Int J Soc Robot
DOI 10.1007/s12369-009-0041-3

Multi-Part People Detection Using 2D Range Data

Oscar Martinez Mozos · Ryo Kurazume · Tsutomu Hasegawa

Accepted: 21 December 2009

Abstract People detection is a key capacity for robotics systems that have to interact with humans. This paper addresses the problem of detecting people using multiple layers of 2D laser range scans. Each layer contains a classifier able to detect a particular body part such as a head, an upper body or a leg. These classifiers are learned using a supervised approach based on AdaBoost. The final person detector is composed of a probabilistic combination of the outputs from the different classifiers. Experimental results with real data demonstrate the effectiveness of our approach to detect persons in indoor environments and its ability to deal with occlusions.

Keywords Laser-based people detection · Multiple cue classification · Sensor fusion · Multi-part object detection

1 Introduction

Detecting people is a key capacity for service robots that have to interact with humans [3, 16, 21]. A robust detection of persons in the environment will improve the ability of these systems to communicate with people and to take decisions. Additional applications of people detection can be found in autonomous vehicles [9, 15]. In this case the main objective is to detect pedestrians to change the behavior of the vehicle accordingly.

In this paper we address the problem of detecting people in indoor environments using 2D laser range finders. These kind of proximity sensors are often used in robotic applications since they provide a wide field of view and a high data rate. In addition, their measurements are invariant to illumination changes. Previous works have used 2D laser range finders to detect people in the environment. Typically, the lasers are located at a height which permits the detection of legs [2–5, 7, 13–16]. Although good classifications rates have been obtained using machine learning techniques [2, 15], there is still the need to improve the robustness of the final detectors. One of the main problems when using lasers at this height is the little information about legs that is provided by the range scans. An example is shown in the bottom right of Fig. 1. Here, the legs of a person are represented by short segments composed of few points. In cluttered environments, like homes or offices, these segments can be easily misclassified with other objects in the environment such as tables, chairs or other furniture. Moreover, occlusions often occur and make the detection of people quite difficult, or even impossible when the legs are hidden.

The key idea of this work is to improve the robustness of people detection systems by taking into account different body parts. Our approach uses 2D laser range scans which are situated at different heights. Each laser is responsible for detecting a different body part like the legs, the upper body or the head. The output of the different detectors is then combined in a probabilistic framework to obtain a robust final classifier. The complete system is shown in the left image of Fig. 1.

O.M. Mozos (✉)
Dept. of Computer Science and System Engineering, University of Zaragoza, Zaragoza, Spain
e-mail: ommozos@unizar.es

R. Kurazume · T. Hasegawa
Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

R. Kurazume
e-mail: kurazume@ait.kyushu-u.ac.jp

T. Hasegawa
e-mail: hasegawa@irvs.is.kyushu-u.ac.jp

Published online: 01 January 2010

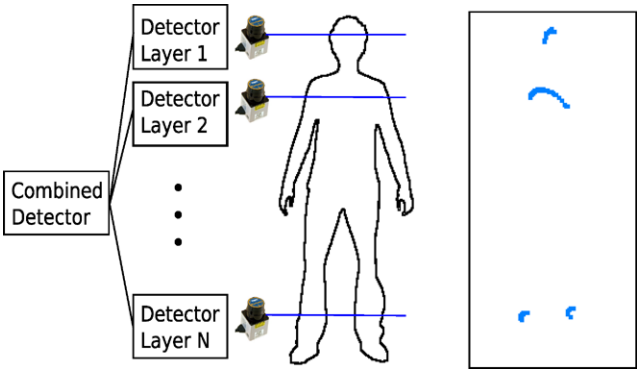

Fig. 1 The *left image* shows the configuration for the complete system with 2D range scans situated at different layers. A classifier is learned for the body part found in each layer. These classifiers are then combined to create a final person detector. The *right image* depicts examples of segments representing body parts at three different heights: legs, upper body, and head (bird's eye view)

The method presented in this paper is based on the classification of segments that represent each body part. For each layer, a classifier is trained using a supervised learning approach based on boosting [2]. The training data for each classifier is composed of the segments that represent the body part of the corresponding layer. In the classification step, each new segment accumulates evidence for its final classification using a probabilistic voting approach [8]. In this technique, the final classification for a specific segment takes into account the classification of all segments in the scene. Moreover, the approach presented in this paper is able to detect a person in the scene using only a single observation. We do not integrate multiple observations over time.

Experimental results shown in this paper illustrate that the resulting classification system can detect persons in cluttered environment with high recognition rates. Moreover, we present results illustrating that the multi-layer classifier improves the detection over single-layer ones. Finally, we show the robustness of the classifier under occlusions.

The rest of the paper is organized as follows. After discussing the related work in Sect. 2, we introduce the single layer people detection in Sect. 3. The multi-layer approach is explained in Sect. 4. Experimental results are shown in Sect. 5. We present a discussion of about our approach in Sect. 6. Finally, we conclude in Sect. 7.

