Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh, UK, pp. 767-773, IEEE 2014

People Detection and Distinction of their Walking Aids in 2D Laser Range Data based on Generic Distance-Invariant Features

Christoph Weinrich, Tim Wengefeld, Christof Schroeter and Horst-Michael Gross

***Abstract*— People detection in 2D laser range data is a popular cue for person tracking in mobile robotics. Many approaches are designed to detect pairs of legs. These approaches perform well in many public environments. However, we are working on an assistance robot for stroke patients in a rehabilitation center, where most of the people need walking aids. These tools occlude or touch the legs of the patients. Thereby, approaches based on pure leg detection fail. The essential contribution of this paper are generic distance-invariant range scan features for people detection in 2D laser range data and the distinction of their walking aids. With these features we trained classifiers for detecting people without walking aids (or with crutches), people with walkers, and people in wheelchairs. Using this approach for people detection, we achieve an F_1 score of 0.99 for people with and without walking aids, and 86% of detections are classified correctly regarding their walking aid. For comparison, using state-of-the-art features of Arras et al. on the same data results in an F_1 score of 0.86 and 57% correct discrimination of walking aids. The proposed detection algorithm takes around 2.5% of the resources of a 2.8 GHz CPU core to process 270° laser range data at an update rate of 10 Hz.**

I. INTRODUCTION

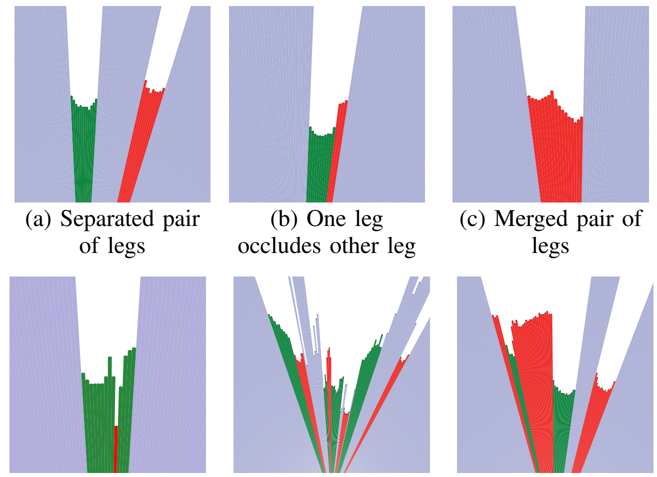
People detection and position tracking are important requirements to improve human-robot interaction (HRI), e.g. for the realization of socially compliant navigation of mobile assistance robots in populated public environments. Furthermore, a distinction of people's walking aids allows specific adaption of the robot behavior. Since many mobile robots are equipped with 2D laser range scanners, this sensor is often used for on-board people detection.

The advantages of laser-based person detection are the sensors' large field of view and the low uncertainties of the hypotheses regarding the distance between person and laser scanner. Still, due to the relatively low amount of data, the computing demand of most laser-based detectors is low as well. This enables high update rates.

However, the information content of laser scans is comparatively low. Most laser scanners perceive just one layer at low altitude above the ground, where objects in the environment are sometimes indistinguishable from persons, resulting in false positive detections. Therefore, people tracking is rarely based solely on laser-based detections. Instead, these detections are often complemented by person hypotheses based on other sensors, like cameras.

This work has received funding from the German Federal Ministry of Education and Research as part of the ROREAS project under grant agreement no. 16SV6133 and from the Federal State of Thuringia and the European Social Fund (OP 2007-2013) under grant agreement N501/2009 to the project SERROGA (project no. 2011FGR0107).

C. Weinrich, T. Wengefeld, C. Schroeter, and H.-M. Gross are with Neuroinformatics and Cognitive Robotics Lab, Ilmenau University of Technology, 98694 Ilmenau, Germany `christoph.weinrich at tu-ilmenau.de`



(a) Separated pair of legs (b) One leg occludes other leg (c) Merged pair of legs

(d) Person with crutch (e) Wheelchair close to the sensor (f) Person with walker

Fig. 1: Range scan details of persons (with walking aids). The individual scan segments, which are reflected by the persons (or aids) are highlighted alternating in green and red.

