

Time-varying Pedestrian Flow Models for Service Robots

Tomáš Vintr¹, Sergi Molina², Ransalu Senanayake³, George Broughton¹, Zhi Yan⁴, Jiří Ulrich¹,
Tomasz Piotr Kucner⁵, Chittaranjan Srinivas Swaminathan⁵, Filip Majer¹, Mária Stachová⁶,
Achim J. Lilienthal⁵ and Tomáš Krajník¹

Abstract—We present a human-centric spatio-temporal model for service robots operating in densely populated environments for long time periods. The method integrates observations of pedestrians performed by a mobile robot at different locations and times into a memory efficient model, that represents the spatial layout of natural pedestrian flows and how they change over time. To represent temporal variations of the observed flows, our method does not model the time in a linear fashion, but by several dimensions wrapped into themselves. This representation of time can capture long-term (i.e. days to weeks) periodic patterns of peoples’ routines and habits. Knowledge of these patterns allows making long-term predictions of future human presence and walking directions, which can support mobile robot navigation in human-populated environments. Using datasets gathered by a robot for several weeks, we compare the model to state-of-the-art methods for pedestrian flow modelling.

I. INTRODUCTION

The advances in artificial intelligence, machine vision and computer science, along with the ever-decreasing prices of hardware allowed the introduction of robots into domestic and office environments. These robots are supposed to share their space with people, interact with them, and perform tasks which are considered to be monotonous, tedious or boring. However, introducing mobile robots into human environments faces several challenges.

One of such challenges is the reliability and safety of long-term operation in environments that change over time due to people activity. Unless properly addressed, the environment changes cause gradual deterioration of robot localisation robustness and, in turn, navigation efficiency. The effect of changes can be suppressed by gradual adaptation of the spatial environment models [1], [2], [3], [4] or by explicit

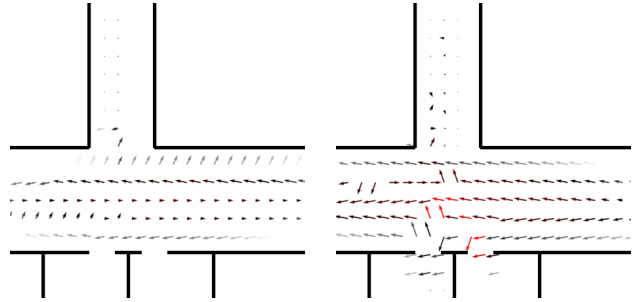


Fig. 1. Directions of pedestrian movement at 9:15 and 16:30 predicted by the proposed model. The arrow lengths correspond to flow intensity, i.e. number of people walking in the directions indicated by the arrows.

representation of time, which allows to model persistence [5], [6], periodicity [7] or more general long-term dynamics [8].

Another challenge is acceptance of the robots by the humans, who might consider the robots to behave in an inappropriate, offensive or even aggressive way. As pointed out in [9], one of the critical aspects of long-term acceptance of mobile robots in human-populated environments was the way they navigate around humans. One of the problems is that nowadays navigation methods represent the environment as a static structure and dynamic objects, such as humans, are treated separately. That assumes a reactive approach, where a robot estimates the people velocities by tracking them and then replans its trajectory. As reported in [10], the errors of state-of-the-art methods exceed 0.4 m for prediction horizons of 1 s, which means that reactive navigation around people still requires a high-speed sense-predict-plan-act loop.

To overcome the limitations of reactive approaches, a robot could learn natural motion patterns from long-term experience [11], [10], and plan its path while anticipating people walking in learned directions even if it does perceive any humans at a given moment. In other words, knowledge of the typical patterns of people movement could improve socially-compliant navigation by planning robot trajectories so that robots would follow the natural flows of people, and avoid congestions and areas where they would cause a nuisance. To address this, several authors [12], [11], [13], [14] proposed models specifically aimed to represent the characteristic movement of people across the operational environment of the robot. The works mentioned above aim at the spatial aspects of pedestrian flows, i.e. they represented the typical directions or velocities at different areas. However, the pedestrian flows themselves are not stationary, but, as shown in [15], [14], their intensities, velocities and

¹Artificial Intelligence Center, Czech Technical University, {name.surname}@fel.cvut.cz

²Lincoln Centre for Autonomous Systems (L-CAS), University of Lincoln, UK, smolinamellado@lincoln.ac.uk

³Stanford University ransalu@stanford.edu

⁴Distributed Artificial Intelligence and Knowledge Laboratory (CIAD), University of Technology of Belfort-Montbéliard (UTBM), France, zhi.yan@utbm.fr

⁵AASS Mobile Robotics and Olfaction Lab, rebro University, Sweden {name.surname}@oru.se

⁶University of Matej Bel in Banská Bystrica, Slovakia maria.stachova@umb.sk

The work is funded by CSF project 17-27006Y STRoLL, the CTU IGA grant No. SGS16/235/OHK3/3T/13, and FR-8J18FR018, PHC Barrande programme under grant agreement No. 40682ZH (3L4AV), and European Union's Horizon 2020 research and innovation programme under grant agreement No. 732737 (ILIAD). Special thanks go to Tomasz Kucner and Chittaranjan Swaminathan, who provided us with the predictions of their CLiFF-map.

directions strongly depend on time. A robot, capable of predicting future distributions of pedestrian flows, would be able to plan its collision-free, socially-compliant trajectories in advance, minimising the likelihood of having to alter its plan in order to avoid collisions.