2 Related Work

In the past, several researchers focused on the problem of detecting/tracking people in range scans. One of the most popular approaches in this context is to extract legs by detecting moving blobs that appear as local minima in the range image. For example, Fod et al. [5] use a combination of background and foreground models to extract and track blobs in an indoor environment using multiple laser range finders. Kleinhagenbrock et al. [7] apply an anchoring processes to link legs of persons with the rest of the body. In this case the legs are detected applying a set of thresholds on the segments obtained from a laser scan. The work by Cui et al. [4] also uses background subtraction to detect the legs of persons in open areas. The leg detection system developed by Scheutz et al. [13] searches for legs of an appropriate width, and for a possible gap between legs scaled by the distance. Schulz et al. [14] detect people in laser scans as local minima in the distance histograms. The set of possible leg patterns from scan data is extended by Topp and Christensen [16]. Finally, Xavier et al. [20] detect legs as arcs containing some geometrical restrictions.

In all these previous works the selection of features used to detect the legs is done by hand. In comparison, our work automatically selects the best features for the different body parts detections, and creates a classifier with them.

The multi-part detection of people has been mainly studied in vision. For example, Leibe et al. [8] use a voting approach to detect people in images with a previous learned codebook. Moreover, the works by Ioffe and Forsyth [6] and Ronfard et al. [12] incrementally assemble body parts detected in a picture. Also Mikolajczyk et al. [10] use a probabilistic assembly of different body part detectors. Wu and Nevatia [19] apply a Bayesian combination of body parts which are detected using edgelet features. Other works combine different sensors to detect people. Spinello et al. [15] use laser and vision sensors to detect people from a car. Also Zivkovic and Kröse [22] combine panoramic images with laser scans. In contrast to these works we only use laser range finders to detect people.

Furthermore, AdaBoost has been successfully used as a boosting algorithm in different applications for object recognition. Viola and Jones [18] boost simple features based on grey level differences to create a fast face classifier using images. Treptow et al. [17] use the AdaBoost algorithm to track a ball without color information in the context of RoboCup. Further, Mozos et al. [11] apply AdaBoost to create a classifier able to recognize places in 2D maps.

The approach presented in this paper is based on the work by Arras et al. [2], which uses boosting to learn a classifier for the detection of leg segments. However, we additionally learn classifiers for other body parts, and we introduce a method to combine the resulting classifications.

3 Single Layer Classification

Our system is composed of several laser range finders located at different heights as shown in Fig. 1. Each laser scan is segmented and a set of geometrical features is extracted for each of the obtained segments. Using this data we learn

a classifier to detect the corresponding body part of a person. Each classifier is trained using a supervised approach based on boosting. In the following sections we explain how to create a classifier for a single layer.

3.1 Boosting

To create the individual classifier $C(x)$ for each layer we follow the approach introduced in [2]. This method uses the supervised generalized AdaBoost algorithm to create a final strong classifier by combining several weak classifiers. The requirement to each weak classifier is that its accuracy is better than a random guessing. The input to the algorithm is a set of training examples (x_n, y_n), $n = 1, \ldots, N$, where each x_n is an example and $y_n \in \{+1, -1\}$ is a label indicating whether x_n is a positive example ($y_n = +1$) or a negative one ($y_n = -1$). The examples are initially weighted according to a distribution D. In a series of rounds $t = 1, \ldots, T$, the AdaBoost algorithm selects the weak classifiers that have a small classification error in the weighted training examples. The weight distribution D_t is changed on each iteration to give more importance to the most difficult examples. The final strong classifier is composed of a weighted majority sum of the selected weak classifiers.

Each weak classifier h_j is based on a single-valued feature f_j and has the form

$$h_j(x) = \begin{cases} +1 & \text{if } p_j f_j(x) < p_j \theta_j, \\ -1 & \text{otherwise,} \end{cases} \tag{1}$$

where θ_j is a threshold, and p_j is either $+1$ or -1 and thus represents the direction of the inequality. In each round t of the algorithm, the values for θ_j and p_j are learned so that the misclassification in the training data is minimized. The final strong classifier is a weighted combination of the best T weak classifiers. The output of the final binary classifier $C(x)$ has two possible values $\{+1, -1\}$ representing a positive and negative classification respectively. The final AdaBoost algorithm modified for the concrete task of this work is shown in Fig. 2.

3.2 Geometrical Features

Each layer in our system is equipped with a range sensor that delivers scan observations. The observation z from one laser sensor is composed of a set of beams $z = \{b_1, \ldots, b_L\}$. Each beam b_j corresponds to a tuple (ϕ_j, ρ_j), where ϕ_j is the angle of the beam relative to the sensor and ρ_j is the length of the beam. Following the approach in [2], each observation z is split into an ordered partition of segments $z = \{s_1, \ldots, s_M\}$ using a jumping distance condition. The elements of each segment $s_m = \{p_1, \ldots, p_n\}$ are points represented by Cartesian coordinates $p = (x, y)$, where $x = \rho \cos(\phi)$ and $y = \rho \sin(\phi)$, and (ϕ, ρ) are the polar coordinates of the corresponding beam.

