

LIDAR-based people detection and tracking for @home Competitions

Álvarez-Aparicio, Claudia
Guerrero-Higueras, Ángel M.
Rodríguez-Lera, Francisco J.
Calvo Olivera, M. Carmen
Matellán Olivera, Vicente

Dept. of Mech, Computer and Aerospace Engineering.
Universidad de León.

León, Spain.

{calvaa, am.guerrero, fjrodl, mcalvo, vmato}@unileon.es

Ginés Clavero, Jonatan
Martín Rico, Francisco

GSyC Dept.

Universidad Rey Juan Carlos.

Fuenlabrada, Spain.

{jonatan, fmartin}@gsyc.es

Abstract—People tracking is a basic capability in almost any robotic application. So it is in robotic competitions, where many robot skills rely on this ability. This problem is still challenging, particularly when are implemented using low definition sensors as Laser Imaging Detection and Ranging (LIDAR) sensors in crowded environments. This paper describes a solution based on a single LIDAR sensor that uses the gait to keep a continuous identification in time and space of the individual. The system described in this article is based on PeTra (People Tracking) package, which uses convolutional neural networks to identify legs in populated environments. Experimental validation proposes a test in an apartment replicating realistic competition arena.

Index Terms—LIDAR, neural networks, people tracking

I. INTRODUCTION

During years, RoboCup@Home has been benchmarking robots skills based on their performance at different tasks. During the competition days, the arena is crossed by several robots that are evaluated doing different tasks. During these evaluations, the referees give scores to different robot abilities such as mapping, navigation, or person tracking which are associated with specific functionalities [1], [2]. Among these abilities, person detection and tracking have been considered a cornerstone since the competition beginning.

This paper proposes a solution for tracking individuals during its interaction with the robot in @home environments using a laser beam. The laser is mainly used in competitions for carrying out navigation process and accomplish SLAM tasks. Usually, mobile robots mount laser scanners in a low position (~ 30 -50 cm) from the ground) to detect dynamic obstacles. They are also used to construct occupancy maps and navigate. The information provided allows estimating the distance at precise angles (resolution of 0.5 degrees). This paper proposes to use the same information acquired by the

navigation sensors for inferring and tracking human position along the time.

Detecting and tracking people is a mandatory skill in @home competition given its importance in human-robot interaction tasks [4]. There are three main issues associated to this skill: proxemics, the human position with respect to the robot position; data acquisition, which sensor is used for perceiving the individual; and data classification, which is the classifier that infers a human from a set of features.

Proxemics, the usual method for tracking a human is maintaining them in the robot front [5], the usual place for robot sensors. This method has as a consequence to decide which part of the human is used for tracking. Usually, it has been selected between head, torso, and legs [6], [7]. Jung et al [8] propose a solution based on torso track. Hirai [9] proposes back of the torso and shoulder, in this case, it is proposed a system based on leg detection. Many researchers propose this approach. Volkhardt [10] proposes a brute-force-trained weak classifiers. Belloto et al. were quite successful in their approach [11] and it was decided to proposed a solution based on their approach. The current RoboCup@home scenarios present cluttered scenarios with objects such as table or chair legs, trunks of plants, etc., may be easily confused with the legs of persons. It is also difficult to keep track of a particular person (i.e., a pair of legs) in a crowded environment because many obstructions can result. Therefore, it is possible to achieve better results given current sensors mounted onboard the robots.

Data acquisition deals with the sensor involved. The most extended approach is to deploy multi-sensor solutions. Vision and range sensors are the most extended approaches as shown in [12]. Cameras are prone to gather noisy information, in particular in crowded environments, so range sensors were chosen in this proposal. In particular, LIDAR sensors were used because they are reliable and affordable, both economical and computationally speaking. These sensors provide information about the environment at good frame rates (~ 20 -30 Hz)

This work has been partially funded by Junta de Castilla y León grant LE-028P17, Comunidad de Madrid grant RoboCity2030-Fase 3.

and are easy to process on real-time because each scan consists of an array of only a few 100 integers.

Data classification presents the logical inference system for tracking a human. It is necessary to apply an inference and classifier solution for processing sensor data. Convolutional Neural Networks (CNNs) as presented by Long et al. [13]. Since the effectiveness of CNN is shown in recognizing individuals in different scenarios, our PeTra (Pople Tracking) package includes an algorithm based on it.

