Received: 24 August 2022 | Revised: 8 March 2023 | Accepted: 7 May 2023

DOI: 10.1002/rob.22195

RESEARCH ARTICLE

WILEY

Multiposture leg tracking for temporarily vision restricted environments based on fusion of laser and radar sensor data

Nils Mandischer | Ruikun Hou | Burkhard Corves

Institute of Mechanism Theory, Machine Dynamics and Robotics, RWTH Aachen University, Aachen, Germany

Correspondence
Nils Mandischer, Institute of Mechanism Theory, Machine Dynamics and Robotics, RWTH Aachen University, Aachen 52062, Germany.
Email: mandischer@igmr.rwth-aachen.de

Abstract

Leg tracking is an established field in mobile robotics and machine vision in general. These algorithms, however, only distinguish the scene between leg and nonleg detections. In application fields like firefighting, where people tend to choose squatting or crouching over standing postures, those methods will inevitably fail. Further, tracking based on a single sensor system may reduce the overall reliability if brought to outdoor or complex environments with limited vision on the target objectives. Therefore, we extend our recent work to a multiposture detection system based on laser and radar sensors, that are fused to allow for maximal reliability and accuracy in scenarios as complex as indoor firefighting with vastly limited vision. The proposed tracking pipeline is trained and extensively validated on a new data set. We show that the radar tracker reaches state-of-the-art performance, and that laser and fusion tracker outperform recent methods.

KEYWORDS

adaptive Kalman filter, laser, leg tracking, multiposture, radar, rescue robotics, sensor fusion

1 | INTRODUCTION

Firefighting is a demanding task for the human rescue operator. In recent efforts, mobile robots and drones are deployed to reduce the risk and improve the efficiency of a firefighting mission (Kruijff-Korbayova et al., 2016). However, autonomous robots are not yet enabled to interact directly with the rescue operators in indoor firefighting missions. This is due to a lack of (a) acceptance and (b) reliability of the overall sensor and actuator system. Our effort is to bring autonomous and self-sufficient robots to indoor firefighting that understand their environment and improve reliability through redundancy and robustness.

Our vision is to deploy rescue robots that can operate close to the firefighters' squad in a human–robot collaboration approach (see Figure 1). The robot follows and interacts with the firefighters directly. However, as optical, auditory, and telecommunication is expected to be mostly impaired, we cannot rely on control interfaces operated by a squad member or the mission control. Hence, the robot needs to interpret human behavior, reason on it, and deduct its own actions accordingly. By this, the robot becomes a collaborative rescue robot. The first step towards achieving this goal is to detect, where firefighters are located relative to the robot. This article proposes a novel leg tracker that can deal with the complex environmental influences and varying postures applied in a firefighting mission. The main contributions of our research presented in this article are:

- Developed a novel multiposture leg tracker based on Random Forests and Kalman Filters (KFs), which incorporates three classes of postures: `Standing`, `Squatting`, and `Background`.

FIGURE 1 Firefighters traversing a building as collaborative squad with a collaborative rescue robot.

- Adapted the multiposture leg tracker for radar and laser sensors, with special emphasize on feature extraction, data association, and track management.
- Generated novel annotated data set containing multiposture radar and laser data in three environments. The data set is available via Zenodo (Mandischer & Hou, 2023).
- Implemented trackers with track-level fusion. Code (in C++) and trained trees are available at *GitHub* (https://github.com/nilsmandischer/multi_posture_leg_tracker).

While leg trackers are commonly used to identify people walking, they are not made to detect other body and leg postures. In firefighting, the rescue operators often choose a squatting over a standing posture, hence, common approaches are impaired. We counteract this by extending the usual binary classifier (leg vs. nonleg) to a three class problem, incorporating classes `Standing`, `Squatting`, and `Background`. Likewise, the environmental influences are much higher in firefighting, as vision restrictions must be accounted for. We deploy radar sensors to deal with vision restricted scenes and fuse in laser data in a track-level fusion approach to improve the radar tracker in case of unrestricted vision. By this, the tracker becomes robust against diverse environmental influences while maintaining a common consensus through the radar tracks. As there is no purely planar approach to multiposture leg tracking in the state of the art, we record a new data set incorporating radar and LiDAR sensor data to train and evaluate our proposed tracking pipelines. All data, and implemented and trained trackers are available open source (see above).

The article first discusses the state of the art in laser- and radar-based leg tracking, including a brief overview of multisensor fusion (Section 2), and the methodological background of the article (Section 3). Afterwards, Sections 4, 5, and 6 present the proposed methods in laser and radar leg tracking, and their track-level fusion, respectively. Section 7 discusses the data set used for classifier training before Section 8 presents the validation of the proposed methodology. Finally, Section 9 summarizes the article and gives an outlook towards future developments.

2 | RELATED WORK

Leg tracking is an established field in mobile robotics, and particularly social and service robotics. Approaches differ in methodology and pipeline architecture. Usually, tracking consists of three steps: (1) detection, (2) tracking, and (3) track management. Detection divides the scan set into potential leg clusters and establishes a consensus over each cluster's class (leg and nonleg). The tracking step updates tracks with the incoming leg detections and predicts further motion. Finally, track management initializes and deletes tracks based on a set of rules. While the methodology may differ, the overall pipeline is similar for radar and laser leg tracking. Our leg tracker is designed for robotic systems without dedicated graphics accelerators. Therefore, this section majorly focuses on traditional learning and heuristic approaches rather than deep learning. In fact, more explainable artificial intelligence may benefit acceptance and trust (Shin, 2021). We still give a brief overview over recent developments in this field of deep learning-based leg tracking to account for recent trends.

2.1 | Laser-based leg tracking

This section covers laser-based leg and dynamic object tracking methods. Lindstrom and Eklundh (2001) find moving objects by observing the environment for "violations", that is, scan points that violate prior observations incorporating the robot motion model. Kluge et al. (2001) detect moving objects by convexity and track them based on Object Flow. Fod et al. (2002) use Blob Analysis to find candidates and track moving objects by distance threshold. However, none of these methods is capable to distinguish legs from other dynamic objects.

Scheutz et al. (2004) use the same approach as Fod et al. (2002) but detect legs based on human walking patterns. Cui et al. (2008) extend these approaches towards leg tracking. Both methods only work with a sequence of scans and not on static frames. Fritsch et al. (2003) define five typical features—including point count, mean, and standard deviations—to identify legs by applying fixed thresholds.

Xavier et al. (2005) improve on the methodology by applying geometric features, namely, lines and circles. Liu et al. (2022) modify the HDBSCAN algorithm (Campello et al., 2013) to extract leg clusters. However, according to Chung et al. (2012) assuming a fixed shape is not sufficient as clothing may alter the appearance of leg clusters.

Arras et al. (2007) are the first to apply ensemble classifiers. They define 14 features and use Adaptive Boosting (AdaBoost) to identify legs. Consequently, their method provides an improved performance compared with the prior heuristic approaches. Spinello and Siegwart (2008) extend the method by merging clusters to compensate for oversegmentation. They define a 50-dimensional feature vector and apply a Support Vector Machine (SVM) as leg detector. Fotiadis et al. (2013) point out that distance-based features may impact the adaptability towards new—in particular outdoor—environments, and propose a new distance-independent feature set. Sharif (2020) uses an SVM-based detector to distinguish movers, that is, cars or pedestrians, in autonomous driving. He combines the detector with a particle filter and the Hungarian (Kuhn, 1955) algorithm for track association. Linder et al. (2016) propose a multimodal detection and tracking framework, and compare classifier methods. They observe that a Random Forest classifier performs best for leg detection. Leigh et al. (2015) define 15 features based on the work of Arras et al. (2007) and detect legs with a Random Forest classifier, followed by the application of a KF for tracking. Bellotto and Hu (2010) focus on tracking, and compare Bayesian filters. They further apply a Mahalanobis gating procedure and Greedy Nearest Neighbor (GrNN) for track association. Linder et al. (2015) compare data association methods, namely, Nearest Neighbor (NN), Global Nearest Neighbor (GNN), Greedy NN, and a multihypothesis tracker. They emphasize the importance of the management logic to run a sophisticated NN-based tracker system.

In recent years, deep learning gained significance, hence, these methods were also applied to leg tracking. Guerrero-Higueras et al. (2019) propose the PeTra detection and tracking network based on the U-Net architecture (Ronneberger et al., 2015). They further propose an extension in Álvarez-Aparicio et al. (2019). PeTra is better understood as a static background filter, which is applied to extract leg clusters. Efstathiou et al. (2021) propose the LTGADnet network for leg detection based on the popular YOLO architecture (Redmon et al., 2016). However, measures are rather weak (accuracy 71%, f1-score 70%) which prevents a widespread use. Kohara and Nakazawa (2019) propose a PointNET AutoEncoder for leg detection based on the PointNET architecture (Charles et al., 2017). The approach reaches good classification scores for one-class leg-detection but is expensive in computational resources. Indeed, deep learning-based trackers, while showing potential in leg tracking, currently require intensive hardware usage.

Jia et al. (2020) propose the Distance Robust Spatial-Attention and Autoregressive Model (DR-SPAAM) for people detection. The algorithm is based on the DROW detector (Beyer et al., 2018), but reaches better robustness and accuracy, while using less computational resources. Jia et al. (2021) improve on the detector using a self-supervised label generation through an additional camera sensor. Hence, DR-SPAAM overcomes the hardware limitation of deep learning in embedded systems, but average precision is still lower than some ensemble classifiers, for example, in Linder et al. (2015). In addition, firmer typically lack explainability, which prevents usage in safety-critical applications like firefighting. As features in ensemble classifiers are usually hand-crafted, they mitigate this challenge.

2.2 | Radar-based leg tracking

This section covers radar-based leg and dynamic object tracking methods. Radar is particularly suited for dynamic object detection as frequency modulation ("Doppler effect") allows one to detect the relative velocity of target objects. Therefore, many human tracking methods utilize this modulation. Exemplary, Narayanan et al. (2014) show that micro-Doppler signatures may be used to detect certain body motions, including breathing.

Zhang et al. (2007) and Ahtiainen et al. (2010) both detect legs by identifying walking patterns in micro-Doppler signatures. Similarly, Rohling et al. (2010) distinguish pedestrians and vehicles. Held et al. (2022) use a kinematics-informed motion model to detect leg and upper body motion based on micro-Doppler signatures. They track legs using an Extended Kalman Filter (EKF). Fang et al. (2022) use a combination of KT-constant false alarm ratio and MUSIC algorithm to detect the human's pose and embed it into a particle filter. They emphasize the usage of radar in degrading conditions, particularly in rain. Zhang et al. (2022) propose a sequential Monte Carlo-based track-before-detection method and analyze the effects in similar conditions as in Fang et al. (2022). Both publications confirm that radar is not affected by degrading environmental conditions. Qian et al. (2020) use Joint Probabilistic Data Association on ultra-wideband (UWB) sensors for people tracking with noise suppression. UWB sensors have also been shown to be well applicable in firefighting (Tiemann et al., 2020).