- Input:
  - Set of N labeled examples $(x_1, y_1), \ldots, (x_N, y_N)$ with $y_n = +1$ if the example x_n is positive, and $y_n = -1$ if the example x_n is negative
  - Integer T specifying the number of iterations
- Initialize weights $D_1(n) = \frac{1}{2l}$ for positive examples, and $D_1(n) = \frac{1}{2m}$ for negative examples, where l is the number of positive examples and m the number of negative ones.
- For $t = 1, \ldots, T$
  1. Normalize the weights $D_t(n)$
     $$D_t(n) = \frac{D_t(n)}{\sum_{i=1}^{N} D_t(i)}.$$
  2. For each feature f_j train a weak classifier h_j using the distribution D_t.
  3. For each classifier h_j calculate
     $$r_j = \sum_{n=1}^{N} D_t(n) y_n h_j(x_n),$$
     where $h_j(x_n) \in \{+1, -1\}$.
  4. Choose the classifier h_j that maximizes $|r_j|$ and set $(h_t, r_t) = (h_j, r_j)$.
  5. Update the weights
     $$D_{t+1}(n) = D_t(n) \exp(-\alpha_t y_n h_t(x_n)),$$
     where $\alpha_t = \frac{1}{2} \log(\frac{1+r_t}{1-r_t})$.
- The final strong classifier $C(x)$ is given by
  $$C(x) = \text{sign}\,(F(x)),$$
  where
  $$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x).$$

Fig. 2 The generalized version of the AdaBoost algorithm for people detection

The set of training examples for the AdaBoost algorithm is then composed of the segments together with their label, and their pre-calculated single-valued features

$$X = \left\{ (s_i, y_i, f_i) \mid y_i \in \{+1, -1\}, f_i \in \mathbb{R}^d \right\}. \quad (2)$$

Here $y_i = +1$ indicates that the segment s_i is a positive example and $y_i = -1$ indicates that the segment s_i is a negative example. The set of positives examples is composed of segments that correspond to body parts of persons. The negatives examples are represented by segments that correspond to other objects in the environment. The vector f_i contains the set of geometrical features extracted from the segments. Each feature is represented by a single real value. In our case we calculate $d = 11$ features selected from the list given in [2]:

1. Number of points in the segment.
2. Standard deviation.
3. Mean average deviation from median.
4. Euclidean distance between the first and last point of a segment.
5. Linearity of the segment.
6. Circularity of the segment.
7. Radius of the circle fit in the segment.
8. Boundary length of the segment.
9. Boundary regularity of the segment.
10. Mean curvature of the segment.
11. Mean angular difference.

4 Multi-Layer Detection

After training the individual classifiers for each body part, our system is able to detect the segments corresponding to a person in each layer. In this section we explain how to combined the output of the different classifiers to obtain a more robust final people detector. We first describe the *shape model* which contains the distance relations between the different body parts. Then we describe the voting approach used to combine the outputs of the classifiers. Finally, we explain how to select the segments corresponding to a person.

4.1 Shape Model

Based on [8], we learn a shape model of a person that specifies the distance relations among the different body parts. Figure 3 shows an example of a shape model for the segments corresponding to the three layers shown in the right image of Fig. 1.

To calculate the distance relations in our shape model, we first project all the segments corresponding to a detected person into the 2D horizontal plane (bird's eye view) as shown in Fig. 3. We then calculate the maximum Euclidean distance from the segments corresponding to a concrete body part to the segments corresponding to the other body parts. In this way, we obtain the distance relations between the different layers. Formally

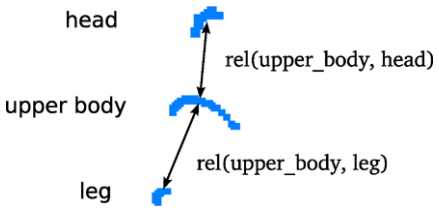

Fig. 3 This figure illustrates two examples of distance relations. In particular, the relations between an upper body segment with respect a head segment, and with respect a leg segment. Segments were projected to the 2D horizontal plane. The distance between the segments has been increased by hand for a better visualization

$$\text{rel}(L_i, L_j) = \max_{\forall s \in S^+} \text{dist}(s_i^+, s_j^+) \mid s_i^+ \in L_i,\ s_j^+ \in L_j, \quad (3)$$

where S^+ indicates the set of segments corresponding to a person. L_i indicates the layer corresponding to body part i (for example the head), and s_i^+ indicates a positive segment of that body part. Finally, dist(,) is a function which calculates the Euclidean distance between the centers of two segments.

The relations between the different layers are learned from a set of positive training examples. A positive training example is composed of the segments obtained when a person is situated in front of the lasers as shown in Fig. 1. The process for obtaining positive examples is explained in more detail in Sect. 5.

Finally, for each relation $\text{rel}(L_i, L_j)$ we create a test function $\delta : S \times S \to \{0, 1\}$ as

$$\delta(s_i, s_j) = \begin{cases} 1 & \text{if dist}(s_i, s_j) \leq \text{rel}(L_i, L_j), \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

This test function indicates whether two new segments s_i and s_j satisfy the distance relation between their corresponding layers.

4.2 Probabilistic Voting

In the detection step, each range sensor delivers an observation z_k which corresponds to the scan taken at layer L_k. This layer may correspond to the legs, upper body, head, or other body part (cf. Fig. 1). After segmenting the observation following the approach of Sect. 3.2, each segment accumulates evidence of being a positive example of the body part corresponding to the layer in which it is located as follows.