The main contribution of this paper is the matching between the laser beam obtained by the robot sensor and the individuals presented in the robot scene. This matching is performed almost in real time and given the nature of the laser, reducing the computational cost associated with similar solutions based on cameras and depth-cameras.

The rest of the paper is organized as follows: Section II presents the system design and development model as well as the experimental description. Section III reports a set of experimental results obtained after our experimental proposal to validate our approach. Section IV presents a discussion of the experimental approach, the measurements, and the final results. Section V presents conclusions and future work associated with this study.

II. MATERIALS AND METHODS

A set of experiments was carried out to evaluate PeTra accuracy. They were conducted in the indoor mock-up apartment at Leon@Home Testbed, a Certified Testbed of the European Robotics league (ERL) located in the Robotics Lab at University of León. Its main purpose is to benchmark service robots in a realistic home environment. Figure 1 shows the apartment plan.

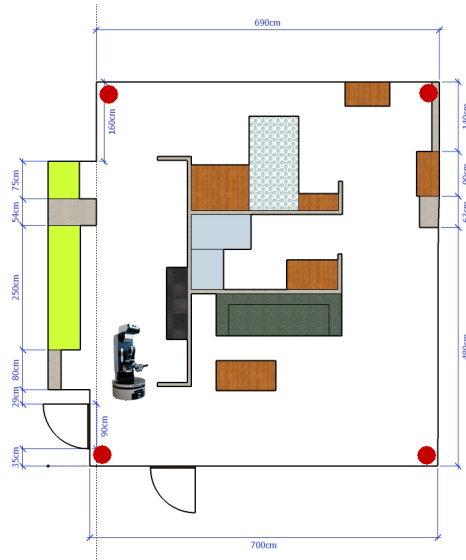


Fig. 1. Mock-up apartment at Leon@Home Testbed

A. Orbi-One robot

Orbi-One, figure 1 (center), is a service robot manufactured by Robotnik. The robot has different sensors, such as an RGB-D camera located in the head or a LIDAR sensor located in its mobile base. Also has a manipulator arm attached to its torso that allows it to interact with its surroundings. Inside, an Intel Core i7 CPU with 8 GB of RAM allows it to run the software based on a ROS framework to control the robot hardware. The LIDAR sensor at its base has been used to develop and test PeTra.

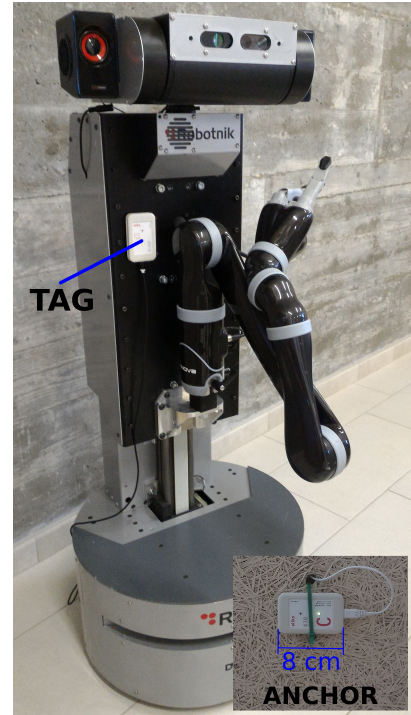


Fig. 2. Orbi-One and KIO RTLS as shown at [14].

B. KIO RTLS

A commercial RTLS (KIO) has been used to provide ground-truth data. KIO calculates the position of a mobile transceiver, called a *tag*, in a two or three-dimensional space. In order to do so, KIO uses radio beacons, called *anchors*, that have been previously located in known positions in the surroundings. Red markers in Figure 1 show the position of the anchors used in the experiments described in the paper. They are attached at the ceiling of the mock-up apartment. The distribution of the anchors has been chosen following the method shown at [15]. Fig 2 show a KIO tag on board of Orbi-One robot and a KIO anchor attached at the ceiling.

KIO tags were carried by the people to be tracked in our experiments. Location estimates provided by KIO have an average error of ± 30 cm according to the manufacturer's specifications. Calibration done by the authors of this paper shows that the error is higher in some areas, and lower in others, but on average, the claims of the manufacturer are correct, see [15].

