

Mensch-Maschine Interaktion auf der Expo.02

Dipl. Wirtsch.-Ing., Dipl.-Ing. Björn Jensen
 Dipl.-Ing. Roland Philippsen
 Dipl.-Ing. Gilles Froidevaux
 Prof. Dr. Roland Siegart

Swiss Federal Institute of Technology, CH-1015 Lausanne
 {bjoern.jensen, roland.philippsen, gilles.froidevaux, roland.siegwart}@epfl.ch

Man-Machine Interaction at Expo.02

The interactive mobile system RoboX is presented. Ten such robots will propose tours to visitors of the swiss national exposition during five month, seven days a week, ten hours per day. Algorithms are to find and track dynamic objects through the exposition are presented and evaluated. Thus, RoboX can initiate a conversation and propose tours to visitors, all the while reacting to their movement.

Kurzfassung

Dieser Beitrag beschreibt das interaktive mobile System RoboX. Auf der Schweizer Nationalausstellung Expo.02 werden zehn solcher Roboter dem Publikum Touren durch eine Ausstellung anbieten. In diesem Beitrag werden Algorithmen zum Finden und Verfolgen von bewegten Objekten vorgestellt und getestet, um zu Besucher gezielt anzusprechen und auf ihre Bewegungen zu reagieren.

1 Einleitung

Mit dem Beginn der Schweizer Expo.02 am 15. Mai 2002 öffnet die Ausstellung "Robotics" ihre Türen. Damit beginnen ebenfalls zehn mobile Roboter des Typs RoboX ihren Dienst. Als interaktive Museumsführer werden die am Autonomous Systems Lab der EPF Lausanne entwickelten Systeme in den folgenden fünf Monaten Tag für Tag mehreren tausend Besuchern der Nationalausstellung die wachsende Nähe und Präsenz robotischer Technologien vermitteln. In verschiedenen Szenarien, darunter auch von Roboter initiierten und geführten Touren wird für den Besuchern robotische Technologie durch intuitive Interaktion zum zentralen Erlebnis.

Das RoboX System wurde gezielt für den Einsatz an der Expo.02 entwickelt. Die Charakteristika dieser Nationalausstellung, wie hohe Besucherzahl, unmittelbarer Kontakt zwischen Roboter und Mensch, flossen direkt in Design und Entwicklung ein. Bereits seit einiger Zeit werden mobile Roboter in Museen eingesetzt [1,2,3,4,5]. Mit RoboX allerdings wurde ein vollständig autonomes System geschaffen, bei dem alle Algorithmen auf den mitgeführten Rechnern laufen. Zudem sollte RoboX leicht an andere Anwendungen anpassbar sein, weshalb Algorithmen und Umgebung unabhängig voneinander betrachtet wurden. Vermieden wurden insbesondere Anpassungen der Umgebung an die Algorithmen. Für die Interaktion ist die einzige notwendige explizite Information über die Umwelt eine topologische Karte der verschiedenen Exponate und deren Verbindungswege.

2 System

RoboX besteht aus einer autonomen Basis mit einem Durchmesser von 65 cm und einer Höhe von 68 cm, sowie einem interaktiven Aufsatz mit einem mechanischen Gesicht (Bild 1). Seine Gesamtgröße von 165 cm ermöglicht Besuchern eine gute Sicht auf den Roboter. RoboX verfügt über zwei Antriebs- und zwei Stützräder, wovon eines gefedert ist. Um trotz dieser nicht-holonomen Konfiguration einen möglichst großen Bewegungsspielraum zu haben, wurde eine oktagonale Form für die Basis gewählt, so daß der Roboter auf der Stelle drehen kann.

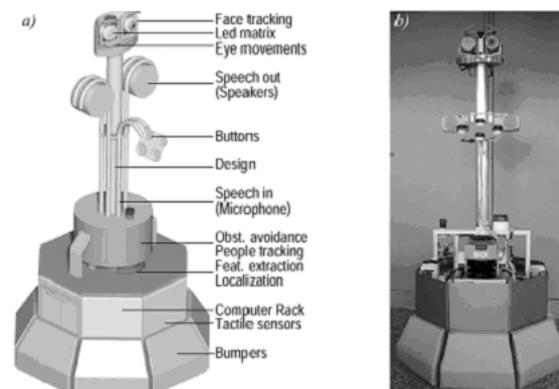


Bild 1: a) Elemente des interaktiven Roboters RoboX, b) der erste Prototyp.

