Multi-Layer People Detection using 2D Range Data

Oscar Martinez Mozos Ryo Kurazume Tsutomu Hasegawa

***Abstract*— This paper addresses the problem of detecting people using multiple layers of 2D range scans. Detecting persons is an important capacity for intelligent systems that have to interact with people. Our approach uses a supervised learning algorithm to train one classifier for each layer, which concentrates in a different body part. The classifiers are then combined in a probabilistic way to create a final robust detector. Experimental results with real data demonstrate the effectiveness of our approach to detect persons in cluttered environments, and its ability to deal with occlusions.**

I. Introduction

Detecting people is a key capacity for intelligent systems that have to interact in populated environments such as service robots [3], [23], [18], autonomous vehicles [17], [10], or ambient intelligence and surveillance systems [6], [16]. A robust detection of persons in the environment will improve the ability of these systems to communicate with people and to take decisions accordingly.

In this paper we address the problem of detecting people 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 [5], [8], [14], [4], [15], [18], [3], [2], [17]. Although good classifications rates have been obtained using machine learning techniques [2], [17], there is still the need to improve the robustness of the final detectors. One of the main problems is the little information that range scans provide about legs. An example is shown in the bottom right of Figure 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 due to the different objects in the environment, such as tables, chairs or other furniture. Finally, 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 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 Figure 1. Our method is based on the classification of segments that represent each body part (right image of Figure 1). 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 [9]. In our method, the voting for a specific segment takes into account the classification of all segments in the scene.

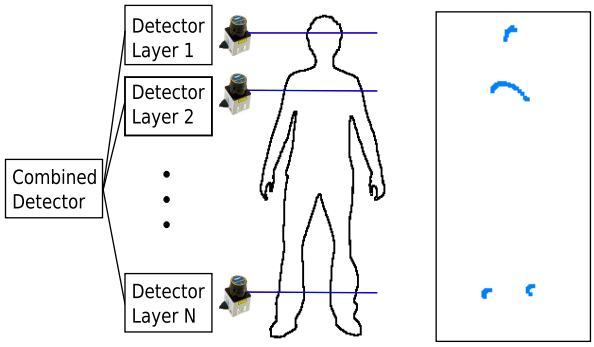

Fig. 1. The left image shows the configuration for the complete multi-layer system with 2D range scans situated at different layers. A classifier is learned for the body part found in each layer. These classifiers are then combined to create a final person detector. The right image depicts examples of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

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.

II. 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 minimum in the range image [5], [8], [4], [14], [15], [18], [4], [22]. Some of these works additionally extract some geometrical or moving features. However, these features are selected by hand. In comparison, our work learns automatically a classifier selecting the best features for the detection. In the

This work was supported by the Canon Foundation in Europe.

Oscar Martinez Mozos is with Dept. of Computer Science and System Engineering, University of Zaragoza, Spain.

Ryo Kurazume and Tsutomu Hasegawa are with Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan.

omozos@gmail.com, kurazume@is.kyushu-u.ac.jp

hasegawa@irvs.is.kyushu-u.ac.jp

work by Arras *et al.* [2], boosting is used to learn a classifier to detect legs segments. In this work we additionally learn classifiers for other body parts, and we introduce a method to combine the classifications.

The multi-part detection of people has been studied mainly in vision. Leibe *et al.* [9] use a voting approach to detect people in images with a previous learned codebook. The works from Ioffe and Forsyth [7] and Ronfard *et al.* [13] incrementally assemble body parts detected in a picture. Also Mikolajczyk *et al.* [11] use a probabilistic assembly of different body part detectors. Wu and Nevatia [21] apply a Bayesian combination of body parts detected using edgelet features. Finally, Zivkovic and Kröse [24] combine different body parts detected using Haar-like features in omnidirectional images.

Other works combine different sensors to detect people. Spinello *et al.* [17] use laser and vision sensors to detect people from a car. Also Zivkovic and Kröse [24] combine panoramic images with laser scans. In contrast to these works we use only laser range finders.

AdaBoost has been successfully used as a Boosting algorithm in different applications for object recognition. Viola and Jones [20] boost simple features based on grey level differences to create a fast face classifier using images. Treptow *et al.* [19] use the AdaBoost algorithm to track a ball without color information in the context of RoboCup. Further, Mozos *et al.* [12] apply AdaBoost to create a classifier able to recognize places in 2D maps. Our application of boosting is similar to [2], although we extended it to other body parts.

III. Single Layer Classification

This section describes the individual classifiers used in each layer. Each classifier is trained to detect a different body part of a person like the legs, the upper body or the head.

