Robust Real-Time Pursuit of Persons with a Mobile Robot Using Multisensor Fusion

Stefan Feyrer and Andreas Zell
Wilhelm-Schickard-Institute, Department of Computer Architecture
University of Tübingen
Köstlinstr. 6, 72074 Tübingen, Germany
{feyrer, zell}@informatik.uni-tuebingen.de

Abstract. Natural man-machine interfaces for mobile service robots presuppose on the part of the robot the ability to perceive humans and their behavior. In this paper we present a system, which enables a mobile robot to detect persons, to track them actively, and to pursue them in real-time. Obstacles are detected and avoided during pursuit. The process of person detection is based on multisensory information. Besides several vision-based methods an additional approach to detect human legs in laser scans is used. Combining both modalities enables a flexible and robust following behavior.

1. Introduction

There is a wide variety of possible application scenarios for mobile service robots in use as personal assistants. In industrial environments production assistants could improve both productivity and working conditions by performing tasks in interaction with humans. Shopping assistants could help customers to find their way around large supermarkets and department stores. In the domestic field nursing robots could assist elderly people by improving their mobility and by serving as a communication platform.

A great number of research projects currently deal with assistance systems of that kind [1] [11] [10]. Regardless of the application robust detection and interpretation of the environment, a high rate of adaptability and a natural man-machine interface are necessary for all of them. One important aspect in this context is the ability of the robot to recognize the human user, to localize him and to direct his attention to him in order to be able to react to verbal commands or gestures in an appropriate way. Ideally the perception of humans should not be decoupled from the current high-level task, it should rather be integrated in it and adapted to the current task context.

This is where our system sets in. It combines a person detection and tracking process with a navigation task aiming to pursue a person and at the same time avoiding stationary and dynamic obstacles. Almost all person recognition systems (see [6] for an overview), even those especially designed for mobile robots [9], are based on vision. In contrast to these and to previous versions of our system [4] [5] we extend the purely vision-based detection module with a laser-based method.

Integration of the laser modality makes it possible to derive advantage from data fusion at model level so that a more robust tracking behavior even in difficult situations can be achieved. Secondly, the limitation of being compelled to face the robot during pursuit is overcome by a flexible automatic weighting of the detection modalities. The new system is able to detect persons in arbitrary orientations and to determine their orientation relative to the robot. Even during pursuit the person can arbitrarily change its orientation without the robot loosing track of it.

2. Visual Detection of Persons

The vision-based component of the person detection module aims to extract head and especially face regions from the camera images. It is based for its own part on several different cues. Besides color and grey-level contours motion information is used to separate moving objects from the stationary background. Additionally, a stereo approach yields three-dimensional scene information and in particular the distance between the robot and the person. The different modules and their mutual influence are described in [4] and [5]. For this reason we confine ourselves here to a short overview.

To identify skin-colored image pixels we use the illumination invariant HS-subspace of the HSI color model. The areas representing skin color are defined in terms of a lookup table which has been derived by statistical examinations. By clustering the skin-colored regions we obtain a set of ROIs as a first rough estimate for head regions.

In a second step, moving objects are detected using a difference picture technique. In case of the robot base being stationary two successive frames are subtracted directly, whereas during robot motion the current frame is compared to a prediction image, which is calculated taking into account the ego-motion of the robot. We do not compensate for the fast pan and tilt movements of the cameras. As a result, the cameras have to move in a saccadic-like manner.

Detection of contours is done using a method called *active rays* [3]. Starting at a reference point within the object boundaries the course of grey levels is examined along straight lines passing through this point. By minimizing an energy function we obtain a chain of contour points approximating the real object contours. The shoulder lines are extracted separately. Comparing the resulting outline with an empirically determined average contour yields a confidence value for being a head/shoulder region.

Depth information needed for localizing the person in the three-dimensional scene and for compensating for ego-motion is provided by a stereo vision module with automatic recalibration.

An evaluation algorithm combines structural and geometrical properties of the extracted regions of interest as well as the results of the motion, contour and stereo analysis. Additionally, the coherence of the image sequence is taken into account. Using a probabilistic framework a confidence value describing the likelihood for containing a head is assigned to each ROI.

Fig. 1. Results of the color, motion, contour and depth analysis.

3. Detection of Persons Using Laser Data

For legal reasons laser range sensors normally have to be attached near the bottom of robot vehicles. Thus, distance information is merely available in a horizontal plane at leg height. As a consequence, discrimination from other objects in the scene can take place exclusively on the basis of certain characteristics of human legs. In our approach characteristic shape information as well as motion information is used.

