In Proc. of Int. Conf. on Field and Service Robotics FSR 2001, June, Helsinki.

INTUITIVE PROGRAMMING OF A MOBILE MANIPULATOR SYSTEM DESIGNED FOR CLEANING TASKS IN HOME ENVIRONMENTS

M. Strobel, J. Illmann, E. Prassler
Research Institute for Applied Knowledge Processing
Helmholtzstr. 16, D-89081 Ulm, Germany
{mstrobel,illmann,prassler}@faw.uni-ulm.de

Abstract: A new service robot designed for cleaning tasks in home environments is introduced. Robot systems will work directly with people in this area, thus placing a central importance on making interactions between people and machines as natural as possible. The main focus of this paper is on the development of a gesture based intuitive programming approach which allows the robot to be easily used by non-experts. As an application example the task of cleaning the workspace in a kitchen area is considered. Gesture recognition and context sensitive interpretation, complete coverage planning as well as motion planning of the mobile manipulator which are the key functionalities to perform the task are highlighted.

Keywords: intuitive robot programming, man machine interface, mobile manipulator system

1 INTRODUCTION

Housekeeping robot assistants focus on the employment of assistive robot systems in everyday domestic settings. There are different motivating factors for the employment of robots at home: on one side, comfort factors and a changing societal framework favor the employment of man-made personnel; on the other side, an increasing number of households include inhabitants that require physical support in day-to-day life due to sickness or age. Robot systems will work directly with people in this area, thus placing a central importance on making interactions between people and machines as natural as possible.

The robot assistant in the home should work together with the user to perform simple housework. In addition to fetch-and-carry duties, this includes tasks such as setting the table or performing cleaning tasks. A new service robot which is able to perform cleaning tasks is currently being built at the FAW. In Chapter 2 of this article the hardware, especially the kinematic design and the sensor setup of this cleaning assistant robot is envisaged.

Human interaction with the robot assistance system will have as its purpose the commanding and teaching of the robot, but it also offers interesting possibilities of augmenting the performance of the entire system. Speech and gestures are a human being's most natural communication channels. Interaction with the robot may further occur over versatile (multimodal), portable control devices. Our approach to an intuitive programming and interaction is highlighted in Chapter 3. We decided to use communication techniques based on hand gestures and optional speech which can be expected to be natural and convenient as they don't require to use any external equipment such as data gloves or helmets. Our approach to gesture recognition and context-sensitive interpretation can be divided into two main categories. In the first category, hand gestures from the human communication partner are executed in the shared workspace and are sensed by a trinocular stereo-vision system. In place of showing items, positions, surfaces etc. in real world, communication techniques based on input via touchscreen are considered in the second category. This will favorably be used when showing a path in an internal representation of the environment, e.g. in a two dimensional floor plan, or when selecting e.g. items in pictures from remote locations.

The robot system should already possess a sufficient base of high-level and specialized skills to avoid expendable and tedious instruction and programming of rudimentary skills. E.g. considering the task of cleaning several surfaces in a kitchen it should be sufficient to show the robot which surfaces it has to clean and not to teach the whole cleaning trajectory. Chapter 4 highlights the high-level skill of the cleaning robot which are required to plan and execute a cleaning task.

2 DESIGN OF THE CLEANINGASSISTANT

The *CleaningAssistant* robot, shown in Figure 1 consists of a mobile base and a manipulator on top of it. The manipulator joints as well as the mobile base are built of modular drive components. Two independently controlled drive components are used to built a differential drive system for the mobile base. The differential drive system is a cheap and robust drive system, but it is underlying a so called non-holonomic constraint which complicates motion planning.

A 7 DOF manipulator is based on a vertical linear axis with the goal to enhance the vertical workspace of the system. Next a SCARA-like chain of revolute joints are mounted on the linear axis. An additional degree of freedom is used to switch between the horizontal and vertical arrangement of the SCARA-like chain. Intermediate configurations are also allowed. The advantage of this arrangement is that it allows to use the advantages of the horizontally mounted chain (low energy consumption and high dynamics for horizontal arm movements) without being limited to only horizontal movements. The end-effector is either a two-finger gripper or a special end-effector, e.g. for cleaning non-textile surfaces like working places in kitchens or sinks. Sensory feedback is provided by a) a new kind of compliant force-torque-sensor, developed at the German Aerospace Center (DLR) (see Meusel and Hirzinger [9]) mounted between the wrist and the end-effector of the manipulator, b) a 2D range laser-scanner (Sick LMS 200) is used for position estimation as well as for obstacle detection and avoidance while navigating the mobile base, c) a trinocular stereo-vision system (Triclops) for gesture and object recognition and localization, d) a touchscreen (Elotouch) for additional gesture input, e.g. for the qualitative path specification or object selection in a displayed scene representation.

