

ROMAN: A Mobile Robotic Assistant for Indoor Service Applications

U. D. Hanebeck, C. Fischer, G. Schmidt
Institute of Automatic Control Engineering (LSR)
Technische Universität München
80290 München, GERMANY
e-mail: {hnb,fischer,gs}@lsr.e-technik.tu-muenchen.de

Abstract

The paper describes design issues of a mobile service robot for health care applications and domestic automation. Key components required for achieving semi-autonomous operation are surveyed, including 1. a highly maneuverable locomotion platform, 2. an anthropomorphic manipulator, 3. a reliable multi sensor system, and a 4. multi-modal huMan-Robot-Interface. In addition, related information processing and control methodologies are presented. Special emphasis is put on a system architecture for integration of the individual components into a full-size service robot. Performance and usefulness of the proposed approaches are demonstrated through experiments in various real-world service scenarios.

1 Introduction

This paper describes ROMAN, a personal robotic assistant for indoor service applications. ROMAN is aimed at relieving humans of time consuming routine tasks. In this context, two main application domains are considered. The first domain, health care services, is approached with the limitation of avoiding major physical interaction with humans. Basic tasks considered are fetch-and-carry, surface cleaning, disinfection, distribution of meals, and simple environment management.

The second domain refers to domestic automation in homes and offices. In this scenario, a robot could perform tasks like mail distribution, house keeping, vacuum cleaning, inspection of buildings, and night watchman duties.

The motivation behind the ROMAN project was to develop key components and integrate them into a full-scale service robot for application to typical service tasks.

Since the robot's workspace is designed for humans, the robot should be human-sized and human-friendly. Furthermore, it should perform anthropomorphic motions. These requirements do not only pose new

challenges to the locomotion, manipulation, perception, and communication capabilities of the robot. In addition, new concepts for the system architecture had to be developed for achieving a satisfying degree of reliability and safety of operation.

The work described in this paper is confined to semi-autonomous operation, since appropriate artificial intelligence techniques for fully autonomous operation are not yet available. However, this approach proves to be very useful and has been demonstrated by various experiments.

To meet the hardware requirements, a human-sized robot has been implemented, which is based on an omnidirectional wheeled locomotion platform. This results in high maneuverability combined with a high maximum velocity of about 2 m/sec. The robot is equipped with an anthropomorphic manipulator arm well suited for human-oriented tasks like door opening or object handling. An onboard multi sensor system is used for object recognition, self-localization, and obstacle detection. For commanding purposes and for monitoring task execution, a multi-modal huMan-Robot-Interface has been developed. It comprises several input/output media and a virtual workspace.

Information processing and control algorithms have been developed for the following tasks.

Coordination of locomotion and manipulation

The execution of a typical service task may be divided into two major phases: 1. Long distance locomotion for connecting local workspaces. 2. Mobile manipulation, i.e., extending the manipulator workspace by support of the locomotion platform. Both phases are supported by sensor information in a closed-loop fashion.

Sensor support of motion/manipulation tasks

A robot self-localization method has been developed, that yields precise real-time posture estimates despite varying environmental conditions and the use of simple sensors. Furthermore, a new approach was

designed for the reliable and fast detection of unknown obstacles in the robot's workspace.

To fulfill the diverse requirements for object recognition during manipulation, two approaches have been considered. On one hand, a feature-based approach is used for extended objects with explicit 3D-models available. On the other hand, a cross-correlation-based approach is used for reliable recognition of small objects with cluttered background.

HuMan-Robot-Interface

Two borderline cases of a huMan-Robot-Interface have been considered: 1. An expert user interface with cockpit-like appearance including visual, acoustic, and tactile feedback. 2. A basic interface for untrained users, which solely relies on a simple natural language dialog to specify desired robot tasks.

2 Information Processing Architecture

The physical components are taken care of by software agents, called expert modules, for

- the huMan-Robot-Interface,
- mobile manipulation,
- object recognition,
- self-localization and obstacle detection,

cf. Fig. 1. Communication between the expert modules takes place via a blackboard server and a dynamic communication switch. The blackboard server is used for posting data of general interest, such as the current robot posture, as supplied by the localization expert. In contrast, the communication switch is used to temporarily establish communication channels between experts depending on current task and situation. During door opening maneuvers, for example, the object recognition expert is connected with the expert for mobile manipulation.