In a first step a raw scan, denoted by $S = (s_1, \ldots, s_n)$, is searched for individual legs. For this purpose scan intervals $I_m = [i_m, j_m]$ with significant differences in depth at their edges are determined. Those intervals, which furthermore stand out against their local environment by smaller depth values, are selected:

$$(s_{i_m-1} - s_{i_m} > \tau) \wedge (s_{j_m+1} - s_{j_m} > \tau) \wedge (\forall i_m \le i < j_m : |s_i - s_{i+1}| \le \tau) \quad (1)$$

Subsequently, a plausibility check is executed, which considers the spatial expansion of the area in tangential direction and the shape of the segment. Segments which fall below a certain convexity measure as well as segments the size of which deviates too much from usual leg extents are rejected.

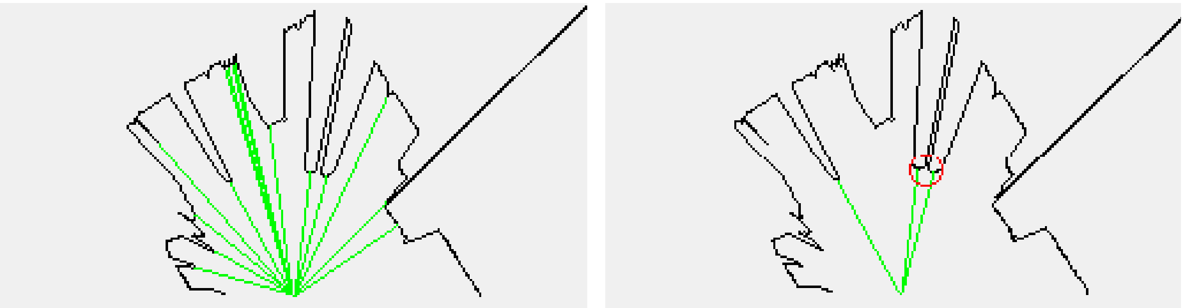
The remaining segments are supplied to an evaluation algorithm, which carries out the transition of individual leg positions to person positions. The evaluation is performed as a function of whether two leg positions are placed in such a way that they can be combined in a person position. Additionally, the position shift d with respect to the preceding detection step is evaluated in combination with the intensity of motion $m \in [0, 1]$ occurred within the respective area:

$$p(d, m) = \alpha \cdot \exp(-(1 - e^{-(d/\sigma)^2} - m)^2) + \beta \quad (2)$$

α and β are scaling factors which restrict the resulting values to the interval $[0, 1]$. The choice of this evaluation function ensures that contradictory indicators – e.g. small position changes during high intensity of motion – lead to low evaluation results. The results together with the ones from the visual detection serve as a basis for the selection procedure in the sensor data fusion component.

Leg motion detection is done using a special correlation procedure, which decouples position and rotation correlation and thus makes it possible to be run in real-time [8]. Therefore, we use the fact that for all regular polygons the position of the maximum of the autocorrelation function is invariant against rotation. As our application is designed for indoor use, we can assume approximately polygonal environments. In order to improve accuracy, odometry data of the robot are integrated into the correlation procedure.

Intermediate results as well as the final result of the detection process based on laser data are depicted in Fig. 2. Fig. 3 shows the system behavior for three different situations.



(a) (b)

Fig. 2. Intermediate and final result of laser-based detection. (a) Result after segmentation. (b) Segments detected as legs and determined position of the person.