Neben den Antrieben enthält die Basis Batterien, die eine Autonomie von zwölf Stunden erlauben, einen PowerPC 750 mit 400 MHz für Navigation [6] und Kollisionsvermeidung unter dem Realtime Betriebssystem XO/2 [9]. Außerdem gewährleisten rund um die Basis angebrachte taktile Elemente mit Schaumstoffkissen ein sofortiges Stoppen des Systems, sollte es trotzdem einmal zu einer Berührung kommen. Ebenfalls in der Basis befindet sich ein Pentium III 700 MHz, 128 MB RAM, 10 GB HD unter Windows 2000 für die Interaktion sowie zwei SICK Laser, deren Daten sowohl für Navigations- als auch für Interaktionsaufgaben verwendet werden.

3 Interaktion

RoboX Aufgabe ist es, den Besuchern Informationen zu den verschiedenen Exponaten zu vermitteln. Innerhalb von rund fünfzehn Minuten führt er Besucher durch eine Ausstellung und kann dabei aus vierzehn Stationen auswählen. Die Interaktion zwischen Mensch und Maschine ist somit zielgebunden, es geht darum, der Besuchergruppe in dieser Zeit die Ausstellung zu zeigen. Dabei gilt es, das Interesse der Besucher zu wecken und aufrechtzuerhalten, sowie dem Roboter die Bewegung durch die Menschenmenge von Station zu Station zu ermöglichen.

Jene Zielgerichtetheit spiegelt sich auch im Interaktionsprogramm von RoboX wieder. Es trennt die eigentliche Tour mit den Präsentationen der Exponate von Ausnahmesituationen, wie: Besucher spielen mit den Tasten des Roboters, der Roboter ist blockiert oder die Besuchergruppe bleibt stehen.

Die Frage „Wo sind die Besucher?“ ist somit zweifach bedeutsam. Erstens kann RoboX so gezielt einzelne Besucher oder Gruppen ansprechen und dann auch führen. Zweitens kann er auf Störungen, wie das Blockieren seines Weges, gezielt reagieren.

Zudem verfügt RoboX über ein ganzes Repertoire von Möglichkeiten, Besucher anzusprechen und sich auszudrücken: Synthetisierte Sprache in französisch, englisch, italienisch und deutsch [12,13], Bewegungen des Roboterkörpers, mechanischen Augen und Augenbrauen zum Imitieren von Gesichtsausdrücken und nicht zuletzt die Symbole auf seiner LED-Matrix (Bild 2).

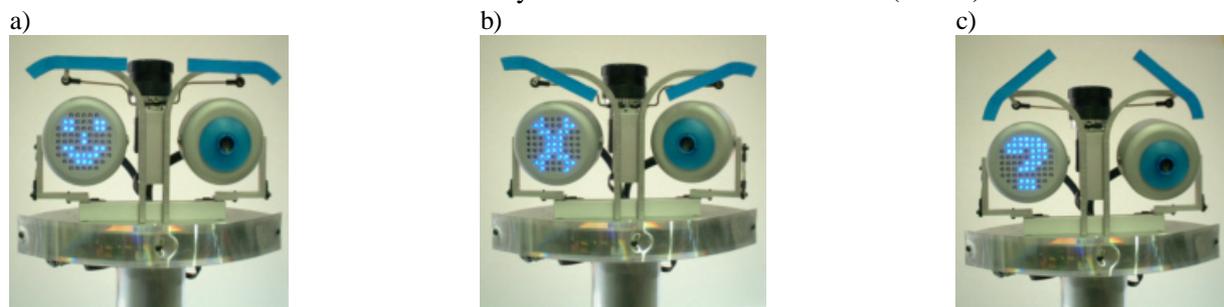


Bild 2 Von RoboX imitierte Gesichtsausdrücke a) neutral freundlich, b) verärgert, c) überrascht.

Um den Besuchern die Möglichkeit zu geben, mit dem System zu kommunizieren haben alle Roboter kapazitive Einagbetaster, die beispielsweise zur Wahl der Sprache oder der nächsten Station verwendet werden können. Darüber hinaus verfügen einige Roboter über eine sprecherunabhängige Einzelwort-Spracherkennung in den oben genannten vier Sprachen [11].

