Fast Face Detection for Mobile Robots by Integrating Laser Range Data with Vision

J. Blanco[1] W. Burgard[2] R. Sanz[1] J.L. Fernández[1]

[1]*University of Vigo, Department of System Engineering and Automation, Spain*
[2]*University of Freiburg, Department of Computer Science, Germany*

Abstract

In many application areas mobile robots need the ability to interact with people. In order to be able to communicate with even untrained users, the interaction should be as natural as possible. One of the preconditions of a natural interaction is that the robot focuses the person it is interacting with. In this paper we present a technique that combines data acquired with a laser range finder and the output of a standard face detection system to determine the positions of persons in the vicinity of the robot. We present experimental results obtained with a real robot which illustrate that our approach can efficiently detect faces of persons and that the overall processing time is reduced compared to a purely vision-based approach.

1 Introduction

Intelligent service robots are developed to assist people in their daily living activities. Over the last few years an increasing number of applications has been reported in a variety of different areas such as hospital service robots [13], museums [4], office buildings [20], or shops [7]. If we want the interaction between such robots and the people in their surrounding to be successful, the robots must behave as natural as possible. One important task during a natural interaction via spoken dialog, for example, is the tracking of the face of the person the robot is interacting with.

In this paper we propose a method allowing a mobile robot to continuously focus the nearest person within a group of surrounding people with a video camera embedded in the robot's human-like face. Our system combines both, vision and laser data to detect persons. It uses the range information obtained with the laser sensor to identify possible candidates in the images. We then apply a standard face detection system to identify faces of persons. Finally, we use the range information to determine the closest person. This approach has several advantages:

- In purely vision-based approaches the person always has to be visible in the current image. Laser range scanners, however, have a much wider field of view so that the robot can adjust the pan of the camera whenever a person enters the perceptual range of the robot.
- The range scanners additionally provide accurate bearing information so that sub-images containing the faces can efficiently computed.
- The segments of the images possibly containing the faces of the persons in the vicinity of the robot can be processed more efficiently than the entire image.
- The proximity information obtained with the laser range scanner can also be used to quickly and reliably extract the distance of the person. This way, a mobile robot can, for example, keep track of he closest person in its vicinity.

Accordingly, our approach combines the advantages of both, vision and laser data. The proximity data is used to extract sub-images that potentially contain faces. We then use a face detection system to verify whether or not a segment contains a face of a person. Experimental results illustrate that by combining both techniques we obtain a serious speed-up in the processing time.

The remainder of this paper is organized as follows. After discussion related work in the following section, we will describe our detection approach in Section 3. In Section 4 we then will present some results obtained with our mobile robots Albert and Rato.

2 Related Work

The problem of detecting and tracking objects or persons has been studied intensively in the area of computer vision [1, 6, 10, 9, 17, 18, 24]. Additionally, there is a variety of vision-based techniques for tracking persons with mobile robots [2, 11, 12, 15, 23]. Existing approaches use color tracking, a mixture of features, or stereo vision to identify single persons and eventually their gestures or actions. All these approaches, however, are solely based on image data and do not exploit distance information that can easily be obtained with a mobile robot equipped

