Detecting and Measuring Human Walking in Laser Scans*

Katerina Zamani[1], Georgios Stavrinos[1], and Stasinos Konstantopoulos[1]

***Abstract*— This paper presents work on detecting and tracking human movement in planar range data. Our method stacks multiple planar scans into a 3D frame where time serves as the third dimension. This representation simultaneously informs about the size and shape of the objects in the scene and their movement, so that no explicit motion models are necessary. The scene is then segmented into 3D spatio-temporal objects which are classified as 'pairs of walking legs' using methods from machine vision. Our main contribution is a novel preprocessing step which aligns the spatio-temporal objects, so that information about the direction and speed of movement is factored out of the representation. The advantage is that the subsequent feature extraction and classification steps are only exposed to movement patterns without reference to direction and speed, which are not relevant to recognizing human walking. The method is empirically evaluated and found to significantly increase classification accuracy.**

I. INTRODUCTION

For many robotics applications, humans are the most relevant and important element of understanding a scene: their position and movement should be taken into account when planning the platform's motion and they are the focus of interaction. Specifically in the work described here, we investigate the application of robotics in *assisted living environments* to collect medically relevant data [1]. In this context, tracking human movement is not only input for motion planning or HRI, but it is also part of the core objective of the application. In particular, tracking human movement is needed to recognize being active around the house and also to measure walking speed, which are then used as behavioural and functional indicators regarding an elderly person's ability to sustain independent living.

The advantages of using range data to detect and track movement are that reliable and accurate measurements can be made, especially in indoors applications, and that the range data (and especially the planar laser scanner data discussed here) can be unobtrusively collected by comparison to wearable motion sensors. What is also interesting is that planar range data carry, by its nature, very little information. This is an advantage in avoiding the privacy issues around collecting and analysing visual or even 3D data, but also makes it practically impossible to extract characteristic features from individual frames. As a consequence, the community has drawn its attention to detecting *moving* objects so that characteristic movement patterns can be extracted from richer object representations that span multiple frames.

This presupposes solving the *data association* problem and especially in situations such as occlusion, data sparsity, and physical proximity of objects. There are two lines of research: The first is based on *Kalman filters* that serve as *motion models* and track objects by estimating the future track position from past observations [2]. Although successful, Kalman filters face the key issue of defining the motion model. To address this, Spinello et al. [3] predefined three different motion models and in each step chose the one with the highest probability. Bennewitz et al. [4] proposed an unsupervised algorithm in which motion patterns were learnt automatically using *expectation-maximization estimation* and Hidden Markov Models. Other approaches are based on assumptions about the environment to simplify the problem. Nemati and Åstrand [5], for example, use hard limits on object size to separate moving humans from moving forklifts in an automated industrial environment and limits on maximum speed and maximum proximity of different objects to segment scan points into objects and to associate objects across scans. More recent methods fuse multiple modalities to improve robustness. Ristić-Durrant et al. [6], for example, fused range data with 3D depth data. Although successful in increasing robustness, the fusion of different modalities voids the privacy argument in favour of planar range data.

An alternative approach (and the one assumed as the basis for the work described here) is to stack multiple planar scans into a 3D *frame* where time serves as the third dimension [7]. This representation simultaneously informs about the size and shape of the objects in the scene and their movement, so that no explicit motion models are necessary. These 3D objects are then clustered based on this spatio-temporal proximity, so that 'clearer' scans inside the frame can help solve occusions and sparsity in more cluttered or distant instances of the same object's track through the frame. After segmentation, these spatio-temporal representations are treated as 3D objects and classified as 'human walking' instances using methods from machine vision (surface modelling, HOG feature extraction, and classification). The fundament assumption is the same as by Nemati and Åstrand [5] that object segmentation and object tracking through time should be based on geometric proximity. However, in the method by Varvadoukas et al. [7] these assumptions are manifested as unsupervised clustering in the Euclidean space whereas Nemati and Åstrand [5] hardwire the definition of 'human leg' and 'forklift' as limits on the width of these objects in the scan.

Although proven to be very successful in associating

*This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 643892. For more details, visit the RADIO project Web site http://www.radio-project.eu

[1]All authors are with Institute of Informatics and Telecommunications, NCSR 'Demokritos', Athens 15310, Greece {kzam,gstavrinos,konstant}@iit.demokritos.gr

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 2, 2017.