Bartsch et al. (2012) emphasize that human gait modeling approaches, like in Ahtiainen et al. (2010), Rohling et al. (2010) or Held et al. (2022), always require a sequence of scans rather than a single frame to be analyzed. Therefore, they heuristically identify people by applying five features based on geometry and Doppler velocity. Castanheira et al. (2020) first segment the scan with DBSCAN (Ester et al., 1996) and label detections using a Random Forest classifier. Majer et al. (2019) define a 62-dimensional feature vector, including geometry, velocity, and intensity features, and apply an SVM classifier.

In some scenarios, Doppler Modulation may not be a feasible source, for example, when using low bandwidth interfaces or pivoting sensors with unknown or low precision self-locomotion, that is, cases, in which complex relative motions occur. Zhai et al. (2018) concentrate on tracking and apply the Sage-Husa Adaptive Kalman Filter (SHAKF) to track people with drones. Zhao et al. (2019) apply DBSCAN combined with a standard KF. Lately, we proposed a method to track people based on a Random Forest classifier with 21 features (without velocity) and a SHAKF (Mandischer et al., 2021).

2.3 | Multisensor tracking and fusion

This section covers sensor-fusion approaches to people tracking using either or both of laser or radar sensors. In literature, many approaches use not only a single sensor source, but also combine multiple with the aim to achieve a more reliable representation of the environment as pointed out by Hackett and Shah (1990). Many fused object tracking approaches originate from driver assistant systems. Möbus and Kolbe (2004) fuse infrared and radar dynamic object trackers at track level. Chavez-Garcia et al. (2014) fuse LiDAR, radar, and camera data at detection-level for dynamic object detection and tracking. Further, Chavez-Garcia and Aycard (2016) improve the framework with uncertainty management and show the potential in a real-world application. Scheutz et al. (2004) fuse vision, laser, and sonar data, whereas each sensor has different objectives in the detection and only laser is used for leg tracking. Linder et al. (2016, 2015) track legs with laser and fuse people detection from RGB-D and monocular vision to get more robust people tracks. They fuse both information on the detection-level. Majer et al. (2019) use a separate LiDAR and radar leg tracker, but use the laser information to improve the radar detector in a lifelong-learning approach. They fuse all information on the track level.

Linder et al. (2021) compare different individual human trackers, based on 2D/3D laser and RGB-D, and cross-validate their usability in multimodal tracking systems. All trackers are fused on the detection-level. They observe that the fused trackers achieve better recall scores while reducing the precision. In their individual domain, 2D laser, RGB-D, and 3D LiDAR, DR-SPAAM (Jia et al., 2020), RGB-D YOLO (Linder et al., 2020), and SECOND-DV (Zhou et al., 2019) taught by RGB-D YOLO reach the best scores, respectively.

3 | FOUNDATIONS OF MULTIPOSTURE LEG TRACKING

The question may arise: Why do we need a multiposture leg tracker? On the one hand, adding additional classes adds redundancy. In fact, squatting leg detection compensates for those cases in which the detector is not able to identify a leg properly due to a lack of adaptability of the underlying classifier. On the other hand, in emergency rescue, squatting postures are far more common and mandatory to be detectable if a continuous track shall be established. Our multiposture tracker is designed with firefighting in mind, but also contributes to civil applications like autonomous driving or service robotics. We define three types of squat: `Deep Squat`, `Partial Kneel`, and `Full Kneel`. The postures are depicted in Figure 2.

Squatting is defined as a posture in which the legs are bent in a certain angle. In `Deep Squat`, both legs enclose an angle of approximately 90° or lower, while neither shin is touching the ground. The angular description is vague, as there is a continuous transition from `Standing` to `Squatting`. In `Full Kneel`, both shins are positioned parallel to the ground plane. Finally, `Partial Kneel` describes the intermediate posture, where one shin is in the air and the other is touching the ground. Both kneeling postures can only be reached by passing a state of `Deep Squat` (see Figure 3). Therefore, it is reasonable to combine these postures into a single class `Squat` or `Squatting`.

The proposed system is composed of multiple components that are mostly interchangeable for both sensor types. However, especially the descriptors are specific for the individual sensor. We

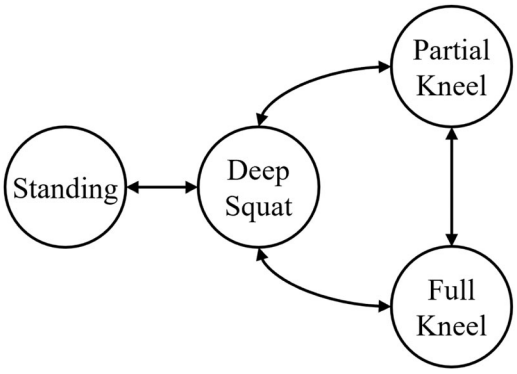

FIGURE 3 Transition graph between `Squat`-related postures and `Standing`.

FIGURE 2 Types of leg posture covered by the methodology. From left: `Standing`, `Deep Squat`, `Partial Kneel` (leg interchangeable), and `Full Kneel`.

design the tracker system with the tracking-by-detection paradigm: Once a scan arrives, the data are segmented into clusters that potentially depict legs. Afterwards, these candidates are labeled using a Random Forest classifier with three classes: `Standing`, `Squatting`, and `Background`. The results are then gated and forwarded to the tracker which uses some type of KF. The following sections discuss the methods used in both tracker subsystems (laser and radar), which are mostly adaptations of established methods and approaches from the state of the art.

3.1 | Candidate legs

In our prior work, we showed that Nearest Neighbor Clustering (NNC) outperforms other clustering algorithms, when it comes to leg extraction with radar (Mandischer et al., 2021). However, this method is also prominently used by other works for laser leg tracking, for example, Leigh et al. (2015). In NNC, every point $\mathbf{p}$ in an exclusive cluster Ω holds a Euclidean distance d_{NN} from any point in another cluster, that is,

$$\min d(\mathbf{p}_i, \mathbf{p}_j) \geq d_{NN}, \quad \mathbf{p}_i \in \Omega_i, \quad \mathbf{p}_j \in \Omega_j. \tag{1}$$

While the underlying algorithm is simple and fast (Konstantinova et al., 2003), the threshold d_{NN} significantly influences the size of the cluster objects. Therefore, d_{NN} needs to be determined carefully in experiments; the results of which are presented in Sections 4 and 5.

3.2 | Ensemble classifier

Usually, mobile robots lack computational power and graphics accelerators in particular. Therefore, deep learning techniques cannot be applied on a broad scale. For a more robust tracker, we choose an ensemble classifier. These classifiers suffer less from data bias or statistical variance compared with, for example, SVMs or Bayesian Learning (Dietterich, 2002). All classifiers in this work utilize a Random Forest (Breiman, 2001). This classifier trains many trees of specified depth, whereas each tree node carries a single feature decision. The individual features are chosen from a subset of features, which is randomly generated per tree. The trees are built and trained according to the Classification and Regression Trees (CART) methodology to maximize information gain on each node. As the Random Forest follows the rule of large numbers (Kolmogorov, 1956), a high accuracy is expected at the disadvantage of slower computational time compared with other traditional learning approaches.

3.3 | Tracking and motion model

In our previous work (Mandischer et al., 2021), we compared different kinds of KFs and a multihypothesis filter for tracking human legs with radar. We concluded that the SHAKF (Sage & Husa, 1969) performs best for radar leg tracking within the given restrictions. Like the original KF, the SHAKF uses a linear motion model, but establishes a more accurate noise estimator. This property provides robustness in applications with high noise.

We formulate the discrete linear system as

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \tag{2a}$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{r}_k, \tag{2b}$$

where $\mathbf{x}_k$ is the state and $\mathbf{z}_k$ the observation at time step k. $\mathbf{F}_k$ and $\mathbf{H}_k$ are system matrices. Further, $\mathbf{q}_k$ and $\mathbf{r}_k$ are process and measurement noise, respectively, modeled as Gaussian distributions with zero mean and known covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$. The SHAKF models the measurement noise covariance as

$$\mathbf{R}_k = (1 - d_k)\mathbf{R}_{k-1} + d_k \left(\mathbf{v}_k \mathbf{v}_k^T - \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T \right), \tag{3}$$

with the fading memory index

$$d_k = \frac{1 - b}{1 - b^k}. \tag{4}$$

Hereby, $b \in [0.95, 0.99]$ is the forgetting factor. Further, $\mathbf{v}_k$ is the prediction error, defined by

$$\mathbf{v}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-. \tag{5}$$

Hereby, $\hat{\mathbf{x}}_k^-$ is the prior state estimate with covariance $\mathbf{P}_k^-$, given by

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_k \hat{\mathbf{x}}_{k-1}, \tag{6a}$$

$$\mathbf{P}_k^- = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1}. \tag{6b}$$

Hereby, $\hat{\mathbf{x}}_k$ is the corrected state estimate with covariance $\mathbf{P}_k$. Using the Kalman Gain

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1}, \tag{7}$$

the corrected state estimate is given by

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \mathbf{v}_k, \tag{8a}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-, \tag{8b}$$

where $\mathbf{I}$ is a unit matrix.

3.4 | Data association and gating

In this work, we use two data association techniques. GNN establishes optimal track associations by minimizing a cost function using a linear assignment function. Here, the Mahalanobis distance between tracks and detections is chosen. To accelerate computation, we apply infinite costs if a pair exceeds a given gating test (compare

Mandischer et al., 2021). For minimization, we utilize the Hungarian method (Kuhn, 1955). This method provides optimal matching in which the sum of costs of all unique detection-to-track pairs is minimal. Second, GrNN uses similar mechanisms as GNN, but allocates pairs on a greedy basis, that is, instead of minimizing the whole candidate space, every iteration finds and keeps the pair with the smallest cost. Therefore, only a suboptimal assignment is established. However, the computation is accelerated compared with GNN (Linder et al., 2015).

4 | LASER SUBSYSTEM

Detecting squatting people in laser scans is a challenging task as the shape varies significantly depending on the exact posture and orientation of a person towards the sensor. Figure 4 shows some possible configurations of leg clusters. What makes the `Squatting` posture challenging is that different body parts (e.g., arms, legs, shins, stomach, and back) can be partially or fully covered by the scan, which also depends on the height the sensor is mounted at. Arms in particular induce random noise to the clusters as they can be positioned almost everywhere in the clusters, while other body parts follow specific rules as to where they may appear.

