Part based people detection using 2D range data and images

Zoran Zivkovic and Ben Kröse

Abstract— **This paper addresses the problem of people detection using 2D range data and omnidirectional vision. The range data is often used in robotics to detect person legs. First, we will develop a reliable 2D range data leg detector using the AdaBoost algorithm. Second, instead of using the leg detector to detect people, we propose a more reliable method that takes into account the spatial arrangement of the detected legs. The method is inspired by the latest results on the "part-based representations" from the computer vision area. Finally, we show how the part-based representations can be used to combine the leg detections from the 2D range data and upper body, lower body and full body detections from omnidirectional camera. The experimental results show highly reliable people detection when the two sensors are combined. We use Haar-like features and the AdaBoost again to construct the omnicam body parts detectors.**

I. Introduction

Robots are moving out of laboratories into public places where the human beings have to be taken into account. Such robots should be able to interact with humans and show aspects of human style social intelligence. This also implies that in addition to the perception required for the conventional functions (localization, navigation, etc.), a "socially interactive" robot needs strong human oriented perceptual capabilities [12]. As a starting point the robot should be able to accurately and robustly detect and localize the persons around it [16].

In this paper we first develop a reliable 2D range data leg detector using AdaBoost. The AdaBoost was used recently in [7] to automatically construct a leg detector from an extensive set of geometric features extracted from 2D scans. The SICK laser scanner as used in [7] and we use the same set of features but now for the URG-04LX 2D range scanner.

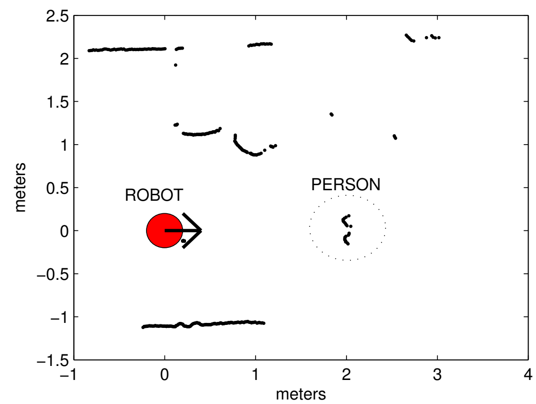
Second, instead of using the leg detector to detect people directly as in [7], we propose a more reliable method that takes into account the spatial arrangement of the detected legs. The method is inspired by the latest results on the "part-based representations" from the computer vision area and the work of Weber, Perona and colleagues [14], [4]. The approach also takes into account that the leg detector might produce false detections or fail to detect legs, for example because of partial occlusion.

Finally, we will show how our part-based model can be used to combine information from 2D range data and omnidirectional vision. In the omnidirectional images we detect persons upper body, lower body and full body using a large set of Haar-like features and again the AdaBoost algorithm [13]. The upper body, lower body and full body detections are then combined with the leg detections from the 2D range scan. In the experimental section we show that the probabilistic combination of part detections performs much better than each part separately. Furthermore, we demonstrate the robustness of the proposed method to partial occlusions.

This paper starts with the related work which is presented in Section 2. It follows with the description of constructing the 2D range scan leg detector using the AdaBoost in Section 3. Furthermore, we present our part-based probabilistic model in Section 4, and the way to combine the information from 2D range data and omnidirectional vision in Section 5. The results from our experiments are presented in Section 6. Finally, conclusions are given in Section 7.

Published at IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007. The work was supported by EU-project Cogniron.

Zoran Zivkovic and Ben Kröse are with ISLA lab, University of Amsterdam, The Netherlands, email: `zivkovic@science.uva.nl`



a) Example 2D range scan from URG-04LX laser range scanner

b) Corresponding panoramic image from the robot omnidirectional camera. The 2D laser range scan points from above are projected onto the image.

Fig. 1. Example data. The camera and the scanner are calibrated: their intrinsic parameters and the relative pose are known.

II. Related work

A 2D laser range scanner is often used in robotics for detecting and tracking people [9], [7]. People are detected by finding their legs in the laser scans. Disadvantages of using the laser scans for people detection are: the persons can be detected only at limited distances from the robot, low detection rate in highly cluttered environments and that the methods fail when the person legs are occluded.