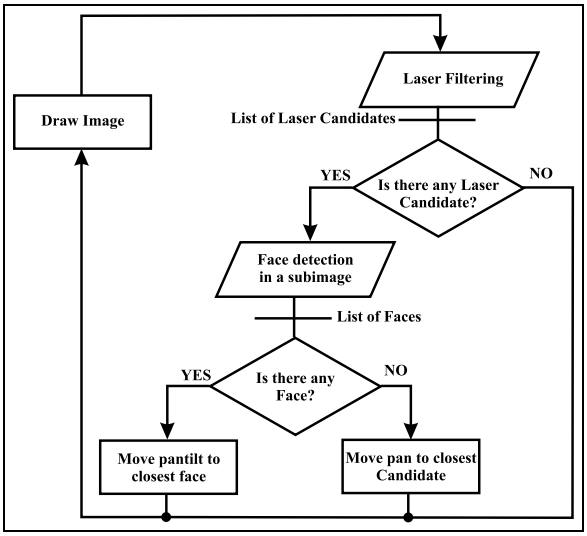

Figure 1: *Flow Chart of the face detection process.*

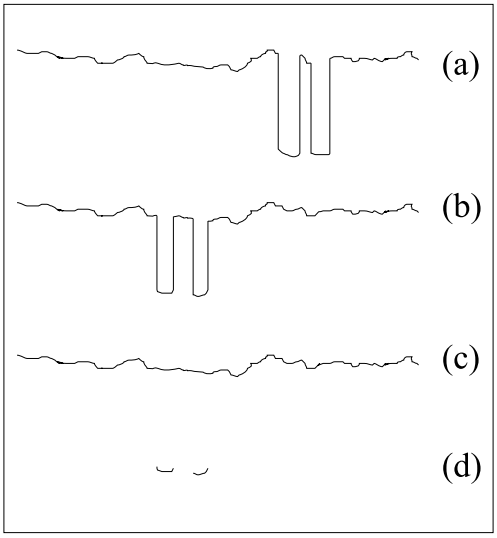

Figure 2: *Two consecutive laser scans (a and b), resulting background histogram (c), and laser candidate histogram (d) for scan b.*

with range sensors such as laser range finders. Additionally, laser range sensors have been used to locate and to track persons in range scans [8, 14, 16, 19]. These systems apply different algorithms to extract features and to represent the uncertainty of the robot about the position of the person in its vicinity.

Both approaches, the purely vision-based and the purely range-data based, have their advantages and disadvantages. 2D laser processing is significantly faster than image processing. It furthermore provides highly accurate distance information which usually has to be extracted from the images in a time-consuming process. The laser sensor covers a larger portion of the robot's vicinity than standard cameras. On the other hand, the laser provides poor features. As reported in [19], in a two-dimensional laser range scan the profile of a human can easily be generated by a trash bin or by a column. Vision sensors are cheap and generate a huge amount of information. However, detecting persons in images is not an easy endeavor and extracting distances from images usually is a time-consuming process.

In this paper we therefore propose a combination of both sensor modalities to robustly detect persons in the vicinity of the robot. Our approach uses a laser range scanner to identify potential candidates. It then extracts portions out of the camera image that correspond to the candidates in the range scan. To detect a person we then use a face detection library [17]. If a face has been detected in the camera image, we determine the distance of the person based on the range information provided by the laser. The resulting information is then used to adjust the pan and the tilt of the camera in order to focus the person with the robot's camera.

3 Person detection with a Mobile Robot

This section describes the method used to detect and to track the faces of moving people in the environment by means of a combination of a laser range and a video camera. A flow-chart of the whole process is illustrated in Figure 1. To accurately identify moving objects, the detection process is started as soon as robot has stopped. Our system operates in two stages: It first applies a filter to the laser data to detect features of moving objects. Next it applies the system described in [17] to find faces in image segments corresponding to the features in the range scans.

3.1 Laser-Based People Detection

The first step in our approach evaluates the current scan obtained with the laser range scanner. Thereby the goal is to determine a list of potential candidates that subsequently are evaluated using the vision module. The key idea of our laser data interpretation component is to use a background histogram that is continuously updated in order to distinguish moving from non-moving objects. The learning of the background histogram starts whenever the robot stops. The background histogram stores for every beam angle the maximum distance measured into this direction. Moving objects are then detected by comparing the most recent scan and the current background histogram.

Figure 2 depicts a typical sequence of laser scans obtained with a laser scanner mounted in a height of 40 cm. At this height, the legs of persons usually appear as local minima in range scans. A typical initial histogram obtained in a situation in which a person walks through the

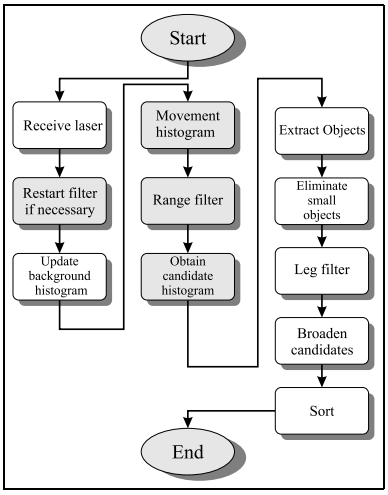

Figure 3: *Flow-chart describing the process of determining person candidates in laser range scans.*