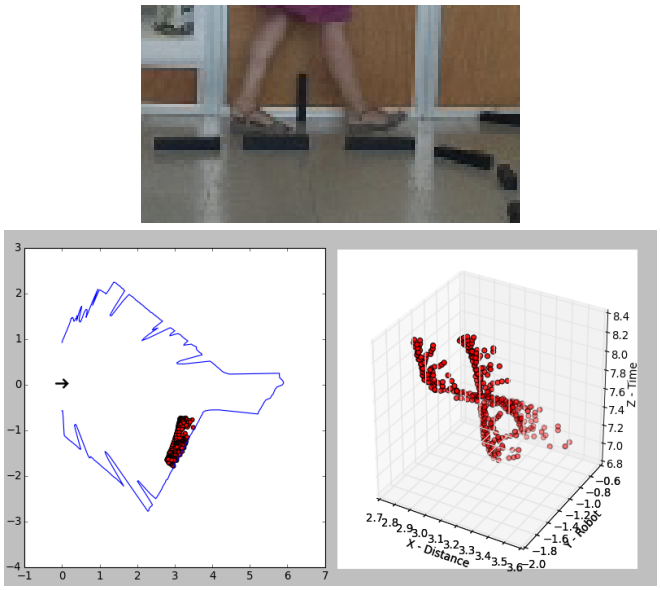


Fig. 1. The 3D spatio-temporal representation (right) and the projection on the map (left) of somebody walking across the front view of the robot. The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), as marked by the black arrow.

data across frames, this method was very sensitive to the direction of movement: different directions produce radically different surface models and, consequently, HOG features. This is due to the fact that the spatial coordinates in the 3D spatio-temporal representation are identical to those in the original space; as a result, the surface grid depends on the orientation (walking direction) of each 3D object. This places an unnecessary heavy burden on the classifier, that has to generalize into a 'human walking' model movement in all posible directions, each producing radically different HOG feature vectors.

In this paper propose a preprocessing step that addresses this shortcoming by aligning the 3D representation so that it is neutral with respect to the direction of movement. The paper is organized as follows: we first give a brief overview of the background on recognizing and measuring human walking that we assume as a basis (Section II) and then proceed to present our alignment method (Section III), present experimental results (Section IV), and conclude (Section V).

II. RECOGNIZING HUMAN WALKING

As already mentioned, the core idea of the *HPR* method is to approach walking pattern recognition as the task of classifying the 3D objects created by stacking consecutive 2D scans into a 3D spatio-temporal representation, where X,Y is the planar data and Z is the time dimension [7].

More specifically, the first step is to remove the points that correspond to the static map created by Simultaneous Localization and Mapping and used to localize the robot in the environment. The remaining scan points correspond to dynamic objects in the environment. These are used to construct 3D frames by translating the polar scans into the 2D Cartesian space, and then buffering multiple such 2D planes for a period of time. The time dimension is translated into space by multiplying with 5 kmph, the average speed of human walking. In this representation, a pair of walking legs creates the characteristic helix-like shape curves shown in Fig. 1.

The 3D data points are separated and grouped into 3D objects using the DBSCAN clustering algorithm [8]. DBSCAN is an established density-based clustering technique which builds clusters of arbitrary shape, and is efficient for low-dimensional data. DBSCAN fits our spatial data very well, because it takes advantage of the proximity of the data points in the plane. Another advantage is DBSCAN's non-linear separation of clusters, facilitating crossing paths resulting in non-linear disjunctions. Finally, DBSCAN is robust to noise due to the triggering and the accuracy of laser scans. In our experiments we use Euclidean distance as the distance metric, with 40 set as the minimum number of points in a cluster.

The next step is to apply *surface modelling* in order to compose a surface that fits the point cloud. The fitting defines the function $y = f(z, x)$, where z is the scaled time dimension and x, y are the Cartesian coordinates with the robot located at $(0, 0)$ facing towards the positive half of the y axis. In this manner, (a) occlusion guarantees that $f(\cdot)$ is a function; and (b) we fit the most informative surface, the one facing the scanner, and not the 'ceiling' view that would only give us movement patterns without showing the motion of the legs.[1]