Let S be the set of all segments found in the scene. Let s_i be a segment in the scene, and let $c_i \in \{+1, -1\}$ be the

classification of segment s_i. Following a similar approach to [8], we calculate the score $V(c_i^+)$ for a positive classification $c_i = +1$ of segment s_i by marginalizing over all segments s_j found in the scene

$$V(c_i^+) = \sum_{s_j \in S} P(c_i^+, s_j) = \sum_{s_j \in S} P(c_i^+|s_j)P(s_j). \tag{5}$$

Here c_i^+ is equivalent to $c_i = +1$. The last term in (5) represents the probability of a positive classification c_i^+ for segment s_i given all segments found in the scene. We further marginalize over the classification of all segments

$$P(c_i^+|s_j) = \sum_{c_j \in \{+1,-1\}} P(c_i^+, c_j|s_j) \tag{6}$$

$$= \sum_{c_j \in \{+1,-1\}} P(c_i^+|c_j, s_j)P(c_j|s_j). \tag{7}$$

In our system, there are two possible values for a segment classification $c_j \in \{+1, -1\}$. These values indicate whether the segment s_j corresponds to a person $c_j = +1$ or not $c_j = -1$. Instantiating the variable c_j in (7) we obtain

$$\sum_{c_j \in \{+1,-1\}} P(c_i^+|c_j, s_j)P(c_j|s_j)$$

$$= P(c_i^+|c_j^+, s_j)P(c_j^+|s_j) + P(c_i^+|c_j^-, s_j)P(c_j^-|s_j). \tag{8}$$

Here c_j^- is equivalent to $c_j = -1$. Substituting in (5), we get the final expression for the score of a positive classification as

$$V(c_i^+) = \sum_{s_j \in S} (P(c_i^+|c_j^+, s_j)P(c_j^+|s_j) + P(c_i^+|c_j^-, s_j)P(c_j^-|s_j))P(s_j). \tag{9}$$

It remains to explain how to calculate each term in (9). The term $P(c_j^+|s_j)$ indicates the probability of a positive classification of segment s_j. This value is obtained directly from the output of the classifier $C_k(x)$ at the layer L_k where s_j was found

$$P(c_j^+|s_j) = \begin{cases} 1 & \text{if } C_k(s_j) = +1 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Using this result, the probability for a negative classification $P(c_j^-|s_j)$ can be obtained as

$$P(c_j^-|s_j) = 1 - P(c_j^+|s_j). \tag{11}$$

The term $P(c_i^+|c_j^+, s_j)$ indicates the probability of a positive classification for segment s_i given there is another segment s_j in the scene which corresponds to a person too, i.e., $c_j = +1$. This value is obtained using the test function of the shape model introduced in Sect. 4.1 as

$$P(c_i^+|c_j^+, s_j) = \delta(s_i, s_j). \tag{12}$$

The expression $P(c_i^+|c_j^-, s_j)$, indicates the probability for a positive classification of segment s_i given there is another segment in the scene which corresponds to an entity which is not a person. In this work, we introduce the following model

$$P(c_i^+|c_j^-, s_j) = \begin{cases} 0 & \text{if } \delta(s_i, s_j) = 1 \\ \theta & \text{otherwise.} \end{cases} \tag{13}$$

This expression indicates that whenever we find a segment in the scene corresponding to an object other than a person, this object cannot fulfill the shape model of a person. This is indicated by the first condition in (13). However, if we find a segment not corresponding to a person but outside the shape model, this segment will contribute with a small vote. This vote is represented by the parameter θ. In our experiments we found that a value of $\theta = 0.05$ gives good results in the detection process.

Finally, we need to obtain a value for the term $P(s_j)$. In our case we used a uniform distribution over the segments in the scene as

$$P(s_j) = \frac{1}{|S|}, \tag{14}$$

where $|S|$ indicates the total number of segments found in the scene.

4.3 Person Detection

After accumulating evidences for all segments found in all layers, we have a distribution of probabilistic votes among the different hypotheses c_i^+. To detect a person in the environment, we look for the hypothesis c_p^+ which maximum positive score

$$c_p^+ = \underset{c_i^+}{\text{argmax}}\, V(c_i^+). \tag{15}$$

The segment s_p^+ corresponding to c_p^+ is then selected as the representative for the person in the scene. Having the shape model, one can additionally obtain the rest of the segments corresponding to the selected person. These segments will correspond to positive segments s_j^+ that fulfill the shape model and thus gave a positive vote to the representative segment s_p. In our experiments we try to detect one person only, and for this reason we apply (15) for selecting the final hypothesis that best represents the person.

Furthermore, to detect several persons one can look for different local maxima in the hypotheses space [8]. The

computational cost of searching for several persons is independent of the proposed voting approach, since the voting is applied simultaneously to all segments in the scene. As a result, each segment will contain an accumulated value indicating its possibility of being part of a person.

More complex data association problems, like distinguishing between persons standing close together, can be solved integrating multiple observations over time. However, data association using multiple observations is outside the scope of this paper.

5 Experiments

The approach presented above was implemented using a system composed of three layers. At each layer, we located an URG-04LX laser range finder with a field of view of 240 degree. The resolution of the lasers was of 0.36 degree. Each laser was situated at a different height and its goal was to detect a different body part. The upper laser was located 160 cm above the floor, and its function was to detect heads. The middle one was located 140 cm above the floor. This laser detected upper bodies. The final one was located 30 cm above the floor, and its task was to detect legs. The complete system is shown in Fig. 4. The experiments were carried out in the Laboratory for Intelligent Robots and Vision Systems at the University of Kyushu in Japan. Although we use three layers in our experiments, the reader should note that the approach presented in this paper can be applied to any number of layers. During the experiments the sensors were kept stationary. In addition, the lasers did not need to be calibrated, as the possible errors in the alignments are included in the relations of the shape model. These relations were automatically learned during the training process.