Due to the geometrical position of the scanning plane, most detectors are actually leg detectors. However, the operational area of our robot is a rehabilitation center for stroke patients [1]. Many patients need aids for locomotion, like walkers, wheelchairs, or crutches. These tools occlude or touch the legs very often. Therefore, we need a detector which also detects legs in combination with these walking aids. Hence, there are particularly consequences for the features. For instance, features which describe the parameters of circular segments [2] are no longer sufficient. Instead, the feature vectors have to be able to describe more complex object shapes. The features proposed in this paper are not designed for the detection of object-specific shapes. Instead, the features are tailored to the characteristics of laser scans. Therefore, we defined two requirements for the features:

1) invariant to the distance between laser scanner and perceived object
2) unspecific to the objects to be classified, by maintaining as much information of the laser scan as possible

In the following, it is described, why these requirements conflict, and how this conflict is handled by the proposed features. For most sensors, the resolution of a certain object's perception reduces with the object's distance to the sensor. Many detection approaches utilize down-scaling of the sensor data to enable distance-invariant feature extraction. The goal

Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh, UK, pp. 767-773, IEEE 2014

is to obtain the same feature vector when a certain view of an object is perceived, independent of the distance between sensor and object. For example, many approaches for visual person detection in monocular camera images use Gaussian pyramids to detect potential objects at a-priori unknown distance to the sensor. However, this kind of down-scaling reduces the high-frequency information content within the images.

A great advantage of laser range data is the explicitly given distance of a sub-segment (for segmentation see Sec. III-A). Therewith, the known real-world object size and the measured distance can directly be used to determine the perceived size of a potential object to be detected. In the approach proposed here, this is used to perform down-scaling and feature extraction efficiently in one processing step. Furthermore, during this processing step both low-frequency content and higher-frequency content is extracted by the features. The features are designed such that the lower-frequency features are independent of the object's distance, and additional information is available in the higher-frequency features for closer objects (while these contain no significant information for distant objects).

The missing specificity of the features to a certain object requires a powerful classifier. Furthermore, due to the high dimensionality of the feature space (including possibly irrelevant dimensions), the training of the classifier should employ feature selection techniques to avoid over-specialization to the training data.

The next section reviews state-of-the-art work, which is related to people detection in 2D laser range data. Thereafter, Sec. III presents our approach, whose innovation are the generic distance-invariant features (GDIF). Sec. IV demonstrates the advantages of the GDIF by detailed experiments.

II. RELATED WORK

There are various approaches for person detection in 2D laser range data, which work on multiple stationary laser scanners [3]. However, for our application only approaches based on laser range scanners on mobile robots are relevant. In [2] approaches for leg detection are classified regarding their usage of motion or geometry features. Since approaches based on motion features (like [4]), are not able to detect standing or sitting people, these approaches are not sufficient for our application as patients in rehabilitation centers are slowed down in their movements and pause often. While those patients need to rest, they are even more vulnerable by a mobile assistance robot due to their limited motion abilities. To show compliant behavior towards these patients, the robot has to robustly detect standing people.

These people are detectable by approaches which are based on geometrical features. In [5], a set of thresholds is used to classify sub-segments of range scan data as leg or non-leg. The focus of [5] is on person tracking, wherefore laser-based detection is just one cue. In contrast, [6] is directly focused on leg detection. The range data is segmented based on jump distances (see Sec. III-A), and each segment is classified based on thresholds of geometrical features. However, the features are selected by the developer, and the classification thresholds have to be set manually.

In [2], the set of geometrical features is extended to 14 geometrical features, which are presumably suitable for leg detection. Then, AdaBoost [7] is used for feature selection and classifier training. Since this approach was designed for the detection of legs, these features are relatively specific to legs (circularity, convexity). Therefore, these features are not sufficient for detecting more complex objects, as for example wheelchairs or walkers.

In contrast, the generic distance-invariant features proposed in this paper are not designed for the detection of object-specific shapes. Instead, the features are tailored to the characteristics of laser scans. Furthermore, in contrast to [2], legs are not detected individually. Instead, our classifier is able to detect (partially occluded or merged) pairs of legs. Thus, the grouping of two individually classified legs to one person hypothesis is unnecessary.

Based on the work of Arras et al., different approaches for multiple 2D range scanners at different height [8], [9] or 3D range scanners [10], [11] have been proposed in recent years. However, on our mobile robot only the range data of one height level is perceived. Nevertheless, the approach presented here could be easily extended to multiple layers, as well.

III. PEOPLE DETECTION BASED ON GENERIC DISTANCE-INVARIANT FEATURES

The input for the people detection approach is a laser range scan $R = \{B_1, B_2, ..., B_b\}$, which consists of a set of b beams, where each beam B_i corresponds to a tuple (ϕ_i, δ_i) of the beam's angle ϕ_i and its measured distance δ_i.

