Proc. 13th Int. Conf. on Intelligent Autonomous Systems (IAS 2014), Padua, Italy, 12 pages, 2014

Generic Distance-Invariant Features for Detecting People with Walking Aid in 2D Laser Range Data

Christoph Weinrich*, Tim Wengefeld, Michael Volkhardt, Andrea Scheidig, and Horst-Michael Gross

Ilmenau University of Technology
Neuroinformatics and Cognitive Robotics Lab
Helmholtz-Platz 5 (Zusebau)
98693 Ilmenau, Germany
christoph.weinrich@tu-ilmenau.de

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. The proposed approach was used to train classifiers for detecting people without walking aids, people with walkers, people in wheelchairs, and people with crutches. By the use of these features, the detection accuracy of people without walking aids increased from an F_1 score of 0.85 to 0.96, compared to the state-of-the-art features of Arras et al. Moreover, people with walkers are detected with an F_1 score of 0.95 and people in wheelchairs with an F_1 score of 0.94. The proposed detection algorithm takes on average less then 1% of the resources of a 2.8 GHz CPU core to process 270° laser range data with an update rate of 12 Hz.

Keywords: person detection, 2D laser range data, rehabilitation robotics

1 INTRODUCTION

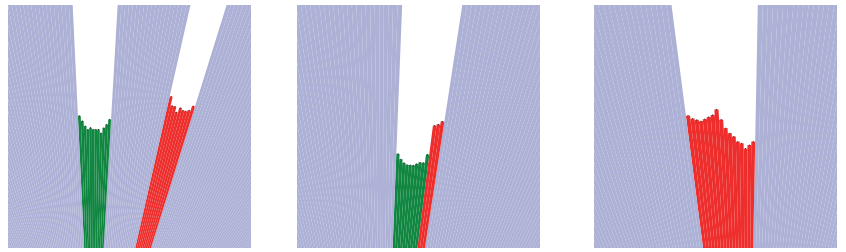
People detection and position tracking are important requirements to improve human-robot interaction (HRI), e.g. for the realization of socially compliant navigation of mobile assistance robots in populated public environments. Since

* 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 Fed. State of Thuringia and the European Social Fund (OP 2007-2013) under grant agreement N501/2009 to the project SERROGA (proj. no. 2011FGR0107).

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 over the floor, whereby 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.



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

(d) Person with crutch (e) Person in wheelchair (f) Person with walker

Fig. 1: Laser range scan details of persons (with walking aids). The individual scan segments, which are reflected by the persons or their 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. 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 [1] are no longer sufficient. Instead, the feature vectors have to be able to describe more complex structures. The proposed features are not designed for the detection

of any certain object's specific shape. 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 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, down-scaling reduces the high-frequency information content.

A great advantage of laser range data is the explicitly given distance of a sub-segment (for segmentation see Sec. 3.1). 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. 3 presents our approach, whose innovation are the generic distance-invariant features (GDIF). Sec. 4 demonstrates the advantages of the GDIF by detailed experiments.

2 RELATED WORK

There are various approaches for person detection in 2D laser range data, which work on multiple stationary laser scanners [2]. However, for our application, only approaches based on laser range scanners on mobile robots are relevant. In [1] approaches for leg detection are classified regarding their usage of motion or geometry features. Since approaches based on motion features (like [3]), 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 [4], a set of thresholds is used to classify sub-segments of range scan data as leg or non-leg. The focus of [4] is on person tracking, wherefore laser-based detection is just one cue. In contrast, [5] is directly focused on leg detection. The range data is segmented based on jump distances (see Sec. 3.1), 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 [1], the set of geometrical features is extended to 14 geometrical features, which are presumably suitable for leg detection. Then, AdaBoost [6] 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 [1], 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 [7, 8] or 3D range scanners [9, 10] have been proposed in recent years. However, on our mobile robot only the range data of one height level is perceived. Nevertheless, this approach could be easily extended to multiple layers, as well.

3 PEOPLE DETECTION BASED ON GENERIC DISTANCE-INVARIANT FEATURES

The input for the people detection approach is a laser range scan $R = \{B_1, ..., B_b\}$, which consists of a set of b beams, whereby 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 $G = \{P_1, P_2, ..., P_p\}$. Whereby each person P_i corresponds to a tuple (x_i, y_i) of the person's center position relative to the laser's coordinate system. Goal of this approach is to detect all people positions G.