The next step is to extract *Histograms of Oriented Gradients (HOG)*, which are descriptors of the gradients in the image. In our experiments we use the sk-image implementation[2] of the HOG descriptors proposed by Dalal and Triggs [9]. The extracted features are then used by classification models, where we have experimented with a *Naive Bayes* classifier with PCA pre-processing for reducing dimensionality to independ features, with *Support Vector Machines* under the *Radial Basis Function (RBF)* kernel, and with *Linear Discriminant Analysis (LDA)*.[3]

The final step is to link the 3D spatio-temporal objects that have been classified as 'human walking', in order to track through time the movement of individual humans in the scene. In order to achieve this, we allow these 3D objects to ovelap in time by placing some 2D scans in both the previous and the next consequtive 3D frame. For each such sequence of partially overlapping frames, we compute the median 3D point from each frame. We then project these points to the spatial plane and translate the temporal coordinate back to time. The end-result is the walking person's trajectory in space as a series of X, Y coordinates annotated with

[1]The implementation used in our experiments is a Python port for ROS of the MATLAB implementation at `http://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit`

[2]See `http://scikit-image.org`

[3]All three, as implemented in Python in the sk-learn library, see `http://scikit-learn.org`

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 2, 2017.

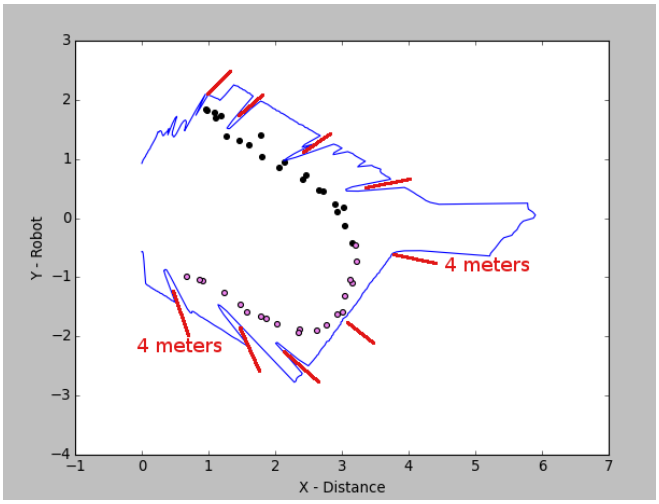


Fig. 2. Median points of clusters projected to the plane. Red arrows show positions of markers placed on the ground every 1 m.

timestamps. Walking speed measurements are made on this final representation (Fig. 2).

This complete pipepine is implemented in Python for ROS Hydro and is publicly available on `https://github.com/roboskel` The background described in this section can be retrieved at the release tagged `1.0`, while the implementation of the new component described and evaluated in the following sections is included in the release under tag `2.0.0`.[4]

III. RANGE DATA CLUSTER ALIGNMENT

The clusters formed by human walking give the characteristic shape shown in Fig. 1. Each of the two traces in that figure tracks a leg and the overall shape results from the human walking pattern where legs alternate between being almost stationary during the *stance* (the parts of the track that rise in the figure) and then moving rapidly during the *swing* (the almost horizontal part of the track in the figure).

The inclination of the imaginary centerline between the two traces gives the walking speed and the projection of this centerline on the spatial plane gives the walking direction. The core idea of our method is that we can improve the classification models by *aligning* movement traces so that their imaginary centerlines have the same pose. If we can do this, we factor speed and orientation out of the point cloud and, subsequently, also out of the resulting surface model and HOG descriptors.

To achieve this, we observe that the data points of each cluster can be approximated by a Gaussian distribution, thus they can be described by a mean value and a covariance matrix. The Gaussian distribution has a direction in the space that can be approached by an ellipsoid. Our aim is to find the main direction of this ellipsoid in order to rotate the data points of the cluster. To do this, we use *Singular Value Decomposition (SVD)*, which decomposes a matrix into three matrices U, Λ, V where Λ is a diagonal matrix and U, V are orthogonal matrices.

SVD fits our approach because we can achieve rotation through the use of the eigenvectors. Specifically, we apply SVD in the covariance matrix of each cluster in order to rotate it in the direction that has the maximum variance. The SVD method is used as a tool for the decomposition of the covariance matrix, to get the eigenvalues and eigenvectors that are necessary for the declaration of the main direction of each cluster.