Some of the beams B_i are reflected by persons $G = \{P_1, P_2, ..., P_p\}$. Each person P_i corresponds to a triplet (x_i, y_i, a_i) of the person's center position x_i, y_i relative to the laser's coordinate system and the type of his/her walking aid $a_i \in \{legs, walker, wheelchair\}$. Goal of our approach is to detect all people positions G and estimate the correct walking aid.

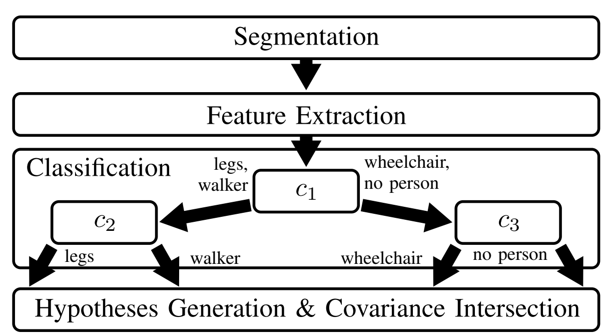


Fig. 2: Processing steps of the proposed approach

A. Segmentation of 2D Range Data

Like in [2], in our approach the beams in the scan R are split into subsets of beams (see Fig. 3). Therefore, the jump distance is applied. The first beam's B_1 index is inserted in a new segment S_1. Iterating over the range scan R from

Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh, UK, pp. 767-773, IEEE 2014

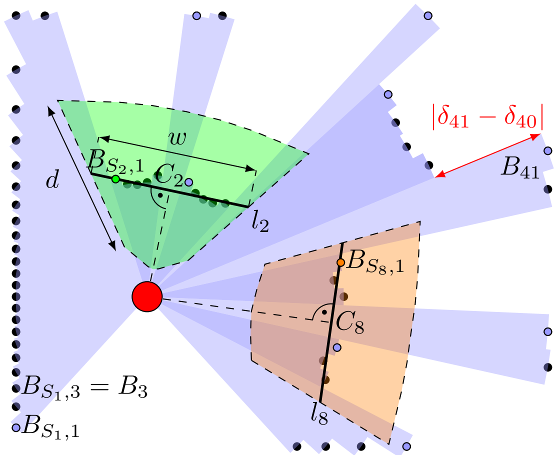
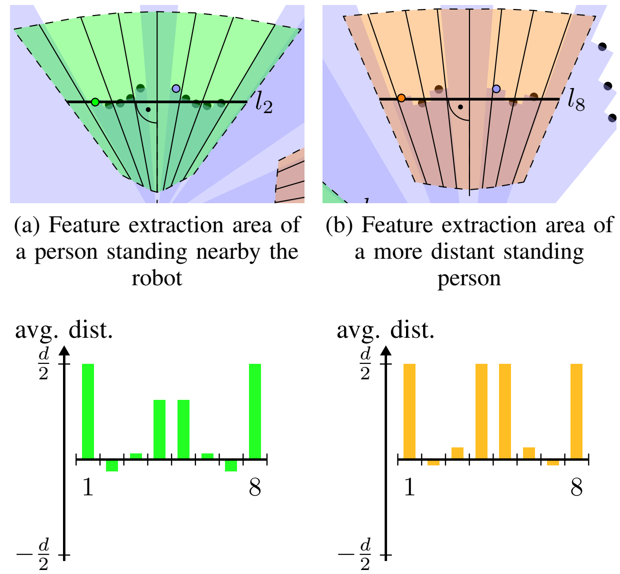


Fig. 3: Schematic illustration showing a range scan of a small room with two persons at different distances from the laser. The jump distance, the decision criterion for defining an new segment, between beam B_{41} and its former beam is exemplarily shown in red. The first beam $B_{S_j,1}$ of each found segment S_j is highlighted as a colored dot. Although each of these beams is used as origin point for the subsequent feature extraction, only these feature extraction areas are shown here in green and orange, whose feature vectors shall be classified as person.



(a) Feature extraction area of a person standing nearby the robot

(b) Feature extraction area of a more distant standing person

(c) Feature vector of a person standing nearby the robot

(d) Feature vector of a more distant standing person

Fig. 4: Visualization of range scan details of Fig. 3, which show the feature extraction areas of a person standing nearby the robot (a) and a more distant person (b). Exemplary, the extracted average distances to the base line for $n = 8$ line segments are visualized as bar histogram (c),(d).

beam B_2 to beam B_b, a new subset is initialized with the beam index i if the difference of the measurements $|\delta_i - \delta_{i-1}|$ of beam B_i to its former beam B_{i-1} is above a certain threshold Δ. Otherwise, the beam index i is added to the current subset. The output of the partitioning procedure is an ordered sequence $P = \{S_1, S_2, ..., S_s\}$ of segments such that $\bigcup_i S_i = \{1, ..., b\}$.

