

3D-Gestenerkennung und Sprache - ICONIC

Handout zum Seminar

“multimodale Mensch-Maschine-Kommunikation”

bei Prof. Ipke Wachsmuth / Timo Sowa

WS 2000/2001

Marco Balke

mbalke@techfak.uni-bielefeld.de

18. Januar 2001

1 Einleitung & Motivation

Die bisherigen Systeme, die Gestenerkennung zur Befehlsgebung unterstützen, können lediglich emblematische Gesten (siehe Kapitel 3) erkennen und auswerten. Diese Methode ist jedoch sehr unflexibel und verlangt vom Benutzer das Erlernen eines *Vokabulars*.

Die hier vorgestellten Ansätze interpretieren die natürliche konverbale Gestik des Menschen und werten deren Bedeutung im Kontext der erkannten Sprache aus. Dadurch soll die Mensch-Maschine-Kommunikation der zwischenmenschlichen Kombination ähnlicher werden.

2 Was heißt Gestenerkennung

Bei der Gestenerkennung wird der sog. Gesture-Space betrachtet, das ist der Raum vor dem Körper, in dem Gesten üblicherweise verübt werden, also Kopfhöhe bis Lendenhöhe vertikal und die Reichweite der Arme horizontal.

Die wichtigsten Werkzeuge sind die Datenhandschuhe, die die Gelenkwinkel der Hand erkennen. Zusammen mit den Sensoren, die die Lage der Körperteile im Raum erkennen, kann man daraus die Gesten errechnen.

Einige Projekte benutzen auch Eyetracker, da durch die Kombination aus Blickrichtung und Geste genauer bestimmt werden kann, welches das referenzierte Objekt ist.

3 Verschiedene Gestenkategorien

Man unterscheidet 4 Arten von Gesten:

- emblematische Gesten(Symbole): Bedeutung per Konvention festgelegt
- deiktische Gesten: Handlungen des Zeigens auf ein Objekt oder eine Region
- mimetische Gesten: Handlungen, die die Interaktion mit einem Objekt imitieren
- ikonische Gesten: Beschreibung der Form oder Lage von Objekten mit den Händen

4 ICONIC

4.1 Allgemeines

ICONIC wurde am MIT von Carlton J. Sparrel und David B. Koons ca. 1994 in Zusammenarbeit mit den anderen hier besprochenen Projekten entwickelt. ICONIC benutzt Spracheingabe und Datenhandschuhe zur Interpretation ikonischer und mimetischer Gesten, wobei die Gestik zum Auflösen von Ambiguitäten der Sprache verwendet wird.

4.2 Ziele des Systems

Ziel der Entwickler war es, ein intuitiv zu bedienendes multimodales System zu schaffen, das die Kommunikation mit dem Computer näher an die zwischenmenschliche Kommunikation heranführt.

4.3 Systembeschreibung

- **Datenhandschuh:** Die Raw-Daten des Handschuhs werden mit einem Zeitstempel versehen und darauf geprüft, ob zu dem vorhergehenden Zustand der Hand eine Änderung eingetreten ist. Ist dieses der Fall, werden die Raw-Daten in eine merkmalsbasierte Zwischendarstellung überführt, mit der dann weiter gearbeitet wird.
- **Sprache:** Der Spracherkenner zerlegt den erkannten Satz in eine Phrasenstrukturgrammatik und ordnet den einzelnen Wörtern Zeitstempel zu.
- **Interpreter:** Der Interpreter erhält die Daten vom Handschuh und dem Spracherkenner zugleich und verarbeitet diese dann parallel. Ist die Referenz der Sprache nicht eindeutig, wird geprüft, ob eine in dem fraglichen Zeitfenster vollführte Geste die Ambiguität auflösen kann.

4.4 gewonnene Erkenntnisse

Dadurch, daß ICONIC die natürliche koverbale Gestik des Benutzers interpretiert, und ihm damit erlaubt sich wie in einer zwischenmenschlichen Kommunikation zu verhalten, ist dieses System wesentlich mächtiger als die bisher betrachteten Ansätze, die auf eine bestimmte Anzahl von vorher festgelegten emblematischen Gesten beschränkt waren.

Dieses ist dadurch möglich, das der Interpreter die Gestik erst im Kontext der Sprache verarbeitet, und somit die gleiche Geste in verschiedenen Situationen einen unterschiedliche Bedeutungsgehalt trägt.

5 Andere Systeme am MIT

5.1 System 1

Dieses System wurde von Sparrel, Koons und Thorisson ca. 1994 entwickelt, und erlaubt dem Benutzer eine Manipulation einer 2D-Karte mittels Eyetracker, deiktischer Gesten und Sprache. Hierbei wird die Blickrichtung des Auges erkannt, um so eine genauere Zielerfassung als nur durch Interpretation einer deiktischen Geste zu ermöglichen. In diesem Ansatz werden die drei Eingangssignale getrennt voneinander geparkt, danach mit Zeitstempeln versehen und dann gemeinsam interpretiert.

5.2 System 2

Dieser Ansatz wurde parallel zu System 1 von den gleichen Entwicklern konstruiert, verfolgt jedoch ein anderes Ziel. In diesem Fall werden Sprache und deiktische, ikonische und mimetische Gesten erkannt, mittels derer der Benutzer eine 3D-Blockwelt verändern kann. Auch hier wird wie in ICONIC von den Raw-Daten des Handschuhs auf eine Zwischendarstellung abstrahiert, die *Gestlets* genannt wird. Diese dienen wiederum dazu, Mehrdeutigkeiten der Sprache aufzulösen.

5.3 System 3

Dieses Projekt wurde auch ca. 1994 am MIT entwickelt, allerdings von Alan Wexelblatt, der jedoch eng mit Koons, Sparrel und Thorisson zusammengearbeitet hat. Ziel war es, Hypothesen zu überprüfen, nach denen die Benutzung von Gestik abhängig von Geschlecht der Person, Inhalt des Wiedergegebenen und der Sprache (Personen, die nicht in ihrer Muttersprache kommunizieren, benutzen häufiger Gesten) ist.

Dazu wurde Versuchspersonen Szenen aus den Filmen *Casablanca* und *Terminator* gezeigt, die dann wiedergegeben werden sollten.

Das Ergebnis zeigte jedoch, daß keine der Annahmen bestätigt werden konnte, und damit keine Vorhersage über den Gebrauch von Gestik gemacht werden kann.

6 Zusammenfassung

Abschließend läßt sich sagen, das die 3D-Gestenerkennung das Potential zu einem mächtigen Werkzeug in der Mensch-Maschine-Kommunikation hat, da vom Benutzer kein Erlernen neuer *Befehle* verlangt wird. Gerade mimetische und ikonische Gesten tragen viele Informationen, die sich verbal nur unständig darstellen lassen. Dabei muß die Interpretation aller Modalitäten jedoch parallel erfolgen, da nur im Kontext erkannt werden kann, welchen Bedeutungsgehalt eine Geste trägt.

Literatur

- [1] Koons, D.B., et al.: Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures. In Maybury, M.T. (Ed.): Intelligent Multimedia Interfaces (pp. 257-276). Menlo Park (CA): AAAI Press/The MIT Press, 1993. [SA]
- [2] Sparrell, C.J., Koons, D.B.: Interpretation of Coverbal Depictive Gestures. AAAI Spring Symposium 1994, pp. 8-12, Stanford University. [F]
- [3] Wexelblatt, A.D.: An Approach to Natural Gesture in Virtual Environments. ACM Transactions on Computer-Human Interaction 2(3), 1995, 179-200