A. Boosting

To create the individual classifier $\mathcal{C}_n$ for layer n we follow the approach introduced in [2]. This method uses the supervised 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. 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. Each weak classifier h_j is based on a single-valued feature f_j and has the form

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

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 $\mathcal{C}_n$ has two values $\{+1, -1\}$ representing the positive and negative classification respectively. More details about this approach are given in [2].

B. Geometrical Features

In this section we describe the segmentation method and the features used in our system. Our system is equipped with several range sensors that deliver observations. The observation z from one laser sensor is composed of a set of beams $z = \{b_1, ..., b_L\}$. 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 $\mathcal{S} = \{s_1, s_2, ..., s_M\}$ using a jumping distance condition. The elements of each segment $s = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ are represented by Cartesian coordinates $\mathbf{x} = (x, y)$, where $x = \rho\cos(\phi)$ and $y = \rho\sin(\phi)$, and (ϕ, ρ) are the polar coordinates of the corresponding beam.

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

$$X = \left\{(s_i, y_i, f_i) \mid l_i \in \{+1, -1\}, f_i \in \Re^d\right\},$$

where $y_i = +1$ 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 dimension d of the feature vector f_i depends on the number of single features extracted from each segment. In our case we calculate eleven features selected from the list given in [2]: number of points, standard deviation, mean average deviation from median, width, linearity, circularity, radius, boundary length, boundary regularity, mean curvature, and mean angular difference.

IV. Multi-Layer Detection

After training the individual classifiers for each body part, our system is able to detect in each layer the segments corresponding to a person. In this section we explain how to combined the output of the different classifiers to obtain a more robust final people detector.

A. Shape Model

Based on [9], we learn a shape model of persons that specifies the geometrical relations among the different body parts. Figure 2 shows an example of a shape model for the segments corresponding to the three layers shown in the right image of Figure 1. To calculate the geometrical relations in our shape model, we first project the segments pertaining to a person into the 2D horizontal plane (bird's eye view). We then calculate the maximum distance of a segment corresponding to a concrete body part with respect to the segments corresponding to the other body parts as

$$\text{rel}(\mathcal{L}_i, \mathcal{L}_j) = \max_{\forall x \in X} \text{dist}(s_i^+, s_j^+) \mid s_i^+ \in \mathcal{L}_i, s_j^+ \in \mathcal{L}_j, \quad (2)$$

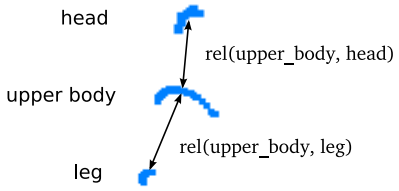

Fig. 2. This figure illustrates two examples of geometrical relations. In particular, the relations between an upper body segment with respect a head segment, and with respect a leg segment. Segments were projected to the 2D horizontal plane. The distance between the segments has been increased by hand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

where $\mathcal{L}_i$ indicates the layer corresponding to body part i (for example the head), and s_i^+ indicates a positive segment of that body part. Finally, $\mathrm{dist}(,)$ is a function which calculates the Euclidean distance between the centers of two segments. These relations are learned from a set of positive training examples. The process for obtaining positive examples is explained in Section V.

Finally, for each relation we create a test function $\delta : S \times S \rightarrow \{0, 1\}$ which indicates whether two new segments s_j and s_j satisfy it

$$\delta(s_i, s_j) = \begin{cases} 1 & \text{if } \mathrm{dist}(s_i, s_j) \leq \mathrm{rel}(\mathcal{L}_i, \mathcal{L}_j) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

B. Probabilistic Voting

In the detection step, each range sensor delivers an observation z_j which corresponds to the scan taken at layer $\mathcal{L}_j$. This layer may correspond to the legs, upper body, head, or other body part (Figure 1). After segmenting the observations (Section III-B), each segment accumulates evidence of being a positive example of the body part corresponding to the layer it was located at.

Let s_i be a segment in the scene, and let l_i be the layer where s_i is located. Now let $c_i \in \{+1, -1\}$ be the classification of segment s_i. Following a similar approach to [9], we calculate the score for a positive classification $c_i = +1$ of segment s_i by marginalizing over all segments found in the scene

$$V(c_i^+) = \sum_j P(c_i^+, s_j) \quad (4)$$

$$= \sum_j P(c_i^+ \mid s_j) P(s_j). \quad (5)$$

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

$$P(c_i^+ \mid s_j) = \sum_{c_j} P(c_i^+, c_j \mid s_j) \quad (6)$$

$$= \sum_{c_j} P(c_i^+ \mid c_j, s_j) P(c_j \mid s_j). \quad (7)$$

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

$$P(c_i^+ \mid s_j) = \quad (8)$$
$$P(c_i^+ \mid c_j^+, s_j) P(c_j^+ \mid s_j) + P(c_i^+ \mid c_j^-, s_j) P(c_j^- \mid s_j).$$