4.1 | Feature extraction

Leg clusters are determined with NNC and a threshold of d_{NN} = 0.2m. To cover the large bandwidth of leg appearances, we extend the features proposed by Leigh et al. (2015) and Linder et al. (2016). Overall we describe 30 laser features f_i as depicted in Table 1. Features f_9–f_{11} and f_{14}–f_{16} are extensions of the linearity and circularity features, respectively. In addition, f_{21} is the mean of angles $\angle(\mathbf{p}_1\,\mathbf{p}_i\,\mathbf{p}_{n_\Omega})$ inscribed by the first, last, and any intermediate point $\mathbf{p}_i \in \Omega$, where $i \in \{2, n_\Omega - 1\}$. Further, feature f_{23} is the ratio of maximal to minimal standard deviation σ in either the x or y direction,

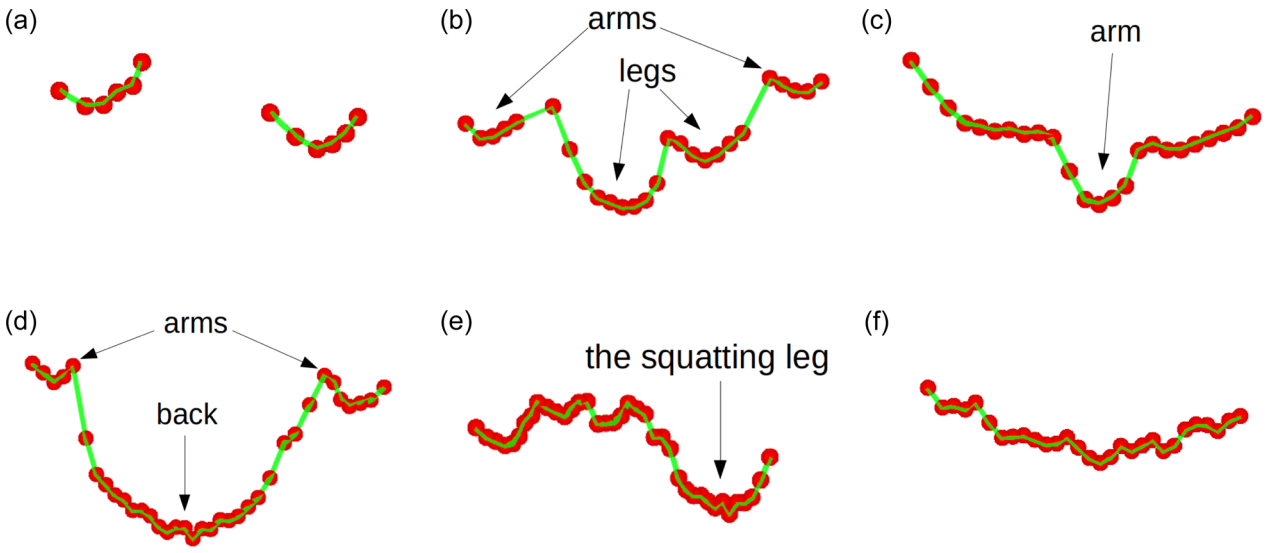

FIGURE 4 Laser clusters of a person in different postures. (a) `Standing` legs, (b) `Deep Squat` frontal, (c) `Deep Squat` sideways, (d) `Deep Squat` backwards, (e) `Partial Kneel` frontal, and (f) `Full Kneel` frontal.

TABLE 1 Laser features f_i.

1. Size of cluster	11. Ratio of f_9 to f_{10}	21. Inscribed angular variance
2. Standard deviation	12. Circularity	22. Standard deviation of inscribed angles
3. Mean deviation from median	13. Radius of best-fit circle	23. Aspect ratio
4. Width	14. Minimum circularity error	24. Kurtosis
5. Minimum distance between points	15. Maximum circularity error	25. Polygon area
6. Maximum distance between points	16. Ratio of f_{14} to f_{15}	26. Right side occluded
7. Ratio of f_5 to f_6	17. Boundary length	27. Left side occluded
8. Linearity	18. Boundary regularity	28. Number of local minima ranges
9. Minimum linearity error	19. Mean curvature	29. Difference of maximum and minimum range
10. Maximum linearity error	20. Mean angular difference	30. Ratio of minimum to maximum range

that is, $\max(\sigma_x, \sigma_y)/\min(\sigma_x, \sigma_y)$. The Kurtosis measures the outliers in a statistical distribution, therefore f_{24} correlates with heavy outliers and establishes knowledge of the noise level.

The feature vector is further extended using the normalization techniques proposed by Fotiadis et al. (2013). As leg clusters may vary depending on their distance from the sensor and the overall scan size (especially in radar scans), normalization is required to achieve a robust and scene-independent classifier. Explicitly, the final 148-dimensional feature vector contains

- the original features f_1–f_{30},
- f_1–f_{30} divided by distance from sensor,
- f_1–f_{30} multiplied by distance from sensor,
- f_2–f_{30} divided by f_1, and
- f_2–f_{30} multiplied by f_1.

4.2 | Tracking

The laser subsystem utilizes a standard KF with a constant velocity model. The usage of a the constant velocity model was experimentally verified. We observed that the update frequencies and the motion velocities applied allow one to linearize local motion, that is, a linear motion model is sufficient to approximate motion locally. We also tested other models for the laser subsystem, but they only allowed minor improvements while impacting the complexity. Hence, the constant velocity model was chosen. The transition matrix of the motion model is defined by

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

where δt is the time increment between steps $k-1$ and k. Its covariance matrix is defined by $\mathbf{Q}_k = \sigma_p^2 \mathbf{I}_4$. As the measurement vector $\mathbf{z}_k = [x, y]^T$ only contains position information, the transition matrix is defined by

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{10}$$

The measurement noise covariance is given by $\mathbf{R}_k = \sigma_m^2 \mathbf{I}_2$, with $\sigma_m < \sigma_p$ as the laser scans are usually reliable and precise.

4.3 | Data association

In track management, we distinguish between person tracks and tentative tracks. Latter are such that may belong to humans or arbitrary objects, that is, where no final class decision has been made. Person tracks are both, `Standing` and `Squatting` people. In track association, initially all person tracks are duplicated, such that they may be associated with two detections simultaneously. This is needed, as two legs or a single leg and any other body part (lower body, arm, etc.) may belong to the same tracked person. Tracks and detections are associated with a gate test using the GNN algorithm. The association step is performed twice. First, the detections are associated with the person tracks and their duplicates. Afterwards, the leftover detections are matched with tentative tracks. This procedure prevents that person tracks may lose detections due to association with tentative tracks, and may consequently be deleted. From all leftover detections after association, a grid map is generated, which is used to prevent malformed track initialization in later iterations. Therefore, the leftover detections are treated as occupied space, or background.

After association, tracks are updated by the associated detections using the KF. Further, the confidence of the track is updated according to the classification result of the detection. For person tracks, the update step varies in two cases:

- If a person track and its duplicate are associated with detections, but only one is labeled `Squatting`, the track is updated with the `Squatting` detection, such that the track follows the torso (instead of a potential arm detection).
- If only one of person track or duplicate is associated with a detection, the track is updated with either the position of the `Squatting` detection or the mean position of `Standing` detection and predicted track.

In all other cases, the person tracks are updated with the mean position of all associated detections. Afterwards, duplicates are deleted and unmatched detections become new tentative tracks.

4.4 | Track management

The initialization logic differs for person tracks labeled `Standing` or `Squatting`. It is to be noted that the following criteria require the tracks to be referenced in a static reference frame, for example, the map frame. A `Standing` person track is confirmed, that is, a pair of tentative tracks becomes a person track, under following conditions (all parameters optimized in parameter studies):

1. The confidence of both tracks exceeds 30%.
2. Both tracks are in free space (local grid map).
3. Both tracks move over 0.5 m without separation.

In contrast, a `Squatting` person track is confirmed from a tentative track if it fulfills following additional conditions, and replacing the separation criterion (3):

3. Total traveled distance is greater than 0.5 m.
4. The track was associated with only `Squatting` detections in four consecutive frames.
5. The track has moved more than 0.3 m from the initial position.
6. The velocity of the track is between 0.2 and 2 m/s.

7. The distance traveled in the current frame is between 0.02 and 0.2 m.

A person track can only be initialized if no other person track is located nearby to avoid creating duplicates of the same person. After initializing a person track, the according tentative tracks are deleted. Further, a tentative track may be deleted over time if it fulfills the deletion criteria in Section 5.5.

If a person track has been associated with more than 150 detections, it is regarded mature, while other tracks are declared young. For track deletion, we define a static and dynamic case, which refers to the motion state of the sensors. A track is deleted under following conditions (all parameters optimized in parameter studies):

1. The covariance exceeds 0.81.
2. The track is not matched for ζ_{del} consecutive frames.
3. The track travels incorrectly for χ_{del} consecutive frames.

We define $\zeta_{del} = \{30, 45, 15, 30\}$ and $\chi_{del} = \{20, 45, 10, 30\}$ (dynamic-mature, static-mature, dynamic-young, and static-young).

5 | RADAR SUBSYSTEM

Leg detections are vastly different in appearance in radar than in laser scans (compare Figures 4 and 5). While laser scans mostly show two separate leg clusters in case of `Standing`, `Deep Squad`, or `Partial Kneel`, radar only shows blobs of different geometrical shape and intensity. While, in theory reflection intensity may be used in both sensors, radar relies more on this information as geometric features become less significant. It is to be noted that the indurad iSDR-C radar sensors used in this work output a normalized intensity distribution.

5.1 | Radar prefiltering

In prior work (Mandischer et al., 2019), we observed that the high level of noise in radar scans deployed in indoor environments may be lowered by applying an adaptive threshold. Therefore, the radar scan is initially thresholded using the approach proposed by Otsu (1979). In Otsu Thresholding, a histogram is split into two classes, while the intra-class variance is minimized and the inter-class variance is maximized. We use the histogram of intensities and declare two classes: foreground and background. Hereby, the less intense background class contains objects that are mostly noise or located behind multiple other objects. The threshold th_{Otsu} is determined by maximizing the Otsu gradient λ, defined by

$$\lambda = \frac{\omega_A(\mu_A - \mu)^2 + \omega_B(\mu_B - \mu)^2}{\omega_A\sigma_A^2 + \omega_B\sigma_B^2}, \tag{11}$$

with two classes A and B and their mean μ, probability distribution ω, and variance σ (no index refers to whole scan), defined by

$$\sigma_A = \frac{1}{M}\sum_{i=\tau_{min}}^{thOtsu}(i - \mu_A)^2 m_i, \tag{12a}$$

$$\sigma_B = \frac{1}{M}\sum_{i=thOtsu+1}^{\tau_{max}}(i - \mu_B)^2 m_i. \tag{12b}$$

Hereby, τ is the discrete (and integer) intensity class, m_i the number of elements in each class, and M the overall scan size. We then use Stochastic Hill Climbing to solve the maximization problem. By this procedure not only the background to detection ratio but also the overall scan size is lowered, which improves computational efficiency of later steps.

5.2 | Feature extraction

As radar clusters are less precise, the NNC threshold is raised to $d_{NN} = 0.3$ m and clusters with less than five points are neglected. As depicted in Figure 5, patterns are not easily found in radar clusters. However, `Squatting` clusters usually contain more points and produce more eccentric geometrical shapes. For feature extraction, we use a similar feature set as in Section 4.1. Within, some geometric features are exchanged for intensity-based features. Further, since less knowledge can be gained from enclosed points, some features describe properties exclusively on the cluster's boundary ($g_2, g_{15}, g_{16}, g_{23}, g_{29}$). The feature set is depicted in Table 2. Again, the set is normalized which results in a 143-dimensional feature vector. To emphasize the difficulty in finding a good feature description in radar scans and consequently making good estimates, Figure 6 shows a comparison of radar and laser detections.

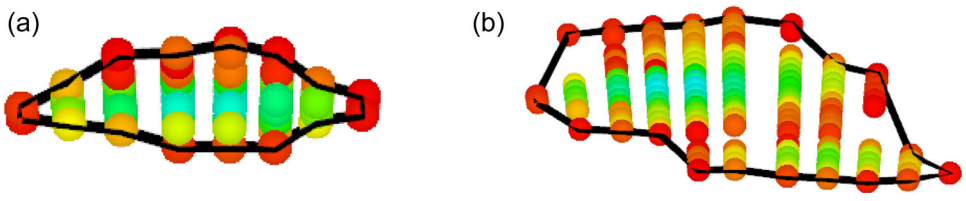

FIGURE 5 Radar clusters of a person in different postures. Colors (from red/low to green/high) indicate reflection intensities. (a) `Standing` person cluster and (b) `Squatting` person cluster.

