Detecting and Measuring Human Walking in Laser Scans

Katerina Zamani
Institute of Informatics and Telecommunications
NCSR 'Demokritos'
Athens, Greece
and Electrical and Computer Engineering, National Technical University of Athens, Greece
kzam@iit.demokritos.gr

Georgios Stavrinos
Institute of Informatics and Telecommunications
NCSR 'Demokritos'
Athens, Greece
gstavrinos@iit.demokritos.gr

Stasinos Konstantopoulos
Institute of Informatics and Telecommunications
NCSR 'Demokritos'
Athens, Greece
konstant@iit.demokritos.gr

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 about 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 pre-processing 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.

CCS CONCEPTS

• **Computing methodologies** → **Vision for robotics**; *Shape inference*; *Supervised learning by classification*; • **Applied computing** → Health informatics;

KEYWORDS

Pattern Recognition, Robotics and Assistive Technologies

ACM Reference Format:
Katerina Zamani, Georgios Stavrinos, and Stasinos Konstantopoulos. 2018. Detecting and Measuring Human Walking in Laser Scans. In *SETN '18: 10th Hellenic Conference on Artificial Intelligence, July 9–15, 2018, Rio Patras, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3200947.3201026

1 INTRODUCTION

For many robotic 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. That alone motivated a considerable body of robot perception work to look into recognizing and tracking human figures in the scene. More recently, and in the context within which this paper operates, this trend is further stimulated by applications from *assisted living environments* where robots are expected to collect medically relevant data, including data for gait analysis [1]. In this context, tracking human movement is not only input for motion planning, obstable avoidance, and human-robot interaction, 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.

Among the perception modalities and methods for recognizing and tracking humans, *range data* collected from laser scanners offer significant advantages for assisted living applications: reliable and accurate measurements can be made, especially in indoors applications, and robot-mounted laser scanners can unobtrusively collect data by remote sensing without requiring wearable motion sensors. What is also interesting is that range data (and especially the planar laser scanner data discussed here) carries, by its nature, very little information. This is an advantage for avoiding the privacy issues around collecting and analysing visual or even 3D data collected in private residences, but it 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 fits nicely with the technical objectives of assisted living applications, since they require measurements regarding walking and, in general, mobility.

However, analysing representations that span multiple frames presupposes solving the *data association* problem and especially in situations such as occlusion, data sparsity, and physical proximity of objects. Several studies in multiple-target tracking track objects by using using *Kalman filters* as *motion models* that estimate the future track position from past observations [2].

An approach applied in indoors and outdoor scenarios too for human detection, tracking and following, is presented by Leigh et al. [7]. Firstly they group together laser points that are withing a predefined threshold, in a way that a person's two legs are been

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
SETN '18, July 9–15, 2018, Rio Patras, Greece
© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
ACM ISBN 978-1-4503-6433-1/18/07…$15.00
https://doi.org/10.1145/3200947.3201026

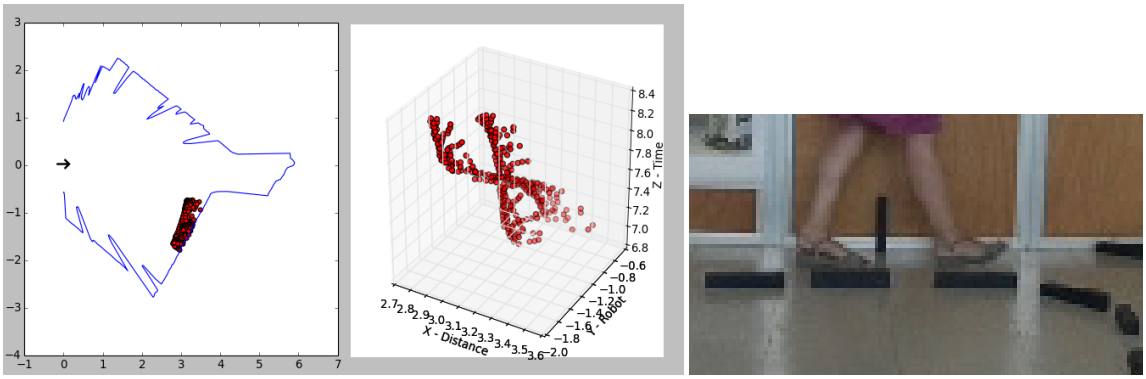


Figure 1: The 3D spatio-temporal representation (middle) and the projection on the map (left) of somebody walking across the front view of the robot (right). The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), as marked by the black arrow.