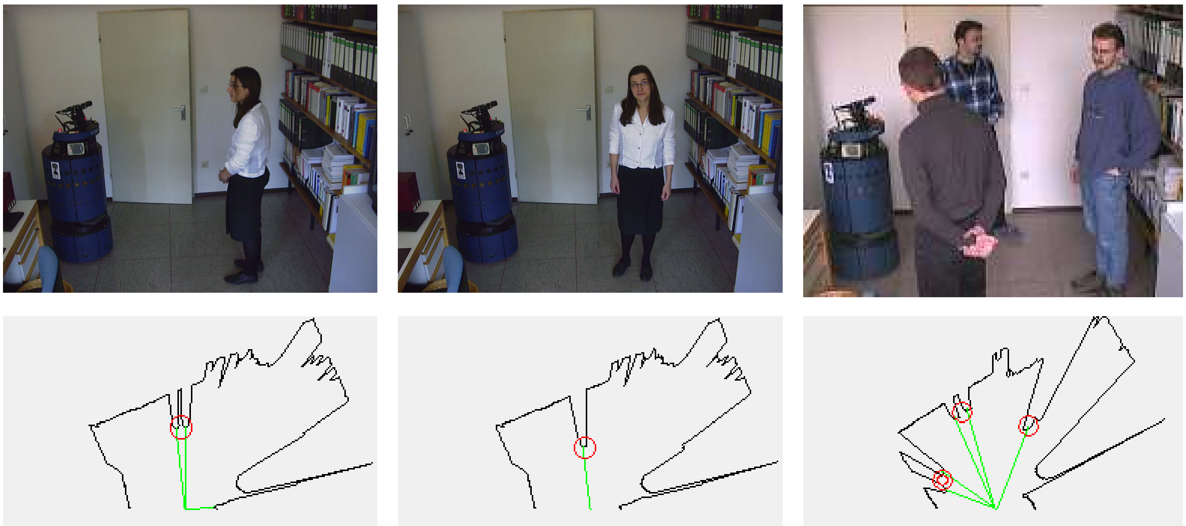


Fig. 3. Situations for with the person detection algorithm works successfully. (a) Person facing the robot. (b) One leg occluded by the other. (c) Multiple persons.

4. Robust Localization by Fusing Multisensory Information

The robustness of the system can be increased by combining the results of the visual and the laser-based detection module. This presupposes however that both detection systems respond to the existence of the person in the scene, which is generally the case. Situations where the laser-based person detection fails are depicted in Fig. 4.

The fusion process itself is realized as a maximum search within a potential field built by overlaying two-dimensional Gaussian functions. Each function represents a position where a person has been detected either by the visual or the laser-based detection component. Furthermore, a Kalman filtered position estimation is added, which takes into account the temporal coherence of the position values. Positions, which fall below a minimum confidence value, are not considered. The standard deviations for the Gaussian functions are chosen according to the accuracy of position estimation. Generally, position values determined by the laser-based component are more accurate than measurements done by stereo analysis. The functions are weighted by the confidence values of the positions for representing a person.

The resulting overall map is used to determine the current position of the person to be pursued. When there are several people near the robot, the system will initially start to track the one with the highest confidence value. During pursuit the integrated position continuity model ensures that the system will not loose the track, even if there are other positions with higher values than the one currently tracked.

The system is able to react in a flexible way to orientation changes of the person going ahead. As soon as the person turns its face away from the cameras, the color-based detection component is no longer able to supply a reasonable contribution to the detection process. In this case the detection process is transferred completely to the remaining components.

5. Pursuit of Persons and Obstacle Avoidance

The results of the recognition component are transferred into robot movements aiming to follow the detected person or to approach it up to a defined distance. For navigation we use a real-time path planning method, which operates on the basis of an artificial potential field [7].

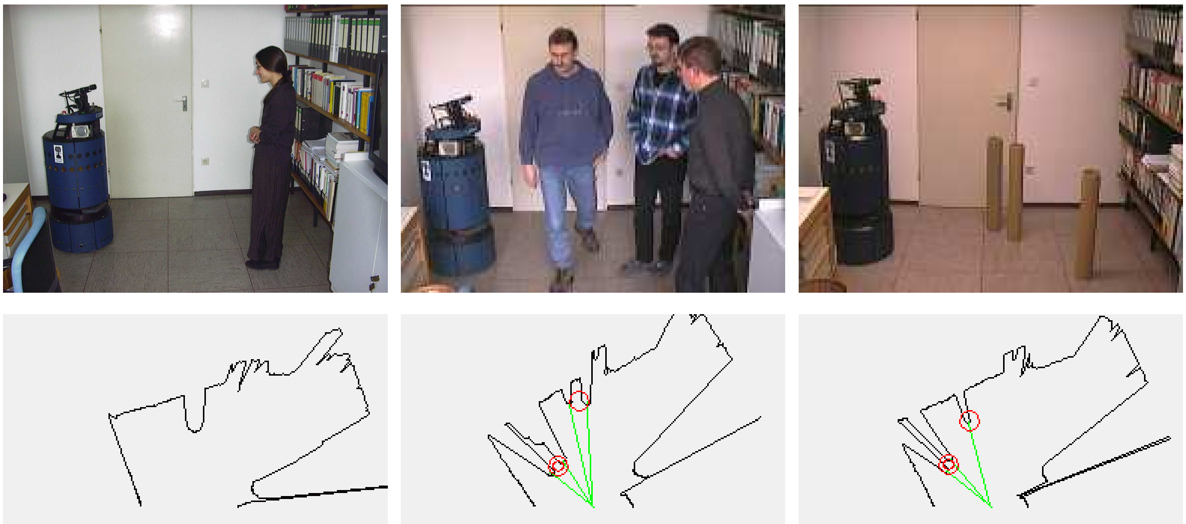


Fig. 4. Some situations where laser-based person detection fails. (a) Very long skirt. (b) One person occluded by the other. (c) Paper rolls looking like legs.