TABLE 2 Radar features g_i.

1. Size of cluster	11. Radius of best-fit circle	21. Aspect ratio
2. Number of points on contour	12. Minimum circularity error	22. Kurtosis
3. Standard deviation	13. Maximum circularity error	23. Polygon area
4. Mean deviation from median	14. Ratio of g_{12} to g_{13}	24. Mean intensity
5. Width	15. Boundary length	25. Median intensity
6. Linearity	16. Boundary regularity	26. Standard deviation of intensity
7. Minimum linearity error	17. Mean curvature	27. Minimum intensity
8. Maximum linearity error	18. Mean angular difference	28. Maximum intensity
9. Ratio of g_7 to g_8	19. Inscribed angular variance	29. Average intensity on contour
10. Circularity	20. Standard deviation of inscribed angle	

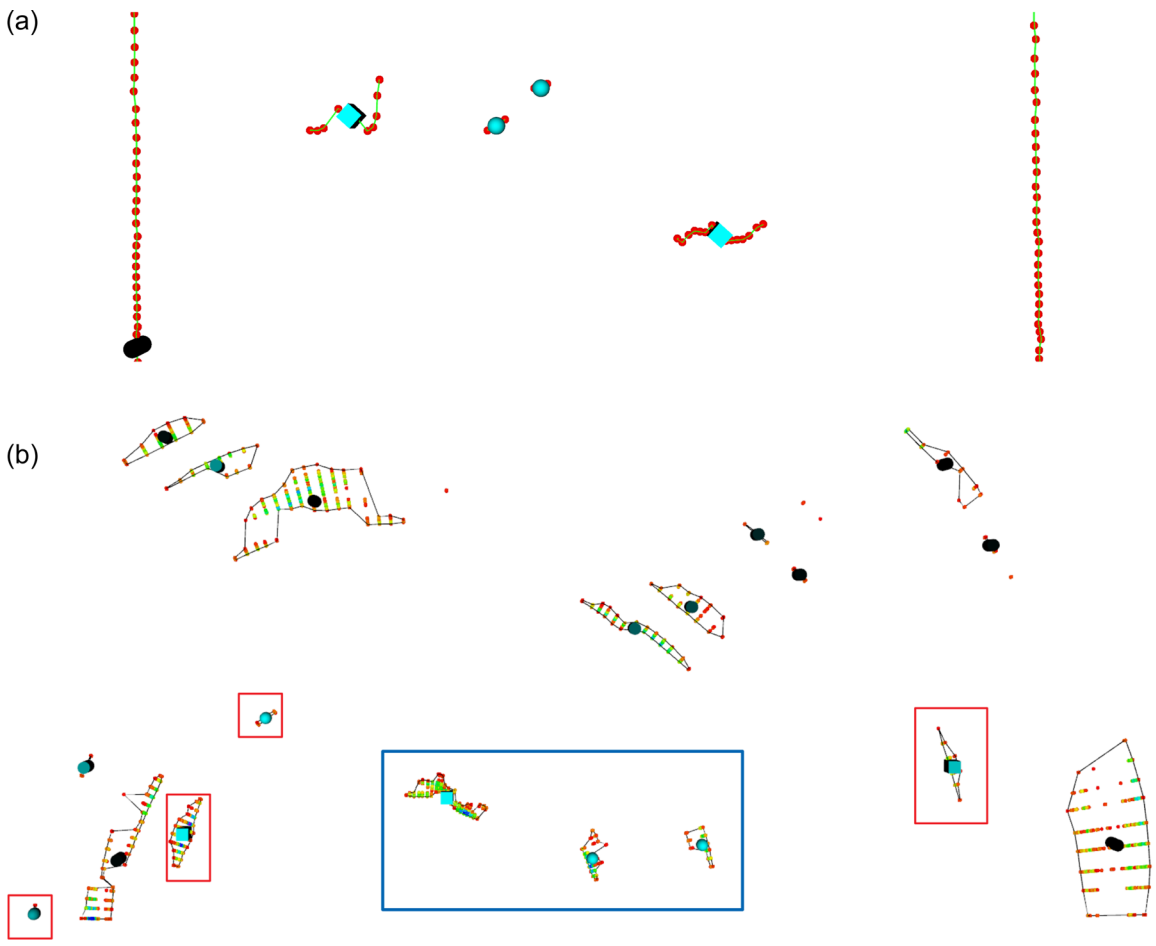

FIGURE 6 Classification results compared in laser and radar scans. Shown are `Standing` (blue sphere), `Squatting` (blue box), and `Background` (black cylinder) detections; both images are excerpts of the entire scene. (a) Laser detections after classification. Clusters are connected by green lines. (b) Radar detections after classification. The blue frame encloses True Positives and the red frames enclose False Positives (compare Section 8.1). Clusters are enclosed by black lines.

5.3 | Tracking

To accommodate for higher noise in the radar scan, the KF is exchanged with the SHAKF described in Section 3.3 and which we already successfully applied in Mandischer et al. (2021). We choose a forgetting factor of $b = 0.95$, which puts much higher weight on more recent scans. Due to the lower update frequencies, the process noise covariance matrix $\mathbf{Q}_k$ also needs to account for acceleration, and is consequently defined by

$$\mathbf{Q}_k = \begin{bmatrix} \frac{\delta t^4}{4} & 0 & \frac{\delta t^3}{2} & 0 \\ 0 & \frac{\delta t^4}{4} & 0 & \frac{\delta t^3}{2} \\ \frac{\delta t^3}{2} & 0 & \delta t^2 & 0 \\ 0 & \frac{\delta t^3}{2} & 0 & \delta t^2 \end{bmatrix}. \quad (13)$$

5.4 | Data association

As radar produces many more clusters than laser, GrNN is used for data association. As the detections are way less precise, cover the same object with multiple data points, and are subject to noise, it is not directly possible to create a local grid map similar to the laser subsystem. In Mandischer et al. (2019), we used the prefilter (Section 5.1) to further reduce the point set to a single point per cluster. However, in experiments we observed that this method may become unstable when applied to the volatile radar tracks. To counteract this instability and to save computational resources, we renounce the usage of a local grid map for the radar subsystem. To cope with the challenges induced by renouncing the local grid map, that is, missing discrimination from static map objects, person tracks are still duplicated, but the data association step is only performed once. Consequently, the detections are more likely to get matched with high certainty person tracks.

5.5 | Track management

Track initialization and deletion uses the same logic as presented in Section 4.4, but with different thresholds. Further, the logic used to initialize `Squatting` laser tracks from tentative tracks is applied to both `Standing` and `Squatting` track initialization in the radar subsystem. In addition, we also distinguish between the static and dynamic case on initialization. A track is initialized under following conditions (static/dynamic; all parameters optimized in parameter studies):

1. The confidence exceeds 30/50%.
2. Total traveled distance is greater than 0.7/1 m.
3. The track has moved more than 0.5/0.8 m from the initial position.
4. The track's velocity is between 0.1/0.3 and 2 m/s.
5. The distance traveled in the current frame is between 0.05/0.08 and 0.5/0.4 m.
6. The track is matched in two/eight consecutive frames.

Initialized tracks use the same convention for maturity as in the laser subsystem (Section 4.4). A track is deleted under following conditions (all parameters optimized in parameter studies):

1. The confidence is lower than ψ_{del}.
2. The track is not matched for ζ_{del} consecutive frames.
3. The track travels incorrectly for χ_{del} consecutive frames.

Here, $\psi_{del} = \{0.2, 0.2, 0.3, 0.3\}$, $\zeta_{del} = \{10, 20, 8, 15\}$, and $\chi_{del} = \{7, 15, 4, 10\}$ (dynamic-mature, static-mature, dynamic-young, and static-young). After initialization, radar tentative tracks are not directly deleted. Instead, they follow a similar deletion logic as the mature or young tracks but with different thresholds. They are deleted if criterion (2) is met with $\zeta_{del} = 8$, or if the track moved more than 0.6 m between two consecutive frames.

6 | TRACKING LEVEL FUSION

The aim of sensor fusion is to utilize the benefits of both tracking subsystems, while balancing their downsides. The laser tracker is precise in classification and positional accuracy, but is majorly impacted by weather or in vision degrading environments. In contrast, the radar tracker is not impacted by environmental conditions and also allows one to detect objects behind thin obstacles, but only reaches average accuracy. Hence, sensor fusion makes the system robust against degrading factors, while maintaining a reasonable accuracy.

The fusion method is based on Bayesian filtering as proposed by Bellotto and Hu (2010). Bellotto et al. (2015) provide a library for the lenient implementation of different Bayesian filter networks (https://github.com/LCAS/bayestracking). The fusion uses an EKF with a constant velocity model for tracking. The track association is then performed using the GrNN variant. However, in contrast to the individual trackers, each person is only represented by at most one input track, that is, detection, as legs are already merged in the subsystems. Tracks are initialized if a set of unmatched detections exists in a certain spatial and temporal space. At least five tracks from radar within 1.5 s or five tracks from a laser within 0.5 s need to be observed, before a track is initialized. Track deletion is performed based on a covariance threshold (sum of squared individual covariances exceeds 1.5). The full system is depicted in Figure 7. The final outputs of the track fusion are in the form of person tracks.

7 | DATA SET

Training and testing data sets share the same environments and setups. We make sure to strictly separate training and testing data in the later evaluation. The total data set consists of 31,300 laser and 9600 radar samples split into three classes `Standing`, `Squatting`, and `Background`. The composition of the data set and the split into training and testing sets are listed in Table 3. All data are collected using a Summit XL Steel mobile robot by Robotnik equipped with an *indurad iSDR-C* pivoting radar sensor and a *SICK S300 Professional* LiDAR. Both sensors provide planar 2D measurements of approximately 0.33 m (LiDAR) and 0.43 m (radar) off ground level. The full setup is depicted in Figure 8. The radar sensor by indurad is special in that it provides full 360° planer scans of the environment. Commonly, radar sensors are restricted in their angular range or 1D sensors are used, for example, in driver assistance systems. In addition, radar usually suffers from low angular resolution in the presence of multiple targets (Rasshofer & Gresser, 2005). The sensor used in this work, has a much higher angular resolution, even in close proximity. This comes with the downsides, that the maximal range is reduced to 16 m and that the sensor offers no Doppler or phase measures. However, the sensor provides standard data otherwise, hence, the transferability of the results to other sensors is given.