3.1 Segmentation of 2D Range Data

Like in [1], in our approach the beams in the scan R are split into subsets of beams (see Fig. 2). Therefore the jump distance is applied. The first beam's B_1

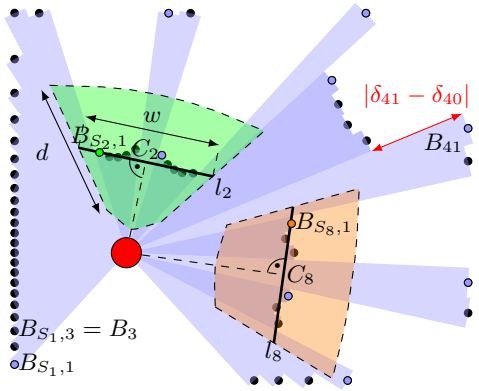


Fig. 2: Schematic illustration showing a range scan of a small room with two persons at different distances from the laser. The jump distance, the decision criterion for defining an new segment, between beam B_{41} and its former beam is exemplarily shown in red. The first beam $B_{S_j,1}$ of each found 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 in green and orange, whose feature vectors shall be classified as person.

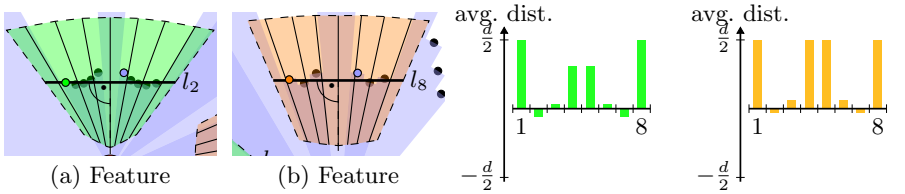
index is inserted in a new segment S_1. Iterating over the range scan R from beam B_2 to beam B_b, a new subset is initialized with the beam index i if the difference of the measurements $|\delta_i - \delta_{i-1}|$ of beam B_i to its former beam B_{i-1} is above a certain threshold Δ. Otherwise, the 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 [1], 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 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 [1] 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 [11] 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.

3.2 Feature Extraction



(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 nearby standing person

(d) Feature vector of a more distant standing person

Fig. 3: Visualization of range scan details of Fig. 2, which show the feature extraction areas of a person standing nearby the robot (a) and a more distant standing person (b). Exemplary, the extracted average distances to the base line for $n = 8$ line segments are visualized as bar histogram (c),(d).

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. 2 and Fig. 3). 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 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.

3.3 Classification

The dimensionality of the feature space F is determined by the product of line segments n and the number of extracted features. Discrete AdaBoost [6] is used for classification of the feature space by the classification function $h : F \rightarrow \{0, 1\}$ into person (1) and non-person (0). However, in contrast to [1] each weak classifier is a binary decision tree [12].

3.4 Hypotheses Generation

After a feature vector is classified as a person (possibly with walking aid), the center of the base line C_j is used as the 2D position for the person hypotheses. The result of the proposed approach are the detected person hypotheses $G' = \{C_j | h(f_j) = 1\}$.

4 EXPERIMENTS

4.1 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 [11][1] is used to evaluate the proposed approach on a publicly available benchmark data set.
2. **HOME:** We captured the HOME data set within living areas 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 Table 1. 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

[1] http://www.informatik.uni-freiburg.de/~spinello/people2D.html
[2] https://www.tu-ilmenau.de/neurob/team/weinrich/

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.

Table 1: 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

4.2 Detectors

To evaluate our proposed approach, we tested our features against two alternative feature spaces in combination with three different classifiers, resulting in overall nine different approaches. The tested feature extractors are:

1. **ARRAS:** Our own reimplementation of the features of [1].
2. **SPINELLO:** The open source implementation[1] of features of [11].
3. **GDIF:** The proposed generic distance-invariant features.

The classifiers 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.
3. **50-10:** An AdaBoost classifier, which combines 50 weak classifiers, whereby 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 [1], SPINELLO-50-1 is similar to the approach in [11], and GDIF-50-10 the proposed approach of this work.