We present a method capable of learning the natural flows of people and *how they change over time*. The core idea of the method is to model the time domain by several dimensions wrapped into themselves, which can efficiently represent periodicities of the pedestrian flow characteristics. Using a real-world dataset spanning over several weeks, we compare the method’s predictive performance to state of the art algorithms for pedestrian flow modelling. To promote reproducible and unbiased comparison, the dataset, code and supporting materials publicly available [16], and the comparisons are performed using data provided by the authors of the aforementioned methods.

II. RELATED WORK

The ability to autonomously move across space, i.e. navigation, is a pivotal competence of mobile robots. For navigation in an efficient, reliable and safe manner, a robot needs to be able to determine its position, position of its destination and it has to be able to plan its trajectory to avoid collisions. Since the accuracy of self-localisation and efficiency of the motion planning depend on the quality of the robot knowledge about its operational environment, i.e. on the fidelity and faithfulness of its internal representation of the surrounding world. Thus, a significant research effort was aimed at methods for building large-scale and accurate maps of the environment [17]. While most of the methods developed so far are aimed to model the environment by static structures, deployment of robots in dynamic or changing environments raised the need to represent the environment dynamics as well.

In the mapping and localisation community, the effects of the environment dynamics were studied mainly from the perspective of localisation reliability, which gradually deteriorates if the environment changes are neglected [18]. To deal with the changes, some approaches proposed to gradually adapt the maps by incrementally replacing their elements [4], by remapping the areas which changed [3], or by allowing multiple representations of the same location [2], by identifying the invariant characteristics of the world [19] or by general schemes to incrementally update continuous maps [20] using Bayesian techniques.

Another stream of the research proposed to exploit the observed dynamics to obtain more information about the environment. For example, Ambrus et al. [21] presented a method that can identify clusters of 3d data which changed between subsequent observations of the same location. Subsequent work demonstrated, that these clusters can be used for autonomous discovery of object models from the spatial changes observed [22].

Other researchers proposed to process the changes observed to obtain information about the temporal aspects of the long-term environment dynamics. For example, Dayoub

et al. [23] and Rosen et al. [24] proposed to interpret the changes in order to obtain models that would characterise the persistence of the environment states. The persistence models were then used to predict which elements of the environment should be used for localisation. The work of Tipaldi et al. [25] proposed to represent occupancy of grid cells by a Hidden Markov model, which also characterises the state persistence. Finally, the work of Krajník et al. [7] proposed to model the probability of environment states in the spectral domain, which captures cyclic (daily, weekly, yearly) patterns of environmental changes, which are often induced by humans. The concept of Frequency Map Enhancement (FreME_n) [7] was applied to a variety of scenarios, and was shown to improve both localisation [7], motion planning [26] and human-robot interaction [27].

With the exception of the FreME_n, the aforementioned works were aimed at the problem of localisation in environments with undergoing a slow change. However, a substantial part of the natural environment dynamics is constituted by the motion of people in the robot vicinity, which requires the robot to plan its trajectory with respect of the people around. As stated in [28], knowledge of the general pedestrian flows will allow the robots to move in a socially compliant manner, increasing not only their efficiency, but also their acceptance by the public. To characterise the flows, the work of [11] proposes to extend an occupancy grid model by propagating information about the changes in the adjacent cells. Another discrete, grid-based model can be found in [12], where authors predict the paths of people based on an input-output Markov model associated with each cell. The authors of [29] assume the pedestrian flows change over time in a periodic fashion, and associate each cell of their grid with directional information enhanced by FreME_n [7].

Other streams of research tried to model the pedestrian movement by continuous representations. For example, in [30] authors show how to modify the plans of the robot motion by taking into account the long-term observations of people movement. However, this approach did not address the multimodality of pedestrian motion distribution making it unable to model motion in opposing directions. Later work [31] improved this particular aspect and presented a method that can model multimodal distributions of pedestrian movement directions. Kucner et al. also improved their approach in [11] and proposed a continuous representation in [32]. To model speed and direction of people, [33] introduced an expectation-maximisation scheme based on the Independent von MisesGaussian distributions [34]. They also showed that the model of the movement of people could be used to achieve more efficient navigation of the robot through human crowds [35].

Similarly to the work [36] and [37], which demonstrate that incorporation of techniques to model periodic aspects of time into continuous spatial models results in powerful predictive representations, and [14], which shows the benefit of periodic temporal representations for pedestrian flow modelling, we propose to model the pedestrian flows by continuous spatio-temporal representation, which allows to

model how the flows change over time.

