Modelling and Predicting Rhythmic Flow Patterns in Dynamic Environments

Sergi Molina[1], Grzegorz Cielniak[1], Tomáš Krajník[2], and Tom Duckett[1]

[1] Lincoln Center for Autonomous Systems, University of Lincoln, UK
{smolinamellado,gcielniak,tduckett}@lincoln.ac.uk
[2] Faculty of Electrical Engineering, Czech Technical University, Prague, Czechia
tomas.krajnik@fel.cvut.cz

Abstract. We present a time-dependent probabilistic map able to model and predict flow patterns of people in indoor environments. The proposed representation models the likelihood of motion direction on a grid-based map by a set of harmonic functions, which efficiently capture long-term (minutes to weeks) variations of crowd movements over time. The evaluation, performed on data from two real environments, shows that the proposed model enables prediction of human movement patterns in the future. Potential applications include human-aware motion planning, improving the efficiency and safety of robot navigation.

Keywords: mobile robots, long-term autonomy, human activities

1 Introduction

Robotic automation has been happening in industry for many years in structured environments. Modern factory floors are changing to work in a more flexible way, where robots such as AGVs or forklifts have to share the same environment with people in a collaborative manner. To improve the efficiency of mobile robot navigation in human-populated areas, a robot should take into account the general flow of the people in the environment [1]. However, the flows of people movement are not constant over time, but are constantly changing.

This work considers building spatio-temporal models of human flow to improve the efficiency of long-term robot operation, assuming that the observed direction of motion follow underlying patterns due to the rhythmic nature of human activities. We model the temporal dimension of activities using periodic functions at different time scales, going from minutes to days. By learning typical motion patterns, the robot can either avoid areas with high density of people movement, or move among the crowd with the expected dominant direction of flow, thus blending better with its surroundings.

Our model applies the Frequency Map Enhancement (FreMEn [2]) to a grid-based discretized representation of space, where each cell contains several hypothesis of people movement direction. To evaluate the proposed model, we performed experiments using two datasets gathered over the course of several weeks: one from a shopping center in Japan [3], and one at University of Lincoln, UK. The experiments show that the proposed model can learn the flow patterns and predict people movement at a given time in the future.

2 Related work

Recently, a number of approaches for mapping environment dynamics have been proposed. One general approach is to store past observations in a compressed way, as in Arbuckle et al. [4] and Mitsou et al. [5], without the actual intention of using the past observations for predictions of future states. Other approaches learn generative models, which can predict people motion direction in order to improve collision avoidance. While these approaches use long-term data for training, they are aimed at short term prediction of people movement based on current observations [6–8]. Works like Wang et al. [9], Kucner et al. [10], and Saarinen et al. [11] treat dynamics as a change of occupancy in grid map cells, which allows to represent characteristic motion patterns at given areas. [10] proposed to model how the occupancy likelihood of a given cell in a grid is influenced by the neighbouring cells and showed that this representation allows to model object movement directly in an occupancy grid. Similarly in [9] and [11] the direction of traversal for each cell is obtained using an input-output hidden Markov model and a Markov chain, respectively. The aforementioned works assume that the typical motion patterns are constant in time, which might not be typical, e.g. human flows at an entrance of an office building at the start and end of a work-day are likely to be in opposite directions.

Thus, one should take into account the temporal domain, which is important in human-populated environments. For example, Dayoub et al. [12] and Rosen et al. [13] used statistical methods modelling short and long-term to create persistence models and reasoned about the stability of the environmental states over time. Tipaldi et al. [14] proposed to represent the occupancy of cells in a traditional occupancy grid with a hidden Markov model for improving robot localisation over the traditional approaches. The idea of identifying periodic patterns in the measured states and using them for future predictions was presented in Krajnik et al. [2]. These approaches demonstrated that by considering temporal aspects such as periodicities in the environment, robotic capabilities including localisation, planning and exploration can be improved [14–16].

Inspired by the cell models to represent dynamics and the power of spectral analysis to create spatio-temporal models capable of long-term predictions, we have developed a model which is able to predict crowd movements. Similar to the work by Jovan et al. [17] based on counting human trajectories during intervals of time, in our approach we count the number of people walking in certain directions in a grid-based representation of the environment.