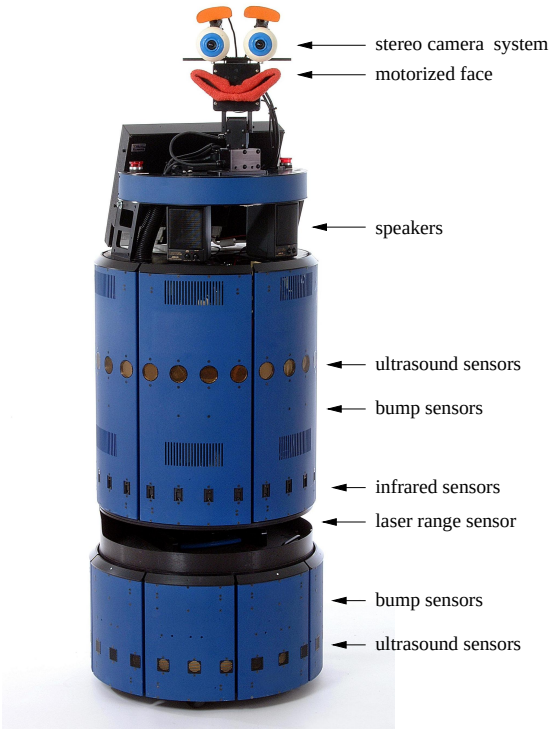

Figure 4: *The mobile robot Albert used to carry out the experiments.*

field of view of the robot is depicted in Figure 2(a). The second plot illustrated in Figure 2(b) shows another histogram obtained for one of the next scans. The resulting background histogram is depicted in Figure 2(c). Given this background histogram, the legs of the person in Figure 2(b) can directly be extracted. The resulting features are depicted in Figure 2(d).

Before we transfer any segments of the image to the face detection module, we apply the filtering procedure illustrated in Figure 3 in order to eliminate features that do not correspond to people. The most important steps of this process are the following. Whenever the robot stops we start computing the background histogram which contains for every beam the maximum distance measured in the corresponding direction. Moving objects can then easily be identified since they produce beams shorter than the maximum distance. By comparing the background histogram and the current scan, we extract those beams which are more than 30cm shorter than the maximum range measurement into this direction. We then remove all features that are too small to be caused by the legs of persons. Afterwards we cluster narrow features. This way, features belonging to a single person are grouped. For example, the individual legs of a person sometimes appear as two features. The clustering process ensures that the coordinates of the segment are computed appropriately. Please note that the clustering may lead to a wide segment possibly containing several persons. In this case, we rely on the face detection system to locate the beams corresponding to the individual persons. Finally we sort the list of candidates according to their distance to the robot. The closest candidate will be at the beginning of the list.

3.2 Integrating Laser Data with the Face Detection module

Once the list of laser candidates has been computed, we proceed with the closest one. Our strategy seeks to minimize the size of images the face detection method has to be applied to. Therefore, we only examine the region of the image that potentially contains the closest person. If no candidate can be identified in the range scan, we do not transfer the corresponding image to the face detection module. The extraction of regions in the camera images is based on a calibration of the camera and laser range scanners [3, 21]. If there is more than one face in the sub-image, there must a group of people inside the candidate. Then we compare each of the face positions to the original laser data, in order to decide which face is the closest to the robot. Next, we move the pan-tilt to it. If the currently considered segment does not contain a face, we proceed with the next segment according to our current order. In the case that no faces are detected the pan is moved to the closest feature. If again no face is detected in the next frame we also change the tilt to a constant height of 1.55m. This way we avoid tilt positions that are too high or too low. This appeared to be important because the cameras are mounted at a height of 1.4m.

Figure 5: *User Interface of the Face Detection System.*

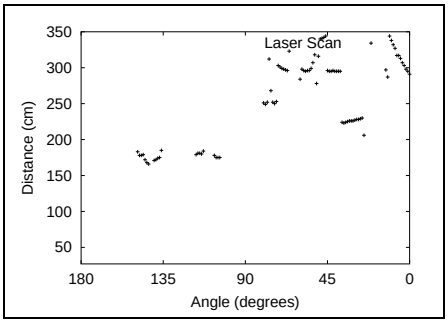

Figure 6: *Laser scan.*

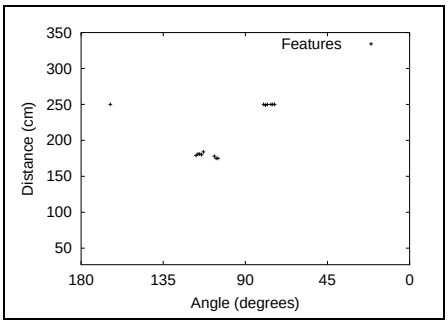

Figure 7: *Features detected in the laser scan depicted in Figure 6.*

4 Experimental Results

In order to evaluate the performance of the person detection technique, we carried out several experiments with the mobile robot Albert in the office environment of the Autonomous Intelligent Systems laboratory of the University of Freiburg. Albert is an RWI B21 robot equipped with a SICK PLS 200 laser-range finder. This sensor measures the distance to obstacles in the surrounding of the robot. The robot face includes a pant-tilt unit, two video cameras and a set of servomotors to control robot face gestures (see Figure 4).