However, in contrast to [2], the feature extraction is not limited to the beams of the individual segments S_i. Instead, the aim is to extract features of the complete object, even if the object is over-segmented into several adjacent segments.

Assuming, that the jump distance between background and a person is above the threshold Δ, the first beam $B_{S_j,1}$ of each segment S_j is used as point of origin for the feature extraction. As shown in the next section, the area considered in the feature extraction is based on these points and the objects' maximal Euclidean width, independent of the segments' size.

In [2] the jump distance threshold Δ influences the size of the range scan details to be classified, possibly leading to over-segmentation. To counter this effect, in [12] Delaunay triangulation is used to generally merge segments whose centers are close. Note, that in the proposed approach only the positions for reference points are determined by Δ, while the range is predefined according to the real-world size of the interesting objects (e.g. persons or aids). Thus, reducing Δ just increases the number of classifications and therewith the calculation effort, without risking over-segmentation of objects.

B. *Feature Extraction*

After segmentation, each remaining origin point $B_{S_j,1}$ is used as starting point for a baseline l_j of fixed width w, which is orthogonal to the line between the baseline's center C_j and the sensor (Fig. 3 and Fig. 4). This baseline l_j is divided into n line segments of equal length. Each line segment covers a certain range of the laser beams. Simple features f_j are extracted from all these beams based on their distances between the beams' actual reflection points and their intersections with the baseline. Note, that these distances are clipped to a fixed range of $[-\frac{d}{2}, \frac{d}{2}]$ before the features are extracted. This clipping is performed to reduce the influence of the distance between the objects to be detected and the background. Therewith, the extraction areas result from the origin points $B_{S_j,1}$, the fixed width w and the fixed depth d. Consequently, the extraction area parameters w and $\frac{d}{2}$ should be above the maximum extension of the objects to detect. So far, as features we use the average distance, the minimum distance, and the maximum distance. Further features, like variance etc., may be supplemented in future. The number of beams per line segment depends on the baseline's distance to the sensor. If there is less than one beam per line segment, the adjacent beams are interpolated. Note, that features like the minimum, maximum, variance etc. do not add any significant information to the average distance, if only one beam is covered by a line segment. These proposed features are characterized by low computing effort.

C. Classification

The dimensionality of the feature space F is determined by the product of line segments n and the number of extracted features. For classification of the feature space F a binary decision tree is used. Each decision is made by discrete AdaBoost classifiers [7], where the weak classifiers are binary decision trees [13]. Fig. 2 shows how each feature vector f_j is classified by application of two of the three binary classifiers. The resulting classification function is $h : F \rightarrow \{\text{legs}, \text{walker}, \text{wheelchair}, \text{no person}\}$.

D. Hypotheses Generation with Covariance Intersection

After a feature vector is classified as one of the person classes, the center of the base line C_j is used as 2D position for the person hypotheses. This results in a set of 2D person hypotheses $G' = \{(C_j, h(f_j))\,|h(f_j) \neq \text{no person}\}$. Additionally, the estimated variance of the pose hypotheses is represented by a covariance matrix. Since the adjacent feature extraction areas usually overlap, one actual person often is detected several times. To prevent generation of multiple hypotheses for the same person, hypotheses are fused using covariance intersection [14] if their Mahalanobis distance is below a threshold. The merged hypothesis' label is determined by majority decision.

Although beyond the scope of this paper, it should be pointed out that these labels are utilized to create hypotheses in 3D space by adding estimates of the heads' vertical positions to the 2D detections C_j. The mean head hight significantly differs between people sitting in wheelchairs and people walking. However, since the head height is not directly observable by the laser scanner, we assign a relatively large variance to the vertical component of the 3D hypotheses. Nevertheless, estimation of the head height supports merging the resulting hypotheses with those from other detectors in a multi-modal person tracker framework presented in [15].

IV. EXPERIMENTS

A. Data Sets

To benchmark the proposed approach, our algorithm and selected reference methods have been evaluated on three data sets:

1) **SPINELLO:** The data set of [12][1] is used to evaluate the proposed approach on a publicly available benchmark data set.
2) **HOME:** We captured the HOME data set within apartments of an assisted living facility of the AWO - Arbeiterwohlfahrt Bundesverband e.V. (German Workers' Welfare Federal Association).
3) **REHA:** We captured the REHA data set in the corridors of a rehabilitation center for stroke patients.