3 Spatio-Temporal Model

The aim of this work is to create a model of human motion which is able to predict the flow patterns of people over time, as well as where and when these flows are happening. The underlying geometric space is represented by a grid, where each cell contains k temporal models, which correspond to k discretized orientations of people motion through the given cell over time. Since the total number of temporal models, which are of a fixed size, is $k \times$ the number of cells (n), the spatio-temporal model does not grow over time regardless of the duration of data collection. This makes the model not only memory efficient, but also allows to make probabilistic predictions of the likely flow of people in a certain direction for a given cell at any instant of time.

3.1 Temporal Framework - FreMEn

The temporal models, which can capture patterns of people movement, are based on the FreMEn framework [2]. FreMEn is a mathematical tool based on the Fourier Transform, which considers the probability of a given state as a function of time and represents it by a combination of harmonic components. The model not only allows representation of environment dynamics over arbitrary timescales with constant memory requirements, but also prediction of future environment states based on the patterns learned.

To illustrate the basic principles of FreMEn, let us consider a binary state such as the presence or absence of a visual feature (see Fig. 1). Treating the measured state as a signal, we decompose it by means of the Fourier Transform, obtaining a frequency spectrum with the corresponding amplitudes, frequencies and phase shifts. Then, transferring the most prominent spectral components to the time domain provides an analytic expression representing the probability of the feature being visible at a given time in the past or future.

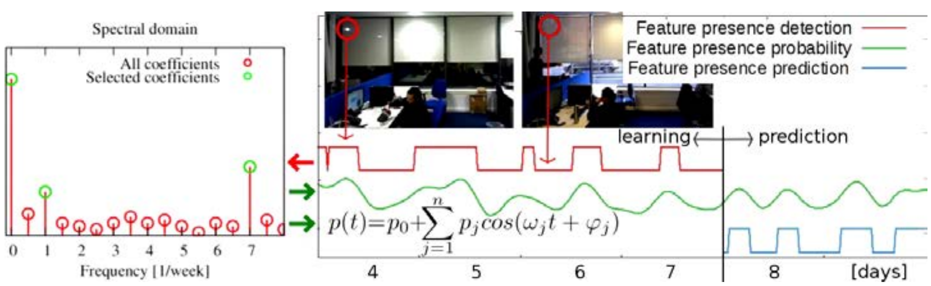

Fig. 1: Example of the FreMEn applied in a binary state [2].

Assuming the direction of people movement is affected by patterns that might be periodic, we apply the FreMEn concept to discretised directions of people movement through a particular cell.

3.2 Building the model

Our model assumes that it is provided with people detection data, containing person position and orientation (x, y, α). At the beginning of the model construction, we associate each cell with k bins, corresponding to the discretised orientation of people motion and to the k temporal models. When building the model, the x, y positions are discretised and assigned to a corresponding cell and the orientation α is assigned to one of the k bins, whose value is incremented by 1. In other words, we count the number of people detections occurring in each orientation bin and cell. After a predefined interval of time, we normalise the bins, and use the normalised values to update the spectra of the temporal models by the scheme described in [2]. Then, we reset the bin values to 0 and start the counting again. Notice that the number of people detected in a particular cell is not a determining factor since we aim to model the relative amount of occurrences among all k orientations.

3.3 Making predictions

To predict the orientation of a human movement through a cell at a future time t, we first calculate the probability for each discretised orientation θ, ($\theta = i\frac{2\pi}{k}$ and $i \in \{0, 1, \ldots, k-1\}$), associated to that cell as

$$p_\theta(t) = p_0 + \sum_{j=1}^{m} p_j cos(\omega_j t + \varphi_j), \tag{1}$$

where p_0 is the stationary probability, m is the number of the most prominent spectral components, and p_j, ω_j and φ_j are their amplitudes, periods and phases. The spectral components ω_j are drawn from a set of ω_s that, in our case, we have selected to cover periodicities ranging from 20 minutes to 1 day with the following distribution:

$$\omega_s = 2\pi \frac{1+s}{24 \cdot 3600}, \qquad s \in 0, 1, 2, 3, 4, ..., 72. \tag{2}$$