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

$$\sum_j \big(\begin{array}{l} P(c_i^+ \mid c_j^+, s_j) P(c_j^+ \mid s_j) \\ + P(c_i^+ \mid c_j^-, s_j) P(c_j^- \mid s_j) \end{array} \big) \cdot P(s_j). \quad (9)$$

It remains to explain how to calculate each term in (9). The term $P(c_j^+ \mid s_j)$ indicates the probability of a positive classification of segment s_j. This value can be obtained directly from the output of the classifier $\mathcal{C}_{l_j}$ at the layer l_j where s_j was found

$$P(c_j^+ \mid s_j) = \begin{cases} 1 & \text{if } \mathcal{C}_{l_j}(s_j) = +1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Thus, the probability for a negative classification is obtained as

$$P(c_j^- \mid s_j) = 1 - P(c_j^+ \mid s_j). \quad (11)$$

The term $P(c_i^+ \mid c_j^+, s_j)$ 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, i.e., $c_j = +1$. This value is obtained using the test function of the shape model (Section IV-A)

$$P(c_i^+ \mid c_j^+, s_j) = \delta(s_i, s_j). \quad (12)$$

Finally we need to obtain a value for the expression $P(c_i^+ \mid c_j^-, s_j)$, which indicates the probability for a positive classification of segment s_i given there is another segment in the scene which corresponds to other object. We call this expression the *occlusion model*, since it indicates the relation of the people with other objects in the scene. In this work, we apply the following model

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

This expression indicates that whenever we find a segment in the scene corresponding to an object other than a person, this object can not fulfill the shape model of a person.

C. 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

$$c_p^+ = \underset{c_i^+}{\mathrm{argmax}}\, V(c_i^+). \quad (14)$$

The segment s_p corresponding to c_p^+ is then selected as the representative for the person in the scene. To detect several persons one can look for different local maximum in the hypotheses space. In our experiments we try to detect one person only, and for this reason we apply (14) for selecting the final hypothesis that represents the person.

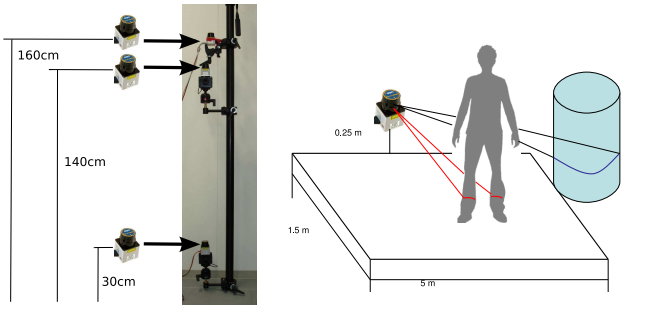

Fig. 3. The left image shows the 3-layer system used in the experiments. Each laser is located at a different height to detect a different body part: head ($160cm$), upper body ($140cm$), legs ($30cm$). The right image depicts the process for obtaining positive training data. A free space ($5m \times 1.5m$) is left 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. 4. First scenario for the experiments. The top pictures were taken from the position were the sensors were located. The blue rubbish in the right image (marked with a white circle) are used for the occlusion experiments. The bottom 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).

V. Experiments

The approach presented above was implemented using a three layer system as shown in Figure 1. At each layer, we located a 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 is situated at a different height and detects a different body part. The upper laser is located $160cm$ above the floor. This laser is thought to detect heads. The middle one is located $140cm$ above the floor. This laser detects upper bodies. The final one is located $30cm$ above the floor, and its task is to detect legs. The complete system is shown in the left image of Figure 3. The experiments were carried out in the Laboratory for Intelligent Robots and Vision Systems at the University of Kyushu in Japan. The sensors were kept stationary during the experiments.

We first explain how to obtain a training set for the learned 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 very 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 for the occlusion model is θ. In our experiments we found that a value of 0.05 gives good results under occlusion situations. Finally, we selected a jump distance of $15cm$ for segmenting the scans.

A. Training Data

The first step in the experiments was to train the classifiers for each layer. As explained in Section III, 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 $5m \times 1.5m$ 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 Section III-B. 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 the right image in Figure 3. This is a straightforward method to obtain training data without the need of hand-labeling.

B. Multi-Layer Classification

In the the following experiments we tested our multi-layer approach in an indoor environment. We first obtained the training data following the procedure explained above. The data was obtained in a location of the laboratory shown in the top images of Figure 4. The training data was composed of 344 multi-layer observations containing 17286 segments. Examples of training scans are shown in the bottom images of Figure 4.