separated into two distinct clusters. The clusters are classified as humans or not by training a random forest classifier with some statistical and geometric features of the clusters, as Arras et al. [3] proposed. The output confidence level of the classifier along with the cluster's location are given as input to the tracking module of their system [7]. Specifically, the position of each detected cluster is used for Kalman filtering with a constant motion model. Then a *Global Nearest Neighbour (GNN)* data association is applied in order to track multiple detected clusters.

Although successful, Kalman filters face the key issue of defining the motion model. To address this, Spinello et al. [10] 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 [8], 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. [9], for example, fused range data with 3D depth data. Although successful in increasing robustness, the fusion of different modalities voids the privacy argument in favor 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 [11]. 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 occlusions 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 modeling, histogram of oriented gradients) to extract classification features. The fundamental assumption is the same as by Nemati and Åstrand [8] that object segmentation and object tracking through time should be based on geometric proximity. However, in the method by Varvadoukas et al. [11] these assumptions are manifested as unsupervised clustering in the Euclidean space whereas Nemati and Åstrand [8] 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 data across frames, this method was very sensitive to the direction of movement: different directions produce radically different surface models and, consequently, histogram of oriented gradients (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 possible 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 2) and then proceed to present our alignment (Section 3) and tracking (Section 4) methods, experimental results (Section 5), and conclusion (Section 6).

2 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 [11].

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 shown in Fig. 1.

The 3D data points are separated and grouped into 3D objects using the DBSCAN clustering algorithm [6]. 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 modeling* 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 [5]. 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 independent features, with *Support Vector Machines* under the *Radial Basis Function (RBF)* kernel, and with *Linear Discriminant Analysis (LDA)*.[3]

3 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; the projection of this centerline on the spatial plane gives the walking direction. The core contribution of the work described in this paper 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, 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, and thus also the main direction of its normal distribution. 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:

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

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 Eq. 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. 2 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. 4a and 4c). Alignment emphasises their similarities and transforms the original problem into one that is more suited for classifying using HOG descriptors (Fig. 4b and 4d). What should be noted is that using the direction of the Gaussian ellipsoid 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

[1] Specifically, we use in our experiments the Gridfit algorithm, as it can be found at `http://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit`
We re-implemented the algorithm in Python for the benefit of ROS integration.
[2] See `http://scikit-image.org`
[3] All three, as implemented in Python in the sk-learn library, see `http://scikit-learn.org`

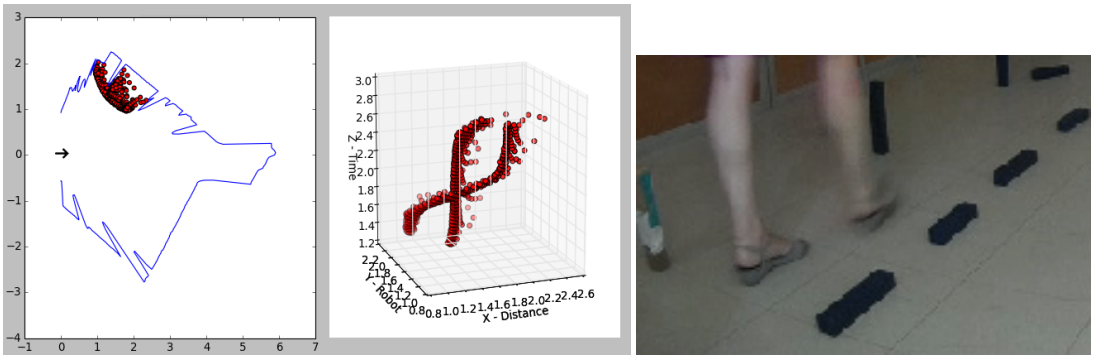


Figure 2: The 3D spatio-temporal representation (middle) and the projection on the map (left) of somebody walking away from the robot (right). The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), the location marked by the black arrow in the map projection.