Although, during our experiments the system is used in a static manner, the lasers can be incorporated to a mobile platform, since the approach does not apply any background subtraction. Moreover, we do not accumulate information over time, because we classify a single multi-layer observation at each point in time.

We first explain how to obtain a training set for the learning step. We then demonstrate how a multi-layer classifier can be learned in an indoor environment to detect people. In addition we show the robustness of this classifier under occlusions and in cluttered environments. Finally, we show the improvements of the detection rates when using our multi-layer detector in comparison to a single-layer system.

One important parameter of the AdaBoost algorithm is the number of weak classifiers T used to form each final strong classifier. We performed several experiments with different values for T and we found that $T = 200$ weak classifiers provide the best trade-off between the error rate of the classifier and the computational cost of the algorithm. Another parameter that has to be set is θ. In our experiments we found that a value of 0.05 gives good results in the different detection situations. Finally, we selected a jump distance of 15 cm for segmenting the scans.

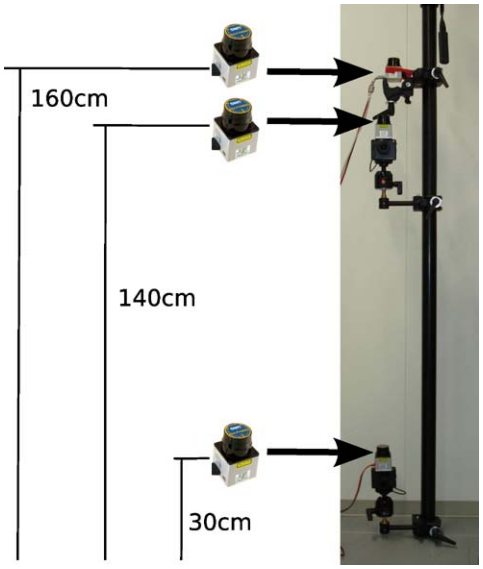

Fig. 4 The image shows the 3-layer system used in the experiments. Each laser is located at a different height to detect a different body part: head (160 cm), upper body (140 cm), legs (30 cm)

5.1 Training Data

The first step in the experiments was to train the classifiers for each layer. As explained in Sect. 3, we used the supervised algorithm AdaBoost to create each classifier. The input to the algorithm is composed of positive and negative examples. The set of positive examples contains segments corresponding to the different body parts: legs, upper body, and head. The set of negative examples is composed of segments corresponding to other objects in the environment such as tables, chairs, walls, etc. We used the same training algorithm for the three layers, with the only difference being the training data used as input.

To obtain the positive and negative examples we left a free space of 5 m $\times$ 1.5 m in front of the lasers. This space did not contain furniture or other objects. We then started recording laser scans while a person was walking randomly inside the rectangle. The obtained scans were segmented following the approach in Sect. 3.2. The segments were then automatically labeled as positive examples of a body part if they were inside the rectangle, and as negative examples if they fell outside the rectangle. This process is shown in Fig. 5. This is a straightforward method to obtain training data without the need of hand-labeling.

5.2 Multi-Layer Classification

In the following experiments we tested our multi-layer approach in an indoor environment. We first obtained the train-

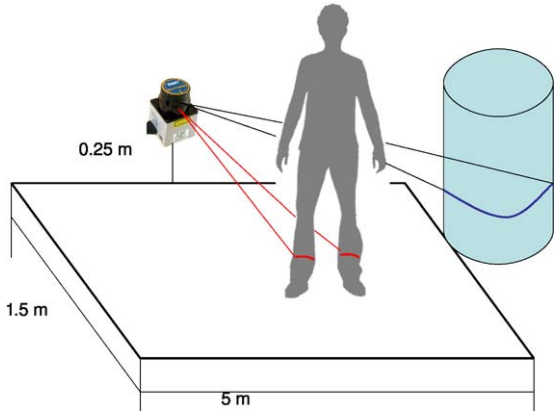

Fig. 5 The image depicts the process for obtaining positive training data. A free space (5 m × 1.5 m) is left empty in front of the lasers. A person walks inside this space and the corresponding segments are automatically labeled as positive examples. The segments falling outside the rectangle are automatically labeled as negative examples

Fig. 6 First scenario for the experiments. The pictures were taken from the position were the sensors were located. The blue trash bins in the *right image* (marked with a *white circle*) were used in the occlusion experiments

Fig. 7 (Color online) The images show examples of scans taken at the different layers. The *left image* corresponds to the lower layer (legs), the *middle image* to the middle layer (upper body), and the *right image* to the top layer (head). *Blue points* indicate segments classified as positive (body parts). *Black points* correspond to segments classified as negative (non-body parts)

ing data following the procedure explained above. The data was obtained in a location of the laboratory shown in the images of Fig. 6. The training data was composed of 344 multi-layer observations containing 17286 segments. Examples of training scans are shown in the images of Fig. 7. This figure depicts scans taken at the different layers when a person was situated in front of the lasers.