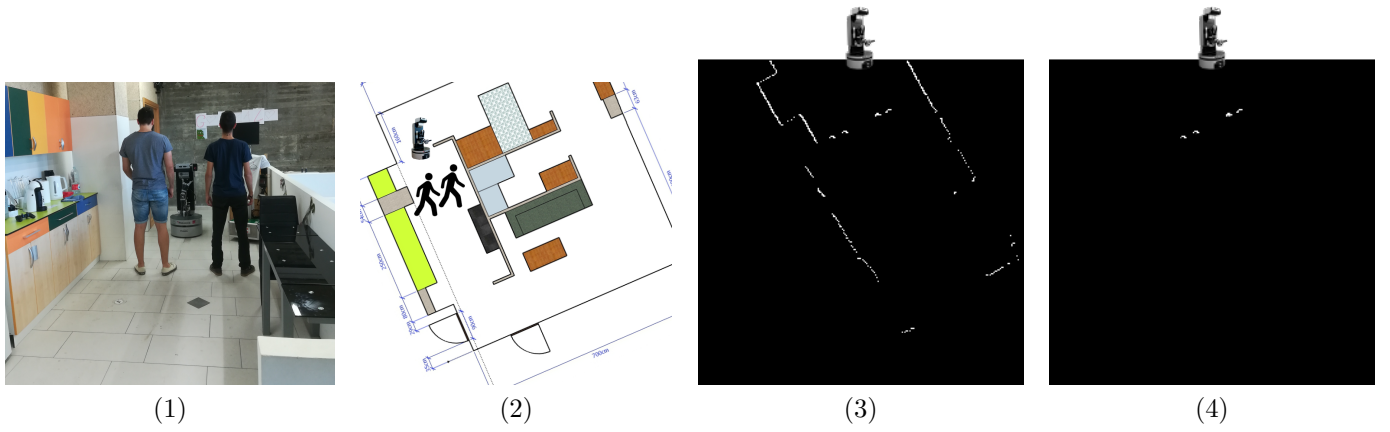


Fig. 3. PeTra operation steps as shown at [16]: Orbi-One and two people at the kitchen of the mock-up apartment (1), same situation over the apartment plan (2), occupancy map built from LIDAR's readings on the above situation (3), and PeTra's output after processing the above occupancy map (4).

C. PeTra

PeTra is a tool for detecting and tracking people developed by the Robotics Group at the University of León. The system is based on a CNN [17] which uses an occupancy map built from LIDAR measurements as input. We explored different alternatives and we have found that the configuration proposed in the U-Net architecture by [18] is the one that better matches our requirements. U-Net architecture was developed for Biomedical Image Segmentation. It consists of a contracting path to capture context and a symmetric expanding path that enables precise localization that we need to look for a specific pattern (pair of legs) in the occupancy map.

PeTra has been developed and later evaluated on a mobile robot based on the Robot Operating System (ROS) framework [16]. It has been designed as a ROS node with the CNN embedded on it. For PeTra's evaluation a dataset known as *Range-based people tracker classifiers Benchmark* (RRID:SCR_01574) was compiled. Further information about the dataset may be found at [19].

Once the network is trained, PeTra is capable to perform the following steps in real time:

- 1) First, the data provided by the LIDAR sensor is processed to build a two-dimensional occupancy map in front of the robot. Figure 3–1 shows Orbi-One and two people at the study area. Figure 3–2 shows the same situation over the apartment plan. Figure 3–3 illustrates the occupancy map obtained from the LIDAR's readings on the above situation. Occupancy map is presented as a picture where white pixels denote positions where the LIDAR scan found an obstacle, and black pixels denote positions where either the LIDAR scan went through without detecting any obstacle or did not go through that position.
- 2) Then, the occupancy map of the previous step is given to the network as input data. The network produces a second occupancy map representing the zones where legs have been detected. Figure 3–4 shows the network

output after processing the occupancy map shown in Figure 3–3.

- 3) Finally, a mass center calculation returns people location which is published as a ROS location message in a specific ROS topic.

In the Youtube channel of the Robotics Group of the University of León readers can find a video demonstration¹ of PeTra working as displayed on Rviz, a well-known tool to display robot information for the ROS framework.