For semi-autonomous execution of service tasks, a human operator issues the desired action by means of natural language commands. In case of speech recognition problems, missing details, or unknown situations, a dialog is initiated by the robot in order to obtain a complete command. This robot command is subsequently split up into basic task primitives like door opening, door passing, or traveling along hallways. The task composed of these stereotypic operations is executed by either the long distance locomotion expert or the mobile manipulation expert. Both are continuously supported by current sensor information.

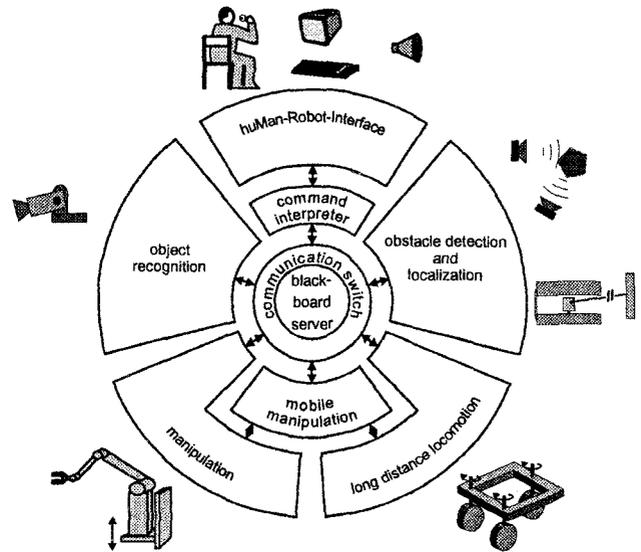


Figure 1: Information processing architecture.

3 Motion Control of a Mobile Manipulator

Service tasks are executed by either of two expert modules: 1. Long distance locomotion for approaching the desired object without performing manipulation tasks. 2. Mobile manipulation for handling objects, which may itself be divided into \diamond *arm preparation phase*, \diamond *path following phase*, and \diamond *arm completion phase*. During the arm preparation phase, the arm is folded out. In the path following phase, the end-effector follows a predefined path, where the limited manipulator workspace is extended by locomotion. Finally, a transport configuration of the manipulator is attained during the arm completion phase.

3.1 Long Distance Locomotion

This expert module is used for connecting local manipulation workspaces via long distance maneuvers.

Path planning is based on a predefined road map, i.e., a topology graph, which connects typical workspaces in the environment. Each node specifies the desired robot position; a directed edge between two nodes defines \diamond robot orientation, \diamond passing direction at the goal node, and \diamond the robot velocity between the nodes.

For connecting two arbitrary robot postures, the closest graph nodes are selected as start and goal respectively. The shortest path between start and goal through the road map is generated using Dijkstra's algorithm. Finally, the resulting via-points are connected by applying two competing methods:

Reference Trajectory Based Motion Primitives lead to an explicit connection of via-points, (Fig. 2 a). For this purpose, cartesian and polar splines

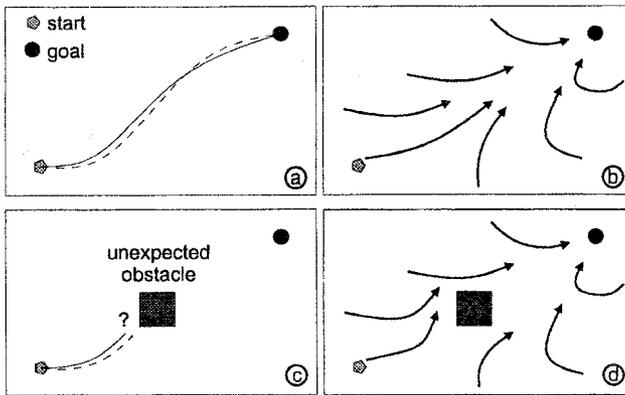


Figure 2: Comparison of a), c) trajectory based and b), d) model based motion primitives: a), b) nominal case, c), d) unexpected obstacle.

of fifth order are used. The main problem with this approach is the need for replanning when obstacles are encountered. However, it is useful for tasks, where an explicitly known path is required, e.g., in floor cleaning operations.