All data has been captured by laser range scanners (SICK S300) with an angular resolution of 0.5°. A substantial difference of SPINELLO to our own data sets[2] is, that the SPINELLO data is recorded by a static laser scanner. Thereby, there is only little variance in the background and the background of training and test data is the same. In contrast, the background of the HOME data set is diversely structured, and different rooms are used for the training and testing data set. Furthermore, the recording of background data in the HOME and REHA environment was paused, when the robot stopped. Thereby, no background view is recorded several times. The challenge of the REHA data set is, that it contains people with walking aids, whose detection was the motivation for this work.

For clarification of the detection task, Fig. 1 shows six range scan details from the REHA data set. The scan details in the top row show three different segmentation cases of pairs of legs. Regarding the segmentation, a pair of legs can result in two different segments, which allows a good description of the segments by circle features. However, one leg can be occluded by the other leg, and legs can even be merged to one segment. The bottom row shows different views of a person with a crutch, a wheelchair, and walker.

A summary of the essential characteristics of the data sets is shown in Tab. I. This table shows the proportion of merged and occluded legs as well. Note, that a smaller robot is used in the HOME environment. This is why the HOME data set is recorded by a laser range scanner in a height of 23 cm above the ground and the REHA data set in a height of 40 cm. The proportion of merged or occluded legs increases with the height of the laser scanner above the floor, because when people take a step, the distance of the legs decreases from the feet to the hip. The test data set of the HOME environment covers 1,250 range scans without persons and 1,250 range scans with legs. The REHA test data comprises 5,000 range scans, because additionally 1,250 range scans with walkers and 1,250 range scans with wheelchairs are included. Pictures showing typical scenes of the HOME and REHA environments are depicted in Fig. 5.

B. Detectors

To evaluate our proposed approach, we tested our features against two alternative feature spaces in combination with three different kinds of classifiers at the nodes of the class decision tree shown in Fig. 2, resulting in nine different approaches overall. The tested feature extractors are:

1) **ARRAS:** Our own re-implementation of the features of [2].
2) **SPINELLO:** The open source implementation[1] of features of [12].
3) **GDIF:** The new generic distance-invariant features proposed here.

The classifiers at the nodes of the class decision tree are:

1) **10-1:** An AdaBoost classifier, which combines 10 weak classifiers. Each weak classifier is a stump.
2) **50-1:** Like 10-1, but combining 50 weak classifiers.

[1] http://www.informatik.uni-freiburg.de/~spinello/people2D.html

[2] https://www.tu-ilmenau.de/neurob/data-sets/

TABLE I: Data sets

	SPINELLO	HOME	REHA
Laser range finder field of view	180°	270°	270°
Laser range finder angular resolution	0.5°	0.5°	0.5°
Laser range finder height above ground	?	23cm	40cm
Recorded range scans	38,994	24,249	30,582
Test data range scans	19,497 (50%)	2,500 (∼10%)	5,000 (∼16%)
Persons without walking aids	just beams labeled	18,022	13,503
Clearly separated legs	?	10,570 (59%)	4,790 (35%)
Occluded legs	?	3,092 (17%)	3,769 (28%)
Merged legs	?	4,360 (24%)	4,944 (37%)
Persons with wheelchairs	0	0	5,093
Persons with walkers	0	0	4,219

3) **50-10:** An AdaBoost classifier, which combines 50 weak classifiers, where each classifier is a decision tree with a maximal depth of ten.

In the following, the combination of feature extractor and classifier are named by concatenation of both specifiers. Accordingly, ARRAS-10-1 specifies the approach in [2], SPINELLO-50-1 is similar to the approach in [12], and GDIF-50-10 is the proposed approach of this work.

For all the detectors the same jump distance $\Delta = 0.1$ m is used. For the proposed features (GDIF) the baselines l_j have a width of $w = 1.0$ m and the clipping distance is set to $d = 3.0$ m. The baselines are divided into $n = 15$ line segments of approx. 6.7 cm each. Since the data sets were captured with laser scanners at different height above the ground (see Tab. I), the appearance of people differs slightly between these data sets. Therefore, we decided to train separate classifier trees for each of the data sets.

C. Detection Quality

In the first experiment, we tested the benchmark approaches on the SPINELLO data set. For evaluation of this data set, we used the same evaluation measure like in [12]. In the ground truth data, each beam is labeled as person or background, depending on what reflected the beam. Therefore, after the segments are classified as person or non-person, the actual evaluation is based on the individual beams, which belong to these segments. Furthermore, there is no *walker* and no *wheelchair* class, meaning that only one classifier c_1 for classification of *legs* and *no person* is necessary.