With the use of SVD method we achieve the alignment of the eigenvalues and the corresponding eigenvectors. By sorting the eigenvalues in a descending order, we use the respective eigenvectors to form the transformation matrix. The maximum eigenvalue represents the main direction of the cluster, thus for its normal distribution too. As the covariance matrix is symmetric, V and U are same. Therefore, U is the linear transformation matrix that we will use for the alignment of the cluster points, as (1) shows.

$$X' = U \cdot (X - \mu) \tag{1}$$

where X, X' are the $l \times N$ initial and transformed data matrices respectively and U is the $l \times l$ transformation matrix. The attribute μ is the mean value of X, thus $(X - \mu)$ specifies a transferring preprocessing step.

As (1) specifies, each transformed data point of the cluster is derived from the projection of the initial data point in the space, where it is declared by the eigenvectors of U matrix. The final alignment of each cluster is computed by transferring its data points to the beginning of the axes and rotating them with the use of the U transformation matrix. The align stage leads to the alignment of the clusters by the same direction, regarding the corresponding variability.

To demonstrate the effect of alignment on the surface fitted to clusters, consider Fig. 3 showing a scene where a person is walking in a different direction relative to the robot than in Fig. 1. Although the resulting clusters look visually similar, their different direction makes their similarity less pronounced in their grid images (Fig. 5a and 5c). Alignment emphasises their similarities and tranforms the original problem into one that is more suited for classifying using HOG descriptors (Fig. 5b and 5d). What should be noted is that using the direction of the Gaussian elipsoid of the point cloud to find the direction of the transformation we are expressing a property of natural, uninhibited human gait. In other movement patterns, such as stepping sideways, the largest variance in the point cloud will not necessarily correspond with the direction of movement and will not maintain direction invariance in the HOG descriptors. This is, however, a positive quality of the approach for our use case, as it removes unnatural and special circumstances from our walking speed measurements.

Finally, it should be noted that alignment does not obscure

[4]The repository for the system described here is `https://github.com/roboskel/HumanPatternRecognition` Version 2.0.0 can be downloaded from `https://github.com/roboskel/HumanPatternRecognition/archive/2.0.0.tar.gz`

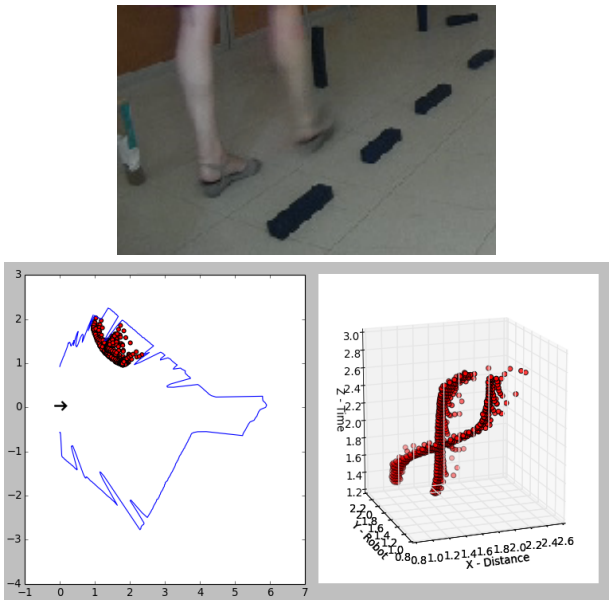


Fig. 3. The 3D spatio-temporal representation (right) and the projection on the map (left) of somebody walking away from the robot. The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), as marked by the black arrow.