Figure 1: Gesture based programming of the FAW CleaningAssistant for home environments.

The central control unit is based on an industrial PC running the operating system LINUX and the software framework *SmartSoft* [11] which is especially designed for the communication between sensors, actors and planning components in the field of robotics. SmartSoft is based on a client-server architecture and a few simple communication primitives which provide unified interfaces. Thus modules can be combined in a flexible way to implement complex operations.

3 INTUITIVE PROGRAMMING APPROACH

According to our conviction, natural instruction, interaction and programming of robots will play a decisive role in the entering of such systems in home environments. Therefore we decided to use techniques based on hand gestures and optional speech which can be expected to be natural and convenient as they don't require to use any external equipment.

Two different methods which are appropriate for an intuitive programming of the *CleaningAssistant* robot are introduced: Section 3.1 is dealing with the visual detection and model based interpretation of pointing gestures given in the real world working area by a human operator. Second, our approach to commanding motions for the mobile base of the *CleaningAssistant* by a touch screen interface is shown in Section 3.2.

3.1 Task Programming or Object Selection by means of Pointing Gestures in Real World

Let us consider the task of showing a *CleaningAssistant* which surfaces it has to clean. Our approach for this purpose is based on the recognition and context-sensitive interpretation of human pointing gestures as follows: First a human is pointing on the surfaces the robot has to clean or shows the robot a specific surface area by means of a sequence of pointing gestures. We concentrated on visual recognition of hand gestures because they are preferable to speech in a number of situations, for example in noisy environments or when communicating quantitative information and spatial relations. Chapter 3.1.1 gives a brief introduction in way we are recognizing hand gestures via vision. When a pointing gesture is recognized a 3D-line is fitted to the corresponding 3D-coordinates of points (received from the trinocular stereo vision system) which are classified as belonging to hand or underarm. The pointing direction information together with geometric environment information is used to select objects to handle or areas to clean. See Section 3.1.2 for our method to find the right (object) correspondence in the scene.

3.1.1 Visual Recognition of Hand Gestures

The visual recognition of static hand gestures is based on an approach which integrates color, disparity and motion information. These different information sources are combined to provide a robust skin segmentation and are used subsequently to compute a normalized view of the segmented hand. Shape and contour features that are extracted from this normalized view are used in the classification stage to compare the gesture with stored gesture models.

The implemented architecture is displayed in Figure 2. This procedure is continuously repeated. If a predefined pointing gesture is recognized the pointing direction is computed by fitting a line to the world coordinates of the points assigned to the hand.

Segmentation Stage: To distinguish a hand region from the background we use a combined color- and disparity-based approach. The color segmentation utilizes a predefined model of skin color. Our color representation is based on hue and saturation. Because gesture recognition should be reasonably insensitive to illumination changes color intensity is ignored. It is known that the distribution of skin color in hue-saturation space can not be described by simple shapes [3]. Therefore we decided to use a histogram for representation of skin color. Histogram entries characterize the membership of a color to the skin model. A first model histogram is computed by extracting hue and saturation values of image points belonging to a certain predefined area in the world. In a bootstrap process this histogram is continuously updated and will eventually represent skin color. To segment a color image the rectified image is preprocessed by a quantization algorithm [2]. By applying this preprocessing step the number of image colors are reduced and are able to get a much better segmentation and clearer segment contours. Using the color histogram the extracted segments are color classified. Concurrently a disparity based image segmentation is applied. To compute disparity images we utilize a calibrated trinocular camera [14]. Corresponding image points in the three images are estimated by computing local correlations of the edge filtered images. The disparity image is segmented based on a-priori information about the expected disparity range of the person. The expectation about the persons disparity is continuously updated to account for slow movements.

