Fälschliche Reaktionen des Programms, beispielsweise das Rückwärtsblättern anstelle des Vorwärtsblätterns, müssen vom Benutzer durch eine passende Folge von Anweisungen kompensiert werden, im Beispiel also durch doppeltes Vorwärtsblättern. Eventuell wäre an dieser Stelle eine Undo-Operation sinnvoll.

Zu Beginn einer Präsentation, wenn sowohl Publikum als auch der Vortragende nicht mit BodyTalk vertraut sind, mag es für den Vortragenden und die Zuhörenden ungewöhnlich sein, ausschweifende Armbewegungen auszuführen. Dieser Überraschungseffekt für das Publikum lässt allerdings rasch nach.

Auf der technischen Seite besteht weiterer Entwicklungsbedarf, um eine kontinuierliche Gestenerkennung zu realisieren und den Fußschalter obsolet zu machen. Inhaltliche Erweiterungen sind insbesondere das Einbringen weiterer Funktionen in die Interaktionsschnittstelle, wie sie bei komplexen Präsentationsumgebungen, etwa den zu Beginn erwähnten vernetzten multimedialen Hörsälen, auftreten. Eine weitere Möglichkeit ist die zusätzliche Interaktion mit der projizierten Präsentation, etwa durch Hinzeigen, beispielsweise in dem Fall, dass sie interaktiv zu bedienende Software zeigt. Dieser anwendungsbezogene Dialog ist von der Navigation mittels BodyTalk unabhängig.

Das BodyTalk System ist im Rahmen der Projektgruppe 345 an der Universität Dortmund im Fachbereich Informatik entstanden. Zu dieser Projektgruppe gehörten Irina Alesker, Ralf Böning, Dirk Försterling, Daniel Herche, Roland Kuck, Alexandra Nolte, Michael Pack, Thomas Rosanski, Birgit Scheer, Konstantin Steuer, Alexander Umanski und Björn Weitzig.

Literaturverzeichnis

- Assan, M., Grobel, K. (1997): Video-based sign language recognition using Hidden Markov models. In: Lecture Notes in Artificial Intelligence 1371, Subseries of Lecture Notes in Computer Science, Gesture and Sign Language in Human-Computer Interaction. Bielefeld: International Gesture Workshop. S. 97-109.
- Cooperstock, J.R., Feld, S.S., Buxton, W., Smith, F.C. (1997): Reactive environments – throwing away your keyboard and mouse. *Communications of the ACM*. 40 (9). S. 65-73.
- Deponte, J., Müller, H., Pietrek, G., Schlosser, S., Stoltefuß, B. (1997): Design and Implementation of a System for Multimedial Distributed Teaching and Scientific Conferences. In: Proc. Virtual Systems and Multimedia (VSMM'97). IEEE Press.
- Hienz, H., Kraiss, K.-F., Bauer, B. (1999): Continuous Sign Language Recognition using Hidden Markov Models. In: Tang, Y. (Hrsg.): ICMI'99 - The Second International Conference on Multimodal Interface. S. IV10-IV15. Hong Kong. <http://www.techinfo.rwth-aachen.de/Veroeffentlichungen/1999.html>
- Hu, M.-K. (1962): Visual Pattern Recognition by Moment Invariants. *RE Trans. Inf. Theory* Vol. 8. S. 179-187.
- KDE Entwickler (1999): About the KOffice. <http://koffice.kde.org>
- Kohler, M., Schröter, S. (1998): Handgestenerkennung durch Computersehen. In: Dassow, J.; Kruse, R., (Hrsg.): Informatik '98, Informatik aktuell, Berlin: Springer-Verlag. Siehe auch: [_](#)
- Kohler, M. (1999): New Contributions to Vision-Based Human-Computer Interaction in Local and Global Environments. Sankt Augustin: Infix-Verlag.
- Lipscomb, J. S. (1991) : A Trainable Gesture Recognizer. *Pattern Recognition* Vol. 24, S. 895-907
- Müller, S., Rigoll, G., Mazurenok D., Willet, D. (1998): Invariante Erkennung handskizzierter Piktogramme mit Anwendungsmöglichkeiten in der inhaltsorientierten Bilddatenbankabfrage. In: 20. DAGM-Symposium Tagungsband. Berlin: Springer-Verlag, S. 271-279.
- OMG Object Management Group (2000): Corba Basics. <http://www.omg.org/gettingstarted/corbafaq.htm>

- Politt, C. (1993): Vergleich des Trennebenenklassifikators mit dem „Nächsten Nachbarn“-Klassifikator. In: Mustererkennung 1993. Berlin: Springer-Verlag.
- Rigoll, G., Eickeler, S., Kosmala, A. (1997): High performance realtime gesture recognition using Hidden-Markov-Models. In: Lecture Notes in Artificial Intelligence 1371, Subseries of Lecture Notes in Computer Science, Gesture and Sign Language in Human-Computer Interaction. Bielefeld: International Gesture Workshop. S. 69-80.
- Rigoll, G., Eickeler, S., Kosmala, A. (1998): Hidden Markov model based continuous online gesture recognition. In: International Conference on Pattern Recognition. Brisbane. S. 1206-1208.
- Stadlbauer, R. (1999): KPresenter. <http://koffice.kde.org/kpresenter/index.html>
- Sun Microsystems (1997): Java IDL.
- Sun Microsystems (2000): StarOffice. <http://www.sun.com/products/staroffice>
- Wood, J. (1995): Invariant Pattern Recognition: A Review. Pattern Recognition Vol 29. S. 1-17

Adressen der Autoren

Christian Leubner / Helge Baier / Jens Deponte / Sven Schröter
Universität Dortmund
Fachbereich VII
Graphische Systeme
Otto-Hahn-Str. 16
44221 Dortmund
leubner@ls7.cs.uni-dortmund.de