For all the detectors the same jump distance $\Delta = 0.1m$ is used. For the proposed features (GDIF) the baselines l_j have a width of $w = 1.0m$ and the clipping distance is set to $d = 3.0m$. The baselines are divided into $n = 15$ line segments of approx. $6.7cm$.

4.3 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 [11]. 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.

The precision-recall curves, generated by variation of the AdaBoost classification threshold θ, are shown in Fig. 4. 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 confirm, that 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.

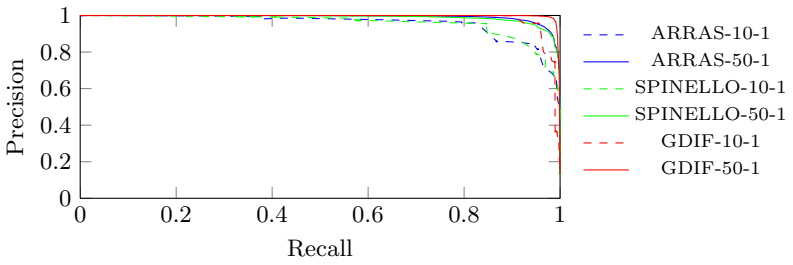


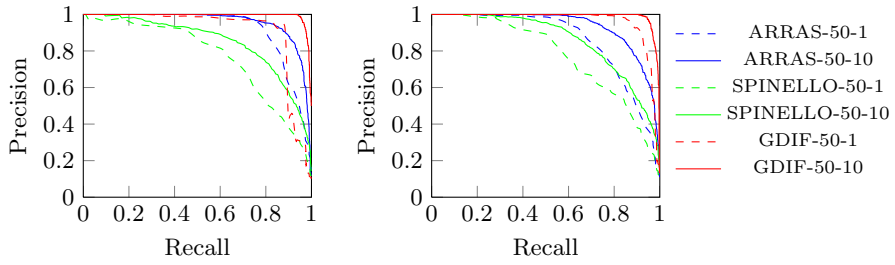
Fig. 4: Precision-recall curve for different combinations of feature extractors and classifiers for laser beam classification on *SPINELLO* data set.

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 GDIF approaches, the minimum distance of the detected persons'

center positions C_j to the present persons are calculated. If the minimum distance is below 0.7 m, this is a true positive. If the closest person is further away, this is a false positive. For the ARRAS and SPINELLO approaches, an additional 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' center points. If a segment is classified positively without a second positively classified segment nearby, the segment's center point is directly treated 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 [13]. The precision-recall curves of this experiment are shown in Fig. 5a. They show, 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.



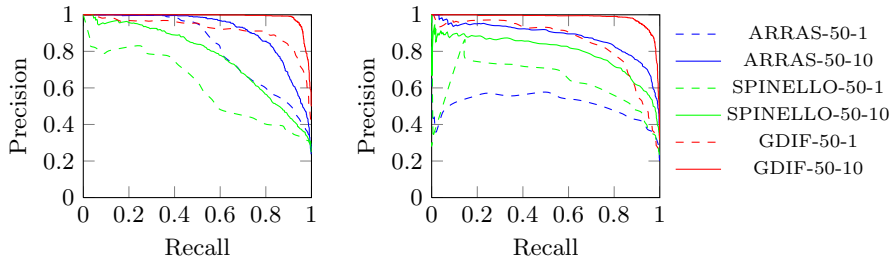
(a) HOME data set (b) REHA data set

Fig. 5: Precision-recall curve for different combinations of feature extractors and classifiers for detection of persons *without walking aids* on our *HOME* and *REHA* data set.

The next three experiments are performed on the REHA data set. Like Fig. 5a, Fig. 5b shows, that the performance of the GDIF is better than the performance of the ARRAS or SPINELLO features for the detection of person without walking aids. The best F_1 score for the ARRAS features on the REHA data set is 0.85, and for the GDIF it is 0.96.

Fig. 6a shows the detection performance of people with walkers. The best F_1 score for the ARRAS features is 0.83, and for the GDIF it is 0.95. The reason for this behavior is, the ARRAS features are designed for the detection of legs and the person's legs are mostly occluded by the walkers.