Reference Model Based Motion Primitives achieve reactive maneuvers using all three degrees-of-freedom (DOF). Appropriate maneuvers for given situations are generated by a fuzzy reference robot model, which is based on user-formulated linguistic rules. To guarantee stable model following by the real robot, a nonlinear controller has been designed based on Lyapunov stability theory. An advantage of this approach is the ease of handling unexpected obstacles, Fig. 2 d).

Both methods are based on the concept of a virtual locomotion platform, Fig. 3. An abstract command interface is used, where incremental robot motions are defined by the instantaneous center of rotation (κ_1, ϕ_1) and the velocity v_1 of the virtual wheel. These commands are converted to steering angles and drive wheel velocity by means of a wheel coordination scheme. This abstract interface allows the expert modules to fully exploit the omnidirectional capabilities of the robot without the need for considering its physical characteristics.

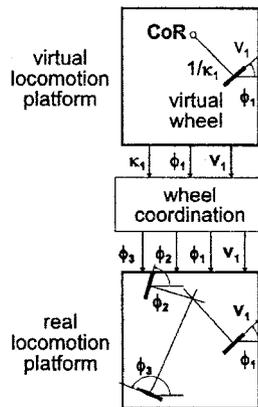


Figure 3: Virtual locomotion platform and wheel coordination.

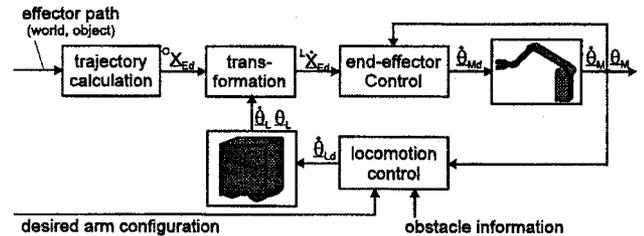


Figure 4: Scheme for control of end-effector trajectory.

In addition, neural network based approaches for vision-guided parking and docking of mobile robots have been developed [1].

3.2 Mobile Manipulation

This section focuses on assisting the manipulator (M) during the *path following phase* by use of the additional DOFs provided by the locomotion platform (L).

A mobile manipulator comprises n_M manipulation DOFs and n_L locomotion DOFs, i.e., a total of $n = n_M + n_L$ DOFs. In ROMAN's case $n_M = 6$ and $n_L = 4$, since the omnidirectional locomotion platform is vertically extended by a high/low linear axis carrying the manipulator arm. With respect to a 6-dimensional task space, $m = 6$, a degree-of-redundancy (DOR) given by $r = n - m = 4$ may be defined for the mobile manipulator ROMAN.

The proposed mobile manipulation strategy is based on the natural separation of the two motion subsystems, locomotion and manipulation. Furthermore, two assumptions are required: 1. The manipulator is capable of operating in the entire local task space. 2. The locomotion platform is omnidirectional. Hence, it is used to extend the manipulator workspace during task execution and ensure high manipulability.

The manipulator joints θ_M are assigned to control the end-effector motion. The locomotion states θ_L control internal motions optimizing the overall motion behaviour [2], while considering environmental constraints, Fig. 4.

The locomotion control strategy is assumed to support the manipulation unit continuously during the path following phase. For this purpose, the following objectives are considered:

- avoidance of static and dynamic obstacles,
- online variation of the desired motion behaviour,
- consideration of joint angle limits,
- human-like, smooth motion.

Typical human behaviour includes the avoidance of ill-posed configurations. To achieve similar performance with ROMAN, the remaining locomotion DOFs are

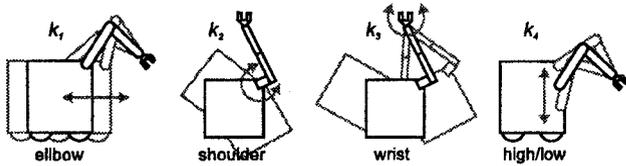


Figure 5: Visualization of orthogonal arm configuration criteria $k_1 \dots k_4$.

controlled with respect to $r = 4$ orthogonal configuration criteria $k_1 \dots k_4$, Fig. 5:

- k_1 : Elbow-angle, i.e., defined by distance between end-effector and locomotion platform.
- k_2 : Shoulder-angle, i.e., arm position with respect to the locomotion platform.
- k_3 : Wrist-bend angle, i.e., angle between end-effector and forearm.
- k_4 : Shoulder-height, i.e., shoulder/floor distance.