After computing the probabilities for every single orientation, we conclude that the dominant orientation within that cell for that instant of time, t, corresponds to the orientation with the highest predicted probability:

$$cell_\theta = argmax(p_\theta(t)), \qquad \theta \in i\frac{2\pi}{k}, \;\; i \in \{0, 1, \ldots, k-1\}. \tag{3}$$

4 Evaluation datasets

In order to evaluate the model, we have run experiments using two datasets. The first one is from the ATC shopping center in Japan [3], and the second one is a dataset recorded in the corridors of our offices at the University of Lincoln, UK (UoL). Both of them provide an environment with complex movements and enough days to the train the models in the long-term, although ATC covers a larger area than UoL. The people detections inside the environment are given in x, y coordinates together with the angle of movement α for every timestamp.

4.1 ATC dataset

The first one is a pedestrian tracking dataset recorded at the ATC shopping center in Osaka, Japan [3]. The perception system consists of multiple 3D range sensors (36 $\times$ Panasonic D-Imager, 11 $\times$ Asus Xtion PRO and 2 $\times$ Velodyne HDL-32E), covering an area of approximately 900 m^2 (Fig. 2), able to detect and track all the people around the place in every instant of time. The data provided was recorded on every Wednesday and Sunday between October 24th, 2012 and November 29th, 2013, resulting in a total of 92 days. From all these data, we selected the first 17 Wednesdays and 17 Sundays, using 15 of each as our training set and the other 4 days as testing data. The recording of each day provides people trajectories starting from approximately 09:00 until 21:00, so for the rest of day we assume there are no occurrences of people, simulating the shopping center being closed.

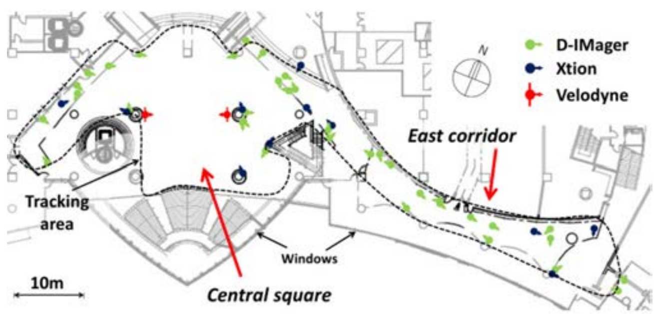

Fig. 2: Shopping center map with the location of all the sensors and the two main zones.

4.2 UoL dataset

The second dataset was collected at one of corridors in the Isaac Newton Building building at the University of Lincoln (UoL). The data collection was performed by a stationary robot equipped with a Velodyne VLP-16 and a 2D laser. The robot was placed in one of the T-shaped junctions in a way that allowed its sensors to scan the three connecting corridors simultaneously, covering a total area of approximately 64 m^2 (see Fig. 3). To detect and localize people in the 3D point cloud provided by the lidar, we used an efficient and reliable person detection method developed by Yan et al. [18]. Our dataset spanned from mornings to late evenings for 14 days, sparsely recorded over a four week period. From these, 13 days were used for training and the remaining one was used to evaluate the model.

(a) (b)

Fig. 3: UoL dataset: Robot location in the corridor (a) and example of a person walking seen by the Velodyne scans (b).

5 Experiments

In our experiments, we discretised the space into 1×1 m cells for the ATC dataset and 0.5×0.5 m cells for the UoL one, which results in a total of $n = 1200$ and $n = 400$ active cells, respectively. Regarding the dynamics, works presented in [10] [9] treat them by means of cell-to-cell transitions. However experiments in our previous work (Molina et al. [19]) showed that the diagonal cell transitions tend to be less likely than the 4 cardinal directions (N/S/E/W). So in order to ensure that all the orientations are treated the same way, we have instead discretized the angles of people traversing a particular cell into a finite set of directions ($k = 8$) as in Fig 4a. There is no theoretical limit on how coarse the granularity of the spatial and angular discretization can be, apart from the increased computational cost. However, setting it too high could lead to not have enough data to fill all states and a loss of generality, while on the other hand, setting it too low could increase the accuracy but result in a much lower model resolution in exchange.