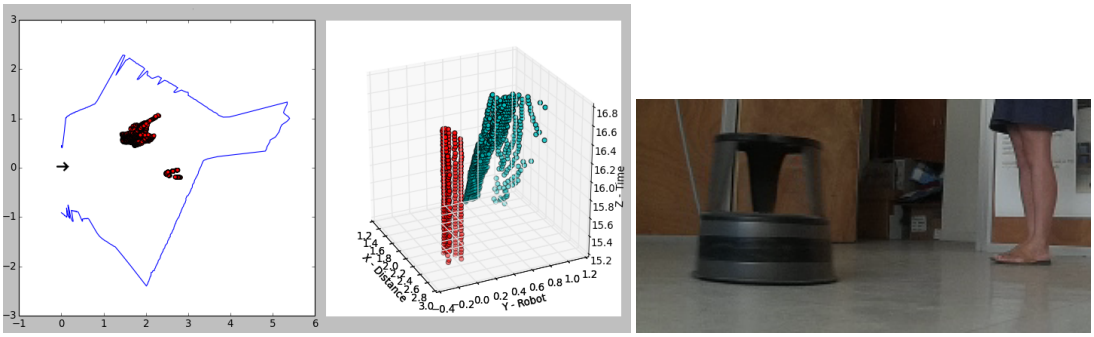
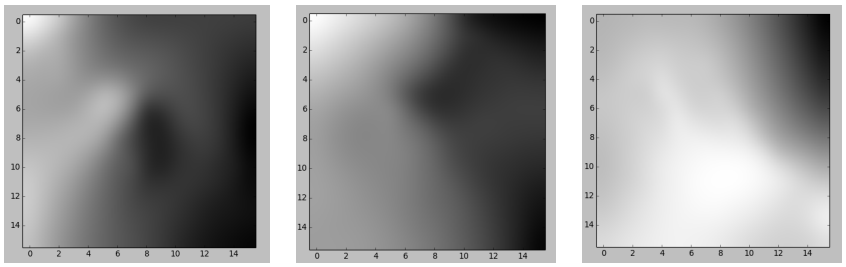


Figure 3: The 3D spatio-temporal representation (middle) and the projection on the map (left) of a stool rolling near a person standing still (right). The robot is at $(X, Y) = (0, 0)$ facing towards the positive half of the horizontal axis (X), the location marked by the black arrow in the map projection.



(a) Raw cluster from Fig. 1 (c) Raw cluster from Fig. 2 (e) Raw cluster from Fig. 3

(b) Aligned cluster from Fig. 1 (d) Aligned cluster from Fig. 2 (f) Aligned cluster from Fig. 3

Figure 4: Grid images of the three clusters.

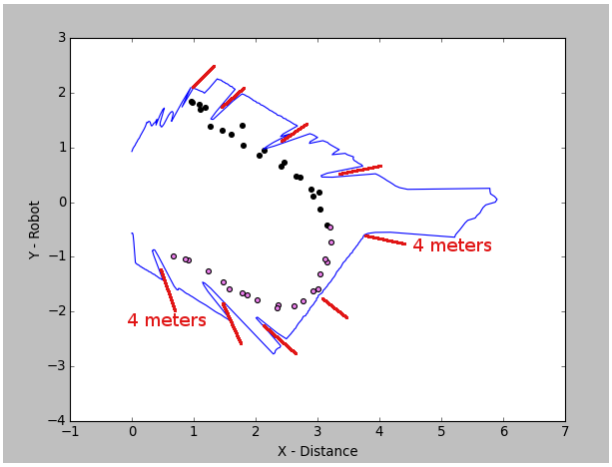


Figure 5: Median points in the plane that show the trajectory of a walking human tracked across several frames. The two different point colors represent the two 4m spans recognized by the system. The ground truth is derived from markers on the ground set 1m appart with their position given here by the red lines.