A set of desired criteria k_d might be prescribed by a higher-level task planner to ensure either high manipulability or to meet task-related constraints. For example, for object recognition purposes, it might be required to point the onboard CCD-camera towards the desired object by means of appropriate platform motion.

A soft locomotion control scheme is employed for supporting the manipulator arm such that the desired configuration criteria are met best possible. The control scheme consists of parallel blocks, where each block represents an independent fuzzy inference mechanisms taking care of a single criterion. Control output provides locomotion state vectors resulting in smooth, human-like motion patterns.

4 Robot Self-Localization

A new approach has been developed for estimating the current location (position and orientation) of a fast mobile robot based upon data from an onboard multi sensor system.

Localization could in principle be performed by dead-reckoning, i.e., by propagating a given initial location with data from incremental sensors like an odometer and/or a gyroscope. However, due to systematic and correlated uncertainties in conjunction with the required integration, dead-reckoning suffers from accumulating errors. Thus, its accuracy is sufficient for small path lengths only.

To keep the localization accuracy within specified bounds though, additional geometry sensors are required for repetitively updating the robot location with respect to certain environmental features. In ROMAN's case, two types of inexpensive and simple

sensors are considered: Angle measurement systems and range sensor arrays with an arbitrary geometry. Their limited perception capabilities are compensated by intelligent sensing strategies.

An overview of the proposed localization system is given in Fig. 6. It comprises two phases: 1. The initialization phase for determining the initial robot location with little a priori knowledge. 2. The continuous localization phase for updating the robot location during motion.

A simple and efficient initialization procedure has been developed [3], which is based on an interpretation tree approach and a new linear solution scheme for converting angle measurements to robot locations. It copes with \diamond indistinguishable landmarks, \diamond erroneous measurements, \diamond landmark occlusions, and, in addition, makes use of prior knowledge.

To ensure accurate and stable in-motion robot localization, despite the limited perception capabilities of the sensors considered, both spatial and temporal continuity conditions are exploited. For that purpose, incremental sensors are used to predict the robot location and to formulate landmark hypotheses. Based on these hypotheses, the sensor system is configured online, i.e., sensing and processing power are focussed on useful landmarks at an early stage. In addition, measurement hypotheses are derived for validating actual measurements. As a result, environmental features can be reliably sensed with high sampling rates [4].

To increase localization accuracy, efficient closed-form solutions have been derived for calculating the robot location based on both angle and distance measurements. These solutions include the consideration of both measurement errors and landmark position uncertainties as an integral part [5, 6].

The measured robot location is fused with the predicted location by set-theoretic methods. The result of fusion serves as the basis for the next prediction step.

One of the unique features of this approach is the rigorous modeling of uncertainties by just specifying error bounds. This is advantageous, when detailed stochastic models are not appropriate, not available, or too complicated. Rather than propagating point estimates, the algorithm propagates all feasible states that are compatible with the a priori error bounds for landmark position uncertainties and measurement noise. The inherently high complexity of this approach is reduced by approximating the sets of feasible states with ellipsoidal sets. A new formalism for ellipsoid calculus yields simple and computationally efficient algorithms for prediction and fusion.