4.1 Implementation Details

The technique described above has been implemented including all necessary communication interfaces and hardware drivers. We additionally implemented interfaces to the image processing software for the face recognition [17]. In order to reduce computation time, we have chosen a fast variant of the face detection method. Figure 5 shows the GUI of this application. It contains several displays to visualize the current data.

In order to simultaneously use both laser range and video camera data, it is necessary to establish geometric relationships between both sensors. Moreover, we need to correlate the data obtained from the camera, in pixels, into the coordinate system, in centimeters. The empirical camera model used was the pin-hole model with radial distortion correction using Tsai's method [22]. The camera parameters have been obtained with Tsai's empirical method. All experiments described below were performed with two PC's. The first one is based on an AMD Athlon, 1.7GHz and 512MB RAM, where we ran the video capture module, the communication server, and the detection program. The second is a Pentium II 300MHz with 128Mb RAM on which the laser and the pan-tilt modules were running. Both computers were connected to each other through a TCP/IP 10Mbps network.

Our system is highly efficient. It allows to process more than 3 frames per second at a resolution of 300x200, which turned out to be detailed enough for a reliable face detection.

4.2 People Detection in a Cluttered Environment

The first experiment illustrates the capabilities of our system to reliably detect persons in even cluttered environments. Figure 6 shows a typical laser scan obtained in a laboratory environment. The features extracted with our filtering technique are depicted in Figure 7. Please note that the system detects three features potentially corresponding to persons. The leftmost feature, however, is not considered, since it is too small. For the two remaining features the system determines two segments. Since the feature in the center of Figure 7 is the one that is closest to the robot, the image processing is only performed for the corresponding feature. Figure 8 shows the output of the face detection routine. Also shown there are the two segments that have been computed by the filtering process. As can be seen from the figure, the result of the overall process is quite accurate. Furthermore, the segments cover only 60% of the whole image and simultaneously contain both candidates.

From the scene in Figure 8 if the person on the right moves forward and the other person moves backwards, the camera would change focus to the person on the right

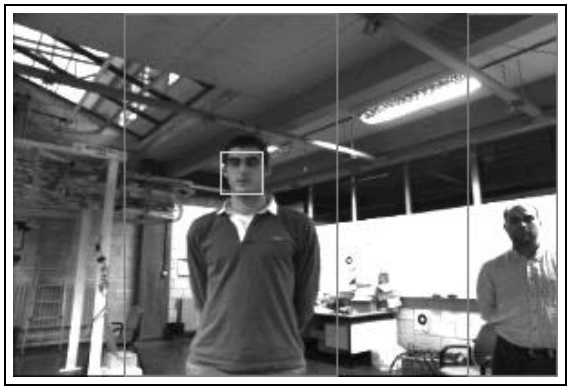

Figure 8: *Image segments for the features shown in Figure 7 and face detected in the closest segment.*

Figure 9: *Person of Figure 8 have changed their position.*

as it is shown in Figure 9.

If a smaller person is standing in front of a taller one, both faces will be detected and the camera will be focused to the face of the smaller person. This make sense since the smaller person must be the closest, otherwise it will be occluded by the taller one.

4.3 Speed-up Obtained by Combining Laser and Vision

The second experiment is designed to demonstrate the speed-up obtained by combining laser and vision information. For this purpose we performed several experiments and measured the processing time needed to determine a face. We compared the time needed by our algorithm to the time required by the face detection system applied to the whole images. The results for different resolutions are summarized in the Tables 1 and 2.

Table 3 contains the overall reduction in computing time. As can be seen, the exploitation of the range data seriously decreases the required computation time. Please note that the processing time of a laser measurement was about 13 ms. At first glance, it seems that one could expect a greater reduction. However, the following points need to be considered

Table 1: *Full image detection: Image processing time for different resolutions of the images*

Resolution	Average Time [ms]	Number of Images
300x200	341.24	89
450x300	702.28	103
600x400	1252.59	58

Table 2: *Combined detection: Image processing time for different resolutions of the images*

Resolution	Average Time [ms]	Number of Images
300x200	228.34	92
450x300	383.71	172
600x400	636.14	69

- In our current system, the search for faces is carried out in a 90 cm wide section in the environment. If the face is close to the camera, the 90 cm range covers a large part of the image. Throughout the experiments described here, the people were relatively close to the robot (and the camera) so that large portions of the images had to be processed.
- Additionally, The range of the camera is approximately 60 degrees whereas the laser sensor we used covers 180 degrees. Our system only has to process a third of all images compared to a situation in which one solely relies on vision. In such a case one would have to rotate the pan-tilt unit in order to capture three pictures at different angles so as to cover the corresponding area.

