Spatio-temporal representation for long-term anticipation of human presence in service robotics

4 authors:

Tomas Vintr
Czech Technical University in Prague
14 PUBLICATIONS   31 CITATIONS

SEE PROFILE

Zhi Yan
Université de Technologie de Belfort-Montbéliard
38 PUBLICATIONS   347 CITATIONS

SEE PROFILE

Tom Duckett
University of Lincoln
206 PUBLICATIONS   4,093 CITATIONS

SEE PROFILE

Tomáš Krajník
Czech Technical University in Prague
108 PUBLICATIONS   1,836 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Swarm Robotics View project

Project COSMOS View project

Spatio-temporal representation for long-term anticipation of human presence in service robotics

Tomáš Vintr[1], Zhi Yan[2], Tom Duckett[3] and Tomáš Krajník[1]

Abstract— We propose an efficient spatio-temporal model for mobile autonomous robots operating in human populated environments. Our method aims to model periodic temporal patterns of people presence, which are based on peoples' routines and habits. The core idea is to project the time onto a set of wrapped dimensions that represent the periodicities of people presence. Extending a 2D spatial model with this multidimensional representation of time results in a memory efficient spatio-temporal model. This model is capable of long-term predictions of human presence, allowing mobile robots to schedule their services better and to plan their paths. The experimental evaluation, performed over datasets gathered by a robot over a period of several weeks, indicates that the proposed method achieves more accurate predictions than the previous state of the art used in robotics.

I. INTRODUCTION

The technology has reached the state where autonomous service robots are deployed for extended time periods in human populated environments, where they assist people in their daily chores [1] or help those who need special care [2].

Hawes *et al.* [1] describe the development of robots deployed for long periods in several different scenarios. In one scenario, a mobile info-terminal was required to offer information to clients of a care home. For example, the robot displayed a lunch menu near a cafeteria or provided directions to visitors around the care home lobby. In both cases, the robot had to reach a given spot before the people arrived there to utilise the information provided. This was done not only to provide the information on time but to avoid problems when navigating between people and causing a nuisance by crossing their path. [1], [3] demonstrated that a robot that is able to anticipate human behaviour and schedule its services accordingly accomplishes richer interactions with humans and is considered to be more helpful by the humans it shares its space with. However, the study [3] pre-determined the locations at which the robot provided its services and allowed it only to select the time when it provided those services. As concluded by [3], it

[1]Artificial Intelligence Center, Czech Technical University. {tomas.vintr, tomas.krajnik}@fel.cvut.cz

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

[3]Lincoln Centre for Autonomous Systems (L-CAS), University of Lincoln, UK. tduckett@lincoln.ac.uk
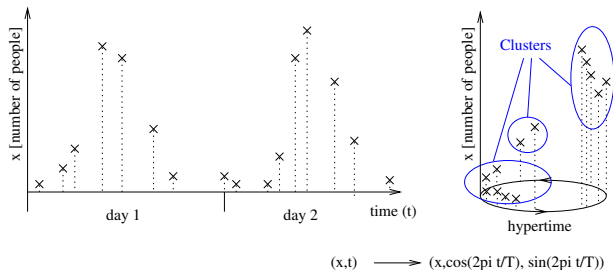
Fig. 1. Example of the warped hypertime projection of 1D data (number of humans in a given area). The numbers $\mathbf{x}_i$ observed at t_i are projected into a 3d vector space as $(\mathbf{x_i}, cos(2\pi t_i/T), sin(2\pi t_i/T))$, where they form clusters because they exhibit a periodic behaviour with a period T. Here, the warped hypertime dimensions define a base of a cylinder, while the numbers $\mathbf{x}_i$ define a cylinder side. Obtained from [11].

would be useful to allow the robot to choose the locations as well according to the typical areas of people presence. This requires a mobile robot to learn a spatio-temporal model that can predict the likelihood of people presence at different locations and times.

Since the planning, localisation and navigation components of most indoor mobile robots represent the robot operational space using an occupancy grid [4], most previous related work aimed to incorporate information about human-induced environment dynamics into occupancy grid models. For example, Kucner *et al.* [5] assumed that changes in occupancy are caused by moving objects and attempted to encode typical motion patterns into the grid. In particular, [5] proposes to associate each cell with a probabilistic model, which can predict the occupancy of the neighbouring cells depending on the direction an object passed through the given cell. A similar approach is proposed in [6], where the direction of traversal over each cell is obtained using an input-output hidden Markov model connected to the neighbouring cells. Another approach [7], [8], [9] associates each cell with a set of temporal models, which predict the direction of people movement at a particular time. The temporal models in [7], [8] are based on an approach presented in [10], which efficiently represents the periodic behaviour of changes caused by humans by employing spectral analysis. In [10], the model is applied to occupancy grids and can predict cell occupancies for a particular time. During the prediction phase, the system simply models the cell occupancies by a set of harmonic functions that capture long-term patterns of the changes observed.