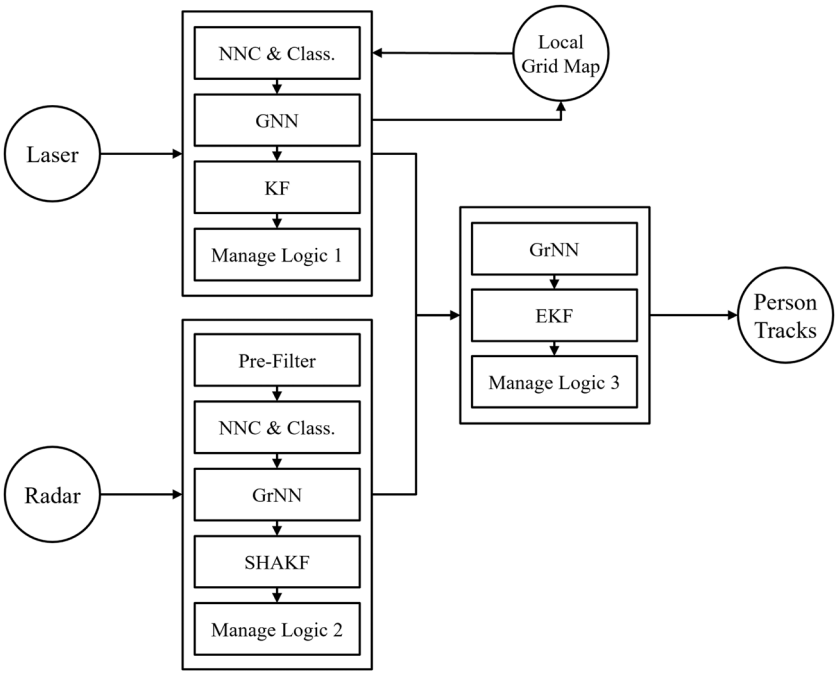

FIGURE 7 Individual and fusion pipeline from sensor input to person tracks. EKF, Extended Kalman Filter; GNN, Global Nearest Neighbor; GrNN, Greedy Nearest Neighbor; KF, Kalman Filter; NNC, Nearest Neighbor Clustering; SHAKF, Sage-Husa Adaptive Kalman Filter.

The following sections describe the reference environments. These are essentially the same as those used in our prior work (Mandischer et al., 2021). The data recorded is available at Mandischer and Hou (2023).

7.1 | Indoor environment

The lab at IGMR (*IGMR Lab*, Figure 9a) is used as an indoor reference environment. Within, many metallic objects induce noise to the radar readings and are, therefore, partially covered by wooden boards. The environment is cluttered and complex for both sensors. People move in direct visibility range of the robot's sensors, but are eventually occluded by a machine bed (Figure 9a, right side). Due to space limitation, a maximum of two people may walk around the robot.

7.2 | Outdoor environments

Outdoor data are generated in two different environments, which mainly differ in their controllability. The first is a walkway between two buildings (*Buildings*, Figure 9b). This environment allows for tests with only controlled influences. However, the lamp posts to the side and the metallic coating of one building induce false leg detections to both subsystems, and noise to the radar sensor in particular. As the walkway is particularly long, higher distances may be realized compared with all other environments. The *Buildings* environment fits up to five people. The second environment is a walkway along a road (*Road*, Figure 9c). The buildings alongside are mostly made of stone and have low noise induction potential. However, cars parking along or passing by on the road may induce noise. This environment is only traversed with a moving robot, while both other incorporate static and moving robot states. Therefore, this environment is only used for testing rather than training, but establishes the most realistic use-case.

TABLE 3 Composition of data sets for classifier training and testing.

Sensor	Set	Standing	Squatting	Background
Laser	Train	7763	5410	9148
	Test	3273	2479	3282
Radar	Train	2085	2195	2719
	Test	1720	1142	1814

8 | RESULTS

In this section, we evaluate the proposed system on the classifier and tracking level. We, further, compare the results to the state of the art, conclude with a discussion of the results, and show the potential and usability for vision-degraded environments. All tests use reference hardware (AMD Ryzen 7 5800H, NVIDIA GeForce RTX 3060 Laptop GPU, 16 GB RAM) and software (Ubuntu 18.04, ROS Melodic, openCV 3.4). All modules are implemented in C++.

8.1 | Classifier evaluation

For classifier evaluation, we consider every correctly labeled detection that overlaps with a real object as True Positive (TP). Along this definition, we define True Negatives (TNs), False Positives

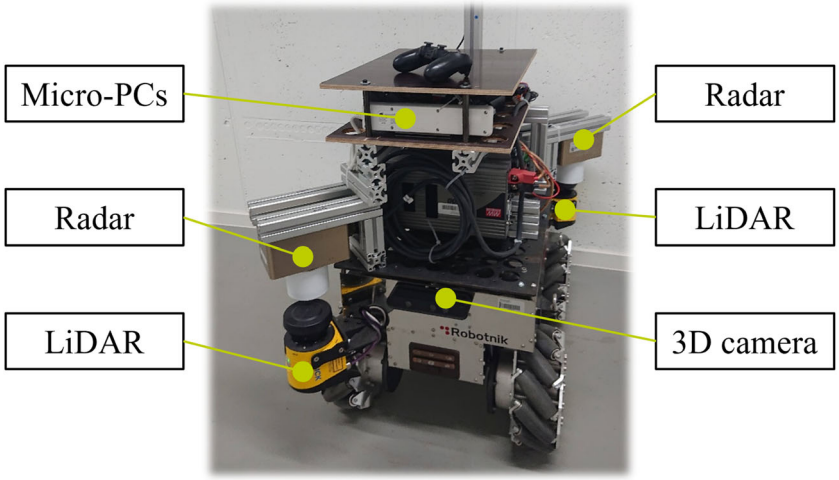

FIGURE 8 Hardware used in data generation and evaluation. The 3D camera Orbbec Astra is used for reference videos in the data sets.

FIGURE 9 Reference environments used for data generation and evaluation. Despite the conditions seen in the images, all outside data was taken in dry weather. (a) *IGMR Lab* (Mandischer et al., 2021), (b) *Buildings*, and (c) *Road*.

(FPs), and False Negatives (FNs). As the proposed system is a multiclass tracker we use macromeasures and accuracy for evaluation. Macro refers to the classes being evaluated individually and then combined into an average value. The measures precision pr, recall re, and f1-score $f1$ are defined by

$$pr = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FP_k}, \tag{14a}$$

$$re = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FN_k}, \tag{14b}$$

$$f1 = \frac{1}{K}\sum_{k=1}^{K}\frac{2TP_k}{2TP_k + FP_k + FN_k}, \tag{14c}$$

where K is the number of classes (here $K = 3$). Further, the accuracy ac is defined over all correct classifications (microaccuracy) by

$$ac = \frac{\sum_{k=1}^{K} TP_k + TN_k}{\sum_{k=1}^{K} TP_k + TN_k + FP_k + FN_k}. \tag{15}$$

To evaluate the influence of certain features on the class decision, Figures 10 and 11 depict the importance of radar and laser features, when no normalization is applied. In laser classification, most important features are width, left-sided occlusion, and circularity error. Cluster width is important to discriminate `Standing` and `Squatting` people, as clusters tend to be wider in `Squatting` posture. Occlusion, however, is atypical in leg detection. In our training set, the majority of occlusion comes from `Background` samples (9090 `Background`, 495 `Standing`, and 114 `Squatting`) as the pedestrians are usually in free sight of the robot and may only be occluded due to other pedestrians passing by. Therefore, the occlusion state is important to distinguish `Background` from other detections. The difference in left- and right-sided occlusion comes from a minor data set bias towards counter-clockwise motion as trial participants were free to choose the direction of movement. In radar classification, the overall features are more streamlined. The distance from the sensor is of major importance, as the ability to distinguish targets close to each other deteriorates at larger distances, hence, candidates are more likely to be detected in closer areas. This is due to the angular resolution of the radar sensor, which particularly takes effect in farther areas. However, the importance of the distance features is lower in the final feature set, as the distance from the sensor is used for

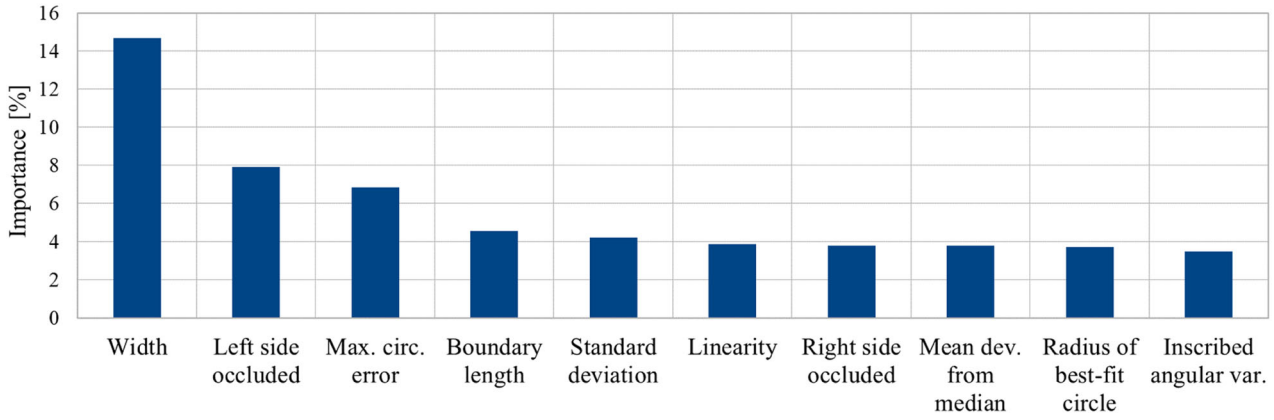

FIGURE 10 Ten most important laser features in class discrimination without normalization.

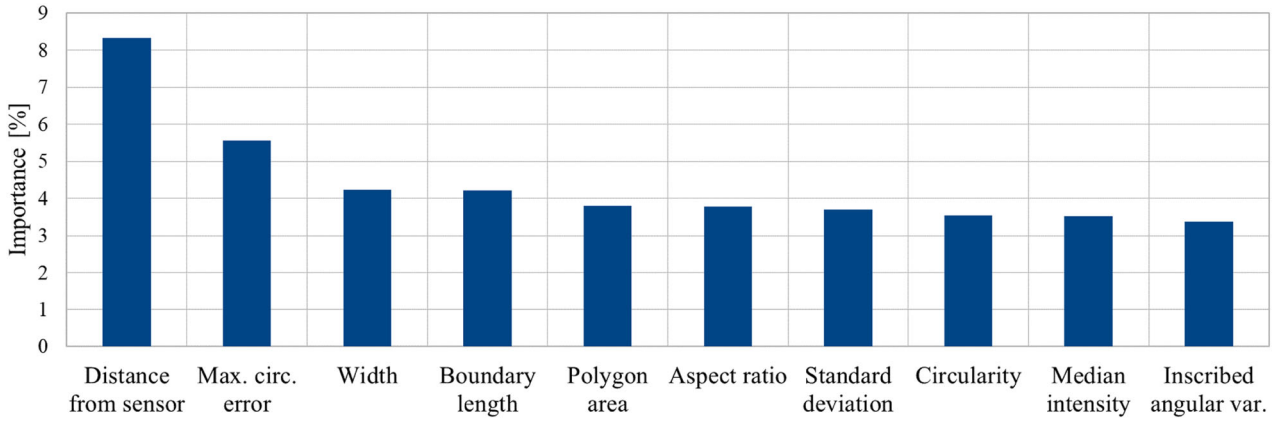

FIGURE 11 Ten most important radar features in class discrimination without normalization.

normalization. Note that the radar sensor used for data generation has a lower maximal range (18 m) than the laser sensor (30 m). Other important features are similar to the laser subsystem.