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 the distinction between human walking and other movement. As an example, Fig. 3 shows a cluster from a stool pushed to roll on its wheels. As can seen in Fig. 4e and 4f, the aligned grid image remains separable from the walking pattern grid images.

4 CLUSTER TRACKING AND SPEED MEASUREMENT

We have so far discussed the categorization of moving objects in laser scans as human walking or other movement, by recognizing walking patterns in frames of a relatively short duration. Our frames-based method circumvents the hardests aspects of the data association problem, but to offer a complete solution we still need to associate the recognized objects between frames. For instance, for our gait analysis use case, the requirement is that walking speed is measured over a distance of 4m. Since frames are shorter than the time it takes to walk 4m, we need to track the movement of each person in the scene and to make measurements over sequences of frames.

To achieve this, we re-defined frames so that they ovelap in time, by placing two (2) scans in both the previous and the next consequtive frame. In this manner, we have a partial overlap that helps us link each 3D object classified as 'human walking' in one frame with the most-overlapping human-walking 3D object in the next frame. We use the Euclidean distance metric to compute proximity. Since the core method described in the previous section has already clarified many of the situations where occlusions appear into dicreet 3D objects for each human, this is a considerably easier task than directly associating point between scans in the raw data.

Besides tracking the same person across frames, we also need to reduce these point clouds to a position on the planar map, in order to make the walking speed measurement. To do this, we slice the clusters to segments (thinner than complete frames) along the time dimention. Each segment is reduced to its median 3D point and projected to the spatial plane, so that we get a series of timestamped positions of the person on the planar map. Walking speed measurements are made on this final representation (Figure 5).

The advantage of measuring walking speed as described here (as opposed to directly using the raw scan) is tracking over longer sequences and also making measurements on moving objects that match the normal human gait patterns. If a person stands still in the middle of a 4m span or needs to walk sideways or in general needs for whatever reason to assume a gait that does not match normal walking, those sequences are disregeraded.

5 EXPERIMENTS AND RESULTS

The methods described above are implemented in Python for ROS Indigo. The complete ROS node pipeline implementing the work described here is publicly available, as well as the the implementation of the background we compare against (cf. Section 2).[4]

Tests are made on data recorded by a stationary TurtleBot2, observing the scene with a Hokuyo UST-10LX range finder. The range finder is mounted 12cm above the ground, collecting scans just above the human's ankle. One scan is obtained every 25msec.

5.1 Cluster Alignment Evaluation

To test our alignment method, we compare the classification accuracy of the pipeline described in Section 2 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 (sitting in chairs, picking and putting down boxes, walking between furniture).

The following parameters were used:

- Each clustering frame includes 40 scans, and lasts 1 sec
- The average speed of human walk is set to 5 kmph. This parameter is used to convert time into distance and create the 3D spatio-temporal frames (Section 2).
- 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:

[4]The repository is
`https://github.com/roboskel/HumanPatternRecognition`
The method can be retrieved at the release tagged `2.0.1`. The method described in Section 2 can be retrieved at the release tagged `1.0`.

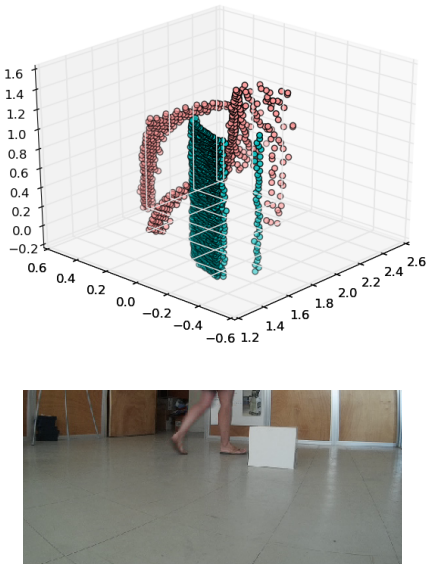


Figure 6: Characteristic instance of a scenario where the human is walking around the back of a box and the clusters produced by this scene.