The primary problem of grid-based models is not only

their memory inefficiency (because of the necessity to model a high number of discrete states separately), but also the necessity to obtain measurements at all of the modeled cells. In [12], the authors argue that continuous models are of better use for robot navigation than discrete ones and can be used to make short-term predictions of object movement [13]. They employ data provided by a range-finder sensor, which identifies not only the occurrence of humans and other objects but allows to determine which space is empty. Using this method the two categories of observations (empty and occupied) allow to employ a classification method with a sigmoid-based output. Later work [14], [15] speeds up model building by using an elegant combination of kernels and optimisation methods. The speed-up achieved by a combination of these methods allows to recalculate the model on a frequent basis, but the model itself does not represent the temporal domain. In [16], [17], [18], [19], [20], the authors incorporate the environment dynamics into the continuous model. [18] suggests using this continuous probabilistic directional map to plan robot paths through space in accordance with the directions imposed by the movement of human crowds. Unlike the aforementioned works, which are aimed primarily at modeling the spatial distribution of objects or predicting their motion to support robot path planning, we aim to build a model capable of long-term prediction of people presence, which should be able to support the scheduling of robot services.

Long-term predictions require analysis of time series, which are typically gathered over long time periods. In the time domain, the evolution of a given variable is usually modelled as a combination of trend, seasonal and cyclic patterns [21], [22]. While the trend represents the long-term decrease or increase, cyclic patterns are understood as rises or falls with no periodicity, and seasonality models the periodic changes. While some of the periodicities are obvious, as they are governed by daily and yearly rhythms, identification of the periodicities from the observed data allows to discover unexpected periodicities, which might significantly improve the predictive power of temporal models [23]. The seasonal patterns are often identified from the data by means of spectral analysis and modeled by Gaussian processes with periodic kernels [24] or Von Mises distributions [25]. In our case, we assume that from a long-term perspective the effect of any trend in people presence is not significant compared to factors caused by human routines. As the cyclic changes are not predictable in general, our method does not attempt to represent them explicitly.

Previous work on time series analysis for long-term robot navigation, which neglects trend and cyclic patterns with unknown periodicity, is presented in Krajnik *et al.* [10]. The authors propose to model the probability of environment states in the spectral domain, which naturally represents the "seasonality" of changes observed by the robot. In a series of papers, the authors demonstrated that application of the spectral-temporal models improves the efficiency of long-term mobile robot localisation [10], planning [26], exploration [27], etc. However, the spectral models of [10]

are designed to model the evolution of binary states over time, and thus are applicable only to individual components of discretised spatial models, e.g., occupancy grid cells [27], landmark visibility [10], traversability of certain areas [26] etc. In the aforementioned works, the spatial interdependence of the environment components is neglected, and the models considered only the temporal but not spatial-temporal relations of the represented environments. The necessity to model a high number of discrete states also results in memory inefficiency, which causes problems especially when representing large areas.

We propose a method that aims to capture periodicities in human presence in the robot's operational area and use them to anticipate human presence in the future. The method can identify the dependence of human presence on time by modeling the spatio-temporal distribution of their occurrences. The problem of finding such distributions is that while the modeled space is constrained, and thus one can gather an arbitrary number of measurements from the same spatial location, time itself unfolds indefinitely and one cannot get an arbitrary number of measurements from the same spatio-temporal area. The lack of the data itself makes estimation of the density of humans at a given time and location rather difficult. To deal with this problem, we search the measured data for periodic patterns, and if we find some, we project the timeline into a multi-dimensional space, curved in a way that reflects the period found. Thus, data from the same phase of a given period (e.g., the number of people occurrences from the same hour of day) are projected approximately into the same area of the curved space, see Fig. 1. After some time, the data points projected onto the same area of the curved space allow for an accurate estimation of the density of people at a given time and location. The principal advantage of the method presented is its continuous nature, which ensures its memory efficiency and scalability both in space and time.

Since there might be more periodicities present in the data, our approach can extend the continuous spatial representations by adding several dimensions representing different periodicities of time. In particular, we transform every time periodicity into two new dimensions that form a circle in its $2d$ subspace and use them to extend the representations that model the people occurrences in $2d$ space. We can understand this projection of time as follows: human behavior is much more similar during two different Sundays than on Sunday and Monday, although two Sundays are one week apart, while Sunday and Monday are next to each other. Similarly, people occurrences during two different mornings are expected to be more similar than the occurrences in the morning and evening. The circles also reflect time continuity – human behaviour five minutes before midnight is probably similar to behavior five minutes after midnight. From this point of view, modeling time as linear has meaning only if we would aim to study a long-term trend. Thus, the model of the proposed method is built over the space extended by *warped hypertime*, which efficiently represents both the structure of the space (given by the environment) and time (given by the human habits). A proof of this concept is given in [28].