Person detection from images is a widely studied problem in the computer vision area. Many of the presented algorithms aim at the surveillance applications [5] and are not applicable to mobile platforms since they assume static camera. There is also a large number of papers considering the people detection without the static camera assumption, e.g. [6], [10], [15], [11]

The people detection can be seen as a part of the more general problem of object detection. A recent modern approach for object detection in the computer vision area is the so called ”part-based representation”. Various part-based representations, e.g. [1], [4], [2], [3], are demonstrated to lead to higher recognition rates. An important advantage of the part-based approach is it relies on object parts and therefore it is much more robust to partial occlusions than the standard approach considering the whole object.

The part-based people detection was considered a number of times. Seemann et al. [10] use SIFT based part detectors but do not model part occlusions. Wu and Nevatia [15] describe the part occlusions but the occlusion probabilities and part positions are learned in a supervised manner. We base our algorithm on a principled probabilistic model of the spatial arrangement of the parts similar to the work of Weber, Perona and colleagues [14], [4]. An advantage of having a proper probabilistic model is that, after constructing the part detectors, the part arrangement and occlusion probabilities can be automatically learned from unlabelled images. In this paper we apply the part-based approach to 2D range data and show how the part-based approach can be used to properly combine information from 2D range data and omnidirectional vision.

III. Detecting legs in 2D range scans

An example 2D range scan is presented in Figure 1a. The appearance of the legs can change drastically and sometimes the legs cannot be separated. Another problem is that in cluttered environments many other objects, e.g. chairs or tables, may produce 2D scan output similar to human legs.

We will follow the common approach for leg detection as in [7]:

- **Step 1:** Split the range scan into a number of segments - abrupt changes in the range scan are used as segment borders. We use the Canny edge detector to detect the abrupt changes in the range scan.
- **Step 2:** Calculate geometric features for each segment - we will use the same set of features as in [7], see Table I. The features that depend on motion are removed since we also want to detect persons when they are standing still.
- **Step 3:** Run a previously trained classifier on the extracted geometric features to decide if the 2D range scan segment resembles a persons leg.

To train the classifier we need a training set of positive and negative scan segment examples. We use AdaBoost algorithm to train the classifier. The Adaboost algorithm is a simple and efficient way to build a classifier from a large set of features when we do not know which features are relevant [13].

feat. no.	description
1	number of points
2	standard deviation
3	mean average deviation from the median
4	jump distance from the preceding segment
5	jump distance from the succeeding segment
6	segment width - distance between the first and the last point
7	linearity - average distance of the points from the fitted line
8	circularity - average distance of the points from the fitted circle
9	radius of the circle fitted to the points
10	segment length
11	segment regularity - standard deviation of the distances of the adjacent points
12	mean curvature - mean of the curvature calculated for each three successive points
13	mean angular difference - mean of the angles formed by each three successive points

TABLE I

GEOMETRIC FEATURES EXTRACTED FROM 2D RANGE SEGMENTS.

IV. Probabilistic part-based person model

The leg detector from previous section can be used to detect people directly [7]. We present here a part-based model that takes into account the possible distance between the persons legs and it is more robust to occlusions and clutter. The presented probabilistic model that be learned automatically from a set of training data.

A. Part detection

A human is detected by detecting P human body parts. In our case $P = 2$ for the legs. The 2D position of one leg will be denoted by $\mathbf{x}_p = (x_p, y_p)$. We will use the Gaussian distribution as the probabilistic model of the leg positions:

$$p_{shape}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mu, \Sigma) \tag{1}$$

where $\mathbf{x} = \begin{pmatrix}\mathbf{x}_1 & \dots & \mathbf{x}_P\end{pmatrix}$ is a $2P$ long vector containing all the 2D part positions, μ is the mean and Σ is a $(2P) \times (2P)$ covariance matrix. If the covariance matrix is diagonal than this model can be seen as describing ”string-like” constraints between the body-part positions [3]. The non-diagonal covariance matrix will express additional relations between the body parts.

Given a range scan and a set of segments extracted from the scan we will apply the AdaBoost classifier from the previous section to detect person’s legs. Let N denote the number of segments classified as legs and let $\mathbf{x}_j$ denote the 2D position of the j-th detection. All leg detections from one scan will be denoted as:

$$\mathcal{X} = \begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N\end{pmatrix} \tag{2}$$

The 2D image position $\mathbf{x}_j = (x_j, y_j)$ of the j-th detection is calculated as the median position of the scan segment points. For different scans the $\mathcal{X}$ can have different lengths. Sometimes the legs cannot be separated or their appearance in the scan might change drastically. Furthermore, in cluttered environments many other objects, e.g. chairs or tables, may produce 2D scan output similar to human legs and some detections might be false detections.