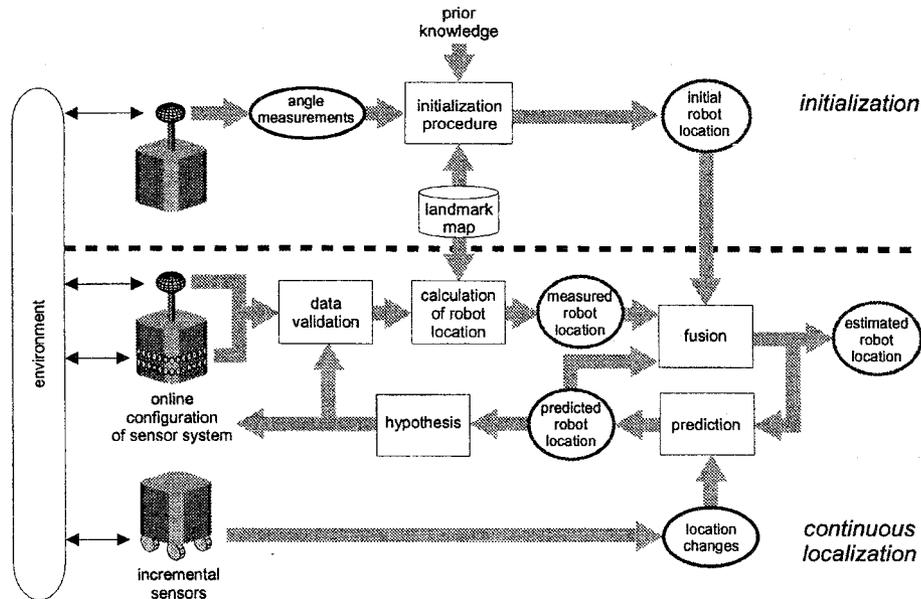


Figure 6: Scheme of ROMAN's localization system.

5 Vision Based Object Recognition

During the execution of service tasks, a collection of a priori known objects needs to be manipulated. These objects are recognized and precisely localized by means of vision techniques.

On the one hand, extended objects like doors, cupboards and drawer boxes are considered. These objects are composed of simple features like straight lines or planes. On the other hand, fetch-and-carry tasks often require grasping of small objects with cluttered background. Because of these diverse requirements, two approaches for object recognition are used:

A **feature-based approach** [7] considers extended objects, for which explicit 3D-models are available. Since recognition is based on a 2D perspective view, in general a full 3D match needs to be performed. This would be prohibitively expensive for real-time applications. Hence, a simplified three-step recognition procedure has been developed.

In a first step, straight lines are extracted from the image by means of standard techniques. These lines are then mapped onto a predefined reference plane by means of an inverse perspective transformation. In case of a drawer box, for example, the lines representing the top side can be recovered without any perspective distortions. Based on these transformed features, a simple 2D matching procedure is used to recognize and localize the object.

This approach is useful, when a reference plane can be defined, which is the case in many practical situations.

A **correlation-based approach** [8] is used for recognizing small objects with cluttered background like cups, glasses, and bottles. For each object, typical scenes are analyzed offline with the result of generating a set of object templates, and in addition a set of background templates.

Online, a search window is shifted through the current image on a pixel-by-pixel basis. After each step, the window is compared with the object and the background templates. The number of required comparisons is reduced by using an optimized binary search tree.

One advantage of this approach is, that no explicit object models are required. Furthermore, even for a large number of typical object views only a small number of comparisons is necessary. As a result, the procedure is capable of real-time operation. To further speed up the procedure and to increase its reliability, a priori knowledge about typical object postures and typical backgrounds is considered while generating templates.

6 HuMan-Robot-Interface

The huMan-Robot-Interface (MRI) developed is used for commanding the service robot, monitoring its operation, and providing goal-oriented support during task specification and task execution [9]. Furthermore, the robot generates comments about the state of task execution and informs people in the workspace about its future actions.

Human-friendly MRI operation is achieved by only using communication resources that are typical for human-to-human-dialogue: **acoustic**, **visual** and **haptic** channels. The achieved quality increase of

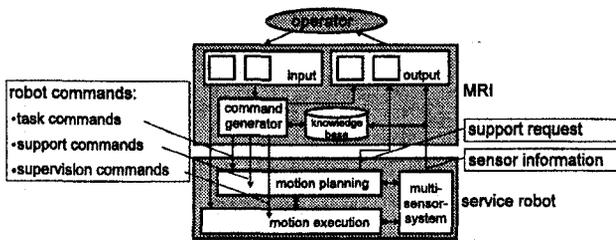


Figure 7: MRI information structure.

the human-robot-dialogue promotes the acceptance of service robots in real-life applications. Resulting implementation requirements can be summarized as follows:

- **A dialogue-oriented natural-spoken command input** renders service robots more accessible even for inexperienced users. Furthermore, the high level of abstraction of natural spoken language allows fast and comprehensive commanding.
- **Visual monitoring and robot support** simplifies situation assessment by the operator and thus, eases operator involvement.
- **Tactile supervisory control during mobile manipulation** may be performed by the operator thanks to realistic force feedback.
- **Voice output during task operation** by means of a speech synthesizer simplifies bidirectional communication.

