

Dialoggesteuerte Handlungskontrolle eines bimanualen Konstruktionsroboters

C. Scheering, B. Hildebrandt, A. Knoll
Universität Bielefeld, Postfach 100131, 33501 Bielefeld

10. August 1998

Keywords: Konstruktionshandlung, Sprachverstehenssystem, Interaktive Robotik, Instruktionen, inkrementelle Sprachverarbeitung, Multisensor-Interaktion, Mensch-Maschine-Interaktion

1 Einleitung

Wir entwickeln ein Robotersystem für Konstruktionshandlungen, das durch spontan gesprochene Anweisungen gesteuert wird. Die Architektur des Systems ist kognitiv motiviert (z. B. Zhang, Collani & Knoll 1998; Hildebrandt, Moratz, Rickheit & Sagerer 1995), wobei die technischen Rahmenbedingungen teilweise enge Grenzen setzen. Um die Vielfalt von Konstruktionsaufgaben einzuschränken, haben wir uns auf ein Szenario konzentriert, das zum einen übersichtlich genug und zum anderen ausreichend allgemeingültig ist. In dem ausgewählten Szenario erteilt der menschliche Instrukteur anhand eines Bauplans einem künstlichen Konstrukteur, d.h. einem Robotersystem, Anweisungen, um aus Holzbauteilen ein Modellflugzeug zu konstruieren.

2 Die Systemkomponenten

Die Hauptkomponenten des Systems sind ein Spracherkennungs- und ein Sprachverstehenssystem sowie ein multi-sensorgesteuertes Zwei-Arm-Robotersystem. Das Spracherkennungssystem basiert auf dem Diktiersystem IBM *Via Voice Gold*. Um die Erkennungsleistung zu erhöhen, wurde das Vollformen-Lexikon dem Vokabular des Szenarios angepaßt und das System um grammatische Bedingungen und einfache Wortkompositionsregeln erweitert. Das Spracherkennungssystem ist als Frontend zur Sprachverstehenskomponente angelegt. Falls eine Äußerung nur partiell korrekt erkannt wurde, ist in der Regel eine Interpretation durch die Sprachverstehenskomponente möglich.

Die Sprachverstehenskomponente basiert auf einer von Mark Steedman entwickelten Variante der Kategorialgrammatik, der *Combinatory Categorical Grammar* (z. B. Steedman 1987, 1996). Um kognitiven Verarbeitungsprinzipien der Sprachrezeption, wie z. B. der inkrementellen Verarbeitung, und der flexiblen Konstituentenstellung im Deutschen zu entsprechen, wurde die Grammatik modifiziert und an das Deutsche angepaßt. Dadurch können selbst komplexe und (lokal) ambige Anweisungen (z. B. *steck die lange Schraube in das zweite Loch von links der siebenlöchrigen Leiste*) effizient interpretiert werden. Aus den sprachlichen Anweisungen leitet die Sprachverstehenskomponente semantische Strukturen ab, aus denen wiederum elementare Operationen für den Roboter generiert werden. Hierbei wird die aktuelle Szenensituation berücksichtigt.

Die Arme des Konstruktionsroboters sind in einer Montagezelle hängend nebeneinander montiert. Dadurch wird ihnen ein den menschlichen Armen ähnlicher Aktionsradius erlaubt. Neben einem externen Multi-Kamera-System zur Objekterkennung und groben Lagepositionierung (Scheering 1998) ist auf jedem Arm eine „Eye-on-Hand“ Kamera montiert, durch die im Nahbereich mittels eines visuellen Regelungsprozesses Bauteile exakt gegriffen werden können (Meinicke 1997). Zusätzlich sind an den Handgelenken der Arme Kraftmomentensensoren montiert. Sie erlauben dem System, komplexe Montageoperationen wie Schrauben und Stecken als kraft-positions-geregelte Bewegungen zu steuern. Zusätzliche visuelle Informationen der Handkameras werden ebenfalls darin einbezogen (Collani 1998).