B. Missing detections and clutter

From a range scan we will extract the collection of person's leg candidates $\mathcal{X}$ but we do not know which are true and which are false detections. To indicate which detections are correct we will use $P = 2$ element vector $\mathbf{h}$ with element $h_p = j$, $j > 0$, indicating that the j-th detection $\mathbf{x}_j$ belongs to the of the p-th body part (leg) and the other detections of that part are false detections. Given $\mathbf{h}$ the 2D positions of the person's legs are composed of the corresponding detections $\mathbf{x} = \begin{pmatrix}\mathbf{x}_{h_1} \mathbf{x}_{h_2}\end{pmatrix}$. We will denote the set of all other detections that belong to the background clutter as $\mathbf{x}^{bg}$.

It is also possible that the part was not detected at all. We use $h_p = 0$ to indicate this. The position of not detected leg is considered as missing data. To make distinction between the missing and the observed parts we will denote the set of missing parts as $\mathbf{x}^m$ and the set of observed parts as $\mathbf{x}^o$. To indicate the fact that there can be missing parts, the probabilistic model of the arrangement of the body parts (1) will be written as: $p_{shape}(\mathbf{x}) = p_{shape}(\mathbf{x}^o, \mathbf{x}^m)$.

C. Probabilistic model

A probabilistic model that considers the possibility of part detector false alarms and missed detections of body parts of a person can be written as joint distribution:

$$p(\mathcal{X}, \mathbf{x}^m, \mathbf{h}) = p(\mathcal{X}, \mathbf{x}^m | \mathbf{h}) p(\mathbf{h}) \tag{3}$$

where $\mathbf{x}^m$ are the missing data. The number of missing parts in $\mathbf{x}^m$ is determined by $\mathbf{h}$.

In order to define $p(\mathbf{h})$ we will introduce two auxiliary variables $\mathbf{b}$ and $\mathbf{n}$. The variable $\mathbf{b} = sign(\mathbf{h})$ is a binary vector that denotes which parts have been detected and which not. The value of the element $n_p \leq N_p$ of the vector $\mathbf{n}$ represents the number of detections of part p that are assigned to the background clutter. We can now write the joint distribution (3) as:

$$p(\mathcal{X}, \mathbf{x}^m, \mathbf{h}, \mathbf{n}, \mathbf{b}) = p(\mathcal{X}, \mathbf{x}^m | \mathbf{h}) p(\mathbf{h} | \mathbf{n}, \mathbf{b}) p(\mathbf{n}) p(\mathbf{b}) \tag{4}$$

where we add the two auxiliary variables $\mathbf{b}$ and $\mathbf{n}$ and assume independence between them.

Furthermore, we have:

$$p(\mathcal{X}, \mathbf{x}^m | \mathbf{h}) = p_{shape}(\mathbf{x}^o, \mathbf{x}^m) p_{bg}(\mathbf{x}^{bg}) \tag{5}$$

where the observed parts $\mathbf{x}^o$, the missing parts $\mathbf{x}^m$ and the false detections from clutter $\mathbf{x}^{bg}$ correspond to the hypothesis $\mathbf{h}$ and the $p_{bg}(\mathbf{x}^{bg})$ describes the distribution of the false detections. We will assume some small constant as the uniform density for the false detections $p_{bg}(\mathbf{x}^{bg})$.

The probability $p(\mathbf{b})$ describing the presence or absence of parts is modelled as an explicit table of joint probabilities. Each part can be either detected or not, so there are in total 2^P possible combinations that are considered in $p(\mathbf{b})$.

We assume here that the background part detections $\mathbf{n}$ are independent of each other and modelled using Poisson distribution with mean M_p [14]. Different M_p-s for different parts admit different detector statistics. The Poisson parameter will be denoted by vector $\mathbf{M} = \begin{pmatrix} M_1 & \dots & M_P \end{pmatrix}$.

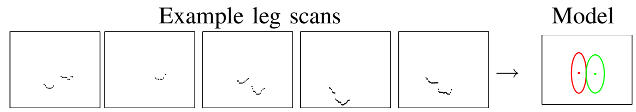


Fig. 2. Example person's legs scans from the data set used to train our probabilistic part-based model and the learned model parameters. For each part we present its mean position contained in the parameter μ. The ellipse represents the 1-sigma uncertainty of the part position as described by the diagonal elements of the covariance matrix Σ.