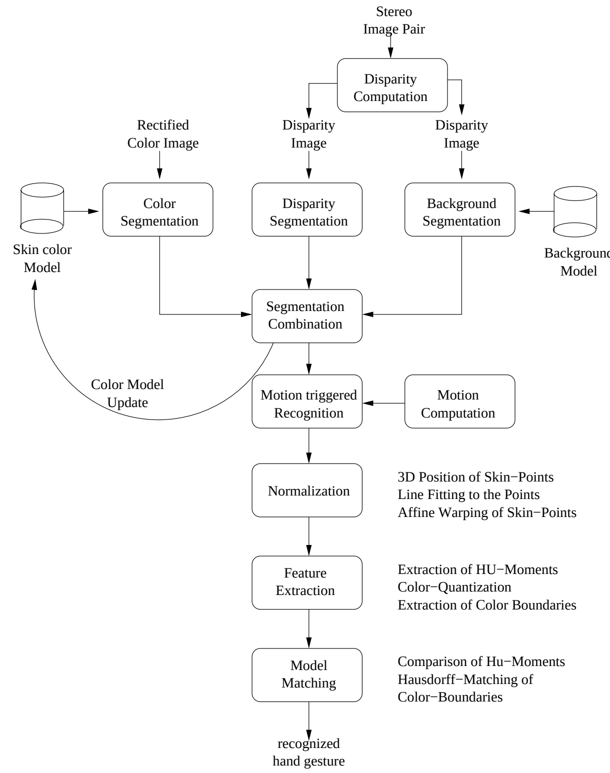

Figure 2: Gesture Recognition Procedure

An additional figure-ground separation stage is based on background information. A scene model containing disparity information of background areas is estimated beforehand. During the recognition process extracted disparity information is compared to this scene background.

Figure 3: Example of a captured gesture image.

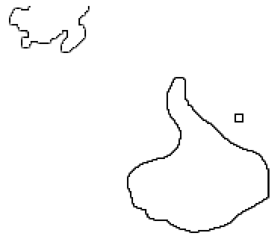

Figure 4: Extracted contours of the normalized hand regions.

Image areas belonging to the background are not taken into further account. Because color information is not explicitly used in this background separation process the segmentation is quite robust. Finally color and disparity segmentation results are combined to provide regions of interest for hand regions which are further investigated. After some simple pre-processing, blob-labelling and validation of candidate regions the model representing skin color is continuously adapted by adding the new skin color values to the color histogram.

Motion computation: Since we want just to recognize static gestures we require a gesture to be kept still for some time. Local image motion is computed simply by temporal image differencing. We require that the temporal difference image shows low motion in the region of interest for one second.

View normalization: To classify extracted candidates for hand regions the view aspect is normalized with respect to distance and orientation of the gesture. For this purpose the world positions of all image pixels which were assigned to a segment are computed and a three dimensional line is fitted to this points. The line is used to normalize the view by affine warping.

Feature Extraction: The classification of extracted skin segments is based on comparison of moment invariants and contour lines with stored models. To compensate for slight illumination variations a local color histogram is extracted from the normalized skin region. This color histogram is back-projected onto the skin region. This local color normalization is followed by an additional local color quantization. This local color normalization led to the best results in slightly changing environments.

Finally the boundaries of the color quantized hand regions are extracted and approximated [13] (see Figure 4).

Model Matching: To classify a gesture the hand region is compared to a set of gesture models using second and third order normalized moments of the extracted regions [5] and a Hausdorff-distance based contour matching [6]. The classifier uses a nearest-neighbor-criteria to select the most similar gesture model. If the difference between model and extracted gesture is too large the gesture is rejected.

3.1.2 Gesture Interpretation

Interpretation of a pointing gesture is done context-sensitive. Considering the example of selecting a surface to clean the system "knows" that it has to find a corresponding surface to the detected pointing gesture. Position and orientation of the pointing hand gesture are calculated as described above. The corresponding surface is determined with the help of the spatial pointing information, an internal three dimensional representation of the scene (especially the geometrical environment description) and algorithms known from computer graphics [4].

In a first stage, it is assumed that a priori information about the environment (here: house and arrangement of the furniture) is available. This restriction can be relaxed in a further stage, in which the system will learn interactively the object distribution in the environment and its coherence.

So far, we are using polyhedrons as fundamental geometrical modelling primitives. A polyhedron is considered as single convex solid. Based on polyhedrons we are using a hierarchical descriptions, called *polytrees* here, for convex or non-convex geometric objects. The depth of a polytree and the branching factor of the individual nodes are arbitrary. The leaves of a polytree corresponds to the underlying convex pieces of solid. For a more detailed description of the modelling, see Mirtich [10].