3 Die Kontrollarchitektur

Zur Steuerung des Roboters verwenden wir die in Abbildung 1 dargestellte, in drei Ebenen strukturierte Kontrollarchitektur:

- *Cognition:*

Hier findet die Interaktion mit dem menschlichen Instrukteur statt. Diese geschieht vornehmlich über gesprochene Sprache, geplant ist darüber hinaus die Integration von Zeigegesten. Hauptaufgabe dieser Ebene ist es, elementare Anweisungen aus der sprachlichen Äußerung unter Berücksichtigung des aktuellen Szenenzustandes für den Roboter zu generieren. Sollte eine Anweisung mehrdeutig sein, wird ein Dialog zur Verifikation angestoßen. Konnte eine Anweisung nicht von der unteren Ebene ausgeführt werden, wird ebenfalls in den Dialog mit dem Instrukteur getreten.

- *Scheduling:*

Auf der Ebene der Ablaufsteuerung wird eine Sequenz primitiver Roboteroperation entsprechend des aktuellen Roboterzustands zur Durchführung elementarer Anweisungen generiert. Aus einer von der Ebene *Cognition* gelieferten *Schraub*-Operation wird z. B. ein

1. `find_contact(nut, bolt)` (Stelle Kontakt her).
2. `find_thread()` (Finde die Gewindebohrung).

3. `find_notch_point()` (Finde den Gewindeeinstieg).
4. `screw()` (Schraube ein).

Bevor ein im Monitormodul erkannter Fehler an die übergeordnete Ebene *Cognition* weitergereicht wird, versucht die Ebene autonom, sie zu beheben. Der durch Anweisungen geänderte Szenenzustand wird automatisch aktualisiert.

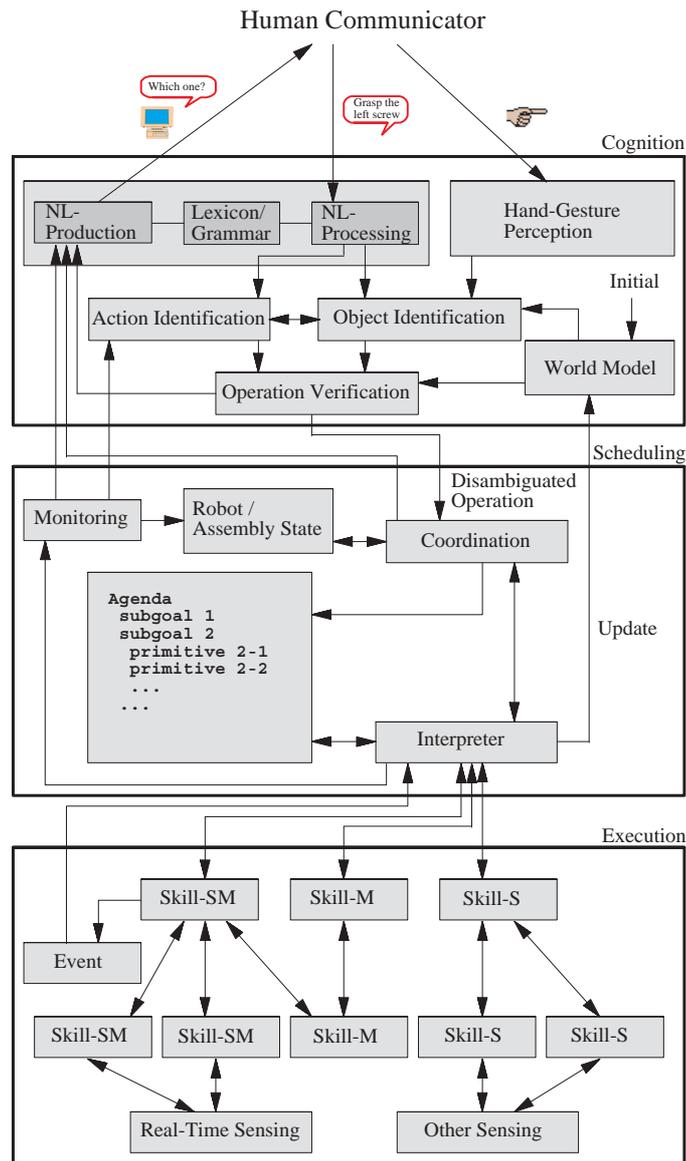


Abbildung 1: Kontrollarchitektur des Konstruktionsroboters

- *Execution:*

Der Interpretierer aus der vorherigen Ebene verwendet sogenannte Konstruktionstechniken (*skills*), um Sequenzen von elementaren Operationen abzuarbeiten. Wir unterscheiden dabei drei Arten von Skills:

- *Motorische Skills* sind direkte einzelne Roboterbewegungen, wie z. B. *open/close hand, move-to, move-joint-to, rotate hand*.
- *Sensorische Skills* stellen Information für die Ablaufebene oder andere sensorische Skills zur Verfügung. Wir unterscheiden dabei echtzeitfähige Skills (z. B. für Kraft/Momentenregelung, visuelle Regelung) von anderen Skills. Beispiele für sensorische Skills sind *get joint, get force, get torque, track object, detect moving robot, grab an image*.
- *Sensomotorische Skills* verbinden Perzeption und Aktion zumeist in Bewegungsregelungen, wie z. B. bei kraftüberwachten Bewegungen.