The density $p(\mathbf{h}|\mathbf{n}, \mathbf{b})$ is defined as:

$$p(\mathbf{h}|\mathbf{n}, \mathbf{b}) = \begin{cases} 1/|\mathcal{H}(\mathbf{b}, \mathbf{n})| & \text{if } \mathbf{h} \epsilon \mathcal{H}(\mathbf{b}, \mathbf{n}), \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

where $\mathcal{H}(\mathbf{b}, \mathbf{n})$ is the set of all hypotheses consistent with the values of $\mathbf{b}$ and $\mathbf{n}$. Here $|\mathcal{H}(\mathbf{b}, \mathbf{n})|$ denotes the total number all consistent part assignment hypotheses. This expresses that these hypotheses are considered equally likely.

D. Learning model parameters

The density distribution (4) will have the following set of parameters $\Omega = \{\mu, \Sigma, p(\mathbf{b}), \mathbf{M}\}$. Therefore we can write the distribution (4) also as:

$$p(\mathcal{X}, \mathbf{x}^m, \mathbf{h}) = p(\mathcal{X}, \mathbf{x}^m, \mathbf{h} | \Omega) \tag{7}$$

The likelihood of a collection of detected parts $\mathcal{X}$ is obtained by integrating over the hidden hypotheses $\mathbf{h}$ and the missing parts:

$$p(\mathcal{X}|\Omega) = \sum_{\text{all possible } \mathbf{h}} \int_{\mathbf{x}^m} p(\mathcal{X}, \mathbf{x}^m, \mathbf{h} | \Omega). \tag{8}$$

We use Gaussian distribution to describe the arrangement of the body parts. Integrating over the missing parts $\mathbf{x}^m$ for the Gaussian distribution can be performed in closed form.

To estimate the parameters of the model we start from a set of L aligned images of persons. The part detectors are applied to each image. The collection of detected parts for i-th image will be denoted as $\mathcal{X}_i$. The maximum likelihood estimate of the parameters Ω is computed by maximizing the likelihood of the data:

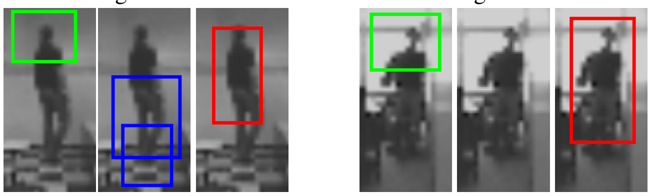
$$\prod_i^L p(\mathcal{X}_i | \Omega) \tag{9}$$

Once we have the part detectors, the part arrangement parameters are estimated using expectation maximization algorithm from a set of unlabelled scans. Details of the algorithm are the same as in [14].

E. Detection

Let us denote the maximum likelihood parameters learned from a set of scans of persons as Ω_{person}, see Figure 4. The parameters of the model can be also learned. These parameters will be denoted as Ω_{bg}. We are now presented with a new scan and extracted the set of detected parts $\mathcal{X}$. The scan is either a scan of a person or some background clutter:

$$p(\mathcal{X}) = p(\mathcal{X}|Person) p(Person) + p(\mathcal{X}|BG) p(BG) \tag{10}$$



Part detections – person 1. Part detections – person 2.

Fig. 3. Example body part detections in low resolution images. In the example on the left the legs are detected at two different positions and scales. In the example on the right side the legs are not detected at all.

Fig. 4. Example images from the data set used to train our probabilistic part-based model and examples of learned part arrangement model parameters. For each part we present its mean position contained in the parameter μ. The ellipse represents the 1-sigma uncertainty of the part position as described by the diagonal elements of the covariance matrix Σ. Here green color represents head, blue are legs and red is the full body detector.

where $p(Person)$ and $p(BG)$ are unknown a priori probabilities that the scan contains a person or background. The a posteriori probability that there is a person is:

$$p(Person|\mathcal{X}) = \frac{p(\mathcal{X}|Person)p(Person)}{p(\mathcal{X})} \approx \frac{p(\mathcal{X}|\Omega_{person})p(Person)}{p(\mathcal{X}|\Omega_{person})p(Person) + p(\mathcal{X}|\Omega_{bg})p(BG)} \tag{11}$$

The last step above is an approximation since we use the maximum likelihood estimates for the model parameters Ω_{person} and Ω_{bg} instead of integrating over all possible parameter values. Calculating $p(\mathcal{X}|\Omega)$ is done using (8).