This potential field consists of an attractive potential pulling the robot to the current goal position and a repulsive potential pushing the robot away from the obstacles. Since we have no global obstacle map, the raw distance values provided by the laser scanner are used to calculate the repulsive potential. To avoid overreaction to lateral obstacles the distance values are weighted according to their angle with respect to the robot.

The current target position is taken from a list containing the positions which are supplied by the localization component at regular intervals. A special selection procedure avoids dead end situations for the robot. More details about this algorithm can be found in [5].

6. Experimental Results

The presented system works on a RWI B21 mobile robot. The robot is equipped with two dualprocessor 400 MHz PentiumII PCs. The main function of one of them is image processing, while the other one is mainly used for navigation and laser-based person detection. The two system components communicate via a fast interprocess communication protocol. Image acquisition is done by two CCD color cameras mounted on a pan-tilt-unit. The built-in 2D laserscanner has a scanning field of 180° and an angular resolution of 0.5°.

As the recognition performance strongly depends on scene structure, scene dynamics and, as described above, on the behavior of the person being tracked, it is difficult to give quantitative results which hold in general. However, analysis of several pursuit experiments has shown that the robustness of the system increases significantly when the two sensor modalities are combined. The first experiment has been carried out under difficult lighting conditions so that the results of the vision component would not have been stable enough to track the person successfully. However, in combination with the laser module the system was capable to track the person without error. During 30 seconds of pursuit the vision module provided 5 incorrect position values, all of them could be corrected though by the laser component. For the second experiment we chose an environment which contained several paper rolls similar to human legs. Some of them can be seen in Fig. 4 (c). Thus, the laser module produced a lot of false positive detections, which could be compensated this time by the results of the vision module.

Fig. 5. Image sequence taken during the pursuit of a person including the passage of a narrow doorway.

As for processing performance we achieve a frequency of 8 Hz at a resolution of 192x144 pixels on the part of image processing and at full resolution on the part of laser data processing. With a maximum velocity of 0.5 m/s the robot is able to follow a person at approximately human walking speed.

References

[1] G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsch, D. Magaritis, M. Montemerlo, J. Pineau, N. Roy, J. Schulte, S. Thrun: Towards Personal Service Robots for the Elderly, *Proceedings of the Workshop on Interactive Robotics and Entertainment*, Pittsburgh, PA, USA, 2000.

[2] J. L. Crowley, F. Bérard: Multi-Modal Tracking of Faces for Video Communications, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 640-645, 1997.

[3] J. Denzler, H. Niemann: Active Rays: A New Approach to Contour Tracking, *International Journal of Computing and Information Technology*, vol. 4, no. 1, pp. 9-16, 1996.

[4] S. Feyrer, A. Zell: Tracking and Pursuing Persons with a Mobile Robot, *Proceedings of the International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time-Systems*, Corfu, Greece, pp. 83-88, 1999.

[5] S. Feyrer, A. Zell: Detection, Tracking, and Pursuit of Humans with an Autonomous Mobile Robot, *Proceedings of the International Conference on Intelligent Robots and Systems*, Kyongju, Korea, pp. 864-869, 1999.

[6] D. M. Gavrila: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82-98, 1999.

[7] O. Khatib: Real-Time Obstacle Avoidance for Manipulators and Mobile Robots, *International Journal of Robotics Research*, vol. 5, no. 1, pp. 90-98, 1986.

[8] A. Mojaev, A. Zell: Online-Positionskorrektur für mobile Roboter durch Korrelation lokaler Gitterkarten, in H. Wörn, R. Dillmann, D. Henrich (eds.): *Autonome Mobile Systeme 1998*, Springer, pp. 93-99, 1998.

[9] C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, R. Wörz: Vision Based Person Tracking with a Mobile Robot, *Proceedings of the 9th British Machine Vision Conference*, Southampton, Great Britain, pp. 418-427, 1998.

[10] R. D. Schraft, C. Schaeffer, T. May, A. Traub: Mobiler Serviceroboter Care-O-bot® – ein Hilfssystem im häuslichen Bereich – Aufbau, Funktionsweise und Einsatzmöglichkeiten, in H. Wörn, R. Dillmann, D. Henrich (eds.): *Autonome Mobile Systeme 1998*, Springer, pp. 234-244, 1998.

[11] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, D. Schulz: MINERVA: A Tour-Guide Robot that Learns, *Proceedings of the 23rd Annual German Conference on Artificial Intelligence*, Bonn, Germany, pp. 14-26, 1999.

[12] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland: Pfinder: Real-Time Tracking of the Human Body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, 1997.