For this work, a new PeTra release has been deployed. The system has been changed in order to correlate people location estimates along time in order to get trajectories. In a first release, a simple approach has been used to correlate location estimates. Each estimate in a specific instant (t) is paired with the closest estimate in $t + \Delta t$. The closest estimate is chosen by computing the Euclidean distance.

D. Evaluation method

In order to evaluate PeTra correlation, location estimates of people can be compared to the ground-truth data provided by KIO RTLS. The accuracy error of both PeTra (e_{PeTra}) in a specific instant of time has been calculated as the Euclidean distance between their location estimates (l_{PeTra} and ground-truth data provided by KIO-RTLS (l_{KIO}). The comparison may be done just in case there is a valid location estimate from PeTra, otherwise e_{PeTra} will get maximum value. Equation 1 show e_{PeTra} calculation.

$$e_{PeTra} = \begin{cases} \text{Max. range value} & \text{if } \nexists l_{PeTra} \\ d(l_{PeTra}, l_{KIO}) & \text{if } \exists l_{PeTra} \end{cases} \quad (1)$$

Equation 2 shows $d(l_{PeTra}, l_{KIO})$ calculation,

$$d(l_{PeTra}, l_{KIO}) = \sqrt{\sum_{i=1}^n (l_{PeTra_i} - l_{KIO_i})^2} \quad (2)$$

¹<https://youtu.be/Qa6eEJzUIRg>

where n is the number of dimensions considered. In our experiments, only X and Y coordinates are considered, so $n = 2$. The Z coordinate is constant since a mobile robot moves on the ground.

To evaluate PeTra correlation, new data were gathered. In order to do so, Orbi-One stood still while one or two people, carrying a KIO tag, moved around it under three different scenarios:

- 1) In the first scenario, just one person moved in front of the robot.
- 2) In the second one, two people moved in front of the robot.
- 3) Finally, in the third scenario, two people moved in front of the robot, but in this case, their trajectories were crossed.

These scenarios have been chosen according to different situations that may occur in robotics competitions such as ERL or RoboCup. A rosbag file² was created for each scenario. Figure 4 shows PeTra working on scenario 2 (with two people in front of the robot) as displayed on Rviz. Yellow markers show LIDAR's readings. The red arrow shows the Orbi-One's location and orientation. The beginning of the arrow matches the robot location. PeTra's location estimates are shown on the center: red markers show legs location estimates, and the green letters between red markers show the people center. Yellow and Pink big markers show PeTra location estimates.

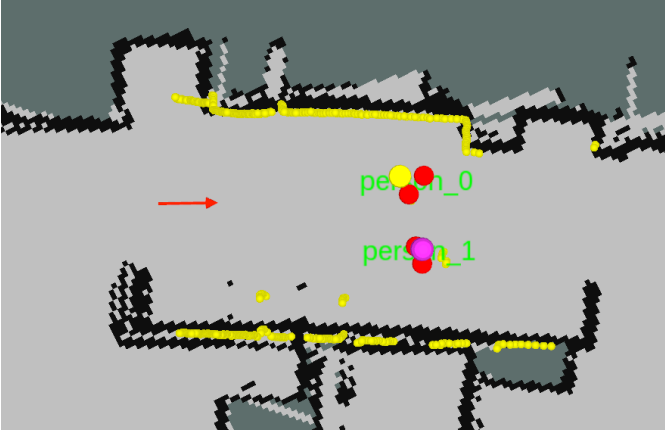


Fig. 4. Information of the robot as displayed on Rviz for scenario 2.

Figure 5 shows the onboard camera data recorded by Orbi-One in the scenario shown in Figure 4.

III. RESULTS

The accuracy of PeTra correlation has been evaluated applying the described method. Table I shows the mean error and standard deviation for the 3 scenarios explained above. As can be seen in the table, the average error of e_{PeTra} on scenario 1 is 0.41m. The standard deviation of e_{PeTra} is 0.22m in this case. Regarding scenario 2, the average error is 0.41m for the

²A rosbag file is equivalent to a recording of the state of the robot in a period and can be used as a dataset.



Fig. 5. On-board camera data recorded by Orbi-One in the scenario shown in Figure 4

first person in the scene, and 0.45m for the second one. The standard deviations are 0.15m and 0.30m respectively. Finally, on scenario 3, the average error is 0.95m for the first person, and 1.04m for the second one, while the standard deviations are 0.66m and 0.91m respectively.