4 Wo sind die Besucher?

Ein wichtiges Element überzeugender Mensch-Maschine Interaktion ist die Detektion von Besuchern. Dadurch kann erreicht werden, das RoboX diese gezielt anspricht und auf ihre Bewegungen entsprechend reagieren kann. Vermieden wird das Problem, daß der Roboter Informationen herunterleiert, ohne daß ihm jemand zuhört.

Es gibt eine Vielzahl von Möglichkeiten, Besucher in einer Ausstellung zu finden. Von den Sensoren des RoboX Systems bieten sich Laser Sensor und Farbkamera an. Wichtig bei der Interpretation der Sensordaten ist, daß möglichst nur Besucher als solche klassifiziert werden. Es muß daher sorgfältig eine Balance gefunden werden zwischen einer hohen Erkennungsrate – RoboX reagiert auf jeden Besucher – und einem geringen Prozentsatz an Fehlern – Säulen, Exponate werden für Besucher gehalten.

Unser Ansatz, Besucher zu finden und zu verfolgen, ist demzufolge zweistufig. Er fusioniert die Informationen von Laser Sensoren und Farbkamera. In einem ersten Schritt werden dabei die Laserdaten untersucht, um Bewegungen in der Umgebung des Roboters zu finden. Dies ermöglicht dem Roboter eine „Rundumsicht“, und er kann auf alle Besucher reagieren.

Sind die Besucher nahe genug, wird die Farbkamera benutzt, um die zu den Bewegungen korrespondierenden Gesichter zu finden. Durch ein Nachführen der Kamera kann deutlich gemacht werden, wer Adressat seiner Handlungen ist. RoboX sieht seinen Gesprächspartner an. Spricht er zu einer Gruppe, so kann er seinen Blick über sie schweifen lassen. Interagiert er mit einem Besucher, so wird er diesen immer im Blick halten.

4.1 Detektion bewegter Objekte

Zur Detektion und Verfolgung von bewegten Objekten in der Umgebung eines Roboters existieren bereits mehrere leistungsfähige Vorschläge [7,8]. Unsere Variante zielt auf eine Bewegungsdetektion mit möglichst geringem Rechenaufwand ab, da alle Berechnungen auf dem von RoboX mitgeführtem PC erledigt werden müssen.

Es wird dabei davon ausgegangen, daß der Fehler beim Schätzen der Positionsänderung des Roboters zwischen zwei Scans vernachlässigbar klein ist. Die Umgebung wird in statische und dynamische Objekte eingeteilt. Diese lassen sich aus den Scandaten ermitteln. Dazu wird für jeden Winkel die maximal gemessene Entfernung in einer statischen Karte abgelegt, welche in Polarkoordinaten auf dem Roboter zentriert ist. In jedem neuen Scan lassen sich dynamische Elemente als diejenigen identifizieren, die eine signifikante Differenz zu den korrespondierenden Elementen der statischen Karte aufweisen. Liegen die aktuellen Elemente näher als die der statischen Karte, werden sie direkt als dynamisch gekennzeichnet. Im umgekehrten Fall findet ein Austausch der Elemente statt: die statische Karte enthält danach das aktuelle Element und der aktuelle Scan das Element aus der Karte, welches nun als dynamisch markiert wird.

Zur Zeit arbeiten beide Scanner mit einer maximalen Reichweite von 8 Metern bei einer Auflösung von einem Millimeter und einer Winkelauflösung von einem halben Grad. Bewegungen können somit durch einen Vergleich von nur 720 Werten gefunden werden.

In einem nächsten Schritt werden von den dynamischen Elementen räumlich zusammenhängende Gruppen gebildet. Es ergibt sich so eine Zuordnung zu den bewegten Objekten. Durch die Platzierung des Lasers auf Kniehöhe kann erwartet werden, daß die Mehrzahl der so gefundenen Objekte Beine der Besucher sind. Der Vergleich der Scandaten mit der statischen Karte erlaubt, auch Objekte zu finden, wenn sie nach einer Bewegung zum Stillstand gekommen sind.

Für einen statischen Roboter können so alle Bewegungen gefunden werden, welche auf den Roboter zu oder parallel zu ihm verlaufen. Bewegungen von ihm weg können nur durch einen Vergleich der aktuellen mit der vorhergehenden statischen Karte, also mit einer Verzögerung, gefunden werden.