The concept of the proposed MRI is shown in Fig. 7. The main command channel is connected to a command generator, which converts user-specified tasks into robot commands. Execution of these commands is monitored, and possibly supported, via the sensor/request channel [10].

Three types of robot commands are used for specifying desired robot actions: 1. *Task Commands*: fetch-and-carry typical items of an indoor environment. 2. *Support Commands*: access low-level robot motion components. 3. *Supervision Commands*: stop or proceed motion, cancel current task, or stop system in case of emergency [11].

A natural language command is converted to robot commands in three steps, Fig. 8: 1. Feature extraction, 2. Semantic decoding, 3. Robot command generation. The first two steps are performed by a sophisticated hybrid system developed at the Institute of Human-Machine-Communication of the Technische Universität München [12]. Currently, the system is limited to natural spoken sentences without subordinate clauses.

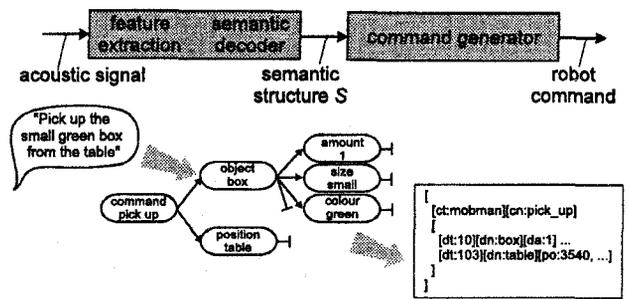


Figure 8: Natural spoken speech interface.

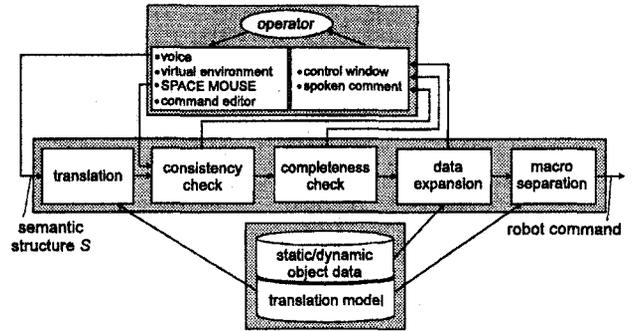


Figure 9: Command generator.

The resulting semantic structure is translated to robot commands by the command generator, Fig. 9. Flexible and safe command generation is guaranteed by:

1. Validation of the commands with respect to consistency and completeness.
2. Inclusion of current environmental information.
3. Operator involvement via different multi-modal MRI resources.

7 Experimental Validation

The key components described so far have been integrated into the full-scale service robot ROMAN. ROMAN is adapted to indoor requirements: width 0.63 m x depth 0.64 m x height 1.85 m, cf. Fig. 11. Maximum velocity is 2 m/sec; the robot's weight is 260 kg.

7.1 Implementation Details

ROMAN is based on an omnidirectional locomotion platform. Three independently steerable wheel systems provide high maneuverability. Standard wheels with a diameter of 0.2 m allow travel across rough surfaces like carpeted floor.

A manipulator with 6 DOFs is mounted on a high/low linear axis at one corner. This allows manipulation of objects located on the floor as well as objects positioned in shelves.

For self-localization of ROMAN, an onboard laser-based angle measurement system is used. An eye-safe laser beam scans the environment in a horizontal plane and determines the azimuth angles to known artificial landmarks, i.e., retro-reflecting tape strips attached