III. METHOD DESCRIPTION

The aim of the proposed method is to find an estimation of Bernoulli distribution of an occurrence of spatio-direction-temporal events at time t_i on position x_i, y_i with the speed v_i at angle ϕ_i . Since it is not possible to obtain multiple data with the same t_i by performing additional observations, one cannot calculate the Bernoulli distribution in a straightforward, frequentist way. This is caused by the fact that the modeled events are sparse, and the process generating them is not stationary. To deal with the problem, we proposed in our previous works to use a “warped-hypertime” projection of the time line into a closed subset of multidimensional vector space, where each pair of dimensions would represent one periodicity [38], [39], [40], [41], [42]. Then, we create a model characterising the probability distribution of spatio-direction-temporal events in the vector space extended by the warped hypertime. To do so, we estimate distributions the spatio-temporal events projected into the higher dimensional vector space using Expectation Maximisation algorithm for estimating Gaussian Mixture Models (EM GMM).

The idea behind the aforementioned projection is that events which occur with the same periodicity will form clusters in the hypertime space even if they are separated by long intervals of time. An intuitive example, shown in Figure 2 for the case of $T = 1\text{day}$, is that hypertime associates the given observations with time of the day. Figure 2 shows an example output of a office building corridor. Here, the detections overlap at mornings and evenings, when people leave and enter work, while the non-detections form clusters around noon and midnight, when people work or the building is vacant.

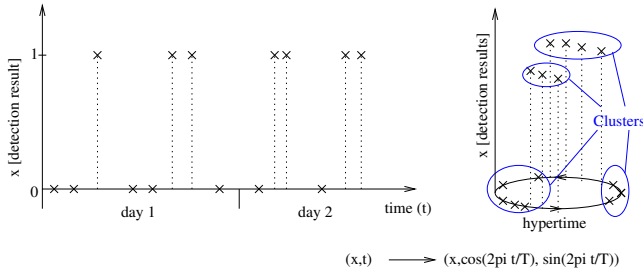


Fig. 2. Example of the warped hypertime projection on binary data (person detection) and one periodicity T . The numbers x_i observed at t_i are projected into a 3d vector space as $(\mathbf{x}_i, \cos(2\pi t_i/T), \sin(2\pi t_i/T))$, where they form clusters because they exhibit a periodic behaviour with a period T . The warped hypertime dimensions define a base of a cylinder, and values of x_i define a cylinder side.

A. Warped Hypertime Projection

Let us assume that the robot pedestrian tracking system robot provides us with vectors $(x_i, y_i, v_i, \phi_i, t_i)$, indicating the detected people positions, velocities and orientations as well as the timestamp of the observation. To avoid complications

caused by the ambiguity of angles, we transform the aforementioned vector to $(x_i, y_i, v_i \cos \phi_i, v_i \sin \phi_i, t_i)$ and denote it as (\mathbf{x}_i, t_i) .

Let us have a set of detections $D(\mathbf{x}_i, t_i)$, $i = 1 \dots n$ of occurrences and non-occurrences of some events at a location \mathbf{x}_i at time t_i , where $D(\mathbf{x}_i, t_i) = 1$ for detected and $D(\mathbf{x}_i, t_i) = 0$ for non-detected occurrences of the studied phenomenon. To determine the parameters the warped hypertime projection, we need to identify the most distinctive temporal periodicities in the provided data. To do so, we create a time series $R(t_i) = D(\mathbf{x}_i, t_i)$ by neglecting spatial components of the detections and apply the spectral decomposition method derived from the Frequency Map Enhancement [7]. First, we estimate the longest periodicity present in the data T_{max} and then we calculate Υ periocities as $T_\tau = T_{max}/\tau$, where $\tau = 1 \dots \Upsilon$. Then, we calculate the prominence of each periodicity as:

$$\gamma_\tau = \frac{1}{n} \sum_{i=1}^n R(t_i) e^{-j2\pi t_i/T_\tau}. \quad (1)$$

Since the experiments performed in [7] indicate that the most accurate predictions in human-populated environments are obtained by modeling 2-3 periodicities, we select two periodicities with the highest γ_τ and denote them as $T_{1,2}$. Then, we project every measurement (\mathbf{x}_i, t_i) into the new vector space by:

$$^@ \mathbf{x}_i = \left(\mathbf{x}_i, \cos \frac{2\pi t_i}{T_1}, \sin \frac{2\pi t_i}{T_1}, \cos \frac{2\pi t_i}{T_2}, \sin \frac{2\pi t_i}{T_2} \right). \quad (2)$$