V. Combining information from range scans and omnidirectional vision

Robots are often equipped with a vision sensor. In this section we will show how the part based model from the previous section can be extended to include part detections from the corresponding image.

A. Part detection in images

We use a set of Haar-like-feature classifier to detect various human body parts in images. Each classifier is trained using AdaBoost algorithm on a large set of example images of the corresponding body part [13]. In this paper the classifiers are trained on upper body, lower body and full body images. The part detectors can lead to many false alarms and missed detections [8], see Figure 3.

B. Extending the part-based model

The part based model from the previous section that was applied to the 2D range leg detections can be easily extended with the human body parts detected in images. We have now $P = 2 + 3 = 5$ body parts and $\mathbf{x} = \begin{pmatrix}\mathbf{x}_1 & \ldots & \mathbf{x}_P\end{pmatrix}$ is a $2P$ long vector containing 2D leg positions and the 2D image positions for the upper body, lower body and full body detected in images.

The positions of all detected parts are summarized in a data structure:

$$\mathcal{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \ldots & \mathbf{x}_{1,N_{leg1}} & \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \ldots & \mathbf{x}_{2,N_{leg2}} & \\ \mathbf{x}_{3,1} & \mathbf{x}_{3,2} & \ldots & \ldots & \mathbf{x}_{3,N_{up-body}} \\ \mathbf{x}_{4,1} & \mathbf{x}_{4,2} & \ldots & \mathbf{x}_{4,N_{low-body}} & \\ \mathbf{x}_{5,1} & \mathbf{x}_{5,2} & \ldots & \ldots & \mathbf{x}_{5,N_{full-body}} \end{pmatrix} \tag{12}$$

with one row per part and where each row contains information about the detections of the corresponding body part. Since we use the same detector for both legs detected in the range scans the first two rows are repeated. The element $\mathbf{x}_{p,j}$ contains the 2D positions for the legs or the 2D image position for the parts detected in images of the j-th detection of the p-th part. The rows of $\mathcal{X}$ can have different lengths and some might be empty if that part is not detected in the image.

Since we do not know which detections are false and which true we will again use the hidden P dimensional assignment vector $\mathbf{h}$ with element $h_p = j$, indicating that the j-th detection of the p-th part $\mathbf{x}_{p,j}$ belongs to the object and other detections of that part are false detections. Given $\mathbf{h}$ the shape of the object is composed of the corresponding detections $\mathbf{x} = \begin{pmatrix}\mathbf{x}_{1,h_1} & \ldots & \mathbf{x}_{P,h_P}\end{pmatrix}$. Note that since we use the same detector for both legs care should be taken not to select the same leg detection for both legs.

The other model equations remain the same and the same procedure can be used to learn now the part based model containing the part detectors from both sensors. Example part based model learned from image detections is presented in Figure 4.

VI. Experimental results

In this section the part detectors and the part based model are evaluated.

A. Leg detection evaluation

In order to train and evaluate our range scan leg detector we recorded a dataset of 2D range scans. During the experiments the URG-04LX 2D range scanner was mounted on our robot at 50cm above the floor. In the first experiment we recorded 1950 scans while driving the robot through the corridors in our building. As result we obtained 2890 scan segments corresponding to person's legs and 6978 segments from the background clutter. In the second experiment we recorded 1580 scans while driving the robot within the cluttered offices in our building, see Figure 1. As result we obtained 2142 scan segments corresponding to person's legs and 7061 segments from the background clutter.

Half of the data was used to train our leg detector using AdaBoost and the other half is used to evaluate the detector. The procedure is repeated 10 times for different random

Environment	True positives (%)	False positives (%)
Corridor	90.1 +- 1.4	2.7 +- 0.6
Office	76.8 +- 1.2	7.2 +- 0.8
Both	80.1 +- 2.5	6.5 +- 0.7

TABLE II

RECOGNITION RESULTS FOR LEG DETECTION FROM THE 2D RANGE SCANS WITHIN DIFFERENT ENVIRONMENTS. THE STANDARD DEVIATION OF THE RESULTS FOR 10 RANDOM TRIALS IS ALSO REPORTED.