Finally, the detection performance for people in wheelchairs is shown in Fig. 6b. The best F_1 score for the ARRAS features is 0.82 and for the GDIF it is



(a) People with walkers (b) People in wheelchairs

Fig. 6: Precision-recall curve for different combinations of feature extractors and classifiers for detection of persons with *walkers* and *wheelchairs* on our *REHA* data set.

0.94. Compared to the detection of people without walking aids or people with walkers, the performance of the ARRAS features reduced significantly. This is probably, because of the complex structure of the wheelchairs, whose information content is reduced greatly by these features.

4.4 Computing Effort

Next to the detection quality of the features in combination with the classifiers, the computing effort is relevant for 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 GDIF-50-10 takes on average $2.224 \cdot 10^3$ cycles which are 0.79 ms on a 2.8 GHz CPU. This corresponds to a maximum detection rate of more than 1.2 kHz on a machine doing person detection only. Accordingly, for a laser range scanner with 12 Hz update rate, the proposed detection algorithm takes less than 1% of one 2.8 GHz CPU core.

5 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 dimensions of the extraction area are based on the proportion of the objects to be detected. Using this features, the leg detection quality increased from an F_1 score of 0.85 to 0.96 compared to the features of [1]. Since the extracted features are really simply

to extract, the computational effort for the feature extraction is lower compared to [1], and overall this approach is able to detect people in laser scans in realtime with no significant computational load on an up-to-date CPU. Furthermore, the proposed generic distance-invariant features allow to detect people with walkers with an F_1 score of 0.95 and for people in a wheelchair an F_1 score of 0.94 is determined compared to F_1 scores of only 0.83 and 0.82 when using the features of [1].

In future, we plan to train different classifiers for legs, walkers and wheelchairs and arrange them in a decision tree. Thereby, a multi-class decision regarding the walking aid might be possible [14].

References

1. Arras, K., Mozos, O., Burgard, W.: Using boosted features for the detection of people in 2d range data. In: Proc. ICRA. (2007) 3402 –3407
2. Kanda, T., Glas, D., Shiomi, M., Hagita, N.: Abstracting peoples trajectories for social robots to proactively approach customers. T-RO **25**(6) (2009) 1382 –1396
3. Schulz, D., Burgard, W., Fox, D., Cremers, A.B.: People tracking with mobile robots using sample-based joint probabilistic data association filters. IJRR **22**(2) (2003) 99–116
4. Kleinehagenbrock, M., Lang, S., Fritsch, J., Lömker, F., Fink, G.A., Sagerer, G.: Person tracking with a mobile robot based on multi-modal anchoring. In: Proc. Workshop ROMAN. (2002) 423–429
5. Xavier, J., Pacheco, M., Castro, D., Ruano, A., Nunes, U.: Fast line, arc/circle and leg detection from laser scan data in a player driver. In: Proc. ICRA. (2005) 3930–3935
6. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Proc. EuroCOLT. (1995) 23–37
7. Carballo, A., Ohya, A., Yuta, S.: Fusion of double layered multiple laser range finders for people detection from a mobile robot. In: Proc. MFI. (2008) 677–682
8. Mozos, O.M., Kurazume, R., Hasegawa, T.: Multi-layer people detection using 2D range data. In: Proc. ICRA Workshop: People Detection and Tracking. (2009)
9. Navarro-Serment, L.E., Mertz, C., Hebert, M.: Pedestrian detection and tracking using three-dimensional ladar data. IJRR **29**(12) (2010) 1516 – 1528
10. Spinello, L., Luber, M., Arras, K.O.: Tracking people in 3d using a bottom-up top-down detector. In: Proc. ICRA. (2011) 1304–1310
11. Spinello, L., Siegwart, R.: Human detection using multimodal and multidimensional features. In: Proc. ICRA. (2008) 3264–3269
12. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Wadsworth International Group, Belmont, CA (1984)
13. Munkres, J.: Algorithms for the assignment and transportation problems. SIAM **5**(1) (1957) 32–38
14. Weinrich, C., Wengefeld, T., Schröter, C., Gross, H.M.: People detection and distinction of their walking aids in 2d laser range data based on generic distance-invariant features. In: Proc. RO-MAN. (2014) to appear