The performance of the classifiers is measured with Equations 14 and 15. Table 4 depicts the overall scores, while Tables 5 and 6 show the individual confusion matrices. The laser classifier discriminates well between all classes as indicated by the high values on the main diagonal. The radar subsystem reaches good results when discriminating `Squatting` from `Background`, but fails to well discriminate postures. However, in a collaborative firefighting scenario with limited vision, the system should track people robustly. Therefore, our objective is less to have a system that can discriminate `Squatting` from `Standing`, but rather identify people against `Background`. Hence, the radar tracker is well suited for this task. On the contrary, subpar discrimination of `Squatting` from `Standing` may limit the usage for high-level reasoning algorithms, for example, to identify the human task performed, which probably requires a human posture estimate, besides other features. The overall scores of the classifiers are impacted by the lack of discrimination between `Squatting` and `Standing` in the radar case.

TABLE 4 Scores of each classifier trained on indoor and outdoor data (runtime refers to average).

Sensor	*ac*(%)	*pr*(%)	*re*(%)	*f*1(%)	Runtime (ms)
Laser	97.17	97.20	97.09	97.14	4.44
Radar	78.40	79.78	74.62	75.07	24.84

TABLE 5 Confusion matrix of laser classifier (absolute and relative distributions).

GT \ Detected	`Standing`		`Squatting`		`Background`	
`Standing`	3174	96.98%	40	1.22%	59	1.80%
`Squatting`	31	1.25%	2407	97.10%	41	1.65%
`Background`	54	1.65%	59	1.80%	3169	96.56%

Abbreviation: GT, ground truth.

8.2 | Tracker evaluation

To measure the tracker performance, we use the CLEAR MOT (Bernardin & Stiefelhagen, 2008) metrics Multiple Object Tracking Accuracy (*MOTA*) and Multiple Object Tracking Precision (*MOTP*), defined by

$$MOTA = 1 - \frac{\sum_k (FP_k + MIS_k + IDS_k)}{\sum_k GTT_k}, \tag{16a}$$

$$MOTP = \frac{\sum_k \sum_{i=1}^{n_k} d_{i,k}}{\sum_k n_k}, \tag{16b}$$

where *MIS* is the number of missed detections (also referred to as FN) and *IDS* is the number of ID switches. We also count the case in which a known person reappears in the line of sight of the sensor and gets assigned a new ID as ID switch. Further, GTT_k is the number of ground truth tracks in the current frame k. In *MOTP*, d is the distance between track and GT track, and n_k is the number of matches. A person is labeled visible if that person generates a cluster of at least three points in the laser scan (compare Leigh et al., 2015), or five points in the radar reading, respectively.

For the feasibility tests, we use a static robot in the *Buildings* environment with 4–5 people. The pedestrians switch from `Standing` to `Squatting` posture arbitrarily and may move out or into the area covered by the sensors. Table 7 compares the tracker performances with varying people count. Within, cycle times are represented as the individual time taken to evaluate the subsystem from input to output. However, as the fused track may be generated from radar and/or laser tracks, its cycle time represents a mean of these subsystems based on the ratio of incoming radar and laser tracks, that is,

$$t_{fusion,total} = \max(t_{laser}, t_{radar}) + t_{fusion,individual} \tag{17}$$

per cycle. The laser tracker reaches a high *MOTA* in all cases with low impact from increasing the people count. However, the cycle time is higher as in the other subsystems due to the update of the grid map. The radar tracker loses almost 10% in *MOTA* between tests. The *MOTP* is sufficient for all test cases. The fused tracker shows good potential to increase the inaccuracies of the radar subsystem. However, due to the delays induced by the low update rate of the radar sensor, synchronizing both tracks spatially is challenging, hence the impaired *MOTP*.

It is notable that the fused tracker has lower overall scores than the laser-only tracker, which indicates that using the latter may improve tracking performance. However, we use radar as the main sensor. This is as radar is available in both, restricted and unrestricted, visual conditions. Therefore, we do not need to observe the visual conditions and can rely on that the radar tracker builds a congruent baseline to the generated human tracks. From the perspective of radar tracking, fusing in laser information increases the *MOTA* significantly in scenes without vision restriction. However, if tracking strategies are to be adapted depending on the observed visual conditions, the tests indicate that a shift from radar-only to laser-only tracking would be beneficial instead of using the fusion approach.

As Table 7 only includes short-term tests, we define a second test case to show the long-term potential of the proposed methodology. Within, two people walk around a static robot for ten minutes, while arbitrarily choosing `Standing` or `Squatting` postures. Two or three people is the standard use-case for firefighting squads in Germany, whereas two is more common in general Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (2008). The test results are depicted in Table 8. Comparably, the method is not impacted by raising the test length. In fact, all trackers slightly increase in *MOTA*.

After showing the good performance of the proposed method in the static case, we test the system in motion, that is, robot moving at 0.5 m/s. The velocity is typical for firefighting applications, as firefighters choose similar or lower velocities. Beforehand, we conducted a video analysis on professional indoor firefighting, which resulted in mean velocities of 0.5 m/s (`Standing`) and 0.3 m/s (`Squatting`). Due to limited data samples of moving firefighters in smoke, the values are nonsignificant, but may be used as a rough estimate.

TABLE 6 Confusion matrix of radar classifier (absolute and relative distributions).

GT \ Detected	`Standing`		`Squatting`		`Background`	
`Standing`	1490	86.63%	190	11.05%	40	2.33%
`Squatting`	544	47.64%	597	52.28%	1	0.09%
`Background`	168	9.26%	0	0.00%	1646	90.74%

Abbreviation: GT, ground truth.

TABLE 7 Tracker performances with one and many people present.

Sensor	People	Frames	*FP*	*MIS*	*IDS*	*MOTA* (%)	*MOTP* (m)	Runtime (ms)
Laser	2	1520	0	147	4	94.37	0.07	24.45
	5	1521	5	659	12	88.51	0.07	23.20
Radar	2	598	40	100	8	82.99	0.06	6.27
	5	598	70	263	23	79.94	0.05	9.07
Fusion	2	2118	27	283	5	91.14	0.16	11.04
	5	2119	22	1006	13	86.40	0.16	12.79

Note: Data captured in *Buildings*. Tests cover similar durations. Runtime refers to the average of the individual system.

Abbreviations: FP, False Positive; IDS, number of ID switches; MIS, number of missed detections; MOTA, Multiple Object Tracking Accuracy; MOTP, Multiple Object Tracking Precision.

As our use-case is to follow single targets, we design the final test such that the robot follows a single person (manually controlled to mitigate input errors), while others pass by arbitrarily. Only the target person may choose the `Squatting` posture. Table 9 shows the results in environments *Buildings* and *Road*. While moving, the scores are impaired compared with the static case. This is due to the altered management logic. In addition, pedestrians walking by are only briefly in the field of view of the sensors. Therefore, they may be completely missed due to restricted track initialization criteria, which is depicted in the high number of *MIS*. However, in both tracks, the target person is tracked continuously without impairment; the major impairments come from other pedestrians. The resulting measures are on par with the state of the art (see Table 10) in the moving robot case, and exceeding the state of the art in the static case. Thus, we conclude that the tracking system is suitable for the defined application.

TABLE 8 Long-term performance with two people in environment *Buildings*.

Sensor	Frames	*FP*	*MIS*	*IDS*	*MOTA* (%)	*MOTP* (m)
Laser	7615	6	677	15	94.62	0.07
Radar	2984	202	408	42	84.33	0.05
Fusion	10,599	105	1176	14	92.44	0.13

Abbreviations: FP, False Positive; IDS, number of ID switches; MIS, number of missed detections; MOTA, Multiple Object Tracking Accuracy; MOTP, Multiple Object Tracking Precision.

8.3 | Discussion

As presented in the prior sections, the overall tracker performance is good compared with the state of the art. Including a `Squatting`

TABLE 9 Performance with moving robot.

Sensor	Environment	Frames	*FP*	*MIS*	*IDS*	*MOTA* (%)	*MOTP* (m)
Laser	*Buildings*	3043	560	141	11	87.65	0.06
	Road	3044	628	1164	11	71.17	0.05
Radar	*Buildings*	1198	184	286	12	77.07	0.06
	Road	1198	327	478	7	53.99	0.05
Fusion	*Buildings*	4241	658	490	2	85.38	0.12
	Road	4242	649	1817	13	69.08	0.13

Abbreviations: FP, False Positive; IDS, number of ID switches; MIS, number of missed detections; MOTA, MultiObject Tracking Accuracy; MOTP, MultiObject Tracking Precision.

TABLE 10 *MOTA* and *MOTP* measures of state-of-the-art one-class detectors as indicated by their respective publications.

Source	Sensor type(s)	Classifier	*MOTA* (%)	*MOTP* (m)
Efstathiou et al. (2021)	Laser	Deep learning	70.81*	0.032**
Mandischer et al. (2021)	Radar	Ensemble	81.19	n.a.
Castanheira et al. (2020)	Radar	Ensemble	97.80*,***	n.a.
Kohara and Nakazawa (2019)	Laser	Deep learning	90.74*	n.a.
Majer et al. (2019)	Laser, radar	Ensemble	n.a.	0.106
Zhao et al. (2019)	Radar	Ensemble	n.a.	0.160
Álvarez-Aparicio et al. (2019)	Laser	Ensemble	n.a.	0.390**,***
Guerrero-Higueras et al. (2019)	Laser	Ensemble	n.a.	0.180**,***
Linder et al. (2016)	Laser, RGB-D	Ensemble	89.40	n.a.
Leigh et al. (2015)	Laser	Ensemble	33.20	0.160
Linder et al. (2015)	Laser, RGB-D	Ensemble	82.20	n.a.
Bartsch et al. (2012)	Radar	Heuristic	95.30*	n.a.
Bellotto and Hu (2010)	Laser	Ensemble	n.a.	0.138**

Abbreviations: MOTA, Multiple Object Tracking Accuracy; MOTP, Multiple Object Tracking Precision; n.a., not available.

*Accuracy instead of MOTA.