The precision-recall curves, generated by variation of the AdaBoost classification threshold θ, are shown in Fig. 6. As stated by Spinello et al., this data set is relatively simple, and the background does not change. This is the reason, why the classifier 50-10 is able to classify this data set almost perfectly, independent of the applied feature extractor, and therefore these curves are not plotted. The plotted curves

(a) HOME environment

(b) REHA environment

Fig. 5: Views of the operating environments where the HOME and REHA laser range data sets were captured.

confirm, that with simple AdaBoost weak classifiers, the ARRAS and SPINELLO features show similar performance, and the GDIF outperform both of them. Furthermore, the use of 50 weak classifiers increases the detection rate compared to the use of ten weak classifiers.

The next experiment was performed on the HOME data set. In contrast to the first data set, we were not interested in the classification of beams, but in the detection of persons. Therefore, the evaluation is based on actual person detections. For the ARRAS and SPINELLO approaches, a grouping of individually detected legs to pairs of legs is necessary. If two positively classified segment centers are closer than 0.8 m, this results in one person hypothesis at the central point between the segments' centers. If a segment is classified positively without a second positively classified segment nearby, the segment's center point is directly treated

Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh, UK, pp. 767-773, IEEE 2014

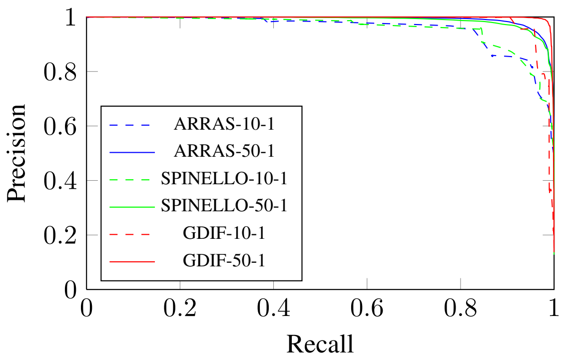


Fig. 6: Precision-recall curve for different combinations of feature extractors and classifiers for laser beam classification on *SPINELLO* data set.

.

as person hypothesis. If a positively classified segment can be assigned to multiple positively classified segments, the assignment of segments is optimized applying the Hungarian method [16]. The GDIF approaches do not detect individual legs, but person hypotheses and therefore the detected center positions C_j can be used directly.

In order to decide whether a detection corresponds to a true person, the distance to the closest person is evaluated. If the distance is below 0.7 m and no other detection is closer to the person, the hypothesis is regarded as successful detection. All other hypotheses are treated as false positive detections. Like for the SPINELLO data set, only one classifier c_1 is trained to distinguish *legs* from *no person*.

The precision-recall curves of this experiment are shown in Fig. 7. They illustrate, that the GDIF outperform the ARRAS and SPINELLO features again. To provide a single measure of quality of a detector D, we use the maximum F_1 score over the detector's AdaBoost threshold θ:

$$\max_{\theta} F_1 = \max_{\theta} 2 \cdot \frac{\text{precision}\,(D_\theta) \cdot \text{recall}\,(D_\theta)}{\text{precision}\,(D_\theta) + \text{recall}\,(D_\theta)} \tag{1}$$

The best F_1 score for the ARRAS features is 0.90, while it is 0.97 for the GDIF.

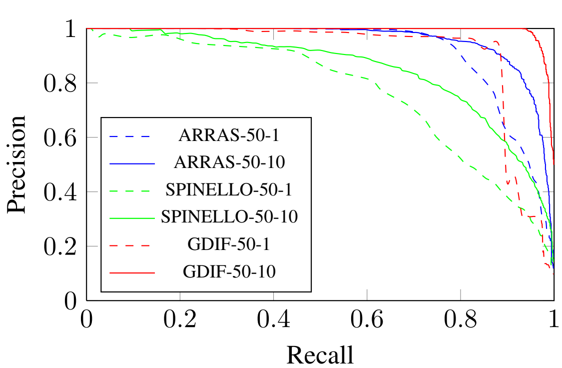


Fig. 7: Precision-recall curve for different combinations of feature extractors and classifiers for detection of persons *without walking aids* on our *HOME* data set.

The next three experiments are performed on the REHA data set. Here, additionally the quality of classification regarding the walking aid is evaluated, by taking the correctness of the label a_j into account. Fig. 8 confirms, that the performance of the GDIF is better than the performance of the ARRAS or SPINELLO features for the separation of *legs* and *walkers* from *wheelchairs* and *no person* (classifier c_1) on REHA data set. The best F_1 score for the ARRAS features on the REHA data set is 0.84, and for the GDIF it is 0.98.