Moreover, the interval used for creating the histograms, used as the input for the model in the training phase, was 10 minutes. Fig. 4b shows an example of a situation where during a random 10-minute interval in a random cell, a total of 61 people detections occurred within the boundaries of that cell. The distribution of detections shown in the histogram states that for those 10 minutes, the people traversing that cell tend to walk at angles in the range of the bin corresponding to θ_2 (i.e. between 67.5 and 112.5 degrees).

5.1 Comparing ground truth against predicted values

To estimate the optimal number of model components (m) to be taken into account for predictions, we need to compare the predictions obtained from the model against the real data. We have defined the error as a percentage ratio between the cells predicted with the wrong orientation when compared with the ground truth and the total number of cells.

(a) (b)

Fig. 4: (a) 8 bins discretizing every 45 degrees the full circumference. (b) Histogram of detections, before normalizing, obtained in a cell during an interval of 10 minutes.

However, predicting the dominant orientation with the real data at a single point in time t is not possible, since for a single instant we cannot count enough detections to determine which orientations obtain the highest number. Instead, the idea is to compare the most dominant orientation obtained during a particular interval of time (Δt). For the real data we chose the orientation in each cell with the higher number of occurrences over the time interval, the same as we did during model building. For the model prediction, we compute the cumulative probability values in the same interval for all the orientations and pick the one with the highest number.

$$cell_\theta = argmax(\int_{\Delta t} p_\theta(t)), \qquad \theta \in i\frac{2\pi}{k}, \;\; i \in \{0, 1, \dots, k-1\}. \tag{4}$$

5.2 Results

In order to evaluate the model we have calculated the error for both testing sets using multiple values of modelled periodicities m, going from 0, which corresponds to a stationary model, to 4, and for different time intervals: of 0.5, 1, 2, 3, 4 and 6 hours. Since there was no movement of people during night-time in both datasets, we calculated the error only for half a day, i.e. 09:00–21:00 at ATC and 8:00–20:00 at UoL. Since for the ATC dataset we have used 4 days for evaluation, the results shown are the average over all days. The error results for all the combinations of model orders (number of periodicities) and time intervals for both datasets are summarized in Table 1.

It is clear that for the ATC dataset, modeling the time domain with at least one periodicity (i.e. model component) provides us with significantly more accurate prediction of the people movement. For the UoL dataset, the improvement in accuracy of the higher order model over the static one (with m=0) is not as prominent as in the ATC case, but it is still statistically significant. This difference is caused by the lack of clear periodicities present in the environment at UoL as shown in Figures 9 and 10.

In both cases, the models with between 2 and 4 periodicities obtain the best results, each having a similar performance. However, when choosing the optimal value m, if there is no substantial change when adding more periodicities, it is better to stay with lower values so the model remains simpler and faster to calculate. So for this case staying with order 2, or even 1, would be the best choice, which is in accordance with the results in Krajnik et al. [2].

Looking at the interval parameter, it is clear that for both datasets, going for higher intervals times means obtaining lower error values, but in return we suffer from bigger temporal granularity, as the predictions also span longer time intervals.

Table 1: Model errors for both datasets using different order models and time intervals [%]

	ATC dataset						**UoL dataset**					
Interval[h]	**0.5**	**1**	**2**	**3**	**4**	**6**	**0.5**	**1**	**2**	**3**	**4**	**6**
Order 0	45.0	44.2	40.8	38.5	37.2	34.6	40.1	37.5	34.3	32.5	32.6	30.6
Order 1	42.3	40.7	36.9	34.5	33.4	30.9	38.2	36.4	33.4	31.2	31.9	29.2
Order 2	42.0	40.3	36.7	34.0	32.1	**29.8**	39.8	37.6	34.0	31.5	31.3	**28.0**
Order 3	42.1	40.2	36.7	34.1	32.4	30.0	38.8	36.2	33.6	31.3	30.7	28.1
Order 4	41.9	39.8	36.2	33.5	32.0	29.8	38.5	36.5	33.3	31.2	30.8	28.4

5.3 Qualitative evaluation

To better understand the approach and results obtained, we have drawn the orientations for the ground truth and the model predictions for both datasets.