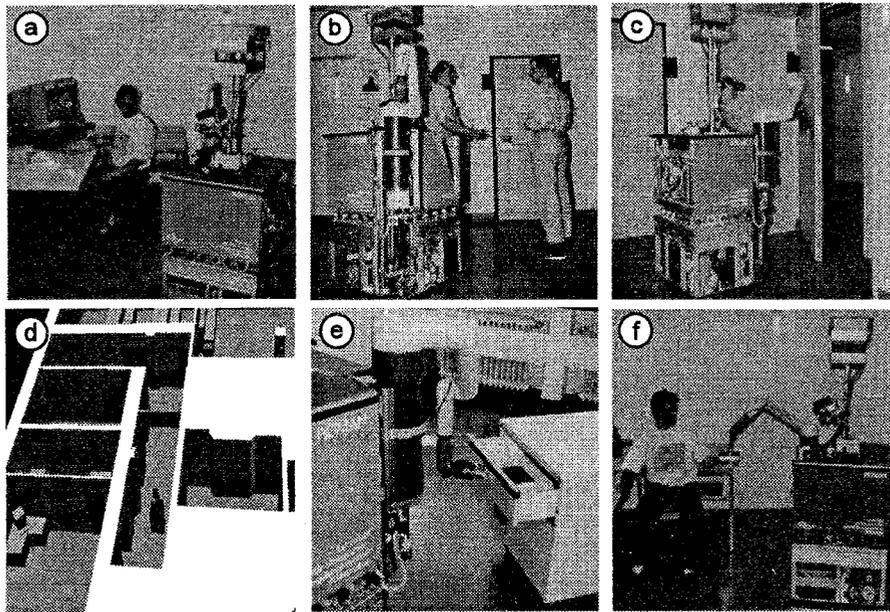


Figure 10: a) Operator at MRI and ROMAN. b) Evasive maneuver. c) Door opening operation. d) Virtual workspace. e) Opening drawer and picking up box. f) ROMAN handing over requested box to operator.

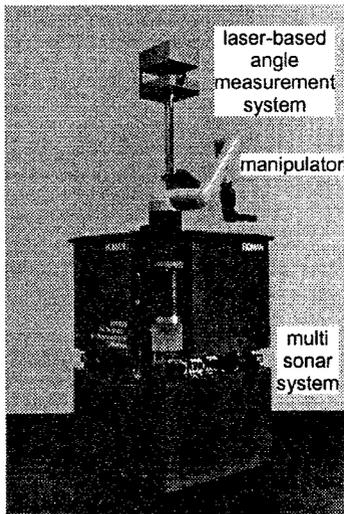


Figure 11: Mobile robotic assistant ROMAN.

to the walls. In addition, ROMAN is equipped with a multi sonar system for obstacle detection/localization and self-localization.

Object recognition is based on a standard video camera. The camera is mounted on a tilt-unit to allow recognition of objects within the entire manipulator workspace.

7.2 Typical Service Task

The sequence of pictures in Fig. 10 shows ROMAN executing a typical service task. The operator issues a natural language command: “Pick up the box contained in the drawer box located in the laboratory”,

Fig. 10 a). This command is correct and complete, i.e., specifies action, object, and location. Based on this information, ROMAN automatically generates a locomotion path and approaches the goal while being continuously supported by the self-localization module. In addition, the robot monitors its surrounding for unexpected obstacles. If an obstacle is detected, Fig. 10 b), ROMAN either performs an evasive maneuver or asks for clearing the way.

When approaching doors, the current state is determined by means of video-based object recognition. If the door is closed, Fig. 10 c), ROMAN opens the door by using all 10 degrees-of-freedom. After opening the door, ROMAN continues approaching the goal.

During task execution, the operator may observe ROMAN moving through the virtual workspace and may issue intervening commands in case of critical situations, Fig. 10 d). At the goal location, Fig. 10 e), ROMAN determines the relative position of the drawer box by means of the object recognition module. The drawer box is then opened and the box is taken out of the drawer. Finally, the object is transported back and handed over to the operator, Fig. 10 f).

The experiment demonstrates the usefulness of the semi-autonomous approach for execution of service tasks, i.e., basic autonomous operation combined with operator intervention.

8 Conclusions

A prototype implementation of a comprehensive mobile service robot system has been introduced. It was

developed with respect to health care applications and domestic automation.

During development, some key components have found to be absolutely necessary for success and have thus been described in some detail:

- highly maneuverable locomotion platform,
- light-weight manipulator arm,
- online coordination of locomotion and manipulation,
- reliable self-localization and obstacle detection,
- vision-based object recognition,
- human robot interaction via multi-modal MRI.

These key components have been integrated into the full-size robot ROMAN and tested in various typical service scenarios in realistic environments. Early testing was found to be essential to open the designers' eyes for application specific problems.

A semi-autonomous approach has been chosen, since appropriate artificial intelligence techniques for achieving fully autonomous behaviour are not yet available. Furthermore, the semi-autonomous approach proves to be closer to the current expectations of potential end-users. Their past reluctance could be partially reduced thanks to already demonstrated capabilities of a state-of-the-art service robot in real-life applications. Even the interest of possible robot manufacturers was stimulated by the ROMAN experiments.