5 Summary and Conclusions

In this paper we presented an approach to combine laser and vision data for robust and efficient face detection with mobile robots. Our robot first extracts features of persons from range scans and determines segments in the images that might contain the faces. It then applies an image processing library to detect the faces in these segments.

Our approach has been implemented and tested on a real robot. The experiments we carried out illustrate that our

Table 3: *Improvement on the processing time*

Resolution	Improvement
300x200	29.27%
450x300	42.00%
600x400	48.18%

approach is able to robustly detect persons even in cluttered environments. The experiments furthermore show that the approach significantly reduces the computation time needed compared to the purely vision-based approach.

Future work will take the results reported here as a basis for people tracking and identification systems like, for example, the one described in [5]. The technique presented in this paper provides an efficient solution to the task of identifying faces in images. We expect that this will make person identification more robust and thus will lead to improvements of human-robot interaction.

Acknowledgments

We would like to thank all the people of the University of Freiburg for helping during the Jorge Blanco's visit to the Autonomous Intelligent Systems laboratory. This work has been partially funded by the Spanish CYCIT project DPI2002-04377-C02-02 as well as by the IST Programme of Commission of the European Communities under contract number IST-2000-29456.

References

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding: CVIU*, 73(3):428–440, 1999.

[2] D. Beymer and Konolige K. Tracking people from a mobile platform. In *IJCAI-2001 Workshop on Reasoning with Uncertainty in Robotics*, 2001.

[3] J. Blanco. People detection with a mobile robot based on laser-range sensor and artificial vision. Master's thesis, Department of System Engineering and Automation, University of Vigo, 2002. In Spanish.

[4] W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2), 1999.

[5] G. Cielniak, M. Bennewitz, and W. Burgard. Where is ...? learning and utilizing motion patterns of persons with mobile robots. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003. To appear.

[6] T. Darrell, B. Moghaddam, and A. P. Pentland. Active face tracking and pose estimation in an interactive room. In *Proc. of the IEEE Sixth International Conference on Computer Vision*, pages 67–72, 1996.

[7] H. Endres, W. Feiten, and G. Lawitzky. Field test of a navigation system: Autonomous cleaning in supermarkets. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 1998.

[8] A. Fod, A. Howard, and M. J. Matarić. Laser-based people tracking. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[9] D.M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), 1999.

[10] I. Haritaoglu, D. Harwood, and L. Davis. A real time system for detecting and tracking people. In *Proc. of the Int. Conference on Automatic Face and Gesture Recognition*, 1998.

[11] I. Horswill. Polly: A vision-based artificial agent. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1993.

[12] Roger E. Kahn, Michael J. Swain, Peter N. Prokopowicz, and R. James Firby. Gesture recognition using the perseus architecture. Technical Report TR-96-04, University of Chicago, 19, 1996.

[13] S. King and C. Weiman. Helpmate autonomous mobile robot navigation system. In *Proceedings of the SPIE Conference on Mobile Robots*, 1990.

[14] B. Kluge, C. Koehler, and E. Prassler. Fast and robust tracking of multiple moving objects with a laser range finder. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2001.

[15] D. Kortenkamp, E. Huber, and R. P. Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1996.

[16] M. Montemerlo, S. Thun, and W. Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[17] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1), 1998.

[18] H. Schneiderman and T. Kanade. Probabilistic modeling of local appearance and spatial relationships for object recognition. In *Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[19] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving objects with a mobile robot. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[20] R. Simmons, R. Goodwin, K. Z. Haigh, S. Koenig, J. O'Sullivan, and M. M. Veloso. Xavier: Experience with a layered robot architecture. *ACM magazine* Intelligence, 1997.

[21] R. Triebel. Generating textured 3d-models with mobile robots. Master's thesis, School of Computer Science, University of Freiburg, 2001. In German.

[22] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.

[23] S. Waldherr, S. Thrun, R. Romero, and D. Margaritis. Template-based recognition of pose and motion gestures on a mobile robot. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1998.

[24] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.