For better visualization and due to space constraints, for the ATC dataset we have plotted the grid only for the central square, although the east corridor also contains relevant flow patterns. The model chosen for presentation in Figure 6 is of order 2, meaning we are taking the two most prominent FreMEn components (corresponding to 24 and 12 hours). The interval time selected is 4 hours, so the ground truth and predictions are calculated from 09:00 to 13:00, from 13:00 to 17:00, and from 17:00 to 21:00. Looking at the ground truth (Fig. 5) one quickly notices that the central area of the square presents a noticeable change in the flow during the day. The flow of people changes from the "West" direction (green) in the morning to the "East" direction (dark blue) in the afternoon. Checking the same intervals for the orientations predicted (Fig. 6), the model was able to capture and learn the periodicities occurring, being capable to predict those changes with a high degree of fidelity for future states.

To understand how the internal representation of a cell works, in Fig. 7 and 8 there are examples of two different cells, one placed in a changing environment like the central square (Fig. 8) and the other one in a location with a constant flow in the same direction during the whole day (Fig. 7), both marked with black circles in Fig. 5 and 6. The plot shows the probabilities calculated for all the orientations of the cell during the day (00:00 to 24:00), taking the same two periodicities in the model as for the representation of the environment. In the dynamic cell, the two predominant orientations which alternate during the day obtain higher probability values than the rest. The "West" one tends to be

Fig. 5: ATC - Ground truth for the intervals (left to right) 09:00 to 13:00, 13:00 to 17:00 and 17:00 to 21:00.

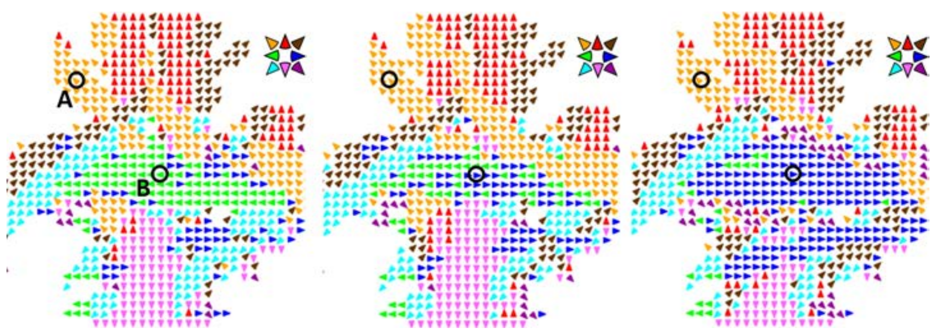

Fig. 6: ATC - Model predictions for the intervals (left to right) 09:00 to 13:00, 13:00 to 17:00 and 17:00 to 21:00.

more likely to occur during the morning, while the "East" orientation overrides the previous one as time passes. This indicates that in the morning the number of people entering the mall tends to be higher than of those leaving, hence the arrows in one direction, but in the afternoon the situation changes to the opposite. However, taking a look at the static cell only one orientation ("North-West") obtained higher probabilities compared with the rest, due to the cell belonging to a location used mainly as an exit corridor.

For the UoL dataset we have chosen to plot also the results taking two periodicities, but in this case picking an interval of 6 hours, obtaining two predictions during the day, from 08:00 to 14:00 and from 14:00 to 20:00. The UoL Figures 9 and 10 do not show a clear rhythmic flow as occurred in the previous dataset, hence the lower increase in performance when going from a static model to one taking into account some periodicities. Despite that fact, our approach is still able to recognize the main flows in the different areas of the corridor where we recorded the dataset.

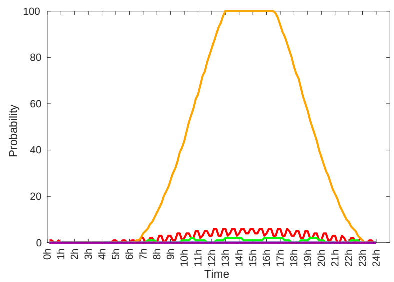

Fig. 7: Probabilities of the 8 orientations for a cell in a corridor with constant flow orientation (cell A in Figs. 5 and 6) during one day.