- *Naive Bayes* with *Principal Component Analysis (PCA)* preprocessing for reducing dimensionality to independent features, as previously used by Varvadoukas et al. [11] 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 1 presents the experimental results with and without the alignment stage.

Table 1: 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%**

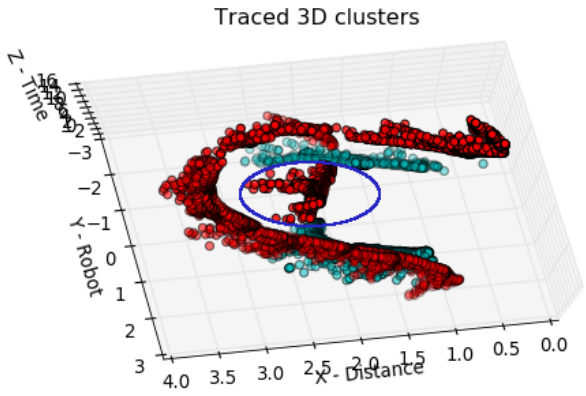


Figure 7: Tracked clusters in a crossed human situation. The different colors represent the different clusters (i.e., people) each scan point was placed into. The circle shows the clustering error.

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 without alignment. What we can also observe in Table 1 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.

5.2 Cluster Tracking Evaluation

Evaluation is based on recordings of scenarios with proximity situations and occlusions, sudden and smooth changes in walking speed and direction or completely stopping and resuming walking, two people crossing each other while walking, and people walking behind or picking up and relocating a box. Figure 6 has a characteristic instance.

Besides the Hokuyo UST-10LX, we have also used a Hokuyo URG-04LX range finder, which is slower and also scans a shorter range. Both are located 12cm above the ground, thus they collect data points above the human's ankle. The following parameters were used:

- The time scale of UST-10LX is set to 25 msec and of URG-04LX to 50 msec.
- The clustering window is set to 40 scans for UST-10LX and to 20 scans for URG-04LX.
- The minimum distance walked to accept a measurement is 4m.
- Based on the experiments presented above, the LDA classifier performs better so the experiments in this section use the LDA model.

The evaluation metrics are the *Mean Absolute Error (MAE)* and the *Mean Relative Error (MRE)*. MAE gives the mean absolute error in the time (sec) to walk 4 meters:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |f_i - y_i| \quad (2)$$

where n is the number of the measurements, where f_i are the system-computed values, and y_i the ground-truth values. MRE indicates how good a measurement is relatively to the size of the measured quantities:

$$MRE = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{|f_i - y_i|}{y_i} \quad (3)$$

Table 2 presents, for both scanners, the MAE and MRE of the whole dataset and of the dataset excluding the sessions with crossing trajectories. Fig. 7 presents the tracked clusters of a crossed human session. We can notice that erroroneus clustering of the points in the circle as 'red' radically changes the median point of the 'green' cluster and thus of the walking speed computations.

Naturally, the higher error rate in the crossing scenarios impacts walking speed measurements. It should be noted, however, that this has little impact for our particular application where we need one, indicative, walking speed value per day. This can be extracted by inference or by statistical aggregation: the system can infer which measurement instances are more reliable by discarding instances where other moving objects are detected nearby; or the system can statistically discard outliers given our use case scenario of health monitoring for a person living alone and where most of the measurements can be safely assumed to be of the primary user without other people in the scene.

6 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 core contribution presented in this paper is the improvement of a previously published method for detecting human walking patterns by adding an alignment preprocessing stage. Alignment converts range data into a representation where position and direction of motion is factored out of the 3D 'walking legs' objects that are used for feature extraction and classification. 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 favorably to classifying the same data without alignment, increasing classification accuracy from 73% to 90%.