Für den bewegten Roboter wird die Karte statischer Elemente auf die neue Roboterposition transformiert. Bei eventuellen Überdeckungen werden immer die dem Roboter nächsten Elemente behalten. Winkel, für die nach der Transformation keine Werte ermittelt werden können werden als frei gekennzeichnet und mit dem nächsten Scan initialisiert. Als letzten Schritt vor der Klassifizierung als dynamisches Element ist nur noch zu überprüfen, ob dieses Element im vorhergehenden Scan eventuell verdeckt war. Ist dies nicht der Fall, kann es als dynamisch markiert werden.

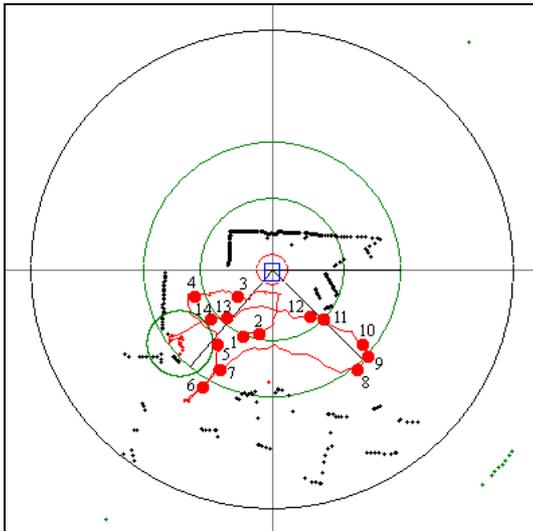
Mit der Unterscheidung der Scanelemente in dynamische und statische Objekte kann RoboX bereits gezielt auf Besucher zufahren. Für eine komplexe Interaktion ist es jedoch wünschenswert, möglichst viel über die Bewegung der Menschen zu erfahren, um entsprechend darauf reagieren zu können.

Diese Verfolgung der Bewegung wird mit einem Kalman-Filter realisiert, der mit der Position eines dynamischen Objektes initialisiert wird. Da dynamische Objekte das linke, das rechte oder beide Beine repräsentieren können, wird hier eine relativ hohe Beobachtungsunsicherheit zugrundegelegt.

Im nächsten Schritt wird mittels eines Nearest-Neighbour Verfahrens aus den dynamischen Objekten das am besten passende für die nächste Positionsschätzung verwendet. Sollte es zu weit entfernt sein, findet keine neue Schätzung statt und die Positionsunsicherheit wird entsprechend erhöht.

Um die Bewegung des Besuchers interpretieren zu können wird die Roboterumgebung in verschiedene Sektoren eingeteilt. Diese können als zusätzliche Information in die Mensch-Maschine Interaktion einfließen. RoboX unterscheidet beispielsweise zwischen zu nah, nah, weit und zu weit, sowie vorne, links, rechts und hinten. Ein Besucher befindet sich immer in einem solchen Sektor. Mittels der Sprachausgabe kann RoboX Besucher bitten, sich wieder vor ihn zu stellen oder weiterhin seiner Tour zu folgen. Bild 3 zeigt, wie ein Besucher über mehrere Minuten verfolgt wurde, sowie die von RoboX generierten Kommentare.

a)



b)

Tracker is searching...

1. Found.
2. close in front of me.
3. close on my right side.
4. on my right side far away.
5. far away in front of me.
6. in front of me, too far away.
7. far away in front of me.
8. in front of me, too far away
9. on my left side, too far away.
10. on my left side far away.
11. close on my left side.
12. close in front of me.
13. close on my right side.
14. on my right side far away.

Bild 3 a) Bewegungsdetektion mittels Laser Scanner Daten. Der Roboter ist in der Bildmitte zu sehen. Die konzentrischen Kreise zeigen die verschiedenen Sektoren (nah, weit, zu weit). Die beiden diagonalen Linien grenzen den linken, den vorderen und den rechten Sektor voneinander ab. Der äußerste Kreis entspricht dem Sensorhorizont (8 m), Punkte, die ausserhalb liegen, wurden als nicht zuverlässige Messungen klassifiziert. Die durchgezogene Linie ist die Bewegungstrajektorie des Besuchers. Der kleine Kreis links unten im Bild gibt die Akzeptanz Region für die nächsten Messungen an. b) Von RoboX generierte Kommentare zur Bewegung des Besuchers.