II. METHOD DESCRIPTION

The aim of the method is to create a spatio-temporal model capable of predicting future human presence across the operational environment of the robot. As the model is created from data gathered by a real mobile robot over long time periods, it has to deal with the uncertainty of the measurements, occlusions, missing data, etc. The input data of the model are the positions of people provided by a state-of-the-art people detection method based on vision or active sensors, such as [29].

Let us assume that every time a robot measures the position of a detected person, it obtains a tuple $(\mathbf{x}_i, t_i)$, where $\mathbf{x}_i$ describes the position and t_i corresponds to the time when the measurement was taken. Thus, our method aims to find a function $\rho(\mathbf{x}, t)$, which characterizes the frequency of occurrence of the vector $\mathbf{x}$ over time. Contrary to the usual approach to the analysis of time series that iteratively decomposes particular patterns, our approach is to create a specific projection of tuples $(\mathbf{x}_i, t_i)$ into a new vector space, and then create a model of the frequency of occurrence in this space. (As said earlier, we assume that there are no long-term trends and neglect cyclic patterns due to the nature of the data.) To project time to the multidimensional vector space, warped hypertime, we need to identify dominant periodicities of the human presence in particular areas using tools related to spectral analysis. Moreover, to locate areas of frequent occurrence of data we use a clustering method. Thus, the presented method combines spectral analysis and clustering. The method has to determine the number and length of the periodicites from the non-uniformly sampled data. Doing so in one step would be susceptible to spectral aliasing [30] with the consequence of finding wrong periodicities. Thus, we choose an iterative approach that comprises of three steps:

1) clustering over spatio-temporal vector space (spatio-temporal clustering),
2) identification of periodicities and
3) spatio-temporal vector space extension.

During the first step, it builds a spatio-temporal model using a Gaussian Mixture Model. Next, it identifies the most prominent periodicity of the residuals between the model and measurements. In the last step, the vector space is extended by the addition of new dimensions and the aforementioned three steps are repeated using this extended space. If the model created over the extended vector space performs worse than the previous one, the method is terminated, and the previous model is returned.

A. Identification of Periodicities

For the sake of notation simplification, we will start the description of our method from the second step. To perform identification of periodicities, we calculate the model error over time as follows. At first, we partition the time line of the training set into K discrete intervals $(t_\kappa, t_{\kappa+1})$. Then, we predict the number of occurrences within these time intervals by integrating the values returned by our model over the entire domain of $\mathbf{x}$ and given time interval $(t_\kappa, t_{\kappa+1})$.

The obtained values form a histogram $\mathbf{m}$ representing the number of predicted occurrences for each time interval $(t_\kappa, t_{\kappa+1})$. Secondly, we count the number of training data $(\mathbf{x}, t)$ occurring during $(t_\kappa, t_{\kappa+1})$, obtaining a histogram $\mathbf{h}$. Then we calculate the differences of these histograms and create a new time series of residuals

$$R(t_k) = m_{(t_\kappa, t_{\kappa+1})} - h_{(t_\kappa, t_{\kappa+1})}, \tag{1}$$

where $t_k \in (t_\kappa, t_{\kappa+1})$, $\kappa = 0, \ldots, K-1$. The frequency spectrum of the time series $R(t)$ is then processed by the non-uniform Fourier transform method *FreMEn* described in [10]. Then we find the most prominent (i.e. with the largest absolute value) component γ_τ of the spectrum obtained from the previous step and calculate the corresponding period T_τ from its circular frequency ω_τ as $T_\tau = 2\pi/\omega_\tau$. The purpose of this step is to identify periodicities across the entire space regardless of their spatial distribution. The sensitivity of the *FreMEn* method is sufficient to detect periodicities even when they affect only a small area of the monitored space.

1) Short Overview of FreMEn: FreMEn [10] is a tool to introduce the notion of time into spatial models used in mobile robotics. The tool, which is available online [31], has been shown to improve the efficiency of autonomous robot operation in long-term deployments. For our purposes, we reimplemented the FreMEn to identify the periodicities in $R(t)$, which correspond to residuals of our model over time. In particular, with the known function ρ we create a histogram m and time series $R(t)$ and process it by our FreMEn implementation, which performs a non-uniform Fourier transform and identifies the dominant periodicities in the $R(t)$.