TABLE I
 e_{PeTra} MEAN ERROR AND STANDARD DEVIATION (m)

Situation	Person	$X_{e_{PeTra}}$
1	1	0.41 (± 0.22)
2	1	0.41 (± 0.15)
	2	0.45 (± 0.30)
3	1	0.95 (± 0.66)
	2	1.04 (± 0.91)

Figure 6 shows the evolution of the accuracy error of PeTra over the time horizon given by rosbag files recorded on scenarios 1, 2, and 3. Red markers represent e_{PeTra} for the first person in the scene. Green markers represent e_{PeTra} for the second person. Red and green lines illustrate e_{PeTra} Moving Average (MA) for both people. MAs were used to smooth out short-term fluctuations and highlight longer-term trends or cycles.

IV. DISCUSSION

Table I shows the mean error and standard deviation for scenarios 1, 2, and 3. As expected, PeTra mean error (e_{PeTra}) has a low value (0.41m³) on scenario 1. This is could be considered a double-check attending results presented at [16]. For scenario 2, results are also consistent with the results presented at [16]. For both people, e_{PeTra} have low values (0.41m and 0.45m, respectively). However, on scenario 3, e_{PeTra} has a very high value for both people (0.95m and 1.04m, respectively). This is due to the fact, that when people trajectories cross, the pairing method does not work properly.

Figure 6 shows the evolution of e_{PeTra} over the time horizon given by rosbag files recorded on scenarios 1, 2, and

³It is important to point out that location estimates provided by KIO have an average error of ± 30 cm as explained at section II-B

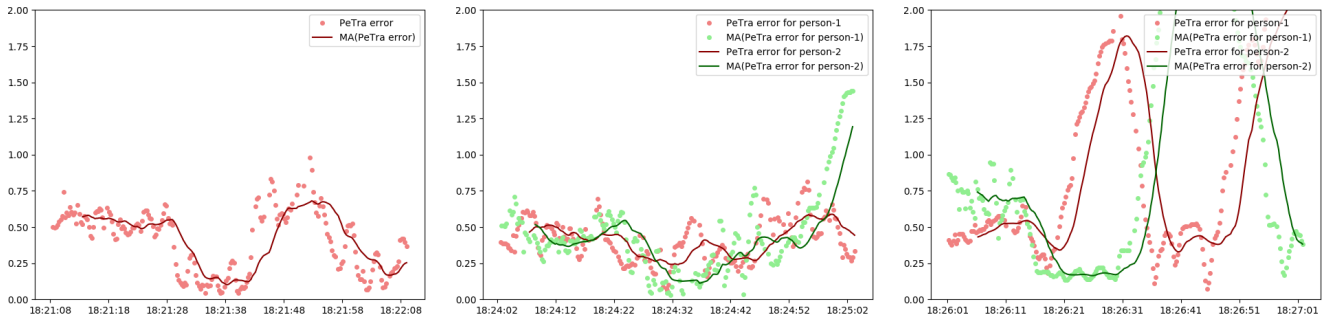


Fig. 6. e_{PeTra} (m) evolution over time (hh:mm:ss) on scenarios 1 (left), 2 (center), and 3 (right). Red markers show e_{PeTra} for the first person in the scene, green markers show e_{PeTra} for the second one. Red and green lines show $MA(e_{PeTra})$ for both people.

3. On scenario 1, we can observe that e_{PeTra} values remains low all the time. On scenario 2, e_{PeTra} values also remain low even for the two people in the scene. On scenario 3, we observe big fluctuations on e_{PeTra} corresponding with the moments where people trajectories cross.

V. CONCLUSIONS AND FURTHER WORK

This paper presented a system to correlate pair of legs (people) based on the PeTra tool. This could be used for different applications, such as improving navigation, facilitate human-robot interaction, or security and safety applications. To evaluate the performance an experiment has been carried out involving several people on different scenarios. As a result, experiments probed that PeTra correlation offers good accuracy in scenarios where people trajectories do not cross.

The main contribution of this work is the PeTra correlation system itself which allows tracking people by processing LIDAR range data in real-time. But in addition, we also want to point out the technical contributions of this work, which include a people tracking system ready to be used by any mobile robot using the ROS framework and a dataset to evaluate it.