4.2 Verfolgung von Gesichtern

Zur Verfolgung von Gesichtern in Bildsequenzen gibt es vielfältige Ansätze. Besondere Herausforderungen beim Einsatz auf einem mobilen System ergeben sich durch die Bewegung der Kamera und der Veränderung der Beleuchtungssituation. Der Algorithmus zur Gesichtsverfolgung für RoboX sind angelehnt an das System in [10]. Er erkennt und verfolgt hautfarbene Regionen in der Bildsequenz. Je nach gewähltem Modus kann die Kamera im Auge von RoboX dann so nachgeführt werden, daß entweder nacheinander alle hautfarbenen Regionen in die Bildmitte kommen oder daß permanent eine Region in der Bildmitte bleibt.



Bild 4 Ergebnisse der Gesichtserkennung, detektierte Gesichter sind im dunklen Rahmen dargestellt.

Der Algorithmus zur Gesichtsverfolgung lässt sich in folgende Schritte teilen:

1. **Klassifizierung nach Hautfarben:** Um Helligkeitsunterschiede mindestens teilweise auszublenden, werden die RGB-Bilder der Kamera zuerst normalisiert. Dazu werden die Verhältnisse von Grün und Blau zu Rot gebildet. Als Hautfarben werden fest definierte Bereiche dieser Verhältnisse akzeptiert und so ein Binärbild erzeugt. Kleinere hautfarbene Regionen, die nicht sinnvoll als Gesicht interpretierbar sind, werden mittels morphologischen Filtern unterdrückt.
2. **Extraktion von Kontouren:** Jede verbleibende Region im Binärbild bekommt einen eindeutigen Identifikator. Um hautfarbene Regionen, die keine Gesichter sind, zu unterdrücken, werden mehrere heuristische Filter verwendet. Sie basieren auf der Größe des Rechtecks, seinem Seitenverhältnis, dem Anteil Hautfarbe an der Fläche des Rechtecks und der Morphologie der Regionen. Das heißt, daß wir für Gesichter an der Stelle von Augen oder Mund innerhalb der hautfarbenen Region Löcher erwarten.
3. **Verfolgung:** Für bereits gefundene Regionen versucht das System die neue Position zu bestimmen. Dazu wird aus den aktuellen Regionen, diejenige ausgewählt, welche der alten Position am nächsten ist. Die aktuellen Regionen, welche nicht zugeordnet werden können werden für das nächste Bild beibehalten. Um angezeigt zu werden müssen sie über mehrere Bilder verfolgt werden.

5 Tests der Verfolgung von Gesichtern

Der Gesichtsverfolgungsalgorithmus wurde, während einer Demonstration des RoboX Systems in unserem Labor umfassend getestet. Beispiele sind in Bild 4 zu sehen. Aus der Sequenz von 2800 Bildern, die mit 4 Hz abgelegt wurden, wurde jedes zwanzigste Bild manuell klassifiziert. Es wurde unterschieden zwischen Bildern mit hoher und niedriger Bildschärfe, sowie dunklen Bildern. Für alle untersuchten Bilder wurden die Anzahl der Gesichter in der Aufnahme, sowie die Anzahl der hautfarbenen Regionen, die keine Gesichter sind ermittelt.

	Anz. Bilder	Anz. Gesichter	Anz. Regionen	Anz. Falsch	Erkannte Gesichter
hohe Bildschärfe	100	584	375	37	64.21%
geringe Bildschärfe	39	193	88	0	45.60%
dunkles Bild	30	270	34	0	12.59%
Gesamt	169	1047	497	37	47.47%
Geschätzt Sequenz	2800	17347	8234	613	47.47%

Tabelle 1 Auswertung einer 11 minütigen Bildsequenz mit 20 Besuchern. Die Bilder wurden unterteilt in hohe und niedrige Bildschärfe, sowie dunkle Bilder um die Performance der Gesichtsverfolgung für verschiedene Situation ermitteln zu können.

Die Auswertung (Tabelle 1) konnte auf jedes zwanzigste Bild beschränkt werden, da direkt aufeinanderfolgende Bilder nur selten Unterschiede hinsichtlich der Anzahl der vorhandenen und der erkannten Gesichter aufweisen. Die dunklen Bilder liegen ausschliesslich zu Beginn der Sequenz (RoboX begrüßt die Besuchergruppe). Hier fanden sich durchschnittlich neun Gesichter pro Bild, während in der gesamten verbleibenden Sequenz zwischen fünf und sechs Gesichter pro Bild zu sehen sind.