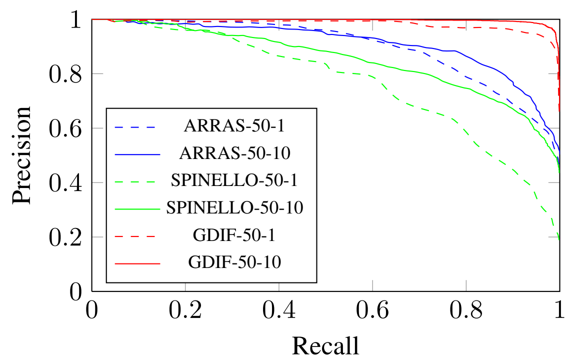


Fig. 8: Precision-recall curve for different combinations of feature extractors and classifiers for separation of *legs* and *walker* from *wheelchair* and *no person* (classifier c_1).

Fig. 9 shows the performance for separation of people with walkers from people without walking aid (classifier c_2). The best F_1 score for the ARRAS features is 0.84, and for the GDIF it is 0.92.

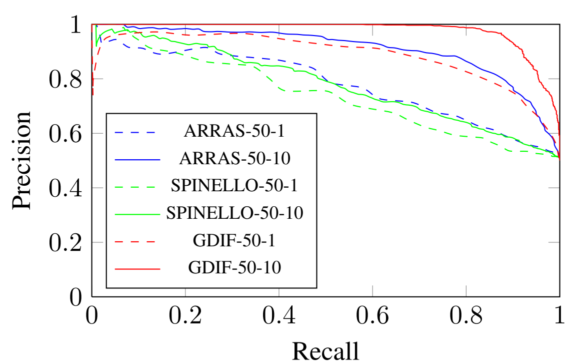


Fig. 9: Precision-recall curve for different combinations of feature extractors and classifiers for separation of *legs* from *walker* (classifier c_2) on REHA data set.

The classification performance of c_3 for classification of *wheelchair* and *no person* is shown in Fig. 10. The best F_1 score for the ARRAS features is 0.82 and for the GDIF it is 0.98.

In the following, the complete decision tree with all three classifiers at their best F_1 scores is evaluated. The resulting confusion matrix for **GDIF-50-10** and ARRAS-50-10 is shown in Fig. 11, with the values for GDIF highlighted in bold. The higher values in the principal diagonal and lower values in the *no person* row show, that the GDIF outperform the ARRAS features.

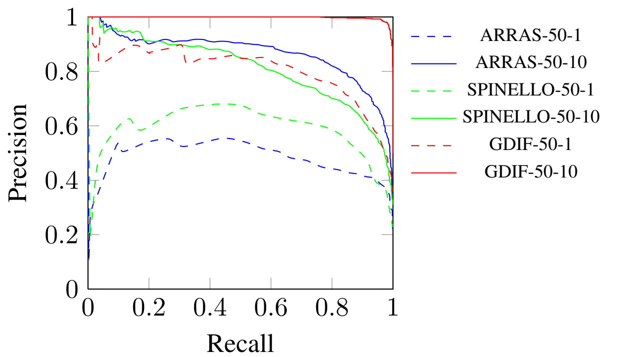


Fig. 10: Precision-recall curve for different combinations of feature extractors and classifiers for separation of *wheelchair* from no person (classifier c_3) on REHA data set.

	Legs		Walker		Wheelchair		No person	
Legs	**1154**	1057	**84**	54	**1**	26	**11**	113
Walker	**163**	262	**1023**	681	**59**	149	**5**	158
Wheelchair	**36**	86	**58**	117	**1137**	873	**19**	174
No person	**28**	392	**18**	71	**26**	206		

Fig. 11: Confusion matrix of **GDIF-50-10** (boldface) and ARRAS-50-10 on the REHA data set

The table shows, that using GDIF-50-10 an F1 score of 0.99 is achieved for detection of people with and without walking aids, and 86 % of the detections are classified correctly regarding their walking aid. For comparison, using ARRAS-50-10 results in an F1 score of 0.86, and 57 % correct discrimination of walking aids.

D. *Computing Effort*

Next to the detection quality of the features in combination with the classifiers, the computing effort is relevant for mobile robotic application. The average number of CPU cycles of our reimplementation for the extraction of the ARRAS features is $180 \cdot 10^3$ (the open source implementation of the SPINELLO features is even slower), and for the GDIF it is $65 \cdot 10^3$. Thus, the extraction of the proposed features takes just 36% of the time. The complete person detection on a 270° laser range scan according to the proposed approach using a decision tree with three GDIF-50-10 classifiers takes $7.075 \cdot 10^3$ cycles on average which are 2.5 ms on a 2.8 GHz CPU. This corresponds to a maximum detection rate of almost 400 Hz on a machine doing person detection only. Accordingly, for a laser range scanner with 10 Hz update rate, the proposed detection algorithm takes less than 2.5% of one 2.8 GHz CPU core.