First, we choose a sufficiently large number representing the maximum length of considered periodicities T_1 and create a set of candidate frequencies $\{\omega_\tau\}_{\tau=1}^{\Upsilon}$, where $\omega_\tau = \tau/T_1$. A period T_τ corresponds to frequency ω_τ as $T_\tau = 1/\omega_\tau$ and T_Υ is the shortest periodicity taken into consideration. Then we calculate the components of the frequency spectrum as:

$$\gamma_\tau = \frac{1}{K} \sum_{k=1}^{K} R(t_k) e^{(-1j)t_k \omega_\tau 2\pi}. \tag{2}$$

Finally, we choose the period corresponding to the most prominent component of the spectrum, i.e., $T_\tau : \max\{|\gamma_\tau|\}$.

B. Spatio-temporal Vector Space Extension

Once we have identified the dominant period T_τ, we can extend the space with two extra dimensions. For each $(\mathbf{x}_i, t_i)$ of the original measured data in space-time, we calculate its projection into the space extended by warped hypertime as follows:

$$(\mathbf{x}_i, t_i) \rightarrow \left(\mathbf{x}_i, \cos \frac{2\pi t_i}{T_\tau}, \sin \frac{2\pi t_i}{T_\tau} \right), \tag{3}$$

where warped hypertime $\left(\cos \frac{2\pi t_i}{T_\tau}, \sin \frac{2\pi t_i}{T_\tau} \right)$ forms a circle in $2d$ plane which represents the periodicity and continuity of the occurrences. Thus occurrences with a similar position relative to periodicity T_τ are projected onto a similar position on the circle. During the iterative process we can extend

the extended vector space with another two dimensions and create a higher dimensional warped hypertime extension of the space. Such projection of the measurements $\mathbf{x}_i$ obtained at time t_i will be denoted $^{@}\mathbf{x}_i$, where:

$$^{@+1}\mathbf{x}_i = \left(^{@}\mathbf{x}_i, \cos\frac{2\pi t_i}{T_{\tau_{@+1}}}, \sin\frac{2\pi t_i}{T_{\tau_{@+1}}} \right), \tag{4}$$

and $^{0}\mathbf{x}_i = \mathbf{x}_i$, $^{1}\mathbf{x}_i = \left(\mathbf{x}_i, \cos\frac{2\pi t_i}{T_{\tau_1}}, \sin\frac{2\pi t_i}{T_{\tau_1}} \right)$, $^{2}\mathbf{x}_i = \left(\mathbf{x}_i, \cos\frac{2\pi t_i}{T_{\tau_1}}, \sin\frac{2\pi t_i}{T_{\tau_1}}, \cos\frac{2\pi t_i}{T_{\tau_2}}, \sin\frac{2\pi t_i}{T_{\tau_2}} \right)$, etc. The spatial component $\mathbf{x}_i$ of every extended vector space $^{@+n}\mathbf{x}_i$ is the same, and the additional dimensions are created purely from the time components t_i. Fig 1 taken from [11] illustrates the warped hypertime projection of a number of human occurrences within a given area.

C. Clustering over Spatio-temporal Vector Space

To form the spatio-temporal model, we use clustering to estimate the spatio-temporal density of the modeled phenomena. This model will allow us to estimate the frequency of occurrence of a given vector $\mathbf{x}_0$ at a given time t_0. In other words, the model establishes a function $\rho(\mathbf{x}_0, t_0)$, which indicates the frequency of occurrence of a given vector $\mathbf{x}_0$ at time t_0. Since we represent this function by several clusters in the "warped hypertime-space", the function $\rho(\mathbf{x}_0, t_0)$ is represented as a sum

$$\rho(\mathbf{x}_0, t_0) = \sum_{j=1}^{c} \alpha_j u_{0j}(\mathbf{c}_j, {}^{@}\mathbf{x}_0), \tag{5}$$

where c is a number of clusters with centers at $\mathbf{c}_j$, u_{0j} is a membership function of vector $^{@}\mathbf{x}_0$ to the jth cluster, and α_j is the cluster weight. Thus, for the purposes of the continuous model creation we need a clustering method that provides a membership degree $u_{0j}(\mathbf{c}_j, {}^{@}\mathbf{x}_0)$ as a function of distance between the cluster center $\mathbf{c}_j$ and projection $^{@}\mathbf{x}_0$ of $(\mathbf{x}_0, t_0)$.

At the end of this step, we also generate a discrete, multi-dimensional histogram $\mathbf{M}$ which approximates the spatio-temporal distribution of $(\mathbf{x}_i, t_i)$ over time-space and compare it to the histogram $\mathbf{H}$, generated from the actual measured data, using root-mean-square deviation [32]:

$$RMSD_{current} = \sqrt{\sum_{bins} (\mathbf{M}_{bin} - \mathbf{H}_{bin})^2}, \tag{6}$$

with *bin* defined according to procedure in section II-C.2. Then, we proceed with the next steps. At this phase of the next iteration, we calculate $RMSD_{new}$ according to (6). If this error $RMSD_{new}$ is lower than the error $RMSD_{current}$ before extending the space by (4), we proceed with the next iteration. Otherwise, we revert to the previous model and terminate the algorithm.