Environment	First five features
Corridor	9, 9, 6, 7, 1
Office	9, 6, 7, 9, 6
Both	9, 9, 6, 7, 8

TABLE III

FIRST FIVE GEOMETRIC FEATURES SELECTED FOR DETECTING PERSON'S LEGS IN THE 2D RANGE SCANS USING ADABOOST.

test and training data sets. In Table II we summarize the recognition results. The recognition is first tested using just the data form obtained in the corridor and then just the data from the office environment. Finally, we evaluated the results on all data. We can observe that the recognition in corridor was reasonably high and comparable to the results from [7]. However in the cluttered office environment the recognition results degraded heavily. The reason might be our difficult office dataset. Furthermore, while SICK 2D range scanner was used in [7], we used a less accurate URG-04LX.

For each scan segment we extract the geometric features reported in Table I. The AdaBoost algorithm will automatically select the relevant features during training. Separate features are used to build the so called "weak" classifier and they are then combined to form the final classifier. We used 25 classifiers to build the final classifier. The number of classifiers was chosen to produce the best recognition results. It is interesting to see which features are selected to be used first as the most relevant ones to build the final classifier. Note that some features might be selected a number of times. The first five features chosen by the Adaboost are reported in Table III. As in [7], the radius of the fitted circle (feature 9) was chosen as the most informative one. The segment width (feature 6) was also often used. This can be explained by the fact that it was more robust representation of the size of the leg since the fitted circle was sometimes influenced by the noise. Furthermore, the feature 7 describing the linearity of the segment was also important since there are usually many linear segments in man-made environments, e.g. walls, that can be then discarded using this feature. In [7] jump distance from the preceding and the succeeding segment (features 4 and 5) were selected as relevant. In our dataset the persons were at various distances from the background and this rendered these features less important. The mean angle along the segment (feature 13) can be used to distinguish between convex and concave segments[7] but in our environment there were not many concave segments so the feature was not considered important.

Detection method	True positives (%)	False positives (%)
Leg detector	87.1	3.7
Part based model (legs in the range scans)	91.2	2.5
Upper body	72.1	15.8
Lower body	55.3	30.7
Full body	69.2	20.5
Part based model (images only)	85.1	7.9
Part based model (images and range scans)	**95.1**	**1.5**

TABLE IV

RECOGNITION RESULTS FOR PEOPLE DETECTION USING LASER DATA, IMAGE DATA AND COMBINATION OF THE TWO SENSORS.

B. People detection using part based model

The Haar-like-feature based image part detectors we used in our experiments were trained on the MIT pedestrian dataset [8] and are available in the Intel OpenCV library.

For training the part-arrangement model, we collected 1200 images of people (56×112 pixels) and corresponding laser scan segments. The images are cut out of the omnidirectional images obtained from our robot in an office environment. 600 images and scans were used to train our part based model and other 600 were used for testing. The automatically learned part based model parameters for the image detections are in Figure 4. We can observe that there is more uncertainty in the positions of the full and lower body than for the upper body region. Learning the part-based model parameters takes around 10 seconds for 600 images in our Matlab implementation.

The recognition results on this data set containing alligned images of people and corresponding scan segments are summarised in Table IV. The AdaBoost leg detector leads to similar recognition results as on the larger dataset form previous section. The part based model applied to the range scan detections takes into account that the leg positions and the recognition results slightly improve. The body part detectors in images lead to poor detection results but when combined the results get much better. Finally, the best results are obtained when we combine the part detectors from images and range scans using our part arrangement model.

C. Recognition from the robot

The part based model detection is implemented on our robot. We assume that people walk over a flat ground floor surface - true for most man-made environments, Figure 6. We define a set of possible 2D floor positions $\mathcal{T}_t$. In our experiments we use a $10m \times 10m$ area around the robot and a grid of possible positions at every $10cm$. This gives 10000 possible floor points $\mathcal{T}_t$ to evaluate our part based model. For each position we extract corresponding part of the omnicam image to get aligned image and properly scaled image, as in Figure 4. We used data from the National Center for Health Statistics (*www.cdc.gov/ nchs/*). For adult humans, the mean height is 1.7m with a standard deviation of 0.085m. We define maximal height of human to be mean plus three

Two correct detections of partially occluded people.

Two correct detections. The persons are in the dark and hardly visible

Fig. 5. Example people detection results in panoramic images recorded from a moving robot.