**Mean error instead of MOTP.

***Best indicated score.

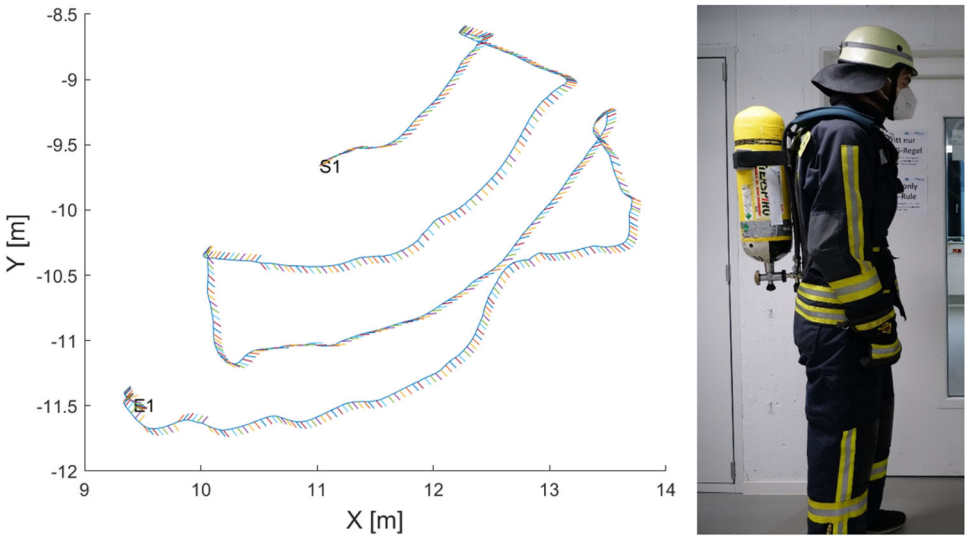

FIGURE 12 Tracked path (left; including orientations; from start "S1" to goal position "E1") of *Tauchertechnik* (Aschenbrenner, 1999) (freely translated: "Diving Search") with firefighter's clothing (right). Tracked entirely in `Squatting` posture, in a new environment, and with a person not included in any data set wearing full personal protective equipment.

posture improves the tracking performance, in particular in the radar case when discriminating people from `Background`. Further, a fusion of laser and radar on the track level can compensate for the downsides of the radar subsystem and will improve the tracking accuracy in visually unrestricted environments. However, tests have shown that the system performs better with static sensors. We assume that more tweaks in the initialization logic will make the system more robust in the general use-case. However, this is not what we desire. The use-case for the system is firefighting in which the robot follows a single target person with one or two other people present and usually occluded by the target person due to the formation of lines. Within, long-term feasibility is of particular interest. When the leg track is lost during a limited vision scene, in which the vision sensors tasked with specific person identification are impaired, and in which the system fully relies on the leg track, the re-detection of the target person may no longer be possible; the system consequently fails. The long-term test supports a good usability in such applications. Further, in this scenario it is more important to establish a robust track than to initialize tracks faster (which would be the more general case). In prior work (Mandischer et al., 2019), we have indicated that radar has no impairments by smoke or dust (see also Fang et al., 2022; Zhang et al., 2022), therefore, we assume that the results are similar to the firefighting use-case. However, future tests will need to support this assumption.

On the contrary, the data—while well suited for the general case—only depict a limited number of participants and clothing. In contrast to our latest work (Mandischer et al., 2021), no female pedestrians are covered, which results in a lack of diversity. Further, only casual clothing is included. The first tests (see Figure 12) show that the tracker trained with the data set also works well with firefighters' clothing, but further evaluation is pending. However, the presented results and the outlook towards firefighting supports the general feasibility in the use-case.

9 | CONCLUSION

In this article, we proposed a novel pipeline for multiposture leg tracking incorporating `Standing` and `Squatting` postures. We first introduced a definition of the postures and showed how to identify leg candidates and track them using KFs. We proposed two novel tracker pipelines for laser and radar sensors, including new feature descriptors and track management logic. Both trackers generate leg candidates using NNC. The laser tracker is based on a combination of GNN data association and a KF with a constant velocity model. The radar tracker is based on Otsu Filtering (Mandischer et al., 2019), GrNN data association, and the SHAKF. Both, trackers are fused on the track level using the Bayesian fusion framework proposed by Bellotto et al. (2015). We evaluated the system profoundly. First, we defined a new data set with 40,000 data samples (Mandischer & Hou, 2023), which is used for classifier training and testing. We evaluated the performance of the classifier and tracker individually, showing good results in all test cases. Finally, we discussed the results and gave an outlook towards real firefighting applications.

ACKNOWLEDGMENTS

We like to thank Marius Gürtler, Sebastian Döbler, Steffen Kaufmann, and Till Waldermann for their support in validation and testing, Maximilian Hilger for his support in implementation, validation and testing, and Vincent Brünjes for support in proof-reading the article. We further like to thank indurad for providing the sensors, and Henkel and Feuerwehrschule Würzburg, particularly Jürgen Schemmel, for providing firefighters' clothing and other materials. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Multi-Posture Leg Tracker Data Set at https://zenodo.org/record/7665007, reference number https://doi.org/10.5281/zenodo.7665006.

ORCID

Nils Mandischer http://orcid.org/0000-0003-1926-4359

REFERENCES

Ahtiainen, J., Terho, S. & Koponen, S. (2010) Radar based detection and tracking of a walking human. *IFAC Proceedings Volumes*, 43(16), 437–442. Available from: https://doi.org/10.3182/20100906-3-IT-2019.00076

Álvarez-Aparicio, C., Guerrero-Higueras, Á.M., Rodríguez-Lera, F.J., Clavero, J.G., Rico, F.M. & Matellán, V. (2019) People detection and tracking using LIDAR sensors. *Robotics*, 8(3), 75. Available from: https://doi.org/10.3390/robotics8030075

Arras, K.O., Mozos, O.M. & Burgard, W. (2007) Using boosted features for the detection of people in 2D range data. In: Hutchinson, S. (Ed.) *IEEE International Conference on Robotics and Automation, 2007*. Piscataway, NJ: IEEE Service Center, pp. 3402–3407. Available from: https://doi.org/10.1109/ROBOT.2007.363998

Aschenbrenner, D. (1999) Suchen und retten. *Brandschutz: deutsche Feuerwehr-Zeitung; Zeitschrift für das gesamte Feuerwehrwesen, für Rettungsdienst und Umweltschutz*, 53(5), 429–439.

Ausschuss Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV) (February 2008) *Feuerwehr-dienstvorschrift fwdv 3: Einheiten im lösch- und hilfeleistungseinsatz*, 2nd edition, Kohlhammer Verlag.

Bartsch, A., Fitzek, F. & Rasshofer, R.H. (2012) Pedestrian recognition using automotive radar sensors. *Advances in Radio Science*, 10, 45–55. Available from: https://doi.org/10.5194/ars-10-45-2012

Bellotto, N., Dondrup, C. & Hanheide, M. (2015) *Bayestracking: The Bayes tracking library v1.0.5*. Zenodo. Geneva, Switzerland: CERN Data Center. Available from: https://doi.org/10.5281/zenodo.15825

Bellotto, N. & Hu, H. (2010) Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of Bayesian filters. *Autonomous Robots*, 28(4), 425–438. Available from: https://doi.org/10.1007/s10514-009-9167-2

Bernardin, K. & Stiefelhagen, R. (2008) Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008, 1–10. Retrieved from: https://jivp-eurasipjournals.springeropen.com/articles/10.1155/2008/246309

Beyer, L., Hermans, A., Linder, T., Arras, K.O. & Leibe, B. (2018) Deep person detection in two-dimensional range data. *IEEE Robotics and Automation Letters*, 3(3), 2726–2733. Available from: https://doi.org/10.1109/LRA.2018.2835510

Breiman, L. (2001) Random forests. *Machine Learning*, 45(1), 5–32. Available from: https://doi.org/10.1023/A:1010933404324

Campello, R.J.G.B., Moulavi, D. & Sander, J. (2013) Density-based clustering based on hierarchical density estimates. In: Hutchison, D., et al. (Eds.) *Advances in knowledge discovery and data mining*. vol. 7819, Berlin: Springer, pp. 160–172. Available from: https://doi.org/10.1007/978-3-642-37456-2_14

Castanheira, J., Curado, F., Pedrosa, E., Gonçalves, E. & Tomé, A. (2020) Machine learning methods for radar-based people detection and tracking by mobile robots. In: Silva, M.F., LuísLima, J., Reis, L.P., Sanfeliu, A. & Tardioli, D. (Eds.) *Robot 2019*. vol. 1093, Cham: Springer International Publishing AG, pp. 379–391. Available from: https://doi.org/10.1007/978-3-030-36150-1_31

Charles, R.Q., Su, H., Kaichun, M. & Guibas, L.J. (2017) PointNet: deep learning on point sets for 3D classification and segmentation. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA: IEEE, pp. 77–85. Available from: https://doi.org/10.1109/CVPR.2017.16

Chavez-Garcia, R.O. & Aycard, O. (2016) Multiple sensor fusion and classification for moving object detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 17(2), 525–534. Available from: https://doi.org/10.1109/TITS.2015.2479925

Chavez-Garcia, R.O., Vu, T.-D. & Aycard, O. (2014) Fusion at detection level for frontal object perception. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA: IEEE, pp. 1225–1230. Available from:Available from: https://doi.org/10.1109/IVS.2014.6856555

Chung, W., Kim, H., Yoo, Y., Moon, C.-B. & Park, J. (2012) The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder. *IEEE Transactions on Industrial Electronics*, 59(8), 3156–3166. Available from: https://doi.org/10.1109/TIE.2011.2170389

Cui, J., Zha, H., Zhao, H. & Shibasaki, R. (2008) Multi-modal tracking of people using laser scanners and video camera. *Image and Vision Computing*, 26(2), 240–252. Available from: https://doi.org/10.1016/j.imavis.2007.05.005

Dietterich, T.G. (2002) Ensemble learning. In: Arbib, M.A. (Ed.) *The handbook of brain theory and neural networks*. Cambridge, MA, USA: The MIT Press, pp. 1–9.

Efstathiou, D., Chalvatzaki, G., Dometios, A., Spiliopoulos, D. & Tzafestas, C.S. (2021) Deep leg tracking by detection and gait analysis in 2D range data for intelligent robotic assistants. In: Asfour, T. (Ed.) *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Piscataway, NJ: IEEE, pp. 2657–2662. Available from: https://doi.org/10.1109/IROS51168.2021.9636588

Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis, E., Han, J. & Fayyad, U. (Eds.) *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Washington, DC, USA: AAAI Publications, pp. 226–231.

Fang, X., Li, J., Zhang, Z. & Xiao, G. (2022) FMCW-MIMO radar-based pedestrian trajectory tracking under low-observable environments. *IEEE Sensors Journal*, 22(20), 19675–19687. Available from: https://doi.org/10.1109/JSEN.2022.3203154

Fod, A., Howard, A. & Mataric, M. (2002) A laser-based people tracker. In: *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*. vol. 3, Piscataway, NJ, USA: IEEE, pp. 3024–3029.