The intersections of a ray starting from the determined position of the operators hand and directing into the gesture's pointing direction and objects from our geometric environment model are used as raw input. Depending on the context this raw input information is further processed, e.g. intersection points may be interpreted as points on surface borders.

3.2 Commanding motion by a touch screen interface

In place of showing items, positions, surfaces etc. in real world, communication techniques based on pointing sequence input via touch-screen are considered too. In our application of

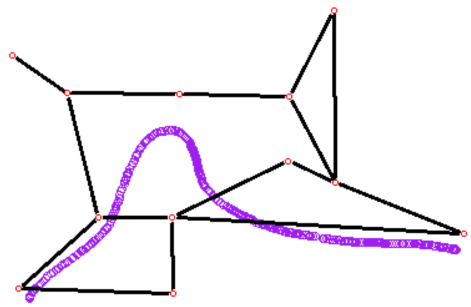

Figure 5: Topological planar graph and touch sequence input.

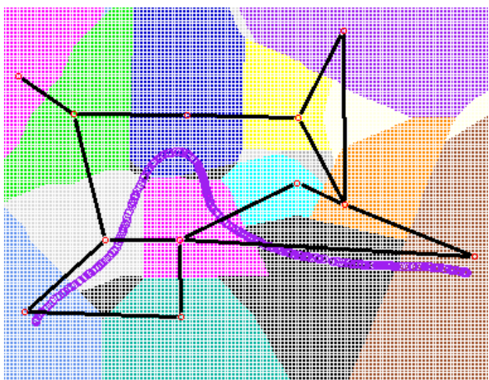

Figure 6: Graph, grid point sequence and approximated graph Voronoi regions.

the CleaningAssistant, this kind of command input will favorably be used when showing a path in an internal representation of the environment, e.g. a topological graph of the environment in which vertices are representing locations (rooms etc.) and edges are representing connections between them which are possible to traverse by the *CleaningAssistant*.

Human input demonstrating trajectories by a sequence of points may be incomplete, distorted etc. These effects are compensated by a transformation of vertex sequences of a regular grid into paths of a planar graph which codes feasible motions. The transformation is based on alteration operations including re-routings and on a so-called graph Voronoi regions which partition the plane according to proximity to vertices and edges. A detailed treatment of the transformation is given in Kämpke and Strobel [7].

Figure 5 shows an example of an input touch sequence of about 200 grid points together with the underlying topological planar graph. Edges are labeled by their Euclidean lengths. Figure 6 depicts approximations of the graph Voronoi regions. The "overshoot" (U-shaped section) of

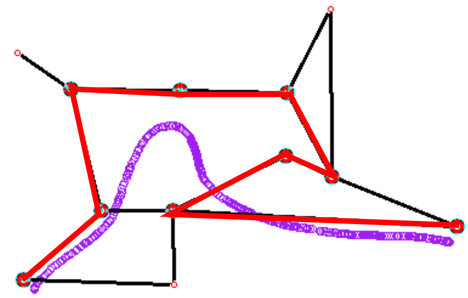

Figure 7: Graph with grid point sequence transformed to a path (bold edges).

the grid point sequence in Figure 5 in the center region of the graph is so large that vertex insertions occur. Figure 7 shows the resulting path.

4 COMPLETE COVERAGE PLANNING AND PATH PLANNING

The main high-level skills of the cleaning robot which are required to plan and execute a cleaning task are full coverage planning and path planning of the mobile base and the manipulator. In our applications, full coverage planning deals with the problem of finding a full coverage path for a given polygonal bounded surface to be cleaned. This means the planning of a path that guarantees that e.g. a tool like a sponge or a special end effector will pass over every point within that given area. Choset and Pignon [1] introduced a complete coverage path planning algorithm which is based on a exact cellular decomposition of the target area. In order to transform the full coverage paths planned in workspace to paths in the robot's configuration space Marrone and Strobel [8] derived a closed form solution of the robot's inverse kinematics. The redundancy of the manipulator is therein used to optimize the inverse kinematics solution in terms of manipulability (see Yoshikawa [15]) and closer posture. So far, only the redundancy of the manipulator is considered to optimize the system's movements. Considering the additional redundancy of the mobile base too is part of our current research. An additional requirement for manipulator path planning is to check whether a specific manipulator configuration is valid or not. Valid means that the configuration joint variables must be within their mechanical bounded range and that the manipulator configuration is not causing a collision with the robot itself and with obstacles in the environment. Our approach to collision detection is based on the polyhedral modelling of the robot and the environment which is also used for the context sensitive interpretation of pointing gestures (see section 3.1.2). The calculation of possible intersections of the polyhedral modelling primitives is done with the help of *V-clip* polyhedral collision detection library from Brian Mirtich [10]. As already mentioned before, at the current state of our work, navigation of the robot's non-holonomic mobile base and manipulator motion planning is considered separately. A detailed treatment of our approach to non-holonomic non-circular mobile robot navigation in narrow environments can be found in Strobel [12].