Note that difference between histograms $\mathbf{m}$ and $\mathbf{M}$ (or $\mathbf{h}$ and $\mathbf{H}$) lies in their dimensionality. While $\mathbf{M}$ and $\mathbf{H}$ divide the whole domain of vectors $(\mathbf{x}, t)$, histograms $\mathbf{m}$ and $\mathbf{h}$ divide only the time domain.

1) Clustering method: Initially, we tried to model the spatio-temporal phenomena by a mix of Gaussian distributions, determined by an Expectation-Maximisation method similar to [14], [18]. This method returns both necessary parameters for our approach: membership functions u_j and cluster weights α_j. However, the results obtained by the multivariate Gaussian mixtures models were unsatisfactory, because the cluster weights are calculated for $^{@}\mathbf{x}$ and not for $(\mathbf{x}, t)$. The difference lies in the domain of these projections. For example, let us consider the projection of $\{t_i\}$ onto the circle that corresponds to a period of one week. Moreover, let us consider data where some of the days (e.g., half of the Fridays) are missing. This will decrease the weights of the clusters corresponding to the Friday occurrences in $^{@}\mathbf{x}$ compared to the case where measurements from all Fridays in $(\mathbf{x}, t)$ are available. Therefore it is necessary to calculate cluster weights α_j in the space-time $(\mathbf{x}, t)$.

2) Cluster Weights: Having obtained the positions of the clusters $\mathbf{c}_j$, their shape as covariance matrices Σ_j and the membership u_{ij} of every vector $^{@}\mathbf{x}_i$ to every cluster by GMM, we need to determine the cluster weights α_j. These weights correspond to the ratio between the number of vectors belonging to the cluster, i.e., $m_j = \sum_{i=1}^{n} u_{ij}$, and its volume v_j in the projection $(\mathbf{x}, t)$. Then we determine the cluster weight as its density

$$\alpha_j = m_j / v_j. \tag{7}$$

To estimate the volume of the j-th cluster v_j we need to define the domain of the function $\rho(\mathbf{x}, t)$ and basic volume unit b. The domain of $(\mathbf{x}, t)$ can be either defined during measurement or it can be estimated as an union of reasonably large neighborhoods of measured vectors. The basic volume of space-time has to be chosen according to the purposes of the model (for example, squared meter hour). Then we can create histograms (namely $\mathbf{M}$, $\mathbf{H}$) over this domain with bins of volume b. For the purposes of cluster volume estimation, we also need a list of space-time vectors $\mathbf{v}_\beta$ that define the position of every bin, for example, the centres of bins. By projecting these centres to the chosen warped hypertime-space and using the chosen model, we obtain a membership value $u_{\beta j}$ of every $^{@}\mathbf{v}_\beta$ to every cluster. Then we can calculate the volumes of clusters as follows:

$$v_j = \sum_{\forall \beta} u_{\beta j}. \tag{8}$$

It should be noted that the basic volume b directly influences the values predicted by the model. During estimation of the number of occurrences at some position $(\mathbf{x}_0, t_0)$, the method returns the number of occurrences in the neighborhood of the vector $(\mathbf{x}_0, t_0)$ with volume b.

3) Poisson Distribution Based Initialization: Different random initializations that are usually understood as default for clustering algorithms led to poor performance of the tested methods, especially with a higher dimension of the warped hyperspace-time. The Poisson distribution based

initialization proposed in [33] notably improved the computational stability of clustering over the space-hypertime. According to the instructions, we chose λc vectors from $^{@}\mathbf{X}$ and eliminated $(\lambda - 1)c$ redundant vectors using Algorithm 1.

Algorithm 1: Elimination of redundant vectors.

Input: λc randomly selected vectors from a set of vectors

Output: c vectors, each representing one cluster

for $i = 1$ **to** $(\lambda - 1)c$ **do**

  Calculate the set of distances between every pair of selected vectors $\{d_{ij}\}$;
  Find $min\{d_{ij}\} = d_{kl} = d_{lk}; i \neq j$;
  Find $min\{\sum d_{il}; \sum d_{ik}\} = \sum d_{im}$;
  Eliminate vector m;