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 Figure 5. The results of the detections are shown in the $Test$ row of Table I. 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,

TABLE I

MULTI-LAYER DETECTION RATES

	True detection	False detection	Total observations
Test	**92.0%** (149)	8.0% (13)	162
Occlusion	**85.8%** (272)	14.2% (45)	317
Hard	**75.2%** (161)	24.8 % (53)	214

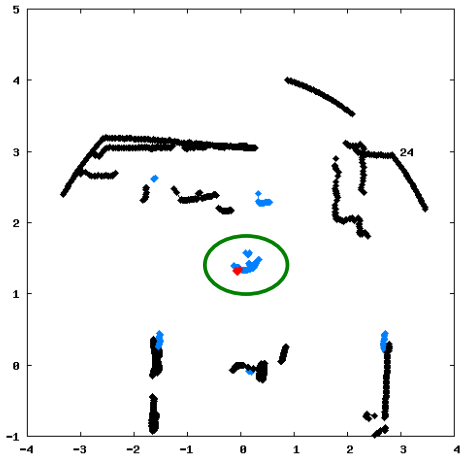

Fig. 5. The image shows an example of a detection for the experiment called $Test$ in Table I. Different colors indicate different classifications. Blue segments are classified as body parts, the red segment is the one with best evidence of been a person. Black segments are classified as other objects. The segments corresponding to the person (ground truth) are marked with a green ellipse. The lasers are located at $(0, 0)$.

a person walked in front of the lasers and, at same point in time, he took two rubbish bins and put them in front of the lasers. The bins are shown in the top right image of Figure 4. Following, the person walked around them, and finally put the bins back in their initial position. In this situation several occlusion problems 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 $Occlusion$ row in table Table I. 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 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 Figure 6.

In a third experiment, we tested the performance of our learned multi-layer detector in a new and very cluttered environment. Figure 7 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 $Hard$ row of Table I. The detection rate decreased to 75.2, however we think this is still a good result for such an extremely challenging scenario. Figure 8 shows a snapshot of this experiment. Videos for the three experiments are available in [1].

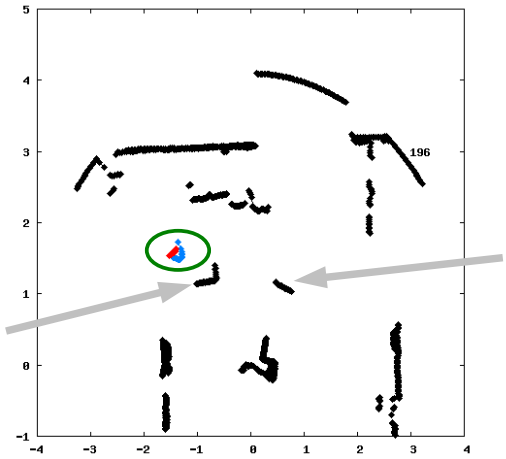

Fig. 6. The image shows an example of a detection for the experiment called $Occlusion$ in Table I. The meaning of the colors are the same as in Figure 5. The position of the bins are pointed with light grey arrows. The person is behind one of the bins with his legs occluded. The lasers are located at $(0, 0)$.

Fig. 7. These images correspond to part of the Laboratory for Intelligent Robots and Vision Systems which is used for experiments. As we can see the location is very cluttered. This scenario is called $Hard$ in Table I.

C. 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 (Section IV-B) in the layer corresponding to the legs. We repeat the detection in the three scenarios from the previous section: $Test$, $Occlusion$, and $Hard$. Results are shown in Table II. For the $Test$ experiment the results are quite similar, since there are no occlusions and the legs are correctly detected. However, we can see the improvement of our method in the experiment $Occlusion$, 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 $Hard$ scenario the single-layer obtained a detection rate of 41.1%, while our multi-layer approach got a rate of 75.2%. This is a very important improvement.

D. Individual Classification Rates

In this last experiment we compare the classification rates for the different layers. In this experiment we used the test set from the $Test$ experiment, and analyzed the performance of each layer when classifying segments. Results are summarized in Table III. 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

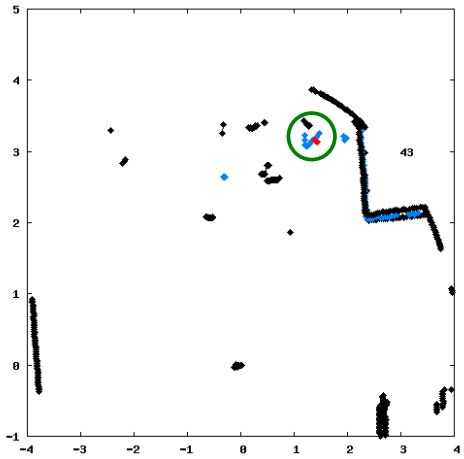

Fig. 8. The image shows an example of a detection for the experiment called $Hard$ in Table I. The meaning of the colors are the same as in Figure 5. The lasers are located at $(0, 0)$.