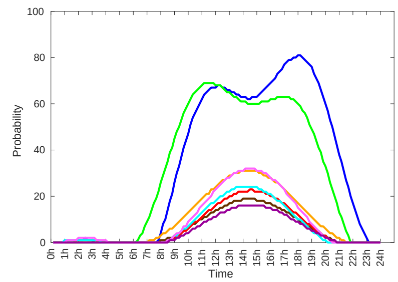

Fig. 8: Probabilities of the 8 orientations for a cell in the central area with changing flow orientation (cell B in Figs. 5 and 6) during one day.

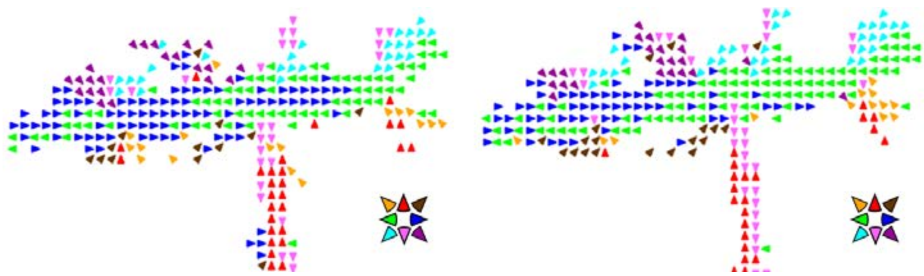

Fig. 9: UoL - Ground truth for the intervals (left to right) 08:00 to 14:00, and 14:00 to 20:00.

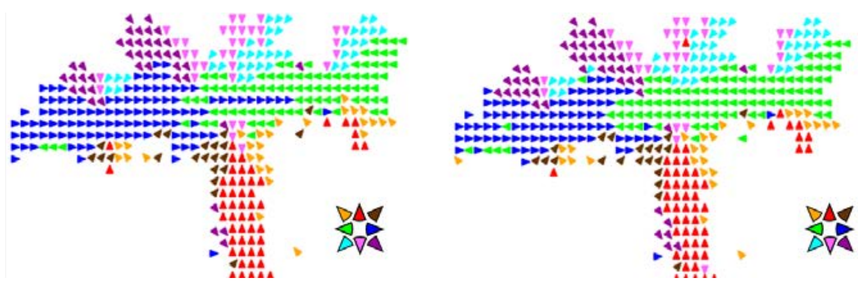

Fig. 10: UoL - Model predictions for the intervals (left to right) 08:00 to 14:00, and 14:00 to 20:00.

6 Conclusion

We have proposed an approach to model the dynamics of human motion in an indoor environment from a long-term perspective. The approach can generate predictions of future crowd movement directions at any given time. The model assumes that human activities are influenced by human habits which follow certain routines. This allows to represent part of the temporal dynamics by means of a combination of harmonic functions. The experiments showed that although not all the dynamics are periodic, taking into account the periodic patterns allows to calculate more accurate predictions of people movement compared to models that neglect the temporal domain.

However, one of the limitations of the developed model lies in the fact that once the number of periodicities m is chosen, all the cells in the map are predicted using the same m. It is unlikely that all cells will experience the same rhythm of activities, so it would be relevant to analyze all the cells individually and choose for each one the best model order to fit the data. This could also be used to cluster the cells into semantic categories according to the types of activities observed.

From a robotics perspective, this representation could be used to take the predicted human motion into account when planning the robot movement as presented in [1]. A navigation algorithm taking into account the people movement would result in more natural and harmonious movement with respect to the crowd, moving with the flow and obtaining fewer face-to-face encounters.

For the experiments presented in this paper, we assumed that the environment is fully observable. However, we know this is unlikely in a real-world scenario, where we would only observe the immediate vicinity of the robot through the robot's own sensors. The capability of FreMEn to work with sparse data [2] allows us to update our models only with the data actually observed by the robot. However, this requires that the robot actively gathers data at locations and times that provide new information to update the spatio-temporal model. Thus, we plan to study robot exploration strategies, which would actively maintain the spatio-temporal representations to keep them up-to-date. Future work will also include testing the model in other scenarios such as warehouse operations to check the method's generalization capabilities.