B. Model of the probability distribution

We assume that the time-dependent occurrences of the phenomenon (\mathbf{x}_i, t_i) projected into the warped hypertime space as $^@ \mathbf{x}_i$ are distributed in a way which allows to model their distribution by Gaussian mixtures. To model the Bernoulli distribution of $D(\mathbf{x}_i, t_i)$, we split the dataset to occurrences and non-occurrences (these are mutually exclusive), and we build the mixture models of occurrences $GMM_1(^@ \mathbf{x}_i)$ and non-occurrences $GMM_0(^@ \mathbf{x}_i)$ in separate using an Expectation Maximisation algorithm. Thus, we obtain two models, characterised by cluster weights $\alpha_{\{0,1\}j}$, cluster centres $\mathbf{c}_{\{0,1\}j}$ and covariances $\Sigma_{\{0,1\}j}$. These allow to determine the probability that a given projected sample $^@ \mathbf{x}$ belongs to a particular cluster using a χ^2 distribution:

$$P_{\{0,1\}j} = 1 - P \left[Q \leq (^@ \mathbf{x} - \mathbf{c}_{\{0,1\}j})^T \Sigma_{\{0,1\}j}^{-1} (^@ \mathbf{x} - \mathbf{c}_{\{0,1\}j}) \right], \quad (3)$$

where $Q \sim \chi^2(d)$ and d is dimensionality of the constructed vector space. The overall probability $M_{\{0,1\}}(^@ \mathbf{x})$ of generating an occurrence of $^@ \mathbf{x}$ by a mix of distributions $GMM_{\{0,1\}}(^@ \mathbf{x})$ is estimated as:

$$M_{\{0,1\}}(^@ \mathbf{x}) = \sum_{j=1}^c \alpha_{\{0,1\}j} P_{\{0,1\}j}. \quad (4)$$

Then, the probability of the occurrence of (\mathbf{x}, t) is given by the following ratio based on its hypertime projection $^@ \mathbf{x}$:

$$M(\mathbf{x}, t) = \frac{M_1(@\mathbf{x})}{M_1(@\mathbf{x}) + M_0(@\mathbf{x})}. \quad (5)$$

IV. EVALUATION

A. Dataset

The approach described above was evaluated using a dataset collected at the department of computer science at the University of Lincoln. The data recording was performed by a Pioneer 3-AT mobile robot equipped with a 3D lidar (Velodyne VLP-16) and a 2D lidar (Hokuyo UTM-30LX), using a reliable person detection method [43].

During the data collection, the robot remained stationary in a T-shaped junction, which allowed its sensors to scan the three connecting corridors simultaneously, covering a total area of around 75 m² (Fig. 3). However, since the robot could not stay at the corridor overnight due to safety rules, and it was needed by other researchers occasionally, we did not collect the data on a full 24/7 basis. Instead, the data collection was performed during ~10 hour long sessions starting before the usual working hours. Recharging of the batteries was performed overnight, where the building is vacant, and there are no people on the corridors.

The resulting dataset is composed of 9 data-gathering sessions recorded over four weeks. A typical session contains approximately 30000 detections of people walking in the monitored corridors. Every detection is represented by a vector $(t, x, y, \phi), v$, the position, orientation, and speed of detected human in time. Similar to [44], we added 70000 “no detection” vectors of the positions, orientations, and speeds, where no human was detected. As some of the methods in comparison do not model the speed, this value was set to $v = 1.0$ for every measurement. For detailed information about individual methods used in comparison, see section IV-C. The structured overview of the properties of individual methods can be seen in Table I.

The 3D lidar has 16 scan channels with a 360° horizontal and 30° vertical field-of-view, and was mounted at the height of 0.8 m from the floor on the top of the robot (Fig. 3 left), which allows us to have a perspective that covers the entire environment for data collection (Fig. 3 right). All people appearing in the corridor are detected and tracked in the 3D lidar’s frame of reference. More specifically, the 3D point cloud generated by the Velodyne lidar is first segmented into different clusters using an adaptive clustering method [43], then an offline trained SVM-based classifier was used for human classification. The 2D positions of the people are subsequently fed into a robust multi-target tracking system [45] using Unscented Kalman Filter (UKF) and Nearest Neighbour Joint Probability Data Association method (NNJPDA), and the human-like trajectories (in XY-plane) are eventually generated and recorded.

B. Evaluation methodology

In accordance with [46], we divided the dataset into a training and testing subset, where the training dataset consisted of seven days from three weeks and the test dataset

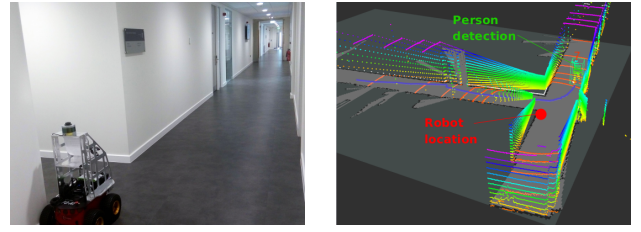


Fig. 3. Photo of the UoL dataset data collection setup: Robot location in the corridor and example of a person walking as seen by the 3D lidar.

consisted of two days measured out of the time interval of training dataset. We chose two different criteria to measure the quality of model. The first is root-mean-square deviation (RMSD) [47] between model predictions $M(\mathbf{x}_i, t_i)$ and test dataset values $D(\mathbf{x}_i, t_i)$

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (M(\mathbf{x}_i, t_i) - D(\mathbf{x}_i, t_i))^2}, \quad (6)$$

which is widely used in the time series forecasting [48].