TABLE II

SINGLE-LAYER DETECTION RATES

	True detection	False detection	Total observations
Test	**92.6%** (150)	7.4% (12)	162
Occlusion	**73.2%** (232)	26.8% (85)	317
Hard	**41.1%** (88)	58.9% (126)	214

classifications decrease to 84%-86%. The classification rates for these body parts are a novelty in this paper.

TABLE III

CONFUSION MATRICES FOR SINGLE LAYERS

		Classification	
	True Label	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% (26)
	No Person	12.5%	**87.5%**

VI. CONCLUSION

This paper presented a novel approach for people detection using multiple layers of 2D range scans. Each laser is responsible for detecting 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. Finally, 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.

REFERENCES

[1] http://www.informatik.uni-freiburg.de/~omartine/publications/mozos2010ijsr.html.

[2] K.O. Arras, O.M. Mozos, and W. Burgard. Using boosted features for the detection of people in 2D range data. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3402–3407, 2007.

[3] M. Bennewitz, W. Burgard, and S. Thrun. Learning motion patterns of persons for mobile service robots. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[4] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Tracking multiple people using laser and vision. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Alberta, Canada, 2005.

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

[6] G.L. Foresti, L. Micheloni, C. Snidaro, and P Remagnino. *Ambient intelligence: a novel paradigm*, chapter Security and building intelligence: from people detection to action analysis, pages 199–212. Springer, 2005.

[7] S. Ioffe and D.A. Forsyth. Probabilistic methods for finding people. *International Journal of Computer Vision*, 43(1):45–68, 2001.

[8] M. Kleinhagenbrock, S. Lang, J. Fritsch, F. Lömker, G.A. Fink, and G. Sagerer. Person tracking with a mobile robot based on multi-modal anchoring. In *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, Berlin, Germany, 2002.

[9] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International journal of computer vision*, 77(1-3):259–289, 2008.

[10] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool. Coupled object detection and tracking from static cameras and moving vehicles. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(10):1683–1698, 2008.

[11] K. Mikolajczyk, C. Schmid, and A. Zisserman. *Computer Vision - ECCV 2004*, chapter Human Detection Based on a Probabilistic Assembly of Robust Part Detectors, pages 69–82. Lecture Notes in Computer Science. Springer-Verlag, 2004.

[12] O.M. Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using AdaBoost. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 1742–1747, Barcelona, Spain, April 2005.

[13] R. Ronfard, C. Schmid, and B. Triggs. Learning to parse pictures of people. In *European Conference of computer Vision*, 2002.

[14] M. Scheutz, J. McRaven, and G. Cserey. Fast, reliable, adaptive, bimodal people tracking for indoor environments. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004.

[15] 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 (IJRR)*, 22(2):99–116, 2003.

[16] L. Snidaro, C. Micheloni, and C. Chiavedale. Video security for ambient intelligence. *IEEE Transactions on Systems, Man and Cybernetics*, 35(1):133–144, Jan. 2005.

[17] L. Spinello and R. Siegwart. Human detection using multimodal and multidimensional features. In *Proc. of The International Conference in Robotics and Automation (ICRA)*, 2008.

[18] E.A. Topp and H.I. Christensen. Tracking for following and passing persons. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2005.

[19] André Treptow and Andreas Zell. Real-time object tracking for soccer-robots without color information. *Robotics and Autonomous Systems*, 48(1):41–48, 2004.

[20] P. Viola and M.J. Jones. Robust real-time object detection. In *Proc. of IEEE Workshop on Statistical and Theories of Computer Vision*, 2001.

[21] Bo Wu and Ram Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *Int. J. Comput. Vision*, 75(2):247–266, 2007.

[22] J. Xavier, M. Pacheco, D. Castro, and A. Ruano. Fast line, arc/circle and leg detection from laser scan data in a player driver. In *Proc. of the IEEE Int. Conference on Robotics & Automation (ICRA'05)*, 2005.

[23] H. Zender, O.M. Mozos, P. Jensfelt, G-J.M. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502, June 2008.

[24] Z. Zivkovic and B. Krose. Part based people detection using 2d range data and images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 214–219, 2007.