Table 2: MAE and MRE for the walking analysis. The error measures are computed for the whole dataset and for the subset that excludes the crossing trajectories sessions.

	Error Metrics	*UST-10LX*	*URG-04LX*
Complete dataset	MAE	0.89	2.29
	MRE	33%	59%
Excluding crossing	MAE	0.44	2.75
	MRE	11%	68%

A secondary contribution is porting the complete processing pipeline from MATLAB to Python, and using the ROS middleware for its data acquisition and its inter-component communication. This facilitated integrating the pipeline into the wider health monitoring system developed by the EU project *RADIO: Robots in Assisted Living Environments as an Unobtrusive, Efficient, Reliable and Modular Solution for Independent Ageing.*

In the context of the RADIO application and use cases, medically relevant information extracted from planar range data is extremely valuable: Remote sensing satisfies requirements for unobtrusive data acquisition and the nature of planar range data satisfies privacy considerations regarding collecting potentially sensitive footage from private residences. Future work within RADIO will use data from evaluation studies with primary users during the RADIO project to identify the best method for extracting the indicative daily value discussed in Section 5.2.

Further work will look into fusing data from multiple robots observing the same scene or using the robot's movement capability to improve the viewing angle with respect to clarifying occlusions.

ACKNOWLEDGMENTS

This project has received funding from the European Union's 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`

REFERENCES

[1] Christos Antonopoulos, Georgios Keramidas, Nikolaos S. Voros, Michael Hübner, Diana Goehringer, Maria Dagioglou, Theodore Giannakopoulos, Stasinos Konstantopoulos, and Vangelis Karkaletsis. 2015. 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 (LNCS)*, Vol. 9040. Springer, 519–530.

[2] Kai Oliver Arras, Slawomir Grzonka, Matthias Luber, and Wolfram Burgard. 2008. 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.*

[3] Kai O Arras, Oscar Martinez Mozos, and Wolfram Burgard. 2007. Using boosted features for the detection of people in 2d range data. In *Robotics and Automation, 2007 IEEE International Conference on.* IEEE, 3402–3407.

[4] Maren Bennewitz, Wolfram Burgard, Grzegorz Cielniak, and Sebastian Thrun. 2005. Learning Motion Patterns of People for Compliant Robot Motion. *I. J. Robotic Res.* 24, 1 (2005), 31–48.

[5] Navneet Dalal and Bill Triggs. 2005. Histograms of Oriented Gradients for Human Detection. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005).* IEEE. https://doi.org/10.1109/CVPR.2005.177

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei et al. Xu. 1996. 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).* 226–231.

[7] Angus Leigh, Joelle Pineau, Nicolas Olmedo, and Hong Zhang. 2015. Person tracking and following with 2d laser scanners. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on.* IEEE, 726–733.

[8] Hassan Nemati and Björn Åstrand. 2014. Tracking of people in paper mill warehouse using laser range sensor. In *Proceedings of the 2014 European Modelling Symposium (EMS 2014).* IEEE.

[9] Danijela Ristić-Durrant, Ge Gao, and Adrian Leu. 2016. Low-level Sensor Fusion-based Human Tracking for Mobile Robot. *Facta Universitatis, Series: Automatic Control and Robotics* 1, 1 (2016).

[10] Luciano Spinello, Rudolph Triebel, and Roland Siegwart. 2008. Multimodal People Detection and Tracking in Crowded Scenes. In *Proc. 23rd AAAI Conf. on Artificial Intelligence (AAAI 2008), Chicago, IL, 13–17 July 2008.* 1409–1414.

[11] Theodoros Varvadoukas, Ioannis Giotis, and Stasinos Konstantopoulos. 2012. Detecting Human Patterns in Laser Range Data. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012) (Frontiers in Artificial Intelligence and Applications)*, Vol. 242.