III. EVALUATION

To evaluate our method, we utilised a dataset of people presence collected at the University of Lincoln. The data were collected by a mobile robot equipped with a Velodyne 3D laser rangefinder. The robot was driven to a location which provided a good overview of one of the T-shaped corridor junctions (Fig. 2). To localize people in the measurements provided by the laser 3D scanner, we used a reliable and efficient person detection and localisation method [29]. Since the robot batteries need to be recharged on a daily basis, we could not collect the data in a continuous, 24/7 manner, but had to remove the robot from the observation spot every night when the building was vacant, and there were no people at the corridors. The collected dataset contains detections from early mornings to late evening from weekdays over several weeks. A typical day contains approximately 32000 people detection measurements, which correspond to a large number of people walking or standing in the monitored corridors. The method [29] provides human detection result as a single vector, e.g., a point in space. However, for the purpose of mobile robot path planning, a human represents an obstacle with a particular spatial volume. For this reason, we preprocessed the dataset by substituting every detection vector by a set of vectors in its spatial neighborhood with a diameter of 0.5m.

The primary purpose of the evaluation is to estimate the predictive capability of the models created by the proposed method. Therefore we split the gathered data into training and test sets and learn the model from the training set only. The training dataset consists of two weeks of measurements and the test dataset consists of two days of measurements from another week (Wednesday and Thursday). Then, we partition the timeline of the test data into a spatio-temporal 3D grid $\mathbf{G}$ and create a 3D histogram $\mathbf{H}$. In particular, for each bin, we count the number of detections in the test data. Then, we predict a histogram $\mathbf{M}$ for the same grid $\mathbf{G}$ using models created from the training data. The overall error of the method is calculated as a root-mean-square deviation
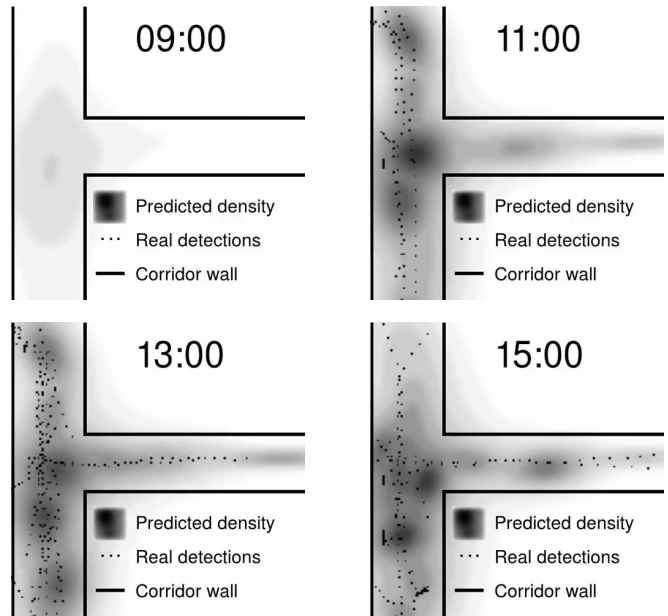

Fig. 2. Model compared to the data. Black dots represent detections from the test dataset and grey areas represent the model (darker area means higher number of predicted occurrences). Black lines represent walls of corridors.

between histograms $\mathbf{H}$ and $\mathbf{M}$, similar to equation (6). For our evaluation, we calculated this error for different spatial and temporal resolutions of the grid $\mathbf{G}$,, i.e., for the different basic volumes b, see Fig. 3.

To evaluate our method, we compare it to three other spatio-temporal models. These three models represent the space in a discrete way. In particular, they associate each cell of the spatial 2d grid with a temporal model: *Mean*, which is simply an average of the all past measurements occurring within a given cell, *Hist*, which splits each day into h intervals and predicts the number of occurrences as an average for the relevant time of a day, and *FreMEn*, which extracts f spectral components from the people occurrence history and uses these periodic components for future predictions. To predict the number of occurrences at a particular location at a particular time, we simply select the grid cell that corresponds to the location and use the cell's temporal model to perform the prediction. In other words, to generate the histogram $\mathbf{M}$, we calculate a separate temporal model for each cell of the grid $\mathbf{G}$ and then predict the values of the cells in $\mathbf{M}$ using these models. On the contrary, our method represents the space in a continuous way and to predict the histogram $\mathbf{M}$, it simply calculates the cell value according to (5). Regardless of the way $\mathbf{M}$ is predicted, the error of the method is calculated as the root-mean-square deviation between the histogram of measurements $\mathbf{H}$ and histogram $\mathbf{M}$ (6). The experimental evaluation is performed by an automated system similar to the one presented in [34]. This system first optimizes each method's (*Mean*, *Hist*, *FreMEn*) parameters and then runs pairwise t-tests to determine which of the compared methods perform statistically significantly better than other ones.
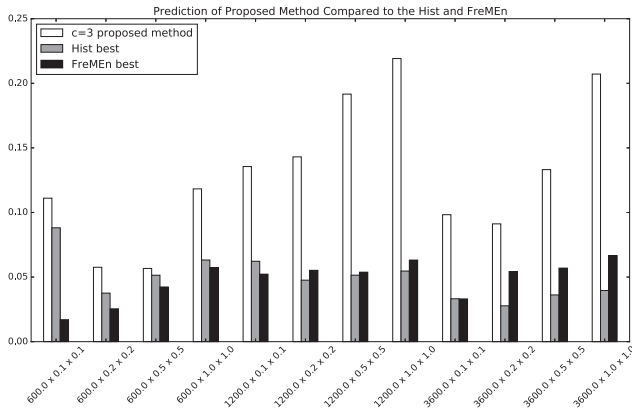
Fig. 3. Comparison of the predictive power of different methods. The graph shows the reduction of the prediction error compared to the *Mean* model, which neglects the temporal properties of the people presence. The indicated values (*y*-axes) are calculated using (9). On the *x*-axis, there are cell sizes or basic volumes. The white bar shows the prediction power of the proposed method using three clusters, grey bar Hist and black bar FreMEn.