Acknowledgement

This work has been supported within H2020-ICT by the EC under grant number 732737 (ILIAD), by CSF grant no. 17-27006Y and OP VVV Research Center for Informatics no. CZ.02.1.01/0.0/0.0/16_019/0000765.

References

1. Palmieri, L., Kucner, T.P., Magnusson, M., Lilienthal, A.J., Arras, K.O.: Kinodynamic motion planning on Gaussian mixture fields. In: IEEE International Conference on Robotics and Automation (ICRA), IEEE (2017) 6176–6181
2. Krajník, T., Fentanes, J.P., Santos, J., Duckett, T.: FreMEn: Frequency map enhancement for long-term mobile robot autonomy in changing environments. IEEE Transactions on Robotics (2017)
3. Brscic, D., Kanda, T., Ikeda, T., T. Miyashita, T.: Person position and body direction tracking in large public spaces using 3d range sensors. IEEE Transactions on Human-Machine Systems **43**(6) (2013) 522–534

4. Arbuckle, D., Howard, A., Mataric, M.: Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2002)
5. Mitsou, N., Tzafestas, C.: Temporal occupancy grid for mobile robot dynamic environment mapping. In: Mediterranean Conference on Control Automation. (2007)
6. Sun, L., Yan, Z., Molina, S., Hanheide, M., Duckett, T.: 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. In: IEEE International Conference on Robotics and Automation. (2018)
7. Rudenko, A., Palmieri, L., Arras, O.: Joint long-term prediction of human motion using a planning-based social force approach. In: IEEE International Conference on Robotics and Automation (ICRA). (2018)
8. Bennewitz, M., Burgard, W., Cielniak, G., Thrun, S.: Learning motion patterns of people for compliant robot motion. IJRR **24**(1) (2005) 31–48
9. Wang, Z., Ambrus, R., Jensfelt, P., Folkesson, J.: Modeling motion patterns of dynamic objects by IOHMM. In: Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, IEEE (2014) 1832–1838
10. Kucner, T., J. Saarinen, J., M. Magnusson, M., Lilienthal, A.J.: Conditional transition maps learning motion patterns in dynamics environments. In: IEEE International Conference on Intelligent Robots and Systems. (2013)
11. Saarinen, J., Andreasson, H., Lilienthal, A.J.: Independent Markov chain occupancy grid maps for representation of dynamic environment. In: IEEE Intelligent Robots and Systems, IEEE (2012) 3489–3495
12. Dayoub, F., Cielniak, G., Duckett, T.: Long-term experiments with an adaptive spherical view representation for navigation in changing environments. Robotics and Autonomous Systems (2011)
13. Rosen, D.M., Mason, J., Leonard, J.J.: Towards lifelong feature-based mapping in semi-static environments. In: International Conference on Robotics and Automation (ICRA), IEEE (May 2016) 1063–1070
14. Tipaldi, G.D., Meyer-Delius, D., Burgard, W.: Lifelong localization in changing environments. IJRR (2013)
15. Fentanes, J.P., Lacerda, B., Krajník, T., Hawes, N., Hanheide, M.: Now or later? predicting and maximising success of navigation actions from long-term experience. In: International Conference on Robotics and Automation. (May 2015) 1112–1117
16. Santos, J.M., Krajnik, T., Pulido Fentanes, J., Duckett, T.: Lifelong information-driven exploration to complete and refine 4d spatio-temporal maps. Robotics and Automation Letters (2016)
17. Jovan, F., et al.: A Poisson-spectral model for modelling temporal patterns in human data observed by a robot. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (Oct 2016) 4013–4018
18. Yan, Z., Duckett, T., Bellotto, N., et al.: Online learning for human classification in 3d lidar-based tracking. In: International Conference on Intelligent Robots and Systems (IROS). (2017)
19. Molina, S., Cielniak, G., Krajnik, T., Duckett, T.: Modelling and predicting rhythmic flow patterns in dynamics environments. In: Robotics and Autonomous Systems: Robots Working for and Among Us. (2017)