Even though the results presented here are promising, there are still improvements for future work. One potential problem with the PeTra algorithm proposed here is the performance under crowded scenarios with people crossing along the time. The inclusion of parallel techniques as Kalman filters, optical flow or deformation fields will have an effect on accuracy, however, the inclusion of these computational costly techniques requires a depth analysis of their impact in the full performance during the competitions.

ACKNOWLEDGMENT

The research described in this article has been partially funded by Junta de Castilla y León, under the grant LE028P17; and by Comunidad de Madrid, under the grant RoboCity2030-Fase 3.

Authors would like to thank A. Campazas their support during the data gathering.

REFERENCES

- [1] D. Holz, L. Iocchi, and T. Van Der Zant, "Benchmarking intelligent service robots through scientific competitions: the robocup@ home approach," in *AAAI Spring Symposium: Designing Intelligent Robots*, 2013, pp. 27–32.
- [2] T. Wisspeintner, T. Van Der Zant, L. Iocchi, and S. Schiffer, "Robocup@ home: Scientific competition and benchmarking for domestic service robots," *Interaction Studies*, vol. 10, no. 3, pp. 392–426, 2009.
- [3] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast yolo: A fast you only look once system for real-time embedded object detection in video," *arXiv preprint arXiv:1709.05943*, 2017.
- [4] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.
- [5] J. Kessler, D. Iser, and H.-M. Gross, "I'll keep you in sight: Finding a good position to observe a person," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 4392–4398.
- [6] O. M. Mozos, R. Kurazume, and T. Hasegawa, "Multi-part people detection using 2d range data," *International Journal of Social Robotics*, vol. 2, no. 1, pp. 31–40, 2010.
- [7] M. Asghari Oskoei, M. Walters, and K. Dautenhahn, "An autonomous proxemic system for a mobile companion robot," in *Proceedings of the AISB 2010 Symposium on New Frontiers for Human Robot Interaction*. AISB, 2010.
- [8] E.-J. Jung, B.-J. Yi *et al.*, "Control algorithms for a mobile robot tracking a human in front," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2411–2416.
- [9] N. Hirai and H. Mizoguchi, "Visual tracking of human back and shoulder for person following robot," in *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*, vol. 1, July 2003, pp. 527–532 vol.1.
- [10] M. Volkhardt, C. Weinrich, and H.-M. Gross, "People tracking on a mobile companion robot," in *SMC*. Citeseer, 2013, pp. 4354–4359.
- [11] N. Bellotto and H. Hu, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 167–181, 2009.
- [12] K. O. Arras, B. Lau, S. Grzonka, M. Luber, O. M. Mozos, D. Meyer-Delius, and W. Burgard, "Range-based people detection and tracking for socially enabled service robots," in *Towards Service Robots for Everyday Environments*. Springer, 2012, pp. 235–280.
- [13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [14] Á. M. Guerrero-Higuera, N. DeCastro-García, and V. Matellán, "Detection of cyber-attacks to indoor real time localization systems for autonomous robots," *Robotics and Autonomous Systems*, vol. 99, pp. 75–83, 2018.
- [15] Á. M. Guerrero-Higuera, N. DeCastro-García, F. J. Rodríguez-Lera, and V. Matellán, "Empirical analysis of cyber-attacks to an indoor real time localization system for autonomous robots," *Computers & Security*, vol. 70, pp. 422–435, 2017.

- [16] Á. M. Guerrero-Higueras, C. Álvarez-Aparicio, M. C. Calvo-Olivera, F. J. Rodríguez-Lera, C. Fernández-Llamas, F. Martín, and V. Matellán, "Security and safety in cluttered environments: Tracking people with a mobile robot from 2d lidar scans using full convolutional neural networks," *Frontiers in Neurorobotics*, vol. 12, p. 85, 2018.
- [17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 3431–3440.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241.
- [19] C. Álvarez-Aparicio, Á. M. Guerrero-Higueras, M. C. C. Olivera, F. J. Rodríguez-Lera, F. Martín, and V. Matellán, "Benchmark dataset for evaluation of range-based people tracker classifiers in mobile robots," *Frontiers in neurorobotics*, vol. 11, p. 72, 2018.

ACRONYMS

CNN—Convolutional Neural Network.
ERL—European Robotics league.
LIDAR—Laser Imaging Detection and Ranging.
ROS—Robot Operating System.
RTLS—Real Time Location System.
MA—Moving Average.