resolution [cm x cm]	proposed method c = 3, @ = 2	proposed method c = 3, @ = 3	FreMEn f = 5	Hist h = 24	Mean
100 x 100	1.7 KiB	2.4 KiB	1.1 MiB	19.3 KiB	3.3 KiB
50 x 50	1.7 KiB	2.4 KiB	4.4 MiB	307.3 KiB	12.9 KiB
20 x 20	1.7 KiB	2.4 KiB	27.2 MiB	1.9 MiB	80.1 KiB
10 x 10	1.7 KiB	2.4 KiB	108.9 MiB	7.7 MiB	320.1 KiB

method is iterative. While FreMEn calculates a temporal model of all the grid cells in less than a minute, our method built the spatio-temporal model in an hour (CPU Intel Core i7-5005U). Moreover, during the model buiding, the method has to discretise the timespace to calculate the histograms **H** and **M**. These occupy memory space comparable to the discrete methods *Mean*, *Hist* and *FreMEn*. However, once model building is finished, these histograms are not needed any more, and the model consists only of the cluster parameters and periodicities.

IV. CONCLUSION

We proposed a novel approach to anticipate human presence for robots that are required to provide people-oriented services. The method uses projections of the time to several new dimensions derived from the natural human behavior. A projection of people detections into the created warped hypertime-space is analyzed using Gaussian Mixture Models. The obtained clusters are used to create spatio-temporal model of the human presence distribution. This model is then used to predict the occurrences in a given time and space.

We evaluated the proposed method on a real dataset from the corridors at a university collected over several weeks. Then, we compared the accuracy of the prediction to state-of-the-art methods used in robotics, and showed that our approach outperforms them in terms of prediction accuracy and memory efficiency. Although the model building is slower and computationally more demanding compared to other tested methods, the model is usually re-built once a day during the robot recharging which usually occurs overnight [1]. The predictive power of the model allows mobile robots to anticipate human presence across its operational area, which improves the future efficiency of robot operation. The efficiency and predictive power of the method allows to create and maintain dynamic models of large human-populated areas and make long-term predictions of human presence.

The performance of the method depends largely on the number of clusters chosen for representation. Our experiment indicates that an appropriate number of clusters can be inferred from the topology of the environment, which is one of the future extensions of this method. Furthermore, we will also evaluate the proposed method on larger datasets, such as [35]. Dataset collection is also in progress at University of Lincoln and at University of Technology of Belfort-Montbéliard.

As the corridor is T-shaped, we decided to choose $c = 3$ clusters for the proposed method. Since the error of prediction is dependent on the partitioning used, we tested the method for grids of various cell sizes (basic volumes b) ranging from 0.1 to 1.0 meters and 10 to 60 minutes. To better visualize the comparison of prediction capabilities of different methods on different resolutions (Fig. 3), we show the ratio between the RMSD of prediction of every compared method to the RMSD of prediction gathered from the *Mean*, i.e.,

$$ratio = 1 - RMSD_{method}/RMSD_{Mean}. \qquad (9)$$

It can be seen in Fig. 3 that the proposed method significantly outperforms the other ones. In most cases it is more than 5% better than *FreMEn* and *Hist*. Fig. 2 also demonstrates that the predicted distributions of the people depend on time and follow the shape of the corridor. Moreover, the proposed model is much more memory efficient compared to FreMEn or the other models. This is because our model is continuous and its complexity is not affected by the number of cells in the underlying grid, consisting of the positions of clusters $\{\mathbf{c}_j\}$, their weights $\{\alpha_j\}$, covariance matrices $\{\Sigma_j\}$, and a set of used periodicities $\{T_\tau\}$. For the two periods and three clusters, it occupies 1.7 KiB of memory space independently to the grid resolution. In contrast, FreMEn needs several numbers for each modelled periodicity of every cell of the grid-based representation. Thus, on our dataset, where the aforementioned grid has a resolution of 100, 50, 20, and 10 cm, the FreMEn model occupies 1.1, 4.4, 27.2, and 108.9 MiB of memory respectively, see Table I. In addition, the resolution of the proposed method is derived from the cluster weights α_j. Therefore it is possible to store different resolutions of one model at almost no memory cost.