5 DISCUSSION

This paper shows the current state of an ongoing project with the main focus on man-machine-interaction. Future steps will concentrate on making the system more robust with respect to environmental changes (e.g. scene background and moving camera in case of image processing and unknown obstacle distribution in case of motion planning). Other goals are the automated (or interactive) building of the so far a priori given polyhedral environment model and the integrated motion planning of mobile base and manipulator.

ACKNOWLEDGEMENT

This work was supported by the German Department for Education and Research (BMB+F) under grant no. 01 IL 902 F6 as part of the project MORPHA. The authors would like to thank Pablo d'Angelo, Uwe Friker, Jens Klopp, Thomas Mack, Ingo Maindoerfer and Fabrizio Marrone for implementation support.

REFERENCES

[1] Choset, H., Pignon, P., *Coverage Path Planning: The Boustrophedon Cellular Decomposition*, In Proceedings of the Field and Service Robotics Conference FSR'97, 1997.

[2] Comaniciu D., Meer, P., *Robust Analysis of Feature Spaces: Color Image Segmentation*, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 750-755, 1997.

[3] Feyrer, S., Zell, A., *Ein integrierter Ansatz zur Lokalisierung von Personen in Bildfolgen* , in Dynamische Perzeption, S. Posch and H. Ritter, eds., no. 8 in Proceedings in Articial Intelligence, pp. 183-190, inx, Sankt Augustin, Germany, 1998.

[4] Foley, van Dam, Feiner, Huges, *Computer Graphics - Principles and Practice*, Addision-Wesley Publishing Company, 1996.

[5] Hu, M., *Visual pattern recognition by moment invariants*, IRE Trans. Information Theory, Vol. IT-8, Num. 2, 1962.

[6] Huttenlocher, D., Noa, J., Rucklidge, W., *Tracking Non-Rigid Objects in Complex Scenes*, Cornell University Computer Science Dept. Tech Report, CUCS TR-92-1320, 1992.

[7] Kämpke, T., Strobel, M., *Graph Voronoi Regions for Interfacing Planar Graphs* , accepted for presentation in the 2001 International Workshop on Computational Geometry and Applications , March, 2001.

[8] Marrone, F., Strobel, M., *CleaningAssistant - A Service Robot Designed for Cleaning Tasks*, accepted for presentation in Advanced Mechatronic Systems (AIM 2001), Como, Italy, July, 2001.

[9] Meusel,P., Hirzinger, G., *Compliant Force-Torque-Sensor for sensory-based Assembly*, Information Sheet at (http://www.robotic.dlr.de/COMPLIANCE/), 1999.

[10] Mirtich, B., *V-clip: Fast and Robust Polyhedral Collision Detection*, Technical Report TR-97-05, Mitsubishi Electric Information Technology Center America, June, 1997.

[11] Schlegel, C., Wörz, R., *The Software Framework SmartSoft for Implementing Sensorimotor Systems* , IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '99, pp. 1610-1616, 1999.

[12] Strobel, M., *Navigation in Partially Unknown, Narrow, Cluttered Space*, IEEE International Conference on Robotics and Automation, Detroit, USA, May 1999.

[13] Teh, C.H. and Chin, R.T, *On Digital Approximation of Moment Invariants*, Computer Vision, Graphics and Image Processing, 33(3):318-326, 1986.

[14] Triclops - Stereo Vision System, Manual Version 2.1, Point Grey Research Inc., Vancouver, British Columbia, 1998.

[15] Yoshikawa, T., *Manipulability of Robotic Mechanisms*, International Journal of Robotics Research, Vol.4, No. 2, pp. 3-9, 1985.