Nevertheless, there are still many open problems. Before the first generation of service robots can be launched in mass quantities, they need to become more robust, more reliable and what is most important *fail-safe*.

A further prerequisite for mass production is a more user-friendly operation that makes the robots more accessible for non-expert users. This includes self-diagnosis in case of operation failures and automatic derivation of appropriate counteractions.

Cost will not be a major issue, since dedicated cost-effective solutions like a mail distributor, a shopping assistant, or a floor cleaner can be derived based on the already developed universal key components.

The next steps of our research include the utilization of dextrous hands to extend the class of graspable objects. In addition, the concept of semi-autonomous operation will be augmented by telepresence capabilities. This will be done by use of standard networking techniques like Internet and standard telecommunication services like ISDN for local and worldwide accessibility. Last, but not least, learning is pursued to obtain a higher level of autonomy, i.e., more intelligent robot behaviour.

References

- [1] W. A. Daxwanger and G. Schmidt, "Neural and Fuzzy Approaches to Vision-based Parking Control," *IFAC Control Engineering Practice*, vol. 4, no. 11, pp. 1607-1614, 1996.
- [2] C. Fischer, M. Buss, and G. Schmidt, "Soft Control of an Effector Path for a Mobile Manipulator," in *Proc. of the Intl. Symp. on Robotics and Manufacturing (ISRAM)*, (Montpellier, France), pp. 299-306, 1996.
- [3] U. D. Hanebeck and G. Schmidt, "Localization of Fast Mobile Robots Using an Advanced Angle-Measurement Technique," *IFAC Control Engineering Practice*, vol. 4, no. 8, pp. 1109-1118, 1996.
- [4] U. D. Hanebeck and G. Schmidt, "A New High Performance Multisensor System for Fast Mobile Robot Applications," in *Intelligent Robots and Systems 1994 (IROS)* (V. Gräfe, ed.), pp. 1-14, Amsterdam: Elsevier Science, 1995.
- [5] U. D. Hanebeck and G. Schmidt, "Set-theoretic Localization of Fast Mobile Robots Using an Angle Measurement Technique," in *Proc. of the 1996 IEEE Intl. Conf. on Robotics and Automation (ICRA)*, vol. 2, (Minneapolis, MN), pp. 1387-1394, 1996.
- [6] U. D. Hanebeck and G. Schmidt, "Closed-Form Elliptic Location with an Arbitrary Array Topology," in *Proc. of the 1996 IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, (Atlanta, GA), pp. 3070-3073, 1996.
- [7] E. Ettelt and G. Schmidt, "Vision Based Guidance and Control of a Mobile Forklift Robot," in *Proc. of the Intl. Conf. on Recent Advances in Mechatronics*, (Istanbul, Turkey), pp. 180-186, 1995.
- [8] E. Ettelt and G. Schmidt, "Vidobasierte Objekterkennung mittels Musterbaumgestützter Kreuzkorrelation," in *Fachgespräch "Autonome mobile Systeme" (AMS)*, (Stuttgart, Germany), 1997.
- [9] C. Fischer, M. Buss, and G. Schmidt, "Human-Robot-Interface for Intelligent Service Robot Assistance," in *Proc. of the IEEE Intl. Workshop on Robot and Human Communication (ROMAN)*, (Tsukuba, Japan), pp. 177-182, 1996.
- [10] C. Fischer and G. Schmidt, "Multi-Modal Human-Robot-Interface for Interaction with a Mobile Service Robot," in *Proc. of the 6th Intl. Workshop on Robotics in Alpe-Adria-Danube Region (RAAD)*, (Cassino, Italy), 1997.
- [11] C. Fischer, M. Buss, and G. Schmidt, "Hierarchical Supervisory Control of Service Robot Using Human-Robot-Interface," in *Proc. of the Intl. Conf. on Intelligent Robots and Systems (IROS)*, (Osaka, Japan), pp. 1408-1416, 1996.
- [12] H. Stahl, J. Müller, and M. Lang, "An efficient top-down parsing algorithm for understanding speech by using stochastic and semantic models," in *Proc. of the Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, (Atlanta, GA), pp. 397-400, 1996.