The second criteria used is the level of similarity between human motion distributions, occurring at certain times and positions, obtained from the 2 test days and the ones predicted by the model. This metric is focused on how well the model is able to predict how a person would move in case it is found, rather than how likely the robot is going to find one.

In order to do that, we have defined a spatial and temporal grid in order to cover the full map and the whole 2 test days. The different approaches are able to provide a probability motion distribution for any point in time/space, however obtaining this same distribution with the ground truth data to do the comparison at a single time instance is not possible. The reason is that for an instance of time we do not have enough detections in order to build a meaningful distribution. Instead, the idea is to compare the distribution obtained from the test data during a defined interval of time. In our evaluations, we have used a spatial grid taking points every 1 meter in x and y directions, and 10 minutes long time intervals.

To make the comparison between the predicted and ground truth histograms for each interval and position, we have used the Chi-square distance. This distance indicates the level of similarity between two discrete distributions or histograms, so the higher the distance the less accurate is our model prediction compared with the test data. The total Chi-square distance of the map for a single interval is defined as:

$$distance_{map} = \sum_{i=1}^n \sum_{b=1}^k \frac{(x_b - y_b)^2}{(x_b + y_b)}, \quad (7)$$

where n is the number of positions, k is the number of angular bins for the direction of people motion in the cells (in our case we have chosen $k = 8$ taking the angles 0, 45, 90, 135, 180, 225, 270 and 315 degrees as values for each bin), x_b is the value of bin b in the predicted orientation histogram, and y_b is the value of the same bin b obtained from the ground truth.

C. Methods compared in the experiment

1) *WHyTe*: There are two parameters, which affect the quality of WHyTe - the number of clusters c and the set of periodicities. The recent experiments showed, that the number of clusters could be relatively small (usually up to 9) [42], and it seems, that the number of clusters is in relation with the topological structure of the space [41]. For this dataset from T-junction we chose $c = 3$ clusters. The second parameters can be derived from data iteratively, but recent experiments showed [42], [41], that the quality of prediction do not usually grow with more than 3 added hypertime circles. We selected the basic set of periodicities as proposed in [7], and found out, that there were two strongly prominent components in the training data, which we used in our method.

2) *STeF-Map*: STeF-Map, which stands for Spatio-Temporal Flow Map, is a representation that models the likelihood of motion directions on a grid-based map by a set of harmonic functions, which capture long-term changes of crowd movements over time. The underlying geometric space is represented by a grid, where each cell contains k temporal models, corresponding to k discretised orientations of people motion through the given cell over time. Since the total number of temporal models, which are of a fixed size, is $k \times n$ where n is the total number of cells, the spatio-temporal model does not grow over time regardless of the duration of data collection. The temporal models, which can capture patterns of people movement, are based on the FreMEn framework [7]. 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 idea is to treat a measured state as a signal, decompose it by means of the Fourier Transform, and obtain 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 likelihood of that state at a given time in the past or future.

This model assumes that it is provided with people detection data, comprising the person position, orientation and timestamp of the detection (x, y, α, t) . When building the model, the x, y positions are discretised and assigned to the corresponding cell and the orientation α is assigned to one of the k bins, whose value is incremented by 1. After a predefined interval of time, the bins values are normalised, and the results are used to update the spectra of the temporal models. Then, the bin values are reset to 0, and the counting starts again.

In order to retrieve the behaviour of human movement through a given cell at a certain time t (which can be at the future or at the past), the likelihood for each discretised orientation associated with a cell can be computed as:

$$p_\theta(t) = p_0 + \sum_{j=1}^m p_j \cos(\omega_j t + \varphi_j), \quad (8)$$

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 covers periodicities ranging from 14 h to 1 week with the following distribution:

$$\omega_s = \frac{3600 \cdot 24 \cdot 7}{s}, \quad s \in 1, 2, 3, \dots, 12. \quad (9)$$

3) *Directional grid maps*: Directional grid maps (DGM) [49] are designed to model the *directional uncertainty* of dynamic environments. The inputs to the model are directions of objects at different locations of the environment, and the outputs are a set continuous probability density functions indicating most probable directions dynamic objects move at various locations of the environment. In order to build a DGM, firstly, the 2D or 3D environment is divided into a fixed-sized grid. Then, a mixture of von Mises distribution is assigned to each cell to model the multimodal angular uncertainty. Analogous to a Gaussian distribution, however with a limited $[-\pi, +\pi]$ support, a von Mises distribution is controlled by its mean angle and concentration (inverse variance) parameters. The number of von Mises components for each mixture is determined by the number of density-wide clusters using the DBSCAN algorithm. Having initialised the von Mises distributions with the cluster centers, the parameters are learned using Expectation-Maximization (EM). In experiments, it takes 1 to 4 iterations to converge the EM. Since the directional grid maps are not designed to deal with the spatiotemporal domain with periodic patterns, in this experiment, the temporal domain is also discretised every 15 minutes in addition to the $2 \text{ m} \times 2 \text{ m}$ spatial discretisation. For the purpose of this experiment, as a proxy, we attempt to estimate the people density by considering the cells where the initial set of mixture parameters change with time. Therefore, the proxy count probabilities are always either 0 or 1 and not the exact people density. In the future, it is possible to replace the von Mises distribution with a Gaussian distribution or replace the Bernoulli likelihood in [20] with a Gaussian likelihood to accurately model such spatial density estimations.