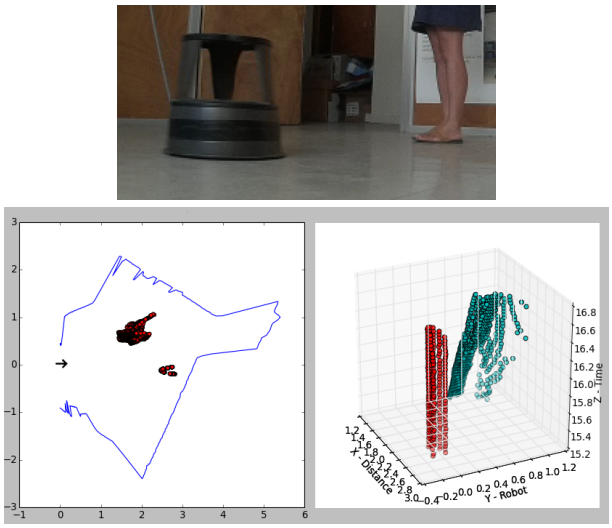
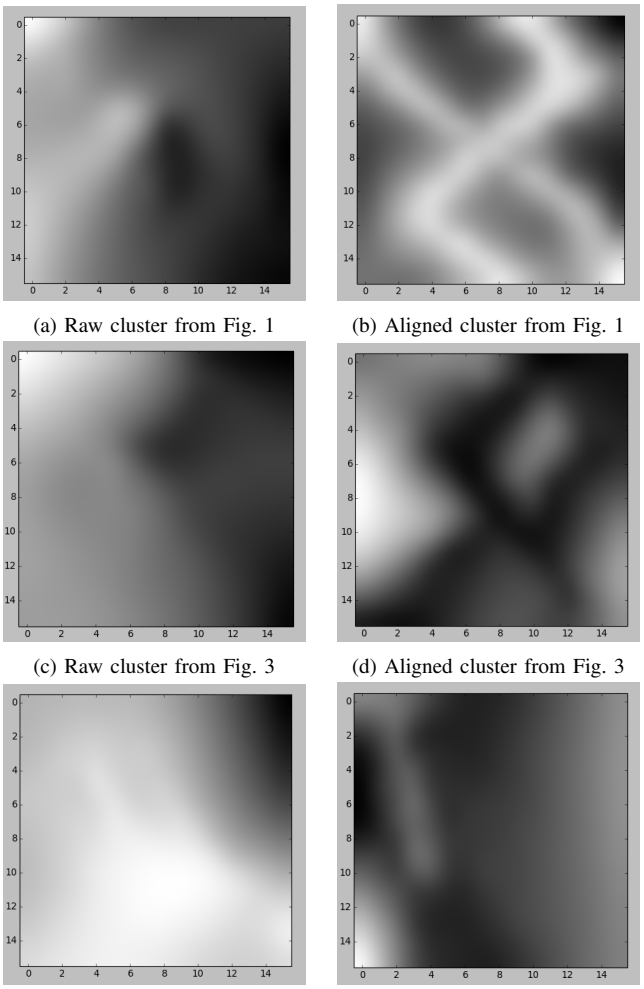


Fig. 4. The 3D spatio-temporal representation (right) and the projection on the map (left) of a stool rolling near a person standing still. The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), as marked by the black arrow.

the distinction between human walking and other movement. As an example, Fig. 4 shows a cluster from a stool pushed to roll on its wheels. As can seen in Fig. 5e and 5f, the aligned grid image remains separable from the walking pattern grid images.

IV. EVALUATION

To test our alignment method, we will compare the classification accuracy of the pipeline desctibed in Section II with and without the alignment stage. We have collected scans where people walk with speed and direction changes, periods of standing still, and interacting with furniture in the room



(a) Raw cluster from Fig. 1 (b) Aligned cluster from Fig. 1

(c) Raw cluster from Fig. 3 (d) Aligned cluster from Fig. 3

(e) Raw cluster from Fig. 4 (f) Aligned cluster from Fig. 4

Fig. 5. Grid images of the three clusters.

(sitting in chairs, picking and putting down boxes, walking between furniture).

The range finder is a Hokuyo UST-10LX mounted on a robot that is stationary throughout data collection. The following parameters were used:

- The range finder is mounted 12cm above the ground, collecting scans just above the human's ankle
- One scan is obtained every 25msec
- Each clustering frame includes 40 scans, and lasts 1 sec
- Clustering is done on Euclidean distance as the distance metric, with a minimum of 40 points in a cluster
- 36D HOG features are extracted from 16x16 surfaces with the use of 6 histogram bins, 8x8 cell size in pixels and with un-normalisation in the histogram's blocks

We collected 811 such frames, with a duration of 1 sec each, split into 12 scenes. These were randomly split into a training set (70% of the data) and a testing set (the rest of the data).

We used the resulting feature vectors to evaluate binary classification between human walking vs. anything else. We experimented with three different classification methods:

- *Naive Bayes* with *Principal Component Analysis (PCA)*

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 2, 2017.

pre-processing for reducing dimensionality to independ features, as previously used by Varvadoukas et al. [7] in their experiments.