In a first experiment, the same person walked in front of the lasers following different trajectories from the training data. In this way we obtained a different test set. We then applied our multi-layer detector to this test. An example of observation with its corresponding detection is shown in Fig. 8. The results of the detections are shown in Table 1, in the row

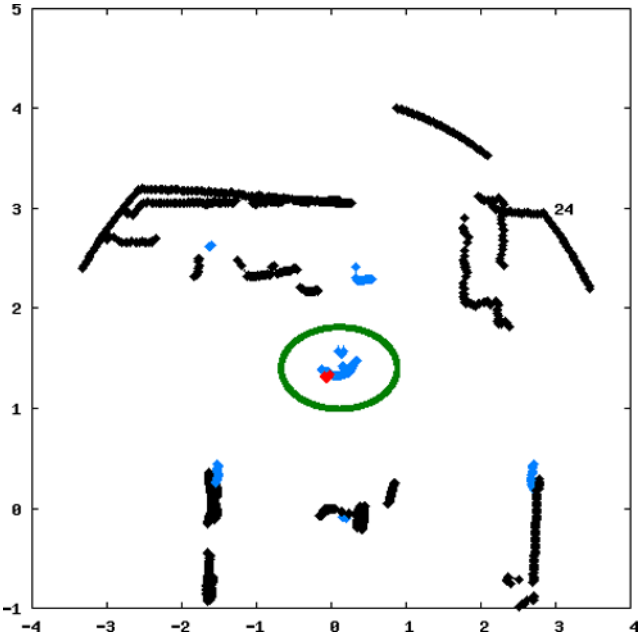

Fig. 8 (Color online) The image shows an example of a detection for the experiment in the first scenario. Different *colors* indicate different classifications. *Blue segments* are classified as body parts, the *red segment* is the one with best evidence of been a person. *Black segments* are classified as other objects. The segments corresponding to the person (ground truth) are marked with a *green ellipse*. The lasers are located at (0, 0)

Table 1 Multi-layer detection rates

	True detection	False detection	Total observations
Scenario 1	**92.0%** (149)	8.0% (13)	162
Scenario 2	**85.8%** (272)	14.2% (45)	317
Scenario 3	**75.2%** (161)	24.8% (53)	214

corresponding to the scenario 1. The detection rate of 92% indicates that we can use our method to detect people with high accuracy in indoor environments.

In a second experiment we tested the performance of our method with partially occluded bodies. In this experiment, a person walked in front of the lasers and, at same point in time, he took two trash bins and put them in front of the lasers. The bins are shown in the top right image of Fig. 6. Following, the person walked around them, and finally put the bins back in their initial position. In this situation several occlusion situations appear. First, while the person was walking around the bins his legs remained occluded. Second, while the person was bending down to take/leave the bins his upper body and his head disappeared.

We applied our detector to this sequence of observations and obtained the results shown in the row corresponding to the second scenario in Table 1. The false positives often occurred when the person was in contact with the bins: taking them, moving them or leaving them. In these situations it

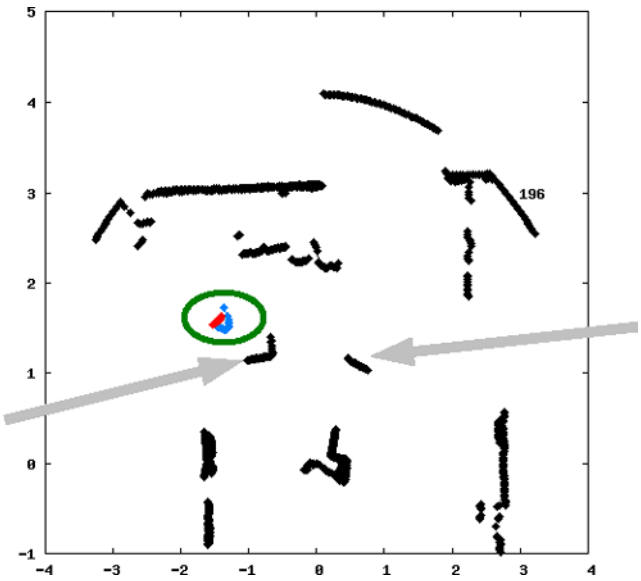

Fig. 9 (Color online) The image shows an example of a detection for the experiment in the second scenario. The meaning of the colors are the same as in Fig. 8. The position of the bins are pointed with *light grey arrows*. The person is behind one of the bins with his legs occluded. The lasers are located at (0, 0)

Fig. 10 Third scenario for the experiments. The pictures were taken from the position were the sensors were located. As we can see the place is very cluttered

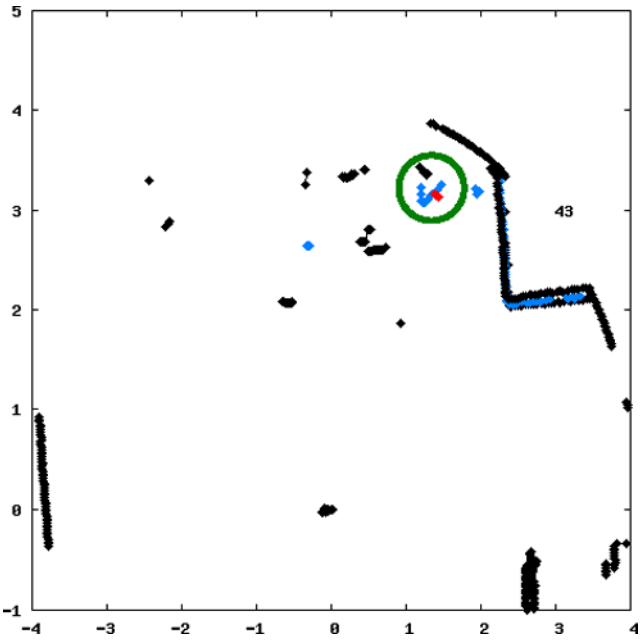

Fig. 11 The image shows an example of a detection for the third experiment. The meaning of the colors are the same as in Fig. 8. The lasers are located at (0, 0)