Zusätzlich werden auf dieser Ebene sogenannte Events generiert, um die Ausführungsebene über unerwartete Ereignisse wie z. B. Kraftüberschreitung, Singularitäten, Ableben eines Moduls etc. zu informieren.

4 Erhöhte Robustheit durch Interaktion und Redundanz

Während der Handlungsinterpretation und -ausführung haben wir es mit einem komplexen Interaktionsgeflecht zu tun. Bei einer unterspezifizierten Anweisung wie *steck die Dreilochleiste auf den Würfel* wird aufgrund semantischen Wissens inferiert, daß ein Instrument benötigt wird. Bezogen auf unser Szenario sind dies Schrauben. Sind mehrere Schrauben in der Szene, dann kann das System im Prinzip nachfragen, ob beispielsweise eine lange oder eine kurze Schraube erforderlich ist. Hält das Robotersystem die erforderlichen Objekte noch nicht „in den Händen“ sind vor der benannten Steckhandlung Greifhandlungen nötig. Erkennt der Instrukteur, daß der Roboter ein ungewolltes Objekt greifen will, kann er sprachlich intervenieren (*stop*) und eine laufende Handlung modifizieren. Eine weitere Intervention kann sich durch den Sensoreinsatz ergeben. Wird z. B. das zu greifende Objekt während der Feinpositionierung durch Einwirkung von Außen verschoben, regelt die visuelle Regelung sofort nach. Darüber hinaus erlaubt das System die Intervention während einer gefügigen Bewegung durch Anfassen des Manipulators. Durch die dabei auftretenden Kräfte wird die Position von außen beeinflusst. Durch den massiven Sensoreinsatz wird die Robustheit der Handlungskomponente stark erhöht. So wird z. B. durch die redundante Auslegung der externen Kameras sowohl eine bessere Positionierung erreicht als auch eine höhere Toleranz gegenüber Bildausfall (z. B. durch Verdeckung) und schlechter Signalqualität. Durch die Handkameras werden Fehler in Lage und Orientierung des Greifers weitgehend ausgeglichen. Trotzdem auftretende Greifungenauigkeiten werden bei den Steck- bzw. Schrauboperation durch die Kraftmomentensensoren ausgeglichen.

Literatur

- Hildebrandt, B., Moratz, R., Rickheit, G. & Sagerer, G. (1995). Integration von Bild- und Sprachverstehen in einer kognitiven Architektur. *Kognitionswissenschaft, 4*, 118-128.
- Meinicke, P. & Zhang, J. (1997). Calibration of a "Self-Viewing" Eye-on-Hand Configuration *Proceedings of the IMACS Multiconference on Computational Engineering Systems Applications, Lille, France, July, 1996*.
- Scheering, C. & Kersting, B. (1998). Uncalibrated Hand-Eye Coordination with a Redundant Camera System. *Proceedings of the IEEE International Conference on Robotics and Automation, Leuven, Belgium, 1998*.
- Steedman, M. (1987). Combinatory grammars and parasitic gaps In N. J. Haddock, E. Klein & G. Morrill (Eds.) *Edinburgh Working Papers in Cognitive Science, Vol.1, Categorical grammar, unification grammar and parsing* (pp. 29-70). Edinburgh: University of Edinburgh.
- Steedman, M. (1996) *Surface structure and interpretation*. Cambridge, MA: MIT Press.
- von Collani, Y., Zhang, J., & Knoll, A. (1998). A neuro-fuzzy solution for fine-motion control based on vision and force sensors *Proceedings of the IEEE International Conference on Robotics and Automation, Leuven, Belgium, 1998*.
- Zhang, J., v. Collani, Y., & Knoll, A. (1998). Development of a Robot Agent for Interactive Assembly *Proceedings of the 4th International Symposium on Distributed Robotic Systems, Karlsruhe, 1998*.