4) *CLiFF-Map Model*: Circular Linear Flow Field map (CLiFF-map) [33] is a technique for encoding patterns of movement as a field of Gaussian mixtures. They can be combined with semi-wrapped Gaussian mixture models (SWGMM) to model multi-modal motion.

$$p(\mathbf{V}|\boldsymbol{\xi}) = \sum_{j=1}^J \pi_j \mathcal{N}_{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j}^{\mathcal{S}\mathcal{W}}(\mathbf{V}) \quad (10)$$

with $\sum_{j=1}^J \pi_j = 1$.

This uses a semi-wrapped normal distribution, distributed along the circumference and height of a cylinder. It's represented as a semi-wrapped normal distribution. It can be derived from:

$$\mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{\mathcal{S}\mathcal{W}}(\mathbf{V}) = \sum_{k \in \mathbb{Z}} \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} \left(\begin{bmatrix} \theta \\ \rho \end{bmatrix} + 2\pi \begin{bmatrix} k \\ 0 \end{bmatrix} \right). \quad (11)$$

where $\mathbf{V} = (\theta, \rho)$ represents the instantaneous velocities, $\theta \in [0, 2\pi)$ is the direction, and $\rho \in \mathbb{R}^+$ the speed.

TABLE I
QUALITATIVE COMPARISON OF METHODS

Name	References	Dynamics		Representation					Complexity	
		long-term	short-term	time	space	intensity	direction	speed	Memory [kB]	Train time [s]
WHyTe	[39]	✓	×	C	C	C	C	C	2	60
STeF	[14]	✓	×	C	D	×	D	×	140	20
DGM	[49]	✓	×	D	D	C	C	×	20	72
CLiFF	[33]	×	×	×	D	C	C	C	6k	>10 ⁴
LSTM	[50]	×	✓	C	C	C	C	C	900	>10 ⁶

Note 1: In the ‘Representation’ columns, C stands for the continuous, and D for the discrete representation of variables provided by method.

Note 2: CLiFF-map was developed using Matlab and other methods are based on the Python language.

5) *LSTM*: We also implemented a deep-learning model for a point of comparison. A long short-term memory [50] neural network was built using Keras atop of the TensorFlow library. It consisted of 4 layers of 50 LSTM units followed by a fully connected layer with 72000 trainable parameters. It was then trained to convergence on the training set. It is important to note however that this method consumed significantly more computing power, both during training and prediction.

D. Evaluation results

The results of the evaluation, which are summarised in Table II, indicate that WHyTe achieved the lowest root mean squared error and STeF achieved the lowest χ^2 error. Although DGM is not designed to estimate the probability of occurrence and it only returns 0 or 1, and therefore RMSE is not a proper measurement for this method, it was comparable with WHyTe in χ^2 statistics. CLiFF-map predicts directions on specific position better than WHyTe. Furthermore, to fit the evaluation procedure, CLiFF-map was discretized into eight orientation bins. WHyTe-0 and STeF-0 are static models without any information about the time during the training phase. We can see the impact on predictive performance when we compare them to their equivalents WHyTe and STeF, which model the periodicities over the time domain. The results in Table II show, that WHyTe models directions with taking the probability of human occurrence into account. Modelling the joint probability of occurrences and directions is the crucial property of the WHyTe, which is supposed to provide apriori knowledge of the dynamics of human-populated environments to the autonomous service robots.

We also include an LSTM model as it is commonly requested by reviewers even for cases that are not suitable for LSTM. LSTM model was trained using four NVIDIA Tesla V100 SXM2 32GB for two and ten hours respectively, and their model sizes were 900 KiB. However, as the LSTM is tailored for short-term prediction only, its predictions quickly converged to the mean probability of people directions across the entire training dataset (both spatially and temporally).

WHyTe, STeF and DGM were developed using Python language, their training on regular personal notebooks lasted about one minute, and the model sizes are 2 KiB, 140 KiB, and 20 KiB respectively, which indicates, that they could be applied in real robotic tasks. It should be noted, that model created by WHyTe is smaller by magnitude(s) to its competitors, which is an important attribute for building models over large areas.