Table 2 Single-layer detection rates

	True detection	False detection	Total observations
Scenario 1	**92.6%** (150)	7.4% (12)	162
Scenario 2	**73.2%** (232)	26.8% (85)	317
Scenario 3	**41.1%** (88)	58.9% (126)	214

was difficult to detect all body parts. However, a detection rate of 85.8% indicates that we still can use our approach to detect partially occluded persons. An example observation taken while the person was behind a bin is shown in Fig. 9.

In a third experiment, we tested the performance of our learned multi-layer detector in a new cluttered environment. Figure 10 shows images of this third scenario. In this experiment, a person walked around and the obtained observations where classified. Results of the detections are shown in the row corresponding to the third scenario in Table 1. The detection rate decreased to 75.2, however we think this is still a good result for such a challenging scenario. Figure 11 shows a snapshot of this experiment. Videos for the three experiments are available in [1].

5.3 Comparison with Single-Layer Detection

In these experiments we analyze the improvement of our multi-layer system in comparison to a single-layer detector. To do this, we apply our probabilistic model (cf. Sect. 4.2) to the layer corresponding to the legs. In this layer we apply the shape model as well. In this case the shape model is only composed of the relations between the legs. The selection of the best segment representing a person is also done following the approach of Sect. 4.3.

We repeat the detection in the three scenarios from the previous section. Results are shown in Table 2. In the first scenario the results are quite similar, since there are no occlusions and the legs are correctly detected. We can see the improvement of our method in the experiments in the second scenario, in which the multi-layer obtains a detection rate of 85.8% in comparison to 73.2% obtained with the single-layer. Finally, in the third scenario the single-layer obtained a detection rate of 41.1%, while our multi-layer approach got a rate of 75.2%.

5.4 Individual Classification Rates

In this last experiment we show the classification rates for each individual layer. In this experiment we used the test set in the first scenario and analyzed the performance of each layer when classifying individual segments corresponding to

Table 3 Confusion matrices for single layers

	True label	Classification	
		Person	Not person
Legs	Person	**94.3%**	5.7%
	No person	7.8%	**92.2%**
Upper body	Person	**84.4%**	15.6%
	No person	11.2 %	**88.8%**
Head	Person	**86.2%**	13.8%
	No person	12.5%	**87.5%**

people. In this experiments we did apply neither the shape model nor the voting approach. Results are summarized in Table 3. We can appreciate that the classification rate for the legs 94.3% is higher than the classification for the other levels. One reason for this is that the person has two legs, and thus we obtain double number of positive training examples. In the upper levels (upper body and head) the classifications decrease to 84%–86%, however they maintain at acceptable levels.

6 Discussion

The approach presented in this paper described a multi-part person detection based on multiple 2D range scans. A first issue regarding the detection of people using range scans is the position of the lasers. For the experiments presented in this paper we decided to situate lasers at three different fixed heights. These heights were selected to cover a range of 155–175 cm. However this work did not take into account other heights (for example children). To solve this problem we propose to increment the number of lasers or to use one laser mounted in a pan tilt unit. Each layer in the system could contain several classifiers in parallel, each of which is able to detect a concrete body part. The relation between the parts can be learned from people of different heights walking in front of the robot, as shown in Fig. 5.

Another behavior which is not showed in this paper is the detection of several people. After the voting approach, all segments in the scene contain a probability of being a person. To select several persons one can apply a method to look for several local maxima as introduced in Sect. 4.3.

Finally, the approach presented in this paper can answer two different questions. First, given we know there is a person in the environment, the system provides the segment with the best evidence of being this person. Second, given we do not know if there is someone in the scene, the system can be used to answer whether there is a person or not. Because our approach is thought to work under occlusion situations, the detection of a person will depend on the level of confidence the user wants to used. For example, if any body part is enough to detect a person, then a low confidence can be used. In contrast, if a person is composed of all parts, then a higher level should be used. The different confidence levels can be modify using a threshold to compare with the evidence $V(c^+)$ of the best segment.

7 Conclusion

This paper presented a novel approach for people detection using multiple layers of 2D range scans. Each laser detects a different body part of a person like the legs, the upper body or the head. For each body part, we learned a classifier using boosting. The output of the different classifiers was combined in a probabilistic framework to obtain a more robust final classifier. In practical experiments carried out in different environments we obtained encouraging detection rates even in very cluttered ones. Moreover, the comparison of our multi-layer method with a single-layer procedure clearly demonstrated the improvement obtained when detecting people using different body parts simultaneously. Finally, although we use three layers in our experiments, the approach presented in this paper is easily extensible to any number of layers.

Acknowledgements This work was supported by the Canon Foundation in Europe.