- *Linear Discriminant Analysis (LDA)*, a linear classifier that is closely related to Naive Bayes. LDA uses Bayes' rule and the model is generated by fitting class conditional densities to the data.
- *Support Vector Machine (SVM)* under the *Radial Basis Function (RBF)* kernel.

The metrics that we use in order to evaluate our classification experiments are: Precision, Recall and Accuracy. Table I presents the experimental results with and without the alignment stage. We can observe that the alignment stage improves performance for all three classification method. Moreover we can conclude that LDA classifier is better suited to this task, achieving an accuracy of 90.46%, which is the highest among all our experiments with and withought alignment.

What we can also observe in Table I is that SVM massively overgeneralizes when presented with non-aligned feature vectors and accepts too many false positives (i.e., low precision). The precision increase under alignment is consistent with our hypothesis that alignment homogenizes and makes separable the HOG feature vectors.

V. CONCLUSIONS

We presented a system that analyses planar range data to recognize human walking patterns and to separate them from patterns of other moving objects. To some extent, natural human walking is also separable from unusual gaits, sideways stepping, and other patterns that diverge from the common forward moving, stance/swing cycle.

The contribution presented in this paper is the addition of an alignment stage which converts range data into a representation where the 3D 'walking legs' objects are aligned to the direction of movement. The advantage of this pre-processing is that walking patterns remain similar regardless of the direction of movement and the resulting features are better descriptors of the movement pattern. Classifying aligned data evaluates favourably to classifying the same data without alignment, increasing classification accuracy from 73% to 90%.

TABLE I
EVALUATION OF CLASSIFICATION PERFORMANCE WITH AND WITHOUT ALIGNMENT

Classifier	Metrics	Without alignment	With alignment
NB + PCA	Precision	100.00%	81.25%
	Recall	22.76%	55.91%
	Accuracy	23.32%	81.27%
SVM	Precision	6.25%	**82.81%**
	Recall	40.00%	55.79%
	Accuracy	76.67%	81.27%
LDA	Precision	12.50%	76.56%
	Recall	28.57%	**80.33%**
	Accuracy	73.14%	**90.46%**

The system is developed in the context of RADIO, an independent ageing project where a robotic home assistant is used to collect clinical observations regarding the functional capability of the home's occupant. One of the clinical requirements for RADIO system is measuring walking speed, while simultaneously satisfying ethical requirements regarding the obtrusiveness of the data collection methods and the privacy considerations regarding the nature of the data.

Our prototype is implemented in Python for ROS Hydro, open source and publicly available at `https://github.com/roboskel/HumanPatternRecognition`

REFERENCES

[1] C. Antonopoulos, G. Keramidas, N. S. Voros, M. Hübner, D. Goehringer, M. Dagioglou, T. Giannakopoulos, S. Konstantopoulos, and V. Karkaletsis, "Robots in assisted living environments as an unobtrusive, efficient, reliable and modular solution for independent ageing: The RADIO perspective," in *Proceedings of the 11th International Symposium on Applied Reconfigurable Computing (ARC 2015), Bochum, Germany, 15–17 April 2015*, ser. LNCS, vol. 9040. Springer, 2015, pp. 519–530.

[2] K. O. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *Proc. of the 2008 IEEE Intl Conf. on Robotics and Automation, (ICRA 2008), May 19-23, Pasadena, CA, USA*, 2008, pp. 1710–1715.

[3] L. Spinello, R. Triebel, and R. Siegwart, "Multimodal people detection and tracking in crowded scenes," in *Proc. 23rd AAAI Conf. on Artificial Intelligence (AAAI 2008), Chicago, IL, 13–17 July 2008*, 2008, pp. 1409–1414.

[4] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *I. J. Robotic Res.*, vol. 24, no. 1, pp. 31–48, 2005.

[5] H. Nemati and B. Åstrand, "Tracking of people in paper mill warehouse using laser range sensor," in *Proceedings of the 2014 European Modelling Symposium (EMS 2014)*. IEEE, 2014.

[6] D. Ristić-Durrant, G. Gao, and A. Leu, "Low-level sensor fusion-based human tracking for mobile robot," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 1, no. 1, 2016.

[7] T. Varvadoukas, I. Giotis, and S. Konstantopoulos, "Detecting human patterns in laser range data," in *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, ser. Frontiers in Artificial Intelligence and Applications, vol. 242, Aug. 2012.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD 1996)*, 1996, pp. 226–231.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*. IEEE, 2005.