TABLE II
PREDICTION ERRORS OF THE EVALUATED MODELS AND DATASETS

Testing sets Criterion	Days		Nights		Days and nights	
	RMSE	χ^2	RMSE	χ^2	RMSE	χ^2
WHyTe-0	0.49	23.0	0.46	0.9	0.48	23.9
WHyTe	0.49	23.4	0.00	0.2	0.40	23.6
STeF-0	0.65	10.0	0.07	16.0	0.53	26.0
STeF	0.57	10.6	0.02	8.1	0.46	18.7
DGM	0.70	25.5	0.83	0.0	0.75	25.5
CLiFF	0.60	15.5	0.16	9.2	0.50	24.7
LSTM	0.57	25.5	0.22	0.0	0.48	25.5

V. CONCLUSION

We propose an approach capable to represent pedestrian flows and how they change over time. Unlike methods, that perform predictions based on recent observations, the model presented can predict the pedestrian flow intensity and direction using observations gathered days to weeks before the prediction. Instead of short-term predictions based on actual observations, which force the robot to react to current situation in an suboptimal manner, a robot utilising our method would be able to anticipate human presence and movement direction from long-term observations and plan its trajectory to minimise the need to evade people. This helps to avoid situations, where the robot interferes with the natural pedestrian flows in its operational area, allowing seamless, socially-acceptable navigation in human-populated environments.

The proposed representation is based on the idea of warped hypertime (WHyTe), which projects the time into a constrained subset of a multidimensional vector space, constructed to reflect the patterns of the human habits.

We evaluated the presented method on a real dataset, gathered over the course of four weeks and compared its predictive accuracy to state-of-the-art methods provided by their authors. Two criteria were used: RMSE, which reflects the ability to predict the joint probability of the presence and the direction of pedestrian movement, and χ^2 statistics, that reflects the ability to predict the conditional probability of the movement direction given that a human is present at some specific position and time. Although the proposed method was not able to compete with STeF and CLiFF methods in χ^2 statistics, it achieved the best prediction in terms of RMSE. This indicates that while the method is not as good in predicting the flow directions, it has a superior performance in predicting intensities of the pedestrian flows. Moreover, we showed, that our method is by the magnitude(s) smaller

compared to the other ones, indicating its suitability to model large-scale environments.

In the future, we will evaluate the impact of the methods used in this comparison on the ability of robots to generate collision-free trajectories in advance.