The downside of the proposed method is its computational complexity, which is caused primarily by the fact that the

REFERENCES

[1] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mu-drová, J. Young, J. Wyatt, D. Hebesberger, T. Kortner *et al.*, "The

STRANDS project: Long-term autonomy in everyday environments," *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.

[2] D. V. Hebesberger, C. Dondrup, C. Gisinger, and M. Hanheide, "Patterns of use: how older adults with progressed dementia interact with a robot," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 131–132.

[3] M. Hanheide, D. Hebesberger, and T. Krajník, "The when, where, and how: An adaptive robotic info-terminal for care home residents," in *ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 341–349.

[4] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, no. 6, pp. 46–57, 1989.

[5] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, "Conditional transition maps: Learning motion patterns in dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 1196–1201.

[6] Z. Wang, R. Ambrus, P. Jensfelt, and J. Folkesson, "Modeling motion patterns of dynamic objects by iohmm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*. IEEE, 2014, pp. 1832–1838.

[7] S. Molina Mellado, G. Cielniak, T. Krajnik, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Towards Autonomous Robotic Systems (TAROS)*, 2018.

[8] 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.

[9] 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.

[10] 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*, 2017.

[11] T. Krajník, "One year of autonomy in everyday environments: The strands project," in *ICRA 2018 Workshop on Long-term autonomy and deployment of intelligent robots in the real-world*, 2018, keynote speech.

[12] 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.

[13] R. Senanayake, L. Ott, S. O'Callaghan, and F. T. Ramos, "Spatio-temporal hilbert maps for continuous occupancy representation in dynamic environments," in *Advances in Neural Information Processing Systems*, 2016, pp. 3925–3933.

[14] F. Ramos and L. Ott, "Hilbert maps: scalable continuous occupancy mapping with stochastic gradient descent," *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1717–1730, 2016.

[15] R. Senanayake and F. Ramos, "Bayesian hilbert maps for dynamic continuous occupancy mapping," in *Conference on Robot Learning*, 2017, pp. 458–471.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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)*, 2018, pp. 3241–3248.

[21] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[22] C. Heylman, Y. G. Kim, and J. Wang, "Forecasting energy trends and peak usage at the university of virginia," in *Systems and Information Engineering Design Symposium (SIEDS), 2015*. IEEE, 2015, pp. 362–368.

[23] V. H. Bennetts, K. Kamarudin, T. Wiedemann, T. P. Kucner, Somisetty, S. Lokesh, and A. J. Lilienthal, "Multi-domain airflow modelling and ventilation characterization using mobile robots, stationary sensors and machine learning," *Sensors*, 2019, to appear.

[24] A. Tompkins and F. Ramos, "Fourier feature approximations for periodic kernels in time-series modelling," in *AAAI Conference on Artificial Intelligence*, 2018.

[25] E. Chinellato, K. V. Mardia, D. C. Hogg, and A. G. Cohn, "An incremental von Mises mixture framework for modelling human activity streaming data," in *Proceedings ITISE 2017*. Leeds, 2017.

[26] J. Pulido 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 *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[27] J. M. Santos, T. Krajnik, J. Pulido Fentanes, and T. Duckett, "Lifelong information-driven exploration to complete and refine 4d spatio-temporal maps," *Robotics and Automation Letters*, 2016.

[28] 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.

[29] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 864–871.

[30] M. W. Maciejewski, H. Z. Qui, I. Rujan, M. Mobli, and J. C. Hoch, "Nonuniform sampling and spectral aliasing," *Journal of Magnetic Resonance*, vol. 199, no. 1, pp. 88–93, 2009.

[31] T. Krajník, "FreMEn: Frequency Map Enhancement for Long-Term Autonomy of Mobile Robots," Open-source implementation: http://fremen.uk, 2018.

[32] 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.

[33] T. Vintr, V. Vintrova, and H. Rezankova, "Poisson distribution based initialization for fuzzy clustering," *Neural Network World*, vol. 22, no. 2, p. 139, 2012.

[34] T. Krajník, M. Hanheide, T. Vintr, K. Kusumam, and T. Duckett, "Towards automated benchmarking of robotic experiments," in *ICRA Workshop on Reproducible Research in Robotics*, 2017.

[35] D. Bršcic, 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.