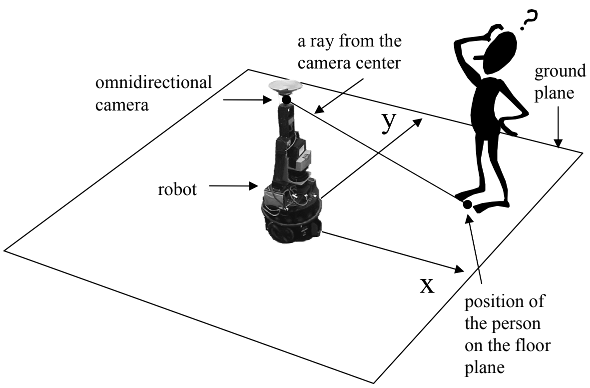


Fig. 6. Schematic representation of our Nomad robot with a calibrated omnidirectional camera on the known floor plane.

standard deviations and the width to be 1/2 of the height. Using these dimensions and having a calibrated camera, each $\mathcal{T}_t$ defines a rectangle region of interest in an image. For each floor position we also extract the corresponding segments from the range scan and use (11) to decide if there is a person at that floor position. Since (11) is computed at a dense grid of ground points, it often has large values for a number of ground points around the position where the person actually is. Therefore the persons are detected as the local maxima of (11).

The first stage of the algorithm where the body parts are detected in the omnidirectional images is the most computationally expensive. Running the three Haar-like-feature based part detectors on a 600×150 panoramic image takes on average 400ms on a 2GHz PC. This is the time needed for checking every image position and all possible part sizes. The possible part sizes start from the initial part size and then the part size is increased 1.1 times while it fits the image. The floor constraint can heavily reduce the number of positions and part sizes to search and detction can be done in around 100ms. Once the parts are detected, detecting persons using our model takes around 15ms. Currently, the people detection can be performed 7.5 times/second in our implementation on a 2GHz single processor. In Figure 5 we present a few panoramic images with the detection results.

VII. Conclusions and further work

People detection from a moving robort using multiple sensors is considered. First, we analyzed the performance of the AdaBoost based classifier for detecting human legs in 2D range scans obtained using the URG-04LX scanner. Furthermore we presented a robust people detection algorithms inspired by the latest results on the "part-based representation" from the computer vision area. The approach is based on the probabilistic combination of fast human body part detectors. The algorithm can be applied to 2D range scan leg detections but also to combine part detectors from various sensors. We used the Haar-feature based cascade classifiers to detect different human body parts in images. These detectors lead often to many false alarms and missed detections. Our principled probabilistic representation combines the part detections and can achieve person detection robust to partial occlusions, part detector false alarms and missed detections of body parts.

References

[1] M.C. Burl, T.K. Leung, and P. Perona. Recognition of planar object classes. *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.

[2] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[3] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *Intl. Journal of Computer Vision*, 61(1):55–79, 2005.

[4] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.

[5] J. Ferryman and J.L. Crowley, editors. *Proc. of the 9th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*. 2006.

[6] D.M. Gavrila and V. Philomin. Real-time object detection for smart vehicles. *In Proc. of the Intl. Conf. on Computer Vision*, 1999.

[7] K.Arras, O.Mozos, and W.Burgard. Using boosted features for detection of people in 2D range scans. *In Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 2007.

[8] H. Kruppa, M. Castrillon-Santana, and B. Schiele. Fast and robust face finding via local context. In *Proc of the IEEE Intl. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003.

[9] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. People tracking with a mobile robot using sample-based joint probabilistic data association filters. *International Journal of Robotics Research*, 22(2):99–116, 2003.

[10] E. Seemann, B. Leibe, K. Mikolajczyk, and B. Schiele. An evaluation of local shape-based features for pedestrian detection. *In Proc. of the British Machine Vision Conference*, 2005.

[11] S.Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(11):1863–1868, 2006.

[12] T.Fong, I. Nourbakhsh, and K.Dautenhahn. A survey of socially interactive robots. *Robots and Autonomous Systems*, 42:143–166, 2003.

[13] P.A. Viola and M.J. Jones. Rapid object detection using a boosted cascade of simple features. *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[14] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. *In Proc. of the European Conf. on Computer Vision*, 2000.

[15] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *In Proc. of the Intl. Conf. on Computer Vision*, 2005.

[16] W.Zajdel, Z.Zivkovic, and B.Krose. Keeping track of humans: have I seen this person before? *In Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 2005.