REFERENCES

- [1] P. Biber and T. Duckett, "Experimental analysis of sample-based maps for long-term slam," *The International Journal of Robotics Research*, vol. 28, no. 1, pp. 20–33, 2009.
- [2] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013.
- [3] K. Konolige and J. Bowman, "Towards lifelong visual maps," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 1156–1163.
- [4] S. Hochdörfer and C. Schlegel, "Towards a robust visual slam approach: Addressing the challenge of life-long operation," in *Advanced Robotics, 2009. ICAR 2009. International Conference on*. IEEE, 2009, pp. 1–6.
- [5] G. D. Tipaldi, D. Meyer-Delius, and W. Burgard, "Lifelong localization in changing environments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1662–1678, 2013.
- [6] D. M. Rosen, J. Mason, and J. J. Leonard, "Towards lifelong feature-based mapping in semi-static environments," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1063–1070.
- [7] T. Krajník, J. P. Fentanes, J. M. Santos, and T. Duckett, "Fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments," *IEEE Transactions on Robotics*, vol. 33, no. 4, pp. 964–977, 2017.
- [8] B. Song, W. Chen, J. Wang, and H. Wang, "Long-term visual inertial slam based on time series map prediction," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.
- [9] D. Hebesberger, T. Koertner, C. Gisinger, and J. Pripfl, "A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults," *International Journal of Social Robotics*, vol. 9, no. 3, pp. 417–429, 2017.
- [10] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett, "3dof pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–7.
- [11] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, "Conditional transition maps: Learning motion patterns in dynamic environments," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 1196–1201.
- [12] Z. Wang, R. Ambrus, P. Jensfelt, and J. Folkesson, "Modeling motion patterns of dynamic objects by iohmm," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), 14-18 Sept. 2014, Chicago, IL*. IEEE conference proceedings, 2014, pp. 1832–1838.
- [13] C. S. Swaminathan, T. P. Kucner, M. Magnusson, L. Palmieri, and A. J. Lilienthal, "Down the cliff: Flow-aware tralatory planning under motion pattern uncertainty," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7403–7409.
- [14] S. Molina, G. Cielniak, and T. Duckett, "Go with the flow: Exploration and mapping of pedestrian flow patterns from partial observations," in *Proc. of Int. Conference on Robotics and Automation (ICRA)*, 2019.
- [15] D. Bršćic, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.
- [16] T. Krajník, "The frequency map enhancement (FreME) project repository," <http://fremen.uk>.
- [17] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [18] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajník, "Artificial intelligence for long-term robot autonomy: a survey," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023–4030, 2018.
- [19] S. Lowry and M. J. Milford, "Supervised and unsupervised linear learning techniques for visual place recognition in changing environments," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 600–613, 2016.
- [20] R. Senanayake and F. Ramos, "Bayesian hilbert maps for dynamic continuous occupancy mapping," in *Conference on Robot Learning*, 2017, pp. 458–471.
- [21] R. Ambrus, N. Bore, J. Folkesson, and P. Jensfelt, "Meta-rooms: Building and maintaining long term spatial models in a dynamic world," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [22] T. Faeulhammer *et al.*, "Autonomous learning of object models on a mobile robot," *Robotics and Automation Letters*, 2016.
- [23] F. Dayoub, G. Cielniak, and T. Duckett, "Long-term experiments with an adaptive spherical view representation for navigation in changing environments," *Robotics and Autonomous Systems*, 2011.
- [24] D. M. Rosen, J. Mason, and J. J. Leonard, "Towards lifelong feature-based mapping in semi-static environments," in *ICRA*, 2016.
- [25] G. D. Tipaldi *et al.*, "Lifelong localization in changing environments," *The International Journal of Robotics Research*, 2013.
- [26] J. P. Fentanes, B. Lacerda, T. Krajník, N. Hawes, and M. Hanheide, "Now or later? predicting and maximising success of navigation actions from long-term experience," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 1112–1117.
- [27] M. Hanheide *et al.*, "The When, Where, and How: An Adaptive Robotic Info-Terminal for Care Home Residents A long-term Study," in *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2017.
- [28] T. P. Kucner, "Probabilistic mapping of spatial motion patterns for mobile robots," Ph.D. dissertation, Örebro University, 2018.
- [29] S. Molina, G. Cielniak, T. Krajník, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Annual Conference Towards Autonomous Robotic Systems*. Springer, 2018, pp. 135–146.
- [30] S. T. O'Callaghan, S. P. N. Singh, A. Alempijevic, and F. T. Ramos, "Learning navigational maps by observing human motion patterns," *2011 IEEE International Conference On Robotics And Automation (ICRA)*, 2011.
- [31] L. McCalman, S. O'Callaghan, and F. Ramos, "Multi-modal estimation with kernel embeddings for learning motion models," *2013 IEEE International Conference On Robotics And Automation (ICRA)*, pp. 2845–2852, 2013.
- [32] T. Kucner, M. Magnusson, E. Schaffernicht, V. Hernandez Bennetts, and A. Lilienthal, "Tell me about dynamics!: Mapping velocity fields from sparse samples with semi-wrapped gaussian mixture models," in *Robotics: Science and Systems Conference (RSS 2016), Workshop: Geometry and Beyond-Representations, Physics, and Scene Understanding for Robotics, University of Michigan, Ann Arbor, MI, USA, June 18-22, 2016*, 2016.
- [33] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1093–1100, 2017.
- [34] A. Roy, S. K. Parui, and U. Roy, "A mixture model of circular-linear distributions for color image segmentation," *International Journal of Computer Applications*, vol. 58, no. 9, 2012.
- [35] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on gaussian mixture fields," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 6176–6181.
- [36] R. Senanayake, O. Simon Timothy, and F. Ramos, "Predicting spatio-temporal propagation of seasonal influenza using variational gaussian process regression," in *AAAI*, 2016, pp. 3901–3907.
- [37] A. Tompkins and F. Ramos, "Fourier feature approximations for periodic kernels in time-series modelling," in *AAAI Conference on Artificial Intelligence*, 2018.
- [38] T. Vintr, S. Molina, J. Fentanes, G. Cielniak, T. Duckett, and T. Krajník, "Spatio-temporal models for motion planning in human populated environments," in *Student Conference on Planning in Artificial Intelligence and Robotics*, 2017.
- [39] T. Vintr, K. Eyisoy, and T. Krajník, "A practical representation of time for the human behaviour modelling," *Forum Statisticum Slovaca*, vol. 14, pp. 61–75, 2018.
- [40] T. Vintr, K. Eyisoy, V. Vintrová, Z. Yan, Y. Ruichek, and T. Krajník, "Spatiotemporal models of human activity for robotic patrolling,"

- in *International Conference on Modelling and Simulation for Autonomous Systems*. Springer, 2018, pp. 54–64.
- [41] T. Vintr, Z. Yan, T. Duckett, and T. Krajník, “Spatio-temporal representation for long-term anticipation of human presence in service robotics,” in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, to appear.
 - [42] T. Krajník, T. Vintr, S. Molina, J. P. Fentanes, G. Cielniak, O. Mozos, and T. Duckett, “Wrapped hypertime representations for long-term autonomy of mobile robots,” *IEEE Robotics and Automation Letters*, 2019, in review.
 - [43] Z. Yan, T. Duckett, and N. Bellotto, “Online learning for human classification in 3d lidar-based tracking,” in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, 2017, pp. 864–871.
 - [44] S. T. OCallaghan and F. T. Ramos, “Gaussian process occupancy maps,” *The International Journal of Robotics Research*, vol. 31, no. 1, pp. 42–62, 2012.
 - [45] N. Bellotto and H. Hu, “Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of bayesian filters,” *Autonomous Robots*, vol. 28, pp. 425–438, 2010.
 - [46] M. Oliveira, L. Torgo, and V. S. Costa, “Evaluation procedures for forecasting with spatio-temporal data,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 703–718.
 - [47] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
 - [48] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.
 - [49] R. Senanayake and F. Ramos, “Directional grid maps: modeling multimodal angular uncertainty in dynamic environments,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3241–3248.
 - [50] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.