Fotiadis, E.P., Garzón, M. & Barrientos, A. (2013) Human detection from a mobile robot using fusion of laser and vision information. *Sensors*, 13(9), 11603–11635. Available from: https://doi.org/10.3390/s130911603

Fritsch, J., Kleinehagenbrock, M., Lang, S., Plötz, T., Fink, G.A. & Sagerer, G. (2003) Multi-modal anchoring for human–robot interaction. *Robotics and Autonomous Systems*, 43(2–3), 133–147. Available from: https://doi.org/10.1016/S0921-8890(02)00355-X

Guerrero-Higueras, Á.M., Álvarez-Aparicio, C., Calvo Olivera, M.C., Rodríguez-Lera, F.J., Fernández-Llamas, C., Rico, F.M. & Matellán, V. (2019) Tracking people in a mobile robot from 2D LIDAR scans using full convolutional neural networks for security in cluttered environments. *Frontiers in Neurorobotics*, 12, 85. Available from: https://doi.org/10.3389/fnbot.2018.00085

Hackett, J.K. & Shah, M. (1990) Multi-sensor fusion: a perspective. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. vol. 2, Los Alamitos, CA: IEEE Computer Society Press, pp. 1324–1330. Available from: https://doi.org/10.1109/ROBOT.1990.126184

Held, P., Steinhauser, D., Koch, A., Brandmeier, T. & Schwarz, U.T. (2022) A novel approach for model-based pedestrian tracking using automotive radar. *IEEE Transactions on Intelligent Transportation*

Systems, 23(7), 7082–7095. Available from: https://doi.org/10.1109/TITS.2021.3066680

Jia, D., Hermans, A. & Leibe, B. (2020) DR-SPAAM: a spatial-attention and auto-regressive model for person detection in 2D range data. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Piscataway, NJ: IEEE, pp. 10270–10277. Available from: https://doi.org/10.1109/IROS45743.2020.9341689

Jia, D., Steinweg, M., Hermans, A. & Leibe, B. (2021) Self-supervised person detection in 2D range data using a calibrated camera. In: *2021 IEEE International Conference on Robotics and Automation*. Piscataway, NJ: IEEE, pp. 13301–13307. Available from: https://doi.org/10.1109/ICRA48506.2021.9561699

Kluge, B., Kohler, C. & Prassler, E. (2001) Fast and robust tracking of multiple moving objects with a laser range finder. In: *Proceedings of the 2001 ICRA. IEEE International Conference on Robotics and Automation*. vol. 2, Piscataway, NJ, USA: IEEE, pp. 1683–1688. Available from: https://doi.org/10.1109/ROBOT.2001.932853

Kohara, Y. & Nakazawa, M. (2019) Human tracking of single laser range finder using features extracted by deep learning. In: *2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, Kathmandu, Nepal: IEEE. Available from: https://doi.org/10.23919/ICMU48249.2019.9006630

Kolmogorov, A.N. (1956) *Foundations of the theory of probability*, 2nd edition, New York: Chelsea Publishing Company.

Konstantinova, P., Udvarev, A. & Semerdjiev, T. (2003) A study of a target tracking algorithm using global nearest neighbor approach. In: Rachev, B. & Smrikarov, A. (Eds.) *Proceedings of the 4th International Conference on Computer Systems and Technologies e-Learning*. New York, NY: ACM, pp. 290–295. Available from: https://doi.org/10.1145/973620.973668

Kruijff-Korbayova, I., Freda, L., Gianni, M., Ntouskos, V., Hlavac, V., Kubelka, V., Zimmermann, E., Surmann, H., Dulic, K., Rottner, W. & Gissi, E. (2016) Deployment of ground and aerial robots in earthquake-struck amatrice in italy (brief report). In: *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Piscataway, NJ: IEEE, pp. 278–279. Available from: https://doi.org/10.1109/SSRR.2016.7784314

Kuhn, H.W. (1955) The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2), 83–97. Available from: https://doi.org/10.1002/nav.3800020109

Leigh, A., Pineau, J., Olmedo, N. & Zhang, H. (2015) Person tracking and following with 2D laser scanners. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ: IEEE, pp. 726–733. Available from: https://doi.org/10.1109/ICRA.2015.7139259

Linder, T., Breuers, S., Leibe, B. & Arras, K.O. (2016) On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ: IEEE, pp. 5512–5519. Available from: https://doi.org/10.1109/ICRA.2016.7487766

Linder, T., Girrbach, F. & Arras, K.O. (2015) Towards a robust people tracking framework for service robots in crowded, dynamic environments. In: *Assistance and Service Robotics Workshop (ASROB-15) at IROS*.

Linder, T., Pfeiffer, K.Y., Vaskevicius, N., Schirmer, R. & Arras, K.O. (2020) Accurate detection and 3D localization of humans using a novel YOLO-based RGB-D fusion approach and synthetic training data. In: Okamura, A. (Ed.) *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ: IEEE, pp. 1000–1006. Available from: https://doi.org/10.1109/ICRA40945.2020.9196899

Linder, T., Vaskevicius, N., Schirmer, R. & Arras, K.O. (2021) Cross-modal analysis of human detection for robotics: an industrial case study. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Piscataway, NJ: IEEE, pp. 971–978. Available from: https://doi.org/10.1109/IROS51168.2021.9636158

Lindstrom, M. & Eklundh, J.-O. (2001) Detecting and tracking moving objects from a mobile platform using a laser range scanner. In: *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the Next Millennium*. Piscataway, NJ: IEEE Operations Center, pp. 1364–1369. Available from: https://doi.org/10.1109/IROS.2001.977171

Liu, P., Yao, H., Dai, H. & Fu, W. (2022) The detection and following of human legs based on feature optimized HDBSCAN for mobile robot. *Journal of Physics: Conference Series*, 2216(1), 012009. Available from: https://doi.org/10.1088/1742-6596/2216/1/012009

Majer, F., Yan, Z., Broughton, G., Ruichek, Y. & Krajnik, T. (2019) Learning to see through haze: radar-based human detection for adverse weather conditions. In: Přeučil, L., Behnke, S. & Kulich, M. (Eds.) *2019 European Conference on Mobile Robots (ECMR)*. Piscataway, NJ: IEEE, pp. 1–7. Available from: https://doi.org/10.1109/ECMR.2019.8870954

Mandischer, N., Charaf Eddine, S., Huesing, M. & Corves, B. (2019) Bots2ReC: Radar localization in low visibility indoor environments. In: *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. Piscataway, NJ: IEEE, pp. 158–163. Available from: https://doi.org/10.1109/SSRR.2019.8848981

Mandischer, N. & Hou, R. (2023) *Multi-posture leg tracker data set*. Zenodo. Geneva, Switzerland: CERN Data Center. Available from: https://doi.org/10.5281/zenodo.7665006

Mandischer, N., Koop, I., Granich, A., Heberling, D. & Corves, B. (2021) Radar tracker for human legs based on geometric and intensity features. In: *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, pp. 1521–1525. Available from: https://doi.org/10.23919/EUSIPCO54536.2021.9616134

Möbus, R. & Kolbe, U. (2004) Multi-target multi-object tracking, sensor fusion of radar and infrared. In: *2004 IEEE Intelligent Vehicles Symposium*. Piscataway, NJ: IEEE Operations Center, pp. 732–737. Available from: https://doi.org/10.1109/IVS.2004.1336475

Narayanan, R.M., Smith, S., Gallagher, K.A. & Qing, X. (2014) A multifrequency radar system for detecting humans and characterizing human activities for short-range through-wall and long-range foliage penetration applications. *International Journal of Microwave Science and Technology*, 2014, 958905. Available from: https://doi.org/10.1155/2014/958905

Otsu, N. (1979) A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. Available from: https://doi.org/10.1109/tsmc.1979.4310076

Qian, H., Yang, X., Zhang, X., Ding, Y. & Zhang, L. (2020) PLA-JPDA for indoor multi-person tracking using IR-UWB radars. In: *2020 IEEE Radar Conference (RadarConf20)*. Piscataway, NJ: IEEE. Available from: https://doi.org/10.1109/RadarConf2043947.2020.9266407

Rasshofer, R.H. & Gresser, K. (2005) Automotive radar and lidar systems for next generation driver assistance functions. *Advances in Radio Science*, 3, 205–209. Available from: https://doi.org/10.5194/ars-3-205-2005

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016) You only look once: unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA: IEEE, pp. 779–788. Available from: https://doi.org/10.1109/CVPR.2016.91

Rohling, H., Heuel, S. & Ritter, H. (2010) Pedestrian detection procedure integrated into an 24 GHz automotive radar. In: *2010 IEEE Radar Conference*, Arlington, VA, USA: IEEE, pp. 1229–1232. Available from: https://doi.org/10.1109/RADAR.2010.5494432

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M. & Frangi, A.F. (Eds.) *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. vol.

9351, Cham: Springer International Publishing, pp. 234–241. Available from: https://doi.org/10.1007/978-3-319-24574-4_28

Sage, A.P. & Husa, G.W. (1969) Adaptive filtering with unknown prior statistics. *Transactions on Automatic Control*, 7, 760–769. Available from: https://doi.org/10.1109/JACC.1969.4169325

Scheutz, M., McRaven, J. & Cserey, G. (2004) Fast, reliable, adaptive, bimodal people tracking for indoor environments. In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. vol. 2, Piscataway, NJ: IEEE, pp. 1347–1352. Available from: https://doi.org/10.1109/IROS.2004.1389583

Sharif, M.H. (2020) Particle filter for trajectories of movers from laser scanned dataset. In: Djeddi, C., Jamil, A. & Siddiqi, I. (Eds.) *Pattern recognition and artificial intelligence*. vol. 1144, Cham: Springer International Publishing AG, pp. 133–148.

Shin, D. (2021) The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *International Journal of Human–Computer Studies*, 146, 102551. Available from: https://doi.org/10.1016/j.ijhcs.2020.102551

Spinello, L. & Siegwart, R. (2008) Human detection using multimodal and multidimensional features. In: Hutchinson, S. (Ed.) *2008 IEEE International Conference on Robotics and Automation*. Piscataway, NJ: IEEE Service Center, pp. 3264–3269. Available from: https://doi.org/10.1109/ROBOT.2008.4543708

Tiemann, J., Fuhr, O. & Wietfeld, C. (2020) CELIDON: supporting first responders through 3D AOA-based UWB ad-hoc localization. In: *WiMob 2020*, Piscataway, NJ: IEEE, pp. 20–25. Available from: https://doi.org/10.1109/WiMob50308.2020.9253377

Xavier, J., Pacheco, M., Castro, D., Ruano, A. & Nunes, U. (2005) Fast line, arc/circle and leg detection from laser scan data in a player driver. In: *2005 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ: IEEE Operations Center, pp. 3930–3935. Available from: https://doi.org/10.1109/ROBOT.2005.1570721

Zhai, G., Wu, C. & Wang, Y. (2018) Millimeter wave radar target tracking based on adaptive Kalman filter. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. Piscataway, NJ: IEEE, pp. 453–458. Available from: https://doi.org/10.1109/IVS.2018.8500505

Zhang, Z., Pouliquen, P.O., Waxman, A. & Andreou, A.G. (2007) Acoustic micro-Doppler radar for human gait imaging. *The Journal of the Acoustical Society of America*, 121(3), EL110–EL113. Available from: https://doi.org/10.1121/1.2437842

Zhang, Z., Wang, X., Huang, D., Fang, X., Zhou, M. & Zhang, Y. (2022) MRPT: millimeter-wave radar-based pedestrian trajectory tracking for autonomous urban driving. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–17. Available from: https://doi.org/10.1109/TIM.2021.3139658

Zhao, P., Lu, C.X., Wang, J., Chen, C., Wang, W., Trigoni, N. & Markham, A. (2019) mID: Tracking and identifying people with millimeter wave radar. In: *15th Annual International Conference on Distributed Computing in Sensor Systems*. Piscataway, NJ: IEEE, pp. 33–40. Available from: https://doi.org/10.1109/DCOSS.2019.00028

Zhou, Y., Sun, P., Zhang, Y., Anguelov, D., Gao, J., Ouyang, T., Guo, J., Ngiam, J. & Vasudevan, V. (2019) End-to-end multi-view fusion for 3D object detection in LiDAR point clouds. In: *3rd Conference on Robot Learning (CoRL 2019)*, Osaka, Japan. Retrieved from: http://arxiv.org/pdf/1910.06528v2

How to cite this article: Mandischer, N., Hou, R. & Corves, B. (2023) Multi-posture leg tracking for temporarily vision restricted environments based on fusion of laser and radar sensor data. *Journal of Field Robotics*, 40, 1620–1638. https://doi.org/10.1002/rob.22195