References

1. http://www.informatik.uni-freiburg.de/~omartine/publications/mozos2010ijsr.html
2. Arras KO, Mozos OM, Burgard W (2007) Using boosted features for the detection of people in 2D range data. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), pp 3402–3407
3. Bennewitz M, Burgard W, Thrun S (2002) Learning motion patterns of persons for mobile service robots. In: Proceedings of the IEEE international conference on robotics and automation (ICRA)
4. Cui J, Zha H, Zhao H, Shibasaki R (2005) Tracking multiple people using laser and vision. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS), Alberta, Canada
5. Fod A, Howard A, Mataric MJ (2002) Laser-based people tracking. In: Proceedings of the IEEE international conference on robotics and automation (ICRA)
6. Ioffe S, Forsyth DA (2001) Probabilistic methods for finding people. Int J Comput Vis 43(1):45–68
7. Kleinhagenbrock M, Lang S, Fritsch J, Lömker F, Fink GA, Sagerer G (2002) Person tracking with a mobile robot based on multi-modal anchoring. In: IEEE international workshop on robot and human interactive communication (ROMAN), Berlin, Germany
8. Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Comput Vis 77(1–3):259–289

9. Leibe B, Schindler K, Cornelis N, Van Gool L (2008) Coupled object detection and tracking from static cameras and moving vehicles. IEEE Trans Pattern Anal Mach Intell 30(10):1683–1698
10. Mikolajczyk K, Schmid C, Zisserman A (2004) Human detection based on a probabilistic assembly of robust Part detectors. In: Computer vision, ECCV 2004. Lecture notes in computer science. Springer, Berlin, pp 69–82
11. Mozos OM, Stachniss C, Burgard W (2005) Supervised learning of places from range data using AdaBoost. In Proceedings of the IEEE international conference on robotics and automation (ICRA), pp 1742–1747. Barcelona, Spain, April 2005
12. Ronfard R, Schmid C, Triggs B (2002) Learning to parse pictures of people. In: Proceedings of the European conference of computer vision
13. Scheutz M, McRaven J, Cserey G (2004) Fast, reliable, adaptive, bimodal people tracking for indoor environments. In Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS), Sendai, Japan
14. Schulz D, Burgard W, Fox D, Cremers AB (2003) People tracking with a mobile robot using sample-based joint probabilistic data association filters. Int J Robot Res 22(2):99–116
15. Spinello L, Siegwart R (2008) Human detection using multimodal and multidimensional features. In: Proceedings of the IEEE international conference on robotics and automation (ICRA)
16. Topp EA, Christensen HI (2005) Tracking for following and passing persons. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems (IROS)
17. Treptow A, Zell A (2004) Real-time object tracking for soccer-robots without color information. Robot Auton Syst 48(1):41–48
18. Viola P, Jones MJ (2001) Robust real-time object detection. In: Proceedings of the IEEE workshop on statistical and theories of computer vision
19. Wu B, Nevatia R (2007) Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors. Int J Comput Vis 75(2):247–266
20. Xavier J, Pacheco M, Castro D, Ruano A (2005) Fast line, arc/circle and leg detection from laser scan data in a player driver. In: Proceedings of the IEEE international conference on robotics and automation (ICRA)
21. Zender H, Mozos OM, Jensfelt P, Kruijff G-JM, Burgard W (2008) Conceptual spatial representations for indoor mobile robots. Robot Auton Syst 56(6):493–502
22. Zivkovic Z, Krose B (2007) Part based people detection using 2d range data and images. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 214–219

Oscar Martinez Mozos is a postdoctoral researcher in he Robotics, Perception and Real Time Group at the University of Zaragoza (Spain). In 2008 he received his Ph.D. from the Albert-Ludwigs-University Freiburg (Germany). In 2004 he received a M.Sc. degree in Applied Computer Science from the same university. In 2001 he received a M.Sc. degree in Bioengineering from the Miguel Hernandez University (Spain). In 1997 he completed a M.Eng. in Computer Science at the University of Alicante (Spain). In 2008 he was a fellow of the Canon Foundation in Europe. His areas of interest include machine learning, artificial intelligence, robotics, and neuroscience.

Ryo Kurazume received his M.E. and B.E. degrees from the Department of Mechanical Engineering Science, Tokyo Institute of Technology, in 1991 and 1989, respectively. His Ph.D. degree was from the Department of Mechanical Engineering Science, Tokyo Institute of Technology, in 1998. He is a Professor at the Graduate School of Information Science and Electrical Engineering, Kyushu University. His current research interests include multiple mobile robots, 3-D modeling, manipulator, walking robots and medical image analysis.

Tsutomu Hasegawa received B.E. degree in 1973 in electronic engineering and Ph.D. degree in 1987, both from the Tokyo Institute of Technology, Tokyo, Japan. He was associated with the Electrotechnical Laboratory of the Japanese Government from 1973 to 1992 where he performed research on robotics. From 1981 to 1982, he was a Visiting Researcher at the Laboratoire d'Automatique et d'Analyse des Systemes (LAAS/CNRS), Toulouse, France. He joined Kyushu University, Fukuoka, Japan, in 1992 and is currently a Professor with the Department of Advanced Information Technology, Kyushu University. His research interests are in dexterous robotic manipulation, geometric modeling, motion planning, and sensing. Dr. Hasegawa received the Franklin V. Taylor Memorial Award from IEEE SMC in 1999. He is a member of the Institute of Electrical Engineers of Japan, the Society of Instrumental and Control Engineers in Japan, the Robotics Society of Japan, and the Japanese Society of Mechanical Engineers.