V. CONCLUSIONS AND FUTURE WORK

This work presents an approach for detecting people in range scan data even when they use walking aids which occlude their legs. Therefore, generic distance-invariant features are proposed. These features are unspecific to the objects to be detected, and the features' extraction area is not dependent on any segmentation algorithm. A jump distance-based segmentation of the range scan is just applied to identify origin points for feature extraction. The length of the extraction area is based on the proportion of the objects to be detected. Using these features, the person detection quality increased from an F_1 score of 0.86 to 0.99 compared to the features of [2]. Since the features are really easy to extract, the computational effort for the feature extraction is lower compared to [2], and overall this approach is able to detect people in laser scans in realtime with no significant computational load on a contemporary CPU. Furthermore, the proposed generic distance-invariant features (GDIF) allow to classify 86 % of the detections correctly regarding their walking aid compared to only 57 % of correct classifications when using the features of [2].

In future, we will investigate whether a replacement of the manually designed decision tree by a data-driven tree structure improves the classification quality or even enables to perform view point estimation based on laser range data similar to [17] on image data.

REFERENCES

[1] H.-M. Gross, K. Debes, E. Einhorn, J. Kessler, S. Mueller, A. Scheidig, C. Weinrich, and A. Bley, "Robotic rehabilitation assistant for walking and orientation training of stroke patients: A report on work in progress," in *Proc. SMC*, 2014, p. to appear.

[2] K. Arras, O. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *Proc. ICRA*, 2007, pp. 3402 –3407.

[3] T. Kanda, D. Glas, M. Shiomi, and N. Hagita, "Abstracting peoples trajectories for social robots to proactively approach customers," *T-RO*, vol. 25, no. 6, pp. 1382 –1396, 2009.

[4] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *IJRR*, vol. 22, no. 2, pp. 99–116, 2003.

[5] M. Kleinehagenbrock, S. Lang, J. Fritsch, F. Lömker, G. A. Fink, and G. Sagerer, "Person tracking with a mobile robot based on multi-modal anchoring," in *Proc. Workshop ROMAN*, 2002, pp. 423–429.

[6] J. Xavier, M. Pacheco, D. Castro, A. Ruano, and U. Nunes, "Fast line, arc/circle and leg detection from laser scan data in a player driver," in *Proc. ICRA*, 2005, pp. 3930–3935.

[7] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. EuroCOLT*, 1995, pp. 23–37.

[8] A. Carballo, A. Ohya, and S. Yuta, "Fusion of double layered multiple laser range finders for people detection from a mobile robot," in *Proc. MFI*, 2008, pp. 677–682.

[9] O. M. Mozos, R. Kurazume, and T. Hasegawa, "Multi-layer people detection using 2D range data," in *Proc. ICRA Workshop: People Detection and Tracking*, 2009.

[10] L. E. Navarro-Serment, C. Mertz, and M. Hebert, "Pedestrian detection and tracking using three-dimensional ladar data," *IJRR*, vol. 29, no. 12, pp. 1516 – 1528, 2010.

[11] L. Spinello, M. Luber, and K. O. Arras, "Tracking people in 3d using a bottom-up top-down detector." in *Proc. ICRA*, 2011, pp. 1304–1310.

[12] L. Spinello and R. Siegwart, "Human detection using multimodal and multidimensional features." in *Proc. ICRA*, 2008, pp. 3264–3269.

[13] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, no, Ed. Belmont, CA: Wadsworth International Group, 1984.

[14] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *American Control Conference, 1997. Proceedings of the 1997*, vol. 4, Jun 1997, pp. 2369–2373 vol.4.

[15] M. Volkhardt, C. Weinrich, and H.-M. Gross, "People tracking on a mobile companion robot," in *Proc. SMC*, Oct 2013, pp. 4354–4359.

[16] J. Munkres, "Algorithms for the assignment and transportation problems," *SIAM*, vol. 5, no. 1, pp. 32–38, 1957.

[17] C. Weinrich, C. Vollmer, and H.-M. Gross, "Estimation of human upper body orientation for mobile robotics using an SVM decision tree on monocular images," in *Proc. IROS*, 2012, pp. 2147–2152.