Die 169 untersuchten Bilder zeigen 1047 Gesichter, wovon 497 richtig erkannt wurden, 37 Regionen korrespondierten nicht mit Gesichtern, was einer Fehlerquote von 3.53% entspricht. Demgegenüber steht eine Erkennungsquote von durchschnittlich 47.47% und sogar von 64.21% bei Bildern mit hoher Bildschärfe. Diese fällt in dunklen Bildern auf 12.59%. Gründe hierfür sind vermutlich in den festen Akzeptanzbereichen hautfarbener Bereiche für die Verhältnisse Grün und Blau zu Rot und der unzureichenden Kompensation für die Helligkeitsänderung zu suchen.

6 Zusammenfassung

Im vorliegenden Beitrag wurde das interaktive mobile Roboter-System RoboX vorgestellt. Da die Aufgabe, Besucher im Umfeld des Roboters zu finden, für eine überzeugende Interaktion ein wichtiges Element ist, wurden Algorithmen zum Finden und Verfolgen von Besuchern entwickelt. Dazu wurden die Information von Laser Sensoren und Farbkameras fusioniert. In Labortests konnten Besucher über mehrere Minuten verfolgt und ihre Bewegung entsprechend kommentiert werden. Gesichter werden im Mittel in 47.47% richtig erkannt. Mit einer Erkennungsquote von 64.21% bei hoher Bildschärfe und 45.60% bei geringer Bildschärfe und einer Fehlerquote von nur 3.53% fälschlich als Gesichter klassifizierten Regionen haben wir ein System, das zuverlässig Besucher ansprechen kann.

Literatur

- /1/ Willeke T., Kunz C., Nourbakhsh I., "The History of the Mobot Museum Robot Series: An Evolutionary study", AAAI, 2001.
- /2/ S.Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: A Second-Generation Museum Tour-Guide Robot", Proc. ICRA, 1999.
- /3/ Burgard W., Cremers A., Fox D., Hänel D., Lakemeyer G., Schulz D., Steiner W., Thrun S., "Experiences with an interactive museum tour-guide robot", AI (114), 1999, pp. 3-55.
- /4/ I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, "An Affective Mobile Educator with a Full-time Job," AI, Vol. 114, No. 1 - 2, October, 1999, pp. 95 - 124.
- /5/ Graf, B., Baum, W., Traub, A., Schraft, R.D.: "Konzeption dreier Roboter zur Unterhaltung der Besucher eines Museums". In VDI-Berichte 1552, pp. 529-536, 2000.
- /6/ Arras K.O., Tomatis N., Jensen B., Siegwart R., "Multisensor On-the-Fly Localization: Precision and Reliability for Applications," Robotics and Autonomous Systems, vol. 34, issue 2-3, pp. 131-143, February 2001.
- /7/ Prassler E., Scholz J., Elfes A., "Tracking People in a Railway Station During Rush-Hour", In Proc. of the First International Conference, ICVS'99, Berlin, Germany, pp. 162-179, 1999.
- /8/ Schulz D., Burgard W., "Probabilistic State Estimation of Dynamic Objects with a Moving Mobile Robot", Robotics and Autonomous Systems 34, pp.107-115 (2001).
- /9/ Brega R., Tomatis N., Arras K.O., "The Need for Autonomy and Real-Time in Mobile Robotics: A Case Study of XO/2 and Pygmalion," In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Takamatsu, Japan October 30-November 5, 2000.
- /10/ Hilti A., Nourbakhsh I., Jensen B., Siegwart R., "Narrative-level visual interpretation of human motion for human-robot interaction", In Proceedings of IROS 2001. Maui, Hawaii, October 29 - November 3, 2001.
- /11/ Prodanov, P.; Drygajlo, A.; "Voice Enabled Interfaces for Interactive Tour-Guide Robots", submitted to IROS 2002.
- /12/ Siebenhaar-Röllli, B., Zellner Keller, B., & Keller, E., "Phonetic and Timing Considerations in a Swiss High German TTS System". In: E. Keller, G. Bailly, A. Monaghan, J. Terken & M. Huckvale, "Improvements in Speech Synthesis", pp. 165-175, Wiley & Sons, 2001.
- /13/ Keller, E., & Zellner, B., "A timing model for fast French. York Papers in Linguistics", University of York. pp. 53-75, 1996.